UNIVERSITY OF BURGUANDY
SPIM doctoral school

PhD from the University of Burgundy in

Computer Science

Presented by:

Hamdi Bouchech

Defense Date: January 26, 2015

Selection of optimal narrowband multispectral images for face recognition

Thesis supervisor:
Dr. Sebti Foufou

Jury:

Frederic Morain-Nicolier, Professeur a I'IUT de Troyes, Rapporteur.
Pierre BONTON, Professeur à l' Université Blaise Pascal, retraité , Rapporteur.
Saida Bouakaz, Professeur à l' Université Claude Bernard Lyon 1, Examinatrice.
Pierre Gouton, Professeur à l' Université de Bourgogne, Examinateur.
Yassine Ruichek, Professeur à l' Université de Technologie de Belfort-Montbéliard, Examinateur.
Sebti Foufou, Professeur à l' Université de Bourgogne, directeur de thèse.

I would like to dedicate this thesis to my loving parents ...

Acknowledgements

I would like to express my special appreciation and thanks to my advisor Dr Sebti Foufou, you have been a tremendous mentor for me. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. Your advice on both research as well as on my career have been priceless.

I would also like to thank my committee members, Dr Yassine Ruichek, Dr Frederic Morain-Nicolier, Dr Pierre BONTON, Dr Saida Bouakaz and Dr Pierre Gouton for serving as my committee members even at hardship. I also want to thank you for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions, thanks to you.

I would especially like to thank Dr Mongi Abidi, Dr Andreas Koschan and Dr Farid Touati for enlightening me the first glance of research by their stimulating discussions and priceless time and materials they have afforded to me during the last three years.

A special thanks to my family. Words cannot express how grateful I am to my mother and father for all of the sacrifices that they have made on my behalf. Your prayer for me was what sustained me thus far.

At the end, i would also like to thank all of my friends who supported me in writing, and incented me to strive towards my goal.

Abstract

Face recognition systems based on 'conventional' images have reached a significant level of maturity with some practical successes. However, their performance may degrade under poor and/or changing illumination. Multispectral imagery represents a viable alternative to conventional imaging in the search for a robust and practical identification system. Multispectral imaging (MI) can be defined as a 'collection of several monochrome images of the same scene, each of them taken with additional receptors sensitive to other frequencies of the visible light or to frequencies beyond the visible light like the infrared region of electromagnetic continuum. Each image is referred to as a band or a channel'[25]. However, one weakness of MI is that they may significantly increase the system processing time because of the huge quantity of data to be mined; in some cases, hundreds of MI are taken for each subject.

In this thesis, we propose to solve this problem by developing new approaches to select the set of best visible spectral bands for face matching. For this purpose, the problem of best spectral bands selection is formulated as an optimization problem where spectral bands are constrained to maximize the recognition accuracy under challenging imaging conditions. We reduce the redundancy of both spectral and spatial information without losing valuable details needed for the object recognition, discrimination and classification. We have investigated several mathematic and optimization tools widely used in the field of image processing.

One of the approaches we have proposed formulated the problem of best spectral bands selection as a pursuit problem where weights of importance were affected to each spectral band and the vector of all weights was constrained to be sparse with most of its elements are zeros. In another work, we have assigned to each spectral band a linear discriminant analysis (LDA) based weak classifier. Then, all weak classifiers were boosted together using an Adaboost process. From this later, each weak classifier obtained a weight that characterizes its importance and hence the quality of the corresponding spectral band. Several other techniques were also used for best spectral bands selection including but not limited to mixture

of Gaussian based modeling, multilinear sparse decomposition, image quality factors, local descriptors like SURF and HGPP, likelihood ratio and so on. These different techniques enabled to build systems for best spectral bands selection that are either static with the same bands are selected for all the subjects or dynamic with each new subject get its own set of best bands. This later category, dynamic systems, is an original component of our work that, to the best of our knowledge, has not been proposed before; all existing systems are only static. Finally, the proposed algorithms were compared to state-of-the-art algorithms developed for face recognition purposes in general and specifically for best spectral bands selection.

Abstract

Les performances des systèmes de reconnaissance des visages en utilisant des images RGB baissent rapidement quand ils sont appliqués dans des conditions d'illumination extrêmes. L'utilisation des images multispectrales représente une alternative prometteuse pour résoudre ce problème. Dans cette thèse on s'intéresse à l'utilisation des images multispectrales visibles pour la reconnaissance des visages humains. Les images multispectrales visibles sont des images capturées à des longueurs d'ondes différentes du spectre visible (band spectral) qui s'étend de 480nm à 720nm. Ces images représentent des caractéristiques qui favorisent la reconnaissance des visages humains dans des conditions particulières comme la présence d'excès d'illumination incidente sur le visage photographié. Notre travail consiste à exploiter ces caractéristiques sur des stages différentes: optimiser le choix du nombre de bandes spectrales à utiliser, optimiser les longueurs d'ondes choisies, optimiser les techniques de fusion des informations extraites à partir des différentes bandes spectrales pour avoir plus d'informations utiles et moins d'informations bruits. Plusieurs nouvelles approches ont été proposées dans le cadre de ce travail avec des résultats encourageants en terme de performances. Ces approches ont exploité plusieurs outils mathématiques pour resoudre les différents problèmes rencontrés, en particulier la formulation de la sélection des bandes spectrales optimales sous formes de problèmes d'optimisation où nous avons utilisé le « basis pursuit algorithm » pour déterminer un vecteur de poids sparse pour représenter l'importance des différentes bandes. Dans d'autres problèmes d'optimisation, nous avons attribué à chaque bande un classifieur faible, puis combiné les classifieurs faibles avec différents poids associés selon l'importance. La méthode Adaboost a été utilisée pour trouver la combinaison optimale. D'autre techniques ont introduites d'une manière originale la décomposition multilinéaire des images de visage pour formuler une sorte de base de données caractérisant les bandes spectrales. Cette base de données a été utilisée avec les nouvelles images, ou image test, pour déterminer les bandes les plus robustes contre une variation importante d'illumination. Le travail présenté dans le cadre de cette thèse est une petite contribution à la reconnaissance des visages en utilisant des images multispectrales, qui est une approche d'actualité, mais qui nécessite encore plus de développement afin de max-

imiser ses performances.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

The demand for very accurate authentication and personal verification systems has grown rapidly in order to satisfy new critical applications such as access control to restricted areas and facilities, authentication for secure banking, law enforcement, video surveillance and automatic screening at airports [52] [119] [120] [168] [115] [84] [37] [77] [83] [127]. Existing personal authentication systems can be largely divided into two main groups: (1) Systems that authenticate personal documents (ID cards, passports, access passwords, etc.) or signatures and hand writing specificities (characteristic letter shape and size or pressure on the paper, etc.) and (2) Systems based on biometric properties of individuals (e.g., fingerprints, retina and iris patterns, hand internal and external geometries, voice, face, etc.). Currently, personal authentication systems within the first group are used far more in practice than those of the second group. The former are trusted by users due to their ease of use and reduced costs. In addition, an ID card is easily portable and can be attributed to a subject within a few minutes using small unsophisticated devices. Signatures, ID cards, or passports can also be verified manually by the specialized authority. This aspect contributes to the level of trust enjoyed by the Group One systems, as it is human nature to trust what one has personally verified. However, two serious problems arise from the first category of authentication systems:

- In applications such as securing areas with valuable assets or controlling high-traffic borders, ensuring accurate personal authentication via human action increases the risk of error and limited objectivity.

- Counterfeit ID cards, hacked passwords, or even multiple people with similar natural

signatures are scenarios frequently encountered in real life.

Given these problems, as well as the continuous appearance of new fraud strategies, it can be considered that the authentication systems in Group One are less reliable than is desired [5], [107]. On the other hand, biometrics-based systems present potential opportunities to overcome the aforementioned two issues. Biometrics in general refers to a person's measureable physiological and behavioral characteristics. The development and use of reliable biometric systems is receiving wide attention from public and private entities. Several different types of biometrics have been explored [73]. To date, iris, fingerprint, and face-based biometric measurements have shown to be the most promising [18], [91], [102].

Biometric authentication focuses on who a person is, rather than what the person has, thereby giving biometric systems the advantage of greater sensitivity to any attempt to fraud. Since the beginning of time, humans have used foot prints, smell or even sounds for tracking animals or persons. In its modern form, biometric authentication dates roughly to the late 19th century when fingerprints were used to identify criminals. Although they are the most commonly used biometrics, iris, fingerprint and face are not the only biometric options in the literature. Jain et al. identified seven factors to be used when assessing the suitability of any trait for use in biometric authentication [71]: (1)Universality, such that every person considered by the system has the defined traits. (2) Uniqueness, such that two different persons should have sufficiently different properties of the defined trait. (3) Permanence, or how the trait varies over time, which is critical for system stability. (4) Measurability, such that the trait is easily quantifiable. (5) Performance, which relates to the accuracy, speed, and robustness of the technology used. (6) Acceptability, which pertains to the degree of tolerability to those undergoing identification. (7) Circumvention relates to the ease with with identification may be avoided by using an artifact or a substitute.

Based on the targeted application, biometrics like fingerprint, iris or face may have different trade-offs in performance, advantages and disadvantages. This may be accentuated by the appearance of additional constraints. For example, subjects to be identified are not always cooperative, and safety considerations may dictate that unauthorized persons who are forbidden to access critical areas should be identified from a distance. For instance, in the case of border passage, the speed of a given biometric approach may be an important consideration; the identification process should be fast enough to avoid creation of long lines. The safety and privacy of the persons involved is also a major concern. Finally, the chosen biometric should ensure stable performance with respect to working conditions as well as building accurate, cheap, portable and easy to set up systems.

The above mentioned constraints, and many others depending on the considered application, have promoted faces as one of the most suitable biometrics for many current applications. Authentication using face images has attracted significant research attention [8], [47], [133], [147], [149], [150], [151], [165]. There are clear advantages to the use of face images; for example, faces can be captured unobtrusively and from a long distance [160], [161], such that persons do not realize that they are surveilled and no contact with a device is required. Also, the existing face-based authentication systems can be accurate, cheap, portable (laptop + camera), easy to set up, and fast enough to allow real-time applications. Given these points, face-based recognition systems are a good fit with respect to the previously mentioned constraints.

While these are significant advantages, a review of the literature of face recognition systems should also reveal the existance of limitations. Face appearance may vary with factors like viewing angle, illumination, or occlusion that may introduce a significant difference between face images captured for the same person and by the same system and thus further complicates the task of face recognition (see [56] for further factors). Most of these factor are totally or partially related to lighting conditions, i.e., lighting direction (frontal, right, left, etc.) and light nature (fluorescent, halogen, daylight, etc.) and thus, their effects vary based on the spectral band in which face images were captured; researchers have proposed to overcome some of these weaknesses in face image based systems by using multispectral images (MI) and multispectral imaging systems (MIS). Multispectral imaging can be defined as a "collection of several monochrome images of the same scene, each of them taken with additional receptors sensitive to other frequencies of the visible light or to frequencies beyond the visible light like the infrared region of electromagnetic continuum. Each image is referred to as a band or a channel" [82]. This definition can be extended to emphasize the objective of capturing images of faces in additional bands, beyond the classic visible spectrum. Additional receptors are added to preserve as many invariant facial features regarding imaging conditions as possible, and to explore spectral band complementarity.

1.1 Motivation

Multispectral imaging has shown a strong capability to cope with the challenges posed by imaging conditions and particularly high illumination variation [23], [24], [50], [90]. Chang showed that MIS strength comes from the intuitive fact that with more spectral bands, more facial information could be obtained [22]. However, one weakness of MI is that they may significantly increase the system processing time because of the huge quantity of data to be

mined; in some cases, hundreds of MI are taken for each subject. To solve this problem, a system for best (optimal) spectral bands selection (abbreviated hereafter as BSBS) must be established to satisfy real time applications in face recognition. The objective of this system is to select a set of spectral bands that maximizes the system overall recognition performance in a given imaging conditions. Very few are research works that proposed such systems for best spectral bands selection. The limited number of publications on this topic may be due to the fact that an imaging system capable of capturing hundreds of spectral bands, as in [125], is much more complex and costly than a system suitable for capturing only two or three bands. In addition, the question of the effect of involving a large number of bands on the system performance arises, as well as whether the addition of components to the recognition system, such as more complex image registration algorithms and an algorithm for band selection, will negatively impact the feasibility of building real-time systems.

1.2 Objectives

From above, we can see that there are three main factors that disrupted the optimal investigation of multispctral images for face recognition. These factors are:

- Building systems to capture hundreds of spectral bands at different wavelenghts is much more complex and costly than a usual imaging system.

- Using multispectral images requires the optimization of systems for best bands selection to overcome the problems of high data dimensionality. However, This task is far from being easily reachable due to the multitude imaging conditions that have to be respected.

- The previous two factors have limited the interest of researcher in using MI and hence caused the quasi- absence of experiences were MI were successfully applied. Hence, the effectiveness of using MI is still not approved.

In this work, we aim to solve the aforementioned weaknesses encountering the use of MI for face matching. We develop a complete set of tools including systems for best spectral bands selection and intensive experimentations involving state-of-the-art algorithms. The different experiments conducted within the scope of this work provide an enough-trusted ground truth that reflects the advantages and drawbacks of using MI for face matching. The efficiency of using systems for best spectral bands selection is emphasized and several approaches have been tested. This later task occupied the largest part of this thesis.

1.3 Thesis Organization

The rest of this thesis is organized as fellow:

- **Chapter 2**. In this chapter we review the most recent works that have investigated multispectral images for face recognition. We review the concepts of MI preprocessing , matching face images from different spectrums (heterogenous face matching) and matching image faces from the same spectrum (homogenous face matching) are reviewed. Finally, existing systems for band selection as well as the most used databases providing multispectral images are also presented.

- **Chapter 3**. In this chapter, we describe the category of static BSBS systems. These systems select the same set of best spectral bands for all subjects presented in the database. For a new subject (test set), the selected bands are then used for face matching. In this chapter, three different static BSBS approaches are proposed. The first system investigates the sparsity concept to optimize a vector of weights W. The elements of W are weights assigned to each spectral band to quantify its suitability to a given imaging condition. W was constrained to be sparse with most of its elements are zeros. The few spectral bands with non zero weights were then selected as best spectral bands. The second approach applied multilinear decomposition on cubic images T (multispectral images). T was decomposed using a 3-mode singular value decomposition (SVD) and three mode matrices for subjects, spectral bands and illuminations were sparsely determined. The spectral bands mode matrix define a sparse vector for each spectral band. Spectral bands having the sparse vectors with the lowest variation with illumination were then selected as the best spectral bands. The final proposed approach attributed a weak classifier to each spectral band. These weak classifiers were built based on the linear discriminant analysis (LDA) tool. All weak classifiers were then boosted (combined) using the SAMME adaboost algorithm. The final stronger classifier assigned a weight to each weak classifier. This weight determined the importance of the corresponding weak classifier and in turn the importance of the corresponding spectral band.

- **Chapter 4**. In this chapter, the second category of BSBS systems, namely dynamic BSBS systems is studied. We propose two approaches. One approach learns a filter adequate to process multispectral images. The filter was adapted to the well known feature extraction tool SURF (Speeded Up Robust Features)) and it has the role to ensure a better selection of best spectral bands invariant to illumination condition. The second approach proposed was a dynamic system for BSBS. This category of systems selects a different set of best spectral bands for each new subject. This way, the set of selected bands adapts with

the variation of imaging condition. We can expect that this category of BSBS systems will have a better performance but a much higher processing time. Our dynamic system is based on several optimization techniques including SURF, Mixture Gaussian analysis (MG), Likelihood ratio and transfer learning.

- **Chapter 5**. In this chapter, we present and analyse the different experimental results of our approaches for BSBS. CMC curves are drawn to show the evolution of recognition rates achieved by each algorithm. The different features extraction algorithms and used databases are described and referenced.

- **Conclusion**. We conclude our thesis with this chapter were we summarize the different approaches proposed to solve the problems highlighted in section on Objectives. We briefly describe each proposed approach, its advantages and weaknesses and the achieved results. Finally, we introduce our future plan to build better systems that overcome the revealed weaknesses.

Chapter 2

Literature Review

2.1 Multispectral Images: Acquisition And Preprocessing

2.1.1 Multispectral Imaging System

Modern spectral image acquisition systems typically rely on combinations of CCD cameras with various types of narrow or broad band filters. The images are then processed using conventional high-capacity computational machinery with software developed to properly treat the spectral data. Ohta et al. used a film-based system for multispectral image capture. They used a mechanical filter wheel with eight gelatin filters to image rigid objects. Tominaga used a six color camera system with six color filters, which had the six spectral channels of the color filters fixed wavelength bands [148]. The VASARI imaging system developed at the National Gallery in London employed a seven-channel multispectral camera to capture paintings [128]. However, these fixed filter systems suffer from three main restrictions [82]. First, the selection of color filters with regard to their central wavelength and their bandwith is limited. Second, filters with a narrow bandpass are difficult to manufacture. Last but not least, misalignments between image bands may occur due to mechanical vibrations of the imaging system when the filter wheel is spinning. To overcome these issues, electronically tunable filters can be used where band selection can be controlled simply by applying voltage, acoustic signals, etc. Further reading about attributes considered in tunable filter conception can be found in [51].

The most commonly used tunable filter is the liquid crystal tunable filter (LCTF). This filter selects its bands by electronically controlling liquid crystal elements to choose a specific wavelength while rejecting all others. It is built from a stack of fixed filters consisting of an interwoven birefringent crystal-liquid crystal combination and linear polarizer. The latter property makes the LCTF sensitive to polarization. Another main property of the LCTF, that should be taken into consideration during system design, is the crystal relaxation time; this is the minimum time for this type of filter to change between two successive bands and is typically 50 ms. With special devices this time can be reduced to 5 ms. Last but not least, the LCTF are characterized by their bandpass, also called spectral resolution, which is in the order of nm.

Several research groups [69] [116] have used LCTFs in conjunction with monochrome cameras to capture multispectral images for different applications. Figure 2.1 shows four bands of a sample face image from the $IRIS-M^3$ database. Note that the images are not gamma corrected for display and as a result appear dark. It is an advantage that LCTFs have no moving parts, and hence problems like misalignments between image bands due to mechanical vibrations are reduced; in addition, LCTFs are very lightweight, which makes them suitable for mobile sensor platforms.

2.1.2 Multispectral Image Preprocessing

The aim of this section is to emphasize the importance of preprocessing steps, in a multispectral context, that generally precede any image-based recognition task. Multispectral imaging has emerged as a potential solution for applications such as long-range recognition in low-light (night) conditions [113], [75], [54], [156], detection of subjects wearing a disguise to alter facial characteristics (fake nose, makeup, or wig) and/or surgical alteration (use of plastic surgery), and recognition in challenging condition like rain, fog, and mist [78] [55]. These challenges have been addressed by the use of different spectra, including shortwave IR (SWIR), midwave IR (MWIR), near infrared (NIR), etc. NIR is robust against low light conditions (good SNR ratio), does not require external illumination sources, is not visible to human eyes (thus, the system can remain unobtrusive and covert) and last but not least explores the intrinsic properties of the human body (thermal emission of the face), which enables the user to clearly distinguish between the scene's different elements and makes face detection easier. However, spectral bands in the IR spectrum suffer from two limitations: (1) Most traditional face databases were captured in the visible spectrum and thus a cross-matching (matching between different spectral bands such as visible and NIR

or mid range and long range NIR) has to be performed; (2) The acquisition of multispectral images involves additional sensors with different properties (field of view, precision of lens alignment, resolution, range, etc.) which leads to the acquisition of non-geometrically identical images. Preprocessing images before matching has been shown to be useful in overcoming the two issues mentioned above.

Multispectral image preprocessing techniques can be divided into two groups: (a) Photometric normalization techniques that reduce illumination variation between matched spectra (the first issue), and (b) Geometric normalization techniques that reduce geometric distortions of images (the second issue). One or both technique groups could be applied, depending on the face database in hand, as different databases involve different preprocessing schemes. In Table 2.1 , we summarize the most commonly used multispectral image databases with general details about the spectral bands contained, the number of persons presented, and the different imaging constraints (conditions) provided. We use the following abbreviations to describe the constraints in the correpending databases: speech (SP), expression (E), pose (P), time delay (TD), illumination (I), distance from the camera (Dis), and presence of eye glasses (GL)

Photometric Normalization

In many applications, problems such as noise amplification in flat regions or the appearance of artifacts at image edges may lead to image misinterpretations. For example, night vision images (NIR, SWIR, etc.) are generally acquired under different illumination conditions than daylight images, which may result in different image brightness and dynamic range. Hence, photometric normalization is recommended. Shan et al. used Gamma intensity correction (GIC) and quotient illumination relighting techniques (QIR) to control the overall brightness of images. They project the resulting images to a pre-defined canonical face image [131], [114] applied the logarithm transform to increase the contrast of SWIR images. The logarithm function redistributes the original darker pixels (most of the pixels in a SWIR images) on a much broader range and compresses the range of the original brighter pixels (a smaller number of pixels). Techniques like Histogram Equalization (HE), contrast limited adaptive Histogram Equalization (CLAHE), single scale retinex with logarithmic (SSRlog) and with arctangent (SSRatan) and single scale self quotient (SQI) were also used for contrast enhancement [10] [16] [3] [142]. They are simple to implement because they treat the images globally instead of locally, and regions with different intensity ranges are modified in the same manner. As a result, crucial details like edges may be lost (smoothness effect). To overcome this issue, multi-scale operators have been used. In [157], [131]

Table 2.1 Main databases used in multispectral face recognition

Name	Spectral bands	No of subjects	Images per subject	No of conditions	Web address
Equinox	LWIR, MWIR, SWIR, Visible	95	From 43 to 516 for LWIR and visible class	I,E,GL	www.equinox sensors .com
Laval University	Visible, NIR LWIR, MWIR	-	-	E,TD,GL	-
West Virginia University	Visible, SWIR	135	7-25	-	-
NIST–BSSR1	Matching scores (not images)	-	-	-	www.nist.gov
XM2VTS	Visible	295	4	E,P,TD	www.ee.surrey.ac .uk
NFRAD	Visible, NIR	50	6	I,P, Dis	-
MBGC	Full visible, Full or partial NIR videos	114	148 NIR images, 148 visible images (total number)	Half face images	www.face.nist.gov
polyU–HFSD	400 nm – 720 nm (33bands)	25	3	P,TD	www4.comp.polyu .edu.hk
CMU	450 nm–1100 nm	54		TD,I	www.ri.cmu.edu
OTCBVS	Visible, NIR, thermal	227	More than 8000 images	I,E,P	www.cse.ohiostate .edu
IRIS	NIR, Visible	82	More than 2000 images	I	www.imaging.utk .edu
ASUMS	Visible,thermal	96	-	GL,P,I	[169]
HFB	Visible, NIR	100	9-10	TD, I,P	[92]
PRE-TINDERS	Visible, 1550 nm	48	4 images for each spectral band	Open mouth and neutral expression at 2 meter distance	www.wvhtf.org
UCH Thermal Face	Thermal(7.5–13.5 um)	155	22 images	P, SP, E, indoor, outdoor	[62]

images were divided into a number of small blocks and histogram equalization was performed within each image block (Block-based Histogram Equalization BBHE). Block sizes should be defined carefully as very small block sizes could introduce noise when applied on homogenous regions. This can be seen when using the adaptive histogram equalization technique AHE. The block size controls the trade-off between the local enhancement and noise amplification. Over-complete bases (discrete wavelet transform DWT, cosine wavelet transform DCT, etc.) have been also used as a solution to divide information into different levels of importance for the purpose of selectively enhancing features of interest. Jin et al. used the wavelet transformation to decompose images into several sub-bands [74]. Afterwards AHE was applied on each sub-band with different local window sizes depending on the sub-band frequency properties. Interesting results can also be obtained when images are transformed to the logarithmic domain before being decomposed. Chen et al. showed that they can compensate for illumination variation by discarding low frequency DCT coefficients in the logarithm domain [28]. Liao proposed to use the difference of Gaussian filters (DoG) to increase similarity between vsible and NIR images in a cross-matching context [93]. Applying the DoG filter is equivalent to two-scale filtering, high pass filtering and low pass filtering. It suppresses highlight areas in the low frequency domain that cause the most differences between the two modalities and it also suppresses noise and aliasing that lie in the high frequency domain. Based on the approaches mentioned above, photometric normalization techniques can be divided into two groups: (1) single scale techniques and (2) multi-scale techniques. Examples and properties of photometric normalization techniques are presented in Table 2.2

Geometric Normalization

The purpose of geometric normalization is to project face images into a unified refer-ence, generally an average face image, for eventual fusion or cross-matching. Two main concepts define geometric normalization, face detection and face alignment. Although these two tasks can be done separately, they are in most cases performed at the same time: faces are aligned while detected. Face detection approaches can be divided into four categories: (a) knowledge-based approaches, (b) feature invariant approaches, (c) template matching, and (d) appearance-based methods [141] [32]. In thermal spectra, faces are easily distinguishable from the background environment, which facilitates robust segmentation. In this case, knowledge-based approaches that encode human knowledge to capture the relationships between facial features perform very well. This was shown, for example, by Zheng et al. for detecting faces in SWIR and MWIR images [170]; they used adaptive thresholding

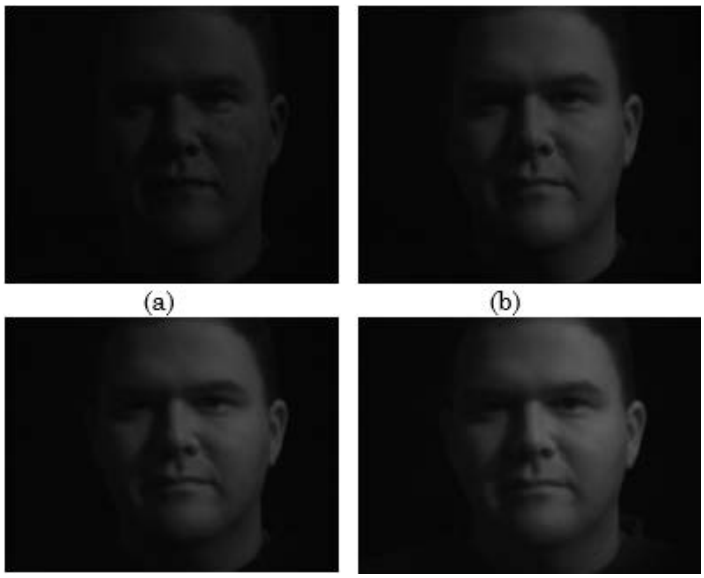


Fig. 2.1 Sample images in a data record in the $IRIS-M^3$ database; spectral images under daylight, side illumination at (a) band 580 nm, (b) band 620 nm, (c) band 660 nm, and (d) band 700 nm.

Table 2.2 Categories of photometric normalization techniques

	Single-scale techniques	Multi-scale techniques
Examples	GIC, QIR, HE, CLAHE logarithm transform, SSRlog, SQI, SSRatan	BBHE, AHE, wavelet based HE , DCT based HE, DoG.
Properties	Simple to implement, Preserve the image gray variation due to monotonic transformations, Loose of details due to smoothness effect	Need more processing time, Non linear transformation that fits well with non linear illumination variation, Enable selective enhancement

based on histogram analysis and projection profile analysis. Once detected, a first set of faces from the database was manually aligned to a reference image. A binary mask learned from the manually aligned faces was then used to automatically detect and align the rest of image sets.

Within the same knowledge-based approach, Bourlai and Cukic detected faces in MWIR images using a simple binarization process followed by blob analysis [16]. They emphasized the specific properties of different spectra of light by using different detection processes for each one. SWIR, MWIR and visible light images were captured in controlled conditions and from short distances between the subject and the camera. Face and eye detection were performed automatically by apply-ing Viola and Jones' AdaBoost algorithm [153]. For SWIR and NIR images cap-tured at distances larger than 60 meters, the eye centers were manually marked. Faces were then aligned automatically using affine transformations like rotation, scaling and translation. In the above approaches, face detection and image alignment are performed as two separate and consecutive steps. Faces that have been previously detected, either manually or automatically, are then aligned using affine transfor-mations that generally start by projecting eye locations to fixed positions, then rotate, scale and translate faces to fit the reference image [131], [10]. Different methods that combine face detection and face alignment have also been proposed within the appearance-based approach. These methods overcome the problem of eye detection in IR images by addressing face detection as a learning-based process instead of a knowledge based process.

In [135], the authors adopted an original approach to minimize the linear and non-linear deformation between visible and LWIR light images in a fusion context. The basic idea of this approach was to register an image V (for Visible) with respect to an image I (for Infrared) by minimizing the normalized mutual information (NMI) between a transformed version of V (T (V)) and I [63]. In brief, mutual in-formation, as a concept from the information theory, is a measure of the amount of information that a signal (or image) contains about another signal (image) [33]. The objective was to optimize the six parameters of T, specifically parameters that define shear, scale, and rotation, and others that define translation transform, so that the NMI between both image modalities is maximized. As geometric deformations that affect images are generally non-linear, the authors first built a Gaussian pyramid from both visible and thermal face images. Then, T was estimated for the coarsest level and used to warp the face image in the next level of the pyramid. The process was iteratively repeated through each level of the pyramid and a final transformed visible face image was obtained at the finest pyramid level.

Since the computation of the NMI and the corresponding optimal values of T is time con-

suming, Zheng proposed to use two cameras located in the same fixture and aimed at the same subject [169]. By doing so, the solution space of Tś parameters is reduced. The author also proposed to use the Fourrier-Mellin transform method (FMT, [27]) for translation, which is much faster than NMI-based registration, while keeping NMI-based methods for scaling and rotation registration. Further readings on appearance based approaches for IR face detection can be found in [129]. Note that feature invariant and template based approaches for face detection and alignment have been adopted to a smaller degree by researchers in the field of multispectral imaging. The former is generally applied when it is necessary to process faces with pose and expression variations, which is typically avoided in order to prevent further complications in face recognition. The latter, template based approaches, are practically infeasible for images like SWIR or LWIR, where face landmarks, like eyes and mouth, are difficult to locate and therefore templates are not easily built.

Finally, multispectral face image preprocessing is both more necessary and more complex than the preprocessing needed for visible light images. For example, locating eyes automatically in thermal images is challenging due to a partial absence of face features in such spectral bands. Illumination problems are also common in multispectral imaging due to person-to-person variations in skin reflectance at different wavelengths.

The techniques for face detection mentioned above should be sufficient for systems dealing with a small number of spectral bands per image. However, when the number of spectral bands becomes larger (hundreds of bands per image in some cases), these techniques may significantly slow the recognition process.

2.2 Face Recognition

Face recognition (identification) techniques can be either homogenous or heterogenous. Homogenous face recognition techniques match a candidate face image, referred to as a probe or genuine image, against a database of images from the same spectral band as the probe. In contrast, heterogenous face recognition techniques match images from different spectral bands against each other. The review of papers published during the past five years that dealt with multispectral face recognition has revealed three major observations:

– The vast majority of researchers have focused their efforts on different methods to optimally exploit frequently used spectral bands, such as NIR and SWIR.

– Published works can be classified into four main categories: (i) work that fuses spectral images at the feature level, which is the most commonly used approach, (ii) work fusing

images at the image level, (iii) work fusing images at the match score level, and (iv) work fusing images at the decision level, which is the least common approach.

– Heterogenous face recognition is receiving increasing attention from researchers due to the intrinsic complexity of the problem (complexity related to the nature of the problem itself, rather than to the working conditions) as well as the desire to match face images from different modalities.

In the following pages, we present an overview of the issues in homogeneous and heterogeneous face recognition.

2.2.1 Face Recognition: Homogenous Context

This section focuses on multispectral face recognition in a homogenous scenario. This task is generally reduced to two issues: (i) selection of the optimal spectral bands to fuse (band selection), and (ii) fusion of the selected spectral bands to obtain as much (multispectral) information as possible. Few research groups have worked on the first issue, band selection, during the past half decade. To the best of our knowledge these groups were (1) Chang et al. at the University of Tennessee [25], (2) Di et al. at the Hong Kong Polytechnic University [41], and (3) Robila at the Montclair State University [125]. The limited number of publications on band selection optimization may be due to the fact that an imaging system capable of capturing hundreds of spectral bands, as in [125], is much more complex and costly than a system suitable for capturing only two or three bands. In addition, the question of the effect of involving a large number of bands on the system's performance arises, as well as whether the addition of components to the recognition system, such as more complex image registration algorithms and an algorithm for band selection, will negatively impact the feasibility of building real time systems. To answer this question about the effectiveness of using hyperspectral imaging systems (hereafter hyperspectral will stand for more than 30 spectral bands, while multispectral will stand for more than 2 spectral bands) and obtain insight into their algorithmic complexity, we will present the main concepts in the works of the three above-mentioned research groups.

Chang et al. [25]. Chang et al. addressed the problem of face recognition in the visible spectrum [25]. They built an imaging system able to capture hyperspectral images in the visible spectrum (from 420nm to 700nm) with wavelength steps of 10nm, thus obtaining 33 visible bands. The authors then attempted to select the most informative bands, based on the observation that different bands contain different amounts of information that is useful

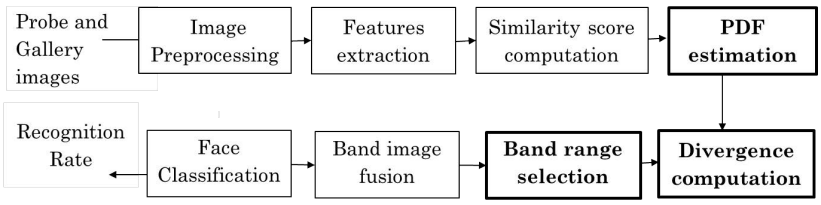


Fig. 2.2 Processing pipeline for band selection in the visible spectrum (adopted from [25])

for face recognition. The pipeline of the proposed system for band selection is shown in Fig. 2.2.

The primary contribution of this work lies in the three steps shown in bold in Fig. 2.2. In the first step, match scores computed between probe and gallery images using a commercial engine (in this case FaceIt and FaceVACS were tested to demonstrate generality of the system) were subdivided into a genuine set G_K (scores between gallery and probes images for the same subject) and an impostor set I_K (scores between gallery and probes images for the different subject). K denotes the visible spectral band used in the matching. In the second step, the probability distribution functions for each of these two sets, namely $p_{G,K}$ and $p_{I,K}$, were estimated using kernel density estimation (KDE). In the ideal case, $p_{G,K}$ and $p_{I,K}$ should cluster at the two opposite parts of the score line - in the high score region and in the low score region, respectively. However, this is not true in the real case, where the two PDFs generally overlap.

