



UNIVERSITY OF BURGUNDY
U.F.R. SCIENCES ET TECHNIQUE
ECOLE DOCTORAL SPIM

THESIS

Presented by:

AHMED BEN SAID

*Submitted in fulfilment of the requirements
for the degree of:*

DOCTOR OF THE UNIVERSITY OF BURGUNDY

**MULTISPECTRAL IMAGING AND ITS USE FOR FACE RECOGNITION:
SENSORY DATA ENHANCEMENT**

**IMAGERIE MULTISPECTRALE ET SON USAGE POUR LA
RECONNAISSANCE DE VISAGE: AMELIORATION DES DONNEES
SENSORIELLES**

Defended on 03-06-2015

Jury :

M. Kacem CHEHDI, Professeur, Université Rennes 1, Rapporteur

Mme. Sylvie TREUILLET, MCF HDR, Université Orléans, Rapporteur

M. Yann GAVET, MCF HDR, Ecole des Mines de Saint Etienne, Examinateur

M. Olivier LALIGANT, Professeur, Université de Bourgogne, Examinateur

M. Sebti FOUFOU, Professeur, Université de Bourgogne, Directeur de thèse.

“when you have eliminated the impossible, whatever remains, however improbable, must be the truth. ”

Sir Arthur Conan Doyle

Abstract

In this thesis, we focus on multispectral image for face recognition. With such application, the quality of the image is an important factor that affects the accuracy of the recognition. However, the sensory data are in general corrupted by noise. Thus, we propose several denoising algorithms that are able to ensure a good tradeoff between noise removal and details preservation. Furthermore, characterizing regions and details of the face can improve recognition. We focus also in this thesis on multispectral image segmentation particularly clustering techniques and cluster analysis. The effectiveness of the proposed algorithms is illustrated by comparing them with state-of-the-art methods using both simulated and real multispectral data sets.

Keywords: multispectral image, sensory data enhancement, segmentation, cluster analysis, denoising

Acknowledgements

First of all, I would like to thank my supervisor Dr. Sebti Foufou for his help, support, guidance, patience, kindness and friendship. He is not just my supervisor but also my mentor.

I would like also to thank Dr Rachid Hadjidj for his important contributions, comments and ideas and invaluable helpfulness. I would like also to thank Dr Mongi Abidi for the outstanding and fruitful collaboration.

M. Kacem Chehdi and Mme Sylvie Treuillet have made me honored by reviewing my thesis. Their remarks and instructions have enormously contributed in enhancing my work. I would like also to thank M. Olivier Laligant and Yann Gavet for being members of my Jury.

All my thanks also go to my teammates: Jamel Hamdi Bouchech and Faten Omri. Being a member of a research team with them for three years has been a wonderful experience. Special thanks go to my friends Mahdi, Seifeddine, Bassem, Rami, Samir and all of the others I do not mention here who definitely deserve my sincere acknowledgments.

Special words for my sweetheart Emna for her love, kindness and patience, she has been and always be my support. Thousands of words won't be enough to express my gratefulness for her.

I would like also to thank my parents and my sister for their unlimited support and care. You have been such a wonderful family.

Résumé

La recherche en biométrie a connu une grande évolution durant les dernières années surtout avec le développement des méthodes de décomposition de visage. Cependant, ces méthodes ne sont pas robustes particulièrement dans les environnements incontrôlés. Pour faire face à ce problème, l'imagerie multispectrale s'est présentée comme une nouvelle technologie qui peut être utilisée en biométrie basée sur la reconnaissance de visage. Dans tous ce processus, la qualité des images est un facteur majeur pour concevoir un système de reconnaissance fiable. Il est essentiel de se disposer d'images de haute qualité. Ainsi, il est indispensable de développer des algorithmes et des méthodes pour l'amélioration des données sensorielles. Cette amélioration inclut plusieurs tâches telles que la déconvolution des images, le defloutage, la segmentation, le débruitage... Dans le cadre de cette thèse, nous étudions particulièrement la suppression de bruit ainsi que la segmentation de visage.

En général, le bruit est inévitable dans toutes applications et son élimination doit se faire tout en assurant l'intégrité de l'information confinée dans l'image. Cette exigence est essentielle dans la conception d'un algorithme de débruitage. Le filtre Gaussien anisotropique est conçu spécifiquement pour répondre à cette caractéristique. Nous proposons d'étendre ce filtre au cas vectoriel où les données en disposition ne sont plus des valeurs de pixels mais un ensemble de vecteurs dont les attribues sont la réflectance dans une longueur d'onde spécifique. En outre, nous étendons aussi le filtre de la moyenne non-local (NLM) dans le cas vectoriel. La particularité de ce genre de filtre est la robustesse face au bruit Gaussien.

La deuxième tâche dans le but d'amélioration de données sensorielles est la segmentation. Le clustering est l'une des techniques souvent utilisées pour la segmentation et classification des images. L'analyse du clustering implique le développement de nouveaux algorithmes particulièrement ceux qui sont basés sur la méthode partitionnelle. Avec cette approche, le nombre de clusters doit être connu d'avance, chose qui n'est pas toujours vraie surtout si nous disposons de données ayant des caractéristiques inconnues. Dans le cadre de cette thèse, nous proposons de nouveaux indices de validation de clusters qui sont capables de prévoir le vrai nombre de clusters même dans le cas de données complexes.

A travers ces deux tâches, des expériences sur des images couleurs et multispectrales sont réalisées. Nous avons utilisé des bases de données d'image très connues pour analyser l'approche proposée.

Mot clés: image multispectrale, amélioration des données sensorielles, segmentation, analyse de clustering débruitage

Contents

Abstract	iii
Acknowledgements	iv
Résumé	v
Contents	vi
List of Figures	ix
List of Tables	xi
Abbreviations	xii
1 Introduction	1
2 Literature review	6
2.1 Introduction	6
2.2 Multispectral image and its applications	6
2.3 Overview of multispectral image denoising	8
2.4 Overview of multispectral image segmentation	12
2.4.1 Clustering for multispectral image segmentation	13
2.4.2 Other techniques for multispectral image segmentation	15
2.4.3 Cluster validity index	17
2.4.3.1 CVI based on membership values	17
2.4.3.2 CVI based on membership and data	19
2.5 Conclusion	20
3 Multispectral image denoising	21
3.1 Introduction	21
3.2 Vector anisotropic Gaussian (VAG) filter for multispectral image denoising	22

3.2.1	Anisotropic Gaussian filter	22
3.2.2	Noise model	23
3.2.3	Vector anisotropic filter and noise estimation	25
3.2.3.1	Vector anisotropic filter	25
3.2.3.2	Sparse matrix transform	25
3.3	Optimized vector Non-Local Mean filter for multispectral image denoising: OVNLM	26
3.3.1	Non-Local Mean filter	26
3.3.2	Vector NLM filter	27
3.3.3	Optimization framework for vector NLM	28
3.3.4	Relevant pixel selection	30
3.3.5	VOLNM algorithm	31
3.4	Conclusion	32
4	Multispectral image segmentation	33
4.1	Introduction	33
4.2	Gravitational Fuzzy C-Means	34
4.2.1	Gravitation theory and interaction between particles	34
4.2.2	Attraction force	34
4.2.3	Escape velocity	35
4.3	Gravitational Weighted Fuzzy C-Means (GWFCM) for multispectral image segmentation	36
4.4	Cluster validity index	37
4.4.1	Cluster validity index based on n-sphere	37
4.4.1.1	n-sphere	37
4.4.1.2	Compactness measure	38
4.4.1.3	Overlap measure	39
4.4.1.4	validity index S	40
4.4.2	Cluster validity index based on Jeffrey divergence	40
4.4.2.1	Motivation	41
4.4.2.2	Cluster validity index based on Jeffrey divergence	41
4.5	Conclusion	46
5	Experimental results	47
5.1	Introduction	47
5.2	Data sets	47
5.2.1	IRIS M^3 multispectral images database	48
5.2.2	HYDICE multispectral images database	48
5.2.3	Machine learning databases	48
5.3	Multispectral image denoising: experimental results	50
5.3.1	Vector anisotropic Gaussian filter: Experimental results	51
5.3.2	Optimized Vector Non-Local Mean filter: Experimental results	54
5.4	Gravitational Weighted Fuzzy C-Means: Experimental results	61
5.4.1	Clusestring	62
5.4.2	Multispectral image segmentation	63
5.5	Cluster validity indexes: Experimental results	66
5.5.1	CVI evaluation methodologies	66

5.5.2 Cluster validity index based on n-sphere	66
5.6 Impact of multispectral image denoising on segmentation	82
5.7 Conclusion	83
6 Conclusions and Future work	85
List of Publications	87
Bibliography	88

List of Figures

2.1	Processing pipeline of face recognition	8
3.1	Anisotropic Gaussian with orientation θ	24
3.2	Normalized histogram of the variance of several famous images	24
4.1	Illustration of Newton's law of attraction forces between two objects	35
4.2	Three clusters generated from Gaussian distributions. Cluster B and C are equally distant to cluster A but cluster C overlaps more with cluster A	41
5.1	Sample of four subjects from IRIS M^3 database: left: 530nm, right: 650nm, top rows: Daylight, bottom rows: Halogen light)	49
5.2	Sample image from HYDICE: (a) 750nm, (b) 761nm (c) 772nm, (d) 859nm	50
5.3	(a) Original Baboon image (b) Corrupted image ($\sigma_{noise} = 20$) (c) 2D anisotropic filter (d) SURE-LET (e) Proposed VAG method	52
5.4	(a) Hydice original image (b) Corrupted image ($\sigma_{noise} = 30$) (c) 2D anisotropic filter (d) SURE-LET (e) Proposed VAG method	52
5.5	(a) IRIS original image (520nm) (b) Corrupted image ($\sigma_{noise} = 30$) (c) 2D anisotropic filter (d) SURE-LET (e) Proposed method	53
5.6	Normalized histogram of the variance of Hydice and IRIS images	54
5.7	Lena image and denoising results (a) Original (b) Corrupted (c) ONLM (d) SURE-LET (e) OVNLM	55
5.8	PSNR variation according to noise standard deviation for Lena image . .	55
5.9	Baboon and denoising results (a) Original (b) Corrupted (c) ONLM (d) SURE-LET (e) OVNLM	56
5.10	PSNR variation according to noise standard deviation for Baboon image .	56
5.11	Multispectral images for subject 1 in 480nm, 560nm and 720nm	57
5.12	Pixel variation along a profile for subject 1	57
5.13	Multispectral images for subject 2 in 480nm, 560nm and 720nm	57
5.14	Pixel variation along a profile for subject 2	58
5.15	Subject 1: (a) Noisy image (b) SURE-LET (c) ONLM (d) Proposed . .	58
5.16	PSNR curves for subject 1	59
5.17	Subject 2: (a) Noisy image (b) SURE-LET (c) ONLM (d) Proposed . .	59
5.18	PSNR curves for subject 2	60
5.19	Variation of PSNR with relative to h	61
5.20	Thyroid - Rand index	63
5.21	Imbalanced Glass - Rand index	63
5.22	Multispectral image organization	64
5.23	Multispectral face image: subject 1	64

5.24 Multispectral face image: subject 2	64
5.25 Segmentation result: subject 1	65
5.26 Segmentation result: subject 2	65
5.27 Data set S1 with 15 Gaussian clusters	73
5.28 CVIs for S1 with different number of clusters (lines are scaled for display purpose)	74
5.29 Data set S2 with 15 Gaussian clusters	74
5.30 CVIs for S2 with different number of clusters (lines are scaled for display purpose)	75
5.31 Data set S3 with 15 Gaussian clusters	75
5.32 CVIs for S3 with different number of clusters (lines are scaled for display purpose)	76
5.33 Data set S4 with 15 Gaussian clusters	76
5.34 CVIs for S4 with different number of clusters (lines are scaled for display purpose)	77
5.35 Data set R15 with 15 Gaussian clusters	77
5.36 CVIs for R15 with different number of clusters (lines are scaled for display purpose)	78
5.37 CVIs for Balance data set with different number of clusters (lines are scaled for display purpose)	78
5.38 CVIs for Banana data set with different number of clusters (lines are scaled for display purpose)	79
5.39 CVIs for Iris data set with different number of clusters (lines are scaled for display purpose)	79
5.40 CVIs for Yale_5 data set with different number of clusters (lines are scaled for display purpose)	80
5.41 CVIs for Yale_7 with different number of clusters (lines are scaled for display purpose)	80

List of Tables

5.1	Ethnicity percentage in IRIS M^3 database	48
5.2	Comparison of different denoising methods for different noise level of the Baboon image (best result is bold)	52
5.3	Comparison of different denoising methods for different noise level of Hydice image (best result is bold)	53
5.4	Comparison of different denoising methods for different noise level of IRIS image (best result is bold)	53
5.5	MSE between original signal and restored signals	60
5.6	CVIs values for subject 1	65
5.7	CVIs values for subject 2	66
5.8	Values of parameter T for each data set	67
5.9	CVIs values for Thyroid data set: best result is bold	67
5.10	CVIs values for Wine data set: best result is bold	68
5.11	CVIs values for Glass data set: best result is bold	68
5.12	CVIs values for Iris data set: best result is bold	68
5.13	CVIs values for Yeast data set: best result is bold	69
5.14	CVIs values for Yale_5 data set: best result is bold	69
5.15	CVIs values for Yale_7 data set: best result is bold	69
5.16	Number of times each index proposes the optimal partition highlighted by the similarity measures as the best one for Thyroid	70
5.17	Number of times each index proposes the optimal partition highlighted by the similarity measures as the best one for Iris	70
5.18	Number of times each index proposes the optimal partition highlighted by the similarity measures as the best one for 1R2RCT_A	70
5.19	Number of times each index proposes the optimal partition highlighted by the similarity measures as the best one for 1R2RT_B	70
5.20	Number of times each index proposes the optimal partition highlighted by the similarity measures as the best one for car5	70
5.21	Similarity measures for balance data set	81
5.22	Similarity measures for banana data set	81
5.23	Similarity measures for Iris data set	81
5.24	Similarity measures for 1R2RCT_A data set	81
5.25	Similarity measures for 1R2RT_B data set	81
5.26	Similarity measures for car5 data set	82
5.27	CVIs values of the noisy segmented images for subject 1	82
5.28	CVIs values of the denoised segmented images for subject 1	83
5.29	CVIs values of the noisy segmented images for subject 2	83
5.30	CVIs values of the denoised segmented images for subject 2	83

Abbreviations

AHC	Agglomerative Hierarchical Clustering
AP	Affinity Propagation
ARI	Adjusted Rand Index
CVI	Cluster Validity Index
EM	Expectation and Maximization
FCM	Fuzzy C-Means
FM	Fox-Mallow
FS	Fukuyama and Sugeno
GWFCMNR	Gravitational Weighted Fuzzy C-Means
HYDICE	HYperspectral Digital Imagery Collection Experiment
Jacc	Jaccard
ML	Maximum Likelihood
MPC	Modified Partitional Clustering
MSE	Mean Square Error
NLM	Non-Local Mean
OVNLM	Optimized Vector Non-Local Mean
PC	Partitional Clustering
PCA	Principle Component Analysis
PDE	Partial Differential Equation
PDF	Probability Density Estimation
PE	Partitional Entropy
PSNR	Peak Signal to Noise Ratio
QP	Quadratic Programming
SER	Signal to ERror
SMT	Sparse Matrix Transform

SNR	Signal to Noise Ratio
SQP	Sequential Quadratic Programming
SURE	Stein's Unbiased Risk Estimator
TV	Total Variation
VAG	Vector Anisotropic Gaussian
VNLM	Vector Non-Local Mean
XB	Xie and Beni

Chapter 1

Introduction

Human beings interact with their environment through their five sensory inputs: eye, ear, nose, tongue and skin. These sensors receive external signal that will be processed by brain to generate the appropriate reaction. All these sensors have their own characteristics and roles. However, most human interactions and analysis are basically conducted using the visual sensor.

Visual analysis allows human to interpret his surrounding, characterize its objects and interact with it based on his reflection. However, the anatomy of the human visual system has its limitation like all its sensors. We can only see in the visible spectrum i.e. in the spectral range of [400nm – 720nm] unlike some other species which have the ability to see outside this range [1].

Face recognition is a simple routine task performed easily by humans which is based on his visible analysis. Following the evolution and wide availability of computers nowadays, interest in incorporating human capacities into computers has grown. Thus, it is natural that the research on automatic face recognition has emerged.

• Motivation

Face recognition is involved in a wide range of applications including person verification and authentication, computer entertainment and security. However, even though human can perform this task easily, accurately, robustly and quickly, none of the existing face recognition systems can do so. The robustness of this task involves recognizing faces under different illumination conditions and various facial expressions. Particularly, the illumination is still an open problem. Indeed, face is a 3D object and, under certain illumination conditions, some area may be shadowed which makes it difficult to locate some critical features. Even with absence of shadows, under a non uniform illumination, a face looks different and the recognition is difficult. Furthermore, under perfect

illumination, features may be distorted with different incident angles [2]. We can conclude that face recognition in the broadband visible spectrum is difficult and requires controlled illumination conditions.

Multispectral images have emerged as a new technology that is able to deal with the various problems encountered with broadband systems. The development of new cameras and filters enables us to see beyond the visible spectrum for example in the Infrared [700nm-1mm], Ultraviolet [10nm-380nm], X-ray [0.01nm-10nm]... Multispectral image is a set of several monochrome images of the same scene each of which is taken at a specific wavelength. Each image is referred as a band or channel. This image may be seen as three dimensional image cube: Spatial dimensions which are the vertical and horizontal axes and a spectral dimension where resides the spectral information. A monochrome image has one spectral band. A pixel is represented by a scalar value. Multispectral image consists of at least two bands. A pixel is represented by a vector of P components where P is the number of spectral bands. Color images may be seen as multispectral image with three spectral bands but the term is commonly used for image with more than three bands. With more than a hundred bands, the common used term is hyperspectral image.

Using multispectral images for face recognition is justified by two main reasons: Narrow spectral bands exhibit more relevant facial information compared to conventional broadband color and black and white images. Indeed, we obtain a unique spectral signature of the facial tissue. Such information can be employed to enhance the accuracy of face recognition. Second, by using multispectral images, we are able to separate the illumination information from object reflectance in contrast to broadband images where it is almost impossible to do so. This separated information can now be used to normalize the images. For example, near-infrared spectral band can be combined with the visible image. This approach has been widely used to construct more effective biometric systems [3–5]. Thermal infrared images have also been widely used. In fact, these sensors detect the heat energy radiation of the face which is independent from the illumination in contrast to the reflectance [6, 7]. Furthermore, thermal infrared is less sensitive to scattering and absorption by smoke or dust and invariant to illumination change [8]. It also allows to reveal anatomical information which is very useful in detecting disguised faces [9].

• Scientific challenges

The quality of the multispectral image has a great implication on the accuracy of the recognition. Environmental corruption such as noise and blur are common phenomena of any captured images due to many factors. In particular, a multispectral image can be subject to quality degradation due to the imperfectness of the sensors [10]. Enhancing

the sensory data leads to performance improvement in problems such as image segmentation, classification and object identification. Basically, accuracy of any recognition system is influenced by these factors [11, 12]. Poor quality images often lead to wrong and missed features which in turn results in false rejection and false acceptance rates. The most common artifacts that are manifested include blurring, illumination, distortion and noise. Blur is a common issue in image processing that is caused by focus incorrectness, motion or environment factors and affects edge information which is crucial for recognition. Uniform illumination is also essential in this task. Any variation affects the recognition accuracy. Optical distortions are related to the functioning of the sensors. Finally, noise is an inevitable part in most real world images. The noise origin is mainly related to environment factors as well as imperfection related to the sensors and an incorrect use of it. Another stage which can be involved in sensory data enhancement is segmentation. This process is widely involved in many computer vision and pattern recognition problems. It consists in extracting important objects of the captured scene. A segmented multispectral image can be used for visualization purpose. In face recognition context, it is extremely useful to determine any details in the face which can be used to enhance the accuracy. These details may be scars in the face, birthmarks...

The sensory data enhancement should take advantage of the additional information brought by the spectral dimension. Indeed, some details may be absent in some bands but appear in other bands. Thus, for a given pixel, we no longer dispose of a single value but a group of values that are grouped in a vector called the pixel vector. Any enhancement process should take into consideration this important fact.

• Contributions

The specific objectives and contributions of this thesis can be summarized as follows:

1. Study the state of the art of segmentation and clustering algorithms and how to adapt them for segmentation of multispectral images.
2. Propose new clustering algorithms.
3. Propose new tools that are able to find the number of segments/clusters of a given data set.
4. Design new filters that are able to ensure maximum noise removal.
5. Proposed filters must ensure a good tradeoff between noise removal and details preservation since accidentally removing any detail would have great implication on the recognition accuracy.

6. These filters should gain benefits from the additional information brought by the spectral dimensionality.
7. Conduct experiments on several color and multispectral images in addition to machine learning data sets.

Throughout this thesis, we focus on cluster analysis as it is a widely known mechanism for image segmentation. We propose new clustering algorithms based on the partitional clustering and new methods to predict the real number of segments of a given data set. In addition, we propose two denoising schemes. A particular attention has been given to ensuring noise removal with details preservation as well as gaining benefit from the new information brought by the spectral dimensionality.

• Document organization

This manuscript includes four chapters and a conclusion. The below paragraphs describe the content of each chapter.

In chapter two, we start by giving an idea about the use of multispectral image in several fields ranging from medicine to military. A general presentation of the process of face recognition is highlighted where its different components are described. We emphasize also the importance of using multispectral image in this application and how it can enhance the accuracy of recognition. In this process, the quality of the images is an important factor. As we are focusing on denoising and segmentation, we study the state of the art of both tasks and describe how the research community has approached these important image processing problems in the context of multispectral images. Denoising algorithms are studied and for multispectral image segmentation, particular attention is given to clustering techniques applied to multispectral image segmentation. We study also the problem of cluster validity and we give some examples of widely used CVIs.

In chapter three, we present our proposed filters for multispectral image denoising. These two filters are designed so that they gain benefit from the additional information brought by the spectral dimension. This chapter is divided into two main parts: we propose in the first one, the first denoising algorithm, called the Vector Anisotropic Gaussian Filter (VAG). It is based on the well-known anisotropic Gaussian filter whose intrinsic property is edge and details preservation. Particular noise model and matrix transform are used to ensure good denoising performance. In the second part, we propose a second denoising algorithm based on the Non-Local Mean (NLM) filter known to its robustness to Gaussian noise. We propose a parametrization of this filter based on an optimization framework where filter parameters are adjusted in a way that guarantees maximum noise

removal.

Chapter four is dedicated to the segmentation task. We present our segmentation algorithm based on cluster analysis and the famous Fuzzy C-Means (FCM) method. This algorithm is inspired from the gravitation theory. We propose also in this chapter two cluster validity indexes. The first one is based on the geometrical shape of n-sphere. The second index is based on Jeffrey divergence computation. We base our thought process on the theory of density of probability to develop a particular index that is able to predict the real number of cluster for various data sets.

The evaluation of the proposed algorithms is the subject of the fifth chapter. Several experiments have been conducted on multiple data sets to assess performances of these algorithms. We use our database of multispectral face images in addition to other data sets from other well-known databases. We compare the performance of our algorithms with different approaches.

The conclusion chapter recalls the main contribution of this thesis and gives few potential expansions of this work.

Chapter 2

Literature review

2.1 Introduction

Multispectral imaging is a powerful tool that involves conventional imaging system, like cameras, augmented with multi-wavelengths capturing enabled devices, like filters. The purpose of this combination is to improve data robustness, distinguishability as well as readability. Hence, we are dealing with P dimensional images, or hypercube, with two spatial dimensions and P spectral dimension. Hyperspectral or Ultraspectral are also possible terminologies that may replace the terminology “multispectral” when the number of captured wavelengths becomes very high. While there are no formal limits between the later three terms, multispectral is generally used for images with 2 to 100 bands. Hyperspectral are with hundreds of bands and Ultraspectral with thousands of spectral bands. Hereafter, we use the terminology ‘multispectral images’ for any cube that includes more than one spectral band.

We focus in this chapter on surveying the research literature in two main enhancement tasks namely image denoising and segmentation. We start in section 2.2 by giving a general idea of applications that use multispectral images. A particular attention is given to face recognition as an application that can use multispectral images. In section 2.3, we give an overview of research works related to multispectral image denoising. In section 2.4, we detail the works related to multispectral image segmentation particularly those based on cluster analysis. We conclude this chapter in section 2.5.

2.2 Multispectral image and its applications

Nowadays, multispectral imaging is involved in many applications for different purposes. In remote sensing, multispectral imaging is used to characterize regions or area

of interests like vegetation, analyze urban growth, survey climate change ... The first multispectral remote sensing image was conducted by Appolo 9 to capture red, green and infrared channels of southern USA. Since that, many multispectral remote sensing systems have been constructed such as NASA's Landsat which provides coverage of the earth surface in the visible, near-infrared, short-wave, and thermal infrared bands. In the captured images, each pixel contains the spectral response representing the light absorption of the area that the pixel represents. The fact that objects or materials in a scene reflect, absorb and scatter electromagnetic wave differently, offers the possibility to identify the area based on the spectrum (also called spectral signature).

Multispectral imaging has been also extensively used in biological and medical fields. Indeed, it is used during surgical operations to enhance tissue visualization and discrimination [13–15]. Furthermore, multispectral imaging is used to evaluate wound healing [16]. In oncology, , fluorescence and diffuse reflective spectroscopy have been jointly used to detect precancerous and cancerous lesions [17].

In food industry, products quality assessment process has been evolved by the integration of multispectral images. Several kinds of defects that may affect products like fruits or vegetables are easier to detect in spectrum beyond the conventional visible spectrum used in RGB or grayscale images [18, 19].

Multispectral imaging is also an obvious choice for military application. Indeed, it has been extensively used for surveillance, defeating camouflage, detecting landmines, target acquisition ...

Numerous applications nowadays rely on human identification. Face recognition is one of the primary biometric based system. The increasing demands for security as well as the development in cameras and sensors have motivated more research activities in face recognition but this problem is still largely unsolved. Since the 1960s, following the emerging of face analysis and modeling techniques, face detection, recognition and tracking have been developed but the reliability of these systems is still argued. Thus, achieving reliable systems is a great challenge to computer vision and pattern recognition researchers. Kanade [20] was the first who developed an automatic face recognition system. A typical face recognition system basically involves four modules as depicted in Fig 2.1: first of all, the face should be detected. This task consists in isolating the face from the background. In case of a video input, a face tracking component should be involved. Next, facial components should be identified and face image should be normalized with respect to some geometrical and photometrical properties such as illumination. After that, we need to extract feature vector which enables us to differentiate between faces of different persons. The technique used for this procedure should be stable with respect to any variation in the geometrical and photometrical properties. Finally, the feature vector is compared and matched against features of faces in the database. The final output of this process is the determination of the identity of the person to a certain

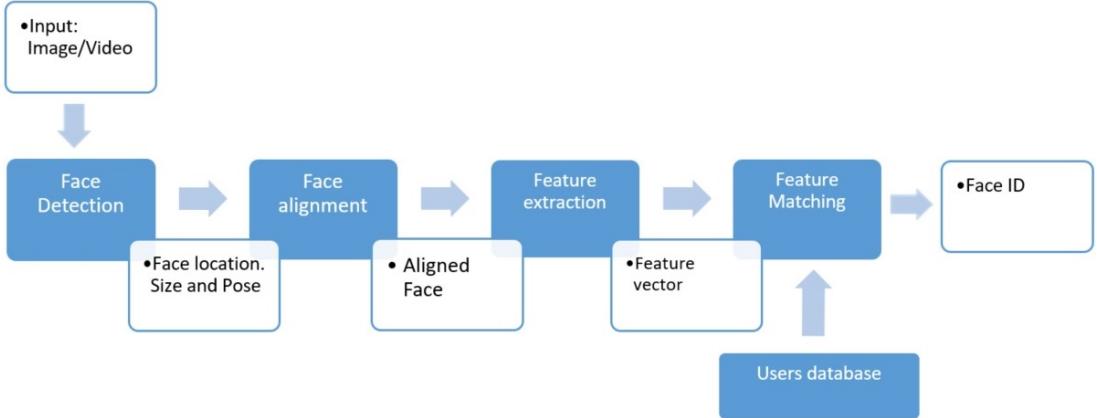


FIGURE 2.1: Processing pipeline of face recognition

degree of confidence. Several aspects interfere in this process and have great influence on the final output particularly the feature vector and the method used to obtain it in addition to the technique used to differentiate between faces. Although research on these topics has achieved great advances, the task has ended up to be difficult. Indeed, classic face recognition systems require a controlled environment where viewpoint, illumination, facial expression, occlusion and so on [21] should be controlled and adjusted prior to any recognition procedure. This also includes a cooperative person that will be recognized. Such luxury is not always available. For example, due to changes in illumination directions and skin colors, significant performance degradation is experienced [22, 23].

To cope with these issues, researchers have studied the potential of multispectral images as an alternative to conventional broadband imaging system and great results have been achieved so far. However, it is commonly known among the community that the quality of images is an important factor in the design of any multispectral based system.

2.3 Overview of multispectral image denoising

The quality of images is an important factor in almost all applications. For example, in MRI images, any detail is essential and can have great implications. In face recognition applications, any detail in the image can improve the accuracy of the recognition. However, noise is inevitable in all real broadband and multispectral images. Thus, it is essential to dispose of techniques that ensure noise removal in order to increase performance in many image processing problems such as classification and segmentation [24, 25].

Several techniques have been proposed to approach this problem. The work of Luisier et al. [26] constitutes the state of the art for multispectral image denoising. Authors

proposed a denoising algorithm parameterized as a linear expansion of thresholds [27]. Optimization is carried out using Stein's Unbiased Risk Estimator (SURE) [28]. The thresholding function is pointwise and wavelet based. A nonredundant orthonormal wavelet transform is applied on the noisy input image. Next, a vector-valued thresholding of individual multichannel wavelet coefficients is performed. Finally, an inverse wavelet transform is applied to obtain the denoised image. Application of an orthonormal wavelet transform is justified by two main properties. First, assuming a white Gaussian noise in the image domain Ω , its wavelet coefficients remain also Gaussian and are independent between subbands. Second, the Mean Square Error (MSE) in Ω is equal to the sum of subbands MSEs.

Another SURE based approach was proposed in [29]. Authors used a generalized form of shrinkage estimate. The optimal choice of the parameters is based on the minimization of the quadratic risk or MSE that depends on the original data which are unknown. Parameters are chosen so that they minimize the obtained risk. Note that the proposed denoising framework was applied by considering wavelet-based approach. Two decomposing schemes were proposed: a decimated M-band wavelet transform and M-band dual-tree wavelet decomposition. For each case, the associated estimator is obtained. Another scheme was proposed in [30]. The algorithm jointly removes noise and blur from image. It is based on the Expectation and Maximization (EM) algorithm [31]. The noisy signal is decomposed into two independent parts: the first one represents the blurring problem while the second represents the denoising one. The latter is performed in wavelet domain. A Gaussian scale mixture is used to model the probability density of the wavelet coefficients. Besides that, a coregistered auxiliary noise-free image of the same scene is included in the framework to improve the restoration process. In fact, it provides an extra prior information to the model.