Chang et al. proposed to use the Jeffrey divergence to quantify this overlap. They claimed that for a fixed illumination condition, this overlap is an intrinsic property of the corresponding band and can be a useful feature for band selection. For each band they obtained a measure Q_K of the divergence between each impostor and gallery PDF. Finally, bands with the highest divergence measure were considered as the best bands for recognition. The Haar wavelet based pixel level fusion method described in [53] was used to fuse the selected bands for face recognition. The proposed algorithm was tested in three scenarios depending on the light condition of the probe and gallery images: (a) Gallery: fluorescent, Probe: halogen, (b) Gallery: daylight, Probe: halogen), (c) Gallery: fluorescent, Probe: daylight. From these three scenarios three sets of bands have been found to give the best recognition results, which are respectively (610 nm, 630 nm, 640 nm), (610 nm, 620 nm, 630 nm) and (680 nm,700 nm,720 nm). Tables 2.3, 2.4, and 2.5 summarize the obtained 'rank one recognition rates' for the three tested lighting configurations.

Di et al. [41]. Di et al. studied the advantages of using hyperspectral images in the visible spectrum for face recognition [41]. $(2D)^2$ PCA was applied for feature extraction [**?**].

Table 2.3 Scenario 1. Gallery: Fluorescent, Probe: Halogen

Used bands	(610nm,640nm,630nm)	(610nm,640nm)	Broad-band
Rank-1 recognition rates	97.14%	94.29 %	88.56 %

Table 2.4 Scenario 2. Gallery: Daylight, Probe: Halogen

Used bands	(610nm,620nm,630nm)	(610nm,620nm)	Broad-band
Rank-1 recognition rates	65.71 %	68.57 %	57.14 %

In brief, $(2D)^2$ PCA is an enhanced, conditionally less expensive variation of the Principal Component Analysis (PCA) where images are projected into a smaller 2D space (in PCA, images are projected into a 1D space). Based on whether G, the data covariance matrix, was computed for the whole set of spectral bands (24 bands were used from 460 nm to 690nm), for each single spectral band or for a subset of the captured spectral bands, specifically $\xi_1 = [530nm, 540nm, 550nm]$ and $\xi_2 = [570nm, 580nm, 590nm]$ the authors developed three algorithms for the task of face recognition: (a) whole set of bands $(2D)^2$ PCA(WB), (b) single band $(2D)^2$ PCA(SB), and (c) band subset fusion-based $(2D)^2$ PCA(BS-xFD). The letter 'x' represents the image fusion technique to be used (either the weighted averaging method WF [134] or the Laplacian pyramid fusion LF [19]. The selection of subsets ξ_1 and ξ_2 was justified by a study of the absorption characteristics of two chemical compounds of the human skin, which are melanin and hemoglobin, under two different light intensities. In WB, G was computed for the whole set of spectral bands, and the final classification is obtained by minimizing the Euclidian distance between a probe and the gallery images measured in the projection space defined by G's eigenvectors. In SB, G was computed for each band. The same measure of similarity is used as in WB, and N decisions were obtained (N is the number of bands to be used). These decisions were then fused using a decision fusion method based on the majority voting strategy (MVD). Finally, images of each subset were fused in BS-xFD using WF or LF. G was computed for each of the obtained fused images and decisions are fused using MVD. Note that SB and WB were also tested on the set $\xi = ((540, 580), \xi_1, \xi_2)$ to study the effect of using parts of the spectral bands where the hemoglobin absorption is higher. Thus, there are two varieties of SB and WB, which are SB(24,1) (Sb on the whole 24 bands), SB(2,3,3) (SB for the set S), WB (24,1)(WB on the whole 24 bands), and WB(2,3,3)(WB for the set S). The above mentioned algorithms were

Table 2.5 Scenario 3. Gallery: Fluorescent, Probe: Daylight

Used bands	(720nm,700nm,680nm)	(720nm,680nm)	Broad-band
Rank-1 recognition rates	97.14 %	97.14	%94.28 %

tested in two experiments: one experiment that includes only algorithms with non fusion steps, and a second experiment that evaluates the efficiency of fusion-based algorithms. The main results can be summarized as:

– SB(2,3,3) and WB(2,3,3) outperformed both WB(24,1) and SB(24,1) with a slightly higher recognition rate for SB(2,3,3); 76 % for SB and 75 % for WB. This emphasizes the fact that contrary to what might be expected, including all the spectral bands may reduce the system efficiency

– BS-WFD and BS-LFD algorithms outperformed all the other algorithms with a recognition accuracy of 79 % and 78 %, respectively. This shows the advantage of image fusion in reducing image noise.

– The BS-xFD system outperformed the conventional RGB-based system.

Robila [125]. Robila developed a PCA based algorithm for face recognition within a hyperspectral context [125]. Database images of 9 subjects were captured. For each subject, five different hyperspectral images were collected in outdoor conditions with various face expressions and postures. Each hyperspectral image was formed of over 100 image bands spanning the spectrum from 400 nm to 900 nm. To achieve an accurate PCA based algorithm within the hyperspectral context and at the same time keep a reasonable computing time (which is not obvious when dealing with images of hundreds of bands per person) Robila proposed three different approaches to apply PCA in a non-classical way and retain the discriminative power of the method. The first approach was to average all bands, for a given hyperspectral image, in one grayscale image; PCA is subsequently applied in the same manner as when using intensity images. In the second approach, PCA was applied for each hyperspectral image (which is formed by hundreds of images), then just one of the eigenfaces (eigenvec-tors) was chosen to represent the corresponding hyperspectral image. The last approach was to apply PCA on each hyperspectral image and average the first three ei-genfaces. In the second approach, one of the first three eigenfaces (eigenfaces with the highest variances) was chosen each time, thus three algorithm were developed within that approach PCA-1, PCA-2 and PCA-3, where the numbers 1, 2 or 3 denote which eigenface was used. Despite that the results were moderate with an accuracy of 70 %, the study emphasized the advantage of using hyperspectral images (ap-proaches 2 and 3) over grayscale images (here represented by approach 1). Robila claimed that spectral images can provide an advantage for face recognition over classic grayscale-image based algorithm.

From the above presentation of the basic concepts used in the three works studied, the following observations can be made:

– Systems involving hyperspectral images together with a band selection algorithms gave better results than did broad-band or gray-scale based systems.

– None of the authors mentioned the time complexity of their systems. Nevertheless, we can predict that it is the highest for the Robila system.

– For [25] the system gave acceptable results (the rank-one recognition rate exceeded 90 % in most cases in a large database with more than 2000 images for 82 subjects, see Table I). Of the three papers presented, this is the only work in which the authors claim that their system could be simplified to reduce the processing time. This is true, since the band selection process can be done only once to select optimal bands for defined working conditions (illumination, expression, etc.). In this case, the band selection process is non adaptive or in short, static. In order to build a system that dynamically selects optimal bands, which is required in many applications, the band selection algorithm should run continually with the proposed system and in this case the computing time will be high for the approach found in [25].

From the above observations we can conclude that much work remains to be done before hyperspectral imaging can be implemented adaptively in real time, particularly when the significant amount of time required for the preprocessing step is added.

Now that band selection has been discussed, the next issue is band fusion. Image or band fusion is the process of combining two or more images of the same scene, possibly taken in different wavelengths, into a single composite image that is more informative and more suitable for visual perception or computer processing. The image fusion process is an effective way to reduce the volume of information while at the same time extracting all the useful information from the source images. Several researchers [103] [39] [24] have proposed image fusion of multispectral images as a potential solution to problems due to changes in imaging conditions, such as changes in illumination and occlusion, as well as time delay and aging. Therefore, image fusion has become a primary component of their face detection-recognition systems. Image fusion can be performed at different levels of the recognition system, including feature level fusion, image level fusion (pixel level or sensor level), match score level fusion, and decision level fusion.

Feature Level Fusion

Feature level based fusion techniques have been the most heavily used fusion techniques during the past half decade. In these techniques, features extracted from the original multi-

spectral images are combined to form a feature vector. The latter is then used for classification. Shao and Wang proposed a new algorithm to fuse NIR and visible images [<sha008>]. Images to be fused were decomposed using a 3-level Laplacian pyramid and local binary patterns (LBP) were extracted from each pyramid level [121]. A histogram H of 11 bins was then formed for each pixel. The first ten bins encode the LBP features and the 11th bin encodes a measure of intensity variation around the considered pixel. The latter was called average grads and was defined as $AG = 1/8.\sum_{p=0}^{8}|g_p - g_c|$. Two measures M and S were defined from AG and H for each pixel i:

$$M = \sum_{0}^{10}(|H_{vis,i} - H_{NIR,i}|/|H_{vis,i} + H_{NIR,i}|) \; and \; S = AG/SD, \tag{2.1}$$

$(H_{vis,i}, H_{NIR,i})$ and (g_p, g_c) are histograms of the corresponding pixels in the visible and NIR bands, respectively, and values of the center pixel and neighboring pixels, where i=0..10, span the histogram bins. S is the salient value of the pixel and SD is the image standard deviation. Adaptive weighted fusion was applied as:

$$I_{fused} = w_1.I_{vis}(i,j) + w_2.I_{NIR}(i,j)$$

and

$$(w_1, w_2) = \begin{cases} w_2 = w_{min} \; and \; w_1 = w_{max}, & \text{if } S_{vis} \geq S_{NIR} \\ w_1 = w_{min} \; and \; w_2 = w_{max}, & \text{otherwise} \end{cases}$$

$$(w_{min}, w_{max}) = \begin{cases} w_{min} = (M-1)/2.(1-a) \; and \; w_{max} = 1 - w_{min}, & \text{if } M \geq 0.75 \\ w_{min} = 0, w_{max} = 1, & \text{otherwise} \end{cases}$$

At last, the proposed algorithm, named AGLBP, was evaluated against two other image fusion algorithms, Simple Choose Max (SCM) and Burt's Algorithm (Burt). Fused images obtained by AGLBP have shown a greater entropy and lower mean cross entropy [105] than have the two other algorithms. Ting et al. proposed a Meyer wavelet based combination of IR and visible information for face recognition [145]. The discrete wavelet decomposition was applied at three levels on both spectra, IR and visible. First, the obtained subbands were tested separately in a face recognition scenario to evaluate their recognition performance. The approximation subband at the first level gave better results. Gabor features with five scales and eight orientations were extracted from these approximation subbands for the two image modalities and combined in one feature vector. Finally, the Fisherface technique [8] was applied on the latter to keep salient details. The similarity between probe and

gallery images was measured using the distance between the corresponding feature vectors. The obtained results were compared to those obtained with two existing approaches: (1) using visible images only, and (2) using image fusion as in [81]. The algorithms were evaluated for the Notre Dame database using the equal error rate (EER). The proposed approach achieved a lower false rejection rate at a given false acceptance rate (mainly for FAR< EER).

The wavelet transform has also been used in [108]. Moon et al. computed the coefficient based activity level (CBA) for each pixel in each decomposition level and for each image modality, namely thermal and visible. Weighted fusion was performed, where weighting coefficients for each pixel were determined based on the latter's CBA. The obtained fused images were then classified using a support vector machine (SVM) with directed acyclic graph (DAG) [67]. Zahran et al. proposed to enhance the face recognition performance using the NIST/Equinox database [164]. LWIR and visible images where summed in the DWT space. Feature vectors for matching were then formed using PCA, and ZM (Zernike moment) de-scriptors applied on cropped face images. Cropping was done using ellipse fitting [46]. ZM was shown to be of minimal redundancy compared to many moment-based descriptors. ZMs are also invariant to rotation and robust against noise. The obtained results showed no superiority of ZM over PCA.

Several subspace techniques in addition to PCA have also been used for feature level fusion. Desa and Hati applied (PCA, KPCA, FLD and KFLD) to build feature vectors from thermal and visible images [38]. A weighted sum was used to fuse feature vectors from both modalities, where optimal weights were determined using a genetic algorithm [139]. Weights determination could be also formulated as an optimization problem. Panda and Naik [117] determined fusion weights in the WDT domain using E. coli bacteria foraging strategy (EBFS) [118]. The fused image's energy in the DWT domain was expressed as a function of the fusion weights, and was then maximized using EBFS. The author evaluated the quality of the obtained fused images using co-occurrence signature, contrast and Inverse Different Moment (IDM).

Particle swarm optimization (PSO) was proposed to fuse visible and infrared images in the DWT domain [124]. In PSO, fusion weights for each wavelet coefficients are initialized then determined iteratively so that a fitness function is minimized with respect to the final fused image. This function was chosen as the overlap area of genuine and impostor match scores estimated densities. Two varieties of the proposed algorithm were evaluated, one variety with binary weights, either 0 or 1, and another variety with real-number valued weights. Equal error rate (EER) measures have proven the superiority of the latter variety over the

binary scheme as well as over unimodal recognition where only one image modality is used. In [167], the authors addressed the problem of face detection in multispectral images having a small quantity of targeted data (in this case, the targeted data are two sets of images at the wavelengths of 850nm and 356nm). The aim of the study was to detect faces in either of the two targeted spectra using a transfer learning based approach. A boosting classifier F was trained using both the targeted data and a larger dataset from the visible spectrum (visible images). The knowledge obtained from the visible spectrum was transferred to the targeted spectrum. The entries for F were formed by combining features extracted from each modality, using MultiBlock Local Binary Pattern, and concatenated in one feature vector. To avoid overfitting due to the small quantity of targeted data, manifold regularization was imposed during the optimization of F. The above algorithm was tested in two scenarios for face detection in the 850 nm and 356 nm spectrum. Several other algorithms, adapted from Dai's work [35] and algorithms in [**?**], have been also implemented for comparison purposes. Receiver operating characteristic (ROC) curves showed the superiority of the proposed algorithm.

Image Level Fusion

In image level fusion, multispectral information is combined before any image transformation; images are fused at their pixel state. In [23] an image level fusion based algorithm was proposed to overcome the problem of illumination variation in face recognition tasks. Multispectral images from NIR and visible spectrum were fused based on physics properties of the LCTF filter used in the imaging system. The fused image was defined as:

$$I_{fused} = (1/C).\sum_{i}^{N}(w_i.I_{\gamma_i}), \tag{2.2}$$

$C = \sum_i w_i^N$ is the number of spectral bands used from both the NIR and the visible spectrum and the fusion weights w_i were set as the reciprocal values of the transmittance of the LCTF filter for each spectral band i. This choice of w_i makes it possible to compensate for the intensity differences caused by the filter transmittance. Other physical information can also be used to define the w_i's such as the spectral power distribution of the light sources, the camera spectral response, or the facial skin reflectance. This physics-based sensor-level fusion algorithm gave better results than methods such as averaging and PCA based fusion proposed in [60]. In [117], fusion weights were determined using the gray level co-occurrence matrix (GLCM) [59]. For each image modality, visible and infrared, the GLCMs were computed and the corresponding information measures M_{vis} and M_{infra} were determined. These de-

fine the weights for each image spectrum as $w_{vis} = max(M_{vis}, M_{infra})/(M_{vis} + M_{infra})$ and $w_{infra} = min(M_{vis}, M_{infra})/(M_{infra} + M_{vis})$. Acceptable recognition rates have also been achieved using weights determined experimentally via trial and error [9]. Several combinations of weights have been tested, specifically (0.7, 0.3), (0.5,0.5) and (0.6,0.4), and the latter combination showed the best performance using the OTCBVS database. In [136] the authors proposed a dynamic algorithm that answers the following questions: when shall we combine the two image modalities, SWIR and Visible, when shall we use just one modality, and which one should it be, when shall we fuse information at image level or match score level, and when shall we combine the two fusion levels in a hierarchical manner ? A dynamic algorithm, which is a variety of the majority voting SVM (mv-SVM) algorithm, was applied to maintain a dynamic selection of the ade-quate recognition algorithm based on two factors: the images quality [152] and the algorithms' recognition prior. Four possible recognition schemes have been selected: (1) using only the visible spectrum for matching, (2) using only the SWIR spectrum for matching, (3) Visible and SWIR are fused at image level as in [135] (4) visible and SWIR match score fusion. The mv-SVM algorithm was improved by two concepts, granular computing (G) and soft labeling (S) [143]. The proposed dy-namic algorithm was evaluated against five scenarios: (1) recognition using only the SWIR images, (2) recognition using only the visible images, (3) fusion of all images, (4) match score fusion, and (5) hierarchical fusion. The latter three algorithms were introduced in [135]. The Equinox image database augmented by the multispectral database captured at the West Virginia University was used for all tests. The ob-tained results showed better recognition accuracy for the SWIR images than for the visible images (85.6 % compared to 78.5 %). Furthermore, the fusion-based algorithms, both at the image level and the match score level, increased the recognition accuracy by at least 10.2-13.7 % and the proposed dynamic algorithm gave the best results with a slightly higher performance than the hierarchical algorithm (99.3 % compared to 99.1 %)

Match Score Level Fusion

In match score level fusion based systems, images are matched against each other and the obtained match scores are combined to get a fused match score vector valid for classification purposes. In [20] [155], the authors addressed the problem of face recognition in challenging conditions such as pose and face expression variation using a sparsity-based assumption: a probe image of a subject when the latter exists in the gallery database could be expressed as a linear combination of a small number of images from the same gallery class to which the subject belongs. Hence, the probe image is said to have a sparse representation

over the gallery. This assumption can be written as $Y = A.X$ where Y is the feature vector of a probe image, A is a projection matrix whose columns form an overcomplete basis and X is the sparse representation of Y , meaning that most elements of X are zeros. The problem of computing sparse linear representations with respect to an overcomplete dictionary of base elements (also called atoms) has been widely addressed in the signal processing field, see for example [42]. The sparsity-assumption based approach was introduced by Wright et al. as potential solution to the face recognition problem [155]. In [20], $X = [x_1...x_N]$ was constrained to minimize the objective function $\|Y - A.X\|_2^2 + \gamma.\|X\|_1$. The norm l1 ensures the sparsity of X and the columns of A were formed by N feature vectors A_i corresponding to the N gallery images. Finally, the match scores between Y and each gallery image A_i were defined by the residual $r_i = \|Y - x_i.A_i\|_2$. To enhance the recognition performance, residuals from visible and thermal images were summed, $r_{i,fusion} = r_{i,vis} + r_{i,Therm}$, and Y was considered to correspond to the gallery image A_i that has the smallest $r_{i,fusion}$. Obtained results, tested for the Notre Dame database, showed the superiority of the proposed algorithm over works in [29].

In [122], a weighted sum was used for match score fusion. For each image modality, visible and thermal, Euclidian distances between query and gallery images, D_{Therma} and D_{Vis}, were computed. The latter were weighted to get a fused distance $D_{fused} = \alpha.D_{Therma} + (1 - \alpha).D_{vis})$. At last, the K-nearest neighbor algorithm (K-NN) was applied on D_{fused} to classify each query image. Besides weighted sums as used above, other untrained score level fusion rules have been investigated by [106]. The authors proposed to fuse visible images with synthesized NIR images. The latter were synthesized from their corresponding visible images using Canonical Correlation Analysis (CCA) [66]. Match scores were computed by normalized correlation. Five rules were tested for fusion: max, min, average, product fusion rules as untrained rules, and SVM. The authors reported a superiority of the average rule over untrained rules, while SVM gave the lowest half total error rate (HTER) in overall. The superiority of SVM emphasizes the importance of learning the statistic properties of match scores (density distribution, covariance, etc.) for fusion purposes. In [20] density distributions of distances between the probe and each gallery class k have been estimated for thermal and visible modalities, namely $S_{k,Vis}$ and $S_{k,Therm}$. Weighted fusion was performed, where the fused distance, $D_{k,Fused}$, was a sum of distances from both image modalities each ponderated by its saliency measure: $D_{k,Fused} = S_{k,Vis}(D_{k,Vis}).D_{k,Vis} + S_{k,Therm}(D_{k,Them}).D_{k,Them}$. The proposed algorithm outperformed either image modality, used alone, in several scenarios including lighting and facial expression variation presented by the Notre Dame database.

The problem of determining the weights within a weighted fusion process can be formulated

as an optimization problem. In [146], a power series model was used to fuse match scores from different sources. In this model, the fusion function g was defined as: $(\alpha, x) = \alpha_0 + \sum_{m=1}^{r} \sum_{j}^{l} .\alpha_{m,j}.x_j^m$, where the x_j's are match scores and $\alpha = [\alpha_{m,j}]$ is a vector of weights to be optimized. The authors proposed two algorithms to estimate α. The first algorithm was based on the minimization of the least square error (LSE) of g, and the second algorithm was based on the minimization of the system Total Error Rate (TER). The proposed power series-based algorithm with TER minimization has shown a better performance against LSE-based version as well as algorithms proposed in [70] and OSU Matlab toolbox [100]. In [4], the authors proposed to determine fusion weights for thermal and visible match scores using probability theory. Weights should sum to one and the fused match score was written as $M_{fused} = \alpha_{vis}.M_{vis} + (1 - \alpha_{vis}).M_{therm}$. The value of α_{vis} was constrained to be the optimal choice of the mixing coefficient α given M_{vis}, hence $\alpha_{vis} = argmax_{\alpha} p(\alpha|M_{vis})$. The density $p(\alpha|M_{vis})$ was determined experimentally. The proposed algorithm was evaluated for the OTCBVS database with subjects wearing eyeglasses. A recognition rate of 97 % has been reported.

In [109], the authors proposed to fuse biometrics data of face, iris, and fingerprint based on a likelihood measure (LR). According to the Neyman-Pearson theorem [85], the LR test is the optimal test to decide whether a match score vector is genuine or impostor. To compute LR, densities of genuine and impostor sets of match scores, namely $f_{genuine}$ and $f_{impostor}$, were estimated using a Gaussian mixture model GMM as proposed in [45]. *LR* was defined as $LR(X) = f_{genuine}(X)/f_{impostor}(X)$, where X is the vector of multi-biometrics match scores for each subject. To improve the performance of their multi-biometric system, the authors have chosen to integrate a measure of match scores quality Q into the densities estimation step. Thus, a quality-based likelihood ratio *QLR* was also defined as $QLR(X,Q) = f_{genuine}(X,Q)/f_{impostor}(X,Q)$, where Q was computed as in [110]. The proposed algorithms were evaluated for three databases: the NIST-BSSR1 database, the XM2VTS Benchmark database, and the WVU-Multimodal database. *LR* and *QLR*-based fusion algorithms were tested against linear SVM, SVM with radial basis, single matcher (matcher based on either of the defined biometrics), and the sum of scores fusion method with scores normalized using the min-max algorithm [72]. ROC curves and 95 % confidence intervals showed a constant superiority of the two proposed algorithms over all the other matchers. SVM based algorithms also gave good performance when compared to *LR* and *QLR*; however, in this kind of algorithm the kernel function and the associated parameters should be carefully chosen. Finally, obtained results have shown an increase of performance when using *QLR* over *LR* (at FAR= 0.001 % obtained results were 85.3 % and 90 % for respectively QLR and LR based fusion).

Decision Level Fusion

Fusing images at the decision level is the task of aggregating decisions obtained from different classifiers in one final decision that determines to which class an image belongs. In general, techniques for decision level fusion [65], [104] are performed along four main steps as illustrated in Fig. 2.3.

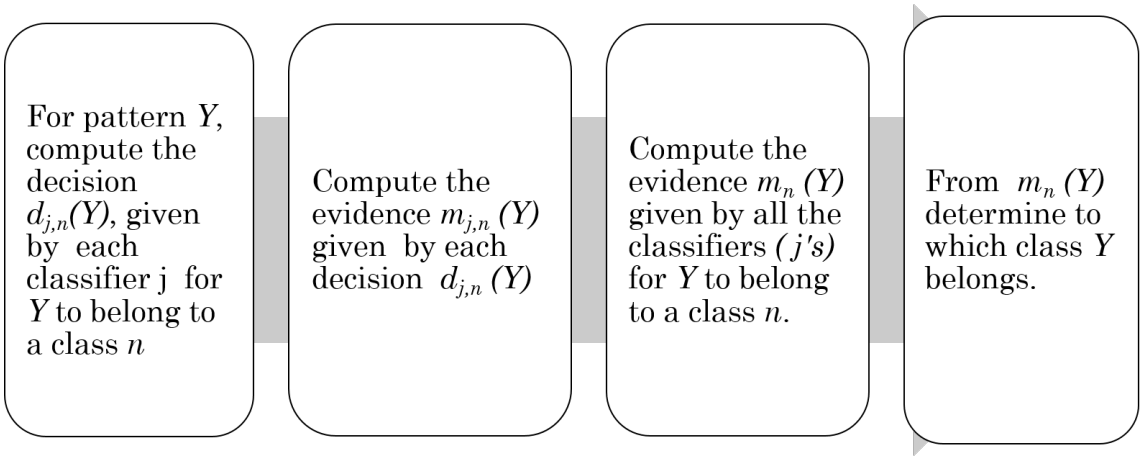


Fig. 2.3 Processing steps in a decision level fusion process.

During the first step, each classifier makes a decision regarding which class a giv-en probe image belongs to. These decisions could be either deterministic, like match scores, ranks and distances, or non-deterministic such as the probability of belonging to a given class. The evidence of each obtained decision is then computed. Each evidence should quantify the belief on the corresponding decision. In [57], thermal and visible images were divided into non overlapping smaller regions called modules. Corresponding modules from probe and gallery images were then projected into a learned eigen space and matched against each other using nearest neighbor technique. Based on the obtained match scores, the number of modules that voted for a given pattern Y as belonging to a class j was computed for each image modality. This number defines the belief on the corresponding classifier (thermal or visible) for classifying Y as class j. The sum of the two evidences coming from thermal and visible classifiers determine the total belief on Y to belong to a class j. Y is then assigned to the class that has the greater total evidence. Ranking decisions have also been presented as a simple, yet effective, approach for evidence computation. A decision rank stands generally for a match score rank, and thus, the smaller the rank is, the greater is our belief on the obtained decision. [130] proposed a comparative study of several rank based decision fusion techniques, specifically minimum ranking fusion, product ranking fusion, and aver-

age ranking fusion [130]. In these techniques, a pattern Y is matched against each class from the gallery database, and the obtained match scores for IR and visible modalities are ranked separately. Ranks of each class are then combined (product, average or minimum) to form its total rank. Finally, Y is assigned to the first ranked class. The study has shown a superiority of the average ranking technique for the Equinox and OTCBVS databases.

Ranks, number of votes, product, average or order statistics operator (min, max, etc.) can be considered as non-learning-based techniques. However, learning-based algorithms have also emerged as a potential approach to decision fusion. In [30] SVM classifiers were trained on LWIR and visible images to determine the probability with which a probe image belongs to a defined class. The latter probability has the meaning of decision evidence. Fuzzy logic was used to aggregate evidences from both modalities into one decision [76]. For each class k, the fuzzy decision evidence F_k, or probability of belonging to k, was computed as:

$$F_k = \begin{cases} max(min(S_{k,vis}, g_{vis}), S_{k,IR}), & \text{if } S_{k,vis} \geq S_{k,IR} \\ max(min(S_{k,IR}, g_{LWIR}), S_{k,vis}), & \text{otherwise} \end{cases}$$

where g_{vis} and g_{LWIR} are Sugeno densities determined via statistical measurement on the recognition rate of the corresponding image modality. $S_{k,vis}$ and $S_{k,LWIR}$ are the probabilities, classification evidence, given by the SVM classifier for visible and LWIR images, respectively, to belong to class k. A probe image is classified to the class that has the largest fuzzy decision value. The proposed approach was compared to other fusion techniques, specifically average matching score, highest matching score, and Bayesian theory coupled with feature extraction and match scores techniques involving PCA, Independent Component Analysis (ICA), K-NN and SVM. Results have shown a superiority of the proposed fuzzy-based algorithm with remarkable stability against illumination and expression variation as well as presence of eyeglasses. In [112], the belief on a decision, $D_k^j(Y)$, made by a Concurrent Self-Organizing Maps classifier j [111] for a pattern Y to belong to a class k was modeled as a decreasing function ϕ of the distance of Y to k. If we denote the defined belief as m_k^j, the latter could be written as:

$$m_k^j(Y) = \phi(-\gamma.D_k^j(Y)), \tag{2.3}$$

The total belief given by the two involved classifiers, infrared and visible, for Y to belong to class k, namely $m_k(Y)$, was obtained using the Dempster and Shafer theory of evidence

[158].

$$m_k(Y) = \prod_j .m_k^j(Y) / \sum_k . \prod_j .m_k^j(Y), \tag{2.4}$$

The class with the largest total belief is then selected for Y. Evidence theory has been also used in [135], where two match score vectors, s_1 and s_2 , were computed using 2D log polar Gabor [137] and local binary pattern (LBP) based matchers. s_1 and s_2 were considered as classifiers' decisions and fused using the Desert Smarandache (DSm) rule of combination [40], [140]. Finally, if the obtained fused decision $m_{fused} = m_{fused,vis} \otimes m_{fused,LWIR}$, where $\otimes$ is the DSm rule operator, is greater than a threshold T, then the probe and gallery images match, otherwise not. The proposed algorithm was evaluated against several existing score fusion techniques; Product rule fusion, Sum rule fusion, SVM fusion, and DST fusion [138]. Results have shown an increase of performance of at least 1 % for the Notre Dame database and 1.57 % for the Equinox database, compared to the DST based fusion algorithm, the latter gave the second best results.

From the above analysis of image fusion approaches, we can see that the fusion of multi-spectral information at the match score level seems to be the most attractive for researchers, as shown by [126]. However, data fusion at the image level has shown an inability to solve problems of illumination, expression or time variation, and has therefore been of less interest.

2.2.2 Face Recognition: Heterogeneous Context

Heterogeneous face recognition refers to recognition of faces captured in different modalities; Visual (VIS), near infrared (NIR), thermal infrared (TIR), etc. Heteroge-neous face recognition arises as a challenge in several situations, for example in nighttime surveillance. To address this challenge, several algorithms have been recently developed that could be divided into two main categories: (1) methods that project all the modalities into a common feature space and apply common face similarity measure methods for recognition and (2) methods that extract modality-invariant features and apply common face recognition techniques. In the following, we review the most recent works within these two categories.

Common Space Based Methods

In common space based methods, face spaces defining the different image modalities could be either 1) projected in a new subspace or 2) analyzed in a synthesis process in which one

of the involved face spaces is kept as reference, and all the others are projected to it. Most of the works on face synthesis were based on a key assumption that an image pixel preserves, roughly, the same geometric relationship with nearby pixels through different image modalities [98]. In [26], the similarity between image patches was defined as a geometric property that would be preserved from NIR to a synthesized visible image. An NIR patch, from a probe gallery, was approximated as a weighted sum of the corresponding K-nearest patches (in the sense of a similarity measure) from the NIR gallery database. Weights were defined as the similarity measures between the corresponding patch and the probe NIR patches. A synthesized visible patch was defined as the same weighted sum of the K visible patches corresponding to the same K-nearest NIR patches. Once all patches are synthesized, the latter are combined to form the synthesized visible image. Sparsity could also be exploited as a form of geometric relationship that is preserved from one face space to another. In [166], VIS and NIR patches were presented as a sparse linear combination α^{vis} and α^{NIR} in learned dictionariesD^{vis} and D^{NIR}, respectively. The authors claimed that if D^{vis} and D^{NIR} are couple-trained using training patches, α^{vis} and α^{NIR} could be considered as equal, α^{vis} = α^{NIR}. Hence, for each NIR patch X^{NIR}_{patch} , from the probe database, a sparse vector α^{NIR} was determined so that $X^{NIR}_{patch} = D^{NIR}.\alpha^{NIR}$. The sparsity condition was ensured by an l_1 norm based optimization, and the synthesized patch was determined as $X^{vis}_{patch} = D^{vis}.\alpha^{vis}$. Cumulative match scores have shown a superiority of the proposed approach against methods in [26] and [98]. In [101] the authors proposed to extend the local sparsity preserving property previously defined at a nearby-pixel-level (small patches) to a more general one, where sparsity is preserved at the image level. Hence, images from different modalities have the same sparse combination over the corresponding gallery database. In their work, a test NIR image X^{NIR} was expressed as a sparse combination α^{NIR} of images from the NIR gallery A^{NIR}, namely $X^{NIR} = A^{NIR}.\alpha^{NIR}$. The corresponding visible image was synthesized as $X^{vis} = A^{vis}.\alpha^{vis}$, where the same sparse vector α^{NIR} is used. Although the proposed method gave better results in comparison to the direct matching between NIR and visible images, image level sparsity that assumes the existence of image level linear mapping function F between different modalities is practically invalid.

This encouraged the use of patch-based methods where F is defined locally. In [154], Lambertian face models were used to define two operators for face synthesis, namely $H_{vis} = (I_{vis}(i) - m(N_i))/\sigma(N_i)$ and $H_{NIR} = (I_{NIR}(i) - m'(N_i))/\sigma'(N_i)$. The authors have shown that with the assumption that F exists within a neighborhood N_i of an image pixel i, both operators are locally equal: $H_{vis} = H_{NIR}$. Hence, synthesized images could be determined as

$$I_{vis}(i) = [(I_{NIR}(i) - m'(N_i))/\sigma'(N_i)].\sigma(N_i) + m(N_i), \tag{2.5}$$

where $\sigma'(N_i)$ and $m(N_i)$ were approximated as the local covariance and mean of the visible gallery image $I_{vis}^{best-match}$ that correspond to the NIR gallery image that best matches $I_{NIR}(N_i)$ within N_i. In [96], a different approach was proposed to determine $\sigma'(N_i)$ and $m(N_i)$; the latter was defined as a weighted sum of the corresponding covariance and mean of gallery visible images within . Weights were learned from the NIR gallery database as the similarity measure between and all the other NIR images. The authors proposed to use Gaussian kernel for similarity measure. Finally, both approaches gave better results compared to NIR-Visible direct matching technique, with better recognition rate for the best-match based technique proposed in [154] over the weighted-based approach given in [96] for different databases.

As mentioned above, heterogeneous face matching could be also performed in a new learned subspace. In this approach, a projection matrix W^k is generally learned for each image modality k. A test image X_k is projected to the new subspace as $Y^k = W^k.X^k$. Several algorithms have been designed to optimize W^k. In [95] a common discriminate feature extraction (CDFE) algorithm was proposed. In CDFE, W^k is constrained to maximize the intra-class compactness and the inter-class dispersion of the projected images (empirical separability) as well as preserving the local consistency (nearby features should have nearby transforms). Hence, W^k should minimize the objective function J defined as:

$$\begin{aligned} J = &\sum_i \sum_j \mu_{ij} \|W^p.X_i^p - W^g.X_j^g\|^2 + \sum_i \sum_j \upsilon_{ij} \|W^p.X_i^p - W^p.X_j^p\|^2 \\ &+ \sum_i \sum_j \zeta_{ij} \|W^g.X_i^g - W^g.X_j^g\|^2 \end{aligned} \tag{2.6}$$

The first term of J ensures empirical separability, while the second and third terms guarantee local consistency. W^p and W^g define the projection matrices for probe and gallery images. In [86] a coupled spectral regression (CSR) approach was proposed to estimate W^k. In this approach, W^k optimization was formulated as a regression problem where the latter defines the regression function to be determined as in [159]. Besides the regression constraint, projection matrices, W^p and W^g, shrank as in ridge regression [61] and are constrained to have minimum difference between each other. The latter constraint is based on the fact that both matrices are operating on the same kind of objects (faces) and thus should not differ too much. The objective function to be minimized by W^p and W^g becomes:

$$J = (1/N).\|Y^g - W^g.X^g\|^2 + (1/N).\|Y^p - W^p.X^p\|^2 \quad + \eta.\|W^g - W^p\|^2 + \lambda.(\|W^p\|^2 + \|W^g\|^2) \tag{2.7}$$

The first two terms are the regression constraints, the third term defines the difference between the two projections matrices, while the last term defines the shrinkage operation. Two

variations of the proposed approach were formulated based on whether W^p and W^g are defined by linear mapping (LCSR) or are modeled as a nonlinear mapping using kernelization (KCSR), i.e. $W^p = \sum \alpha_i^p.\phi(X_i^p) = \phi(X^p).A^p$ and $W^g = \sum \alpha_i^g.\phi(X_i^g) = \phi(X^g).A^g$. Here ϕ is the kernel function and $A^i = [\alpha_1^i..\alpha_N^i], i \in \{p,g\}$ are the new parameters for J. The study reported close results between LCSR, KCSR and CDFE algorithms with LCSR slightly outperforming KCSR. An improved CSR (ICSR) was proposed by the same team in [88]. The authors used projection matrix learning coupled from both modalities:

$$\begin{aligned} W^p &= \sum \alpha_i^p.\phi(X_i^p) + \sum \alpha_i^g.\phi(X_i^g) = \phi(X).A, \\ W^p &= \sum \beta_i^p.\phi(X_i^p) + \sum \beta_i^g.\phi(X_i^g) = \phi(X).B \end{aligned} \tag{2.8}$$

where $A = [\alpha_1^p..\alpha_N^p, \alpha_1^g..\alpha_N^g]$ and $B = [\beta_1^p..\beta_N^p, \beta_1^g..\beta_N^g]$. Each projection matrix, W^p or W^g, was modeled using data from both modalities, which was not adopted in CSR. A second enhancement over CSR was also proposed, in the same work, by making A and B parameter vectors preserving local consistency. The objective function of the latter constraint was formulated as $A^T.L.A + B^T.L.B$, where L is the Laplacian matrix over the samples. Hence, the new global objective function that aims to optimize A and B and optimizes W^g and W^p becomes:

$$\begin{aligned} J = (1/N).\|Y^g - W^g.X^g\|^2 + (1/N).\|Y^p - W^p.X^p\|^2 \\ + \eta.\|W^g - W^p\|^2 + \lambda.(A^T.L.A + B^T.L.B), \end{aligned} \tag{2.9}$$

The study reported a superiority of ICSR over CDFE, LCSR, and KCSR methods. In [68] researchers mention that the above objective functions defining CSR or ICSR do not consider the class relationship during the regression from the face space to the low-dimensional embedding space. Class relationship refers to the empirical separability used in CDFE. Thus, they proposed a new method, called discriminative spectral regression (DSR), where two new terms defining the class separability are included into the objective function which becomes:

$$\begin{aligned} J = \sum_i^R \sum_j (1/l_{ij}) \|Y_j^i - W^i.X_j^i\|^2 + \lambda. \sum_{s.c} \|W^i.X_j^i - W^k.X_l^k\|^2 \\ + \lambda. \sum_{d.c} \|W^i.X_j^i - W^k.X_l^k\|^2 + \lambda. \sum_i^R \|W^i\|^2, \end{aligned} \tag{2.10}$$

The abbreviations 's.c' and 'd.c' denote 'same class' and 'different class,' and the corresponding terms ensured minimizing differences between low-dimensional representations of the same class and maximizing those between different classes. 'R' stands for the number of modalities involved, which enabled DSR to deal with multi-modality (more than 2 modalities) problems. As in ICSR, the authors proposed two variations of DSR, linear

DSR (LDSR) and kernelized DSR (KDSR). The reported results showed a superiority of the two proposed methods over LCSR, KCSR and CDFE, with KDSR slightly outperforming LDSR. However, the DSR ability to deal with more than two modalities was not tested in the study. In [163] and [97], the authors defined projection matrices W^k maximizing the correlation between projected data corresponding to the same subject. While Yi et al. applied CCA at the image level [163] as in the global sparsity mentioned above, work in [97] adopted a patch-based CCA, where a different projection was defined for each image patch from the gallery database. Experiments were done for the HFB database using different patch sizes. The study reported a superiority of the patch based CCA over the global CCA, which emphasize again the effectiveness of patch based algorithms. In [80], a kernel prototype approach was proposed for heterogeneous face matching. The idea was to represent a face as a kernel similar to a collection of prototype images from the gallery database. Hence, for images from the two different modalities, X^p and X^g, two projection vectors were obtained as:

$$\begin{aligned} \phi(X) &= [K(X^p, X_1^p)...K(X^p, X_N^p)], \\ W^p &= [K(X^g, X_1^g)...K(X^g, X_N^g)] \end{aligned} \tag{2.11}$$

where $[X_1^i..X_N^i], i \in \{p,g\}$ are chosen prototypes and K is a kernel for the similarity measure. In practice the similarity measure between X^p and a prototype X_i^p is not as close as needed to that between X^g and X_i^g. The authors proposed to learn a transformation R that compensates for this difference and better aligns the probe similarities with the gallery similarities. Thus, projection vectors become:

$$\begin{aligned} \phi'(X^p) &= R.\phi(X^p), \\ \phi'(X^g) &= R.\phi(X^p) \end{aligned} \tag{2.12}$$

LDA and random subspace methods [64] were used for data dimensionality reduction, and a cosine similarity measure was used to match between $\phi'(X^p)$ and $\phi'(X^g)$. The proposed method called 'prototype random subspace'(or P-RS) was compared to an older method proposed by the same research team in [79], where LDA and random subspaces were applied directly on heterogeneous images without kernel prototype projection(D-RS). Comparison involved also the commercial matcher FaceVACS, and the obtained results have shown a superiority of P-RS and D-RS over FaceVACS both when applied separately or when results of P-RS and D-RS are fused at the match score level. In [87], the authors proposed to use LDA and random subspaces as in D-RS but with two main enhancements. First, covariance matrices were defined on the projected samples rather than the original samples as in D-RS. Second, kernelization and locality constraint were introduced in LDA learning as in ICSR. The proposed algorithm was evaluated against CDFE, LCSR, KCSR, ICSR and global CCA for the CASIA-HFB database. Although ICSR gave the best overall performance, the pro-

posed algorithm outperformed all the other algorithms.

Invariant Feature-Based Methods

The objective of invariant feature-based methods is to investigate the multitude of ex-isting image descriptors, either locally or globally, to build a feature vector as neutral as possible regarding images modalities. This approach typically involves three main processing steps: 1) Preprocessing of images (or appearance normalization), where images are normalized photometrically and geometrically and multiple filters are applied to enhance common features in heterogeneous images while reducing their differences. 2) Modality-invariant local and global (holistic) operators are applied to extract feature vectors from the preprocessed images. 3) Similarity is measured using these vectors. In [93] DoG filters were applied to normalize appearance between NIR and visible face images. DoG filters, like most differential operators in image processing, encode local structure and relationship between neighboring regions, which are invariant with spectral bands. A modified block-based version of the LBP operator, MB-LBP [94], can then be applied for feature extraction. To minimize redundancy from the obtained feature vectors and reduce their size, a gentle AdaBoost classifier followed by the LDA algorithm were used in [94]. The same system was later improved in [99], where the MB-LBP operator was replaced by three descriptors which are: Histogram of Gaussian (HoG), generalized Laplacian of Gaussian (GLoG) and scale-invariant feature transform (SIFT) [36]. The enhanced system showed better performance for the HFB database. Several authors presented a first attemp to perform face matching in the SWIR domain [75] [16] [17]. A number of comparative experiments were conducted including testing different SWIR spectral bands (1550nm, 1350nm, 1150nm, 950nm), several filters like DoG, CLAHE, SSRlog, SSRatan, and SSRlog or SSRatan followed by CLAHE, as well as many score fusion and image matching techniques like sum score fusion, LR based matching and Bayesian classifier using maximum liklihood or the maximum a posteriori hypothesis [144]. The proposed algorithms showed promising results against commercial matchers like FaceIt G8 and verilok.

SWIR spectral bands were further studied in [113], [114]. The authors showed that matching visible images against SWIR images is still challenging and produces lower performances compared to matching visible against NIR images. In their work, the magnitude and phase of a 16 Gabor filter, applied on data images, were encoded: magnitudes were encoded using simplified Weber local descriptor (SWLD) and uniform LBP, while Gabor phases were encoded using generalized LBP operator. The obtained three feature vectors were concatenated and matched against the gallery database using symmetric I-divergence

distance [34]. In [162] NIR and visible images were encoded using the Laplacian of Gaussian filter (LoG). The filtered images were converted into binary images that are locally partioned into small patches. Distances between corresponding patches from both modalities were measured using the Hamming distance. Li proposed to learn a linear filter that best fits with LBP operators [36]. To the best of our knowledge this is the first work that investigated filter learning for LBP-based heterogenous face matching. The image filter vector ω was constrained to maximize the ratio of the between class scatter matrix to the within class scatter matrix. The latter two matrices were defined for the LBP extracted features as in [87]. Extracted features were matched us-ing the LCSR algorithm seen above. From the above algorithms, several techniques, like LDA and Gentle AdaBoost were proposed to overcome the problem of feature vectors' high dimension that generally characterize invariant feature based methods. However, these techniques may cause system overfitting due to the small number of training images for each class/subject, often encountered in face recognition tasks. To solve this issue, Klare and Jain proposed to use random subspaces method [79] where sub-feature vectors were randomly sampled and used to compute the LDA projection matrix.

2.3 Conclusion

The literature review of research works that investigated MI for face recognition (see Table. 2.6 for a summary of reviewed works) highlighted both the effectiveness and the complexity of this approach. Involving spectral bands at different wavelengths introduce several challenging problems not encountered in the monospectral case. The main of these challenges are:

- The involvement of hundreds of spectral bands raise the problem of high data redundancy/diemnsionality.
- Matching between spectral bands from different modalities (spectrums) requires specifically tailored algorithms and optimization tools to guarantee a reasonable performance
- The investigation of multi-sources/multi-sensors information requires the development of adequate non-usual fusion techniques.

While the two last issues encountering the use of MI are being solved (see sections 2.2.1 and 2.2.2), the first issue related to data dimensionality is still by far a challenging problem for researcher.

In the next section, we present the main approaches we have developed to solve the problem of data dimensionality when using MI images. We propose several systems to select the best spectral bands from a given face database and under a given imaging conditions. Our systems for best spectral bands selection (BSBS) are either static with the same set of bands is chosen for all subjects (see chapter 3) or dynamic with different bands are selected for each new subject (see chapter 4).

Table 2.6 Reviewed papers categorized by topic

Name (s)	**Categories**	**Descriptions**
[131]	Multispectral image preprocessing	Gamma intensity correction (GIC) and quotient illumination relighting techniques (QIR) were applied to control the overall brightness of images
[114]		Applied the logarithm transform to increase the contrast of SWIR images
[10] [3] [142]		Histogram Equalization (HE), contrast limited adaptive Histogram equalization (CLAHE), single scale retinex with logarithmic (SSRlog) and with arctangent (SSRatan) and single scale self-quotient (SQI) were used for contrast enhancement.
[157], [131]		Applied Block based and Adaptive Histogram equalization (BBHE and AHE)
[74]		Applied wavelet transform to decompose images on several sub-bands, and AHE was applied on each sub-band with dif-ferent sizes of local windows
[28]		Applied Discrete cosine transform DCT to compensate for il-lumination variation.
[170]		Applied adaptive thresholding based on histogram analysis and projection profile analysis for face detection

Continued on next page

Table 2.6 – continued from previous page

Name (s)	Categories	Descriptions
[16]		Detected faces and eyes for SWIR, MWIR and Visible images automatically using Viola and Jones AdaBoost algorithm. For SWIR and NIR images captured at distance greater than 60 meters, eye centers were manually annotated. Faces are aligned automatically using affine transformation like rotation, scaling and translation.
[135]		Compensate the linear and non-linear deformation between visible and LWIR images by minimizing the Normalized Mutual Information (NMI) between both modalities.
[169]		Applied NMI and the Fourrier-Mellin transform for translation, scaling, and rotation registration between different modalities.
[23]	Band Selection	Used techniques like PDF estimation (KDE) and Jeffrey divergence calculation for band selection, Combinations of selected bands were tested against each other and against broad band images.
[41]		$(2D)^2$ PCA and weighted fusion was used for feature extraction. Bands were selected based on the absorption characteristic of two chemical compounds of the human skin: Melanin and hemoglobin
[125]		PCA was used to measure discrimination ability of each spectral band, three PCA-based algorithms were developed and results showed a superiority of using MI over grayscale images.
[132]	Feature level fusion	NIR and visible images were fused based on extracted LBP features and adapted weighted fusion. Weights were determined based on pixels' salient values and images standard deviation.

Continued on next page

Table 2.6 – continued from previous page

Name (s)	Categories	Descriptions
[145]		IR and visible images were decomposed using Meyer wavelet filter. Gabor filter was used for features extraction. Both modality features were combined into one vector and the Fisher face technique was used for classification.
[108]		
[108]		Pixels' coefficient based activity levels (CBA) were computed and weighted fusion was applied. Weights were determined for each pixel based on its CBA. SVM with direst acyclic graph was used for classification.
[38]		PCA, KPCA, FLD, and KFLD were used for feature extraction. Weighted fusion was used and weights were determined using genetic algorithm.
[117]		Determined fusion weights in the WDT domain using E.coli bacteria foraging strategy.
[164]		LWIR and visible images were fused at WDT domain. PCA and Zernike moments were applied on fused images for classification.
[124]		Particle swarm optimization PSO was proposed to fuse visible and infrared images in the DWT domain.
[167]		LBP feature vectors from different modalities were concatenated and used to train a boosting classifier for face detection.
[23]	Image level fusion	NIR and visible images were fused. Pixel weights were determined based on physics properties of the LCTF filter used in the imaging system.
[117]		Weighted fusion was applied and fusion weights were deter-mined using the gray level co-occurrence matrix(GLCM).

Continued on next page

Table 2.6 – continued from previous page

Name (s)	Categories	Descriptions
[9]		Fusion weights were chosen experimentally based on the corresponding accuracy. Several couples of weights $(w_v is, w_t herm)$ were tested: (0.7, 0.3), (0.5,0.5) and (0.6,0.4).
[136]		Proposed a mv-SVM classifier with granular computing and soft labeling to define image fusion level as well as used image modalities dynamically chosen.
[20], [155]	Match Score level fusion	Match scores for a gallery image Y were defined as the dif-ference between Y and a sparsely projected version of Y for an over¬complete dictionary. Match scores from both modalities of Y were summed.
[109]		Fused multi-biometrics data, faces, iris, and fingerprints, using a likelihood ratio (LR) based method. Densities of genuine and impostor sets of match scores that define LR were estimated using a Gaussian mixture model GMM.
[122]		Match scores from both modalities, defined as the Euclidian distance between probe and gallery images, were summed. K-NN was applied on fused distance for classification.
[106]		Max, min, average, and product fusion rules as untrained rules, as well as SVM were used to fuse match scores from visible and synthesized NIR images.
[21]		Weighted fusion was performed, where match scores of each image modality, defined as Euclidian distance, were deter-mined based on their saliency. Saliencies were extracted from the match scores estimated density distributions.
[146]		Power series model was used to fuse match scores from different sources.
[4]		Fusion weights for thermal and visible match scores were determined using probability theory.

Continued on next page

Table 2.6 – continued from previous page

Name (s)	Categories	Descriptions
[57]	Decision level fusion	The sum of evidences coming from thermal and visible classifiers determined the total belief on an image Y to belong to class j. Modality evidence was determined using a voting scheme.
[130]		Minimum ranking fusion, product ranking fusion, and average ranking fusion were evaluated for the OTCBVS database.
[31]		SVM classifiers were trained on LWIR and visible images to determine the probability with which a probe image belongs to a defined class. The latter probability had the meaning of decision evidence. Fuzzy logic was used to aggregate evidence from both modalities into one decision.
[112]		The belief on a decision was modeled as a decreasing function of the distance between gallery and probe images. Dempster and Shafer theory of evidence was used to fuse obtained beliefs.
[135]		Final decision was obtained by summing decision beliefs obtained by Desert Smarandache theory.
[26]	Common space based methods	Synthesized visible images from NIR images by preserving local geometric relationships between images patches.
[166]		Synthesized visible images from NIR images using sparse projection of NIR patches (local sparsity) in NIR learned dictionaries. Sparse projection was transferred to the visible patches for the visible learned dictionary.
[101]		Synthesized visible images from NIR images using sparse projection of NIR images for the NIR gallery (global projection).Sparse projection was transferred to the visible images.

Continued on next page

Table 2.6 – continued from previous page

Name (s)	Categories	Descriptions
[154]		A modality invariant operator was determined for visible im-age synthesis using a Lambertian model of the faces. The operator parameters, data mean and covariance, were determined using 'best-match' logic.
[96]		The same modality invariant operator as in the above work were determined. However, parameters were determined using an averaging process of total data means and covariances.
[95]		Defined a common discriminate feature extraction (CDFE) algorithm to project data into a common space. Empirical separability and local consistency constraints were included into the algorithm's objective function.
[86]		Defined the coupled spectral regression (CSR) approach to project images modalities in a common space. Shrinkage as in ridge regression, and local consistency were included into the algorithm's objective function.
[88]		Proposed an improvement of the CSR algorithm (ICSR). Projection matrices' learning was made coupled from both modalities. Kernelization and local consistency were included into the algorithm's objective function.
[68]		A discriminative spectral regression (DSR) algorithm was proposed to enhance CSR and ICSR by including the class separability constraints into the algorithm's objective function.
[163], [97]		Projection matrices were constrained to maximize the corre-lation between projected data corresponding to the same subject. Canonical correlation analysis CCA was used for correlation computation.

Continued on next page

Table 2.6 – continued from previous page

Name (s)	Categories	Descriptions
[80]		A kernel prototype approach was proposed, where faces were presented as the kernel similarity to a collection of prototype images from the gallery database. LDA and random subspace methods were used for data dimensionality reduction and a cosine similarity measure was used to match projected faces.
[87]		Projection matrices were estimated using LDA and random subspaces defined on the projected space.
[93]	Invariant feature based methods	DoG filter was applied to normalize appearance between NIR and visible face images. A modified block-based version of the LBP operator, MBLBP, was used for feature extraction and LDA was applied to reduce data dimensionality.
[75] [17] [16]		Several filters like DoG, CLAHE, SSRlog, SSRatan, and SSRlog or SSRatan followed by CLAHE, as well as many score fusion and image matching techniques like sum score fusion, LR based matching and Bayesian classifier using maximum likelihood or the maximum a posteriori hypothesis [144]were evaluated for several SWIR spectral bands.
[113], [114]		Magnitude and phase of a 16 Gabor filters, applied on data images, were encoded: magnitudes were encoded using simplified Weber Local descriptor (SWLD) and uniform LBP, while Gabor phases were encoded using generalized LBP operator. The obtained three feature vectors were concatenated and matched against the gallery database using symmetric I-divergence distance.
[162]		NIR and visible images were encoded using LoG. Filtered images were converted into binary images that are locally partioned into small patches. Distance between corresponding patches from both modalities were measured using the Hamming distance.

Continued on next page

Table 2.6 – continued from previous page

Name (s)	Categories	Descriptions
[89]		A linear filter that best fits LBP operators was learned for matching purpose. An image filter vector was constrained to maximize the ratio of the data between the class scatter matrix and the within class scatter matrix. The two matrices were defined for the LBP extracted features.

Chapter 3

Static Best Spectral Bands Selection

3.1 Introduction

We say that the set of best spectral bands is chosen statically when the same spectral bands are selected for all subjects provided by the Face database. Static means non-dymanic, and the same learned set of best bands is used with all the new subjects (test subjects) for face matching. In the literature, the existent few systems for BSBS are static [25] [115]. This has the advantage to reduce the processing time of the overall system by selecting the best bands only one time during the training phase.

In this chapter, we present four new algorithms that belongs to the category of static systems for BSBS. The first two algorithms are based on a sparsity assumption and formulate the problem of BSBS as a pursuit problem. The third approach used a linear decomposition of the multispectral images (called also cubic images) to determine which spectral bands are less affected by the imaging conditions including particularly high illumination variation. The final approach assigned to each spectral band a week classifier. Week classifiers are then boosted using the Adaboost technique and the spectral bands corresponding to the classifiers with the greatest weights were selected as being the most useful.

3.2 Spectral Bands Selection Based On A Sparsity Assumption

3.2.1 Math Background

In this section, math tools used for our sparsity assumption based BSBS systems are detailed. These tools are the basis pursuit optimization algorithm and the kernel trick.

The Basis Pursuit Problem

The basis pursuit problem (BP) defined in Eq. 3.1 is an optimization problem that belongs to the family of second-order cone programs (SOCPs). In this problem, the unknown variable W is constrained to be sparse with most of its elements are zeros. This condition is ensured by the minimization of the l_1 norm of W. BP is mostly used in signal processing to recover sparse vectors from a small number of linear measurements [7].

$$\begin{aligned} W = \underset{W}{\operatorname{argmin}} \quad & \|W\|_1 \\ \text{subject to} \quad & \|A_{I^{prob}} \times W - B\|_2^2 \leq C \\ & for\ each\ image\ I^{prob}\ from\ the\ probe\ set \end{aligned} \tag{3.1}$$

In the following, we present the solution proposed by Becker et al. in [7] to solve the BP problem using the log-barrier method.

Beker: The problem of Eq. 3.1 could be written as:

$$\begin{aligned} W = \underset{W}{\operatorname{argmin}} \quad & < c_0, W > \\ \text{subject to} \quad & A_0.W = B \\ & f_{I^{prob}}(z) \leq 0 \quad for\ each\ image\ I^{prob}\ from\ the\ probe\ set \end{aligned} \tag{3.2}$$

where each f_i describes a second order conic constraint:

$$f_i(W) = (1/2).\|A_I.W\|_2^2 - (< c_i, W >_{d_i})^2$$

A_i are matrices, c_i are vectors and d_i are scalars. The standard log-barrier method transforms

Eq. 3.2 into a series of linearly constrained programs:

$$\begin{aligned} W = \underset{W}{\text{argmin}} \quad & < c_0, W > +(1/\tau^k) \sum_i -log(-f_i(W)) \\ \text{subject to} \quad & A_0.W = B \end{aligned} \tag{3.3}$$

The BP problem, now formulated as in Eq. 3.3, could be easily solved by few iteration of the Newton's method. At log-barrier iteration k, Netwton's method proceeds by forming a series of quadratic approximations to 3.3, and minimizing each by solving a system of equations. The quadratic approximation of the functional

$$f_0(W) =< c_0, W > +(1/\tau^k) \sum_i -log(-f_i(W))$$

around a point W is given by:

$$f_0(W + \Delta W) \simeq W + < g_W, \Delta W > +(1/2). < H_W.\Delta W, \Delta W >:= q(W + \Delta W)$$

where g_W is the gradiant:

$$g_W = c_0 + (1/\tau). \sum_i (1/-f_i(W) \bigtriangledown .f_i(W))$$

and H_W is the Hessian matrix:

$$H_W = (1/\tau). \sum_i (1/f_i(W)^2 \bigtriangledown .f_i(W). \bigtriangledown .f_i(W)^2) + (1/\tau). \sum_i (1/-f_i(W) \bigtriangledown^2 .f_i(W))$$

given that W is feasible (that $A_0.W = B$ in particular), the ΔW that minimizes $q(W + \Delta W)$ subject to $A_0.W = B$ is the solution to the set of linear equations:

$$\tau. \begin{bmatrix} H_W & A_0^T \\ A_0 & 0 \end{bmatrix} = -\tau.g_W \tag{3.4}$$

The vector τ can be interpreted as the Lagrange multipliers for the quality constraints in the quadratic minimization problem. As A_0 is equal to zero (no equal constraint), the system 3.4 is symmetric positive definite, and thus can be solved using convergent gradiant (CG) when the problem is "large scale". Once ΔW, the Newton step direction, is determined, we need to determine the step lenght s. s is chosen so that:

$$f_i(W + s.\Delta.W) \leq 0 \; for \; all \; i = 1..M$$

the functional has decreased sufficiently:

$$f_0(W + s.\Delta.W) \leq f_0(W) + \alpha.W.\Delta.W. < g_W, \Delta.W >$$

where α is a user specified parameter and M is the number of quadratic constraints. The final W^k obtained determines the solution of the BP problem.

The Kernalization Trick

Real world data is usually non linearly separable; the data is generally collected from different sensors (multi-sensors data) presenting different properties (amplitude, phase, variance,etc). The descrimination between each kind of data is hence chalenging and at the same time indisponsble for any further processing. One solution to this problem is to map the original data to a second higher dimensional space where its linearly separable. Call ϕ the mapping function. However, when the mapped data is to be used with a machine learning algorithm involving inner products, the high dimensionality of the new space is an obstacle and the processing time increases exponentially. To solve this problem, researchers have proposed to manually construct ϕ such that the inner product of two mapped feature vectors X and X' could be expressed as

$$< \phi(X).\phi(X') >= k(X, X')$$

The Kernel trick essentially is to define k in function of original space such that k has a large number of dimensions to ensure linear separability into the new space and simultaneously could be determined without having to know the explicit expression of ϕ. Several families of kernels have been investigated including:

- **Polynomial kernel**

$$k(x,z) = (< x,z > +\theta)^d,\ ford \leq 0$$

- **Radial Basis function**

$$k(x,z) = \exp \frac{\|x-z\|^2}{2.\sigma^2},\ \|x\| = \sqrt{< x,x >}$$

- **Sigmoid kernel**

$$k(x,z) = tanh(\eta. < x,z >) + \theta)$$

In the next sections we use the most known kernel which is the radial basis kernel or called also the Gaussian kernel.

3.2.2 The Straight Sparsity Based Approach

In this section, we present our new approach to select the best spectral bands using a sparsity assumption [13]. As mentioned above, each image cube is formed by several spectral bands (say N) captured from different parts of the light spectrum. However, not all of these spectrums are useful for recognition purposes, and some of them may include significant amount of redundancy with respect to each other and the lighting conditions. To cope with this problem, optimal spectral bands are to be selected. If we consider I_p and I_g two multispectral images from the probe and gallery sets respectively, and $I_{p,i=1..N}$ to be the i^{th} spectral band of I_p , then the most intuitive method of measuring the difference (The distance) between I_p and I_g is to take a sum the differences between feature vectors extracted from the corresponding spectral bands, as shown in Eq. 3.5:

$$Ms(I_p, I_g) = Ms(V_p, V_g) = \sum_i .w_i.Ms(V_p^i, V_g^i) \tag{3.5}$$

The symbol Ms designates the match score measure between the two images I_p and I_g. Ms was determined as the inverse of the City Block distance normalized to be between 0 and 1. The later was experimentally found to give better results than the Chi-Square or Euclidian distances. V^i designates the feature vector extracted from spectral band $i = 1..N$ using a pre-defined algorithm for feature extraction like Multiblock Local Binary Pattern (MBLBP), Histogram of Gabor Phase Patterns (HGPP), Local Gabor Binary Pattern Histogram Sequence (LGBPHS),etc. The superscript i represents the spectral band used, while $w_i = 1 \ldots N$ are weights that determine the contribution, and hence the importance, of each spectral band in the final difference computed between the two images. In the sparsity approach, we assume that the vector of weights $W = [w_1, w_2..w_N]$, is sparse, with most of its elements are zeros, and could thus be optimized by solving the basis pursuit problem (P) defined bellow:

$$\begin{aligned} (P) : W = \underset{W}{\operatorname{argmin}} \quad & \|W\|_1 \\ \text{subject to} \quad & \|A_{I^{prob}} \times W - B\|_2^2 \leq C \\ & \textit{for each image } I^{prob} \textit{ from the probe set} \end{aligned} \tag{3.6}$$

$A_I = [Ms(V_I^j, V_i^j)]_{ij}$ is the matrix of match scores between the probe image I and all the gallery images $V_i = 1..M$, where M is the number of gallery images. For example, the row number i of A_I contains match scores between the image I and the gallery image I_i in each spectral band j (j spans the columns of A_I). We can see that each element of the vector $A_I.W$ is in fact the sum of match scores between I and the corresponding gallery image over all spectral bands as defined in Eq. 3.5. Finally, C is an empirically defined coefficient that controls the tolerated error from the ideal case. As only one image from the gallery database correspond to our probe image I, say the galley image k, then ideally each vector should be of the form $A_I.W = B = [0,..0,1_k,0..0] \in \mathbb{R}^M$, where all its elements are zeros except the k^{th} element, which is 1 (the maximum match between the two corresponding gallery and probe images). Hence, to minimize the recognition error in the non-ideal case, we constrain W to minimize the difference between $A_I.W$ and B for all probe images I. As described before, the l_1 norm is used to ensure a sparse solution for W. Experiments we conducted have shown that the solution of problem (P), namely W, is not generalizable, i.e. for a new images (test images), recognition results decrease significantly. We say that our system has overfitted the training data. The problem of overfitting is often encountered in optimization problems, and several approaches have been proposed to solve it. One of these approaches is to relax the problem constraints (in our case l_2 norm inequalities) by injecting slackness variables ($\zeta_i = 1..M$) into (P). This approach has been successfully applied with SVM classifier. In our method, we use the same approach to improve the generalizability of our algorithm. To the best of our knowledge, this is the first time slackness variables have been used to relax a pursuit problem. Hence problem P becomes:

$$\begin{aligned} (RelaxedP) : W = \underset{W,\zeta}{\operatorname{argmin}} \quad & (\|W\|_1 + \lambda.\|\zeta\|) \\ \text{subject to} \quad & \|A_{I^{prob}}.W - B\|_2^2 \leq C + \zeta_{I^{prob}} \\ & for\ each\ image\ I^{prob}\ from\ the\ probe\ set \end{aligned} \tag{3.7}$$

The injection of slackness variables into the inequality constraints enabled the latter to be violated to some error ζ_I, which enlarges the solution space and hence ensures that W will not fit the training data. In turn, the sum of all slackness variables (sum of tolerated error) should be minimized. This has been ensured by adding the term $\|\zeta\|_1$ to the objective function of problem (P). Finally, the parameter λ controls the tradeoff between $\|\zeta\|_1$ and $\|W\|_1$. Several Matlab libraries have been proposed to solve pursuit problems. In our work, we have used the cvx library.

One great advantage of our algorithm over existent algorithms like the one proposed by

Chang et al. in [25] for example, is the following: in the existent systems for BSBS, the recognition performance of each spectral band is studied separately. This prevents us from knowing how these spectral bands will behaves when fused with each other. In our algorithm, the performance of each spectral band is studied within the whole set of spectral bands, which is ensured by the weighting coefficient vector W. To illustrate this point, suppose that a given spectral bands A and B were found, by the Chang et al. algorithm, to be less discriminative then the spectral band C. The Chang et al. algorithm cannot tell us whether we could find a linear combination of A and B, say $w_1.A + w_2.B$, that is more discriminative then C and thus, whether spectral bands A and B should be selected instead of C and combined with the same coefficients w_1 and w_2. This mutual relation between spectral bands is ensured in our algorithm.

Sparsely decomposing a multispectral image on the space of spectral bands is, as any other projection based approach, affected by the intrinsic similarity of the original data; the projection of a data set presenting a high similarity between its elements, (or a low variances), is usually inefficient. To solve this problem, the original data set should be projected to a second space, usually a higher dimensional space, where the similarity between its elements is increased. This projection has been effectively implemented using the kernel trick. We propose to enhance the performance of the straight sparsity based approach using the kernalized projection. The new algorithm is hereafter referred as the kernalized sparsity based approach.

3.2.3 The Kernalized Sparsity Based Approach

Projecting vectors from their original space to a second space where they are easily classified is an effective approach in machine learning. Since the dimension of the new space may be high enough to make computation impossible. Aizerman et al. in [2], proposed to use a kernel function based trick where the product of projected images of vectors X and Y, namely $\phi(X).\phi(Y)$, is reduced to the inner product $\phi(X).\phi(Y) = K(X,Y) =< X,Y >$, ϕ is the mapping function, while K is the kernel function. Hence, the discrimination between projected vectors is enhanced due to the mapping function, while the computation cost remains moderate since the inner product is computed in the original space. We propose to use the asset of kernel trick to enhance our algorithm for BSBS [14].