A partial differential equation denoising system is proposed in [32]. It is based on the Total Variation (TV) denoising method used in [33] which proposes an objective function to be minimized. For this purpose, authors used the time marching method [34]. The denoising task is then modified to a partial differential equation (PDE) problem. Authors injected in this PDE problem an auxiliary image as a prior as seen with the previous method. This approach is justified by the fact that edge directions and texture information of the auxiliary image are similar to the noisy image. Thus, a smoothing term that takes into account the contribution of this prior information is added.

A non-local multidimensional TV model is recently proposed in [35]. Authors presented the denoising problem as a minimization of a mean square cost function that depends on a regularization term. The non-local property is not restricted to patches from one band but also to other bands with high correlation. Thus, for a given pixel, the similarity between patches from other bands is considered in the computation of the weight. The multichannel image is first divided into many groups. For a given band, bands with

high correlation are grouped together. In addition, the regularization parameters are computed adaptively for each band. They are derived from the estimated noise standard deviation using the coefficient of the highest frequency wavelet subband. Obtained minimization problem is solved using Bregmanized operator splitting [36] which introduce an auxiliary variable. The unconstrained problem is approached using Bregmann iteration method which leads to an update algorithm where Gauss-Seidel and shrinkage methods are used. The proposed framework was jointly used for multichannel image denoising and inpainting.

Zhao et al [37] proposed a denoising framework based on sparse presentation and low rank constraint. Authors analysed the difference in rank between clean and noisy image and concluded that the rank of the clean image is far smaller than the size of the multichannel image which is not true for noisy image. Thus, an assumption is made: a low rank is a characteristic of noise free multichannel image. This information is incorporated in the cost function. Furthermore, the cost function requires patch extraction. To avoid the problem of curse of dimensionality and large error, authors suggested to reshape the 3D spectral cube into a 2D matrix by converting each band into a vector and then patches are extracted. The optimization with respect to some variables is carried out by fixing some other variables. The overall complexity of this framework is $O(HLP^2 + HLP + L)$ with H and L are the size of the spatial dimension.

Yuan et al. [38] studied the noise presented in multichannel image. Authors concluded that there are two types of noise distribution: one distribution in spatial domain and one in spectral domain. Thus, two TV models are used: one applied for multichannel image denoising in spatial domain and the other model is applied in spectral domain. The two models are both optimized with the split Bregman method. Authors studied also the complementary nature of both schemes and concluded that both denoising results can complement each other and a fusing process can bring additional improvement. By using the metric Q proposed in [39], a fusion scheme between bands from each denoising result is proposed and the final denoised multichannel image is obtained.

Yuan et al. [40] proposed also another denoising method. The regularization term in the cost function is often approached by a kernel model. However, it has three main drawbacks when applied for multichannel image denoising. First, the spectral information is not considered. Second, since the spatial resolution is lower than the spectral resolution, this approach is inefficient. Finally, noise differs from one band to another. This fact is not considered. Given these challenges, authors suggested two strategies. In the first one, a spectral-spatial kernel model where the spatial and spectral information are simultaneously used. In the second strategy, the noise distributions in spectral bands are considered different and a local kernel is used to balance the contributions between bands.

A denoising framework based on the Bayesian least squares optimization problem is proposed in [41]. This framework requires computation of the posterior distribution based on Monte Carlo sampling [42]. Given the noisy pixel, the procedure starts by choosing some neighbor pixels. Then, the acceptance probability of the sampled pixel given the noisy one is used to decide whether the sampled pixel is considered or not. This decision is based on a comparison between the acceptance probability and the random variable drawn from the uniform distribution. After selecting sample pixels, the importance-weighted Monte Carlo posterior estimate is computed using the weighted histogram approach proposed in [43] and finally the denoised pixel is obtained.

Peng et al. [44] proposed a vector version of the bilateral filter. The basic assumption behind this filter is that pixels which have influence on the restored pixels are not just neighbor pixels but neighbor pixels with similar values. Typically, in a similar way to Gaussian filter, bilateral filter is defined as a weighted average of neighbor pixels. However, in order to preserve edges, it takes into account the difference in value with the neighbor pixels. In their work, Peng et al. extended the bilateral filter to the vector form. The dissimilarity measure is now expressed as a multivariate Gaussian function. For simplification purposes and to avoid the computation of the noise covariance matrix, data are projected into subspace using Principle Component analysis (PCA) and noise variances of individual channels are computed using the median absolute deviation method [45]. However, this adhoc method makes the results enormously dependent on the choice of the filter parameters.

Authors in [46, 47] proposed an optimization framework for the vector bilateral filter using SURE. Authors have proven that within a neighborhood of the corresponding edge pixel, a high Signal to Error (SER) measure is obtained by maximizing the weight attributed to neighbor pixels with similar values and minimizing the weight given to pixels with significant different values. Authors have also demonstrated that the SER of the vector version of the bilateral filter is always greater than the component wise 2D bilateral filter. The optimization scheme is based on the minimization of the MSE. However, the underlying difficulty of this measure is that it involves the original image which is unknown. MSE is seen as a random variable that depends on the noise. Its expected value is called the risk. To overcome this issue, SURE is applied and the filter parameters are obtained by minimizing the expression of SURE. Obtained minimization problem is non-linear and is solved numerically using the Sequential Quadratic Programming (SQP). Experiments on color and multispectral images have been conducted and comparison using the Peak Signal to Noise Ratio (PSNR) is presented.

Manjon et al. [48] have recently proposed a new algorithm for multispectral image denoising. It is based on the Non-Local Mean (NLM) filter [49]. NLM filter is designed so that it takes advantage from the redundancy exhibited in the images. This redundancy is no longer pixel based but window based. In other words, every small window centered

around a pixel is supposed to have many similar windows. These windows can be located anywhere in the image domain Ω and are no longer restricted to the neighborhood. In the multispectral framework, information from various bands are combined and a new weight is proposed. This filter is highly dependent on the choice of three parameters: The radius of the search window, the radius of the neighborhood window and a smoothing parameter that controls the degree of the smoothing. The latter is very important. Indeed, with a small value, little noise will be removed. On the other hand, with a high value, image will be blurred. Authors have set these parameters manually.

2.4 Overview of multispectral image segmentation

Segmentation of multispectral image is one of the topics that has started to gain attention among the research community [50–54]. Image segmentation in general is a crucial task in many computer vision applications. It consists of partitioning an image to its constituent objects. Image segmentation is an ill-posed problem [55] as the definition of an object in an image is ambiguous. Besides that, the definition of what is not an object is also ambiguous. Last but not least, there are no measures that can assess segmentation result objectively.

Segmentation can be considered as a clustering problem [56]. The applicability of clustering algorithms for image segmentation has been widely studied [57]. Clustering is an important unsupervised learning process. It is involved in many applications such as pattern recognition, data mining, and image segmentation and classification [58, 59]. It consists of regrouping input data into structures called clusters. Members of each clusters are very similar to each other compared to members of other clusters. Several issues are related to multispectral image clustering [50]. These issues can be categorized as computation and complexity related, and quality and accuracy related. The increased resolution thanks to the improvement of sensors has led to increase in image sizes. Thus, we may end up easily with million of pixels. For many clustering algorithms that require for example distance matrix computation, this is a time and memory consuming process. Another issue related to computation is the feature dimension. Indeed, acquisition can be done on different bands. Thus, for one scene, we have different images in different wavelengths. Therefore, the dimensionality of the feature grows which causes also more computation and memory consumption. The so called mixed pixel must be also considered when clustering algorithms are applied. These pixels contain spectral responses of different component. Including fuzzy logic in clustering process may be a solution to this problem. The presence of overlapping clusters is also a major issue in cluster analysis. Two objects that may belong to different clusters may have similar

features. Thus, excluding spatial information in the clustering process may lead to erroneous results. One of the most critical problems in cluster analysis is the number of clusters or segments. More often, clusters may have different densities of population. In some cases, a small cluster can contain important data but it is difficult to find it or may be considered as outliers. Last but not least, we may dispose of a data set without priori knowledge of its number of clusters. For this purpose, researchers have proposed what is called Cluster Validity Index (CVI). CVIs are tools that allow to predict the real number of clusters of a given data set with unknown number of clusters. In the following paragraph, we review the state of the art of clustering algorithms applied for multispectral image segmentation and give an overview of recent works on CVIs. We finalize by presenting an overview of segmentation techniques other than clustering.

2.4.1 Clustering for multispectral image segmentation

Several clustering algorithms are proposed for multispectral image clustering. In [51], authors jointly use K-means clustering [60] and Agglomerative Hierarchical Clustering (AHC) [61] to take advantages of each of them and add a refinement process based on spatial information to eliminate their limitations. K-means is used first as a preprocessing step to end up with a high number of clusters. This step prevents AHC from operating on large data set. AHC is then used to merge clusters based on a proximity matrix calculated for all clusters. The refinement process consists in separating between large and non well separated classes. With the assumption that misclassified pixels are located in boundary of clusters, these pixels will be assigned to the closest adjacent cluster which will lead to a smoothing in the spatial domain.

In [62], authors introduced an objective function based on a modification of the possibilistic objective function proposed in [63] by adding a spatial regularization term forcing neighborhood membership to be as close as possible to the pixel.

Authors in [64] proposed a size weighted objective function in order to deal with the problem of unequal cluster size. In fact, in FCM objective function, big clusters are more influential than small clusters. In worst case, small clusters will be missed. To avoid this situation, after some distance measures and clusters cardinalities computation, a weight is computed and associated to each point.

In [65], Affinity Propagation (AP) [52] is applied with fuzzy statistics for multispectral image clustering. AP is used for exemplars learning. The center of a cluster is called exemplar if it is selected from data points. Application of fuzzy statistics is justified by the fact that even though multispectral images have similar information and spatial structures, each one has its own spectral signature. A fuzzy statistical similarity measure allows a proper membership assignment of uncertain pixels and gives an accurate

estimation of how much similar two pixels vectors are.

A new AHC algoithm was proposed in [66] in which a combination of indices is used to find the most similar classes and merge them. The starting point is a clustered image. These indices use spatial and spectral criterions:

- Spectral Similarity index: The normalized Mahalanobis distance between the mean vectors of every pair of classes.
- Spatial Boundary Index: Two classes with significant common boundary are more likely to merge. By considering 8 neighbors for each pixel, the total boundary counts for every pair of classes is computed and merging decision is taken based on its value.
- Spatial Compactness index: This index is derived from the previous one. This index is computed for each class, . Merging decision of two classes will be taken based on combination of their spatial compactness indices.
- Class Size index: Class size index of a class is the fraction of the number of pixels in a class. The Class size index of a pair of classes is the product of their class size indexes.

The aggregation Index -on which the merging decision is based- is a weighted combination of the four indices. However, finding the weight of each index is a delicate task. For this reason, the authors propose to run the algorithm several times until satisfactory result is obtained. This method extremely depends on the clustering algorithm used at the beginning.

In [53, 54], Soltani et al. studied the partitional method applied for multichannel image segmentation and classification particularly the AP algorithm. The focus is on overcoming the drawbacks of such algorithm especially the high sensitivity to the choice of its parameter and its quadratic complexity. This is conducted by a pre-precessing procedure that aims at reducing the number of pixels to be classified before applying AP. The multichannel image is decomposed into blocks and the reduction is applied on each block separately which results in a less memory usage. In addition, to correctly estimate the real number of clusters, a criterion related to the inter-class variance is maximized. Cariou and Chehdi [67] proposed a novel fully unsupervised clustering algorithm based on the nearest neighbors with application for multichannel image segmentation and classification. The objective is to assign a label to each pixel vector based on a modified maximum likelihood criterion. This is done by partitioning the data into clusters by sampling from local posterior distribution. Specifically, for each data object, a set of objects and its label are chosen using the k -nearest neighbors. The local posterior label

distribution is chosen based on a Gaussian kernel and in a way that the solution won't be trapped into a local minimum or a saddle point. Furthermore, the stopping criterion for this algorithm is well adjusted using a conditional entropy derived from kozachenko-Leonenko entropy estimation. Knowing that this measure reaches an upper limit at the convergence, the stopping criteria is set up from its relative magnitude variation. Experiments on synthetic and multichannel images were conducted and comparison with other unsupervised clustering methods was conducted.

2.4.2 Other techniques for multispectral image segmentation

Besides clustering techniques, several other methods are proposed for multichannel image segmentation.

Many algorithms rely on Bayesian framework for segmentation. Li et al [68] proposed a semisupervised segmentation algorithm based on multinomial logistic regression with active learning. The problem is formulated as assigning labels to image regions and maximizing the posterior distribution. This density is formulated as a multinomial logistic regression model [69] which introduces a vector of fixed functions of the input and a set of logistic regressors for each class label. Authors suggested to use a radial basis function kernel for the vector of functions. By using a semisupervised approach, the posterior density is reformulated. The regressors are learned using both labels and through a graph based technique. The maximum a posteriori segmentation is computed by the α -expansion min-cut-based integer optimization algorithm [70].

Li et al. [71] proposed a second algorithm that is inspired from the previous one which is based on variable splitting and augmented Langrangian algorithm [72] to learn class posterior probability. Multilevel logistic prior is then used to segment images. In order to reduce complexity, the active learning is introduced to avoid acquiring large training data sets. The α -expansion min-cut-based integer optimization algorithm is also used to obtain the maximum a posteriori segmentation.

In a third algorithm, Li et al [73] used subspace multinomial logistic regression and Markov random field for multichannel image segmentation. The subspace projection is conducted to better characterize highly mixed pixels and noise. Multilevel logistic Markov random field is used to include contextual information. The α -expansion min-cut-based integer optimization algorithm is again used to carry out the optimization process.

The use of the multinomial logistic regression classifier in the previous algorithms is justified by its ability to learn the class distributions themselves which has contributed in many multichannel image classification problems.

Authors in [74] proposed a segmentation algorithm based on one-class support vector

machine (SVM) [75]. From a subset of multichannel image, clustering and validation procedures are conducted in order to find the optimal number of clusters. Then, segmentation is performed using K-nearest neighbor classification. The training sample for the classifier is the sample obtained from the first step. This sample is obtained by conducting a band selection procedure for the purpose of dimension reduction instead of using transformation-based dimension reduction techniques. The cluster validity procedure is conducted using a novel algorithm called the support vector power of spectral discrimination that is derived from the one-class support vector machine.

Mitra et al. [76] proposed a segmentation algorithm for multispectral image based on the active support vector machine. The need to solve a quadratic programming (QP) problem in common SVM results in huge memory consumption and computational time. To circumvent these shortcomings, authors broke the large QP problem into series of smaller QP problems. This approach exploits the fact that removing the points that correspond to zero Lagrange multipliers (non-support vectors) doesn't change the solution of the problem. At each step of the active support vector, the most informative point not belonging to the current support vector is queried with its label. This strategy aims at minimizing the total number of labeled points required by the learning algorithm. We note that for k classes, we need to run this routine k times and that the label of the queried point is determined by human expert. Although this algorithm presented good results, it still requires human intervention and highly sensitive to wrong labeling which would result in performance degradation.

One of the famous techniques used for classic image segmentation is the watershed technique which is based on the mathematical morphology. Watershed transform considers the image as a topographic relief. For each pixel, its value is considered as its elevation. From a topographic view, the pixel with a minimum is considered as a catchment basin. This transformation is generally applied on the gradient image. Therefore, at homogeneous regions, we have a minima. On the other hand, the highest values are located at the borders between these regions. An efficient watershed transform was proposed in [77] that is based on flooding simulation. The output of this transform is an image composed of regions.

Authors in [78] proposed a watershed segmentation algorithm for multichannel image segmentation. The algorithm starts by extracting features from the multichannel image. Depending on the output of this procedure (single or multi-band image), one of four techniques used to obtain gradient image is chosen: difference between the dilation and erosion of the image in case of one band or to compute the vectorial gradient, a multidimensional gradient or a watershed segmentation maps a posteriori in case of multi-band image. Next, a segmentation and classification procedure is applied which results in a final segmentation map where every region in the image is identified and

classified. A major drawback of this approach is that small spatial structures are not identified as separate regions.

2.4.3 Cluster validity index

Most of the clustering algorithms require the number of clusters. Knowing that clustering is an unsupervised learning process, users don't have any prior knowledge about the number of clusters. With a lower number of clusters than the real one, we end up with an under-partitioned data set. Conversely, a higher number of clusters leads to an over-partitioned data set. Almost all the proposed CVIs are derived from the partitional clustering particularly the famous Fuzzy C-Means (FCM) clustering algorithm. Given a data set $X = \{X_j\}_{j=1}^N$ with $X_j = (x_{j1}, x_{j2}, \dots, x_{jd}) \in \Re^d$, FCM clustering seeks to minimize the following objective function:

$$J_{FCM}(U, V) = \sum_{i=1}^c \sum_{j=1}^N \mu_{ij}^m d_{ij}^2 = \sum_{i=1}^c \sum_{j=1}^N \mu_{ij}^m \|X_j - v_i\|^2 \quad (2.1)$$

where c is the number of clusters, $U = (\mu_{ij})$ is the $c \times N$ fuzzy membership matrix where each element $\mu_{ij} \in [0, 1]$ represents the degree of belonging of the j^{th} data to the i^{th} cluster, $V = (v_1, v_2, \dots, v_c)^T$ is the $c \times d$ matrix of cluster centers, $m > 1$ is the fuzzification degree and $\|\cdot\|$ is the Euclidian norm.

The problem of finding the number of clusters is called cluster validity. Many CVIs are proposed in the literature; those that are widely used are reviewed in the next paragraphs. They can be roughly divided into two categories: CVIs based on membership values and CVIs based on membership values and data set.

2.4.3.1 CVI based on membership values

In crisp environment, data point belongs to one and only one cluster. Conversely, in fuzzy context each data point belongs to all clusters with a degree of membership called the membership value. Bezdek proposed in [79] his famous Partition Coefficient (PC) which is the sum of the squared membership values of all data points over the number of data points.

$$PC = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^N u_{ij}^m \quad (2.2)$$

The best partition is obtained with the maximum value of PC.

Bezdek also proposed his second CVI, the Partition Entropy (PE) [79, 80] which measures the amount of fuzziness of the data.

$$PE = -\frac{1}{N} \sum_{i=1}^c \sum_{j=1}^N u_{ij} \cdot \log(u_{ij}) \quad (2.3)$$

The minimum value of PE is obtained with the optimal partition.

Dave [81] modified PC index to reduce its monotonicity and proposed the Modified Partition Coefficient (MPC) index.

$$MPC = 1 - \frac{c}{c-1}(1 - PC) \quad (2.4)$$

The optimal partition is given by the maximum value of MPC.

Chen proposed validity index P in [82]. It is composed of two terms. The first one determines the compactness within a cluster. It reflects the closeness of the data point to the center of the cluster. The second term reflects the separation between clusters. The maximal value of P points to the best partition.

$$P = \frac{1}{N} \sum_{j=1}^N \max_i(u_{ij}) - \frac{1}{K} \sum_{i=1}^{c-1} \sum_{j=i+1}^c \left[\frac{1}{N} \sum_{l=1}^N \min(u_{il}, u_{jl}) \right] \quad (2.5)$$

With $K = \sum_{i=1}^{c-1} i$.

Recently, Žalik proposed a CVI for the estimation of the number clusters with different sizes and densities [83]. This index aims at quantifying two properties of fuzzy clusters: overlap and compactness. The overlap measure quantifies the degree of separation between clusters. When overlap is low, separation between clusters is high and vice versa. The compactness measure is used to quantify the variation or scattering of the data within a cluster. We notice that the first measure is computed between clusters while the second one is within the cluster itself. Thus, the optimal data partition is characterized by a higher compactness and lower overlap. The proposed compactness measure is calculated as a sum of the compactness rates of all data points. The compactness rate of a data point i in the j^{th} cluster is equal to the membership of this point in this cluster if the difference between this membership value and the other membership values in other clusters is higher than a predefined constant. Otherwise, it is equal to 0. The overlap measure between two clusters is computed based on the overlap degrees of each data point associated strong enough to both clusters.

$$CO = \frac{1}{N} \sum_{i=1}^N \left(\sum_{j=1:c}^c C_{ij} - \sum_{a=1}^{c-1} \sum_{b=a+1}^c O_{abj} \right) \quad (2.6)$$

Where:

$$C_{ij} = \begin{cases} \mu_{ij} & \text{if } \mu_{ij} - \mu_{ik} \geq T_c, k = 1, \dots, c, k \neq i \\ 0 & \text{otherwise} \end{cases}, k = 1, \dots, c, k \neq i \quad (2.7)$$

and:

$$O_{abj} = \begin{cases} 1 - (\mu_{aj} - \mu_{bj}) & \text{if } \mu_{aj} - \mu_{bk} \geq T_o, a \neq b \\ 0 & \text{otherwise} \end{cases} \quad (2.8)$$

T_c and T_o are two constants.

As these indexes are membership based, they exhibit some drawbacks:

- Monotonicity to the number of clusters.
- Indexes are very sensitive to the degree of fuzziness.
- No contribution for the data.

2.4.3.2 CVI based on membership and data

One of the well known CVIs that is based on the data and membership is the Fukuyama and Sugeno (FS) CVI [84]. FS index is a combination of two terms: the first is a measure of the compactness that mixes the fuzzy membership with the data set and the second mixes the fuzzy membership, the distances between cluster centers, and the mean of these centers. The best partition is the one with the minimum of FS.

$$FS = \sum_{i=1}^c \sum_{j=1}^N \mu_{ij}^m \|X_j - v_i\|^2 - \sum_{i=1}^c \sum_{j=1}^N \mu_{ij}^m \|v_i - \bar{v}\|^2 \quad (2.9)$$

Where $\bar{v} = \frac{1}{c} \sum_{i=1}^c v_i$.

Xie-Beni proposed their index called XB in [85]. XB is computed as the ratio of the within cluster compactness to cluster separation computed as the minimum distance between cluster centers.

$$XB = \frac{\sum_{i=1}^c \sum_{j=1}^N u_{ij}^2 \cdot \|x_j - v_i\|^2}{N \cdot \min_{i \neq l} \|v_i - v_l\|^2} \quad (2.10)$$

Pakhira [86] proposed another CVI that mixes compactness and separation called the *PBMF* index.

$$PBMF = \left(\frac{1}{c} \cdot \frac{E_1}{J_m} \cdot D_c \right)^2 \quad (2.11)$$

with $J_m = \sum_{i=1}^c \sum_{j=1}^N u_{ij} \cdot \|x_j - v_i\|$, $E_1 = \sum_{j=1}^N \|x_j - v_1\|$ and $D_c = \max_{1 \leq i, j \leq c} \|v_i - v_j\|$. Žalik [87] proposed a new validity index based on the separation between clusters and

the concept of clusters overlap instead of the compactness measure.

$$OS = \frac{\sum_{i=1}^N \sum_{x_j \in C_i} O_{x_j}}{\sum_{i=1}^c \min_{j=1..c, i \neq j} \|v_i - v_j\|} \quad (2.12)$$

with:

$$O_{x_j} = \begin{cases} \frac{a}{b} & \text{if } \frac{b-a}{b+a} < 0.4 \\ 0 & \text{otherwise} \end{cases} \quad (2.13)$$

$$a = \frac{1}{|c_i|} \sum_{x_l \in C_i} \|x_j - x_l\| \quad (2.14)$$

$$b = \frac{1}{|c_i|} \sum_{x_l \notin C_i} \|x_j - x_l\| \quad (2.15)$$

with $|c_i|$ is the cardinality of the i^{th} cluster C_i . This CVI doesn't involve membership values. It is based on distance computation between data objects for the overlap measure and the distance between cluster centers as separation measure. The best partition is the one which has low intercluster overlap and high separation thus low value of OS .

2.5 Conclusion

In this chapter, we reviewed state of the art of literature of multispectral image denoising and segmentation. Almost all the denoising algorithms are basically designed to denoise broadband grayscale and color images. These algorithms have been then extended to the vector case. The parametrization has been conducted either with adhoc means or using an optimization framework. As for multispectral image segmentation, we notice that almost all the proposed algorithms originate from the FCM algorithms. The number of clusters is a major problem that has been widely addressed and still an important challenge in cluster analysis.

On the basis of this literature review, we propose new algorithms for multispectral face images denoising as well as new segmentation methods based on the FCM algorithm. New CVIs are also proposed to determine the number of clusters even with complex data sets.

Chapter 3

Multispectral image denoising

3.1 Introduction

Following the widespread of applications relying on image analysis, the need for high quality images has grown. No matter how good your camera or sensors are, there is always an improvement that could be done. In this context, there is an overlap between image enhancement and restoration. While the first procedure is largely subjective, the second one is objective. As an example, contrast stretching is applied to give the image a pleasant aspect to the viewer. Thus, it is considered as an enhancement procedure. On the other hand, deblurring and denoising are considered as restoration techniques. Multispectral images are prone to these imperfection particularly the noise. Basically, during image acquisition, many factors affect the sensors such as the level of the light and sensor temperature [88]. Furthermore, image can be contaminated by noise during transmission. Noise appears to be inevitable and thus, it is essential to apply techniques that ensure maximum noise removal. Another problem arises here: denoising algorithms are generally accompanied by a loss of important features in images. In context of face recognition, any detail would have a great implication on the accuracy of the system. Thus, it is essential to design filters that are able to give good denoising performance while preserving the maximum of details. In this chapter, we extend well known filters to the vector case and propose two new filters adapted for denoising multispectral images. The choice of these filters is triggered by the fact that these filters, by definition, are designed to preserve details and reduce the loss of information. We also consider the following additive noise:

$$I_{in}(m, n) = I_{or}(m, n) + N(m, n) \quad (3.1)$$

With I_{or} is the original pixel and $N(m, n)$ the Gaussian noise, (m, n) are the pixel coordinates in the spatial domain.

3.2 Vector anisotropic Gaussian (VAG) filter for multispectral image denoising

We first recall the definition of anisotropic Gaussian filter then we will describe how to extend it to the vector case. More details about the classical anisotropic filters can be found in classical image processing and computer vision references such as [89–92].

3.2.1 Anisotropic Gaussian filter

Gaussian kernel is a low pass filter frequently used in image processing. It is characterized by its strict positiveness and exponential decay in both signal and frequency domains. While Gaussian filters can have different shapes, the most widely used are the isotropic ones. However, one may argue that isotropic filter can deteriorate details of the image such as edges and lines. Indeed, with isotropic filtering, parallel lines are highly likely to be blurred [93]. Furthermore, with isotropic smoothing, crossing lines are not well detected [94]. Thus, for a good design of a filter, one may prefer to ignore the distortion along a line or edge while cumulating more information about them using the orientation. To cope with this problem, the anisotropic Gaussian filter has been proposed.

The general case of the anisotropic Gaussian filter is the convolution of two Gaussians. It is given by:

$$H(u', v', \sigma_{u'}, \sigma_{v'}, \theta) = \frac{1}{\sqrt{2\pi}\sigma_{u'}} \exp\left\{-\frac{u'^2}{2\sigma_{u'}^2}\right\} * \frac{1}{\sqrt{2\pi}\sigma_{v'}} \exp\left\{-\frac{v'^2}{2\sigma_{v'}^2}\right\} \quad (3.2)$$

where “*” is the convolution and:

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{pmatrix} u \\ v \end{pmatrix} \quad (3.3)$$

u and v represent the horizontal and vertical axes respectively. u' and v' are the axes in the direction of θ and orthogonal to θ respectively. $\sigma_{u'}$ and $\sigma_{v'}$ are the standard deviations of major axis and minor axis respectively. The decomposition of the filter is given as follows [93]:

$$H(u', v', \sigma_{u'}, \sigma_{v'}, \theta) = \frac{1}{2\pi\sigma_{u'}\sigma_{v'}} \exp\left\{-\frac{1}{2} \left(\frac{(ucos\theta + vsin\theta)^2}{\sigma_{u'}^2} + \frac{(-usin\theta + vcos\theta)^2}{\sigma_{v'}^2} \right) \right\} \quad (3.4)$$

Figure 3.1 illustrates an anisotropic Gaussian filter with orientation θ . If an edge lies along the direction of u' -axis, the edge pixels are highly likely to be included in the spatial neighborhood. Consequently, the edge will be preserved while denoising the image [92]. The angle θ is given by:

$$\theta' = \arctan \left(\frac{G_v(u, v)}{G_u(u, v)} \right)$$

$G_v(u, v)$ and $G_u(u, v)$ represent the gradient in the vertical direction (v-axis) and horizontal direction (u-axis) respectively.

$$\theta = \theta' + 90^\circ$$

Thus, the anisotropic filter is expressed as follows:

$$H(u', v', \sigma_{u'}, \sigma_{v'}, \theta) = \frac{1}{2\pi\sigma_{u'}\sigma_{v'}} \exp \left\{ -\frac{1}{2} \left(\frac{(-u\sin\theta' + v\cos\theta')^2}{\sigma_{u'}^2} + \frac{(-u\cos\theta' - v\sin\theta')^2}{\sigma_{v'}^2} \right) \right\} \quad (3.5)$$

Along a window Υ , at a given position (m, n) , we have $u = p - m$ and $v = q - n$ with $p, q \in \Upsilon$ and the resulting filtered image is given by:

$$I_{out}(m, n) = \frac{\sum_{p,q \in \Upsilon} \exp \left(-\frac{\Phi_1 + \Phi_2}{2} \right) \exp \left(-\frac{(I_{in}(p, q) - I_{in}(m, n))^2}{2\sigma_r^2} \right) I_{in}(p, q)}{\sum_{p,q \in \Upsilon} \exp \left(-\frac{\Phi_1 + \Phi_2}{2} \right) \exp \left(-\frac{(I_{in}(p, q) - I_{in}(m, n))^2}{2\sigma_r^2} \right)} \quad (3.6)$$

with I_{out} and I_{in} are the restored and noisy pixel respectively, Φ_1 and Φ_2 are given respectively as:

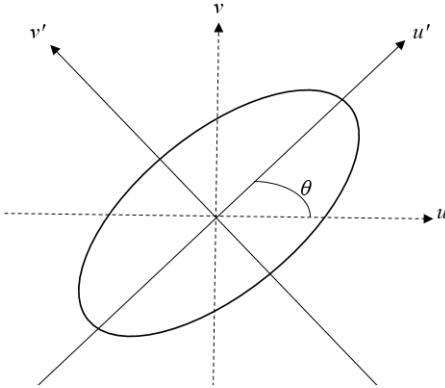
$$\Phi_1 = \frac{[-(p - m)\sin\theta' + (q - n)\cos\theta']^2}{\sigma_{u'}^2} \quad (3.7)$$

$$\Phi_2 = \frac{[-(p - m)\cos\theta' - (q - n)\sin\theta']^2}{\sigma_{v'}^2} \quad (3.8)$$

3.2.2 Noise model

By assuming the independency of the noise from the original image and within a local window, the local variance is given by:

$$\sigma_{in}^2(m, n) = \sigma_{or}^2(m, n) + \sigma_{noise}^2 \quad (3.9)$$

FIGURE 3.1: Anisotropic Gaussian with orientation θ

Where $\sigma_{in}^2(m, n)$ is the local variance of the noisy image and $\sigma_{or}^2(m, n)$ is the local variance of the original image. Within a local region, we can assume that the original image is uniform, thus its variance tends to be zero. Consequently, the local variance of the degraded image is reduced to the noise variance: $\sigma_{in}^2(m, n) \approx \sigma_{noise}^2$.