As defined in section 3.2.2, the (relaxed P) optimization problem involves the computation of matrix A_I , which in turn involves the computation of l_2 distances between feature vectors extracted from the probe and gallery images. We propose to use the Gaussian kernel (ϕ) to

map the extracted feature vectors to the new kernel space. The kernel trick is then applied to compute the new kernalized match score matrix $A_{I-kernel} = [Ms(\phi(V_I^j), \phi(V_i^j))]_{ij}$. In this later, the inner products between mapped feature vectors, used for the computation of the Euclidian distances defining $A_{I-kernel}$ are performed as in Eq. 3.8. Note that superscripts p and g stands for probe and gallery images respectively.

$$\begin{aligned} Ms(\phi(V^p), \phi(V^g))) = \|\phi(V^p) - \phi(V^g)\|^2 = \phi(V^g)^T.\phi(V^g) + \phi(V^p)^T.\phi(V^p) \\ - 2.\phi(V^p)^T.\phi(V^g) = K(V^p, V^p) + K(V^g, V^g) - 2.K(V^g, V^p) \end{aligned} \tag{3.8}$$

The new optimization problem becomes:

$$\begin{aligned} &(kernalizedP): \ W = \underset{W,\zeta}{\operatorname{argmin}}(\|W\|_1 + \lambda.\|\zeta\|) \\ &\text{subject to} \qquad \|A_{I-kernel}.W - B\|_2^2 \leq C + \zeta_{I^{prob}} \\ &\qquad\qquad \|W\|_2^2 = 1 \ for\ each\ image\ I^{prob}\ from\ the\ probe\ set \end{aligned} \tag{3.9}$$

The constraint $\|W\|_2^2 = 1$ on the norm of W ensures that the elements of W keep moderate values and do not become very high. Once the vector of weights W is determined, the spectral bands with the greatest weights are selected as best spectral bands and fused. Feature vectors extracted from the fused images are then mapped to the kernel space using the same kernel function ϕ and matched using the Euclidian distance. The motivation behind mapping extracted feature vectors to a third space is to maximize the similarity between vectors corresponding to the same subject, also called within class covariance S_w , while minimizing the similarity between vectors from different subjects, called the between class covariance S_b . This minimization-maximization process is known in the literature as Fisher criterion [43]. Comparing the Fisher criterion of both non-kernalized and kernalized matrices of match scores, namely A_I and $A_{I-kernel}$ is, hence, a good tool to evaluate the efficiency of our kernalization process: a successful kernalization should increase the Fisher criterion of the matrix of match scores. This later fact would, in turn, make the optimized vector of weights W sparser and ensures as well a faster convergence of the (Kernalized P) problem. To make the Fisher criterion adequate to evaluate the efficiency of our kernalization process, we make a first modification of the former by including the $l_2 norm$ as shown in Eq. 3.10:

$$J_I = \left(\frac{\|S_w^I\|_2}{\|S_b^I\|_2}\right)^2 \tag{3.10}$$

The modified Fisher criterion J_I is computed for each image I by considering all of its N spectral bands as being from the same class as well as the N spectral bands of the subject

corresponding to I from the gallery database. The other gallery multispectral images are then considered as being from the other classes (subjects) and are used to compute the between class matrix S_b^I.

By expressing the between class and within class matrices using the entries of matrix A_I , namely:

$$\begin{aligned} \|S_w\|_2^2 &= \sum_{j=1..N} Ms(V_I^j, V_I^j)^2 = \sum_{i=1..N} (A_I)_{Ij}^2 \\ \|S_w\|_2^2 &= N^2 . \sum_{i=1..N, i\neq I} \sum_j . Ms(V_I^j, V_i^j)^2 = N^2 . \sum_{i=1..N, i\neq I} \sum_j . (A_I)_{ij}^2 \end{aligned} \tag{3.11}$$

we can reduce J_I to the simple expression:

$$J_I(A_I) = (\frac{\|S_w^I\|_2}{\|S_b^I\|_2})^2 = \frac{\sum_{i=1..N} (A_I)_{Ij}^2}{N^2 . \sum_{i=1..N, i\neq I} \sum_j . (A_I)_{ij}^2} \tag{3.12}$$

In Eq. 3.11, the expression of between class and within class matrices considered only the relation between feature vectors from the same spectral band. The relations between feature vectors from different spectral bands, even when being within the same class are ignored since they don't figure in the expression of A_I and hence do not affect our evaluation. This simplification reduce significantly the processing time during the computation of J_I.

Schematically, $J_I(A_I)$ is the ratio of sum of the square of elements inside the green region in Fig. 3.1, by the sum of square of elements inside the red region in the same figure.

$$A_I = \begin{bmatrix} Ms(V_I^1, V_{I_1}^1) & Ms(V_I^2, V_{I_1}^2) & \cdots & \cdots & Ms(V_I^{25}, V_{I_1}^{25}) \\ Ms(V_I^1, V_I^1) & Ms(V_I^2, V_I^2) & \cdots & \cdots & Ms(V_I^{25}, V_I^{25}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Ms(V_I^1, V_{I_M}^1) & Ms(V_I^2, V_{I_M}^2) & \cdots & \cdots & Ms(V_I^{25}, V_{I_M}^{25}) \end{bmatrix}$$

Fig. 3.1 Between class (red region) and within class (green region) matrices for modified Fisher criterion computation. N=25

We can see that the match score matrix with the greatest $J_I(A_I)$ should has, by definition, the elements with the smallest values inside the red region (region of between class elements) and greatest values inside the green region (region of within class elements) . Hence, the highest the Fisher criterion of a matrix A_I is, the sparsest this matrix will be. However, the sparsity of A_I , despite being of great benefit for the computation of W , is not sufficient to

measure the effect of kernalization on A_I . We would like to ensure also that the elements inside the green region of this matrix are also sparse, which means most of them are zeros. This would increase both the convergence speed of the (relaxed P) problem and the sparsity of W. For this purpose, we propose a second modification on the expression of J by including a measure of the saparsity of the green row in Fig. 3.1. We refer to this row as $W(A_I)$ (within class elements of A_I), and we measure its sparsity using the $l_1 norm$. Hence, the Modified Fisher Criterion(MFC) J becomes:

$$J_I(A_I) = (\frac{\|S_w^I\|_2}{\|S_b^I\|_2} + \|W(A_I)\|_1^2)^2 = \frac{\sum_{i=1..N}(A_I)_{Ij}^2}{N^2 \cdot \sum_{i=1..N, i\neq I}\sum_j \cdot (A_I)_{ij}^2 + \|W(A_I)\|_1^2} \tag{3.13}$$

Using our discrimination measurement criterion MFC, we can prove the effectiveness of using the kernalization trick on the sparsity of the projected feature vectors. This sparsity, should enhance the performance of our BSBS algorithm and hence provide a better set of selected spectral bands.

3.3 Spectral bands Selection based on a Multilinear decomposition

3.3.1 Math Background

In this section, math tools used for our multilinear decomposition based BSBS system are detailed. The main tool used is the Domain Adaptive Dictionary Learning Algorithm (DADL).

Domain Adaptive Dictionary Learning Algorithm

The Domain Invariant Dictionary Learning (DADL) algorithm [123] is an algorithm that enables to decompose a 3-order tensor T(a cubic image for example), using 3-mode singular value decomposition (3-mode SVD), in the form of

$$T = Z \times U_{param1} \times U_{param2} \times U_{param3}$$

where param1, param2 and param3 are the problem parameters (for example lighting intensity, used spectral bands, etc). The particularity of DADL is that the mode matrices

U_{param1}, U_{param2}, *and* U_{param3} are all sparse matrices, i.e, most of their entries are zeros. The proposed algorithm begins by writing the decomposition of T in the form of flattened matrices:

$$T_{(3)} = D_{(3)}^{T_1}.U_{param1(3)}^{T_2}.U_{param2(3)}^{T_3}.U_{param3(3)}^{T_4}$$

where $T_{i=1..4}$ are vector transpose operators that ensure the agree of dimensions between multiplied matrices. Then, two iterative algorithms are proposed to learn the base dictionary D and the sparse matrices $U_{param1(3)} = [K_{i,j}] \in \Re^{N_1 \times N_1}$, $U_{param2(3)} = [S_{i,j}] \in \Re^{N_2 \times N_2}$, and $U_{param3(3)} = [I_{i,j}] \in \Re^{N_3 \times N_3}$. Hence, a new image $y_{k,i}^s$ with $param1 = k$, $param2 = i$ and $param3 = s$ could be written as

$$y_{k,i}^s = D_{(3)}^{T_1}.U_{param1(3)}^{T_2}.U_{param2(3)}^{T_3} \times S_s$$

where $D_{(3)}^{T_1}.U_{param1(3)}^{T_2}.U_{param2(3)}^{T_3}$ is the domain dictionary adapted to the param3 domain and S_s is the column vector number s of S corresponding to $param3 = s$. As we said, S_s determine the sparse decomposition of $y_{k,i}^s$ on the param3 adaptive domain dictionary. The authors have shown that for the same s, S_s is invariant when k and/or i changes, i.e, the decomposition of an image $y_{k,i1}^s$, captured at a different $param2 = i_1$ but for the same s, on the domain adaptive dictionary corresponding to s which is $D_{(3)}^{T_1}.U_{param1(3)}^{T_2}.U_{param2(3)}^{T_3}$, is the same, namely S_s. The same properties of domain invariance hold for the other domain matrices.

3.3.2 The Multilinear Decomposition Based Approach

The multilinear decomposition based approach that we propose [12] is based on the results obtained by Qui et al. in [123] (see section 3.3.1). By considering the decomposition of T as:

$$T = Z \times U_{subject} \times U_{spectralbands} \times U_{illumination} \tag{3.14}$$

The mode matrices $U_{subject}$, $U_{spectralbands}$ and $U_{illumination}$ are sparse matrices, i.e, most of their entries are zeros. The decomposition of T is written in the form of flattened matrices $T_{(3)} = D_{(3)}^{T_1}.U_{subject(3)}^{T_2}.U_{spectralbands(3)}^{T_3}.U_{illumination(3)}^{T_4}$. Then, two iterative algorithms are proposed to learn the base dictionary D and the sparse matrices $U_{subject(3)} = [K_{i,j}] \in \Re^{N_1 \times N_1}$, $U_{spectralbands(3)} = [S_{i,j}] \in \Re^{N_2 \times N_2}$, and $U_{illumination(3)} = [I_{i,j}] \in \Re^{N_3 \times N_3}$. N_1, N_2 and N_3 are respectively the number of subjects, spectral bands and illumination conditions provided by the face database. Hence, a new image $y_{k,i}^s$ of subject k at illumination i and spectral

band s could be written as

$$y^s_{k,i} = D^{T_1}_{(3)}.U^{T_2}_{subject(3)}.U^{T_3}_{illumination(3)} \times S_s \tag{3.15}$$

Where $D^{T_1}_{(3)}.U^{T_2}_{subject(3)}.U^{T_3}_{illumination(3)}$ is the domain dictionary adapted to the spectral bands domain and S_s is the column vector number s of S corresponding to spectral band s. As we said, S_s determine the sparse decomposition of $y^s_{k,i}$ on the spectral bands adaptive domain dictionary. We know that for the same spectral band s, S_s is invariante when the saubject k and/or the illumination i changes, i.e, the decomposition of an image $y^s_{k,i1}$, captured at a different illumination $i1$ but for the same spectral band s, on the spectral bands domain adaptive dictionary $D^{T_1}_{(3)}.U^{T_2}_{subject(3)}.U^{T_3}_{illumination(3)}$, is the same, namely S_s.

However, as shown by the authors of DADL (see Fig. 13 from [123]), and confirmed by our experimentations, the vector S_s varies slightly but continuously when varying the lighting condition. The authors of [123] considered this slight variation as a negligible error due to the iterative approximation process used by DADL to determine S_s. We show in this section that the error on S_s could be modeled as the sum of two errors: the mean square error dE due to the iterative process of DADL, and an error dK that quantifies the robustness of each spectral band against illumination variation. We propose to measure dK to determine the best spectral bands.

In section 3.3.1 , we have obtained the following expressions:

$$\begin{aligned} y^s_{k,i} &= D^{T_1}_{(3)}.U^{T_2,k}_{subject(3)}.U^{T_3,i}_{illumination(3)} \times S_s \\ \Rightarrow\ S_s &= (D^{T_1}_{(3)}.U^{T_2,k}_{subject(3)}.U^{T_3,i}_{illumination(3)})^{-T}.y^s_{k,i} \\ &= U^{T_a,i}_{illumination(3)}.U^{T_b,k}_{subject(3)}.D^{T_c}_{(3)}.y^s_{k,i} \end{aligned} \tag{3.16}$$

Where $T_a = T_1 \circ -T$, $T_b = T_2 \circ -T$ and $T_c = T_3 \circ -T$. Then dS_s could be written as:

$$\begin{aligned} dS_s|_{illumination} = {} & d(U^{T_a,i}_{illumination(3)})|_{illumination}.U^{T_b,k}_{subject(3)}.D^{T_c}_{(3)}.y^s_{k,i} \\ & + U^{T_a,i}_{illumination(3)}.d(U^{T_b,k}_{subject(3)})|_{illumination}.D^{T_c}_{(3)}.y^s_{k,i} \\ & + U^{T_a,i}_{illumination(3)}.U^{T_b,k}_{subject(3)}.D^{T_c}_{(3)}.d(y^s_{k,i})|_{illumination} \end{aligned} \tag{3.17}$$

We define dE and dK as:

$$\begin{aligned} dE = {} & d(U^{T_a,i}_{illumination(3)})|_{illumination}.U^{T_b,k}_{subject(3)}.D^{T_c}_{(3)}.y^s_{k,i} \\ & + U^{T_a,i}_{illumination(3)}.U^{T_b,k}_{subject(3)}.D^{T_c}_{(3)}.d(y^s_{k,i})|_{illumination} \\ dK = {} & U^{T_a,i}_{illumination(3)}.d(U^{T_b,k}_{subject(3)})|_{illumination}.D^{T_c}_{(3)}.y^s_{k,i} \end{aligned} \tag{3.18}$$

Hence dS_s could be written as:

$$dS_s = dE + dK \Rightarrow dK = dS_s - dE \tag{3.19}$$

In Fig. 3.2, we determine the variation of dS, dE and dK. dS is determined by varying the illumination and subjects (all combinations of illumination and subjects are considered) and measures the average error on S_s, while dE is determined by varying only the illumination and keeping the subjects and spectral bands fixed. dS, dE and dK are determined for three spectral bands at 670nm, 710nm and 720nm provided by the $IRIS-M^3$ face database (see the Experimental part of this thesis). From Fig. 3.2 we can see that i) dK never become null (either for each combination (subject/illumination) or in average) and varies from one spectral band to another and that ii) dK is roughly constant. Hence, the value of dK could be used to characterize spectral bands. On the other hand, the expression of dK is only function of $d(U^{T_b,k}_{subject(3)})$, and hence should become null whenever the same subject is used. This contradictory results between theory and experiments could be explained as follows: The traits of a given subject at a given spectral band are affected by illumination variation so that for a given recognition system, the identity of that subject is like being changed and hence $d(U^{T_b,k}_{subject(3)})$ does not vanish. This explanation is consistent with the roughly constant value of dK; the subject variation due to illumination happens only one time and without reversibility. Easily, we can see from the expression of dK, that the spectral band with the smallest dK is the less affected by illumination and hence the best for face recognition. we call dK the robustness to illumination factor or RIF. By computing the RIF of all spectral bands, we can determine the spectral bands the less sensitive to illumination variation.

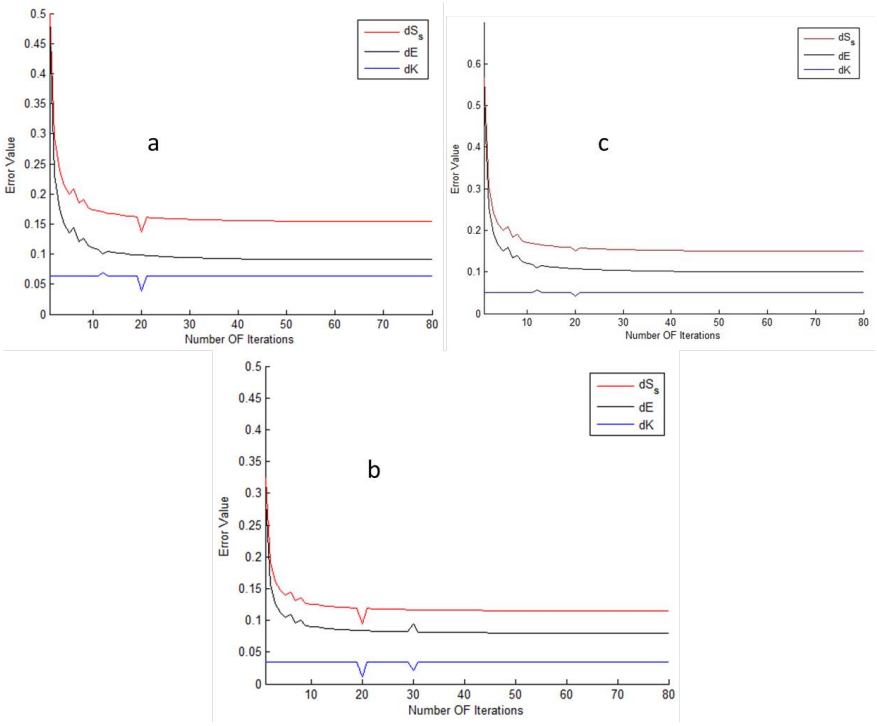


Fig. 3.2 Variation of different errors with the number of iterations for Spectral bands a) SB24, b) SB25 and c) SB20

3.4 Spectral bands Selection based on the classifiers Boosting approach

3.4.1 Math Background

In this section, math tools used for our boosted LDA based BSBS system are detailed. The main tool is the Boosting algorithm.

The Boosting Algorithm: The Strength Of Union

Since it was first proposed by Freund and Schapire in [48], the Adaboost algorithm has shown a particular effectiveness to build strong predictive models (classifiers) from weak classifiers. Despite several variety of the algorithm have appeared since then, the heart of the Adaboost approach remained the same. Given a training data set $(x_1, y_1)...(x_m, y_m)$ where $x_{i=1..m}$ are elements of a domain space X and y_i are the corresponding labels in a

some label set Y. Y is generally the binary set $-1,1$. Let D be a given distribution and $D_t(i)$ the weight of this distribution on training example i on round t. For each round $t = 1..T$, the Adaboost algorithm learn a week hypothesis $h_t : X \rightarrow Y$ that map the training data to the right labels while preserving a minimum error ε_t. A new week hypothesis is computed at each round t and a new error ε_t is determined. The Adaboost algorithm varies the values of D_{t+1} so that to increase the weights of the incorrectly classified data at round t. This variation of data weights has the aim to force the weak classifier at round t to focus on the still challenging samples of the data set. The final classifier H is obtained by a weighted sum of the learned week classifiers (hypothesis):

$$H(x) = sign(\sum_{t}^{T} \alpha_t.h_t) \tag{3.20}$$

Here after we outline the diferent steps of the basic Adaboost classifier as proposed by Freund and Schapire in [48]:

Given $(x_1,y_1)...(x_m,y_m)$ where $x_i \in X$ and $y_i \in -1,1$
Initialize $D_1(i) = 1/m$
For t=1..T

1. Train week learned using the distribution D_t.

2. get weak hypothesis $h_t : X \rightarrow Y$ with error

$$\varepsilon_t = Pr_{i\sim D_t[h_t(x_i)\neq y_i]}$$

3. Choose $\alpha_t = ln(\frac{1-\varepsilon_t}{\varepsilon_t})$

4. Update

$$D_{t+1} = \frac{D_t i.exp(-\alpha_t.y_i.h_t(x_i))}{Z_t}$$

   Z_t is a factor chosen so that D_{t+1} will be a distribution

5. Output the final hypothesis

$$H(x) = sign(\sum_{t}^{T} \alpha_t . h_t)$$

There are several variants of the Boosting approach that have been proposed. These variants differ by the way the weak classifiers are learned.

Real AdaBoost The real Adaboost was first proposed by Friedman, Hastie and Tibshirani in [49]. The weak hypothesis h_t was modeled as:

$$h_t = 1/2.ln(\frac{x}{1-x})$$

LogitBoost LogitBoost is an application of the logistic regression to the Adaboost appraoch. Weak classifiers are constrained to minimize errors with respect to

$$z_t = \frac{y* - p_t(x)}{2.p_T(x)(1-p_t(x))}, \; where \; p_t(x) = \frac{exp(h_{t-1}(x))}{exp(h_{t-1}(x)) + exp(-h_{t-1}(x))} \; and \; y* = \frac{y+1}{2}$$

Gentle AdaBoost In this variation of the Boosting approach, the function h_t is chosen to minimize the quantity

$$\sum_{i} D_t(i).(y_i - f_t(x_i))^2$$

Finally, another variant of the Boosting approach will get a particular attention in this thesis which is the SAMME adaboost. The theory behing this approach will be highlighted and investigated in the next section.

3.4.2 Boosted LDA

In this section, we present our approach to select the optimal spectral bands called Boosted LDA. Each image cube is formed by N spectral bands ranging from $k = 1$ to $k = N$ the number of available spectral bands. In our approach, we associate a weak classifier H_k to each spectral band k and then, using the multi-class Adaboost approach, we search for the optimal combination of these weak classifiers into a final stronger classifier H. We constraint H_k to be a weak classifier that performs face recognition over the considered databases using only the corresponding spectral band k. Hence, if within the final expression of the stronger classifier H, a weak classifier H_m was affected a weight smaller than the

weight affected to a weak classifier H_n , we can deduce, based on the definition of the boosting process, that the spectral band 'n' associated with H_n had better performance over the considered database than the spectral band m associated with H_m. Formally, H could be written as:

$$H(I_p) = sign(\sum_k \omega_k . \tau(H_k(I_p = C))) \tag{3.21}$$

I_p is the probe image to be classified. $H_k(I_p = C)$ is the decision given by H_k for I_p to belong or not to a class C. In our case, this decision is made based on the normalized match scores between I_p and the gallery images forming the class C. The operator gives 1 if H_k decides that I_p belongs to C and 0 otherwise. This operator ensures the vote aspect of Eq. 3.21: the final decision or vote of H is a weighted vote given by all weak classifiers. As it is well established by the Adaboost approach [172], weights are directly related to the recognition performance of the corresponding weak classifier and hence to the corresponding spectral band: weak classifiers with higher recognition performances are given higher weights. Hence, their votes are more important. This linear combination of weak classifiers with the most effective classifiers/spectral bands that are given higher weights enables one to determine the best spectral bands as well as the optimal combination of the latter to ensure an error rate less than or equal to a fixed amount, $err_{tolerated}$.

The choice of the family of weak classifiers that may fairly reflect the quality of each spectral band is crucial for the success of our approach. These classifiers should fill out at least three properties to be used, which are:

1. Weak classifiers should be easy to set up: this condition is required by the Adaboost approach, where the final goal is to use a combination of simple classifiers to build a more complex and efficient classifier.

2. As required by the Adaboost approach, weak classifiers should also be parameterizable: we should be able to assign different weights of importance to the training data so that different configurations of the learned classifier can be obtained. This requirement will be clarified when we define our Adaboost algorithm.

3. The performance of each weak classifier H_k should be directly related to the quality of its corresponding spectral band k: the better the quality of a spectral band k is for recognition purposes, the better the performance of H_k should be.

While it is easy to verify the validity of the first and second properties for a given family of classifiers, the third property is not always obvious. We cannot determine whether the performance of a classifier increases with the quality of the corresponding spectral band unless we have prior knowledge about the quality of each spectral band. As a result, prior

knowledge is the final goal of our approach and, therefore, could not be achieved at this stage. To solve this problem, we have chosen to check for the presence of the third property based on the results presented by Chang et al. in [25] for the $IRIS-M^3$ database. In this work, the authors analyzed the quality of different spectral bands and proposed a band classification based on their effectiveness in a face recognition context. We propose to use results obtained from these two works to form prior knowledge about the quality of different spectral bands. Thus, a family of classifiers $H_k = 1..N$ is considered to fill out the third property if for each $(m,n) \in [1..N]^2$ we have the relation:

$$if\ SB_m \geq SB_n\ then\ H_m \geq H_n \tag{3.22}$$

Where SB_m designates the m^{th} spectral band, and the symbol $\leq$ stands for 'better than' and is determined based on results obtained from [25]. In other words, Eq. 3.22 could be read as follows: if the recognition quality of spectral band SB_m is "better than" the recognition quality of spectral band SB_n , based on [25] or [41], then the performance of H_m should be better than the performance of H_n . However, we point out here that the use of results from [25] to judge the quality of spectral bands is only proposed as a solution to choose our weak classifiers, and, hence, does not detract in any way from the importance of our approach to selecting best spectral bands. Actually, if a spectral band SB_l was considered, for example, by Chang et al. algorithm as having better recognition quality than SB_m and SB_n , Chang et al. algorithm still does not tell us whether we will find a combination $\omega_m.SB_m + \omega_n.SB_n$ that gives better performance than SB_l alone. In this case SB_m and SB_n should be kept instead of SB_l with the same combination. One advantage to our approach is that we can search for these combinations of best spectral bands.

Based on the above set of considerations that our weak classifier should meet, we have chosen to build the latter using two sample projection matrices W_g and W_p , a distance for match scores computation and a statistic optimization tool, which is LDA.

LDA Based Weak Classifiers

For two spectral images I_p^k and I_g^k from the probe and gallery databases respectively, we define the distance between these images as:

$$d(I_p^k, I_g^k) = d(V_p^k, V_g^k) = \|W_g^{k^T}.V_g^k - W_p^{k^T}.V_p^k\|_2^2 \tag{3.23}$$

Where $V^k_{i\in\{p,g\}}$ are feature vectors extracted from the probe and gallery images at wavelength k using a feature extraction method as will be explained in the experimental section of this thesis. Actually, for each algorithm, we determine the set of best spectral bands that maximizes its accuracy. The decision given by a weak classifier $H_k(I_p = C)$ for an image I_p to belong or not to belong to a gallery class C can be defined as:

$$H_k(I_p = C) = \begin{cases} True, & \text{if } d_k(I^k_p, I^k_{g_c}) = min_l\ d_k(I^k_p, I^k_l) \\ False & \text{otherwise} \end{cases}$$

From the last two equations, we can see that different projection matrices are defined for the gallery and probe image sets, as well as for each spectral band k. The superscript k that figures in both equations designates the spectral band used for decision computation. Finally, the subscript g_c stands for the gallery image forming the class C. The objective of using projection matrices is to project extracted feature vectors from their original space, space of the gallery and probe images, to a third space more convenient for the recognition purpose. In our work, we constraint the new space to have two properties: first, the covariance between feature vectors of the same subject, namely the within class covariance matrix S_w, should be minimized; second, the covariance between feature vectors of different subjects, namely the between class covariance matrix S_b, should be maximized. These two conditions are the heart of the linear discriminant analysis (LDA) method. We propose to use LDA to build our weak classifiers by learning the projection matrices defined in Eq. 3.23. LDA, being a simple and efficient tool for parameters optimization [87], maximizes the ratio of within class matrix divided by the between class matrix defined as:

$$J_k = \frac{S^k_b}{S^k_w} \tag{3.24}$$

To obtain the expression of J_k in function of W^k_b and W^k_w , we start by expressing S^k_b and S^k_w as in [87]:

$$\begin{aligned} S^k_w &= S^{gp,k}_w = S^{pg,k}_w \\ S^k_b &= S^{gp,k}_b + S^{pg,k}_b + S^{pp,k}_b + S^{gg,k}_b \end{aligned} \tag{3.25}$$

Where we have :

$$\begin{aligned} S^{gp,k}_b &= \sum_1^N C\alpha_i.(W^{k^T}_g.V^k_{g,i} - W^{k^T}_p.m^k_p)(W^{k^T}_g.V^k_{g,i} - W^{k^T}_p.m^k_p)^T \\ S^{gp,k}_w &= \sum_1^N C\alpha_i.(W^{k^T}_g.V^k_{g,i} - W^{k^T}_p.V^k_{p,i})(W^{k^T}_g.V^k_{g,i} - W^{k^T}_p.V^k_{p,i})^T \end{aligned} \tag{3.26}$$

with

$$m_p^k = \frac{1}{NC}.\sum_1^{NC} V_{p,j}^k$$

m_p^k and m_g^k are the mean vectors of modalities g and p for the spectral band k. By expanding Eq. 3.26 we obtain:

$$\begin{aligned}
S_b^{gp,k} &= \sum_1^N C.\alpha_i.(W_g^{k^T}.V_{g,i}^k - W_p^{k^T}.m_p^k)(W_g^{k^T}.V_{g,i}^k - W_p^{k^T}.m_p^k)^T \\
&= \sum_1^{NC} \alpha_i.W_g^{k^T} V_{g,i}^k.V_{g,i}^{k^T}.W_g^k + \sum_1^{NC} \alpha_i.W_p^{k^T} V_{p,i}^k.V_{p,i}^{k^T}.W_p^k \\
&- \sum_1^{NC} \alpha_i.W_g^{k^T} V_{g,i}^k.V_{p,i}^{k^T}.W_p^k - \sum_1^{NC} \alpha_i.W_p^{k^T} V_{p,i}^k.V_{g,i}^{k^T}.W_g^k \\
&= W_g^{k^T}(\sum_1^{NC} \alpha_i.V_{g,i}^k.V_{g,i}^{k^T}).W_g^k) + W_p^{k^T}(\sum_1^{NC} \alpha_i.V_{p,i}^k.V_{p,i}^{k^T}).W_p^k) \\
&- W_p^{k^T}(\sum_1^{NC} \alpha_i.V_{p,i}^k.V_{g,i}^{k^T}).W_g^k) - W_g^{k^T}(\sum_1^{NC} \alpha_i.V_{g,i}^k.V_{p,i}^{k^T}).W_p^k) \\
&= W_g^{k^T} L_{gg}^k.W_g^k + W_p^{k^T} L_{pp}^k.W_p^k - W_p^{k^T} L_{pg}^k.W_g^k - W_g^{k^T} L_{gp}^k.W_p^k \\
S_w^{gp,k} &= \sum_1^N C.\alpha_i.(W_g^{k^T}.V_{g,i}^k - W_p^{k^T}.V_{p,i}^k)(W_g^{k^T}.V_{g,i}^k - W_p^{k^T}.V_{p,i}^k)^T \\
&= \sum_1^{NC} \alpha_i.W_g^{k^T} V_{g,i}^k.V_{g,i}^{k^T}.W_g^k + \sum_1^{NC} \alpha_i.W_p^{k^T} m_p^k.m_p^{k^T}.W_p^k \\
&- \sum_1^{NC} \alpha_i.W_g^{k^T} V_{g,i}^k.m_p^{k^T}.W_p^k - \sum_1^{NC} \alpha_i.W_p^{k^T} m_p^k.V_{g,i}^{k^T}.W_g^k \\
&= W_g^{k^T} P_{gg}^k.W_g^k + W_p^{k^T} P_{pp}^k.W_p^k - W_p^{k^T} P_{pg}^k.W_g^k - W_g^{k^T} P_{gp}^k.W_p^k.
\end{aligned} \tag{3.27}$$

Where

$L_{pg}^k = \sum_1^{NC} \alpha_i V_{p,i}^k.V_{g,i}^{k^T}$, $P_{gg}^k = \sum_1^{NC} \alpha_i V_{g,i}^k.V_{g,i}^{k^T}$, $P_{pp}^k \sum_1^{NC} \alpha_i m_p^k.m_p^{k^T}$, $P_{pg}^k = \sum_1^{NC} \alpha_i m_p^k.V_{g,i}^{k^T}$ and $P_{gp}^k = \sum_1^{NC} \alpha_i V_{g,i}^k.m_p^{k^T}$

Substituing Eq. 3.27 into Eq. 3.25 we obtain:

$$\begin{aligned}
S_w^k &= W_g^{k^T} L_{gg}^k.W_g^k + W_p^{k^T} L_{pp}^k.W_p^k - W_p^{k^T} L_{pg}^k.W_g^k - W_g^{k^T} L_{gp}^k.W_p^k \\
S_b^k &= W_g^{k^T} K_b[1]^k.W_g^k + W_p^{k^T} K_b[2]^k.W_p^k - W_p^{k^T} K_b[3]^k.W_g^k - W_g^{k^T} K_b[4]^k.W_p^k
\end{aligned} \tag{3.28}$$

with:

$$K_b^k[1] = K_b^k[2] = 2.(P_{gg}^k + P_{pp}^k)$$
$$K_b^k[3] = K_b^k[4] = P_{gg}^k + P_{pp}^k + P_{gp}^k + P_{pg}^k$$

Finally,by defining $A^k = [W_g^k, W_p^k]$, we can write:

$$\begin{aligned} S_w^k &= A^{k^T}.K_w^k.A^k \\ S_b^k &= A^{k^T}.K_b^k.A^k \end{aligned} \tag{3.29}$$

with:

$$K_w^k = \begin{bmatrix} L_{gg}^k & L_{gp}^k \\ L_{pg}^k & L_{pp}^k \end{bmatrix},\ K_b^k = \begin{bmatrix} K_b[1]^k & K_b[4]^k \\ K_b[3]^k & K_b[2]^k \end{bmatrix} \tag{3.30}$$

Hence, the objective function J^k becomes:

$$J^k = \frac{S_b^k}{S_w^k} = \frac{A^{k^T}.K_b^k.A^k}{A^{k^T}.K_w^k.A^k} \tag{3.31}$$

The vector of projection matrices A^k that minimizes the expression in Eq. 3.31 is then obtained by solving the generalized eigenvalue problem:

$$K_b^k.A^k = \lambda.K_w^k.A^k \tag{3.32}$$

A^k is then appropriately split to obtain W_g^k and W_p^k. The inclusion of weights $\alpha_{i=1..NC}$ into Eq. 3.26 enables varying the importance of the corresponding feature vector during boosting process as required by the Adaboost algorithm. The LDA based weak classifier defined above (Eq. 3.23 to Eq. 3.32) is also evidently easy to set up. Hence, the first and second properties that we require for our classifiers (see the top of this section) are well fulfilled. We still have to check for the third property, where our classifiers should have accuracy directly related to the corresponding spectral bands' quality. In Fig. 3.3, we have, respectively, plotted the rank-1 recognition rates of all weak classifiers $H_{k=1..25}^{IRIS}$ upon both databases $IRIS - M^3$ for the three studied algorithms, namely MBLBP, HGPP, and LGBPHS. The later features extraction algorithms as well as the two face databases will be defined and investigated in the experimental part of this thesis. Meanwhile, we use them to evaluate our weak classifiers.