To make such approximation more concrete, we show in Fig. 3.2 the histogram of local variance of some famous images (intensity values between 0 and 255) widely used in image processing (Camera man, House, Lena, Baboon and Peppers) computed using a window of 3×3 . We notice that the shape of the histograms is exponentially decreasing for all images. Most of variance values are close to zero. We note that in case of a textured image, results will be totally different.

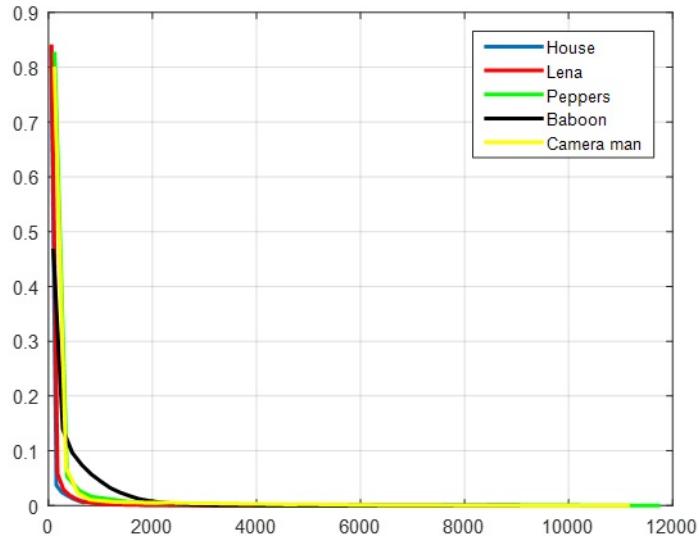


FIGURE 3.2: Normalized histogram of the variance of several famous images

3.2.3 Vector anisotropic filter and noise estimation

In this section, we extend the filter of section 3.2.1 to define a new filter called vector anisotropic Gaussian (VAG) filter adapted to the case of multispectral image and we present how to use of the sparse matrix transform for covariance matrix estimation.

3.2.3.1 Vector anisotropic filter

In order to extend the vector anisotropic Gaussian filter to the vector case, we use the multivariate Gaussian function given by:

$$G_{\Sigma}(x) = \frac{1}{(2\pi)^{P/2}|\Sigma|^{1/2}} \exp\left(-\frac{(x^T \Sigma^{-1} x)}{2}\right) \quad (3.10)$$

Where Σ is a positive definite matrix, P is the dimensionality of x .

The vector anisotropic Gaussian filter at a position (m, n) takes the form:

$$I_{out}(m, n) = \frac{\sum_{p,q \in \Upsilon} \exp\left(-\frac{\Phi_1 + \Phi_2}{2}\right) G_{\Sigma_S}(I_{in}(p, q) - I_{in}(m, n)) I_{in}(p, q)}{\sum_{p,q \in \Upsilon} \exp\left(-\frac{\Phi_1 + \Phi_2}{2}\right) G_{\Sigma_S}(I_{in}(p, q) - I_{in}(m, n))} \quad (3.11)$$

where $I_{in}(m, n)$ and $I_{in}(p, q)$ represent the pixel vectors of dimension P -the number of spectral bands- at positions (m, n) and (p, q) . Σ_S is the covariance matrix that may be set as the noise covariance matrix as suggested in [44].

Since we are operating within a sliding window Υ and by adopting the noise model seen previously, the covariance matrix is estimated. However, within Υ , the number of observations is much less than the dimensionality P . Thus, a simple computation of the covariance matrix will give misleading results. We suggest to apply the sparse matrix transform (SMT) [95] to estimate the noise covariance matrix.

3.2.3.2 Sparse matrix transform

Given a P -dimensional data sample $X = [x_1, x_2, \dots, x_m] \in R^{P \times m}$ of covariance matrix $R^{P \times P}$, the likelihood of this m samples is given by:

$$L(R, X) = \frac{|R|^{-m/2}}{(2\pi)^{Pm/2}} \exp\left(-\frac{1}{2} \text{trace}(X^T R^{-1} X)\right) \quad (3.12)$$

The covariance matrix R is decomposed as the product of an orthogonal matrix of eigenvectors E and a diagonal matrix Λ of eigenvalues such as $R = E \Lambda E^T$. The likelihood is then maximized with respect to E and Λ . The maximum likelihood (ML) estimates is

given by:

$$\hat{E} = \operatorname{argmin}_{E \in \Delta} \{ |diag(E^T SE)| \} \quad (3.13)$$

$$\hat{\Lambda} = diag(\hat{E}^T S \hat{E}) \quad (3.14)$$

Where $S = \langle xx^T \rangle = \frac{1}{m} XX^T$ is the sample covariance matrix and Δ is the set of allowed orthogonal transforms. Thus, we obtain the ML estimate of the covariance as $\hat{R} = \hat{E} \hat{\Lambda} \hat{E}^T$. To regularize the estimate of the covariance matrix, the Sparse Matrix Transform (SMT) restricts the set Δ to the Givens rotation which is the most sparse nontrivial orthogonal transform. Givens rotation is a rotation in the plane by an angle θ spanned by two coordinates axes x and y . It is given by $E = I + \Xi(x, y, \theta)$ where:

$$\Xi(x, y, \theta)_{m,n} = \begin{cases} \cos(\theta) - 1 & if \quad m = n = x \quad or \quad m = n = y \\ \sin(\theta) & if \quad m = x \quad and \quad n = y \\ -\sin(\theta) & if \quad m = y \quad and \quad n = z \\ 0 & otherwise \end{cases} \quad (3.15)$$

We denote E_k a Givens rotation. The product of orthogonal rotations $E_k E_{k-1} \dots E_1$ is also orthogonal. We denote Δ_K the set of orthogonal matrices which can be expressed as a product of K Givens rotation. The SMT is obtained by restricting the set Δ in Eq. (3.13) to Δ_K .

3.3 Optimized vector Non-Local Mean filter for multispectral image denoising: OVNLM

3.3.1 Non-Local Mean filter

The Non-Local Mean (NLM) filter is a well known edge and details preservation filter. Such characteristic has been proven in [96, 97]. Following the same noise model of eq. (3.1), the basic assumption behind the definition of the NLM filter is that we need to take advantage of the high degree of redundancy in the image: the neighborhood of a pixel s is any set of pixels p in the image domain Ω such that a local window surrounding s is similar to the local window surrounding p [49]. Let a pixel at a position (m, n) expressed as I_s where $s = (m, n) \in \Omega$. The general case of NLM filter is given by:

$$I_{out}(s) = \sum_{p \in \Omega} \omega(s, p) I_{in}(p) \quad (3.16)$$

$\omega(s, p)$ is the weight calculated for each pixel. It is computed based on a similarity measure between the pixels in position s and p . $\omega(s, p)$ satisfies the following constraints:

$$\begin{aligned} 0 &\leq \omega(s, p) \leq 1 \\ \sum_{p \in \Omega} \omega(s, p) &= 1 \end{aligned} \quad (3.17)$$

The similarity between two pixels s and p is measured as a decreasing function of the Gaussian weighted Euclidean distance $\|\cdot\|_{2,a}^2$ where $a > 0$ is the standard deviation of the Gaussian kernel. Let $N(s)$ and $N(p)$ be the the pixel vectors of the intensity gray level within a squared neighborhood centered at positions s and p respectively.

$$\omega(s, p) = \frac{1}{C_i} \exp \left(-\frac{\|N(s) - N(p)\|_{2,a}^2}{h^2} \right) \quad (3.18)$$

h^2 acts as a smoothing parameter. C_i is a normalization constant which ensures that $\sum_{p \in \Omega} \omega(s, p) = 1$.

$$C_i = \sum_{p \in \Omega} \exp \left(-\frac{\|N(s) - N(p)\|_{2,a}^2}{h^2} \right) \quad (3.19)$$

The Gaussian weighted Euclidean distance is given by:

$$\|N(s) - N(p)\|_{2,a}^2 = \sum_{k \in K} G_a(k) (N(s - k) - N(p - k))^2 \quad (3.20)$$

Where K is a local window and $G_\alpha(k)$ is defined as,

$$G_\alpha(k) = \frac{1}{2\pi a^2} \exp \left(-\frac{k_1^2 + k_2^2}{2a^2} \right), k = (k_1, k_2) \quad (3.21)$$

Thus, we can distinguish two main characteristics: the restored pixel is obtained by taking into account the contribution of pixels in the whole image and the weight computation is based on similarity between local windows. Such characteristics have triggered researchers to design various novel methods [49].

3.3.2 Vector NLM filter

To take benefit from the additional information brought by the spectral dimension, we extend the NLM filter to the vector case. In multispectral context, we dispose of the reflectance at a given position in different spectral bands. Thus we are operating on a set of pixel vectors $I = \{(I_s)/s \in \Omega\}$. We define the vector NLM (VNLM) filter as:

$$I_{out}(s) = \sum_{p \in \Omega} \omega(s, p) I_{in}(p) \quad (3.22)$$

With the new formulation of the weight between two pixels at position s and p defined as:

$$\begin{aligned} \omega(s, p) = & \\ \frac{1}{C_i} \exp\left(\frac{-1}{h^2} \sum_{k \in K} (I_{in}(s - k) - I_{in}(p - k))^T \Phi^{-1} (I_{in}(s - k) - I_{in}(p - k))\right) \end{aligned} \quad (3.23)$$

If $\Phi = I$, I is the identity matrix-, we have the classical Euclidean distance.

$$\begin{aligned} C(i) = & \\ \sum_{p \in \Omega} \exp\left(-\frac{1}{h^2} \sum_{k \in K} (I_{in}(s - k) - I_{in}(p - k))^T \Phi^{-1} (I_{in}(s - k) - I_{in}(p - k))\right) \end{aligned} \quad (3.24)$$

3.3.3 Optimization framework for vector NLM

In our framework design, we target two main objectives: optimizing the filter parameters and reducing the computation complexity. First we use both the classical Euclidean distance $\|\cdot\|_2^2$ as suggested in [98] and Mahalanobis distance $\|\cdot\|_\Phi^2$ where Φ is a covariance matrix. In addition, we preselect for each pixel a subset of the most similar pixels based on a probabilistic similarity measure.

The filter depends on two parameters: the smoothing parameter h and the covariance matrix Φ . Thus, we have:

$$I_{out}(s) = f(I_{in}(s), \Theta) \quad \text{with} \quad \Theta = (h, \Phi) \quad (3.25)$$

Where f is a non linear estimator and Θ is the filter parameter.

Our aim is to optimize the filter parameter Θ so that we can ensure the best parametrization of the filter in order to obtain the best denoising result. The performance of the estimator is generally evaluated using the mean square error (MSE):

$$MSE = \frac{1}{HL} \sum_{s \in \Omega} \|I_{out}(s) - I_{or}(s)\|^2 \quad (3.26)$$

However, the problem of such estimator is that the ground truth image $I_{or}(s)$ is unknown. MSE can be seen as a random variable of the noise. Its expected value is designated as the Risk R_θ and expressed as:

$$R_\theta = E(MSE) \quad (3.27)$$

The problem of estimating the risk without the need to dispose of the underlying image $I_{or}(s)$ is approached by Stein's Unbiased Risk Estimator (SURE) [26, 29]. Thus, we

have [28]:

$$E \left(\|I_{out}(s) - I_{or}(s)\|^2 \right) = E \left(\|I_{out}(s)\|^2 \right) - 2E \left(I_{out}(s)^T I_{or}(s) \right) + E \left(\|I_{or}(s)\|^2 \right) \quad (3.28)$$

and:

$$\begin{aligned} E \left(I_{out}(s)^T I_{or}(s) \right) &= E \left(f(I_{in}(s), \Theta)^T (I_{in}(s) - n_s) \right) \\ &= E \left(I_{out}(s)^T I_{in}(s) \right) - E \left(f(I_{in}(s), \Theta)^T n_s \right) \end{aligned} \quad (3.29)$$

If we consider a zero mean multivariate Gaussian noise, we have [26]:

$$E \left(f(I_{in}(s), \Theta)^T n_s \right) = E \left(\text{trace} \left\{ \Psi^T \nabla_{I_{in}(s)} f(I_{in}(s), \Theta) \right\} \right) \quad (3.30)$$

With Ψ is the noise covariance matrix. By combining eq. 3.28 and eq. 3.29, we end up with an expression without $I_{or}(s)$:

$$\begin{aligned} E \left(\|I_{out}(s) - I_{in}(s)\|^2 \right) &= E \left(\|I_{out}(s) - I_{in}(s)\|^2 \right) - \text{trace}(\Psi) \\ &\quad + 2E \left(\text{trace} \left\{ \Psi^T \nabla_{I_{in}(s)} f(I_{in}(s), \Theta) \right\} \right) \end{aligned} \quad (3.31)$$

Thus, the risk \hat{R}_θ is the unbiased risk estimator of MSE in eq. 3.26 and is given by:

$$\begin{aligned} \hat{R}_\theta &= \frac{1}{HL} \sum_{s \in \Omega} E \left(\|I_{out}(s) - I_{in}(s)\|^2 \right) - \text{trace}(\Psi) \\ &\quad + 2 \frac{1}{HL} \sum_{s \in \Omega} E \left(\text{trace} \left\{ \Psi^T \nabla_{I_{in}(s)} f(I_{in}(s), \Theta) \right\} \right) \end{aligned} \quad (3.32)$$

Where $\nabla_{I_{in}(s)} f(I_{in}(s)) = J_{f(I_{in}(s))}$ is the Jacobian matrix with respect to $I_{in}(s)$. $J_{f(I_{in}(s))}$ is given by [47]:

$$\begin{aligned} (J_{f(I_{in}(s))})_{i,j} &= \frac{\partial f_i(I_{in}(s), \Theta)}{\partial I_{in}(s_j)} = \\ &\frac{\sum_{p \in \Omega} \frac{\partial \chi(p)}{\partial I_{in}(s_j)} I_{in}(s_i) + \delta_{i,j}}{\sum_{p \in \Omega} \chi(p)} - \frac{\left(\sum_{p \in \Omega} \frac{\partial \chi(p)}{\partial I_{in}(s_j)} \right) \left(\sum_{p \in \Omega} \chi(p) I_{in}(s_j) \right)}{\left(\sum_{p \in \Omega} \chi(p) \right)^2} \end{aligned} \quad (3.33)$$

Where $\delta_{i,j}$ is the delta function and $\chi(p)$ is defined as:

$$\chi(p) = \exp \left(-\frac{1}{h^2} \sum_{k \in K} (I_{in}(s-k) - I_{in}(p-k))^T \Phi^{-1} (I_{in}(s-k) - I_{in}(p-k)) \right) \quad (3.34)$$

With the derivation of $\chi(p)$, we obtain:

$$(J_{f(I_{in}(s))})_{i,j} = \frac{\sum_{p \in \Omega} \chi(p) \left((I_{in}(p) - I_{in}(s))^T \frac{1}{2h^2} (\Phi^{-1} + \Phi^{-1T}) \right)_j^T I_{in}(s_i) + \delta_{i,j}}{\sum_{p \in \Omega} \chi(p)} - \frac{\left(\sum_{p \in \Omega} \chi(p) \left((I_{in}(p) - I_{in}(s))^T \frac{1}{2h^2} (\Phi^{-1} + \Phi^{-1T}) \right)_j^T \left(\sum_{p \in \Omega} \chi(p) I_{in}(s_i) \right) \right)}{\left(\sum_{p \in \Omega} \chi(p) \right)^2} \quad (3.35)$$

Demonstration:

$$\begin{aligned} \chi(p) &= \exp\left(-\frac{1}{h^2} \sum_{k \in K} (y_{s-k} - y_{p-k})^T \Phi^{-1} (y_{s-k} - y_{p-k})\right) \\ &= \exp\left(-\frac{1}{h^2} (y_s - y_p)^T \Phi^{-1} (y_s - y_p)\right) \cdot \\ &\quad \exp\left(-\frac{1}{h^2} \sum_{\substack{k \in K \\ k \neq 0}} (y_{s-k} - y_{p-k})^T \Phi^{-1} (y_{s-k} - y_{p-k})\right) \\ &\implies \frac{\partial \chi(p)}{\partial y_{s_j}} = \exp\left(-\frac{1}{h^2} \sum_{\substack{k \in K \\ k \neq 0}} (y_{s-k} - y_{p-k})^T \Phi^{-1} (y_{s-k} - y_{p-k})\right) \cdot \\ &\quad \frac{\partial}{\partial y_{s_j}} \left(\exp\left(-\frac{1}{h^2} (y_s - y_p)^T \Phi^{-1} (y_s - y_p)\right) \right) \\ &\quad \frac{\partial \chi(p)}{\partial y_{s_j}} = \chi(p) \left((y_p - y_s) \frac{1}{h^2} (\Phi^{-1} + \Phi^{-1T}) \right)_j^T \end{aligned} \quad (3.36)$$

Finally, we formulate the problem of vector NLM filter as a constrained optimization problem:

$$\begin{cases} (h_{opt}, \Phi_{opt}) = \arg \min_{h, \Phi} (\hat{R}(h, \Phi)) \\ s.t. : h > 0, \Phi \geq 0 \end{cases} \quad (3.37)$$

Note that in case of using the Euclidean distance, the only parameter to be optimized is h .

3.3.4 Relevant pixel selection

Back to eq. 3.22, we can clearly see that in order to restore every pixel, we need to go through every other pixel in the domain Ω . This is obviously a time consuming process. To cope with the computation burden of the proposed OVNLM, we propose to preselect for each processed pixel, a subset of the most relevant pixel based on a

similarity measure proposed in [99]. It is based on a probabilistic approach to measure the similarity between two pixels based on the noise distribution. In grayscale case, the similarity measure is defined as:

$$S(x_s, x_p) = \frac{1}{4\sigma|\Omega|\sqrt{\pi}} \cdot \exp\left(-\frac{(x_s - x_p)^2}{4\sigma^2}\right) (\operatorname{erf}\left(\frac{2x_{s0} - x_s - x_p}{2\sigma}\right) + \operatorname{erf}\left(\frac{x_s + x_p}{2\sigma}\right)) \quad (3.38)$$

Where x_{s0} is the maximum value of the true intensity, $|\Omega|$ is a constant and $\operatorname{erf}(\cdot)$ is the error function defined as

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \quad (3.39)$$

In case of RGB color images, the similarity between two pixels $I_s = [r_s, g_s, b_s]$ and $I_p = [r_p, g_p, b_p]$ is defined as:

$$S(I_s, I_p) = S(r_s, r_p) \cdot S(g_s, g_p) \cdot S(b_s, b_p) \quad (3.40)$$

We generalize this similarity measure for the multispectral case. The similarity measure between $I(s) = (I(s_i)_{i=1 \dots P})$ and $I(p) = (I(p_i)_{i=1 \dots P})$ is defined as:

$$S(I_s, I_p) = \prod_{i=1}^P S(I(s_i), I(p_i)) \quad (3.41)$$

Thus, the proposed NLM filter becomes:

$$I_{out}(s) = \sum_{\substack{p \in \Omega \\ S(I_{in}(s), I_{in}(p)) \neq 0}} \omega(s, p) I_{in}(p) \quad (3.42)$$

3.3.5 VOLNM algorithm

The proposed framework is detailed below. We solve the constrained non-linear optimization problem using Sequential Quadratic Programming. Giving a noisy image and noise covariance matrix which can be estimated with the median absolute deviation method [100], we minimize the risk value based on an optimal choice of parameters $\Theta = (h, \Phi)$ until we reach the maximum number of iteration iter_max or the decreasing of the risk value is less than a preset threshold ξ . We implement this framework with MATLAB and the minimization is carried out using $fmincon$ function.

Input ($I_{in}(s)_{s \in \Omega}$), Ψ
Output Optimal ($I_{out}(s)_{s \in \Omega}$) with minimal $\hat{R}_{h,\Phi}$

1- Initialize Ψ , h , Φ , iter=0, maximum iteration number $iter_max$ and stopping threshold ξ

2- Iteration: **do**

a- Calculate ($I_{out}(s)_{s \in \Omega}$) using Eq. (3.42) and

b- Calculate R_{iter} using (3.32)

c- $iter = iter + 1$

d- Update h with SQP

e- Update Φ with SQP

f- Compute $R_{iter+1} = \hat{R}(h_{iter+1}, \Phi_{iter+1})$

While ($iter \prec iter_max$ or $R_{iter} - R_{iter+1} \preceq \xi$)

3.4 Conclusion

Two denoising schemes: vector anisotropic Gaussian filter (VAG) and the Optimized Vector Non-Local Mean filter (OVNLM) have been designed for the multispectral case. The parameters of the VAG filter are chosen heuristically. The parametrization of the VONLM filter conducted using an optimization framework. So, while the VAG filter is simple and does not show high complexity, OVNLM is more complex and requires more computation power due to two factors: the NLM algorithm is by nature time consuming adding that to the optimization process that requires also more computation. We have applied the two denoising procedures on simple color images as well as face multispectral images. Obtained results are presented and discussed in Chapter 5 and comparison with results of some of the algorithms presented in Chapter 2 is also provided.

Chapter 4

Multispectral image segmentation

4.1 Introduction

Image segmentation is a broad term that covers a wide variety of techniques. A basic view of segmentation is that we intend to find in data structures that share similar characteristics and are different from other structures. This problem is called also clustering. This process can be roughly conducted using two main strategies: hierarchical and partitional. Hierarchical clustering yields in general to a hierarchy of clusters. This approach falls in two types: agglomerative (bottom-up) and divisive (top-down). In other words, either we consider at the beginning that each data point is a cluster or that all data points are one cluster, then we merge or split. On the other hand, partitional clustering seeks to divide data into a given number c of clusters C_1, C_2, \dots, C_c by minimizing an objective function. The best known partitional clustering is the K-means algorithm [60]. K-means minimizes the sum of squared distances between features and cluster centers. The cost function of K-means is:

$$J = \sum_{i=1}^c \sum_{x_j \in C_i} \|x_j - v_i\|^2 \quad (4.1)$$

A fuzzy version of K-means, called the Fuzzy C-Means (FCM), was proposed by Bezdek [79]. The fuzzy objective function J_{FCM} is given by eq. 2.1. Various algorithms have been derived from K-means and FCM. In particular, weighting methods are widely discussed [101]. They are based on the idea that attributes in the objective functions e.g. data points or features do not have the same contribution and influence on the clustering process. Thus, a weight is introduced in the objective function targeting data points and/or features. Weighting methods are around prototypes [102] where information about the cluster size is incorporated, features where each feature has a specific weight

[103–105], fuzzy membership degree [106] and distance [107]. Various partitional algorithms require a predefined number of clusters. Cluster validity indexes (CVIs) are used to look for the optimal number of clusters that best fits the data, and to discriminate between clustering algorithms.

We propose in this chapter a new clustering algorithm based on the weighting approach and the gravitation theory. In addition to that, two new CVIs are also proposed. Experiments have been conducted (see Chapter 5) on complex data sets to prove the outperformance of the proposed methods.

4.2 Gravitational Fuzzy C-Means

4.2.1 Gravitation theory and interaction between particles

The proposed algorithm is based on gravitation theories. We assimilate the data set to be clustered as a set of particles and each one has a specific mass. The optimization of the clustering process is based on minimizing the squared distance between centroids and data points which means that we are looking for centers and data points to be as close as possible to each other. The idea here is to take into account the interaction between centroids and data points. In physics, objects of given masses interact between each others. This interaction comprises an attraction force and an escape velocity. This model has been introduced in cluster analysis particularly in hierarchical clustering and is known as gravitational clustering [108]. Indeed, initially, each data point is seen as a cluster then data points are merged according to their interactions leading to the creation of new data points. These new particles have a mass equal to the summation of masses of the merged particles. Iteratively, a dendrogram is formed and clusters are obtained hierarchically [109, 110].

Our approach in this is to include this interaction into the partitional clustering procedure by weighting data points accordingly.

4.2.2 Attraction force

According to Newton’s law of universal gravitation, two objects of masses m_1 and m_2 respectively, exert a gravitational attraction force on each other. This force is proportional to the product of masses of the objects and inversely proportional to the squared distance between them. Fig. 4.1 shows the Newton’s law for gravitational attraction force. Two objects with masses m_1 and m_2 separated with distance r exert attraction

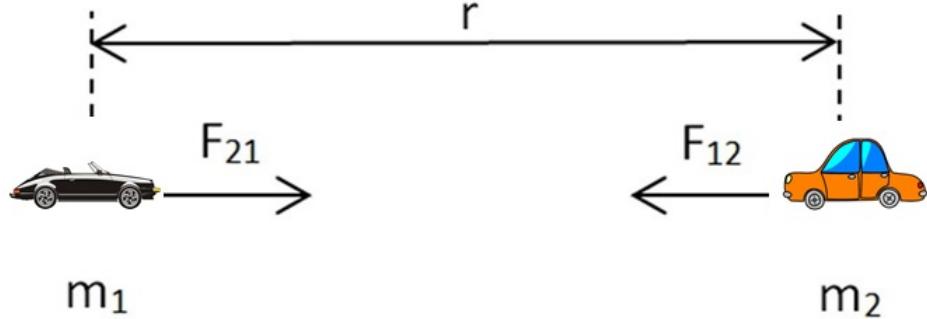


FIGURE 4.1: Illustration of Newton's law of attraction forces between two objects

forces $|F_{12}| = |F_{21}|$ on each other:

$$|F_{12}| = |F_{21}| = G \frac{m_1 m_2}{r^2} \quad (4.2)$$

where G is the universal gravitational constant.

A Markovian model of gravitational attraction between two objects is proposed in [111] where pixels are considered as particles of mass equal to 1 that attract each other:

$$|F_{12}| = |F_{21}| = G \frac{m_1 m_2}{|Z_1 - Z_2|} (Z_1 - Z_2) \quad (4.3)$$

where Z_1 and Z_2 are the location vectors of object 1 and 2 respectively.

In our work, we use the first expression of the gravitational attraction force.

4.2.3 Escape velocity

Escape velocity V_e is the speed with which an object needs to be traveling to break free from a planet and leave it. For example, a spacecraft needs to be going 11.2 km per second to leave the earth without falling back. Escape velocity depends on the mass of the planet the spacecraft is intending to escape from:

$$V_e = \sqrt{\frac{2GM}{r}} \quad (4.4)$$

where G is the gravitational constant, M is the mass of the object to be escaping from and r is the Euclidean distance between the object and the escape point.

4.3 Gravitational Weighted Fuzzy C-Means (GWFCM) for multispectral image segmentation

The interaction between data points and cluster centers is taken into account and incorporated in the objective function as a weight. The attraction $attr_{ji}$ between the j^{th} data point and the i^{th} cluster center is computed using equation 4.2. Then, attraction normalization is conducted:

$$Attr_{ji} = \frac{attr_{ji}}{\sum_{i=1}^c attr_{ji}} \quad (4.5)$$

The escape esc_{ji} of the j^{th} data point from the i^{th} cluster center is computed using equation 4.4 then normalized as:

$$Esc_{ji} = \frac{esc_{ji}}{\sum_{i=1}^c esc_{ji}} \quad (4.6)$$

The proposed weight is the ratio of the degree of attraction to the degree of escape:

$$\omega_{ji} = \frac{Attr_{ji}}{Esc_{ji}} \quad (4.7)$$

By incorporating the proposed weight, the objective function becomes:

$$J_{GWFCM} = \sum_{i=1}^c \sum_{j=1}^N \omega_{ji} \mu_{ij}^m d_{ij}^2 = \sum_{i=1}^c \sum_{j=1}^N \omega_{ji} \mu_{ij}^m \|X_j - v_i\|^2 \quad (4.8)$$

Using the Lagrangian multiplier to solve the above equation, we obtain the following updates:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{\|X_j - v_i\|^2}{\|X_j - v_k\|^2} \right)^{\frac{1}{m-1}}} \quad (4.9)$$

$$v_i = \frac{\sum_{j=1}^N \omega_{ji} \mu_{ij}^m X_j}{\sum_{j=1}^N \omega_{ji} \mu_{ij}^m} \quad (4.10)$$

With FCM, the centers of clusters are functions of the membership degree μ_{ij} , m and X . However, for GWFCM, besides μ_{ij} m and X , a set of weights ω_{ji} are incorporated into the updating iteration and together they affect the final results as illustrated below [112]:

1- Initialize $U = [\mu_{ij}] = U^{(0)}$, stopping criterione

2- At iteration k:

a- Calculate ω_{ji}

$$\text{b- Calculate } v_i^{(k)} = \frac{\sum_{j=1}^N \omega_{ji} \mu_{ij}^m X_j}{\sum_{j=1}^N \omega_{ji} \mu_{ij}^m}$$

$$\text{c- Update } U^{(k+1)} \quad \mu_{ij}^{(k)} = \frac{1}{\sum_{k=1}^c \left(\frac{\|X_j - v_i\|^2}{\|X_j - v_k\|^2} \right)^{\frac{1}{m-1}}}$$

d- If $\|U^{(k+1)} - U^{(k)}\| \leq \epsilon$ then STOP, else go to step 2

4.4 Cluster validity index

We propose two new cluster validity indexes, one based on the geometrical form and mathematical formulation of n-spheres, and the second is based on Jeffrey divergence.

4.4.1 Cluster validity index based on n-sphere

We propose a novel cluster validity index computation algorithm based on the n-sphere shape. Classical measures of separation and compactness perform well only with a specific characteristics of the data set such as absence of overlap or absence of noise. The proposed CVI is a summation type in which compactness and overlap of the clusters are taken into account.

4.4.1.1 n-sphere

We base our design of the proposed validity index on the geometric shape of the n-sphere, also called the hypersphere. An n-sphere is a generalization of the ordinary sphere in n-dimensional space. It refers to the set of points whose distance from a central point is equal to r [113].

$$S^n = \{x \in R^{n+1} \mid \|x\| = r\} \quad (4.11)$$

Therefore, it is an n-dimensional manifold in the $(n+1)$ dimensional space. Particularly, if we set $n = 1$, we obtain the 1-sphere which is the circle of radius r . For $n = 2$, we have the ordinary 2-sphere which is the 3D sphere. The volume of the n-sphere is given by:

$$V_n = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2} + 1)} r^n \quad (4.12)$$

Where Γ is the gamma function given by:

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx \quad (4.13)$$

4.4.1.2 Compactness measure

After applying a clustering algorithm, one will see each cluster as a hypersphere. The proposed compactness measure is defined as the volume encompassed between two n-spheres. These two spheres are centered around the center of the cluster. Let x_c be the closest point to the center v of the considered cluster. The volume of the hypersphere of radius $r_c = \|x_c - v\|$ is V_c .

In order to determine the second hypersphere, we need to sort the distance between cluster center and the points belonging to the cluster. Let n be the number of points inside the cluster and T be a parameter. We compute the distances $d_i = \|v - x_i\|, i = 1, \dots, n$. We sort now the obtained distances and we have now $d_j \geq d_{j-1} \geq \dots \geq d_1$. The radius of the second hypersphere is defined as the distance between the cluster center v and the point p_k for which $d_k - d_{k-1} > T, k = 2 \dots j$. Let x_f be this particular point and V_f the volume of the hypersphere of radius $r_f = \|x_f - v\|$.

We could define x_f as the farthest point to v but in the case of an outlier or a noisy point, this could cause misleading results. So a control over the amount of progress of the distance to v will prevent from being in this particular case.

The value of parameter T is chosen as follows: We use the outlier measure as defined in MATLAB R14a¹. T is defined as the ratio of the number of points which have an outlier measure superior than the mean of all outlier measures over the number of points. The outlier measure of an observation is computed by taking an inverse of the average squared proximity between this observation and other observations. Then normalization for all the outlier measures is applied by subtracting the median of their distribution, taking the absolute value of this difference, and dividing by the median absolute deviation. A high value of the outlier measure indicates that this observation is an outlier.

The proposed compactness measure of the i^{th} cluster C_i is defined as:

$$Comp_i = V_f - V_c = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2} + 1)} (r_f^n - r_c^n) \quad (4.14)$$

¹<http://www.mathworks.com/help/stats/compacttreebagger.outliermeasure.html>

The total compactness measure of a given partition composed of k clusters is defined as:

$$Comp = \sum_{i=1}^c Comp_i \quad (4.15)$$

4.4.1.3 Overlap measure

As we consider each cluster as a hypersphere, the overlap measure is defined as the volume of intersection between hyperspheres. In this case, we define d as the distance between two cluster centers. Let r_i and r_j be the radii of the two hyperspheres associated with each cluster and $OVR(i, j)$ the measure of overlap between cluster C_i and C_j . We have three cases:

- if $d \geq r_i + r_j$: There is no overlap and clusters are completely separated.

$$OVR(i, j) = 0 \quad (4.16)$$

- if $d \leq |r_i - r_j|$: One hypersphere is inside the other and the volume of the hypersphere of the smaller radius is the total amount of overlap between clusters.

$$OVR(i, j) = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2} + 1)} (min(r_i, r_j))^n \quad (4.17)$$

- Otherwise, the volume consists of two hyperspherical caps. The heights of the caps' bases are the signed distances computed as:

$$c_i = \frac{d^2 + r_i^2 - r_j^2}{2d} \quad (4.18)$$

$$c_j = \frac{d^2 - r_i^2 + r_j^2}{2d} \quad (4.19)$$

The volume of a hyperspherical cap of a signed height a in a hypersphere of radius r is [114]:

$$V_{r,a>0}^{cap} = \frac{1}{2} \frac{\pi^{n/2}}{\Gamma(\frac{n}{2} + 1)} r^n I_{1-\frac{a^2}{r^2}}\left(\frac{n+1}{2}, \frac{1}{2}\right) \quad (4.20)$$

$$V_{r,a<0}^{cap} = \frac{1}{2} \frac{\pi^{n/2}}{\Gamma(\frac{n}{2} + 1)} r^n - V_{r,-a>0}^{cap} \quad (4.21)$$

Where I is the regularized incomplete beta function defined as:

$$I_x(a, b) = \frac{B(x, a, b)}{B(a, b)} \quad (4.22)$$

With

$$B(x, a, b) = \int_0^x t^{a-1} (1-t)^{b-1} dt \quad (4.23)$$

and

$$B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt \quad (4.24)$$

Then, the volume of intersection of two hyperspheres is:

$$OVR(i, j) = V = V_{r_i, c_i}^{cap} + V_{r_j, c_j}^{cap} \quad (4.25)$$

The overlap measure is computed as follows:

$$OVR = \sum_{i=1}^c \sum_{j=1; j \neq i}^c OVR(i, j) \quad (4.26)$$

4.4.1.4 validity index S

The proposed validity index is defined as:

$$S = \frac{1}{c} (OVR + Comp) \quad (4.27)$$

Index S represents the average sum of the compactness and overlap measures. The best partition is the one with less compactness and overlap, thus it is highlighted by the minimum of index S .

4.4.2 Cluster validity index based on Jeffrey divergence

We propose a new CVI based on a new separation measure. The classical separation measure which uses distances between features and cluster centers does not reflect the real separation especially when it comes to clusters with different sizes and densities. One of the possible solutions is to design a distance metric that is adapted to each data point as proposed in [115] to ensure a significant separation measure. This unsupervised distance metric learning approach aims at incorporating the maximum of discriminative information to design a metric that is able to regroup similar data samples in one class and dissimilar samples in different classes. The learning process can be global or local. With a global learning, distance between pairs is minimized according to the equivalence constraints. Separation of the data pairs is conducted using the inequivalence constraints. However, in case of classes that exhibit multimodal distributions, a conflict between theses equivalence and inequivalence constraints may occur. In [116], authors proposed a local discriminative distance metrics algorithm that is not only capable of

dealing with the problem of the global approach but also incorporate multiple distance metrics unlike the previous work [117–119] where a single distance metric is learned on the whole data sets. Our approach is based on considering each cluster as a density to be estimated. Thus we suggest a new separation measure between clusters based on Jeffrey divergence [120] between clusters.

4.4.2.1 Motivation

The classical separation measure is based on computing distances between cluster centers. However, such measure has drawbacks. In Fig. 4.2, three clusters are generated from Gaussian distributions. Cluster A and B are well separated. The distance between their centers is $d_{AB} = 14.5$. Meanwhile, an overlap between cluster A and C is present but the distance between their centers is $d_{AC} = 13.2$. This case demonstrates the shortcoming of the separation measure based on distances between cluster centers.

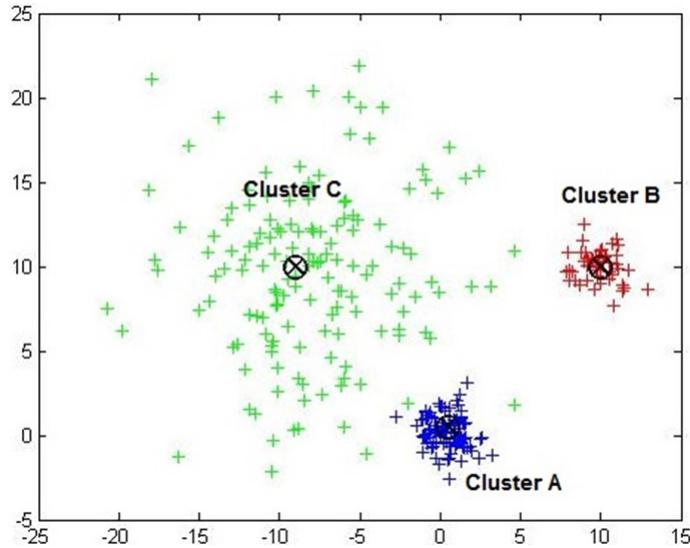


FIGURE 4.2: Three clusters generated from Gaussian distributions. Cluster B and C are equally distant to cluster A but cluster C overlaps more with cluster A

4.4.2.2 Cluster validity index based on Jeffrey divergence

The proposed validity index uses a separation measure based on the computation of Jeffrey divergence. The Jeffrey divergence (JD) between two distributions P and Q is given by:

$$JD(P, Q) = \sum_x (P(x) - Q(x)) \ln \left(\frac{P(x)}{Q(x)} \right) \quad (4.28)$$

JD is computed as the product of two terms. $(P(x) - Q(x))$ is proportional to the distance between two probability densities. $\ln(\frac{P(x)}{Q(x)})$ which is also proportional to the level

of separation between probability densities. The closer $P(x)$ and $Q(x)$ to each other, the higher the overlap is and the lesser the value of Jeffrey divergence is. Divergence allows to evaluate the extent to which two Probability Density Functions (PDF) differentiate. Jeffrey divergence, unlike the Kullback-Leiber, is symmetric which reduces computational time in addition to being widely used in pattern recognition and computer vision applications [121, 122]. It is also numerically stable and robust with respect to noise and the size of the bins [123]. In our experiments, after applying a clustering algorithm, the density of each cluster is computed using PDF estimation techniques, then JD is computed between pairs of clusters. Finally, the separation measure is determined.

- ***Cluster density estimation***

Let x_1, x_2, \dots, x_{N_i} be the N_i features of the i^{th} cluster C_i .

We assume first that data are generated from multivariate Gaussian distribution given by:

$$p(x|\mu, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right) \quad (4.29)$$

Where μ and Σ are the mean vector and covariance matrix of the data respectively. T is the transpose operator and $| |$ is the determinant operator. In order to determine the PDF, we need to estimate parameters (μ, Σ) . We apply a maximum likelihood estimation [124] and we obtain the following estimations:

$$\hat{\mu} = \frac{1}{N_i} \sum_{i=1}^{N_i} x_i \quad (4.30)$$

$$\hat{\Sigma} = \sum_{i=1}^{N_i} (x_i - \hat{\mu})^T (x_i - \hat{\mu}) \quad (4.31)$$

Demonstration:

In all the expressions below, x is a vector of random variables whose mean vector and covariance matrix are given by: $E(x) = \mu$ and $E((x-\mu)(x-\mu)^T) = \Sigma$ where E means the expectation. Using matrix properties:

- $\frac{\partial A^T \cdot x}{\partial x} = \frac{\partial x^T \cdot A}{\partial x} = A$
- $\frac{\partial x^T \cdot A \cdot x}{\partial x} = A^T + A$
- $\frac{\partial}{\partial A} \log|A| = (A^{-1})^T$
- $\frac{\partial}{\partial A} \text{tr}[AB] = \text{tr}[BA] = B^T$

The log likelihood of the multivariate Gaussian distribution is given by:

$$L(x|\mu, \Sigma) = \frac{-nD}{2} \log(2\pi) - \frac{n}{2} \log(|\Sigma|) - \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) \quad (4.32)$$

The estimates of the mean and covariance matrix are determined by computing the derivatives of $L(x|\mu, \Sigma)$ with relative to μ and Σ and setting it equal to zero.

$$\begin{aligned} \frac{\partial L(x|\mu, \Sigma)}{\partial \mu} &= \frac{\partial}{\partial \mu} \left(\sum_{i=1}^n (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) \right) \\ &= \frac{\partial}{\partial \mu} \left(\sum_{i=1}^n (x_i^T \Sigma^{-1} x_i - \mu^T \Sigma^{-1} x_i - x_i^T \Sigma^{-1} \mu + \mu^T \Sigma^{-1} \mu) \right) \\ &= \sum_{i=1}^n \left(\Sigma^{-1} x_i + (\Sigma^{-1})^T x_i \right) - N \left(\Sigma^{-1} + (\Sigma^{-1})^T \right) \mu \end{aligned} \quad (4.33)$$

$$= 0$$

$$\Rightarrow \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\begin{aligned} \frac{\partial L(x|\mu, \Sigma)}{\partial \Sigma^{-1}} &= \frac{\partial}{\partial \Sigma^{-1}} \left(-\frac{N}{2} \log(|\Sigma|) - \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^T \Sigma^{-1} (x_i - \mu) \right) \\ &\propto \frac{\partial}{\partial \Sigma^{-1}} \left(-\frac{N}{2} \log(|\Sigma|) - \frac{1}{2} \sum_{i=1}^n \text{tr}[\Sigma^{-1} (x_i - \mu) \cdot (x_i - \mu)^T] \right) \end{aligned} \quad (4.34)$$

$$= \frac{\partial}{\partial \Sigma^{-1}} \left(\frac{N}{2} \log(|\Sigma^{-1}|) - \frac{1}{2} \text{tr}[\sum_{i=1}^n \Sigma^{-1} (x_i - \mu) \cdot (x_i - \mu)^T] \right)$$

$$= 0$$

$$\Rightarrow \hat{\Sigma} = \sum_{i=1}^n (x_i - \hat{\mu})(x_i - \hat{\mu})^T$$

The Jeffrey divergence between two clusters C_i and C_j generated from Gaussian distributions $N(\mu_i, \Sigma_i)$ and $N(\mu_j, \Sigma_j)$ is:

$$JD(C_i, C_j) = \frac{1}{2} \left(\text{trace}(\Sigma_i^{-1} \Sigma_j) + \text{tr}(\Sigma_j^{-1} \Sigma_i) \right) + \frac{1}{2} \left((\mu_i - \mu_j)^T (\Sigma_i^{-1} + \Sigma_j^{-1}) (\mu_i - \mu_j) \right) - d \quad (4.35)$$

Demonstration:

- $E(x^T \cdot A \cdot x) = \text{tr}(A \cdot \Sigma) + \mu^T \cdot A \cdot \mu$
- $\langle \cdot \rangle$ is the expectation symbol.

$$p(x|\mu_1, \Sigma_1) = \frac{1}{(2\pi)^{d/2} |\Sigma_1|^{1/2}} \exp \left(-\frac{1}{2} (x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1) \right) \quad (4.36)$$

$$q(x|\mu_2, \Sigma_2) = \frac{1}{(2\pi)^{d/2} |\Sigma_2|^{1/2}} \exp \left(-\frac{1}{2} (x - \mu_2)^T \Sigma_2^{-1} (x - \mu_2) \right) \quad (4.37)$$

$$JD(p, q) = KL(p/q) + KL(q/p) \quad (4.38)$$

Where KL is the Kullback-Leiber divergence.

$$\begin{aligned} KL(p/q) &= \int \log(p(x) - q(x)) p(x) \\ &= \int \left(\frac{1}{2} \log \left(\frac{|\Sigma_2|}{|\Sigma_1|} \right) - \frac{1}{2} (x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1) \right. \\ &\quad \left. + \frac{1}{2} (x - \mu_2)^T \Sigma_2^{-1} (x - \mu_2) \right) p(x) \\ &= \frac{1}{2} \log \left(\frac{|\Sigma_2|}{|\Sigma_1|} \right) - \frac{1}{2} \langle (x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1) \rangle \\ &\quad + \frac{1}{2} \langle (x - \mu_2)^T \Sigma_2^{-1} (x - \mu_2) \rangle \quad (4.39) \\ &= \frac{1}{2} \left(\log \left(\frac{|\Sigma_2|}{|\Sigma_1|} \right) - \text{tr}(\Sigma_1^{-1} \Sigma_1) + \text{tr}(\Sigma_2^{-1} \Sigma_1) \right) \\ &\quad + \frac{1}{2} ((\mu_1 - \mu_2)^T \Sigma_2^{-1} (\mu_1 - \mu_2)) \\ &= \frac{1}{2} \left(\log \left(\frac{|\Sigma_2|}{|\Sigma_1|} \right) - d + \text{tr}(\Sigma_2^{-1} \Sigma_1) \right) \\ &\quad + \frac{1}{2} ((\mu_1 - \mu_2)^T \Sigma_2^{-1} (\mu_1 - \mu_2)) \end{aligned}$$

Thus :

$$\begin{aligned} JD(p, q) &= \frac{1}{2} (tr(\Sigma_1^{-1}\Sigma_2) + tr(\Sigma_2^{-1}\Sigma_1)) + \\ &\quad \frac{1}{2} ((\mu_1 - \mu_2)^T(\Sigma_1^{-1} + \Sigma_2^{-1})(\mu_1 - \mu_2)) - d \end{aligned} \quad (4.40)$$

To generalize this formulation to the non-Gaussian distribution case, let us assume that data are generated from any arbitrary distribution. To estimate the PDF, we rely on the multivariate density distribution using the Gaussian kernel and define the density function as:

$$\hat{p}_H(x) = \frac{1}{N_i} \sum_{i=1}^{N_i} K_H(x - x_i) \quad (4.41)$$

Where H is the bandwidth and K is a kernel function:

$$K_H(x) = |H|^{-1/2} K(H^{-1/2}x) \quad (4.42)$$

We use a Gaussian kernel for the density estimation:

$$K(x) = (2\pi)^{d/2} \exp(-\frac{1}{2}x^T x) \quad (4.43)$$

- ***Separation measure***

Based on the computation of JD, we set up the separation measure as follows: after applying a clustering algorithm, we estimate the density of each cluster C_i . Divergence between C_i and $C_{j,j=1,\dots,k,j \neq i}$ is computed. After that, we take the minimum of calculated divergences. The separation measure is:

$$S = \sum_{i=1}^k Sep_i \quad (4.44)$$

With:

$$Sep_i = \min_{j=1,\dots,k,j \neq i} (JD(C_i, C_j)) \quad (4.45)$$

It is the sum of the minimum of overlap, e.g. Jeffrey divergence, between each cluster and other clusters. By choosing the minimum value, we are taking the least overlap degree for each cluster which is an indicator of its separability. A high separation value is an indicator of a good partition.

- ***Compactness measure***

The compactness measure is computed as the summation for all clusters of the squared maximal distance of a feature x_j belonging to the i^{th} cluster C_i to its

center v_i . We denote the compactness measure V .

$$V = \sum_{i=1}^k \max_{x_j \in C_i} \|x_j - v_i\|^2 \quad (4.46)$$

A low compactness measure indicates a good partition.

- ***Proposed cluster validity index: I***

The proposed cluster validity index I is the ratio of the proposed separation measure to the compactness measure [125].

$$I = \frac{V}{S} \quad (4.47)$$

A good partition is characterized by a high separation value and a low compactness. Thus, a low value of I is an indicator of a good partition.

4.5 Conclusion

In this chapter, we have proposed new clustering algorithm and two CVIs based on different concepts. To emphasize the performance of the proposed methods, we conducted experiments on multispectral images and some machine learning data sets from different databases and with different complexities: simple data sets where clusters are easy to identify and complex data sets where clusters are completely indistinguishable. Our focus on cluster analysis for multispectral image segmentation is justified by the variety of challenges and problems that can be investigated in addition to the complexity of handling the multispectral data. We note that different approaches for multispectral image segmentation can be subject for further studies such as active contours and graph cuts.

Chapter 5

Experimental results

5.1 Introduction

We conducted several experiments for multispectral image segmentation and denoising. The segmentation experiments are conducted as follows: we applied the proposed gravitational FCM in addition to several other algorithms and compared their performances using different CVIs. Furthermore, we conducted a clustering task and we compared performances of the algorithms. The proposed CVIs are used to determine the real number of clusters of several data sets and we compared results with several other CVIs. For the denoising part, we contaminated images with different levels of Gaussian noise and applied the proposed denoising algorithms in addition to other denoising methods. We calculated the Peak Signal-to-Noise Ratio and compared the obtained results.

In our experiments, we used the IRIS M^3 multispectral face images database in addition to HYDICE remote sensing database. Famous color images are also used e.g. Lena and Baboon images. For the CVIs comparison, we used various machine learning databases e.g. UCI repository with different complexity to ensure good performance analysis. We end this section by investigating the impact of multispectral image denoising on the segmentation results.

5.2 Data sets

This section presents the image and machine learning databases used in our experiments to test the proposed algorithms and metrics, and compare their performances with the state of the arts tools.

5.2.1 IRIS M^3 multispectral images database

This database is constructed by the IRIS Lab from University of Tennessee [126]. The database was collected between August 2005 and March 2006. It consists of 2624 multispectral face images taken along the visible spectrum in addition to thermal images with a resolution of 640×480 . RGB images are also generated with a resolution of 2272×1704 . These images are taken in different lightening conditions: Halogen light, daylight and fluorescent. The total size of the database is 8.91GB. A total of 82 participants were involved from different genders (76% male, 24% female), ethnicities as depicted in Table 5.1, ages, facial expression, genders and hair characteristics. Samples from this database taken in halogen and daylight illuminations are illustrated in Fig. 5.1

TABLE 5.1: Ethnicity percentage in IRIS M^3 database

	Caucasian	Asian	Asian Indian	African descent
%	57%	23%	12%	8%

5.2.2 HYDICE multispectral images database

This database has been collected by HYperspectral Digital Imagery Collection Experiment (HYDICE) sensor, which has been in use since 1995. It is used to collect hyperspectral imagery of natural scene and man-made object with a spatial and spectral resolution of one meter and 10 nm respectively. which results in 210 channels from 400nm to 2500nm. HYDICE airborne sensor has been used by several organizations to demonstrate the usefulness of the hyperspectral imagery. Figure 5.2 illustrates samples from this database.

5.2.3 Machine learning databases

- **UCI repository:** The UC Irvine Machine Learning repository is hosted by the Center for Machine Learning and Intelligent Systems at UC Irvine. It consists of databases, domain theories and data generators that are widely used by the machine learning community. The 189 data sets are not categorized but listed according to their popularity. Among these data sets, we can find the famous iris data set from Fisher, wine data set, the Wisconsin Breast Cancer (diagnostic) data set



FIGURE 5.1: Sample of four subjects from IRIS M^3 database: left: 530nm, right: 650nm, top rows: Daylight, bottom rows: Halogen light)

- ***Speech and Image Processing Unit synthetic database***: This database is hosted by the University of Eastern Finland. We used synthetic 2D data sets of 5000 points generated from a Gaussian distribution resulting in 15 clusters, these are characterized by their different degrees of overlap between clusters which makes them quite interesting for CVIs comparison [127]. In addition, we used a special shaped synthetic data set [128].
- ***The extended Yale Face Database B*** [129]: This database is composed of 16128 images of 28 human subjects. Images are acquired under 64 illumination conditions with 9 poses.
- ***Hopkins 155***¹: This data set consists of 156 sequences of two and three motions

¹<http://www.vision.jhu.edu/data/hopkins155/>

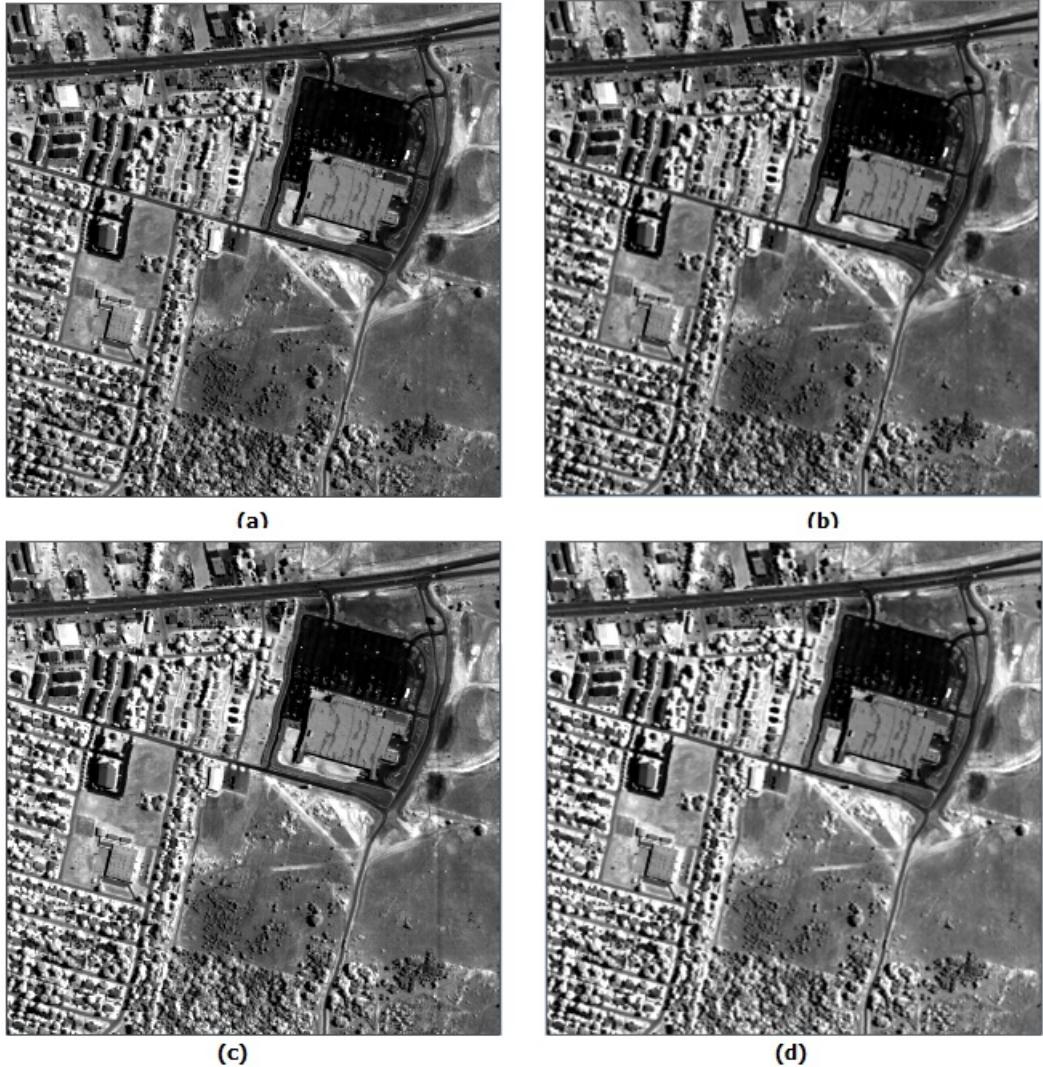


FIGURE 5.2: Sample image from HYDICE: (a) 750nm, (b) 761nm (c) 772nm, (d) 859nm

which can be divided into three categories: checkerboard, traffic and articulated sequences. Each sequence is a segmentation task. The checkerboard category consists of 104 sequences of indoor scenes taken with a handheld camera under controlled conditions. The checkerboard pattern on the objects is used to assure a large number of tracked points. Traffic category consists of 38 sequences of outdoor traffic scenes taken by a moving handheld camera. Articulated sequences display motions constrained by joints, head and face motions, people walking, etc.

5.3 Multispectral image denoising: experimental results

Experiments on multispectral image denoising is basically conducted by contaminating the original image with additive Gaussian noise with different levels. Then, we apply

denoising algorithms on the noisy images. We calculate certain metrics that are able to assess the denoising performance and compare results of algorithms. Typically, the used metric is either the Signal to Noise Ratio (SNR) or the Peak Signal to Noise ratio both expressed in dB: The PSNR is expressed as follows.

$$PSNR = 10 \log_{10} \left[\frac{\max_{i,j} \{I_{or}(i,j)\}^2}{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N |I_{or}(i,j) - I_{out}(i,j)|^2} \right] \quad (5.1)$$

where $I_{or}(i,j)$ and $I_{out}(i,j)$ are the pixel values at position (i,j) in the original and output images respectively. PSNR is the ratio of the maximum possible value of the signal in term of power and the power of the distortion caused by the noise. PSNR is expressed in term of the logarithmic decibel scale. The higher the PSNR is, the better the result is.

5.3.1 Vector anisotropic Gaussian filter: Experimental results

We have applied the vector anisotropic Gaussian filter method proposed in section 4.3 for color and multispectral image denoising. We conducted the experiments on a color image (Baboon, see Fig. 5.3) which has a wide range of colors and multispectral images from HYDICE set and IRIS M^3 databases. Eleven images are extracted for experimentation from HYDICE, and corrupted with an additive zero-mean white Gaussian noise of different levels. Two denoising algorithms are used for comparison with the proposed method: (i) Vector SURE-LET multichannel image denoising algorithm [26] (ii) the anisotropic filter applied on each band image separately. These algorithms have been chosen for comparison purposes. SURE-LET multichannel algorithm has presented good performance for multichannel image denoising. We apply the 2D anisotropic filter to show the importance of including the spectral information in the denoising process. Figure 5.3 shows the denoising results of the Baboon image for $\sigma_{noise} = 20$. Table 5.2 represents the PSNR comparison between the three algorithms. We can see that the proposed method outperforms other methods for different noise levels. Even in the case of heavily corrupted image ($\sigma_{noise}=100$, 8.13 dB), the proposed method presents good performance compared to other methods. With these results, we confirm two things. First, VAG presents good performance for color image denoising. Second, results confirm the outperformance of the denoising algorithms that take into account the spectral information compared with the 2D denoising algorithm.

Figure 5.4(a) represents image from HYDICE database taken at 490nm and. We corrupted it with a white additive Gaussian noise. The noisy image is illustrated in 5.4(b). Results of the three denoising algorithms are illustrated in Fig.5.4(c), (d) and (e).

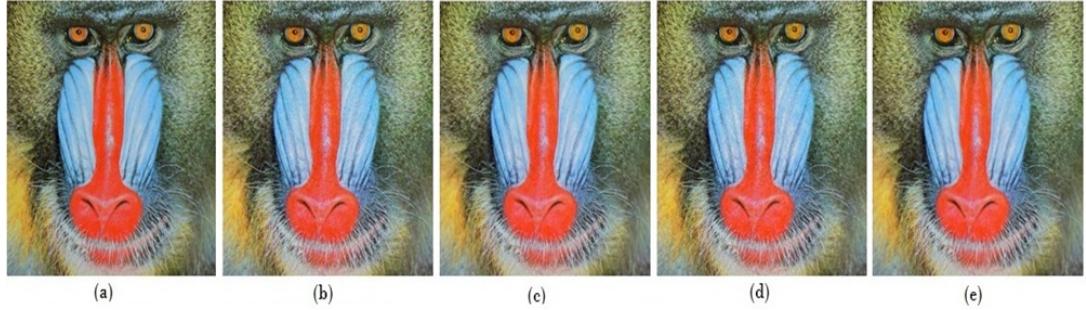


FIGURE 5.3: (a) Original Baboon image (b) Corrupted image ($\sigma_{noise} = 20$) (c) 2D anisotropic filter (d) SURE-LET (e) Proposed VAG method

TABLE 5.2: Comparison of different denoising methods for different noise level of the Baboon image (best result is bold)

Noise Standard deviation per channel	Input PSNR (dB)	2D anisotropic filter	Vector SURE-LET	Proposed vector anisotropic filter
20	22.18	24.56	26.65	28.85
30	18.77	21.92	24.69	27.25
40	16.44	20.02	23.43	26.01
50	14.71	18.56	22.55	24.67
100	10.22	14.04	20.32	21.15

PSNR Results in Table 5.3 demonstrate the outperformance of the proposed VAG algorithm particularly with high level of noise: with ($\sigma_{noise}=100$, 10.62 dB) we obtained a PSNR=20.15 dB compared to 14.10 dB and 11.07 dB for vector SURE-LET and 2D anisotropic filter respectively. Fig. 5.5(a) represents a sample of the multispectral face image from IRIS M^3 database. Fig. 5.5(b) illustrates the noisy image with ($\sigma_{noise}=30$, 18.62 dB). Denoising results of the 2D anisotropic filter, vector SURE-LET and VAG filters are presented in Fig. 5.5(c),(d) and (e) respectively. PSNR results from Table 5.4 demonstrates that VAG filter outperforms the other algorithms even with high level of noise.