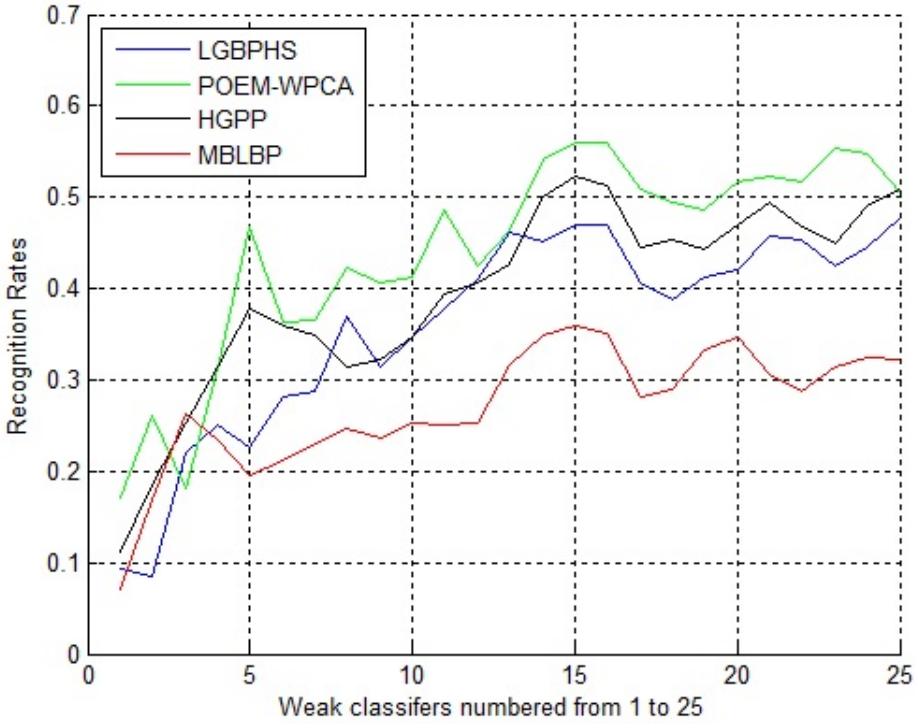


Fig. 3.3 Rank-1 recognition performances of weak classifiers upon the the $IRIS-M^3$ database

We can see that the 'x' axis represents the weak classifier used for classification. We remind the reader that we use numbers from 1 to 25 for the $IRIS-M^3$ database to equivalently symbolize spectral bands and the corresponding weak classifiers. For the $IRIS-M^3$ database, the number 1 equivalently designates the spectral band at 480nm and the corresponding weak classifier . Two main results could be observed in Fig. 3.3. First, for the $IRIS-M^3$ face database, weak classifiers that correspond to spectral bands numbers 14 (610nm), 15 (620nm) and 16 (630nm) gave the best recognition performances for all algorithms. The same spectral bands have been chosen by Chang et al. (Fig.16 from [25]) as being the best spectral bands. With this conformity of results between our analysis of the recognition performance of different weak classifiers and results reported in [25] for the best spectral bands, we can conclude that our weak classifiers, as they are defined, preserve well the intrinsic properties of each spectral band and hence preserve their ranking of superiority. The third required property is verified: the best weak classifier corresponds to the best spectral band. The second result that can be observed in Fig. 3.3 is that weak classifiers gave a medium recognition performance upon the $IRIS-M^3$ database ($\leq 57\%$). This second observation explains the need to boost these weak classifiers by adequately combining them

to get better results. In the next section, we use the Adaboost algorithm to determine the optimal combination of our weak classifiers. Once this is done, we can determine a new classification of best spectral bands based on the contribution of their corresponding weak classifiers in the final strong classifier and the equivalence bands quality/weak classifiers performance ensured by the third property verified above.

Adaboost Algorithm For Stronger Classifiers

In this section we explain how to combine our weak classifiers, defined in the previous section, using an Adaptive boosting algorithm, or shortly, Adaboost. Since it was first proposed by Freund and Schapire in [48], several varieties of the Adaboost algorithm [44] have been successfully investigated in many applications including object detection, face recognition, text categorization, and others. One variety that we adopt in this work is Stagewise Additive Modeling using a Multi-Class Exponential Loss Function (SAMME) [172]. SAMME is both simple and efficient for multi-class problems. The use of a multi-class exponential loss function was proposed to approximate the multi-class Bayes optimal classification rule. We propose to combine our weak classifiers using SAMME as follows:

---

1. Initialize the observation weights $\alpha_i = \frac{1}{NC}, i = 1..NC$
2. For $k = 1..N$
   (a) Learn the weak classifier H_k using weights $\alpha_i = \frac{1}{NC}, i = 1..NC$
   (b) Compute the recognition error made by H_k

$$err_k = \sum_{i=1}^{N} C\alpha_i.\tau(H_k(I_P^i \bar{=} C_g^i))/\sum_{i=1}^{N} C\alpha_i$$

   (c) Compute

$$\omega_k = log(\frac{1-err_k}{err_k}) + log(NC-1)$$

   (d) Update data weights as:

$$\alpha_i \rightarrow \alpha_i.exp(\omega_k.\tau(H_k(I_P^i) \neq I_g^i))$$

   , for $i = 1..NC$

(e) Renormalize $\alpha_{i=1..NC}$

3. Outpute the final classifier:

$$H(I_p) = argmax_c(\sum_k \omega_k.\tau(H_k(I_p = C))))$$

In the above outline of our Adaboost process, we have included all weak classifiers, $k = 1..N$, into the final strong classifier and with different weights. However, we can still stop our boosting process once we get an error rate lower then $err_{tolerated}$ ($err_k \leq err_{tolerated}$). The final stronger classifier, represented by H, is a weighted combination of different weak classifiers H_k with weights that determine the importance of each one. The superiority between weak classifiers, determined by the superiority of their affected weights, implies the same superiority between the corresponding spectral bands. By learning H, do we learn the classification of best spectral bands? One question that arises from the definition of SAMME is which weak classifier we must start our boosting algorithm with. To answer this question, we show in the experimental part of this work that with an $err_{tolerated}$ small enough, the weight affected to a given weak classifier will remain roughly constant regardless with which classifier we have started SAMME.

3.5 Conclusion

In this chapter, we have presented four static algorithms for BSBS. These algorithms were the Non-kermalized/kernalized sparsity based approach, the multilinear decomposition based approach and finally the Boosted LDA based approach. The theory behind each algorithm has been detailed and some preliminary results has been displayed. The proposed algorithms are independent of any face database and so can be applied with any system for face recognition. The particularity of statistic BSBS algorithms as mentioned in the introductory part of this chapter is that the set of best bands is chosen only one time during the training process. This fact is double edged, and such systems are extremely sensitive to the variation of imaging conditions. As for real world application, outdoor imaging conditions for example are uncontrolled and may show significant variation in short time, static BSBS algorithms are very limited when implemented in outdoor environment. To solve this problem, the optimization parameters have to be tuned with each new subject. This dynamic aspect enables to select different spectral bands for different subjects and hence preserves a certain level of

minimum global performance (The recognition accuracy on a set of subjects).

In the next chapter, we present our unique algorithm to dynamically select the best spectral bands from a given dataset. However, as the image quality is essential for the success of such algorithms including the image brightness, level of bluring,etc. We propose first a new image filter adapted to work on MI images. The new filter is aimed to enhance the discrimination performance of any given features extraction algorithm. As a proof of correctness, we determine the parameters of our filter to adequate the use of a known features extraction technique which is SURF (Speed Up Robust Features). The optimized filter and the dynamic BSBS system that will be presented in the next chapter will be further experimented independently in the experimental section of this thesis.

Chapter 4

Dynamic Best Spectral Bands Selection

4.1 Introduction

In this chapter, the second category of BSBS systems, namely dynamic BSBS systems are studied. We propose two approaches. One approach that learn a filter adequate to process multispectral images. The filter was adapted to the well known feature extraction tool SURF (Speed Up Robust Features) and it has the role to ensure a better selection of best spectral bands invariant to illumination conditions. The second approach is then presented were a dynamic system for BSBS is proposed. This category of systems select a different set of best spectral bands for each new subject. This way, the set of selected bands adapts with the variation of imaging conditions. We can expect that this category of BSBS systems will have a better performance but a much higher processing time. Our dynamic system will be based on several optimization techniques including SURF, Mixture of Gaussian analysis (MG), Likelihood ratio and transfer learning.

4.2 Image Filter Learning

4.2.1 Math Background

In this section, math tools used for our dynamic BSBS system are detailed. The main tool used is the SURF descriptor.

SURF descriptor

SURF or Speed-Up Robust Features, is a keypoints based descriptor that was invented by Bay et al. in [6]. Two main phases are defining SURF: Keypoints detection and keypoints description. In the following, we outline the main steps of these two phases.

Keypoints Detection In this first step of SURF, we compute what is called Blob Response Map (hereafter abbreviated as BRM). The latter is a matrix with the same size of the original image I (the image from which we want to extract keypoints). Each element of BRM, namely $BRM(x,y)$ is the determinant of the hessian matrix $H(x,y)$ of pixel $I(x,y)$ from the original image. $H(x,y)$ is defined as:

$$H(x,y,\sigma) = \begin{bmatrix} D_{xx} & D_{yx} \\ D_{xy} & D_{yy} \end{bmatrix} \tag{4.1}$$

Terms D_{xx}, D_{yy}, D_{xy} and D_{yx} are obtained by the convolution of the Gaussian second order derivative with the image I at point (x,y) in the x, y and $x-y$ direction respectively. The use of Gaussian filters enabled computing the Hessian matrix for different scales σ. The determinant of H was defined by Herbert as:

$$BRM(x,y,\sigma) = D_{xx}.D_{yy} - (0.9.D_{xy})^2 \tag{4.2}$$

All the elements of BRM are then computed using Eq. 4.2 for several values of σ (octaves). Keypoints are obtained as the pixels with the corresponding $BRM's$ value is a local maxima of BRM. A $3 \times 3 \times 3$ non-maximum algorithm was used to find these local maxima.

Keypoints Description The second step of SURF is to extract a feature vector from each keypoint. All feature vectors are then concatenated to form the feature vector of the whole image. First, a circle around each determined keypoint is selected. Haar wavelet filter is then applied on each pixel within this circle, in both direction x and y. Hence, for each of these pixels we get two Haar responses dx and dy. dx and dy responses within a window of size $\pi/3$ sliding around the keypoint are then summed separately to get dX and dY. Then the vectorial sum of dX and dY gives a vector of orientation (red arrow in Fig. 4.1) corresponding to each sliding window (gray window in Fig. 4.1) We compute the orientation vector (arrow) of each sliding window and the vector with the greatest lenght is defined as the orientation of the keypoint. We call this vector V_{orient}. Second, a square oriented in the direction of V_{orient} is defined around the considered keypoint (see Fig. 4.2). Haar filter is again applied on the later square and the obtained dx and dy responses ,as well as $|dx|$ and

$|dy|$, for each 2×2 pixels square are summed. Hence, for each small square we get 4 values $\sum dx$, $\sum dy$, $\sum |dx|$ and $\sum |dy|$. Finally, for the whole keypoint we get $4 \times 4 \times 4 = 64$ values that define its feature vector.

4.2.2 Filter Optimization

Several factors related to imaging conditions or cameras characteristics may affect the quality of captured images. This may complicate further the task of spectral bands selection by increasing data redundancy: For example, two blurred regions in two images corresponding to two different subjects may look similar due to blurring. This similarity between two different subjects is an undesirable redundancy that has to be removed. To do so, the most intuitive way is to apply a filter. Choosing the adequate filter is usually challenging due to the high number of factors to tune even for the simplest filter (filter type, dimension, scales, number of banks, etc). The formulation of this problem of filter construction as an optimization process has been proved to be faisable and effective to build the right filter. We adopt this approach and we learn a filter tailored to enhance the performance of SURF descriptor [58]. The same learning process we present here could be followed to learn filters for other features extraction algorithms and achieve a comparable image quality enhancement.

Schematically, our learned filter, that we call hereafter W, is coupled with the feature extraction algorithms to form the final enhanced feature extraction system as shown in Fig. 4.3.

keyregions instead of keypoints

The quality of keypoints selected by the SURF descriptor, subject of our test for our learned filter, and their concentration on different regions of the image space are very sensitive to the descriptor parameters. This makes SURF incompetent for face recognition compared to descriptors that are based on local features, such as Local Binary Patter (LBP). We propose to reduce this SURF weakness by selecting keyregions instead of keypoints. A keyregion is a region with high concentration of keypoints. We first superpose keypoints obtained from all the training images as shown in Fig. 4.4.a.

Then, using the mean shift technique, we cluster these points into a Gaussian-like distribution (see Fig. 4.4.b). The centers of obtained clusters (small circles in Fig. 4.4.b) have a high density of keypoints and are selected as the centers of our keyregions. In a second step, we compute the two-dimensional covariance vector $\sigma_k = [\sigma_{x,k}, \sigma_{y,k}]$ of each cluster of center

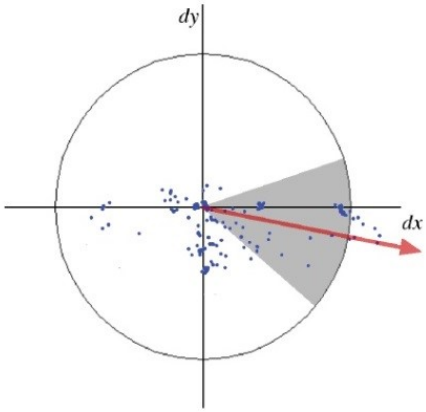


Fig. 4.1 Determination of keypoints orientation for SURF

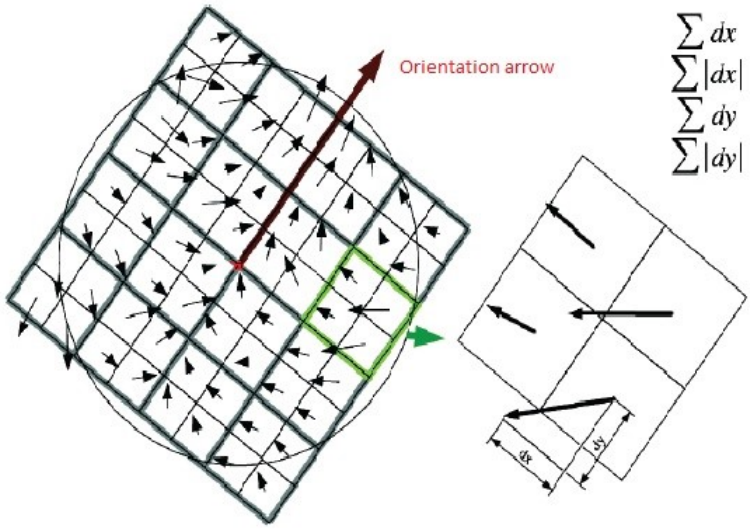


Fig. 4.2 Haar responses of pixels inside the oriented square

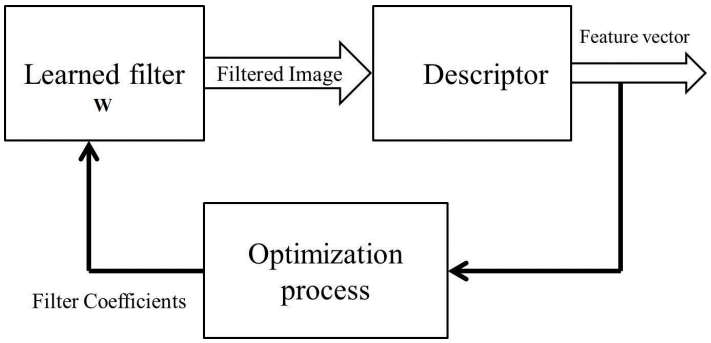


Fig. 4.3 Filter implementation inside the whole recognition system

$C_{k=1..M}$. M is the number of keyregions. The keyregion R_k of center C_k is then determined as the circle of center C_k and radius $r_k = min(\sigma_{x,k}, \sigma_{y,k})$. The set of keypoints inside R_k is then selected as the most interesting, and the keypoint locations are registered. During the test phase, pixels at these locations will be used to match their corresponding keyregions. These keyregions were defined as circular in order to preserve the rotational invariance that characterizes the SURF descriptor. We can see that keyregions are less sensitive to SURF parameters than keypoints; the importance of a particular region inside an image is decided by a vote from the keypoints extracted from all the training images. For example, even if SURF missed an important region in a given face image A due to a bad setting of its parameters, it will have detected that region in a sufficient number of other face images from the gallery set. Thus, the missed region will be used to match face image A. We say that the voting rule, with which keyregions are selected, reduces the number of interesting regions missed by SURF in a given face image. On the other hand, selected keypoints are small in number and of maximum importance. During the test phase, pixels at the same locations of best key points learned during the training phase are used for face matching. The SURF descriptions of all pixels of a given keyregion R_k are determined and linked together in one vector V_k that represents the feature vector of R_k. Finally, the distance D between a probe image I^p and a gallery image I^g is defined as the sum of distances between selected key regions:

$$D(I^p, I^g) = \sum_k D(R_k^p, R_k^g) = \sum_k \|V_K^P - V_K^g\|_2 \tag{4.3}$$

Discriminant Filter learning

In this section, we propose to learn a 3×3 sized linear filter $W_{k=1..M}$ for each keyregion k. Let $dx^f(x,y)$ be the Haar filter response on a filtered keypoint located at (x,y). Using integral image, $dx^f(x,y)$ could be expressed as [6]:

$$dx^f(x,y) = [C_2 - B_2 - D_2 + A_2] - [C_1 - B_1 - D_1 + A_1] \tag{4.4}$$

The letters in Eq. 4.4 are defined in Fig. 4.5. The value of C_2, for example, could be written as:

$$C_2 = Intg_f(m,n) = \sum_{j \le m, e \le n} I^f(l,e) \tag{4.5}$$

The location (m,n) is where C_2 is computed, while $Intg_f$ designates the integral image of the filtered image I^f. $I^f(l,e)$, the value of pixel (l,e) in the filtered image, could be written as: $I^f(l,e) = \sum_j \omega_j I(j,l,e)$, where $I(j,l,e)$ for $j = 1..($ are the nine (3×3) neighborhoods

of pixel $I(l,e)$ in the non-filtered image and $\omega_{i=1..9}$ ate the filter components. Note that pixel neighborhoods are counted from the top left pixel, decreasing row by row to the last neighborhood, which is the lowest, most-right pixel. Hence, Eq. 4.5 becomes:

$$C_2 = \sum_{j\leq m,e\leq n} \sum_{j=1}^{9} \omega_j I(j,l,e) = \sum_{j=1}^{9} \omega_j \sum_{j\leq m,e\leq n} I(j,l,e) \tag{4.6}$$

For $j=1$, for example, the term $\sum_{j\leq m,e\leq n} I(1,l,e)$is the sum of pixels highlighted in yellow in Fig. 4.6.This is simply the value of the integral image of the non-filtered image at pixel (m-1, n-1). We can write:

$$\sum_{j\leq m,e\leq n} I(1,l,e) = Intg(m-1,n-1) \tag{4.7}$$

$Intg$ designates the integral image of the non-filtered image I. Similarly, we can write:

$$\begin{aligned}
&\sum_{j\leq m,e\leq n} I(2,l,e) = Intg(m-1,n)\\
&\sum_{j\leq m,e\leq n} I(3,l,e) = Intg(m-1,n) - Intg(m-1,1)\\
&\sum_{j\leq m,e\leq n} I(4,l,e) = Intg(m,n-1)\\
&\sum_{j\leq m,e\leq n} I(5,l,e) = Intg(m,n)\\
&\sum_{j\leq m,e\leq n} I(6,l,e) = Intg(m,n) - Intg(m,1)\\
&\sum_{j\leq m,e\leq n} I(7,l,e) = Intg(m,n-1) - Intg(1,n-1)\\
&\sum_{j\leq m,e\leq n} I(8,l,e) = Intg(m,n) - Intg(1,n)\\
&\sum_{j\leq m,e\leq n} I(2,l,e) = Intg(m,n) - Intg(1,n) - Intg(m,1) + Intg(1,1)
\end{aligned} \tag{4.8}$$

In Eq. 4.8 the symbol $Intg$ is replaced with the symbol C_2 for consistency and the expression

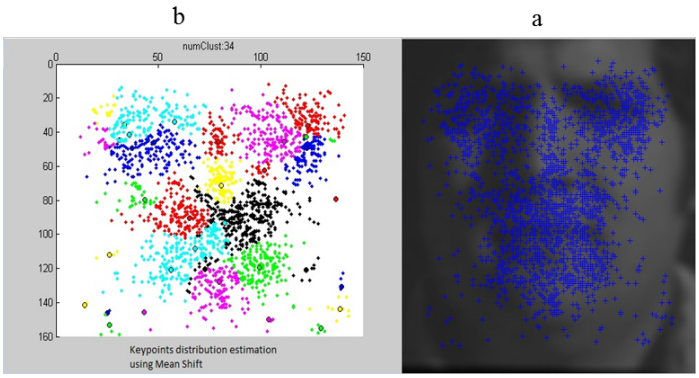


Fig. 4.4 Keyregions determined using SURF keypoints

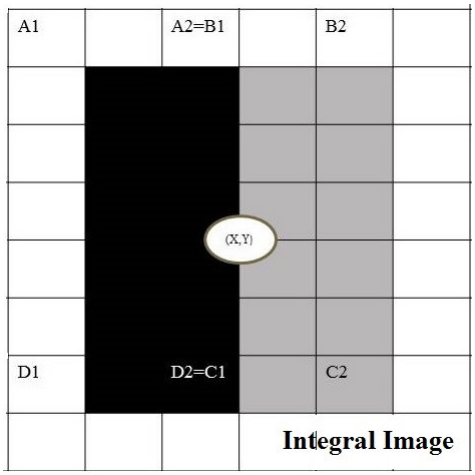


Fig. 4.5 Response of Haar filter on pixel (x,y)

I(1,1,1)		*I(1,1,n-2)*	*I(1,1,n-1)*	*I(1,1,n)*
⋮	⋱	⋮	⋮	⋮
I(1,m-1,1)	…	*I(1,m-1,n-2)*	*I(1,m-1,n-1)*	*I(1,m-1,n)*
I(1,m,1)	…	*I(1,m,n-2)*	*I(1,m,n-1)*	*I(1,m,n)*

Fig. 4.6 In yellow are summed pixels for j=1

of C_2 becomes:

$$\begin{aligned} C_2 = Intg_f(m,n) &= \omega_1.C_2(m-1,n-1) + \omega_2.C_2(m-1,n) + \omega_3.[C_2(m-1,n) \\ &- C_2(m-1,1)] + \omega_4.C_2(m,n-1) + \omega_5.C_2(m,n) + \omega_6.[C_2(m,n) - C_2(m-1,1)] \\ &+ \omega_7.[C_2(m,n-1) - C_2(1,n-1)] + \omega_8.[C_2(m,n) - C_2(1,n)] + \omega_9.[C_2(m,n) \\ &- C_2(1,n) - C_2(m,1) + C_2(1,1)] \end{aligned} \tag{4.9}$$

Similarly, we can determine the values B_2, D_2, A_2, C_1, B_1, D_1 and A_1 supposed located at pixels (p,q), (t,s), (g,h), (m1,n1), (p1,q1), (t1,s1) and (g1,h1) respectively. After grouping similar terms, Eq. 4.4 becomes:

$$\begin{aligned} dx^f(x,y) &= \omega_1.([C_2(m-1,n-1) - B_2(p-1,q-1) - D_2(t-1,s-1) + A_2(g-1,h-1)] \\ &- [C_1(m1-1,n1-1) - B_1(p1-1,q1-1) - D_1(t1-1,s1-1) + A_1(g1-1,h1-1)]) \\ &+ \omega_2.([C_2(m-1,n) - B_2(p-1,q) - D_2(t-1,s) + A_2(g-1,h)] - [C_1(m1-1,n1) \\ &- B_1(p1-1,q1) - D_1(t1-1,s1) + A_1(g1-1,h1)]) + ... \end{aligned} \tag{4.10}$$

The expression of Eq. 4.10 is very long and we only write the first two terms. The other terms could be easily determined by a simple factorization of the filter components. We can see that the first term of $dx^f(x,y)$ (the long expression multiplied by ω_1) is the Haar response at pixel (x-1, y-1) from the non-filtered image (see Fig. 4.7), while the second term is the Haar response at pixel (x-1, y)of the same image. The other non-written terms of Eq. 4.10 are expressed similarly and $dx^f(x,y)$ becomes:

$$\begin{aligned} dx^f(x,y) &= \omega_1.dx(m-1,n-1) + \omega_2.dx(m-1,n) + \omega_3.[dx(m-1,n) \\ &- dx(m-1,1)] + \omega_4.dx(m,n-1) + \omega_5.dx(m,n) + \omega_6.[dx(m,n) - dx(m-1,1)] \\ &+ \omega_7.[dx(m,n-1) - dx(1,n-1)] + \omega_8.[dx(m,n) - dx(1,n)] + \omega_9.[dx(m,n) \\ &- dx(1,n) - dx(m,1) + dx(1,1)] \\ &= \sum_i \omega_i.a_x^i(x,y) \end{aligned} \tag{4.11}$$

The symbol $a_x^{i=1..9}$ in Eq. 4.11 is used for abbreviation. This equation shows that the Haar response of a filtered image at location (x,y) is a weighted sum of the Haar response at a translated version of (x,y) of the non-filtered image. Similarly, we can express the Haar response at (x,y) on the y direction, namely $dy^f(x,y)$. The two other components of the SURF feature vector, $|dx^f(x,y)|$ and $|dy^f(x,y)|$ could not be expressed with the same weighted sum since $|\sum_i \omega_i.a_x^i(x,y)| \neq \sum_i |\omega_i.a_x^i(x,y)|$. To solve this problem, we propose to

reduce the SURF descriptor to be the concatenation of only $dx^f(x,y)$ and $dy^f(x,y)$. Then, the obtained filtered and reduced SURF (or FR-SURF) could be written as:

$$FR-SURF(x,y) = \sum_i \omega_i R-SURF^i(x,y) \tag{4.12}$$

Where $R-SURF^i(x,y)$, for $i = 1..9$, are reduced SURF vectors at the translated version of (x,y). For example,$R-SURF^1(x,y) = [a_x^1, a_y^1]$. FR-SURF is used to learn the filter components, while the full SURF will be used in Eq. 4.3 for face matching. For a given keyregion k, FR-SURF vectors of all its keypoints are concatenated in one vector V_k^f of the form $V_k^f = W_k.V_k$. V_k is the matrix formed by R-SURF vectors. Finally, filter components of region k, namely W_k, are optimized using LDA: the within-class and between-class covariance matrices are written as a function of V_k^f in the form of $W_k^T.M_k.W_k$ and $W_k^T.N_k.W_k$ respectively. M and N are expressed using V_k, and W is then optimized as the solution of the generalized eigen-value problem $N.W = \lambda.M.W$. During the test phase, SURF vectors of keypoints of each keyregion k are multiplied by the corresponding filter W_k and concatenated to form the feature vector of k. Then Eq. 4.3 is used for face matching.

4.3 Dynamic Best Spectral Bands selection

In the previous section, we have seen how we have enhanced the quality of our multispectral images using a tailored filter adapted as un example for the SURF descriptor. We believe that image filtering is much more crucial for dynamic BSBS systems then for static BSBS systems; the parameters that define the former are continuously tuned with each new subject. Hence, preserving a stable quality of images by applying robust filters may reduce the processing time an reduce the number of selected bands.

In the following, we present our algorithm for dynamic best spectral bands selection (DBSS) [15]. In our approach, we assume the existence of two probability density functions (PDFs) F_{good} and F_{bad} that determine the probability of a given spectral band k to belong to the set of good spectral bands and bad spectral bands respectively. We assume also that the two PDFs are function of the quality Q_k of the band k and could be written as: $F_{good} = f_{good}(Q_k)$ and $F_{bad} = f_{bad}(Q_k)$. We propose to use the finite Mixture of Gaussians (MG) technique to estimate the defined two PDFs. Then for a new spectral band, we determine its quality Q, and based on the computed likelihood ratio $F_{good}(Q)/F_{bad}(Q)$, we determine if this spectral band is good enough to be selected for recognition purposes.

The two PDFs F_{good} and F_{bad} are determined over two sets of spectral bands which are the set of good spectral bands S_{good}, and the set of bad spectral bands S_{bad}. These two sets are determined from the training database. The training database is firstly divided into gallery and probe sets. Then for each subject I^{prob} from the probe set, the k^{th} spectral band I_k^{prob} of I_{prob} is matched against the k^{th} spectral bands of all subjects from the gallery database, namely I^gal_k,i=1..M. The obtained vector of match scores $Vm_k^{alg} = [Ms^{alg}(I_k^{prob}, I_{k,1}^{gal})..Ms^{alg}(I_k^{prob}, I_{k,M}^{gal})]$ is then used to determine the recognition performance RP_k^{alg} of band k as follows:

$$RP_k^{alg} = \frac{Ms^{alg}(I_k^{prob}, I_{k,self}^{gal})}{\sum_j^M Ms^{alg}(I_k^{prob}, I_{k,j}^{gal})} \tag{4.13}$$

Ms^{alg} designates the match score determined using Euclidian distance as follows:

$$Ms^{alg}(I_k^{prob}, I_{k,j}^{gal}) = \frac{1}{D_{euc}(V_{I_k^{prob}}^{alg}, V_{I_{k,j}^{gal}}^{alg})} \tag{4.14}$$

$V_{I_k^{prob}}^{alg}$ and $V_{I_{k,j}^{gal}}^{alg}$ are feature vectors extracted from I_k^{prob} and $I_{k,j}^{gal}$ respectively using the evaluated algorithm designated by the superscript "alg". The numerator of Eq. 4.3 is the match score between the spectral band I_k^{prob} and the spectral band $I_{k,j}^{gal}$ from the gallery database corresponding to the same subject (self-similarity). RP_k^{alg} is seen as the ratio of this self-similarity by the sum of match scores between band k and all the gallery images at the same spectrum. We can see that RP_k^{alg} is bounded between 0 and 1.

The second parameter to be determined is the quality Q_k of k. This factor aims to characterize the robustness of the considered band against the variation of incident light. A band with a good quality should show up enough continuity of different image edges that discriminate between different face landmarks like eyes, eyebrows, noise, mouth, etc. As most of the algorithms evaluated in our work are scale invariant, the quality factor should also be scale invariant. To fulfill this later requirement, we adopt the same quality factor proposed by Vasta et al. in [152].This factor was defined using redundant discrete wavelet transform and its value is bounded between 0 and 1. The value 0.5 represents the best quality image while 0 and 1 represent the worst quality image.

Hence, two factors are defined for each spectral band k from the probe set which are the band's recognition performance and its quality, namely $P_k^{alg} = (RP_k^{alg}, Q_k)$. In Fig. 4.8, we plot all characteristic points P_k^{alg} corresponding to all spectral bands of all subjects from

the probe set. In this figure, the algorithm used was MBLBP (Which means that RP was computed using feature vectors extracted with MBLBP). We can see that the plotted points could be divided into three regions. Points belonging to the region with RP greater than 0.7 (region of good bands highlighted in green), points belonging to the region with RP lower than 0.2 (region of bad bands highlighted in red) and a third region with RP between 0.2 and 0.7 (highlighted in cyan). Similar plots could be obtained for the algorithms HGPP and LGBPHS. As in our approach a spectral band is either good or bad, points from the third region should be assigned to one of the green or red regions. To do so, we use a transfer learning technique proposed by Zhou et al. in [171]. This semi-supervised technique affects points to the correct region, red or green, based on two assumption:1) Neighbor points are likely to have the same label (green or red) and 2) Points belonging to the same cluster are likely to have the same label. The result of applying the algorithm proposed by Zhou et al. is shown in Fig. 4.9. Now, F^{alg}_{good} could be determined over Q values of points from the green region, while F^{alg}_{bad} is determined over Q values of red points. Classifying characteristic points into good and bad points based on their recognition performance *RP* and then deduce from this classification the distribution of Q values of the two sets S_{good} and S_{bad}, is the heart of the tricky approach we are proposing in this work. Our PDFs are modeled as a finite mixture of Gaussians each, and are defined as:

$$\begin{aligned} F^{alg}_{good}(Q_k) &= \sum_{i}^{L} .p^{alg}_{j} g(Q_k, m^{j,good}_{alg}, \sigma^{j,good}_{alg}) \\ F^{alg}_{bad}(Q_k) &= \sum_{i}^{L} .q^{alg}_{j} g(Q_k, m^{j,bad}_{alg}, \sigma^{j,bad}_{alg}) \end{aligned} \tag{4.15}$$

where

$$g(Q,m,\sigma) = \frac{1}{\sigma.\sqrt{2.\pi}}.exp(-0.5(\frac{\|Q-m\|}{\sigma})^2)$$

The variational Bayesian inference is then used to determine the model parameters $m^{j,good}_{alg}$, $\sigma^{j,good}_{alg}$, $m^{j,bad}_{alg}$, $\sigma^{j,bad}_{alg}$, p^{j}_{alg} and q^{j}_{alg} for all $j = 1..L$, as well as the number of Gaussians involved L. As an example, we plot in Fig. 4.10, the estimated mixture of Gaussians for three algorithms for features extaraction which are MBLBP [1], HGPP [**?**] and LGBPHS [**?**]. These algorithms will be studied further in the last section of this thesis where we present our results. We can see that the overlapping between the two PDFs F^{alg}_{good} and F^{alg}_{bad} for each algorithm is important, which indicate that using only one of the two PDFs to decide if a new band is good or bad is not accurate. So, we propose to use instead the Likelihood ratio test. For a new multispectral image, we compute the quality Q_k of each of

its 25 spectral bands. Then the corresponding likelihood ratio LR_k is computed as:

$$LR_k(Q_k^a lg) = \frac{F_{good}^{alg}(Q_k)}{F_{bad}^{alg}(Q_k)} \tag{4.16}$$

The two spectral bands with the greatest LR values are then selescted as the best spectral bands and are kept for the next face matching step as it will be explained in the experimental part of this thesis.

4.4 Conclusion

Throughout this chapter we have proposed a complete system for dynamic BSBS formed by a learned filter and a system for optimal bands selection. The built filter was learned to enhance the performance of one particular algorithm which is SURF descriptor. The aim of the filtering step is to enhance the quality of multispectral images provided by the $IRIS-M^3$ face database. Once MI images were filtered by our learned filter, they were used to build the final dynamic BSBS system. The main techniques used for this purpose were mixture of gaussian, transfer learning and likelihood ratio. As will be depicted and discussed in section 5 on experimental results, the performances obtained using the dynamic system have shown a remarkable superiority over the static system.

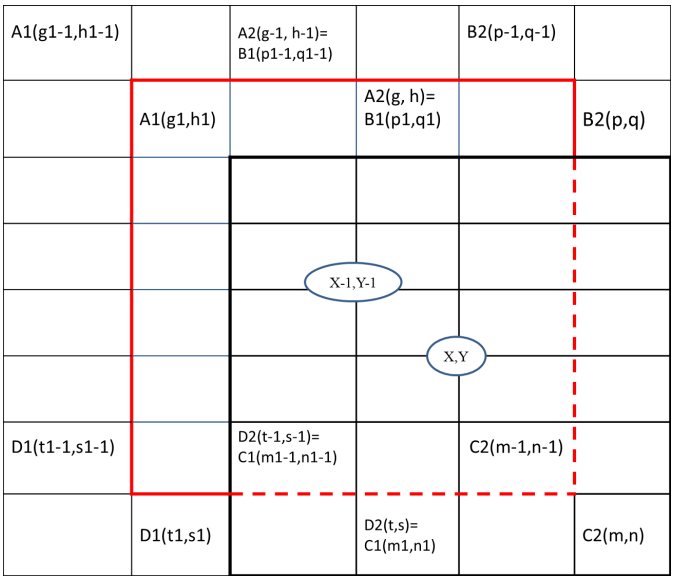


Fig. 4.7 In red, is the first term of the filtered Haar response

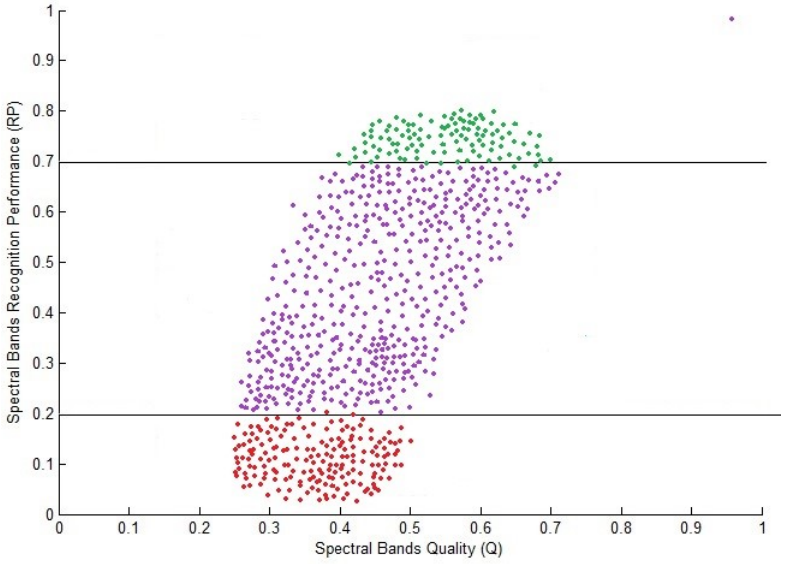


Fig. 4.8 Distribution of characteristics points obtained with MBLBP

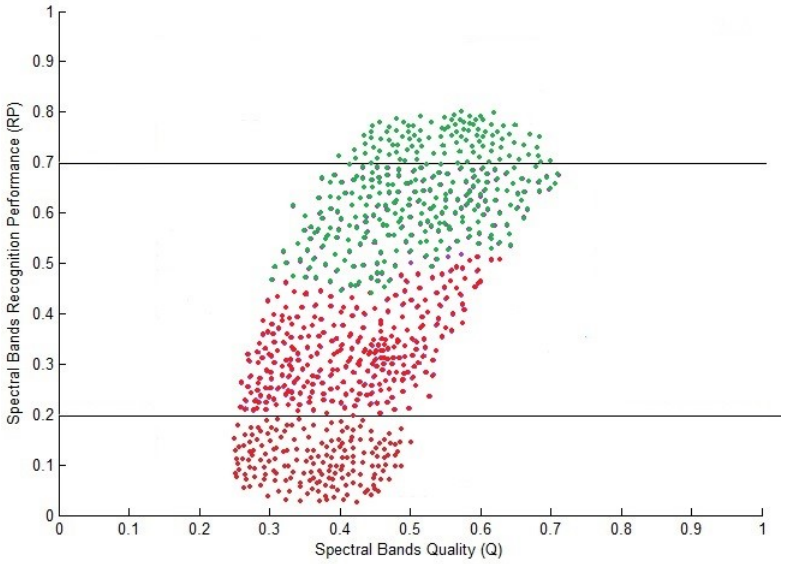


Fig. 4.9 Transfer learning effect on the distribution of characteristic points for MBLBP algorithm

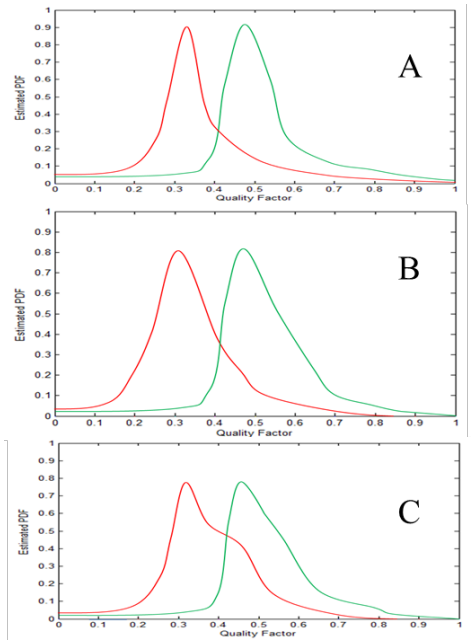


Fig. 4.10 Estimated F_{good}(green)and F_{bad}(red) PDFs for A) LGBPHS, B) HGPP and C) MBLBP

Chapter 5

Results and Discussions

5.1 Introduction

So far , we have presented the different approaches we have built to solve the problem of BSBS for face recognition tasks. Static and dynamic algorithms have been proposed including two approaches based on the sparsity theory, one approach that investigated the Adaboost algorithm and the final approach that selected the set of best spectral bands dynamically using different techniques like mixture of Gaussian, Likelihood ratio, etc.

In this final chapter of our thesis , we experiment the performances of these algorithms upon the $IRIS-M^3$ face database. The main metrics we use to compare the robustness/effectivness of our approaches are the Rank-1 recognition performance and the Cumulative Match Curves (CMC curves). The different bands selection algorithms will be challenged to enhance the performance of exiting state-of-the-art algorithms for features extraction which are Multiblock Local Binary Pattern (MBLBP), Local Gabor Binary Pattern Histogram Sequence (LGBPHS) and Histogram of Gabor Phase Patterns (HGPP). The later algorithms will be applied on the spectral bands selected by each BSBS algorithm and the best results obtained determine the BSBS algorithm that much better each features extraction approach.

In figure 5.1, we see the different components of our final system. Bands selection, bands fusion and match score computation are the main and are experimented separately in the rest of this chapter.

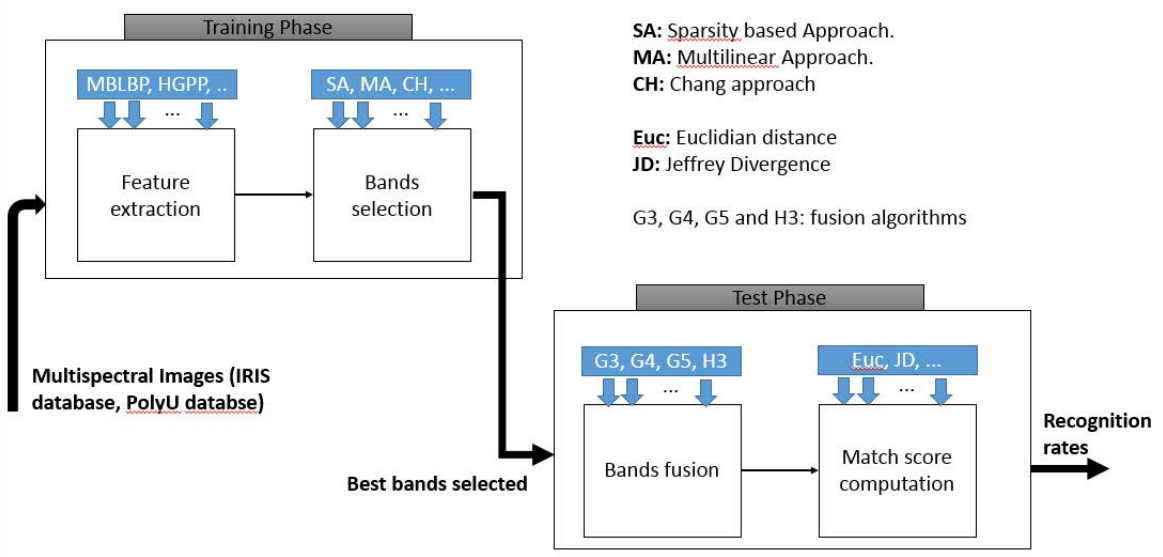


Fig. 5.1 Pipeline of a Face recognition system augmented with a process for bands selection

Bands selection: We start by determining the different sets of optimal spectral bans for each BSBS system. The different settings of our systems including the optimization constants are fixed. Our BSBS approaches were applied on each feature extraction algorithm by solving the corresponding optimization problem. The obtained best bands were then stored for the next bands fusion process.

Bands fusion: The second main step is band fusion. The aim of this small but relevant section is to answer the question of how difficult and critical is the choice of the fusion technique within a MI based environment and its impact on the final recognition result ? and to raise the question on what else can be done to enhance this step ?. The different fusion techniques proposed are generally used with grayscale images and are not optimized for MI images. Although this later issue is important to solve and make a fusion process tailored to MI images, our systems has shown an acceptable performances and the fused images brought significant information over grayscale images.

Match scores computation: The final section of this chapter investigates the outcomes of the previous two sections, namely bands selection and bands fusion, to enhance the performance of MBLBP, HGPP and LGBPHS feature extraction algorithms and prove the effectiveness of the MI based approach. In this section, we extract the relevant face features using the mentioned three algorithms, we combine them into one feature vector that is used to compute the match score between database images. The obtained match distances are used to get the recognition performance over the whole database for each feature extraction technique.

Fig. 5.2 Sample images in a data record in the IRIS-M3(i-p)database:(i) Day lighted gray image, day lighted images at (j) 590nm, (k) 640nm, (l) 700nm, (m) halogen lighted gray image, halogen lighted images at (n) 590nm, (o) 640nm and (p) 700nm

5.2 Face Databases And Algorithms For Features Extraction

5.2.1 The $IRIS-M^3$ Face Database

In the IRIS-M3 face database [25], there are a total of 82 participants of different ethnicities, ages, facial and hair characteristics, and genders with a total number of 2624 face images. The image resolution is 640×480 pixels and the interocular distance is about 120 pixels. The database was collected in 11 sessions between August 2005 and May 2006 with some participants being photographed multiple times. The database is comprised of 76% male and 24% female subjects, and the ethnic diversity is defined as a collection of 57% Caucasian, 23% Asian (Chinese, Japanese, Korean, and similar ethnicity), 12% Asian Indian, and 8% of African descent. For each subject, three groups of images have been captured depending on the lighting condition, including a group of day lighted images, a group of halogen lighted images and a group of fluorescent lighted images. In turn, each image group was formed by two images: a gray image and a multispectral image (image cube) formed by 25 spectral bands captured in the visible spectrum (from 480nm to 720nm) with a step of 10nm. For our experiments, day lighted images and halogen lighted images are going to be matched against each other. Image samples used for our experiments are displayed in Fig. 5.2.

5.2.2 The MBLBP Algorithm

The MBLBP algorithm was first proposed by Ahonen et al. in [1]. It consists of dividing each face image into non-overlapping subregions (Blocks) and then extracting local binary pattern (LBP) features from each block. Feature vectors from all blocks are then concatenated to form the final face image feature vector. For each image pixel (central pixel with value 54 in Fig. 5.3), LBP features are extracted by comparing the value of the former with that of its N neighbors within a circle of radian R around the central pixel. The result of comparing the central pixel with a neighbor pixel is a binary number that equals 1 if the central pixel value is lower than the value of the neighbor pixel and 0 otherwise. Grouping all N binary numbers resulting from comparing all N neighbors, we obtain a binary pattern $LBP_{R,N}$ that characterizes the relation between the central pixel and its neighbors. For N=8, for example, $LBP_{R,8}$ is, for each pixel, a word (octet) with 8 binary numbers (for example, $LBP_{R,8} = 11001011_b = 203_{decimal}$, see Fig. 5.3). The histogram of all $LBP_{R,8}$ values, with size equaling $2^8 = 256$, is then computed for each block to form its feature vector. In the same work, Ahonen et al. proposed to use only uniform patterns for LBP computation, namely $LBP^{u2}_{R,N}$. Uniform patterns are binary patterns that contain at most two transitions from 1 to 0 (or from 0 to 1). For example, 11011111 is a uniform pattern, while 11001011 is not. Hence, for N=8, the size of the histogram of LBP features becomes 59: 58 bins for all the possible uniform patterns and 1 bin for the rest of patterns. In this thesis, we adopt the same settings proposed by the authors of MBLBP, namely R=1 and N=8, and each face image is divided into 7×7 subregions.

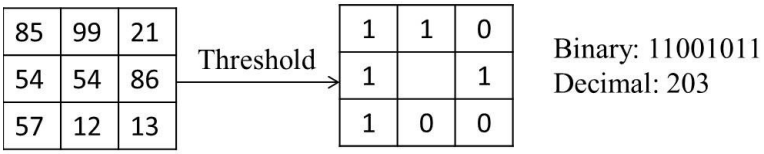


Fig. 5.3 LBP feature extraction for R=1, N=8 (adopted from [1])

5.2.3 The HGPP Algorithm

HGPP was first proposed in [**?**]. Gabor responses at 5 scales and 8 orientations have been computed and encoded for each image. Two encoding approaches have been proposed: one global approach called Global Gabor Phase Patterns (GGPP) and one local approach called Local Gabor Phase Patterns (LGPP). For each image pixel Z, if its Gabor complex response is at scale 'v' and orientation 'u' as $G_{u,v}(Z)$, then the two encoding patterns are defined as

follows:

$$\begin{aligned} GGEP_v^{Re}(Z) &= [P_{0,v}^{Re}, P_{1,v}^{Re}, .., P_{7,v}^{Re}] \\ GGEP_v^{Im}(Z) &= [P_{0,v}^{Im}, P_{1,v}^{Im}, .., P_{7,v}^{Im}] \end{aligned} \tag{5.1}$$

Where "Re" and "Im" designate the real and imaginary part and

$$P_{u,v}^{Re} = \begin{cases} 0, & \text{if } Re(G_{u,v}(Z)) > 0 \\ 1 & \text{if} Re(G_{u,v}(Z)) \leq 0 \end{cases}$$

$$P_{u,v}^{Im} = \begin{cases} 0, & \text{if } Im(G_{u,v}(Z)) > 0 \\ 1 & \text{if} Im(G_{u,v}(Z)) \leq 0 \end{cases}$$

And

$$LGPP_{u,v}^{Re} = [P_{u,v}^{Re}(Z) XORP_{u,v}^{Re}(Z_1), P_{u,v}^{Re}(Z) XORP_{u,v}^{Re}(Z_2) .. P_{u,v}^{Re}(Z) XORP_{u,v}^{Re}(Z_8)]$$

$$LGPP_{u,v}^{Im} = [P_{u,v}^{Im}(Z) XORP_{u,v}^{Im}(Z_1), P_{u,v}^{Im}(Z) XORP_{u,v}^{Im}(Z_2) .. P_{u,v}^{Im}(Z) XORP_{u,v}^{Im}(Z_8)]$$

XOR is a Boolean operator that generates 0 if the two compared numbers are of the same sign and 1 otherwise. As we can see, *GGPP* is computed for all orientations 'u' at a defined scale 'v', while *LGPP* is computed for each orientation 'u' and scale 'v' separately. Finally, $[Z_1..Z_8]$ is the set of neighbor pixels of Z. *GGPP* and *LGPP* patterns for all scales and orientations, for real and imaginary parts, are then concatenated in one bigger vector to form the HGPP feature vector at pixel Z. As in MBLBP, images were subdivided into 64 non-overlapping regions and HGPP is computed for each one of them. All HGPPs are then linked to form the HGPP feature vector of the whole image. For our implementation of HGPP, we have adopted the same configuration proposed in the original work, namely Gabor filters with 5 scales (v=0..4) and 8 orientations (u=0..7), and images were divided into 64 non-overlapping blocks.

5.2.4 The LGBPHS Algorithm

The LGBPHS was first proposed by Zhang et al. in [**?**]. The algorithm is a direct application of MBLBP to magnitude responses of Gabor filters applied at 5 scales and 8 orientations. Feature vectors (histograms) obtained by applying MBLBP to each of the 40 generated images (5 scales $\times$ 8 orientations) are linked to form the final feature vector. Histogram intersection is then used to measure the similarity between images from the gallery and probe database. In our implementation, we have used the same settings as in the origi-

nal work; namely, Gabor filters were applied at 5 scales and 8 orientations and each image (Gabor response) was divided into non-overlapping subregions of size 4 by 8 pixels.

5.3 Results and Discussion

5.3.1 Selected Best Spectral Bands

The Sparsity Based Approach

As mentioned in section 3.2, our sparsity based algorithms BSBS and K-BSBS (non-kernalized and kernalized respectively) select a set of best spectral bands for each of the three studied algorithms, and that by solving the corresponding (relaxed P) and (kernalized P) problems (for example to obtain the non-kernalized best spectral bands for the MBLBP algorithm, the latter should be used to extract feature vectors V that figure in the definition of the (relaxed P) problem (matrix A_I), hence, we find the vector of weights determining the importance of each spectral band when used with MBLBP). We have experimentally chosen to set the same parameters, namely $C = 0.01$ and $\lambda = 0.7$, for all three studied algorithms with and without kernel. Hence, six different vectors of weights; W_{MBLBP}, $W_{MBLBP_{Kernel}}$, W_{HGPP}, $W_{HGPP_{Kernel}}$, W_{LGBPHS} and $W_{LGBPHS_{Kernel}}$ were computed. The bar graphs in Fig. 5.4 show the values of weights per spectral band for each algorithm with and without kernel. From these graphs, we can see that the importance of the involved spectral bands differs significantly from one to another; with most of them having weights lower than 50% of the maximum weight. This fact proves the sparsity aspect claimed for the vector of weights. For our tests, the first two best spectral bands, i.e. bands with the greatest weights, are kept for the upcoming face recognition tasks. Hence, six sets of bets spectral bands are chosen: $(24,25)_{HGPP}$, $(24,25)_{HGPP_{Kernel}}$, $(20,25)_{LGBPHS}$, $(20,25)_{LGBPHS_{Kernel}}$, $(24,25)_{MBLBP}$ and $(24,25)_{MBLBP_{Kernel}}$. Note that we have used numbers from 1 (equivalent to band at 480nm) to 25 (equivalent to band at 720nm) to designate spectral bands, instead of using their wavelengths. Note also that despite we got the same best channels for algorithms HGPP and MBLBP (and the corresponding kernalized versions as well), the corresponding weights of each set are different: (W(24)= 0.134, W(25)= 0.146) for $(24,25)_{HGPP}$ and (W(24) = 0.138, W(25) = 0.179) for $(24,25)_{MBLBP}$. The second major revelation from the bar graphs is that the kernel based algorithms gave the same set of best spectral bands but with two main differences:

1. Weights assigned to spectral bands, are different between kernel based and non-kernel

based algorithms. We will see later how this fact will affect the recognition performance during weights based spectral bands fusion.

2. The variances of vector of weights of the kernalized algorithms are higher, (except for HGPP), than the variances of the same vectors with non-kernalized algorithms, see Table 5.1. Which means that in the projected space, obtained using the Gaussian kernel, the discrimination between different spectral bands is more effective; best spectral bands got significantly higher weights then the rest of bands. Note that the variances are computed for the last 10 elements of each vector (corresponding to band 16 to band 25). This restriction enables to compute variances around the selected best spectral bands only, and hence, make the variance measure more informative regarding the differences between weights obtained for the selected best spectral bands and the other neighborhood bands. The selected spectral bands for each algorithm are then fused in different ways for face matching, which is studied in the next section.

The Multilinear Decomposition Based Approach

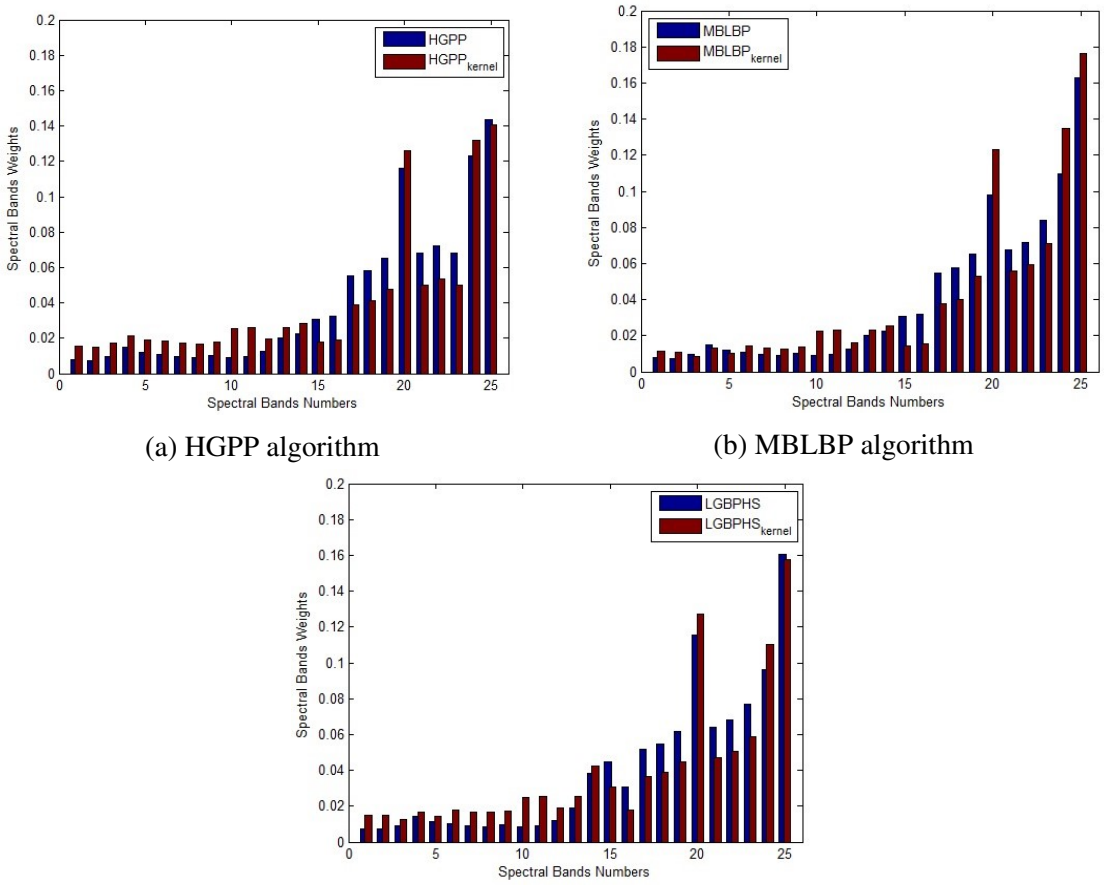
In section 3.3, we have revealed that the spectral bands with the smallest value of dK is the less affected by illumination and hence the best for face recognition. we call dK the robustness to illumination factor or RIF. After computing the RIF of all spectral bands providd by the $IRIS-M^3$ face database, we found out that spectral bands number 25(720nm) and 20(690nm) had the lowest RIF for both involved algorithms MBLBP and HGPP while for the LGBPHS algorithm the spectral bands with the lowest RIF were 25 and 23(700nm). Hence, these later bands were chosen as the best spectral bands for the corresponding algorithms under the defined illumination conditions(daylighted images). For MBLBP and HGPP, the obtained results are consistent with those obtained in [13] (see section 5.3.1). However, for LGBPHS the spectral band number 23 is now selected as being better then band 20 obtained previously by the sparse approach (see section 5.3.1). This may be explained by the high sensitivity of the RIF factor to noise. Which recommend the use of filtering component before computing the RIF factor for each spectral band.

The Boosted LDA Approach

Our boosted LDA algorithm selects a different set of best spectral bands for each of the four studied algorithms and for each face database; learning weak classifiers, determining boosting weights, and selecting best spectral bands are all processes that are repeated for

Table 5.1 Table of variances (var) of obtained vectors of weights (10^{-4})

Vectors	W_{MBLBP}	$W_{MBLBP_{Kernel}}$	W_{HGPP}
Variances	18	19	16
Vectors	$W_{HGPP_{Kernel}}$	W_{LGBPHS}	$W_{LGBPHS_{Kernel}}$
Variances	14	17	19



(a) HGPP algorithm

(b) MBLBP algorithm

(c) LGBPHS algorithm

Fig. 5.4 Weights assigned to spectral bands for algorithms (a) HGPP, b) MBLBP and c) LGBPHS

each algorithm. In our approach, a spectral band is judged better than another spectral band if the weight of the corresponding weak classifier, inside the final strong classifier, is the greatest. This rule is not applicable unless we can demonstrate the equivalent relationship between bands' quality and weak classifiers' performances (third property). With one face database and three algorithms, we built 3 strong classifiers H using our Adaboost algorithm as defined in section 3.4. Three vector of weights that determine the importance (weight) of each spectral band inside the corresponding classifier and for the $IRIS-M^3$ face database. Bar graphs of these vectors of weights are shown in Fig. 5.5. The later figure shows the bar graph of a fourth feature extraction algorithm which is POEM-WPCA. This algorithm is beyond the scope of this thesis and was included just for comparison raisons (work not published yet). We can see from figure 5.5 that some weak classifiers have considerably higher weights than others. For example, the vector of weights obtained for the HGPP algorithm upon the $IRIS-M^3$ database has spectral bands/weak classifiers number 25 (720nm), 24 (710nm) and 20 (670nm) with weights values equal to 0.142, 0.122, and 0.114 respectively. The sum of these values is 0.368, which is 36.80 % the sum of all the 25 weights defining the corresponding vector of weights (weights are normalized to equal 1). On the other hand, the rest of all spectral bands have weights lower than the half of the maximum weight affected in this case to spectral band number 25. Having this concentration of weight in only three spectral bands of 25 explain our motivation either to select the best spectral bands for face recognition or to use Adaboost for this purpose. The latter enabled a clear separation of recognition abilities between different spectral bands and hence makes best bands selection easier. Having the greatest weights, we have selected spectral bands number 25, 24 and 20 as the best spectral bands for HGPP upon the $IRIS-M^3$ database. With the same analysis of Fig. 5.5 and by searching for the first three greatest weights, we determine the set of best spectral bands for each algorithm and with each database. Obtained sets are: $MBLBP = (25,24,20)$, $HGPP = (25,24,20)$ and $LGBPHS = (25,20,24)$.

The earlier a number is in its set, the best the corresponding spectral band is. For example, in the *LGBPHS* set, the number 20 is written before the number 24, so the spectral band number 20 has greater weight. As a result, its recognition quality is better than that of spectral band number 24. From the selected sets of best bands, we can see that the same best spectral bands have been chosen for all algorithms, but with different weights. These bands are (25,24,20). However, spectral bands that have shown the best individual recognition performance upon the $IRIS-M^3$ database for algorithms MBLBP, HGPP, and LGBPHS, namely, bands numbers 14, 15 and 16, have not been selected within the set of best spectral bands when included in our Adaboost algorithm. This observation may be explained by the fact that the difference in individual performances between the best individual spectral bands

(spectral bands tested alone) and the best boosted spectral bands (selected sets) upon the $IRIS-M^3$ database was small (for example, the spectral band number 15 had an advantage of 1% accuracy over spectral band number 24). This small difference was not able to persist during the boosting process and spectral band number 24 was weighted higher than spectral band number 15.

The Dynamic Approach

When capturing face images in outdoor conditions (sun lighted images), lighting direction and intensity may change by time, and hence, best spectral bands selected for the first imaged person are not necessarily those optimal for the last imaged person. A dynamic system that selects different bands for different persons is required and should intuitively give better results than a static system. In Table 5.2, we show the best bands selected for the first two subjects from the $IRIS-M^3$ face database, during the test experiment, with the corresponding likelihood ratio LR . Note that spectral bands in Table 5.2 are presented by their numbers with the corresponding LR value in parenthesis. We can see that the greatest LR coefficients were obtained for the HGPP algorithm. This is related to the fact that the overlapping region between the two PDFs obtained for this algorithm is the smallest between the other algorithms (see Fig. 4.10): the smaller the overlapping region is, the bigger the difference between the two probabilities obtained for a given spectral band to be good and to be bad is. And so the bigger their likelihood ratio will be. The size of the overlapping region between the two PDFs, F_{good} and F_{bad}, is mainly related to the definition of the quality factor Q. A good quality factor that discriminates well between good and bad images should keep the overlapping region at its smallest size. In this case, F_{good} should be centred on the value 0.5, while F_{bad} should be centered on the values 1 or 0. Since LGBPHS, HGPP and the quality factor Q were all defined using multi-scale wavelets transformation (Gabor wavelets for LGBPHS and HGPP, and redundant discrete wavelet transform for Q) overlapping regions obtained for these algorithms were smaller than that obtained for MBLBP, while their LRs were greater. We say that the definition of Q performs better with LGBPHS and HGPP than with MBLBP.

5.3.2 Fusion Of Selected Spectral Bands

After best spectral bands have been selected, the second step to enhance the accuracy of our three studied algorithms is to choose the best fusion technique to fuse these bands. Four weighted fusion approaches are considered; fusion using three different configuration of

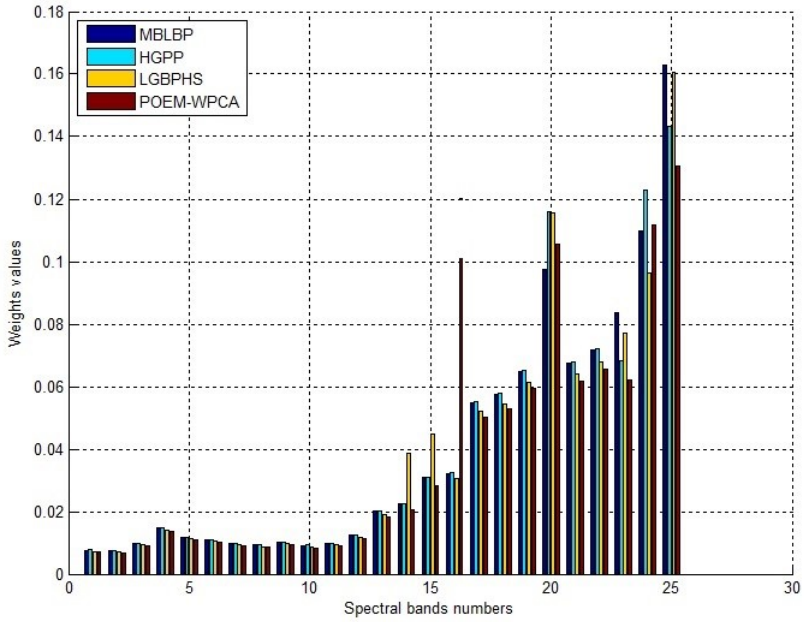


Fig. 5.5 Weights affected to each spectral band (weak classifier) of the $IRIS-M^3$ database and for each algorithm

Table 5.2 Best spectral bands selected for the first two subject with the corresponding LR

	Subject 1		**Subject 2**	
Algorithms	BSBS	second BSBS	BSBS	second BSBS
HGPP	25(2.3)	23 (2.1)	25 (2.33)	24 (2.12)
LGBPHS	25 (2.1)	23 (1.98)	24 (2.1)	23 (2.1)
MBLBP	25 (1.76)	22 (1.73)	24 (1.89)	22 (1.80)

Gabor wavelets (G_3), (G_4) and (G_5), and fusion using 3-level Haar wavelets decomposition (H_3). Weighted fusion at wavelet space is widely used in the literature and several approaches to determine the optimal weights have been proposed [108] [164] [117]. We believe that weights determined during best spectral bands selection, namely the vector W, characterize well the recognition performance of their corresponding spectral bands and, hence, could be used to fuse these later. Our four fusion techniques have the same principle: the wavelet decomposition technique is first applied at a given number of levels (scales), then, for two spectral bands to be fused, pixels at the same level and same position are summed with weights of their corresponding spectral bands. Formally, if we note $P_k^n(x,y)$ and $P_{k1}^n(x,y)$ two pixels at decomposition level n and position (x,y) corresponding to the two spectral bands k and k1 to be fused. With these later having weights of importance w_k and w_{k1}, respectively, the fused pixel at position (x,y) could be defined as:

$$P_{fused}^n(x,y) = w_k.P_k^n(x,y) + w_{k1}.P_{k1}^n(x,y) \tag{5.2}$$

The inverse wavelet decomposition is then applied to get the final fused spectral band. Note that for each algorithm, only selected best spectral bands are fused, which reduce significantly the processing time and maximizes the amount of relevant information retrieved. Intuitively, the best fusion techniques differ with the set of spectral bands used which is related to the BSBS system to be used. In this section, we adopt the sparsity based approach kernalized and non-kernalized and we determine the corresponding best fusion technique. For the other BSBS systems proposed in this thesis, the optimal fusion methods could be determined similarly. Rank-1 recognition rates obtained with the different fusion techniques are summarized in Table 5.3 and Table 5.4 for non-kernalized and kernalized algorithms. Several results can be highlighted:

1. Gabor wavelets based fusion outperformed Haar wavelets based fusion. The literature revealed the similarity of Gabor functions to the mammalian visual cortex, which make these functions optimal for image analysis.

2. Gabor decomposition at 5 scales and 8 orientations outperformed the other Gabor based fusion approaches. The higher the number of scales is, the better the performance would be. However, the processing time may increase enough to make the final system non-real time.

3. Kernelized algorithms outperformed their non-kernelized versions. The superiority was between 2% and 4% when using the G_5 fusion. In this stage, we can see how the difference of obtained weights for each algorithm, had increased the recognition

performance of these later.

Based on the above results and unless mentioned, the fusion technique G_5 will be used with all BSBS approaches.

5.3.3 Enhancing The Performances Of Features Extraction Algorithms

The Sparsity Based Approach

In section 3.2, we have defined our modified fisher crietrion (MFC) that measures the discrimination of the matrix of match scores. The obtained measures of MFC for the three algorithms MBLBP, HGPP and LGBPHS are displayed in Fig. 5.6 , with M, number of probe images, is equal to 70. We can see that the MFCs of the kernalised match score matrices are higher than in the non-kernalized versions. In this later, we can note that their MFC curves (green curves) are oscillating around the value 1 (blue curves), which indicates that the within class and between class matrices S_w and S_b are continuously close to each other. However, the black MFC curves of the kernalized versions are far from the level 1 and hence show up a clear discrimination between both covariance matrices. These observation prove the effectiveness of the kernalization approach to increase the sparsity of the match score matrices, which will have a direct impact on the sparsity of W and the time for convergence of the (kernalized P) problem.

The set of spectral bands selected by the sparsity based approach can be investigated in several tasks like face detection, image fusion, image enhancement or the subject of this section which is face matching. Enhancing face matching of the studied algorithms MBLBP, HGPP and LGBPHS is performed by applying these algorithms on the right (best) spectral bands instead of applying them either on the whole set of spectral bands provided by the database, which is time consuming, or on usual gray/RGB images.