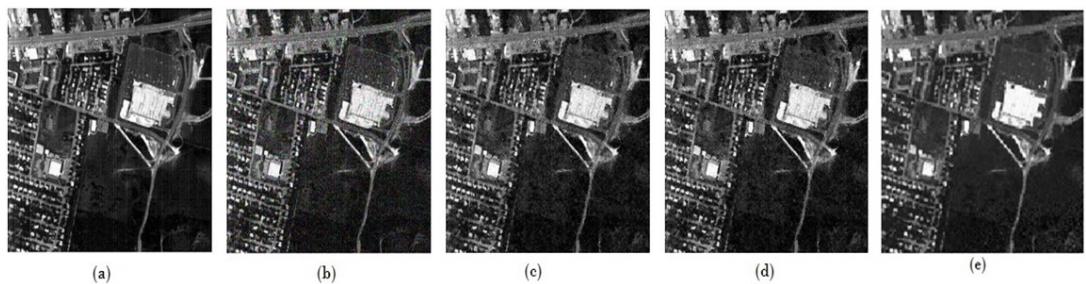


FIGURE 5.4: (a) Hydice original image (b) Corrupted image ($\sigma_{noise} = 30$) (c) 2D anisotropic filter (d) SURE-LET (e) Proposed VAG method

TABLE 5.3: Comparison of different denoising methods for different noise level of Hydice image (best result is bold)

Noise Standard deviation per channel	Input PSNR (dB)	2D anisotropic filter	Vector SURE-LET	Proposed vector anisotropic filter
20	22.46	23.74	27.83	29.83
30	19.25	20.52	24.41	28.67
40	17.04	18.27	21.98	28.01
50	15.37	16.56	20.07	27.33
100	10.62	11.07	14.10	20.15

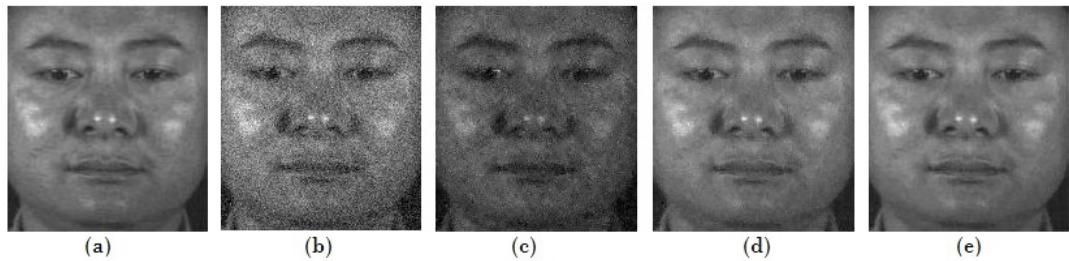


FIGURE 5.5: (a) IRIS original image (520nm) (b) Corrupted image ($\sigma_{noise} = 30$) (c) 2D anisotropic filter (d) SURE-LET (e) Proposed method

TABLE 5.4: Comparison of different denoising methods for different noise level of IRIS image (best result is bold)

Noise Standard deviation per channel	Input PSNR (dB)	2D anisotropic filter	Vector SURE-LET	Proposed vector anisotropic filter
20	22.11	26.65	29.10	31.16
30	18.62	23.25	24.37	29.61
40	16.20	20.70	22.69	25.12
50	14.40	17.86	21.11	24.49
100	9.98	11	13.96	14.12

To emphasize the adopted noise model seen in paragraph 3.2.2, we have computed the normalized histogram of local variance of all multispectral images of each data set. Results are depicted in Fig. 5.6. We notice the same decreasing exponential curve seen in Fig. 3.2. Most of the values are around zero as expected.

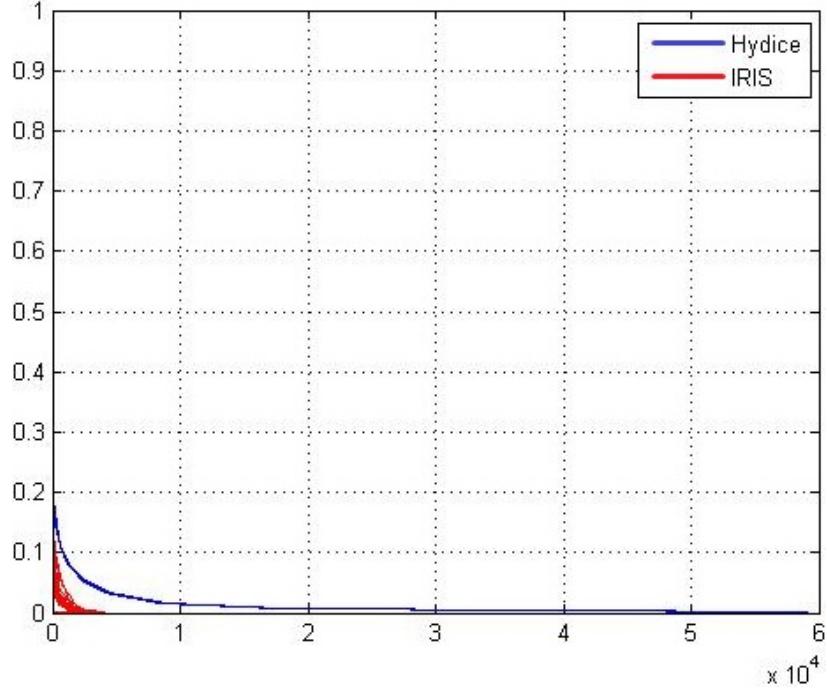


FIGURE 5.6: Normalized histogram of the variance of Hydice and IRIS images

5.3.2 Optimized Vector Non-Local Mean filter: Experimental results

We have applied the proposed OVNLNM for color and multispectral face images and we compare it with (*i*) the vector SURE-LET [26] and (*ii*) ONLM proposed in [48]. The latter is also inspired from the NLM filter and is adapted for multispectral image denoising where parametrization is conducted using ad-hoc means. Figure 5.7 illustrates Lena image corrupted with white Gaussian noise and the outputs of the three denoising algorithms. By analysing Fig. 5.8 which describes the variation of the PSNR according to the noise variance, we notice the outperformance of the proposed framework especially in area of high level of noise. For an input PSNR of 10.27 dB, we have an output PSNR of 28.04 dB for the proposed algorithm compared to 23.16 dB and 18.45 dB for SURE-LET and ONLM respectively.

For Baboon image, Fig. 5.9 exhibits the original, the noisy and the output images of three denoising algorithms. PSNR curves shown in Fig. 5.10 demonstrate the outperformance of OVNLNM compared to other algorithms. With a high level of noise: Input PSNR=8.14 dB, we have obtained 22.57 dB for OVNLNM compared to 19.76 dB and 17.96 dB with SURE-LET and ONLM.

Figures 5.11 and 5.13 illustrate the multispectral image samples used in our experiments. We corrupt these images with different levels of noise then application of the three denoising algorithms gives the results in Fig 5.15 and 5.17 for subject 1 and 2 respectively with noise standard deviation $\sigma = 30$.

To emphasize the performance of the proposed framework, we present the pixel along a

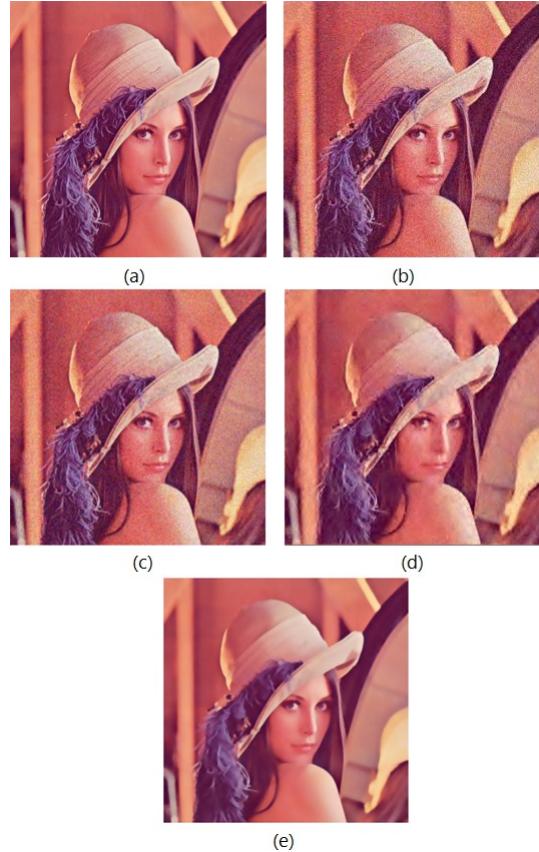


FIGURE 5.7: Lena image and denoising results (a) Original (b) Corrupted (c) ONLM
(d) SURE-LET (e) OVNLM

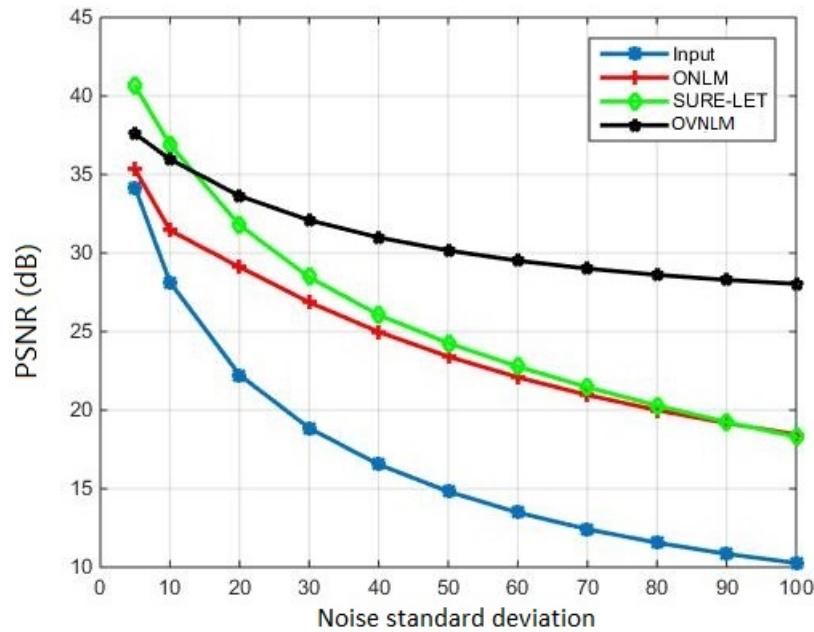


FIGURE 5.8: PSNR variation according to noise standard deviation for Lena image

cross-section in the 25th spectral band for both subjects as illustrated in Fig. 5.12 and Fig. 5.14. We compute the MSE between the cross-section of the original signal and the

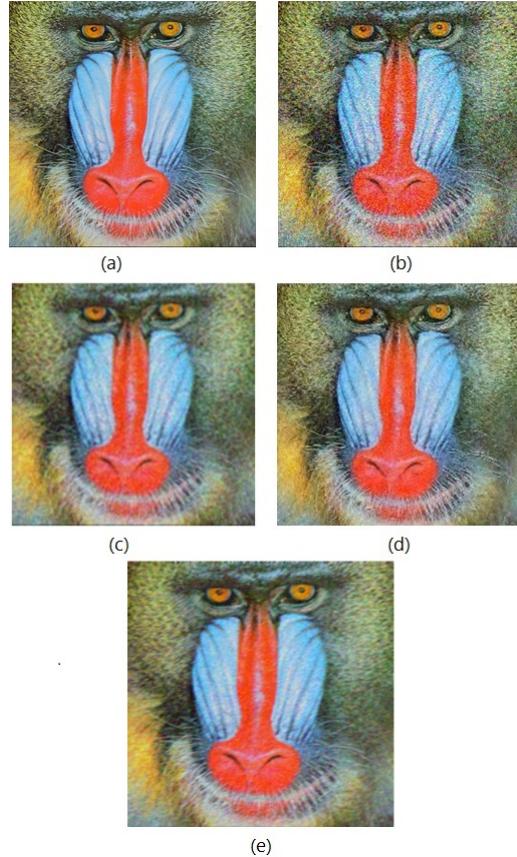


FIGURE 5.9: Baboon and denoising results (a) Original (b) Corrupted (c) ONLM (d) SURE-LET (e) OVNLM

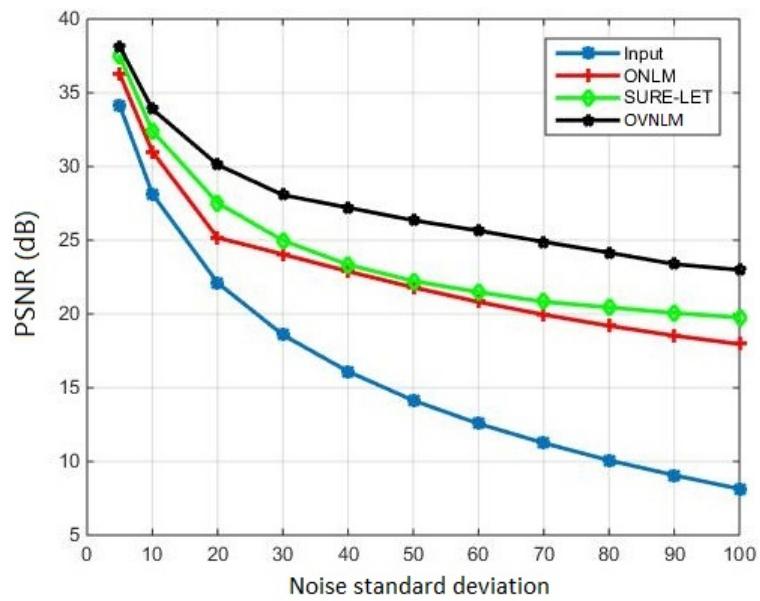


FIGURE 5.10: PSNR variation according to noise standard deviation for Baboon image

output of the three algorithms as illustrated in Table 5.5.

Results demonstrate that the proposed framework presents the best performance with the closest signal to the original one for both subject 1 and subject 2. PSNR results

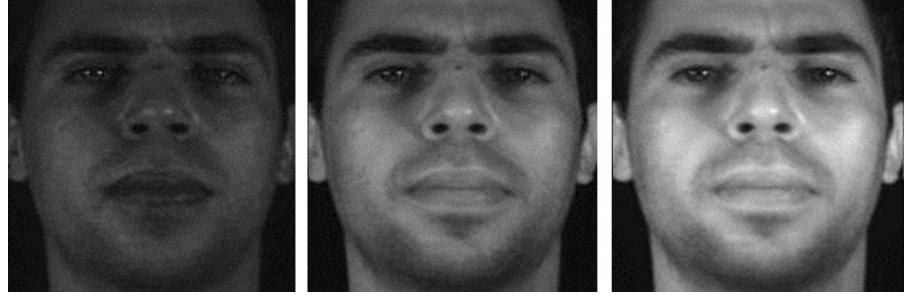


FIGURE 5.11: Multispectral images for subject 1 in 480nm, 560nm and 720nm

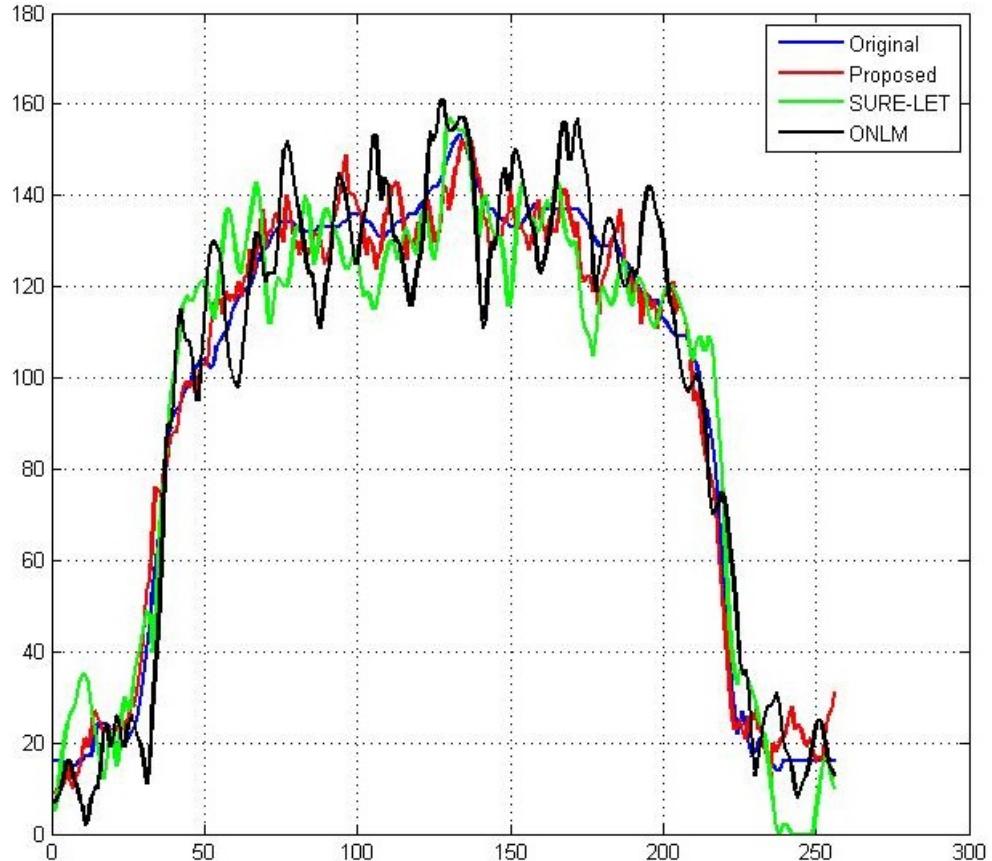


FIGURE 5.12: Pixel variation along a profile for subject 1



FIGURE 5.13: Multispectral images for subject 2 in 480nm, 560nm and 720nm

depicted in Fig. 5.16 for subject 1 demonstrates that OVNLM algorithm exhibits good performance especially in area of high level of noise. Indeed, with $\sigma = 100$, the OVNLM



FIGURE 5.14: Pixel variation along a profile for subject 2

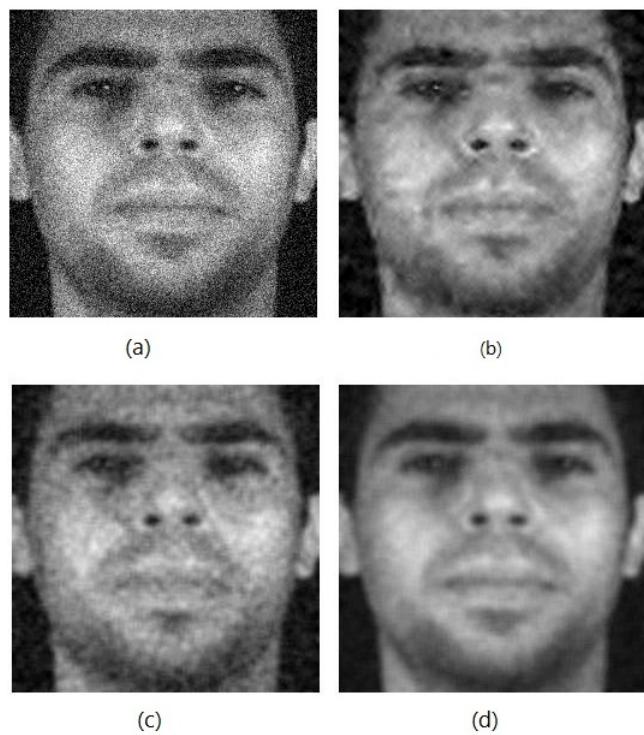


FIGURE 5.15: Subject 1: (a) Noisy image (b) SURE-LET (c) ONLM (d) Proposed

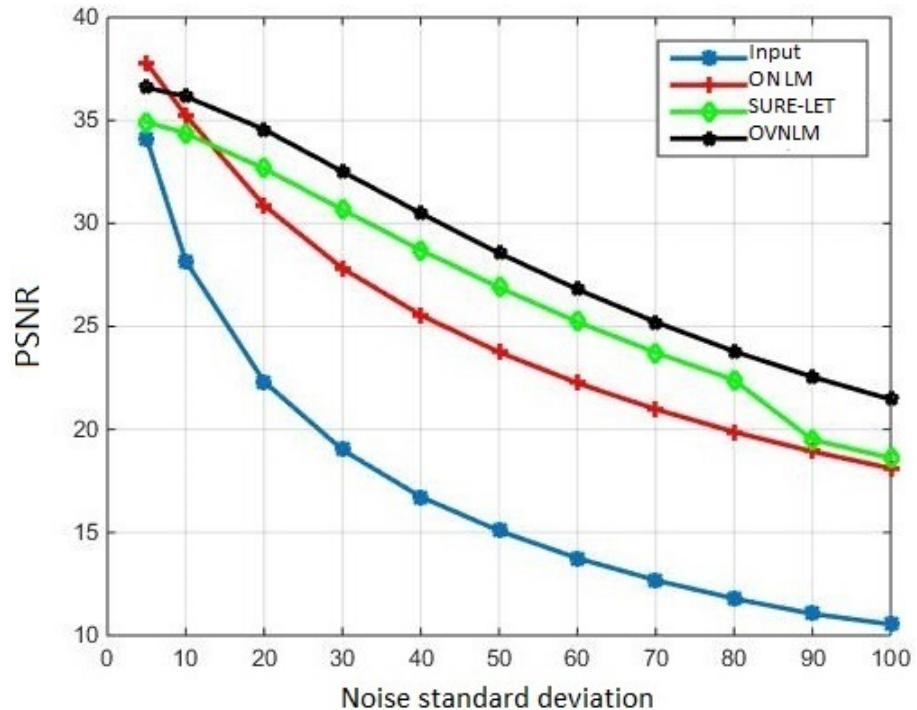


FIGURE 5.16: PSNR curves for subject 1

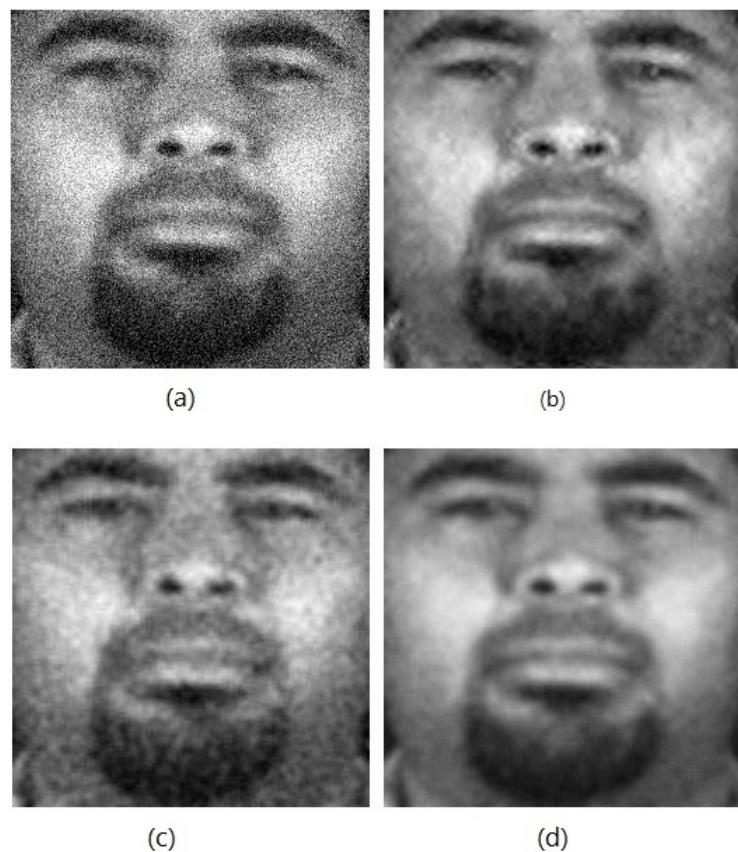


FIGURE 5.17: Subject 2: (a) Noisy image (b) SURE-LET (c) ONLM (d) Proposed

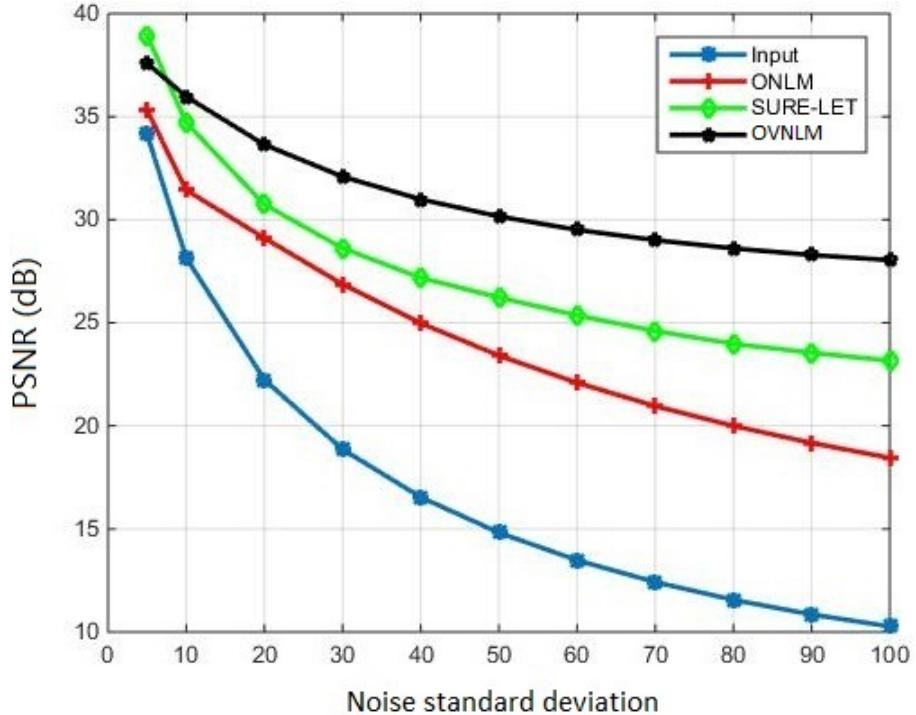


FIGURE 5.18: PSNR curves for subject 2

TABLE 5.5: MSE between original signal and restored signals

	SURE-LET	ONLM	Proposed
Subject 1	6.0796	6.64	3.7
Subject 2	9.73	9.71	4.98

has an output PSNR of 23.03 dB compared to 18.62 dB and 18.11 dB obtained with SURE-LET and ONLM respectively. PSNR results for subject 2 are illustrated in Fig. 5.18. OVNLM presents also good performance particularly with heavily corrupted images ($\sigma = 100$). Indeed, we obtain for OVNLM a PSNR of 23 dB compared to 18.58 dB and 17 dB with vector SURE-LET and ONLM.

To emphasize the importance of a good parametrization of the proposed filter, we analyse its performance by varying parameter h which controls the degree of smoothing for a fixed covariance matrix Φ and with different noise levels as shown in Fig. 5.19. We notice that the resulting PSNR depends enormously on the choice of parameter h which also implies a good choice of parameter Φ . Thus, we can deduce the importance of an optimization scheme to parameterize the filter.

The proposed denoising filters have demonstrated their outperformance compared to other algorithms. Experiments on real color and multispectral images have shown good results in terms of PSNR. We have tried while denoising to take advantage from the spectral information brought by the imaging system and achieve a good tradeoff between good denoising performance and edge and line preservation by extending filters

i.e. anisotropic Gaussian filter and NLM filter, whose intrinsic property is details and fine structures conservation, to the vector case.

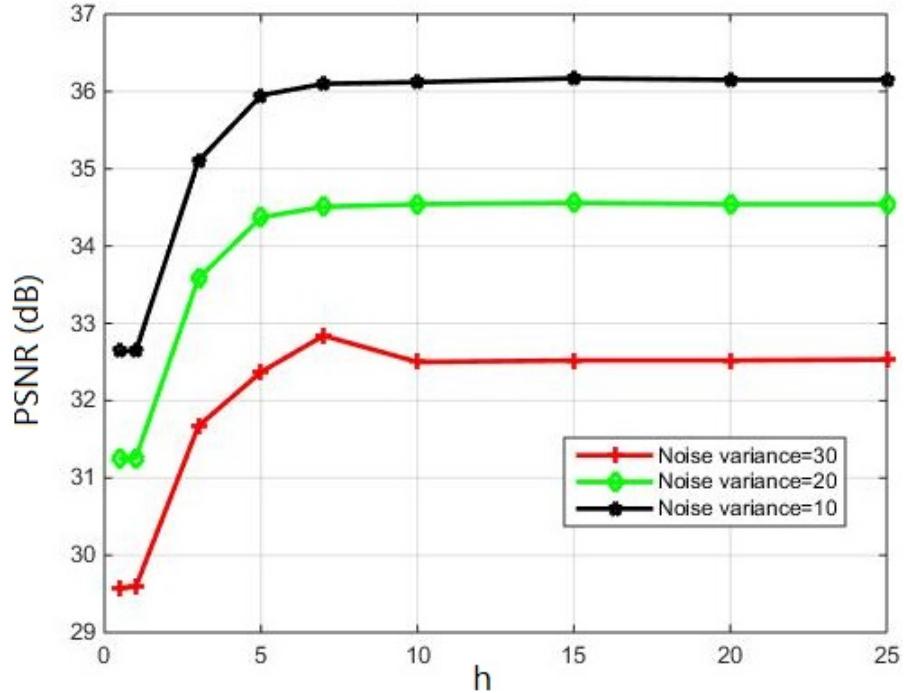


FIGURE 5.19: Variation of PSNR with relative to h

5.4 Gravitational Weighted Fuzzy C-Means: Experimental results

To validate our Gravitational Weighted Fuzzy C-Means (GWFCM) method proposed in section 4.3, we conduct two kinds of experiments: (*i*) clustering two data sets from UCI repository (*ii*) segmenting multispectral face images from IRIS Lab. The proposed method is compared to four clustering algorithms: FCM, bootstrap weighted FCM referred to as BFCM [104], W-K-means [105] and the data weighted clustering DWGK [101]. These four algorithms have been chosen for comparison purposes as they presented good performance for color image segmentation and can be easily applied on multispectral images.

We use throughout these experiments the following similarity indexes defined as follows: Let P_1 and P_2 two partitions. We define a as the number of object pairs that belong to the same clusters in both P_1 and P_2 . Let b be the number of object pairs that belongs to different clusters in both pairs. Let c be the number of object pairs that belong to the same clusters in P_1 but in different clusters in P_2 . Finally, let d the number of pairs that belong to different clusters in P_1 but belong to the same cluster in P_2 . Rand,

Fowlkes-Mallows, Jaccard and Adjusted Rand indexes are defined as:

$$Rand = \frac{a+b}{a+b+c+d} \quad (5.2)$$

$$Fowlkes - Mallows = \frac{a}{\sqrt{(a+d)(a+c)}} \quad (5.3)$$

$$Jaccard = \frac{a}{a+c+d} \quad (5.4)$$

$$Adjusted - Rand = \frac{\frac{a - \frac{(a+d)(a+c)}{a+b+c+d}}{\frac{a+b+c+d}{2} - \frac{(a+d)(a+c)}{a+b+c+d}}}{\frac{a+b+c+d}{2}} \quad (5.5)$$

5.4.1 Clusetring

We use Thyroid data set from UCI repository. It consists of five features and two or three classes according to patient status: normal and sick (suffering from hyperthyroidism, and suffering from hypothyroidism). In addition, we use the Glass data set which consists of nine features and originally seven clusters. We use an imbalanced version of this data set where four clusters are merged into one cluster and the three others into a second cluster. For each data point $X_j = (x_{j1}; x_{j2}; \dots; x_{jd})$, the mass m_j is chosen to be 1.

We run the four algorithms on these data sets with different number of clusters and we compute the Rand index defined in eq. 5.2 to assess the clustering results. The higher Rand is, the better the clustering algorithm is. We take into account the maximum Rand index [130] value obtained after multiple runs. From results in Fig. 5.20, we notice that the weighted algorithms e.g. BFCM, W-K means and GWFCM outperform the classic FCM algorithms which demonstrates the utility of such approach. Results shows also that the proposed GWFCM outperforms the other clustering algorithms. In fact, it attains the best values in all cluster numbers and best results are obtained with the real number of clusters (2 or 3). However, for BFCM, best results coincide only with two clusters. Same behavior is noticed for FCM and W-K algorithms.

Figure 5.21 represents results for Glass data sets. W-K means algorithm showed the lowest performance among all the clustering algorithms with Rand index value of 0.5 for $c = 2$. The proposed GWFCM has the best performance with 0.82 for two clusters. For the rest of cluster numbers, GWFCM keeps his outperformance over other clustering algorithms.

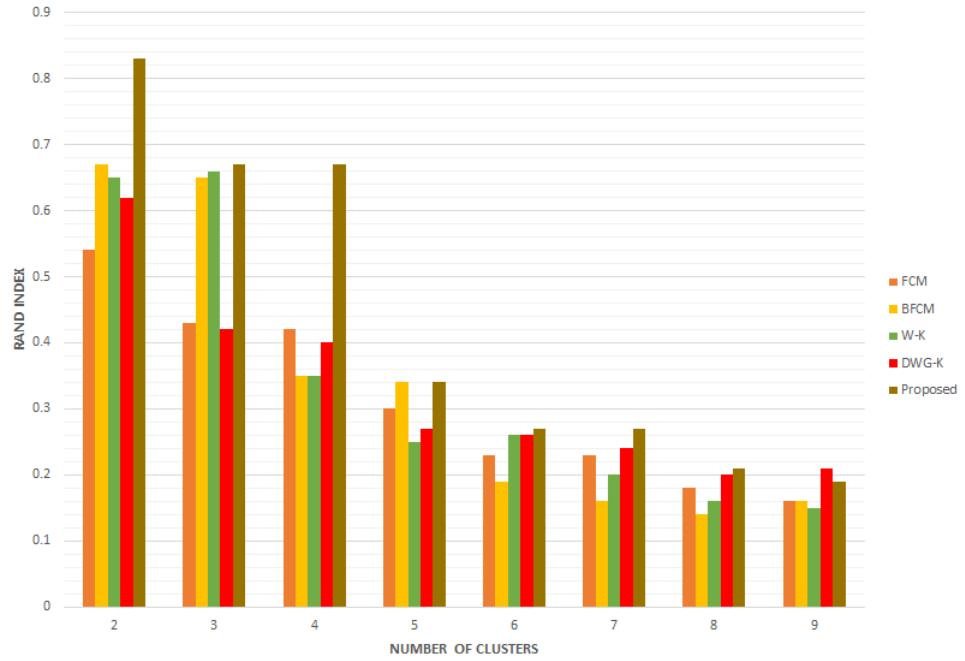


FIGURE 5.20: Thyroid - Rand index

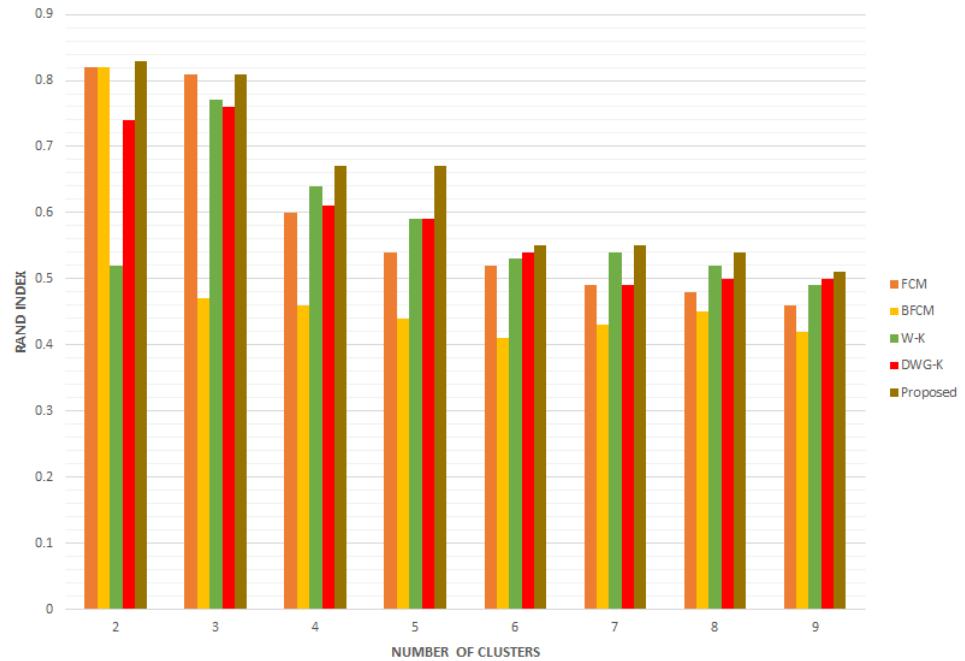


FIGURE 5.21: Imbalanced Glass - Rand index

5.4.2 Multispectral image segmentation

To investigate more the performance of GWFCM, we run it for the purpose of multispectral image segmentation using images from IRIS M^3 database. In this context the data point is the pixel vector as illustrated in Fig. 5.22 [131] $p_j = (p_{j1}, p_{j2}, \dots, p_{j25})$ where $p_{ji}, i=1\dots 25$ is the value of the j^{th} pixel at the i^{th} spectral band. The mass of each

pixel is defined as the maximum value it can take at a given spectral band:

$$m_{pj} = \max_i \{p_{ji}\} \quad (5.6)$$

To compare with other clustering algorithm, we use the following cluster validity indexes

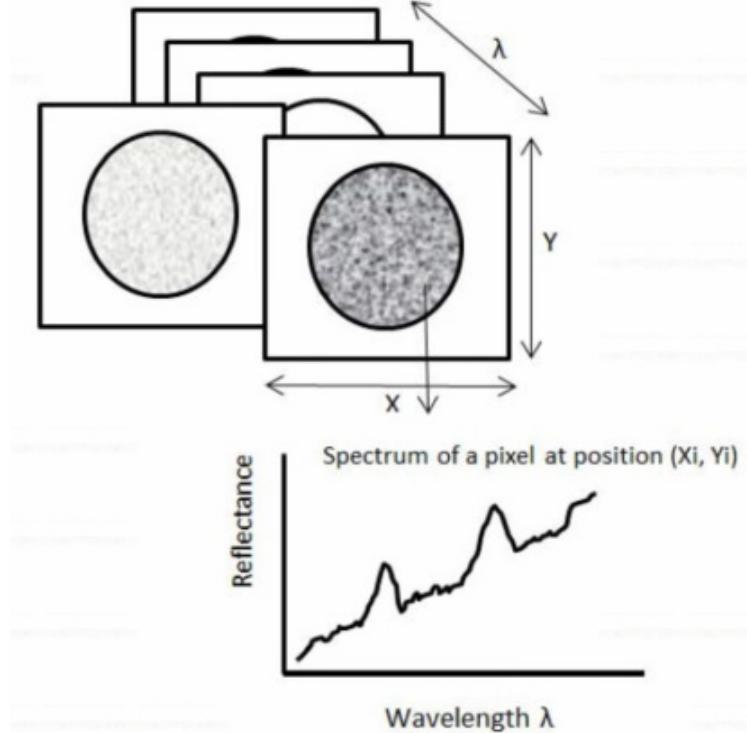


FIGURE 5.22: Multispectral image organization

presented in Chapter 2: XB (see eq. 2.10), FS (see eq. 2.9), PBMF (see eq. 2.11) and CO (see eq. 2.6) to assess the clustering results. Fig. 5.23 and Fig. 5.24 represent examples from IRIS M^3 database taken in the Halogen light (first twelve bands). Figure 5.25



FIGURE 5.23: Multispectral face image: subject 1



FIGURE 5.24: Multispectral face image: subject 2

and Fig. 5.26 show segmentation results for the five algorithms discussed in the previous

section. Table 5.6 presents CVIs values for each clustering algorithm applied on images of Fig. 5.23. We notice that the proposed method presents the best performance with the smallest values of FS and XB indexes. With PFMF and CO, the optimal partition is highlighted by the maximum value. According to these two CVIs, the best partition is obtained with GWFCM. Results for segmentation of Fig. 5.24 in Table 5.7 also show that GWFCM outperforms the other clustering algorithm according to the minimum value of FS index and maximum value of PBMF and CO indexes. The optimal partition according to XB index is obtained with DWG-K. However, GWFCM provides the best separation between clusters centers: $1.9 \cdot 10^5$ compared to $1.8 \cdot 10^5$ obtained by applying DWG-K. Thus the obtained results confirms that GWFCM is superior to the other four clustering algorithm.



FIGURE 5.25: Segmentation result: subject 1

TABLE 5.6: CVIs values for subject 1

CVI	FCM	BFCM	W-K	DWG-K	Proposed
XB	0.09	0.09	0.09	0.08	0.08
FS (10^8)	-2.57	-2.60	-2.39	-0.1	-3.17
PBMF (10^4)	0.25	0.26	0.25	0.39	1.06
CO	6.56	6.70	5.99	5	6.79

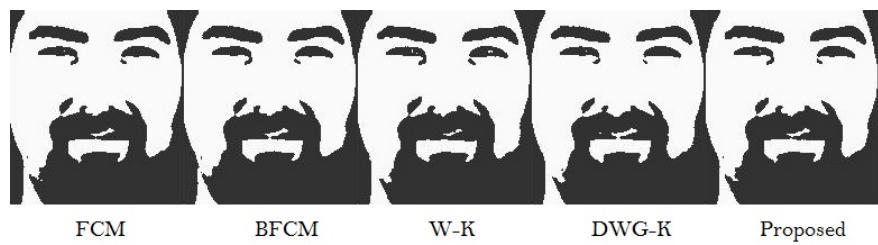


FIGURE 5.26: Segmentation result: subject 2

TABLE 5.7: CVIs values for subject 2

CVI	FCM	BFCM	W-K	DWG-K	Proposed
XB	0.07	0.07	0.07	0.04	0.06
FS (10^8)	-6.7	-6.7	-6.6	-7	-7.4
PBMF (10^4)	0.86	0.88	0.82	0.77	1.5
CO	7.2	7.4	6.9	7.3	7.4

5.5 Cluster validity indexes: Experimental results

5.5.1 CVI evaluation methodologies

The classic CVI evaluation methodology consists in applying a clustering algorithm such as FCM on each data set while varying the number of clusters and computing CVIs for each partition. The best CVI determines that the real number of clusters is the one that best fits the data.

The classic methodology is based on the assumption of perfection of the clustering algorithm which is not always true. In the alternative methodology, the definition of the best partition is different [132]. Indeed, the best partition is the one which is most similar to the perfect partition. Similarity is quantified using similarity measures such as Rand or adjusted Rand index. The best CVI is the index which determines that the most similar partition to the perfect partition is the best one.

5.5.2 Cluster validity index based on n-sphere

In order to demonstrate the effectiveness of this CVI, we conduct experimentation on five data sets from UCI repository. We use: the Tyroid data set, the Wine data set which is composed of 178 thirteen dimensions vectors, the Glass (balanced version of 7 clusters) data set where each dimension represents a constituent of the three types of the studied wine, the famous Iris data set which contains 150 data vectors described by four features: sepal length, sepal width, petal length and petal width and consists of three clusters: two overlapping clusters and one linearly separable from the others, the Yeast data set which is composed of 8 attributes and 10 clusters. We use also two data sets from the eYale B database: Yale_5 and Yale_7 composed of 5 and 7 clusters respectively.

Table 5.8 represents the values of parameter T used to choose a specific point for the compactness measure. Our index denoted S is compared to the following CVIs: PC (see eq. 2.2), PE (see eq. 2.3), XB (see eq. 2.10), FS (see eq. 2.9) and PBMF (see eq. 2.11). Table 5.9 summarizes the results obtained for the thyroid data set. The optimal

number of cluster i.e. 3 is highlighted only by the proposed CVI S and XB while other indexes fail. CVIs such as PC and PE, whose computation is based only on membership values of each data vector in each cluster, are known by their monotonous dependency on the number of clusters, hence it is not surprising if 2 was determined as the optimal number of clusters. This behavior is confirmed by the results obtained for the wine data set in Table 5.10. Indeed and as predicted, PC and PE failed to determine 3 as the optimal number of clusters. Only the proposed CVI, S, and XB performed well. For the glass data set, only the FS and S indexes perform well as shown in Table 5.11. The Iris data set is more complicated due to the presence of overlapping clusters. In this case, only index S was able to determine 3 as the optimal number of clusters as shown in Table 5.12. Behavior of PC and PE is also confirmed. In addition, from Table 5.13 for yeast data set, only index S determined 10 as the optimal number of clusters. For the eYale data sets, only the proposed CVI was able to predict 5 and 7 as the optimal number of clusters as demonstrated by results in Tables 5.14 and 5.15. Experimental results showed that the proposed CVI outperforms the other indexes for all data sets particularly those where overlap between clusters is present. In addition the proposed CVI is also efficient in case of high dimensional data e.g. wine data set.

TABLE 5.8: Values of parameter T for each data set

data	Iris	Glass	Thyroid	Wine	Yeast	Yale_5	Yale_7
T	0.09	0.001	0.01	0.05	0.01	0.04	0.002

TABLE 5.9: CVIs values for Thyroid data set: best result is bold

c	PC	PE	XB	FS	PBMF	S
2	0.7478	0.4019	0.3774	4.8857	0.0281	7.82
3	0.7459	0.4772	0.1564	-2.6151	0.0208	7.06
4	0.5881	0.7425	0.8178	-3.9952	0.0387	9.7
5	0.4622	0.9962	1.4409	-4.8197	0.0072	22.77
6	0.5007	0.9826	0.7670	-7.3917	0.0234	17.74
7	0.4253	1.1637	1.3790	-6.5266	0.0157	15.82
8	0.3609	1.3440	1.8672	-5.8464	0.0020	16.31

We use besides Rand index (see eq. 5.2), other similarity indexes: Fowlkes-Mallows (see eq. 5.3), Jaccard index (see eq. 5.4), and Adjusted Rand (see eq. 5.5). Jaccard index ranges in $[0, 1]$ with 1 corresponds to totally similar sets. With Fowlkes-Mallows index, a higher value indicates a greater similarity between the objects. The adjusted Rand index is derived from the Rand index. It ranges in $[-1, 1]$ and with a value of 1, we have a perfect match between objects. We use Iris and Thyroid data sets in this case.

TABLE 5.10: CVIs values for Wine data set: best result is bold

c	PC	PE	XB	FS	PBMF	S
2	0.62	0.55	0.48	33.33	0.109	2.05
3	0.51	0.83	0.34	8.39	0.07	0.98
4	0.38	1.14	2312	5.68	0.048	5.80
5	0.30	1.36	43.88	5.03	0.031	34.43
6	0.26	1.52	2189	4.00	0.021	57.28
7	0.22	1.68	1.1e+4	3.11	0.015	49.71
8	0.19	1.83	3.8e+7	2.69	0.011	94.97
9	0.17	1.96	1.1e+4	2.48	0.010	30.23

TABLE 5.11: CVIs values for Glass data set: best result is bold

c	PC	PE	XB	FS	PBMF	S
2	0.82	0.29	0.13	-7.93	0.051	72.71
3	0.71	0.51	0.36	-9.85	0.044	58.22
4	0.68	0.61	0.24	-19.28	0.030	59.56
5	0.58	0.83	0.91	-14.46	0.015	51.22
6	0.51	0.99	1.34	-13.20	0.006	55.83
7	0.57	0.93	0.34	-20.06	0.002	45.33
8	0.48	1.09	1.86	-16.35	0.004	50.39
9	0.46	1.19	1.83	-14.88	0.003	68.12
10	0.45	1.22	1.63	-16.34	0.002	69.33

TABLE 5.12: CVIs values for Iris data set: best result is bold

c	PC	PE	XB	FS	PBMF	S
2	0.84	0.26	0.09	-9.87	0.047	1.61
3	0.72	0.49	0.20	-12.76	0.02	1.45
4	0.64	0.66	0.30	-13.02	0.01	2.07
5	0.55	0.88	0.57	-13.28	0.009	3.49
6	0.54	0.91	0.41	-12.31	0.009	3.42
7	0.50	1.08	0.43	-14.37	0.004	3.55
8	0.49	1.09	0.43	-12.17	0.002	5.72
9	0.46	1.20	0.29	-13.72	0.003	6.3

Results for thyroid data set shown in Table 5.16, demonstrate that the proposed index S and XB have the best performances which is consistent with the results in Table 5.9. In fact, the best partition with classic and new methodologies coincides 100 times over 100 executions. For Iris data set which comprises two overlapping clusters, results are different. Results are illustrated in Table 5.17. With the Rand index, this specific case occurs 58 times, 100 times with Jaccard and Fowlkes-Mallows and only 7 times with the

TABLE 5.13: CVIs values for Yeast data set: best result is bold

c	PC	PE	XB	FS	PBMF	S
8	0.17	1.92	3496	9.59	0.002	8.69
9	0.15	2.04	3457	8.36	0.001	7.82
10	0.13	2.15	451.31	7.42	0.001	5.6
11	0.12	2.25	476.17	6.71	0.001	6.74
12	0.11	2.34	3.2e+4	6.09	9.7e-4	6.78
13	0.10	2.42	1615	5.62	6.1e-4	6.60
14	0.10	2.48	3824	5.03	7.6e-4	36.54

TABLE 5.14: CVIs values for Yale_5 data set: best result is bold

c	PC	PE	XB	FS	PBMF	S
2	0.788	0.9	42.2	-5.8e+7	745	5.8e+3
3	0.754	1.08	42.2	-1.3 e+7	615	4.1e+3
4	0.732	2.21	45.3	-6e+7	513	5.3e+3
5	0.718	2.57	48.7	-4.4e+7	522	2.8e+3
6	0.712	2.77	47.2	-2.8e+7	510	3e+3
7	0.700	2.90	50	-1.6e+7	500	4.2e+3
8	0.710	2.97	43.8	-1.2e+7	500	6.5e+3
9	0.713	2.97	41	-1e+7	500	6.8e+3

TABLE 5.15: CVIs values for Yale_7 data set: best result is bold

c	PC	PE	XB	FS	PBMF	S
3	0.662	0.69	12	-3.2e+7	131	2.1e+3
4	0.640	0.82	9	-1.e+7	116	3.6e+3
5	0.590	0.91	22	-0.7e+7	109	5.3e+3
6	0.567	0.99	16	-1.1e+7	102	1.9e+3
7	0.534	1.04	19	-0.6e+7	101	1.6e+3
8	0.530	1.09	15	-0.4e+7	103	2.2e+3
9	0.520	1.13	12.4	-0.5e+7	100	4.7e+3
10	0.516	1.17	14	-0.6e+7	100	5.2e+3

Modified Rand. Results for 1R2RCT_A, 1R2RT_B and car5 from Hopkins 155 database also demonstrate the superiority of index S over the other indexes. From Table 5.18 for 1R2RCT_A data set, index S exhibited the best performance particularly with the Fowlkes-Mallows index. For 1R2RT_B data set, results in Table 5.19 also demonstrate the outperformance of the proposed index. The other indexes particularly PC and PE completely failed. Same results are obtained for the car5 data set as illustrated in Table 5.20.

TABLE 5.16: Number of times each index proposes the optimal partition highlighted by the similarity measures as the best one for Thyroid

Similarity measure	PC	PE	XB	FS	PBMF	S
Rand	0	0	100	0	40	100
Jaccard	0	0	100	0	44	100
Fowlkes-Mallows	0	0	100	0	37	100
Modified Rand	0	0	100	0	44	100

TABLE 5.17: Number of times each index proposes the optimal partition highlighted by the similarity measures as the best one for Iris

Similarity measure	PC	PE	XB	FS	PBMF	S
Rand	0	0	0	1	0	58
Jaccard	0	0	0	0	0	100
Fowlkes-Mallows	0	0	0	0	0	100
Modified Rand	0	0	0	13	0	7

TABLE 5.18: Number of times each index proposes the optimal partition highlighted by the similarity measures as the best one for 1R2RCT_A

Similarity measure	PC	PE	XB	FS	PBMF	S
Rand	0	0	0	26	0	45
Jaccard	0	0	0	0	0	17
Fowlkes-Mallows	0	0	0	0	0	82
Modified Rand	0	0	22	13	17	55

TABLE 5.19: Number of times each index proposes the optimal partition highlighted by the similarity measures as the best one for 1R2RT_B

Similarity measure	PC	PE	XB	FS	PBMF	S
Rand	0	0	0	10	8	32
Jaccard	0	0	0	0	0	56
Fowlkes-Mallows	0	0	0	0	0	8
Modified Rand	0	0	54	11	19	63

TABLE 5.20: Number of times each index proposes the optimal partition highlighted by the similarity measures as the best one for car5

Similarity measure	PC	PE	XB	FS	PBMF	S
Rand	0	0	0	0	8	12
Jaccard	0	0	0	0	0	9
Fowlkes-Mallows	0	0	0	17	0	48
Modified Rand	0	0	0	22	14	54

- Cluster validity index based on Jeffrey divergence

- Synthetic data sets

We use four two-dimensional synthetic data sets S1, S2, S3 and S4 consisting of 5,000 data vectors generated from 15 Gaussian clusters. These data sets are characterized by clusters of different shapes: circular and elliptical. In addition, clusters in data set S1 are well-separated compared to the other data sets where clusters overlap more and more until becoming indistinguishable in data set S4. Furthermore, we use R15, a specially shaped data set. R15 consists of 15 Gaussian clusters in 2D. Figures 5.27, 5.29, 5.31, 5.33, and 5.35 show data distribution in each set. We apply FCM with a number of clusters varying from 10 to 20. The best CVI should identify 15 as the right number of clusters is the one that outperforms the others. In our experiments, we use PC, PE, XB, PBM [86], PBMF_FVG [133], OSV and P indexes for evaluation.

Results for each set are represented in Fig. 5.28, 5.30, 5.32, 5.34 and 5.36 where the proposed index is referred as I. The best result for each CVI is highlighted with black mark. The findings demonstrate that the proposed validity index successfully determines the correct number of clusters for all data sets. On the other hand, indexes such as PC and PE are able to determine the correct number of clusters when there is a low overlapping between clusters such as in data set S1. But we can notice that there is a slight variation in their values as overlapping degree gets higher unlike the proposed index I and the XB index where discrimination between their values for each number of cluster is obvious. Note that the remaining indexes such PBM and PBMF_FVG completely fail to determine the correct number of clusters in all cases and the best result is always found to be the lowest cluster number.

- Real data sets

We use the Balance, Banana and Iris data sets from UCI repository and Yale_5 and Yale_7 data sets from eYale database. First, to test the performance of the indexes, we use the classic methodology presented in section 5.5.1.

Results in Fig. 5.37 for balance data set demonstrate that only the proposed index and the P index are able to determine the optimal number of clusters. The behavior of indexes such as PC and PBM_FVG seen with synthetic data sets is confirmed with the real data set, i.e. the lowest number of clusters is chosen as the optimal number. These indexes represent a monotonic tendency with the number of clusters. In Fig. 5.38 for Banana data set, the proposed validity index estimated that the real number of cluster c^* is the right number of cluster. A large variation in index I is noticed. This gives an information about the correct number of clusters.

Almost all the validity indexes were able to estimate c^* . The monotonic tendency of some indexes such as PE and PC is also confirmed. The presence of overlap in the Iris data set makes it difficult to estimate the optimal number of cluster. Only index I was able to determine $c^* = 3$ as the optimal number of clusters as shown in Fig. 5.39. PC and PE always exhibit the monotonic tendency for the number of clusters. PBM and PBMF_FVG showed the same behavior as for the previous data sets. With eYale database, index I exhibits good performance. Indeed, for Yale_5 data set as shown in Fig. 5.40, only the proposed index was able to predict the real number of cluster as well as for Yale_7 data set according to the results in Fig. 5.41.

For the alternative methodology, We use the Balance, Banana and Iris data sets from UCI repository and 1R2RCT_A, 1R2RT_B and car5 data sets from Hopkins 155 database. We use the four similarity measures namely the Rand index, the Fowlkes-Mallows index, the Jaccard index and the adjusted Rand (ARI) index. We run FCM 100 times to cluster the data sets and we compute how many times each index proposes the optimal partition highlighted by these similarity measures as the best one. Results for Balance, Banana and Iris are shown in Tables 5.21, 5.22 and 5.23. For the balance data set, index I presents good performance. In fact, it is able to detect the best partition for all the similarity measures unlike other indexes which perform well with some similarity measures but fail with others such as PC and PE known for their monotonic tendency. For the Banana data set, the proposed index also presents good performance. It was able to determine the optimal partition with all similarity measures except Rand. For Iris data set, only the proposed index I was able to determine the best partition in case of Rand, Jaccard and adjusted Rand indexes. Such result goes with the result obtained for Iris data in the classic methodology. For the Hopkins 155 data, a classic preprocessing step before using these data is to project it onto a subspace using PCA. In fact, it is known for these data that the feature trajectories of n motions in a video almost perfectly lie in a $4n$ dimensional subspace. Thus, PCA projection will reduce the dimensionality with structure preservation [134]. Results are shown in Tables 5.24, 5.25 and 5.26. The overall performance of the proposed index is better than the other indexes for 1R2RCT_A for all the similarity measures except ARI. Other indexes completely fail with Rand, FM and Jacc measures. With 1R2RT_B, index I presents also good performance with a success of 100 % for Rand, FM and Jacc. PC and XB in this case present also the same performance. OSV and P perform well only with ARI. Results for car5 demonstrate also the outperformance of index I with Rand FM and ARI. In general, results on Hopkins 155 demonstrate that the proposed index I presents the best success rate compared to other indexes.

- **Discussion**

The proposed validity index is highly useful especially for data sets where overlap between clusters is present. While some indexes completely fail in the presence of a small overlapping, the proposed index I is robust and able to estimate the optimal number of cluster even when clusters highly overlap such as in the case of Iris and S4 data sets, where clusters are hardly distinguishable. The design of the separation measure based on the density of clusters is an indicator of the degree of overlap between clusters which could not be determined with the classical separation measure based on the distance between cluster centers. However, in case of high-dimensional data, the computation of the proposed cluster validity index may be computationally expensive. For example, in case of Gaussian clusters, the computation of Jeffrey divergence requires the inversion of the covariance matrix of size $d \times d$. Using Cholesky decomposition, this operation has a complexity of $O(d^3)$.

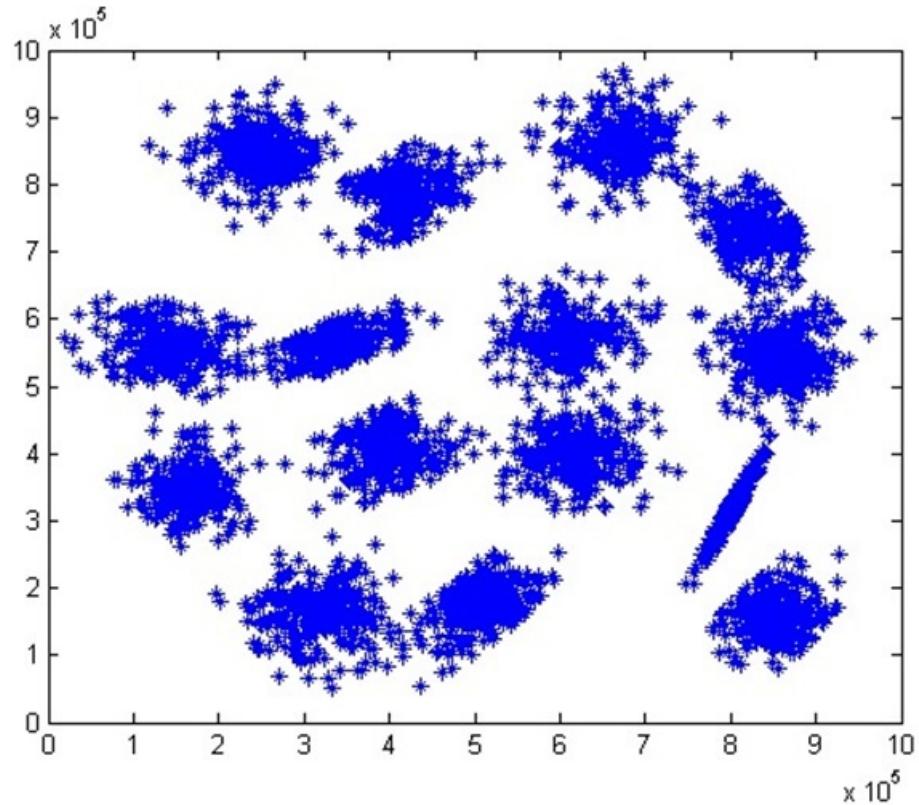


FIGURE 5.27: Data set S1 with 15 Gaussian clusters

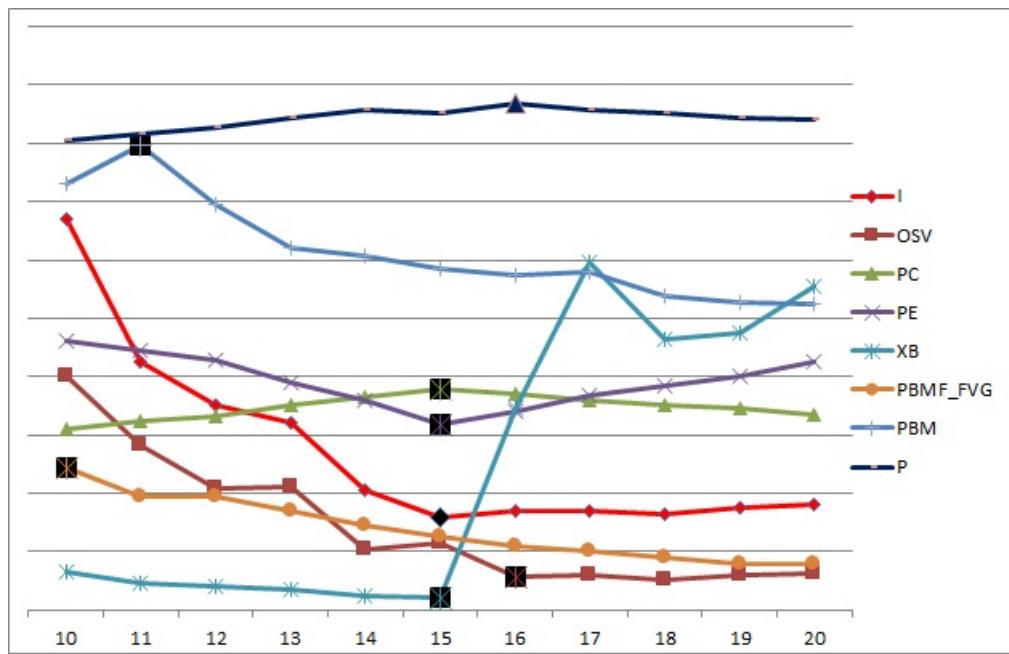


FIGURE 5.28: CVIs for S1 with different number of clusters (lines are scaled for display purpose)