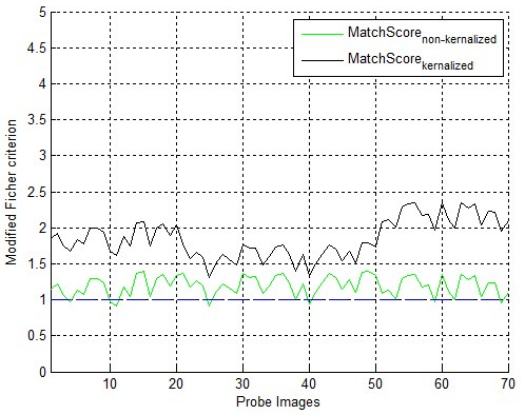
As the performances of the studied algorithms were evaluated, in their original work, upon the FERET face database, we first had to test our implementations of these algorithms upon the same database. In this way, we ensure that our implementations are consistent with those proposed by the original authors. As we can see in Table 5.5, our implementations gave slightly better results than those reported in the original works for the same sets of images (image sets Fb and Fc from FERET face database). This difference is due to the changes we have made on the parameters of the algorithms (Images size, Gabor filters size, etc). Hence, our implementations are verified and could be used for further tests.

Table 5.3 Rank-1 recognition rates (%) for different fusion methods and without kernalization

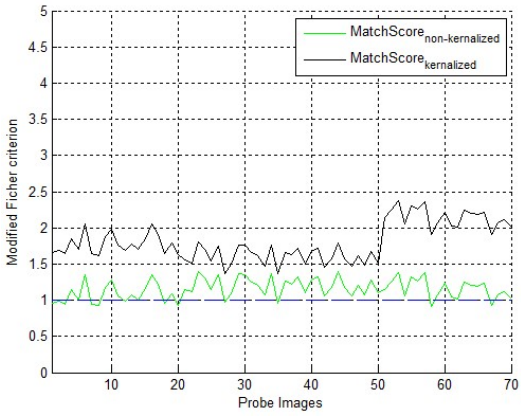
	Fusion Method			
Algorithms	G_3	G_3	G_3	H_3
MBLBP	63.24	64.90	66.15	62.00
HGPP	69.90	70.01	71.10	69.15
LGBPHS	64.45	64.70	65.15	63.95

Table 5.4 Rank-1 recognition rates (%) for different fusion methods with kernalization

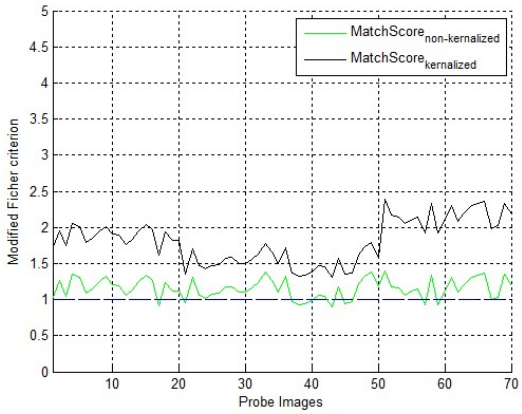
	Fusion Method			
Algorithms	G_3	G_3	G_3	H_3
MBLBP	65.34	67.50	69.18	64.90
HGPP	72.42	72.89	73.25	72.19
LGBPHS	68.60	69.19	69.75	66.23



(a) HGPP algorithm



(b) MBLBP algorithm



(c) LGBPHS algorithm

Fig. 5.6 MFC obtained for the (a) HGPP, b) MBLBP and c) LGBPHS

The studied algorithms, were then tested on the gray images (daylighted images presenting high illumination variation) provided by the $IRIS-M^3$ face database. Rank-1 recognition results are summarized in Table 5.6. We can see that a decrease of at least 46% of accuracy is detected compared to tests upon the Fb image set. Direct corollaries of this result are:

1. the $IRIS-M^3$ face database is more challenging than FERET database, and could be used, once enlarged, as a good simulation of the real world imaging conditions, particularly imaging in sun lighted environment.
2. it is important to test the existing state-of-the-art algorithms upon more challenging databases to determine the limit of their effectiveness in determined imaging conditions (sun lighting, pose variation of more than 60 degree, occluded face landmarks.
3. the three tested algorithms, with accuracies lower than 54%, are very far from being suitable to work upon the IRIS-M3 face database. Therefore, we proposed to enhance them using the multispectral images based approaches.

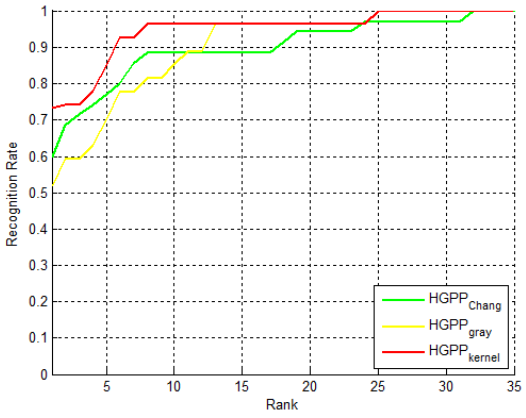
Finally, the performances of the studied algorithms obtained using multispectral images selected by the kernalized approach are compared to two other scenarios: using gray images provided by the $IRIS-M^3$ face database (see Table. 5.6), and using the S_{Chang} set of best spectral bands obtained by Chang et al. The aim of the first scenario is to demonstrate the effectiveness of using visible multispectral images instead of gray images, while the second test shows the superiority of our kernalized algorithm over Chang et al. approach. For the S_{Chang} set, as no weights are affected to the selected spectral bands, we adopt the technique for image matching as proposed by Chang et al. in [25](see section 2.2.1). For each image from the database, chosen spectral bands are fused using the Haar wavelet transform. Feature vectors are then extracted from the fused image by the tested algorithm (HGPP, LGBPHS, or MBLBP) and matched against each other by the Euclidian distance. For our algorithm, selected spectral bands forming the set S_{kernel} are fused using the G5 approach and then match scores are computed using the Euclidian distance as well. The obtained Cumulative Match Curves (CMC) are shown in Fig. 5.7 while the rank-1 recognition rates are displayed in Table. 5.7 As we have expected, both approaches based on multispectral images gave better results than those based on gray images. This could be explained by an intrinsic robustness of the selected best spectral bands against extreme lighting variation compared to broad band images. The former are less sensitive to the illumination conditions and preserve much more face signatures from lighting condition to another. We can see, also, that our approach surpasses the Chang et al. algorithm in all tests. As mentioned above, the theoretical advantage of our approach, here proven experimentally, is that we can evaluate the performance of each spectral band based on its behavior inside the whole set of

Table 5.5 Comparison of rank-1 recognition rates (%) between the original and our implementations of the studied algorithms

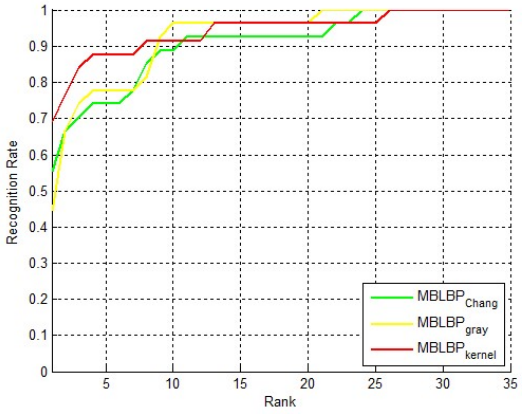
	Original Implementation		**Our Implementation**	
Algorithms	Fb	Fc	Fb	Fc
HGPP	97.60	98.90	98.51	99.12
LGBPHS	94	97	93.03	96.93
MBLBP	93	51	93	51

Table 5.6 Comparison of rank-1 recognition rates (%) between the original and our implementations of the studied algorithms

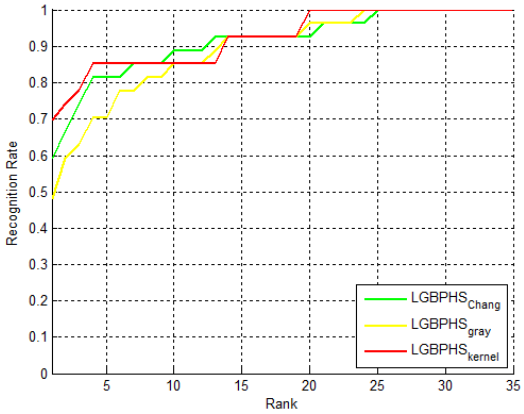
	Studied Algorithms		
Data Base	MBLBP	LGBPHS	HGPP
$IRIS - M^3$	44.44	48.15	51.85



(a) HGPP algorithm



(b) MBLBP algorithm



(c) LGBPHS algorithm

Fig. 5.7 CMC curves obtained using gray images, S_{Chang} set and S_{kernel} set for algorithms (a) HGPP, b) MBLBP and c) LGBPHS

spectral bands, while in the Chang et al. algorithm, each spectral band is tested separately. We can see that the performance of the studied algorithms HGPP, LGBPHs and MBLBP increased with 16.15%, 14.81% and 21.57% respectively when using multispectral bands selected with K_{kernel} instead of using gray images. This significant enhancement demonstrates the effectiveness of the multispectral based approach for illumination related problems in face recognition. Finally, the maximum accuracy obtained was 74.77%, which is still by far lower than the results obtained on the FERET database. This difference indicates the high level of difficulties that sun-lighted faces presents to the exiting state-of-the-art algorithms, even when augmented with multispectral images and projection tools like the kernelisation process.

The Multilinear Decomposition Based Approach

In section 5.3.1, we have used our multilinear decomposition based approach to select the best spectral bands from the $IRIS-M^3$ face databased. The two spectral bands selected were band SB25 (720nm) and band SB20 (670nm) for both algorithms HGP and MBLBP while for LGBPHS selected bands were SB25 and SB23(700nm). To compute match scores, the distance D between two multispectral images $I^p = (I^p_{SB25}, I^p_{SB20})$ and $I^g = (I^g_{SB25}, I^g_{SB20})$ from the probe and gallery databases respectively, was computed as:

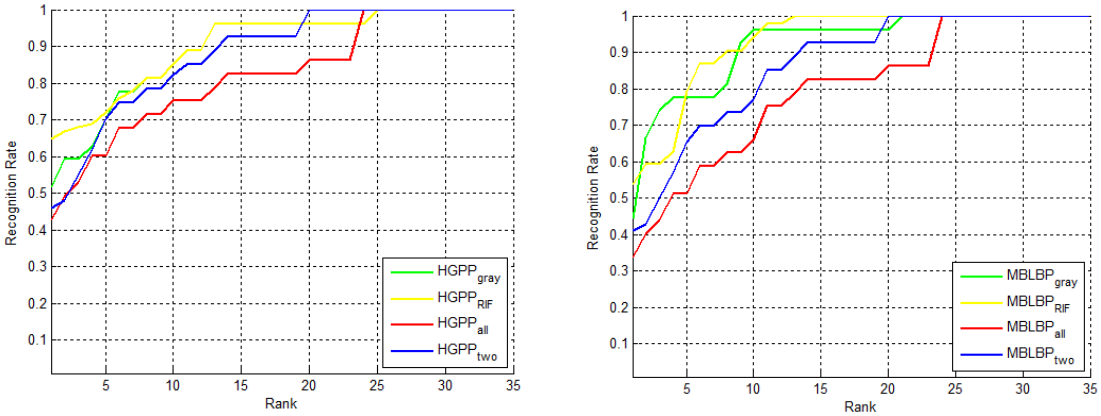
$$D(I^g, I^p) = \omega_{SB25}.\|I^g_{SB25} - I^p_{SB25}\|_2 + \omega_{SB20}.\|I^g_{SB20} - I^p_{SB20}\|_2 \tag{5.3}$$

$\omega_{SB25} = 0.035$ and $\omega_{SB20} = 0.05$ are RIFs obtained for SB25 and SB20 respectively. The proposed approach with its two steps: best bands selection using RIF and spectral bands fusion at match score level using Eq 5.3, is compared to three other basic approaches which are : using gray images, using randomly selected two spectral bands S_{two} (in our case we have chosen $SB12$ and $SB19$) and using all the 25 spectral bands S_{all} for face matching. For the multispectral images based approaches, each spectral band is weighted by its RIF and summed similarly as in Eq. 5.3. CMC curves in Fig. 5.8 and rank-1 recognition rates in Table 5.8 summarize the obtained results.

we can see that using all or randomly chosen spectral bands gave bad results compared to using broad band images, while using selected best spectral bands gave the best performances. We conclude from this, that a multispectral images based face recognition system is inefficient and may be very time consuming, unless its preceded by a good system/phase for best spectral bands selection. Our approach has increased the recognition performance with 10% and 14% for MBLBP and HGPP algorithms respectively, which promote the use

Table 5.7 Rank-1 recognition rates (%) for different selected channels

	Selected Channels			
Algorithms	Gray	S_{Chang}	S_{kernel}	Performance increase over using Gray images
MBLBP	44.44	55.56	**62.29**	17.85
HGPP	51.85	60	**68**	16.15
LGBPHS	48.15	59.26	**62.96**	14.81



(a) HGPP algorithm (b) MBLBP algorithm

Fig. 5.8 CMC curves obtained using different set of selected spectral bands for(a) HGPP and b) MBLBP

Table 5.8 Rank-1 recognition rates (%) of studied algorithms

	RIF	gray	two	all
HGPP	65	51	46	43
MBLBP	54	44	41	34

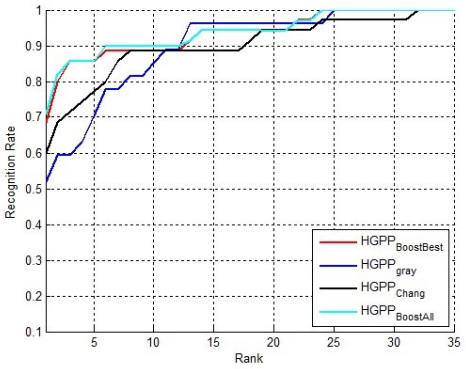
of multispectral images for illumination related problems in face recognition.

The Boosted LDA Approach

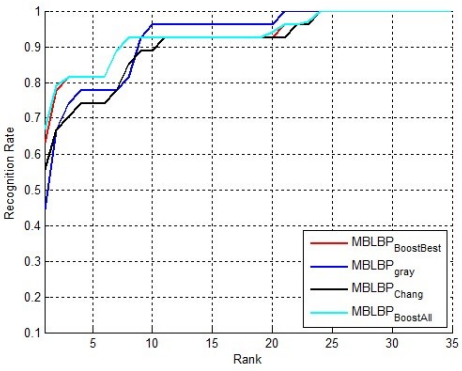
The sets of best spectral bands selected by our algorithm in section 5.3.1 and the $S_{Chang} = (610nm, 620nm)$ set selected by the Chang algorithm are used to study the advantages of the multispectral approach over the gray images-based approach. For the S_{Chang} image set, selected spectral bands at 610nm and 620nm are fused using Haar wavelet transform, then feature vectors are extracted from the fused image using algorithms HGPP, LGBPH or MBLBP based on the tested algorithm. The city block distance is used then to measure distances/match scores between extracted features vectors. Approaches based on the S_{Chang} set and gray images are to be compared to each other and against our boosted LDA-based strong classifiers. The latter, as shown in section 3.4.2, are a weighted combination of spectral bands-based weak classifiers built using LDA. Our stronger classifiers could be used in two different ways that we investigate in this study: either we use all spectral bands with their corresponding weights (referred to as 'Boost-All'), or we use only the selected sets of best spectral bands with their weights too (referred to as 'Boost-Best'). Evidently, Boost-All will consume much more processing time than Boost-Best, since the former process N (N=25 for the IRIS-M3 database) spectral bands for each subject, while the latter process only three spectral bands per subject. With the available assortment of possible configurations and image sets, 12 (4×3) systems are to be studied is this section, which are: $X^{IRIS}_{Boost-Best}$, $X^{IRIS}_{Boost-All}$, X^{IRIS}_{Chang} and X^{IRIS}_{gray}. The letter X designates the algorithm used for feature extraction: HGPP, LGBPH or MBLBP.

From Fig. 5.9, we show the CMC curves obtained for the above studied systems grouped by algorithms. While in Table 5.9, rank-1 recognition rates of the same systems are displayed. We can see that in all tests, algorithms based on the multispectral approach, either with Chang et al. algorithm or with our algorithm gave better results than using gray images. This observation highlights the effectiveness of combining spectral images at different wavelengths to enhance system accuracy in different challenging conditions. The analysis of CMC curves also shows that results obtained by our boosted LDA systems over-passed those obtained by the Chang et al. algorithm in all tests: this superiority proves the advantage of giving different weights of importance to different spectral bands as in our Adaboost-based approach, instead of giving the same importance to all selected spectral bands as proposed by the Chang et al. algorithm.

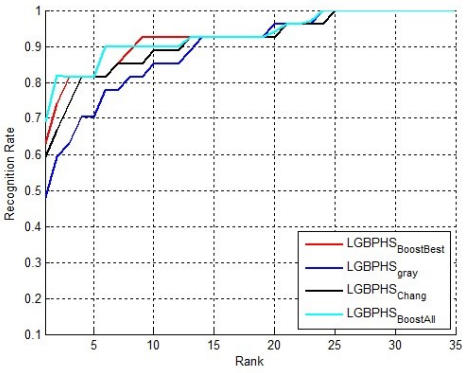
In Table. 5.10, we have reported the rank-1 recognition rates for 'Boost-All' and 'Boost-



(a) HGPP algorithm



(b) MBLBP algorithm



(c) LGBPHS algorithm

Fig. 5.9 CMC curves obtained upon the $IRIS-M^3$ database for algorithms (a) HGPP, b) MBLBP and c) LGBPHS

Table 5.9 Rank-1 recognition rates (%) of boost-all and boost-best based algorithms upon the $IRIS-M^3$ database

	HGPP	MBLBP	LGBPHS
Boost-Best	68.00	62.96	62.29
Boost-All	71.12	69.00	67.00
Difference of accuracy	3.12	6.04	4.71

Best' based algorithms upon both databases. We can see that results are very close with an expected continuous superiority of Boost-All algorithms over Boost-Best algorithms. The difference in accuracy between these two configurations of our boosted LDA systems have not exceeded, in most cases, 5 % for the IRIS-M3 database. On the other hand, Boost-Best based algorithms are requiring much less processing time than Boost-All based algorithms. Hence, using only a small set of well-selected spectral bands as in Boost-Best based algorithms is sufficient to ensure a good accuracy while keeping a reasonable processing time. For algorithm LGBPHS, when applied to the $IRIS-M^3$ database, a significant decrease of more than 6% accuracy is detected when passing from using Boost-All configuration to using Boost-Best configuration. In this case, using only best spectral bands is not efficient and all spectral bands should be included with their corresponding weights determined during boosting process.

As the Adaboost algorithm is injecting weak classifiers one by one during strong classifier building, an important question arises: Does the order of injection of a given weak classifier influences its importance within the final strong classifier? In other words, does the weight of a given weak classifier change based on whether the latter was injected first, third or tenth, for example? To answer this question we have plotted, in Fig. 5.10, the variation of values of best three weights affected to the best three weak classifiers, when the order of injection of the latter varies from 1 to 10 during the boosting process, with an $err_{tolerated} = 103$. Reported values of weights correspond to the HGPP algorithm as an example and are determined for both databases. Two main results can be revealed. The first is that regardless of which weak classifier we started with, the same ranking of these later was obtained. The second result is that weights affected to these best classifiers preserved, roughly, constant values when varying the order of injection of their corresponding weak classifiers. With these two results verified also for all the other 2 algorithms MBLBP and LGBPHS, we can say that our Adaboost algorithm is converging to a fixed configuration independently with which weak classifier we have started boosting.

The Dynamic Approach

In this section, our dynamic BSBS system (DBSS) is evaluated against two scenarios; matching faces using gray images, and matching faces using spectral bands selected by the sparsity based approch seen in section 5.3.1 . The DBSS algorithm has a training phase and a test phase. The repartition of face images involved in each phase is displayed in Table 5.11, with one MI image for each subject. During the training phase, PDF functions are determined as explained in section 4.3. During the test phase, each new multispectral

Table 5.10 Rank-1 recognition rates (%) using sets of best spectral bands upon the $IRIS-M^3$ database

	Selected Sets			
Algorithms	Gray	S_{Chang}	$S_{Boost-Best}$	$S_{Boost-All}$
MBLBP	44.44	55.56	62.29	67
HGPP	51.85	60	68.00	71.12
LGBPHS	48.15	59.26	62.29	69

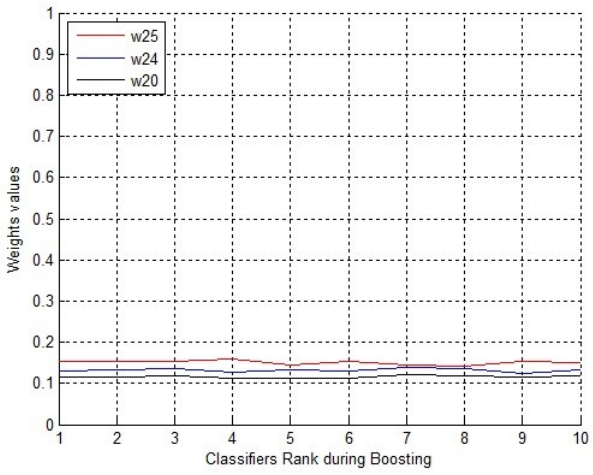


Fig. 5.10 Variation of values affected to the best three weights (w_{25}, w_{24}, w_{20}) with the inclusion order of their corresponding weak classifers: case of HGPP algorithm upon the $IRIS-M^3$ database

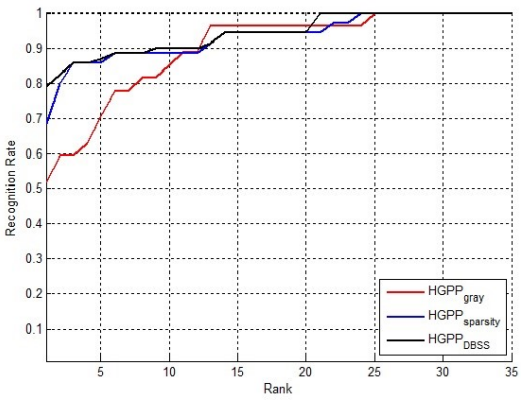
image , with 25 spectral bands, is presented to the system separately. The likelihood ratio *LR* of each of its spectral bands is computed and the two bands with the two greatest *LRs* are then slected. Selected bands were then fused using G_5 and the match scores were computed using the Euclidian distance. The obtained CMC curves are plotted in Fig. 5.11 and Rank-1 recognition results are summarized in Table 5.12. We can see that the performances of the studied algorithms were enhanced with the use of multispectral images instead of gray images: 21.66 % increase of accuracy for MBLBP, 27.25 % for HGPP and 21.85 % for LGBPHS. This fact supports the choice of MI to solve issues related to high light variation in face recognition tasks. The CMC curves shows also the superiority of our DBSS algorithm over the sparsity based algorithm with an increase of 3.81 % for MBLBP, 11.10 % for HGPP and 7.04 % for LGBPHS. This superiority is due to the fact that when capturing face images in outdoor conditions (sun lighted images), lighting direction and intensity may change by time, and hence, best spectral bands selected for the first imaged person are not necessarily those optimal for the last imaged person. A dynamic system that selects different bands for different persons is required and should intuitively give better results than a static system.

5.4 A Comparative Study of the Different BSBS systems

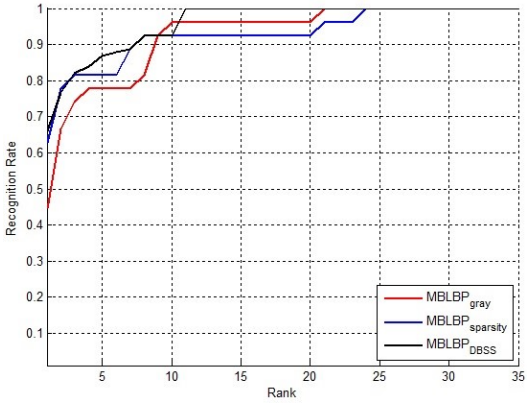
This is the final section of our experimental part and is dedicated to compare the performance of the proposed BSBS systems under the same conditions [11]. The aim of this section is clear, we provide to our systems a common environment of lighting conditions and face databases and determine which technique will perform better. For each algorithm MBLBP, HGPP and LGBPHS, the set of selected best spectral bands, using each of the BSBS algorithms, are displayed in Table 5.13. Abbreviations SA, MA, CHA and DA that figure in this table designates respectively the sparsity, the multilinear decomposition, the Chang and the dynamic algorithms for BSBS. Note that CHA selects the same best spectral bands, namely bands 14 and 15, independently of the algorithm used. On the other hand, the MA algorithm, originally applied on MBLBP and HGPP only, is used in this work, with the same settings as in [13], to select the best spectral bands for the LGBPHS algorithm. For the DA, as different spectral bands are selected for each subject, we display only those determined for the first subject in the database. We have chosen to remove the adaboost approach from this comparative study since this approach is not yet published and may be subject of further enhancement. Despite we have tested our systems on more then one databse, the $IRIS-M^3$ face database was found to be more relevant and challenging, this

Table 5.11 Repartition of multispectral images between training and test sets

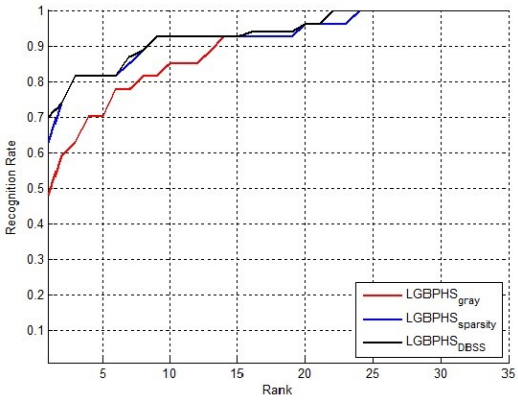
	Training Data	Test Data
Probe Set	15 day lighted + 20 fluorescent lighted	20 day lighted + 15 fluorescent lighted
Gallery Set	35 Halogen lighted	35 Halogen lighted



(a) HGPP algorithm



(b) MBLBP algorithm



(c) LGBPHS algorithm

Fig. 5.11 CMC curves obtained upon the $IRIS-M^3$ database for algorithms (a) HGPP, b) MBLBP and c) LGBPHS

Table 5.12 Rank-1 recognition rates(%) using different algorithms for best spectral bands selection

	Selected Channels			
Algorithms	Gray	DBSS	Sparsity approach	Performance increase over using Gray images
MBLBP	44.44	66.10	62.29	21.66
HGPP	51.85	79.10	68	27.25
LGBPHS	48.15	70.00	62.96	21.85

way we reduce the validation time and ensure a more realistic imaging conditions provided by the daylighted images of the database. The different sets of best spectral bands are then

Table 5.13 Best spectral bands selected for each studied algorithm.

	Approaches for best spectral bands selection			
	SA	MA	CHA	DA
MBLBP	24,25	25,20	14,15	25,22
HGPP	24,25	25,20	14,15	25,23
LGBPHS	20,25	25,23	14,15	25,23

used for the corresponding algorithm MBLBP, HGPP or LGBPHS to match faces from the $IRIS-M^3$ database.

The distance D used for match scores computation between two multispectral images $I^p = (I_1^p, I_2^p)$ and $I^g = (I_1^g, I_2^g)$ from the probe and gallery databases respectively, is computed as fellows:

$$D(I^g, I^p) = \frac{1}{2}.(\|I_1^g - I_1^p\|_2 + \|I_2^g - I_2^p\|_2) \tag{5.4}$$

the subscripts 1 and 2 that figure in the Eq. 5.4, designates the two best spectral bands selected for the considered algorithm. Rank-1 recognition rates for each algorithm are shown in Table. 5.14, while CMC curves are displayed in Fig. 5.12 to Fig. 5.14.

The results of applying the three studied algorithms on gray images, provided by the same database, are also included in our tests to show up the advantage of using multispectral images instead of broad band images to enhance the recognition performances under high illumination variation. As expected, the results obtained using the dynamic BSBS algorithm outperformed all the other best bands selection algorithms. The sparsity based algorithm gave the second best rank-1 recognition rates, while multilinear and Chang algorithms gave closer results. The best static BSBS algorithms determined for MBLBP, HGPP and LGBPHS were respectively JB, SB and SB. We can see that the performance of the BSBS algorithm and the recognition algorithms like MBLBP or HGPP are mutually related. This suggest a careful selection of both components of face recognition systems .

5.5 Conclusion

The experimental tests conducted in this chapter were classified into three groups; tests that determines the set of best spectral bands for each features extraction algorithm using the

Table 5.14 Rank-1 recognition rates (%) using different sets of selected best spectral bands

	Approaches for best spectral bands selection				
	SA	MA	CHA	DA	gray
MBLBP	62,29	54	55,56	66.10	44,44
HGPP	68	65	60	79.10	51,85
LGBPHS	62,96	59	59,26	70	48,15

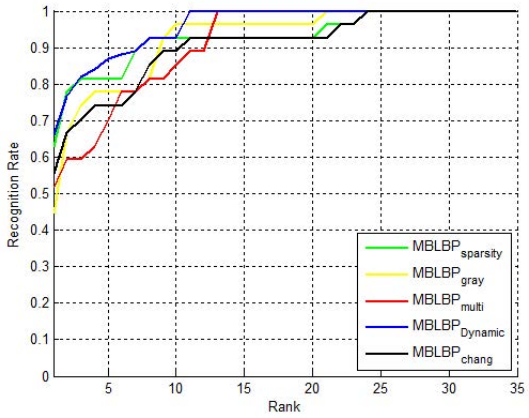


Fig. 5.12 CMC curves obtained for MBLBP using the different BSBS approaches

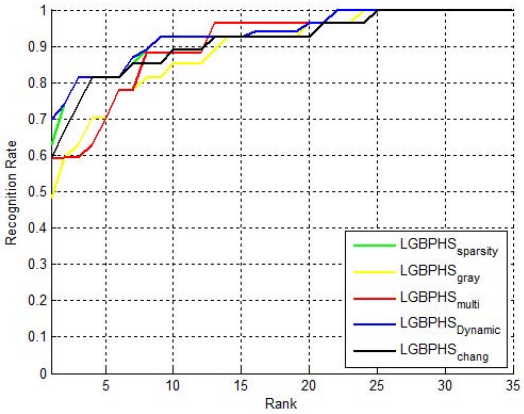


Fig. 5.13 CMC curves obtained for LGBPHS using the different BSBS approaches

different BSBS systems, tests that study the different fusion approaches that could be investigated with MI images to enhance the images quality and utility and finally a group of tests that used the fused images to enhance the performance of MBLBP, HGPP and LGBPHS algorithms. Each tests group was made independent of the other two groups intentionally to ease the reading of this chapter. The results highlited in this chapter are interesting and can be used for different recognition systems involving not only face matching but different other biometrics like fingerprints and so on.

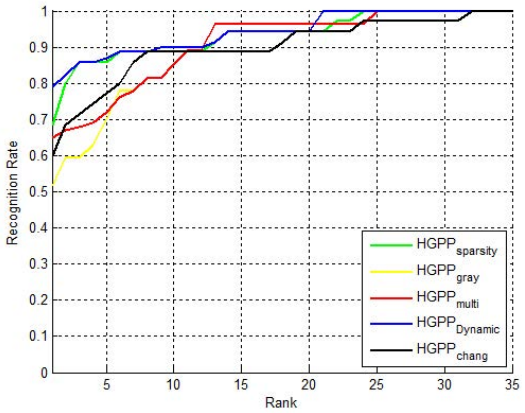


Fig. 5.14 CMC curves obtained for HGPP using the different BSBS approaches

We start by the main outcome of our work , which is the different BSBS systems proposed. These systems were based on a set of optimization techniques like the basis pursuit algorithm and linear discriminant analysis. Once these optimization problems were solved, the different spectral bands were affected different weights of importance based on the considered imaging conditions. The fact that the importance of each band is quantized with a weight, which is a simple real value, made possible the use exting optimization tools that are not originally made for images processing. This independence of the type and size of images used make from our BSBS systems applicable to any kind of tasks involving bands selection even outside the environment of object recognition. Four static algorithms and one dynamic algorithms have been proposed for the selection of best bands for each feature extraction approach. The obtained best bands were highlighted in table. 5.13. Two sets of best bands have shown to be the best, the set of bands 14 and 15 obtained by the Chang algorithm and the set of bands between 20 and 25 obtained by systems. The selection of spectral bands from these two intervals of the light spectrum guarantee a good performance in extreme lighting conditions and is advised for systems working in outside environments (sun light).

The section on fusion MI images was a small but interesting step toward understanding the particularity of such problem. The experimented fusion techniques were challenged against each others to see their performances and determine the suitable one for each feature extraction technique. Four fusion approaches have been tested which are G_3 , G_4 and G_5 that used Gabor wavelet decomposition at three levels, four levels and five levels respectively, and a Haar wavelet decomposition based approach at three levels. The performances of Gabor based techniques has shown a superiority over the Haar base approach. The performance increase with the number of decomposition levels and G_5 achieved the best accuracy. How-

ever, the computation time increases as well with the number of decomposition level and the tradeoff between time complexity/recognition results should be considered.

The last set of experiments had the objective to finally validate the advantages of using MI images instead of grayscale images. The algorithms MBLBP, HGPP and LGBPHS are state of the art and have been used for long time under controlled conditions. The performances of these algorithms dropped significantly when tested with images under daylight illumination. We proposed the use of MI images to solve this issue and the obtained results were extremely significant. The increase of performance was in several times higher then 10% over the use of usual broad band images. This is extremely promising. However, the overall performance is still insufficient with a maximum of recognition performance around 70%. Wich raise the need to further enhance our systems for better accuracy.

Chapter 6

Conclusion

In this thesis, we have studied the problem of face recognition using multipsectral images (MI) captured in the visible light spectrum. The problem has been divided into two main challenges; first optimizing the set of spectral bands selected for face recognition for each imaging condition. Second, conducting extensive experiments that involves state-of-the-art algorithms to prove the usefulness of using MI images to solve problem of high light illumination variation. This later problem, despite not being the only problem that encounter tasks of face recognition, is by far the most challenging and disrupting for the success of such systems in real world situations.

The first challenge, namely the selection of best spectral bands, is the key to any eventual success an MI based system try to achieve. However, this challenge has been avoided by most of researchers due to multiple raisons including the complexity of building systems to capture hundreds of spectral bands at different wavelengths and the intrinsic complexity of the problem itself where the optimization process should consider an important number of constraints/ parameters.