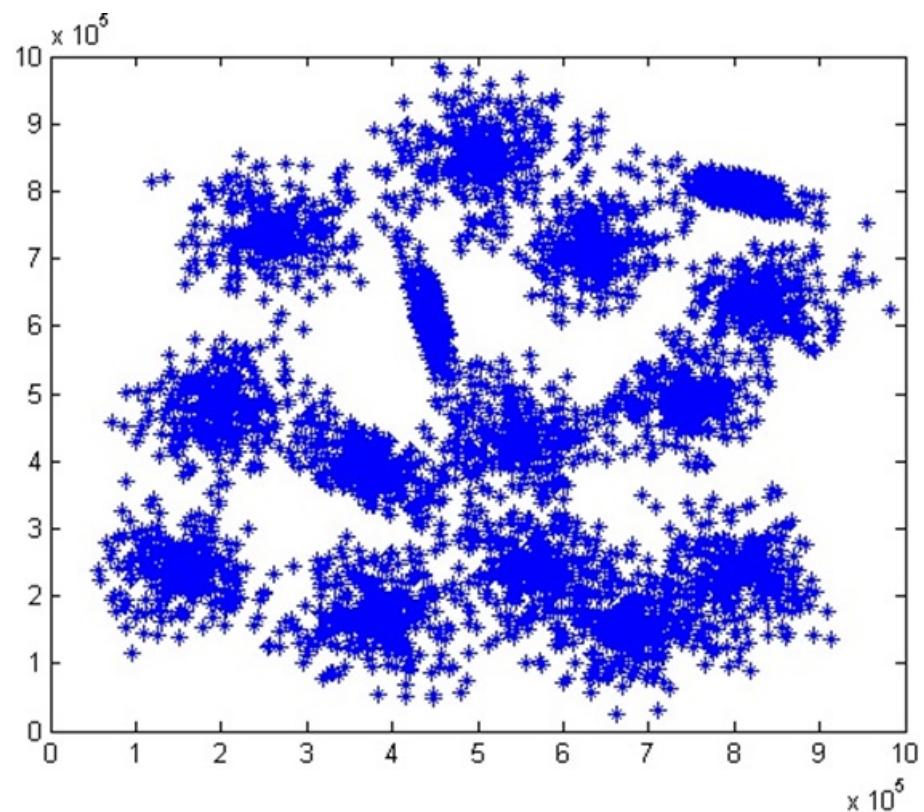


FIGURE 5.29: Data set S2 with 15 Gaussian clusters

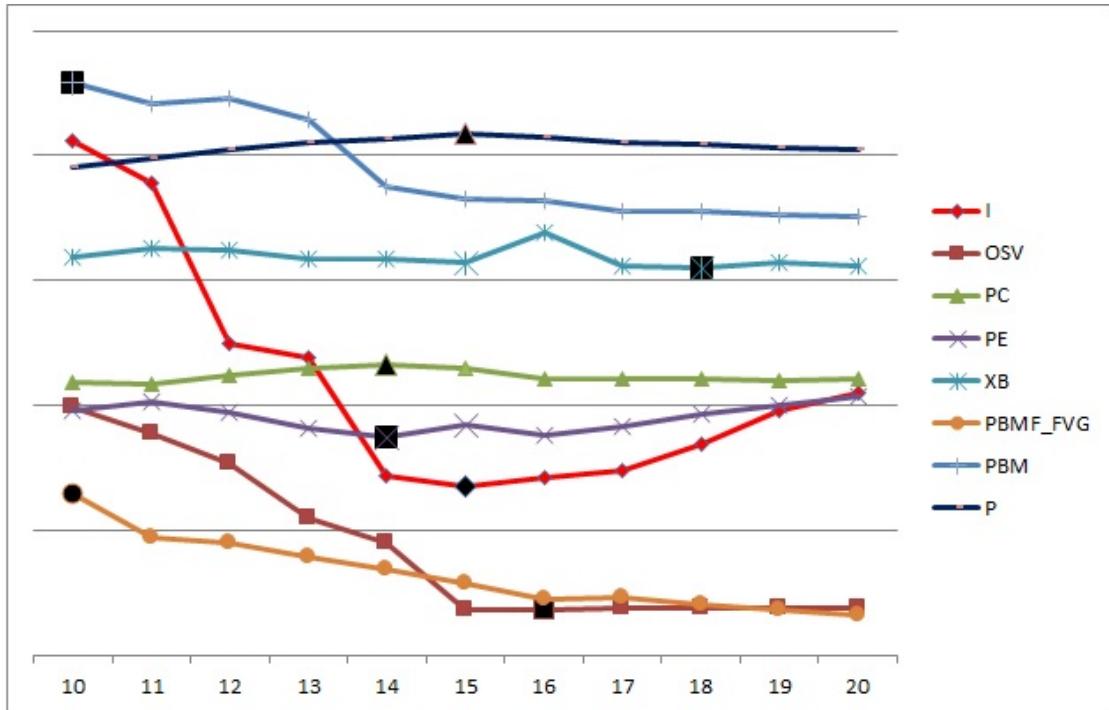


FIGURE 5.30: CVIs for S2 with different number of clusters (lines are scaled for display purpose)

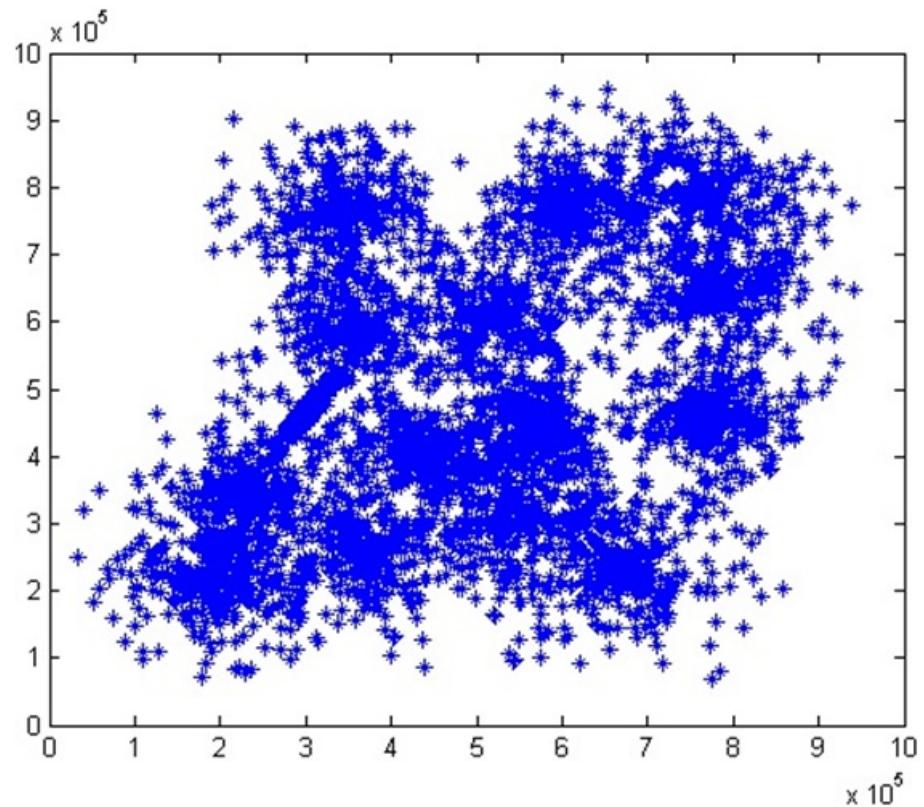


FIGURE 5.31: Data set S3 with 15 Gaussian clusters

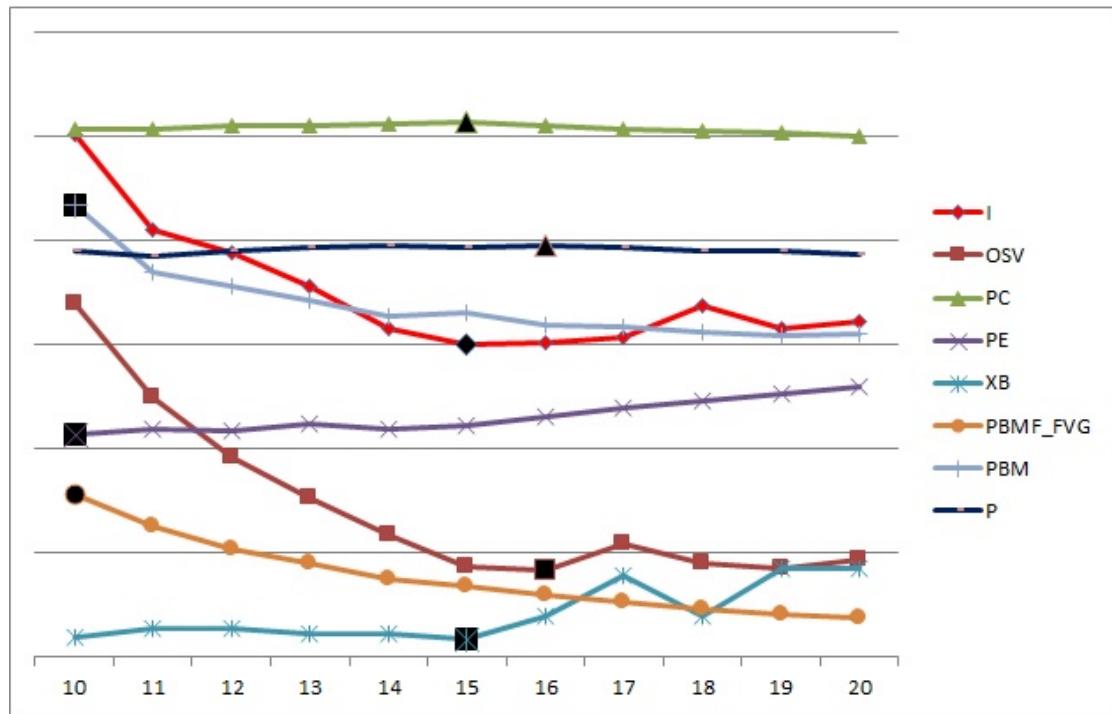


FIGURE 5.32: CVIs for S3 with different number of clusters (lines are scaled for display purpose)

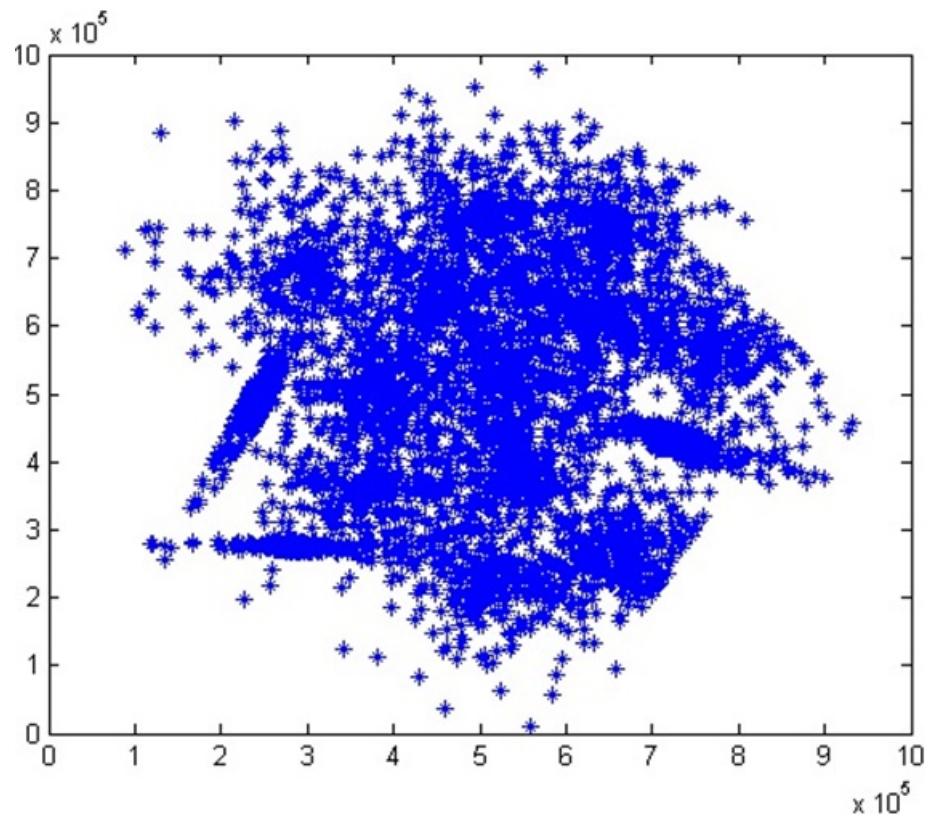


FIGURE 5.33: Data set S4 with 15 Gaussian clusters

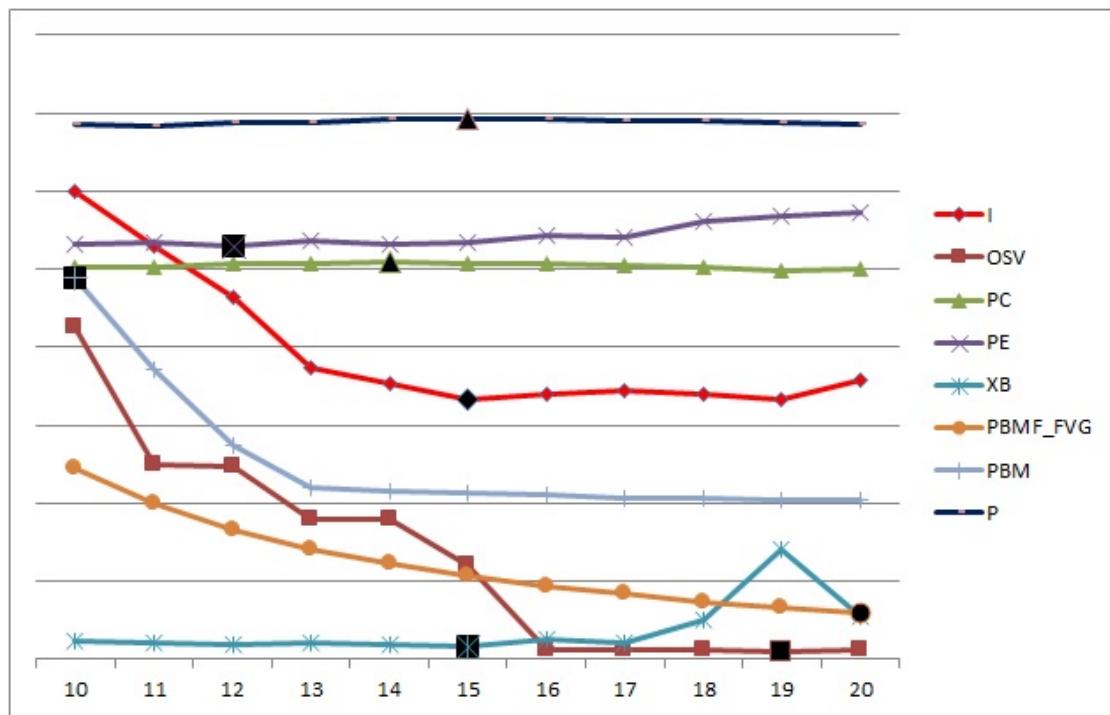


FIGURE 5.34: CVIs for S4 with different number of clusters (lines are scaled for display purpose)

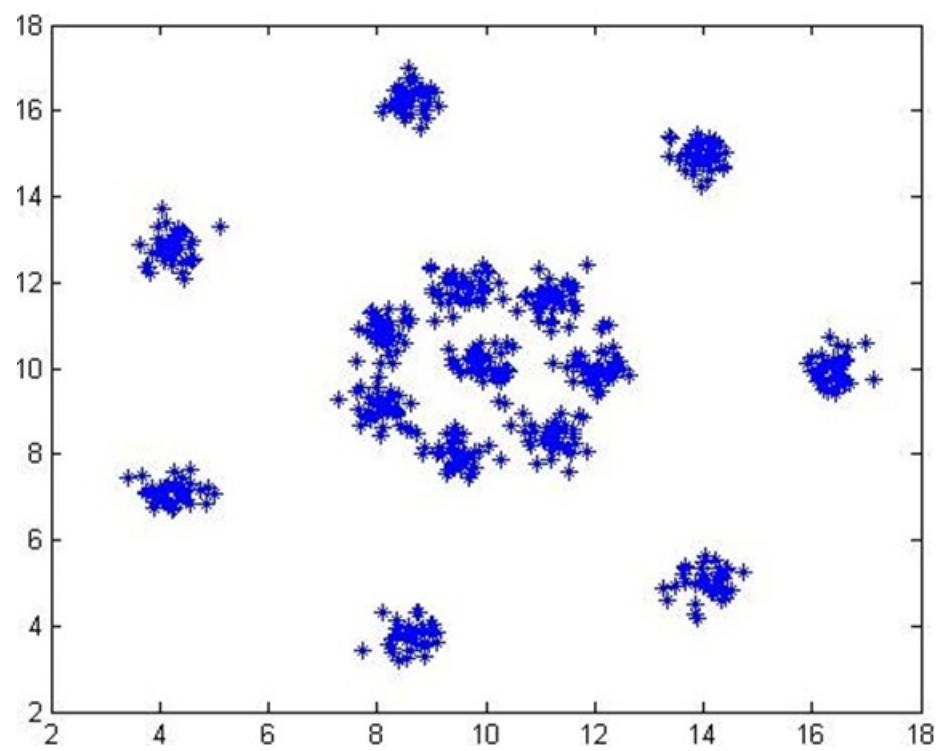


FIGURE 5.35: Data set R15 with 15 Gaussian clusters

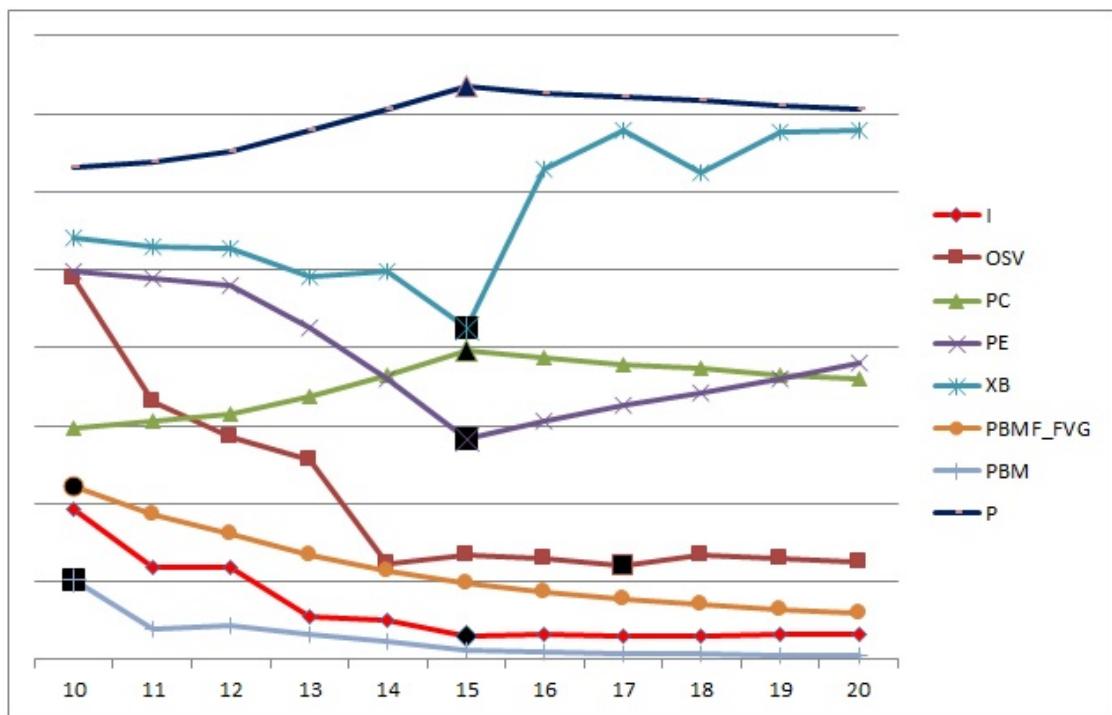


FIGURE 5.36: CVIs for R15 with different number of clusters (lines are scaled for display purpose)

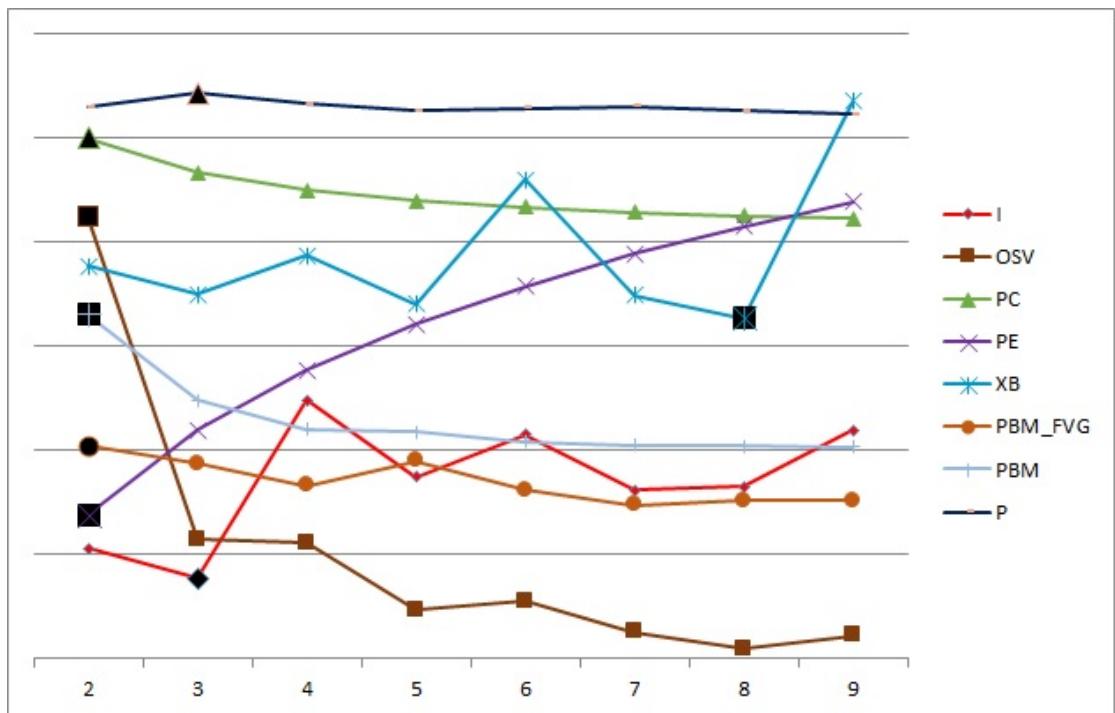


FIGURE 5.37: CVIs for Balance data set with different number of clusters (lines are scaled for display purpose)

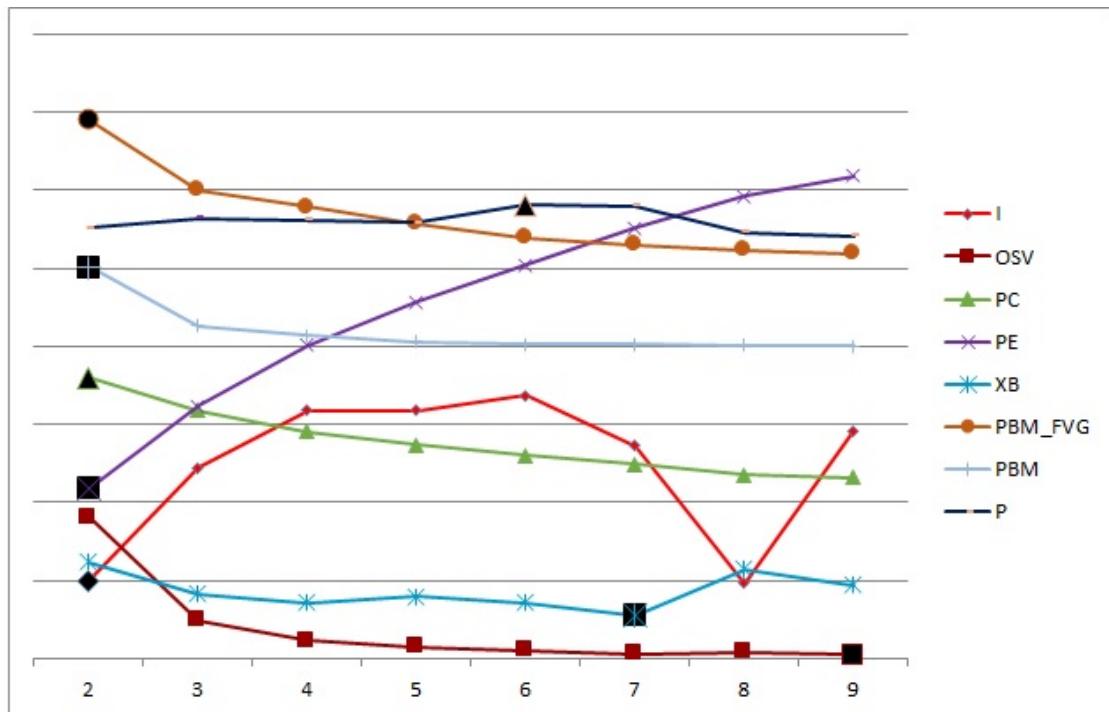


FIGURE 5.38: CVIs for Banana data set with different number of clusters (lines are scaled for display purpose)

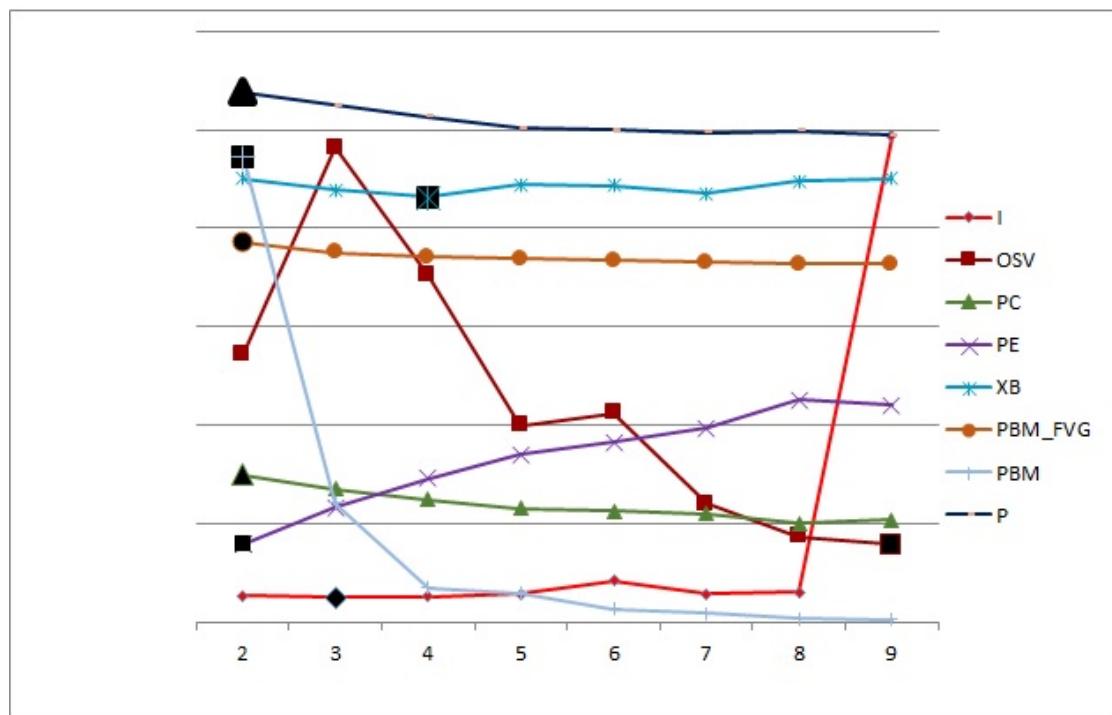


FIGURE 5.39: CVIs for Iris data set with different number of clusters (lines are scaled for display purpose)

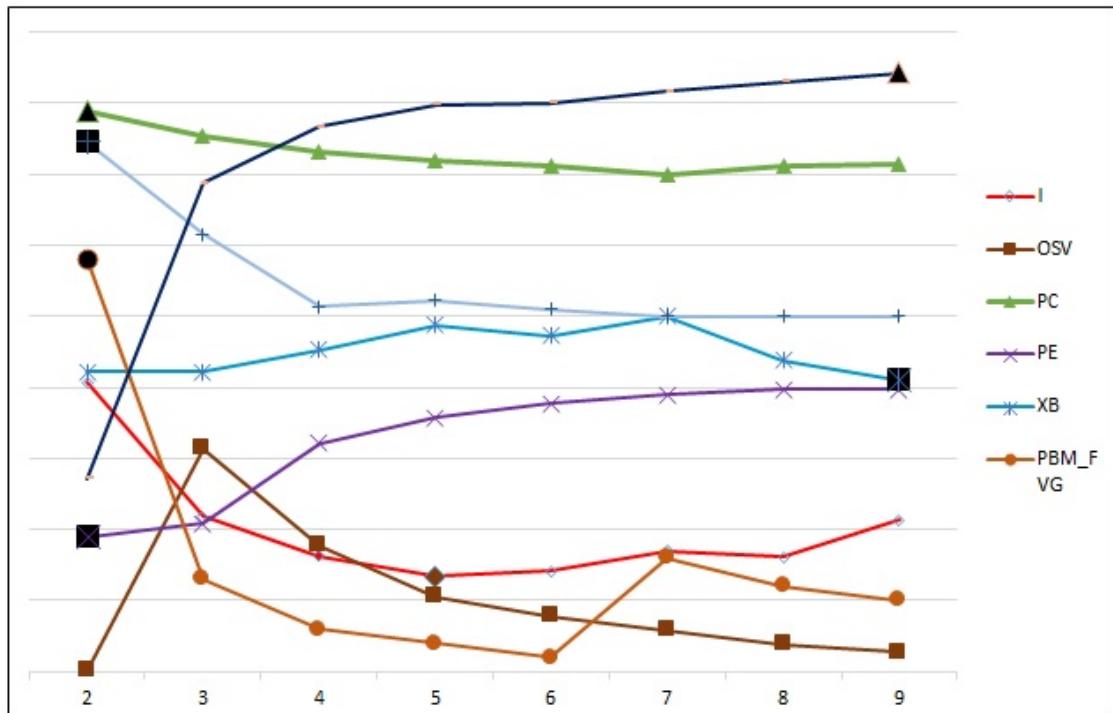


FIGURE 5.40: CVIs for Yale_5 data set with different number of clusters (lines are scaled for display purpose)

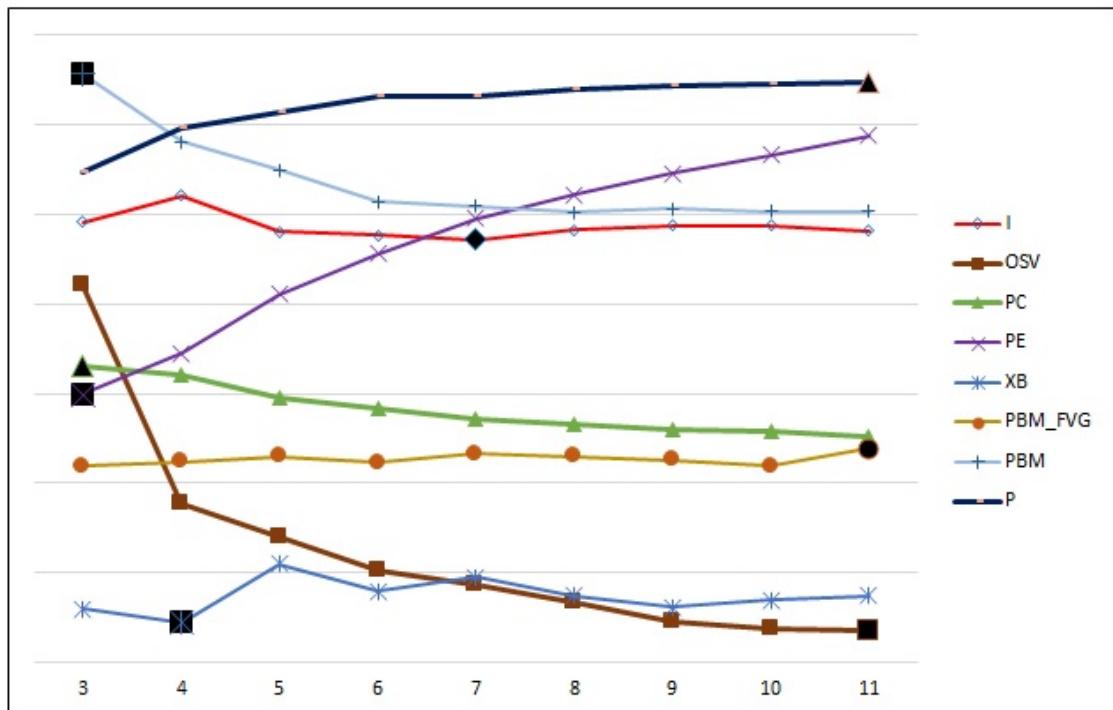


FIGURE 5.41: CVIs for Yale_7 with different number of clusters (lines are scaled for display purpose)

TABLE 5.21: Similarity measures for balance data set

Similarity measure	PC	PE	P	XB	PBM	PBM_FVG	OSV	I
Rand	0	0	0	0	0	0	75	9
Fowlkes-Mallows	100	100	0	0	100	100	0	69
Jaccard	100	100	0	0	100	100	0	67
Adjusted Rand	0	0	0	0	0	0	79	11

TABLE 5.22: Similarity measures for banana data set

Similarity measure	PC	PE	P	XB	PBM	PBM_FVG	OSV	I
Rand	0	81	8	58	0	0	29	0
FM	100	100	0	0	100	100	0	100
Jacc	100	100	0	0	100	100	0	100
ARI	100	0	0	0	0	0	0	100

TABLE 5.23: Similarity measures for Iris data set

Similarity measure	PC	PE	P	XB	PBM	PBM_FVG	OSV	I
Rand	0	0	0	0	0	0	0	17
FM	100	100	0	100	100	100	0	0
Jacc	0	0	0	0	0	0	0	15
ARI	0	0	0	0	0	0	0	84

TABLE 5.24: Similarity measures for 1R2RCT_A data set

Similarity measure	PC	PE	P	XB	PBM	PBM_FVG	OSV	I
Rand	0	0	0	0	0	0	0	80
FM	0	0	0	0	0	0	0	80
Jacc	0	0	0	0	0	0	0	80
ARI	100	100	0	0	100	100	0	0

TABLE 5.25: Similarity measures for 1R2RT_B data set

Similarity measure	PC	PE	P	XB	PBM	PBM_FVG	OSV	I
Rand	100	0	0	100	0	0	0	100
FM	100	0	0	100	0	0	0	100
Jacc	100	0	0	100	0	0	0	100
ARI	0	100	90	0	0	0	66	0

TABLE 5.26: Similarity measures for car5 data set

Similarity measure	PC	PE	P	XB	PBM	PBM_FVG	OSV	I
Rand	0	0	0	76	0	0	0	96
FM	41	41	0	61	0	41	41	82
Jacc	99	99	0	0	0	99	99	0
ARI	0	0	0	0	76	0	0	96

5.6 Impact of multispectral image denoising on segmentation

In this section, we deploy two scenarios. In the first one, we apply the segmentation algorithm GWFCM on multispectral image contaminated with additive Gaussian noise. In the second scenario, we apply one of the proposed filters to reduce the noise and then apply the GWFCM algorithm. In both scenarios, we calculate several CVIs (including the proposed ones) to assess the segmentation results. We choose the images illustrated in Fig. 5.11 and Fig. 5.13 and the proposed NLM filter for denoising. Table 5.27 and Table 5.28 illustrate the values of several CVIs used in previous experiments for subject 1. We can clearly see the difference in values between CVIs in presence and absence of noise. For example, with the proposed indexes, we have 20.2 for S and 5.57e+8 or I after segmenting the noisy image with $\sigma = 100$. After applying the NLM filter, we have 10.4 and 3.96e+7 for S and I respectively. Table 5.29 and Table 5.30 of subject 2 confirm the previous results. With the same noise conditions, we have S=40.1 and I=8.12e+7. However, with NLM filter, we have S=31.6 and I=5.59e+7.