Our thesis has studied this first challenge and comes out with several algorithms that determine the best set of spectral bands for a given imaging conditions. Our algorithms have been portioned into static algorithms that select the same set of best bands during the training phase, which is used then for tests, and dynamic algorithms that selects different bands for different subjects. In the static category, the problem of best spectral bands selection was formulated as a pursuit problem where weights of importance were affected to each spectral band and the vector of all weights was constrained to be sparse with most of its

elements are zeros. In another work, we have assigned to each spectral band an LDA based weak classifier. Then, all weak classifiers were boosted together using an Adaboost process. From this later, each weak classifier obtained a weight that characterizes its importance and hence the quality of the corresponding spectral band. Finally, an approach based on the decomposition of the Face image cube using multi-linear sparse decomposition was proposed. In this later, the projection of image lighting mode into the learned space determined the set of spectral bands the less sensitive to the lighting conditions. For the dynamic category, an approach has been proposed using transfer learning, mixture of Gaussians and Likelihood ratio to model the relation between the intrinsic quality of each spectral band and its recognition performance. This relation is used then to select the best spectral bands of each new subject.

The different approaches we have proposed have been experimented with three sate-of-he-art algorithms which are Multiblock Local Binary Pattern (MBLBP), Histogram of Gabor Phase Patterns (HGPP) and Local Gabor Binary Pattern Histogram Sequence (LGBPHS). Each of these algorithms was used to extract feature vectors from face images. These feature vectors were then injected into our best bands selection algorithms to determine the set of best bands for each feature extraction algorithm. By applying the studied feature extraction algorithms on selected bands instead of broad band images (gray or RGB images), their performances has considerably increased (up to 14% in some cases) which prove the usefulness of using MI based systems for face matching. The set of experiments we performed were conducted on the $IRIS-M^3$ data base. This database was an obstacle for most of feature extraction algorithms we have used. Our systems for best bands selection have solved this issue and a recognition performances of more then 70% are reachable now. However, the generalzability of our approaches either static or dynamic has to be boosted further by applying them on other databases like the polyU-HSFD database built by the Biometric Research Centre (UGC/CRC) at The Hong Kong Polytechnic University.

At this point, its elligible to remove the veil on our future work to enhance the accuracy of our systems and apply them on real world problems. The first of these planned enhancement will be the use of the kernalization trick for all our BSBS systems. Kernelization has shown a great advantage with our Adaboost based approach and the obtained best bands were much more robust against extreme lighting conditions. Second, we plan to apply our algorithms on moving objects like peoples to see the effect of this movement on the behavior of spectral bands. Last but not least, we plan to upgrade our imaging system buy using different filter that capture images from infrared spectrum. This is done, we can make face recognition in night time conditions. We expect build a first prototype of our imaging system to see if

it can be commercialized and used in airports or stadiums for the world cup 2022 held in Qatar.

References

[1] Ahonen, T., Hadid, A., and Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(12):2037–2041.

[2] Aizerman, M. A., Braverman, E. A., and Rozonoer, L. (1964). Theoretical foundations of the potential function method in pattern recognition learning. In *Automation and Remote Control,*, number 25 in Automation and Remote Control,, pages 821–837.

[3] Akhloufi, M. and Bendada, A. (2010). Multispectral infrared face recognition: a comparative study. In *10th International Conference on Quantitative InfraRed Thermography*, Quebec (Canada).

[4] Arandjelović, O., Hammoud, R., and Cipolla, R. (2010). Thermal and reflectance based personal identification methodology under variable illumination. *Pattern Recogn.*, 43(5):1801–1813.

[5] Baron, R. (1981). Mechanisms of human facial recognition. *Int. J. Man Mach. Stud.*, 15(2):137–178.

[6] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *Comput. Vis. Image Underst.*, 110(3):346–359.

[7] Becker, S., Candes, E., and Grant, M. (2011). Templates for convex cone problems with applications to sparse signal recovery. *Mathematical Programming Computation*, 3(3):165–218.

[8] Belhumeur, P., Hespanha, P., and Kriegman, D. (1997). Eigenfaces vs fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720.

[9] Bhowmik, M. K., Bhattacharjee, D., Nasipuri, M., Basu, D. K., and Kundu, M. (2010). Classification of fused images using radial basis function neural network for human face recognition. *CoRR*, abs/1007.0631.

[10] Bolme, D. S., Beveridge, J. R., Teixeira, M., and Draper, B. A. (2003). The csu face identification evaluation system: its purpose, features, and structure. In *Proceedings of the 3rd international conference on Computer vision systems*, ICVS'03, pages 304–313, Berlin, Heidelberg. Springer-Verlag.

[11] Bouchech, H. and Foufou, S. (2014). A comparative study of best spectral bands selection systems for face recognition. In *IEEE International Conference on Computer Systems and Applications, Qatar*.

[12] Bouchech, H., Foufou, S., and Abidi, M. (2014a). Multilinear sparse decomposition for best spectral bands selection. In *accpetd to the International Conference on Image and Signal Processing ICISP'14*.

[13] Bouchech, H., Foufou, S., Koschan, A., and Abidi, M. (2013). Studies on the effectiveness of multispectral images for face recognition: Comparative studies and new approaches. In *Signal-Image Technology Internet-Based Systems (SITIS), 2013 International Conference on*, pages 58–64.

[14] Bouchech, H., Foufou, S., Koschan, A., and Abidi, M. (2014b). A kernelized sparsity-based approach for best spectral bands selection for face recognition. *Multimedia Tools and Applications*, pages 1–24.

[15] Bouchech, H. J., Foufou, S., and Abidi, M. (2014c). Dynamic best spectral bands selection for face recognition. In *Information Sciences and Systems (CISS), 2014 48th Annual Conference on*, pages 1–6.

[16] Bourlai, T. and Cukic, B. (2012). Multi-spectral face recognition: Identification of people in difficult environments. In *ISI*, pages 196–201.

[17] Bourlai, T., Kalka, N., Ross, A., Cukic, B., and Hornak, L. (2010). Cross-spectral face verification in the short wave infrared (swir) band. In *Proceedings of the 2010 20th International Conference on Pattern Recognition*, ICPR '10, pages 1343–1347, Washington, DC, USA. IEEE Computer Society.

[18] Burge, M. J. and Bowyer, K. W. (2012). *Handbook of Iris Recognition*. springer.

[19] Burt, P. and Adelson, E. (1983). The laplacian pyramid as a compact image code. *IEEE Trans. Commun*, COM-31(4):532–540.

[20] Buyssens, P. and Revenu, M. (2010). Ir and visible face identification via sparse representation. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1 – 6.

[21] Buyssens, P., Revenu, M., and Lepetit, O. (2009). Fusion of ir and visible light modalities for face recognition. In *Proceedings of the 3rd IEEE international conference on Biometrics: Theory, applications and systems*, BTAS'09, pages 252–257, Piscataway, NJ, USA. IEEE Press.

[22] Chang, H. (2008). Multispectral imaging for face recognition over varying illumination. In *Ph.D. dissertation*, Univ. Tennessee, Knoxville.

[23] Chang, H., Koschan, A., and Abidi, B. (2008). Fusing continuous spectral images for face recognition under indoor and outdoor illuminants. *Machine Vision and Application*, 19(4):1432–1769.

[24] Chang, H., Koschan, A., Abidi, B., and Abidi, M. (2010). Fusing continuous spectral images for face recognition under indoor and outdoor illuminants. *Machine vision and applications*, 21:201–215.

[25] Chang, H., Yao, Y., Koschan, A., Abidi, B., and Abidi, M. (2009). Improving face recognition via narrowband spectral range selection using jeffrey divergence. *Trans. Info. For. Sec.*, 4(1):111–122.

[26] Chen, J., Yi, D., Yang, J., Zhao, G., Li, S. Z., and Pietikäinen, M. (2009). Learning mappings for face synthesis from near infrared to visual light images. In *CVPR*, pages 156–163.

[27] Chen, Q., Defrise, M., and Deconinck, F. (1994). Symmetric phase-only matched filtering of fourier-mellin transforms for image registration and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(12):1156–1168.

[28] Chen, W., Er, M. J., and Wu, S. (2006). Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. *Trans. Sys. Man Cyber. Part B*, 36(2):458–466.

[29] Chen, X., Flynn, P. J., and Bowyer, K. W. (2005). Ir and visible light face recognition. *Comput. Vis. Image Underst.*, 99(3):332–358.

[30] Chen, X., Jing, Z., and Xiao, G. (2007a). Nonlinear fusion for face recognition using fuzzy integral. *Communications in Nonlinear Science and Numerical Simulation*, 12(5):823 – 831.

[31] Chen, X., Jing, Z., and Xiao, G. (2007b). Nonlinear fusion for face recognition using fuzzy integral. *Communications in Nonlinear Science and Numerical Simulation*, 12(5):823 – 831.

[32] Colmenarez, A. and Huang, T. (1997). Face detection with information-based maximum discrimination. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition*, CVPR '97, pages 782–, Washington, DC, USA. IEEE Computer Society.

[33] Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*. John Wiley & Sons, Inc, second edition.

[34] Csiszar, I. (1975). I-divergence geometry of probability distributions and minimization problems. *Annals of Probability*, 3(1):146–158.

[35] Dai, W., Yang, Q., Xue, G. R., and Yu, Y. (2007). Boosting for transfer learning. In *24th Int. Conf. Machine Learning*, Corvallis, OR.

[36] Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1 of *CVPR '05*, pages 886–893, Washington, DC, USA. IEEE Computer Society.

[37] Derras, M., Debain, C., Berducat, M., Bonton, P., and Gallice, J. (1994). Unsupervised regions segmentation: Real time control of an upkeep machine of natural spaces. In Eklundh, J.-O., editor, *Computer Vision ECCV '94*, volume 801 of *Lecture Notes in Computer Science*, pages 207–212. Springer Berlin Heidelberg.

[38] Desa, S. M. and Hati, S. (2008). Ir and visible face recognition using fusion of kernel based features. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1 – 4.

[39] Deshmukh, M. and Bhosale, U. (2010). Image fusion and image quality assessment of fused images. *International Journal of Image Processing (IJIP)*, 4:484–508.

[40] Dezert, J. (2002). Foundations for a new theory of plausible and paradoxical reasoning. *Information and Security Journal*, 9:2002.

[41] Di, W., Zhang, L., Zhang, D., and Pan, Q. (2010). Studies on hyperspectral face recognition in visible spectrum with feature band selection. *IEEE Transactions On Systems, Man, And Cybernetics–Part A: Systems And Humans*, 40(6):1354 – 1361.

[42] Donoho, D. (2006). For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. *Comm. Pure and Applied Math*, 59(6):797–829.

[43] Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern Classification*. John Wiley and Sons, second edition.

[44] Ferreira, A. (2007). Survey on boosting algorithms for supervised and semi-supervised learning. Technical report, Instituto Superior de Engenharia de Lisboa.

[45] Figueiredo, M. and Jain, A. K. (2002). Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3):381–396.

[46] Fitzgibbon, A., Fisher, R., and Pilu, M. (1999). Direct least square fitting of ellipses. *IEEE trans. On pattern analysis and machine intelligence*, 21(5):476 – 480.

[47] Fleming, M. and Cottrell, G. (1990). Categorization of faces using unsupervised feature extraction. In *Proc. IEEE International Joint Conference on Neural Networks*, pages 65–70.

[48] Freund, Y. and Schapire, R. E. (1996). Experiments with a new boosting algorithm. In *Thirteenth International Conference on Machine Learning*, pages 148–156, Bari, Italy.

[49] Friedman, J., Hastie, T., and Tibshirani, R. (1998). Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 28:2000.

[50] Gao, Y., Leung, M., Hui, S., and Tananda, M. (2003). Facial expression recognition from line-based caricatures. *IEEE Trans. Syst., Man, Cybern. A, Syst.,Humans*, 33(3):407–412.

[51] Gat, N. (2000). Imaging spectroscopy using tunable filters: A review. *The International Society for Optical Engineering*, pages 50–64.

[52] Georghiades, A., Kriegman, D. J., and Belhumeur, P. (2001). From few to many: Generative models for face recognition under variable pose and illumination. *IEEE Transition on Pattern Analysis and Machine Intelligence*, 23(6):643–660.

[53] Gonzalez, R. C. and Woods, R. E. (2002). *Digital Image Processing*. Upper Saddle River, NJ: Prentice-Hall,, second edition.

[54] GoodRich (2011). Surveillance using swir night vision cameras. online.

[55] Goodrich (2012). Lightweight swir sensor for target detection on board uav equipment. online. Available: http://www.defensefile.com/Customisation/News/Manufacturing and Materials/Special cameras and vision system/Lightweight SWI Sensor For Target Detection On Board UAV Equipment.asp.

[56] Gross, R., Shi, J., and Cohn, J. (2001). Quo vadis face recognition? In *Third Workshop on Empirical Evaluation Methods in Computer Vision*.

[57] Gundimada, S., Asari, V. K., and Gudur, N. (2010). Face recognition in multi-sensor images based on a novel modular feature selection technique. *Inf. Fusion*, 11(2):124–132.

[58] H. Bouchech, S. F. and M. Abidi (Qatar, December 14-17, . (2014). Strengthening surf descriptor with discriminant image filter learning: application to face recognition. In *26 International Conference on Microelectronics, Qatar*.

[59] Haralik, R. M., li, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE trans. on Systems, Man and Cybernetics*, SMC-3:610–621.

[60] Hardeberg, J. Y., Schmitt, F., and Brettel, H. (2002). Multispectral image capture using a liquid crystal tunable filters. *Optical Engineering*, 41(10):2532–2548.

[61] Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Predic-tion*. Springer-Verlag.

[62] Hermosilla, G., Ruiz-del Solar, J., Verschae, R., and Correa, M. (2012). A comparative study of thermal face recognition methods in unconstrained environments. *Pattern Recogn.*, 45(7):2445–2459.

[63] Hill, D., Studholme, C., and Hawkes, D. (1994). Voxel similarity measures for automated image registration. In *Proceedings of the Third SPIE Conference on Visualization in Biomedical Computing*, pages 205–216.

[64] Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(8):832–844.

[65] Ho, T. K., Hull, J. J., and Srihari, S. N. (1994). Decision combination in multiple classifier systems.

[66] Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 8:321–377.

[67] Hsu, C. and Lin, C. (2002). A comparison of methods for multiclass support vector machines. *Trans. Neur. Netw.*, 13(2):415–425.

[68] Huang, X., Lei, Z., Fan, M., Wang, X., and Li, S. Z. (2013). Regularized discriminative spectral regression method for heterogeneous face matching. *IEEE Transactions on Image Processing*, 22(1):353–362.

[69] Imai, F. H., Rosen, M. R., and Berns, R. S. (2001). Multi-spectral imaging of van gogh self portrait at the national gallery of art. In *ISTś 2001 PICS*, pages 185–189.

[70] Jain, A., Nandakumar, K., and Ross, A. (2005a). Score normalization in multimodal biometric systems. *Pattern Recogn.*, 38(12):2270–2285.

[71] Jain, A. K., Bolle, R., and Pankanti, S. (1999). *Biometrics: Personal Identification in Networked Society*. springer.

[72] Jain, A. K., Nandakumar, K., and Ross, A. (2005b). Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285.

[73] Jain, A. K., Patrick., F., and Arun, R. (2008). *Handbook of Biometrics*. springer.

[74] Jin, Y., Fayad, L., and Laine, A. (2001). Contrast enhancement by multi-scale adaptive histogram equalization. In *Wavelets: Applications in Signal and Image Processing IX, Andrew F. Laine; Michael A. Unser; Akram Aldroubi; Eds.*, volume 4478, pages 206–213.

[75] Kalka, N. D., Bourlai, T., Cukic, B., and Hornak, L. (2011). Cross-spectral face recognition in heterogeneous environments: A case study on matching visible to short-wave infrared imagery. In *Proceedings of the 2011 International Joint Conference on Biometrics*, IJCB '11, pages 1–8, Washington, DC, USA. IEEE Computer Society.

[76] Keller, J. M. and Osborn, J. (1996). Training the fuzzy integral. *International Journal of Approximate Reasoning*, 15(1):1 – 24.

[77] Khan, R. A., Meyer, A., Konik, H., and Bouakaz, S. (2013). Framework for reliable, real-time facial expression recognition for low resolution images. *Pattern Recognition Letters*, 34(10):1159 – 1168.

[78] Kirschner, J. (2011). Swir for target detection, recognition, and identification. Online. Available: http://www.photonicsonline.com/doc.mvc/ SWIR-For-Target-Detection-Recognition-And-0002, accessed Jul. 5, 2012.

[79] Klare, B. and Jain, A. K. (2010). Heterogeneous face recognition: Matching nir to visible light images. In *ICPR*, pages 1513–1516.

[80] Klare, B. and Jain, A. K. (2012). Heterogeneous face recognition using kernel prototype similarities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99):1.

[81] Kong, S., Heo, J., Boughorbel, F., Zheng, Y., Abidi, B., Koschan, A., Yi, M., and Abidi, M. (Jan. 2007). Adaptive fusion of visual and thermal ir images for illumination-invariant face recognition. *International Journal of Computer Vision*, 71(2):215–233.

[82] Koschan, A. and Abidi, M. (2008). *Digital Color Image processing*. Wiley, New York.

[83] Kouassi, R., Gouton, P., and Paindavoine, M. (2001). Approximation of the karhunen loeve transformation and its application to colour images. *Signal Processing: Image Communication*, 16(6):541 – 551.

[84] Landre, J., Morain-Nicolier, F., and Ruan, S. (2009). Ornamental letters image classification using local dissimilarity maps. In *Document Analysis and Recognition, 2009. ICDAR 09. 10th International Conference on*, pages 186–190.

[85] Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer.

[86] Lei, Z. and Li, S. Z. (2009). Coupled spectral regression for matching heterogeneous faces. In *CVPR'2009*, pages 1123 – 1128.

[87] Lei, Z., Liao, S., Jain, A. K., and Li, S. Z. (2012a). Coupled discriminant analysis for heterogeneous face recognition. *IEEE Transactions on Information Forensics and Security*, 7(6):1707–1716.

[88] Lei, Z., Zhou, C., Yi, D., Jain, A. K., and Li, S. Z. (2012b). An improved coupled spectral regression for heterogeneous face recognition. In *ICB*, pages 7–12.

[89] Li, S. Z. (2012). Discriminant image filter learning for face recognition with local binary pattern like representation. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, CVPR '12, pages 2512–2517, Washington, DC, USA. IEEE Computer Society.

[90] Li, S. Z., Chu, R., Liao, S., and Zhang, L. (2007). Illumination invariant face recognition using near-infrared images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(4):627–639.

[91] Li, S. Z. and Jain, A. K. (2011). *Handbook of Face Recognition*. springer, second edition.

[92] Li, S. Z., Lei, Z., and Ao, M. (2009). The hfb face database for heterogeneous face biometrics research. In *6th IEEE Workshop on Object Tracking and Classification Beyond and in the Visible Spectrum (OTCBVS, in conjunction with CVPR 2009)*, pages 1005–1010.

[93] Liao, S., Yi, D., Lei, Z., Qin, R., and Li, S. Z. (2009a). Heterogeneous face recognition from local structures of normalized appearance. In *Proceedings of the Third International Conference on Advances in Biometrics*, ICB '09, pages 209–218, Berlin, Heidelberg. Springer-Verlag.

[94] Liao, S., Zhu, X., Lei, Z., Zhang, L., and Li, S. Z. (2009b). Learning multi-scale block local binary patterns for face recognition. In *Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, Springer, Heidelberg (2007)*, volume 4642, pages 828–837.

[95] Lin, D. and Tang, X. (2006). Inter-modality face recognition. In *A. Leonardis, H. Bischof, and A. Pinz (Eds.): ECCV 2006, Part IV, LNCS 3954, Springer-Verlag Berlin Heidelberg 2006*, pages 13–26.

[96] Liu, M., Xie, W., Chen, X., Guo, Y. M. Y., Meng, J., Yuan, Z., and Qin, Q. (2011). Heterogeneous face biometrics based on guassian weights and invariant features synthesis. In *Computing, Control and Industrial Engineering (CCIE), 2011 IEEE 2nd International Conference on*, pages 374 – 377.

[97] Liu, M., Yuan, Z., Ma, Y., Chen, X., and Yin, Q. (2012a). Heterogeneous face recognition and synthesis using canonical correlation analysis. *Journal of Convergence Information Technology(JCIT)*, 7(8).

[98] Liu, Q., Tang, X., Jin, H., Lu, H., and Ma, S. (2005). A nonlinear approach for face sketch synthesis and recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1005–1010.

[99] Liu, S., Yi, D., Lei, Z., and Li, S. Z. (2012b). Heterogeneous face image matching using multi-scale features. In *ICB*, pages 79–84.

[100] Ma, J., Zhao, Y., and Ahalt, S. (2002). Osu svm classifier matlab toolbox (ver 3.00).

[101] Ma, Y., Xie, W., Chen, X., Liu, M., Guo, Y., Meng, J., Yuan, Z., and Qin, Q. (2011). An analysis-by-synthesis method based on sparse representation for heterogeneous face biometrics. In *Multimedia Technology (ICMT), 2011 International Conference on*, pages 148 – 151.

[102] Maltoni, D., Maio, D., Jain, A. K., and Prabhakar, S. (2009). *Handbook of Fingerprint Recognition*. springer, second edition.

[103] Malviya, A. and Bhirud, S. G. (2009). Image fusion of digital images. *International Journal of Recent Trends in Engineering*, 2(3):146–148.

[104] Mangai, U., Samanta, S., Das, S., and Chowdhury, P. R. (2010). A survey of decision fusion and feature fusion strategies for pattern classification. *IETE Tech Rev*, 27:293–307.

[105] Mannor, S., Peleg, D., and Rubinstein, R. Y. (2005). The cross entropy method for classification. In *ICML*, pages 561–568.

[106] Mavadati, S. M., Sadeghi, M. T., and Kittler, J. (2010). Fusion of visible and synthesised near infrared information for face authentication. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 3801 – 3804.

[107] Miller, B. (1994). Vital signs of identity. *IEEE Spectrum*, 31(2):22–30.

[108] Moon, S., Kong, S. G., Yoo, J. H., and Chung, K. (2006). Face recognition with multiscale data fusion of visible and thermal images. In *IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety*, pages 24 – 27, Alexandria, VA, USA.

[109] Nandakumar, K., Chen, Y., Dass, S. C., and Jain, A. K. (2008). Likelihood ratio based biometric score fusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):342 – 347.

[110] Nandakumar, K., Chen, Y., Jain, A. K., and Dass, S. (August 2006). Qualitybased score level fusion in multibiometric systems. In *International Conference on Pattern Recognition ICPR*, pages 473–476, Hong Kong, China.

[111] Neagoe, V. and Ropot, A.-D. (2002). Concurrent self-organizing maps for pattern classification. In *Proceedings of the 1st IEEE International Conference on Cognitive Informatics*, ICCI '02, pages 304–312, Washington, DC, USA. IEEE Computer Society.

[112] Neagoe, V. E., Ropot, A. D., and Mugioiu, A. C. (2007). Real time face recognition using decision fusion of neural classifiers in the visible and thermal infrared spectrum. In *Proceedings of the 2007 IEEE Conference on Advanced Video and Signal Based Surveillance*, AVSS '07, pages 301–306, Washington, DC, USA.

[113] Nicolo, F. and Schmid, N. A. (2011). A method for robust multispectral face recognition. In *Proceedings of the 8th international conference on Image analysis and recognition - Volume Part II*, ICIAR'11, pages 180–190, Berlin, Heidelberg. Springer-Verlag.

[114] Nicolo, F. and Schmid, N. A. (2012). Long range cross-spectral face recognition: Matching swir against visible light images. *IEEE Transactions on Information Forensics and Security*, 7(6):1717–1726.

[115] Pan, Z., Healey, G., and Prasad, M. (2003a). Face recognition in hyperspectral images. *IEEE Transition on Pattern Analysis and Machine Intelligence*, 25(12):1552–1560.

[116] Pan, Z., Healey, G. E., Prasad, M., and Tromberg, B. J. (2003b). Face recognition in hyperspectral images. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:334.

[117] Panda, R. and Naik, M. K. (2012). Fusion of infrared and visual images using bacterial foraging strategy. *WSEAS transaction on Signal processing*, 8(4).

[118] Passino, K. (2002). Biomimicry of bacterial foraging for distributed optimization and control. *IEEE Control Sys. Mag*, pages 52–67.

[119] Phillips, P. J. (2002). Human identification technical challenges. In *IEEE Int. Conference on Image Processing*, volume 1, pages I–49–I52.

[120] Phillips, P. J. (2003). Progress in human id. In *Proc. IEEE Conf. on Advanced Video and Signal Based Surveillance*, pages 1–.

[121] Pietikinen, M., Hadid, A., Zhao, G., and Ahonen, T. (2011). *Computer Vision Using Local Binary Patterns*. Springer Publishing Company, Incorporated, 1st edition.

[122] Pop, F. M., Gordan, M., Florea, C., and Vlaicu, A. (2010). Fusion based approach for thermal and visible face recognition under pose and expresivity variation. In *Roedunet International Conference (RoEduNet), 2010 9th*, pages 61 – 66.

[123] Qiu, Q. and Chellappa, R. (2013). Compositional dictionaries for domain adaptive face recognition. *CoRR*, abs/1308.0271.

[124] Raghavendra, R., Dorizzi, B., Rao, A., and Hemantha Kumar, G. (2011). Particle swarm optimization based fusion of near infrared and visible images for improved face verification. *Pattern Recogn.*, 44(2):401–411.

[125] Robila, S. A. (2009). Quo vadis face recognition: Spectral considerations. In *LISAT '09*, pages 1 – 5.

[126] Ross, A. A., Nandakumar, K., and Jain, A. K. (2006). *Handbook of Multibiometrics (International Series on Biometrics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

[127] Ruichek, Y. (2005). Multilevel- and neural-network-based stereo-matching method for real-time obstacle detection using linear cameras. *Intelligent Transportation Systems, IEEE Transactions on*, 6(1):54–62.

[128] Saunders, D. and Cuppitt, J. (1993). Image processing ate the national gallery:the vasari project. Technical Report 14:72, National Gallery.

[129] Selinger, A. and Socolinsky, D. (2002). Appearance-based facial recognition using visible and thermal imagery: a comparative study. Technical Report 02-01, Equinox Corporation.

[130] Shahbe, M. D. and Hati, S. (2008). Decision fusion based on voting scheme for ir and visible face recognition. *MG&V*, 17(4):403–424.

[131] Shan, S., Gao, W., Cao, B., and Zhao, D. (2003). Illumination normalization for robust face recognition against varying lighting conditions. In *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, AMFG '03, pages 157–, Washington, DC, USA. IEEE Computer Society.

[132] Shao, T. and Wang, Y. (2008). The face images fusion based on laplacian pyramid and lbp operator. In *Signal Processing, 9th International Conference on*, pages 1165 – 1169.

[133] Shlizerman, I. K. and Basri, R. (2011). 3d face reconstruction from a single image using a single reference face shape. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 33(2):394 – 405.

[134] Sims, R. and Phillips, M. (1997). Target signature consistency of image data fusion alternatives. *Opt. Eng*, 36(3):743–754.

[135] Singh, R., Vatsa, M., and Noore., A. (2008a). Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition. *Pattern Recognition*, 41:880–893.

[136] Singh, R., Vatsa, M., and Noore, A. (2008b). Multiclass mi-granular soft support vector machine: A case study in dynamic classifier selection for multispectral face recognition. *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1 – 4.

[137] Singh, R., Vatsa, M., and Noore, A. (2009). Face recognition with disguise and single gallery images. *Image Vision Comput.*, 27(3):245–257.

[138] Singh, R., Vatsa, M., Noore, A., and Singh, S. K. (2006). Dempster-shafer theory based classifier fusion for improved fingerprint verification performance. In *Proceedings of the 5th Indian conference on Computer Vision, Graphics and Image Processing*, ICVGIP'06, pages 941–949, Berlin, Heidelberg. Springer-Verlag.

[139] Sivanandam, S. N. and Deepa, S. N. (2010). *Introduction to Genetic Algorithms*. Springer Publishing Company, Incorporated, 1st edition.

[140] Smarandache, F. and Dezert, J. (2006). Advances and applications of dsmt.

[141] Sung, K. and Poggio, T. (1998). Example-based learning for view-based human face detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(1):39–51.

[142] Tan, X. and Triggs, B. (2007). Enhanced local texture feature sets for face recognition under difficult lighting conditions. pages 168–182.

[143] Tao, Q., Wu, G., Wang, F., and Wang, J. (2005). Posterior probability support vector machines for unbalanced data. *IEEE Transaction on Neural Network*, 16(6):1561–1573.

[144] Teixeira, M. (2003). The bayesian intrapersonal/extrapersonal classifier. Master's thesis, Colorado State University, Fort Collins, Colorado 80523 U.S.A.

[145] Ting, L. D., Dan, Z. X., and wen, W. C. (2008). Wavelet-based multispectral face recognition. *OPTOELECTRONICS LETTERS*, 4(5):451–462.

[146] Toh, K., Kim, Y., Lee, S., and Kim, J. (2008). Fusion of visual and infra-red face scores by weighted power series. *Pattern Recogn. Lett.*, 29(5):603–615.

[147] Tolba, A. S., El-Baz, A. H., and El-Harby, A. A. (2006). Face recognition: A literature review. *International Journal of Information and Communication Engineering*, 2(2):88–104.

[148] Tominaga, S. (1996). A multi-channel vision system for estimating surface and illuminant functions. *J.optical society of America A*, 13(11):2163–2173.

[149] Tong, Y., Wang, Y., Zhu, Z., and Ji, Q. (2007). Robust facial feature tracking under varying face pose and facial expression. *Pattern Recognition journal*, 40(11):3195–3208.

[150] Toyama, K., Feris, R., Gemmell, J., and Kruger, V. (2002). Hierarchical wavelet networks for facial feature localization. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pages 118–123.

[151] Turk, M. and Pentland, A. (1991). Face recognition using eigenfaces,. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–591.

[152] Vatsa, M., Singh, R., and Noore, A. (2008). Quality-augmented fusion of level-2 and level-3 fingerprint information using dsm theory. *International Journal of Approximate Reasoning*.

[153] Viola, P. A. and Jones, M. J. (2001). Rapid object detection using a boosted cascade of simple features. In *CVPR (1)*, pages 511–518.

[154] Wang, R., Yang, J., Yi, D., and Li, S. Z. (2009). An analysis-by-synthesis method for heterogeneous face biometrics. In *Proceedings of the Third International Conference on Advances in Biometrics*, ICB '09, pages 319–326, Berlin, Heidelberg. Springer-Verlag.

[155] Wright, J., Yang, A., Ganesh, A., Sastry, S. S., and Ma, Y. (2009). Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(2):210 – 227.

[156] WVHTCF (2011). Tactical imager for night/day extended-range surveillance. online. Available: http://www.wvhtf. org/departments/advanced tech/projects/tinders.asp.

[157] Xie, X. and Lam, K. (2005). Face recognition under varying illumination based on a 2d face shape model. *Pattern Recognition*, 38(2):221–230.

[158] Yager, R. R. and Liu, L. (2010). *Classic Works of the Dempster-Shafer Theory of Belief Functions*. Springer Publishing Company, Incorporated, 1st edition.

[159] Yan, S., Xu, D., Zhang, B., Zhang, H., Yang, Q., and Lin, S. (2007). Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(1):40–51.

[160] Yao, Y., Abidi, B., Kalka, N. D., Schmid, N., and Abidi, M. (2006). High magnification and long distance face recognition: Database acquisition, evaluation, and enhancement. In *Biometric Consortium Conference, 2006 Biometrics Symposium: Special Session on Research at the*, pages 1 – 6.

[161] Yao, Y., Abidi, B. R., Kalka, N. D., Schmid, N. A., and Abidi, M. A. (2008). Improving long range and high magnification face recognition: Database acquisition, evaluation, and enhancement. *Computer Vision and Image Understanding*, 111(2):111–125.

[162] Yi, D., Liao, S., Lei, Z., Sang, J., and Li, S. Z. (2009). Partial face matching between near infrared and visual images in mbgc portal challenge. In *Proceedings of the Third International Conference on Advances in Biometrics*, ICB '09, pages 733–742, Berlin, Heidelberg. Springer-Verlag.

[163] Yi, D., Liu, R., Chu, R., Lei, Z., and Li, S. Z. (2007). Face matching between near infrared and visible light images. In *Proceedings of the 2007 international conference on Advances in Biometrics*, ICB'07, pages 523–530, Berlin, Heidelberg. Springer-Verlag.

[164] Zahran, E. G., Abbas, A. M., Dessouky, M. I., Ashour, M. A., and Sharshar., K. A. (2009). High performance face recognition using pca and zm on fused lwir and visible images on the wavelet domain. In *Computer Engineering & Systems. ICCES 2009. International Conference on*, pages 449 – 454.

[165] Zhang, L., Razdan, A., Farin, G., Femiani, J., Bae, M., and Lockwood, C. (2006). 3d face authentication and recognition based on bilateral symmetry analysis. *Visual Comput*, 22:43–55.

[166] Zhang, Z., Wang, Y., and Zhang, Z. (2011). Face synthesis from near-infrared to visual light via sparse representation. In *Proceedings of the 2011 International Joint Conference on Biometrics*, IJCB '11, pages 1–6.

[167] Zhang, Z., Yi, D., Lei, Z., and Li, S. Z. (2012). Regularized transfer boosting for face detection across spectrum. *IEEE SIGNAL PROCESSING LETTERS*, 19(3):131 – 134.

[168] Zhao, W., Chellapa, R., and Phillips, P. J. (2003). Face recognition: A literature survey. *ACM Computer Surv.*, 35(4):399–458.

[169] Zheng, Y. (2011). An exploration of color fusion with multispectral images for night vision enhancement. Available from: http://www.intechopen.com/books/image-fusion-and-its-applications/an-exploration-of-color-fusion-with-multispectral-images-for-night-vision-enhancement.

[170] Zheng, Y., Zhang, C., and Zhou, Z. (2012). A wavelet-based method for multispectral face recognition. In *SPIE Proceedings, Independent Component Analyses, Compressive Sampling, Wavelets, Neural Net, Biosystems, and Nanoengineering X, Harold Szu, Editors*, volume 8401.

[171] Zhou, D., Bousquet, O., Lal, T. N., Weston, J., and Scholkopf, B. (2004). Learning with local and global consistency. In *Advances in Neural Information Processing Systems 16*, pages 321–328. MIT Press.

[172] Zhu, J., Zou, H., Rosset, S., and Hastie, T. (2009). Multi-class adaboost. *Statistics and Its Interface*, 2:349–360.