Results of these experiments confirm the necessity of applying a denoising algorithm and how it can improve the segmentation results.

TABLE 5.27: CVIs values of the noisy segmented images for subject 1

	$\sigma = 5$	$\sigma = 10$	$\sigma = 20$	$\sigma = 30$	$\sigma = 40$	$\sigma = 50$	$\sigma = 100$
PC	0.71	0.63	0.53	0.50	0.50	0.49	0.53
PE	0.50	0.63	0.79	0.84	0.85	0.85	0.80
XB	0.12	0.19	0.36	0.40	0.35	0.30	0.25
PBM	43	78	79	94	104	111	221
CO	1.79	1.27	0.95	0.93	0.95	0.94	1.13
S	3.8	4.8	5.3	10.1	13.4	16.7	20.2
I	4.21e+5	1.78e+6	1.01e+7	1.93e+7	3.23e+7	6.67e+7	5.57e+8

TABLE 5.28: CVIs values of the denoised segmented images for subject 1

	$\sigma = 5$	$\sigma = 10$	$\sigma = 20$	$\sigma = 30$	$\sigma = 40$	$\sigma = 50$	$\sigma = 100$
PC	0.78	0.78	0.76	0.73	0.70	0.67	0.56
PE	0.38	0.39	0.43	0.47	0.52	0.57	0.74
XB	0.08	0.08	0.09	0.11	0.13	0.15	0.29
PBM	418	415	410	414	423	428	354
CO	2.56	2.47	2.23	1.93	1.69	1.49	1.06
S	0.5	1.3	2.1	3.3	5.7	8.3	10.4
I	7.56e+4	4.25e+5	1.46e+6	1.83e+6	3.78e+6	8.07e+6	3.96e+7

TABLE 5.29: CVIs values of the noisy segmented images for subject 2

	$\sigma = 5$	$\sigma = 10$	$\sigma = 20$	$\sigma = 30$	$\sigma = 40$	$\sigma = 50$	$\sigma = 100$
PC	0.74	0.71	0.64	0.58	0.54	0.51	0.49
PE	0.45	0.51	0.62	0.71	0.77	0.82	0.86
XB	0.11	0.13	0.19	0.23	0.33	0.46	0.33
PBM	331	319	305	247	361	459	538
CO	2.04	1.73	1.30	1.08	0.98	0.91	0.91
S	25.1	27.5	28.4	30.9	33.8	37.6	40.1
I	2.67e+5	9.49e+5	4.87e+6	8.56e+6	1.22e+7	5.01e+7	8.12e+7

TABLE 5.30: CVIs values of the denoised segmented images for subject 2

	$\sigma = 5$	$\sigma = 10$	$\sigma = 20$	$\sigma = 30$	$\sigma = 40$	$\sigma = 50$	$\sigma = 100$
PC	0.80	0.79	0.77	0.75	0.72	0.69	0.57
PE	0.36	0.37	0.41	0.74	0.50	0.54	0.72
XB	0.06	0.07	0.08	0.09	0.11	0.13	0.25
PBM	654	649	640	632	618	595	522
CO	2.83	2.72	2.42	2.10	1.83	1.62	1.11
S	20.3	22.1	25.4	27.9	28.1	29.2	31.6
I	7.44e+4	2.57e+5	1e+6	3.34e+6	5.47e+6	1.17e+7	5.59e+7

5.7 Conclusion

Our experiments have been conducted on several data sets from different databases with different complexity. The proposed denoising algorithms ensure a great tradeoff between noise removal and details preservation. The intrinsic property of the anisotropic Gaussian filter guarantees this important property. The NLM filter is also known by its high performance on Gaussian noise removal. Our approach for this task is based on extending these particular filters to the vector cases and we have been able to ensure good results in term of PSNR.

Segmentation experiments, which are similar to the approach used in cluster analysis, have proven the efficiency of the proposed algorithms. We have tried to adapt the gravitational theory, which is previously used in hierarchical clustering, to the partitional clustering. In addition, experiments on several machine learning data sets have proven the outperformance of the proposed CVIs. We have approached the problem of cluster validity index by using the n-sphere shape and probability density estimation.

Chapter 6

Conclusions and Future work

• Conclusion

The face recognition problem has been widely addressed in the general context of biometric applications. The emerging of multispectral imagery and its applications have contributed in advancing the face recognition. However, many technical and theoretical challenges have raised and should be faced to maximize the use of these images. The imaging system construction and database acquisition need particular attention. In the midst of this, the quality of multispectral images is an important factor which can be influenced by the surrounding conditions particularly the light and noise. Thus, enhancing the sensory data is a must. This enhancement involves noise suppression without a loss of important details and fine structures. We have adopted a noise model where the local variance of image is considered as negligible and that noise variance is the only source of variation within a local window in the noisy image. Our approach for image denoising is to extend the use of some filters to the vector case so that we gain benefit from the additional spectral information. We have used the anisotropic Gaussian filter known by its ability in edge preservation. We have also proposed another denoising algorithm based on the NLM filter. An optimization framework is adopted using the SURE method to choose the filter parameters. Experiments on multispectral images have proven the efficiency of the proposed algorithms in term of the Peak Signal to Noise Ratio.

We have also studied segmentation of multispectral images. This problem has been brought to the general context of cluster analysis. Hence, a novel clustering algorithm was proposed and applied to multispectral image segmentation. In addition, two CVIs are proposed using the n-sphere shape and probability density estimation. Experiments on several machine learning databases have been conducted and we have obtained results.

The problem addressed in this thesis is critical and has a great implication on the downstream processes of face recognition and should also get benefits from the upstream

processes particularly band selection. One of the major encountered issues is the high dimensionality of the data . Knowing that the information in the images is highly likely to be redundant, relevant spectral bands can be chosen resulting in dimensionality reduction and consequently alleviating the computation duration which is an important factor towards building a real time face recognition system.

• Future work

The usefulness of multispectral image has made it popular in many applications. This fact has contributed in advancing research activities in this topic. As we stated throughout this report, enhancing the sensory data is a critical stage toward the building of reliable and accurate face recognition system.

Possible future expansion of this work is to investigate deblurring techniques and propose new algorithms. Indeed, the change in wavelength passing through the filter causes different level of blurring in multispectral images. Applying unsharpness algorithm via Laplacian filter could be an interesting option. Particular attention will be also given to image restoration by deconvolution. Particularly, the adaptively estimated point spread function has shown great results. Extending it to deal with multispectral case could another future expansion of this work. The challenge here is to propose a good parameter selection since,in general, it is application dependant.

Another expansion of this work is to investigate other possible techniques for multispectral image segmentation such as the graph cuts method which is a reliable technique widely used for a classical 2D image segmentation. In addition, Contour detection methods are worth studying particularly the active contour models especially the parameterized one since we know what we need to detect in our images i.e. human face with all its features.

List of Publications

1. A. Ben Said, R. Hadjidj and S, Foufou, 'Cluster validity index based on Jeffrey divergence', *Springer Journal of Pattern Analysis and Applications*, pages: 1-11, Jan. 2015. DOI: 10.1007/s10044-015-0453-7
2. A. Ben Said, R. Hadjidj and Sebti Foufou, 'Optimized vector Non-Local Mean filter for multispectral image denoising', (Under revision)
3. A. Ben Said, R. Hadjidj and S. Foufou, 'Gravitational weighted fuzzy c-means with application on multispectral image segmentation', *4th International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages: 1-5, Paris-France, Oct. 2014. DOI: 10.1109/IPTA.2014.7001937
4. A. Ben Said, S, Foufou and M. Abidi, 'A FCM and SURF based algorithm for segmentation of Mmultispectral face images', *9th International Conference on Signal-Image Technology and Internet-Based Systems (SITIS)*, pages:65-70, Kyoto-Japan, Dec. 2013. DOI: 10.1109/ISBI.2004.1398786
5. A. Ben Said, S, Foufou and M. Abidi, 'Cluster validity index based on n-sphere', *11th ACS/IEEE International Conference on Computer Systems and Applications (AICSSA)*, pages: 144-147, Doha-Qatar, Nov. 2014
DOI: 2014. 10.1109/AICCSA.2014.7073191
6. A. Ben Said, S, Foufou and R. Hadjidj, 'Vector anisotropic Gaussian filter for multispectral image denoising', *12th International Conference on Quality Control by Artificial Vision (QCAV)*, Le Creusot-France, June 2015 (Accepted).
7. A. Ben Said, 'Multispectral face recognition: sensory data enhancement', Presented in *11th ACS/IEEE International Conference on Computer Systems and Applications (AICSSA) Ph.D. symposium*, Doha-Qatar, Nov. 2014

Bibliography

- [1] JonathanT. Erichsen and J.Margaret Woodhouse. Human and animal vision. In BruceG. Batchelor, editor, *Machine Vision Handbook*, pages 89–115. Springer London, 2012.
- [2] Jingu Heo. Fusion of visual and thermal face recognition techniques: A comparative study. In *University of Tennessee*, 2005.
- [3] Abdallah Meraoumia, Salim Chitroub, and Ahmed Bouridane. Biometric recognition systems using multispectral imaging. In Aboul Ella Hassanien, Tai-Hoon Kim, Janusz Kacprzyk, and Ali Ismail Awad, editors, *Bio-inspiring Cyber Security and Cloud Services: Trends and Innovations*, volume 70, pages 321–347. Springer Berlin Heidelberg, 2014.
- [4] Zhihong Pan, G. Healey, M. Prasad, and B. Tromberg. Face recognition in hyperspectral images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(12):1552–1560, Dec 2003.
- [5] Zhihong Pan, Glenn E. Healey, Manish Prasad, and Bruce J. Tromberg. Hyperspectral face recognition under variable outdoor illumination, 2004.
- [6] SeongG. Kong, Jingu Heo, Faysal Boughorbel, Yue Zheng, BesmaR. Abidi, Andreas Koschan, Mingzhong Yi, and MongiA. Abidi. Multiscale fusion of visible and thermal ir images for illumination-invariant face recognition. *International Journal of Computer Vision*, 71(2):215–233, 2007.
- [7] Seong G. Kong, Jingu Heo, Besma R. Abidi, Joonki Paik, and Mongi A. Abidi. Recent advances in visual and infrared face recognition—a review. *Computer Vision and Image Understanding*, 97(1):103 – 135, 2005.
- [8] Diego A. Socolinsky, Andrea Selinger, and Joshua D. Neuheisel. Face recognition with visible and thermal infrared imagery. *Computer Vision and Image Understanding*, 91(1–2):72 – 114, 2003.

- [9] I. Pavlidis and P. Symosek. The imaging issue in an automatic face/disguise detection system. In *Computer Vision Beyond the Visible Spectrum: Methods and Applications, 2000. Proceedings. IEEE Workshop on*, pages 15–24, 2000.
- [10] B. R. Corner, R. M. Narayanan, and S. E. Reichenbach. Noise estimation in remote sensing imagery using data masking. *International Journal of Remote Sensing*, 24(4):689–702, 2003.
- [11] Yi Chen, SaratC. Dass, and AnilK. Jain. Fingerprint quality indices for predicting authentication performance. In Takeo Kanade, Anil Jain, and NaliniK. Ratha, editors, *Audio- and Video-Based Biometric Person Authentication*, volume 3546 of *Lecture Notes in Computer Science*, pages 160–170. Springer Berlin Heidelberg, 2005.
- [12] H. Fronthaler, K. Kollreider, and J. Bigun. Automatic image quality assessment with application in biometrics. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06. Conference on*, pages 30–30, June 2006.
- [13] D. Nouri, Y. Lucas, and S. Treuillet. Infrared hyperspectral imaging for enhanced tissue visualization and discrimination during surgical operation. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 5137–5141, Oct 2014.
- [14] Dorra Nouri, Yves Lucas, and Sylvie Treuillet. Efficient tissue discrimination during surgical interventions using hyperspectral imaging. In Danail Stoyanov, D.Louis Collins, Ichiro Sakuma, Purang Abolmaesumi, and Pierre Jannin, editors, *Information Processing in Computer-Assisted Interventions*, volume 8498 of *Lecture Notes in Computer Science*, pages 266–275. Springer International Publishing, 2014. ISBN 978-3-319-07520-4.
- [15] Dorra Nouri, Yves Lucas, and Sylvie Treuillet. Calibration and test of a hyperspectral imaging prototype for intra-operative surgical assistance. In *Proc. SPIE 8676, Medical Imaging 2013: Digital Pathology*, page 86760P, February 2013.
- [16] Dorra Nouri, Yves Lucas, Sylvie Treuillet, Romuald Jolivot, and Franck Marzani. Colour and multispectral imaging for wound healing evaluation in the context of a comparative preclinical study. In *Medical Imaging 2013: Image Processing*, page 866923, February 2013.
- [17] Darren Roblyer, Cristina Kurachi, Vanda Stepanek, Richard A. Schwarz, Michelle D. Williams, Adel K. El-Naggar, J. Jack Lee, Ann M. Gillenwater, and

- Rebecca Richards-Kortum. Comparison of multispectral wide-field optical imaging modalities to maximize image contrast for objective discrimination of oral neoplasia. *Journal of Biomedical Optics*, 15(6):066017–066017–9, 2010.
- [18] M. Ruiz-Altisent, L. Ruiz-Garcia, G.P. Moreda, Renfu Lu, N. Hernandez-Sanchez, E.C. Correa, B. Diezma, B. Nicolaï, and J. García-Ramos. Sensors for product characterization and quality of specialty crops—a review. *Computers and Electronics in Agriculture*, 74(2):176–194, 2010.
- [19] Devrim Unay. *Multispectral Image Processing and Pattern Recognition Techniques for Quality Inspection of Apple Fruits*. PhD thesis, Catholic University of Louvain, 2006.
- [20] Takeo Kanade. Picture processing system by computer complex and recognition of human faces. In *doctoral dissertation, Kyoto University*. November 1973.
- [21] Ralph Gross, Jianbo Shi, and Jeffrey Cohn. Quo vadis face recognition? Technical Report CMU-RI-TR-01-17, Robotics Institute, Pittsburgh, PA, June 2001.
- [22] P.J. Phillips, Hyeyonjoon Moon, S.A. Rizvi, and P.J. Rauss. The feret evaluation methodology for face-recognition algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(10):1090–1104, Oct 2000.
- [23] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone. Face recognition vendor test 2002: Evaluation report, 2003.
- [24] Mark C. Abrams and Stephen C. Cain. Sampling, radiometry, and image reconstruction for polar and geostationary meteorological remote sensing systems, 2002.
- [25] B.R. Corner, R.M. Narayanan, and S.E. Reichenbach. Noise estimation in remote sensing imagery using data masking. *International Journal of Remote Sensing*, 24:689–702, 2003.
- [26] F. Luisier and T. Blu. Sure-let multichannel image denoising: Interscale orthonormal wavelet thresholding. *Image Processing, IEEE Transactions on*, 17(4):482–492, April 2008.
- [27] Thierry Blu and Florian Luisier. The SURE-LET Approach to Image Denoising. *IEEE Transactions on Image Processing*, 16:2778–2786, 2007.
- [28] Charles M. Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 9(6):pp. 1135–1151, 1981.

- [29] C. Chaux, L. Duval, A. Benazza-Benyahia, and J. Pesquet. A nonlinear stein-based estimator for multichannel image denoising. *Signal Processing, IEEE Transactions on*, 56(8):3855–3870, Aug 2008.
- [30] A. Duijster, P. Scheunders, and S. De Backer. Wavelet-based em algorithm for multispectral-image restoration. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(11):3892–3898, Nov 2009.
- [31] Robert V Hogg and Allen T Craig. *Introduction to mathematical statistics; 4th ed.* Macmillan, New York, NY, 1978.
- [32] Peng Liu, Fang Huang, Guoqing Li, and Zhiwen Liu. Remote-sensing image denoising using partial differential equations and auxiliary images as priors. *Geoscience and Remote Sensing Letters, IEEE*, 9(3):358–362, May 2012.
- [33] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, 2004. ISSN 0924-9907.
- [34] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, November 1992.
- [35] Jie Li, Qiangqiang Yuan, Huanfeng Shen, and Liangpei Zhang. Hyperspectral image recovery employing a multidimensional nonlocal total variation model. *Signal Processing*, 111(0):230 – 248, 2015.
- [36] Xiaoqun Zhang, Martin Burger, Xavier Bresson, and Stanley Osher. Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. *SIAM Journal on Imaging Sciences*, 3(3):253–276, 2010.
- [37] Yong-Qiang Zhao and Jingxiang Yang. Hyperspectral image denoising via sparse representation and low-rank constraint. *Geoscience and Remote Sensing, IEEE Transactions on*, 53(1):296–308, Jan 2015.
- [38] Qiangqiang Yuan, Liangpei Zhang, and Huanfeng Shen. Hyperspectral image denoising with a spatial-spectral view fusion strategy. *Geoscience and Remote Sensing, IEEE Transactions on*, 52(5):2314–2325, May 2014.
- [39] Xiang Zhu and P. Milanfar. Automatic parameter selection for denoising algorithms using a no-reference measure of image content. *Image Processing, IEEE Transactions on*, 19(12):3116–3132, Dec 2010.
- [40] Y. Yuan, X. Zheng, and X. Lu. Spectral-spatial kernel regularized for hyperspectral image denoising. *Geoscience and Remote Sensing, IEEE Transactions on*, 53(7):3815–3832, July 2015.

- [41] L. Xu, F. Li, A. Wong, and D. Clausi. Hyperspectral image denoising using a spatial-spectral monte carlo sampling approach. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, PP(99):1–14, 2015.
- [42] Christiane Lemieux. *Monte Carlo and Quasi-Monte Carlo Sampling*. Springer-Verlag New York, 2009.
- [43] Alexander Wong, Akshaya Mishra, Wen Zhang, Paul Fieguth, and D. A. Clausi. Stochastic image denoising based on markov-chain monte carlo sampling. *Signal Processing*, 91:2112 – 2120, 2011.
- [44] Honghong Peng and R. Rao. Hyperspectral image enhancement with vector bilateral filtering. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 3713–3716, Nov 2009.
- [45] P. J. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection*. John Wiley & Sons, Inc., New York, NY, USA, 1987. ISBN 0-471-85233-3.
- [46] Honghong Peng, R. Rao, and S.A. Dianat. Optimized vector bilateral filter for multispectral image denoising. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 2141–2144, Sept 2012.
- [47] Honghong Peng, R. Rao, and S.A. Dianat. Multispectral image denoising with optimized vector bilateral filter. *Image Processing, IEEE Transactions on*, 23(1):264–273, Jan 2014.
- [48] J. V. Manjon, M. Robles, and N. A. Thacker. Multispectral mri denoising using nonlocal means. *Medical Image Understanding and Analysis*, pages 41–46, 2007.
- [49] A. Buades, B. Coll, and J. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005.
- [50] Thanh N. Tran, Ron Wehrens, and Lutgarde M.C. Buydens. Clustering multispectral images: a tutorial. *Chemometrics and Intelligent Laboratory Systems*, 77:3–17, 2005.
- [51] Thanh N. Tran, Ron Wehrens, and Lutgarde M.C. Buydens. Sparef: a clustering algorithm for multispectral images. *Analytica Chimica Acta*, 490(1–2):303 – 312, 2003.
- [52] Brendan J. Frey and Delbert Dueck. Clustering by passing messages between data points. *Science*, 315(5814):972–976, 2007.
- [53] M. Soltani and K. Chehdi. *Affinity Propagation for partitioning of hyperspectral image of high spatial dimension*. PhD thesis, Rennes 1 University, 2014.

- [54] M Soltani, K. Chehdi, and C. Cariou. Affinity propagation for large size hyperspectral image classification. *Proc. SPIE Image and Signal Processing for Remote Sensing*, 8892:88920P–88920P–10, 2013.
- [55] Nhat Bao Sinh Vu. *Image Segmentation with Semantic Priors: A Graph Cut Approach*. PhD thesis, University of California, Santa Barbara, Sep 2008.
- [56] Azriel Rosenfeld and Avinash C. Kak. *Digital Picture Processing*. Academic Press, Inc., Orlando, FL, USA, 2nd edition, 1982.
- [57] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: A review. *ACM Comput. Surv.*, 31(3):264–323, September 1999.
- [58] Charles Romesburg. *Cluster Analysis for Researchers*. Lulu Press, 2004.
- [59] Brian S. Everitt, Sabine Landau, Morven Leese, and Daniel Stahl. *Cluster Analysis*. John Wiley & Sons Ltd, 2011.
- [60] Anil K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8):651 – 666, 2010.
- [61] David J. Hand. Data clustering: Theory, algorithms, and applications by guojun gan, chaoqun ma, jianhong wu. *International Statistical Review*, 76(1):141–141, 2008.
- [62] Vincent Barra, Véronique Delouille, and Jean-François Hochedez. Segmentation of extreme ultraviolet solar images via multichannel fuzzy clustering. *Advances in Space Research*, 42(5):917 – 925, 2008.
- [63] R. Krishnapuram and J.M. Keller. The possibilistic c-means algorithm: insights and recommendations. *Fuzzy Systems, IEEE Transactions on*, 4(3):385–393, Aug 1996.
- [64] M. Hasanzadeh and S. Kasaei. A multispectral image segmentation method using size-weighted fuzzy clustering and membership connectedness. *Geoscience and Remote Sensing Letters, IEEE*, 7(3):520–524, July 2010.
- [65] Chen Yang, L. Bruzzone, Fengyue Sun, Laijun Lu, Renchu Guan, and Yanchun Liang. A fuzzy-statistics-based affinity propagation technique for clustering in multispectral images. *Geoscience and Remote Sensing, IEEE Transactions on*, 48(6):2647–2659, June 2010.
- [66] A. Marcal and L. Castro. Hierarchical clustering of multispectral images using combined spectral and spatial criteria. *Geoscience and Remote Sensing Letters, IEEE*, 2(1):59–63, Jan 2005.

- [67] C. Cariou and K. Chehdi. Unsupervised nearest neighbors clustering with application to hyperspectral images. *IEEE Journal of Selected Topics in Signal Processing*, PP(99):1–12, 2015.
- [68] J. Li, J.M. Bioucas-Dias, and A. Plaza. Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *Geoscience and Remote Sensing, IEEE Transactions on*, 48(11):4085–4098, Nov 2010.
- [69] Dankmar Böhning. Multinomial logistic regression algorithm. *Annals of the Institute of Statistical Mathematics*, 44(1):197–200, 1992.
- [70] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1222–1239, Nov 2001.
- [71] Jun Li, J.M. Bioucas-Dias, and A. Plaza. Hyperspectral image segmentation using a new bayesian approach with active learning. *Geoscience and Remote Sensing, IEEE Transactions on*, 49(10):3947–3960, Oct 2011.
- [72] J. Bioucas Dias and M. Figueiredo. Logistic regression via variable splitting and augmented lagrangian tools. Technical report, Instituto Superior Técnico, TULisbon, Lisbon, Portugal, 2009.
- [73] Jun Li, J.M. Bioucas-Dias, and A. Plaza. Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields. *Geoscience and Remote Sensing, IEEE Transactions on*, 50(3):809–823, March 2012.
- [74] G. Bilgin, S. Erturk, and T. Yildirim. Segmentation of hyperspectral images via subtractive clustering and cluster validation using one-class support vector machines. *Geoscience and Remote Sensing, IEEE Transactions on*, 49(8):2936–2944, Aug 2011.
- [75] Bernhard Schölkopf, John C. Platt, John C. Shawe-Taylor, Alex J. Smola, and Robert C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Comput.*, 13(7):1443–1471, July 2001.
- [76] Pabitra Mitra, B. Uma Shankar, and Sankar K. Pal. Segmentation of multispectral remote sensing images using active support vector machines. *Pattern Recognition Letters*, 25(9):1067–1074, 2004.
- [77] L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 13(6):583–598, Jun 1991.

- [78] Y. Tarabalka, J. Chanussot, and J.A. Benediktsson. Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recognition*, 43(7):2367–2379, 2010.
- [79] James C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press. New York and London, 1981.
- [80] James C. Bezdek. Cluster validity with fuzzy sets. *Journal of Cybernetics*, 3: 58–73, 1974.
- [81] Rajesh N. Dave. Validating fuzzy partitions obtained through c-shells clustering. *Pattern Recognition Letters*, 17(6):613 – 623, 1996.
- [82] Min-You Chen and D.A. Linkens. Rule-base self-generation and simplification for data-driven fuzzy models. *Fuzzy Sets and Systems*, 142:243–265, 2004.
- [83] Krista Rizman Žalik. Cluster validity index for estimation of fuzzy clusters of different sizes and densities. *Pattern Recognition*, 43(10):3374 – 3390, 2010.
- [84] Fukuyama and Y. Sugeno. A new method of choosing the number of clusters for the fuzzy c-means method. In: *Proceedings of 5th Fuzzy System Symposium*, pages 247–250, 1989.
- [85] Xuanli Lisa Xie and Gerardo Beni. A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:841–847, 1991.
- [86] Malay K. Pakhira, Sanghamitra Bandyopadhyay, and Ujjwal Maulik. Validity index for crisp and fuzzy clusters. *Pattern Recognition*, 37:487 – 501, 2004.
- [87] Krista Rizman Žalik and Borut Žalik. Validity index for clusters of different sizes and densities. *Pattern Recognition Letters*, 32:221–234, 2011.
- [88] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.
- [89] Christoph H. Lampert and O. Wirjadi. An optimal nonorthogonal separation of the anisotropic gaussian convolution filter. *Image Processing, IEEE Transactions on*, 15(11):3501–3513, Nov 2006.
- [90] C.H. Lampert and O. Wirjadi. Anisotropic gaussian filtering using fixed point arithmetic. In *Image Processing, 2006 IEEE International Conference on*, pages 1565–1568, Oct 2006.
- [91] S. Yiu Man Lam and B.E. Shi. Recursive anisotropic 2-d gaussian filtering based on a triple-axis decomposition. *Image Processing, IEEE Transactions on*, 16(7): 1925–1930, July 2007.

- [92] J. M. Geusebroek, A. W. M. Smeulders, and J. van de Weijer. Fast anisotropic gauss filtering. *IEEE Transactions on Image Processing*, 12(8):938–943, 2003.
- [93] J.-M. Geusebroek, A.W.M. Smeulders, and J. van de Weijer. Fast anisotropic gauss filtering. *Image Processing, IEEE Transactions on*, 12(8):938–943, Aug 2003.
- [94] Pietro Perona. Steerable-scalable kernels for edge detection and junction analysis. *Image Vision Comput.*, 10(10):663–672, December 1992.
- [95] Guangzhi Cao, L.R. Bachega, and C.A. Bouman. The sparse matrix transform for covariance estimation and analysis of high dimensional signals. *Image Processing, IEEE Transactions on*, 20(3):625–640, March 2011.
- [96] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 60–65 vol. 2, June 2005.
- [97] Xuming Zhang, Yi Zhan, Mingyue Ding, Wenguang Hou, and Zhouping Yin. Decision-based non-local means filter for removing impulse noise from digital images. *Signal Processing*, 93(2):517 – 524, 2013.
- [98] P. Coupe, P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot. An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images. *Medical Imaging, IEEE Transactions on*, 27(4):425–441, April 2008.
- [99] Y. Matsushita and S. Lin. A probabilistic intensity similarity measure based on noise distributions. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8, June 2007.
- [100] P.J. Huber, J. Wiley, and W. InterScience. *Robust statistics*. Wiley New York, 1981.
- [101] Chenglong Tang, Shigang Wang, and Wei Xu. New fuzzy c-means clustering model based on the data weighted approach. *Data & Knowledge Engineering*, 69(9):881 – 900, 2010.
- [102] Vicenç Torra and Sadaaki Miyamoto. On the use of variable-size fuzzy clustering for classification. In Vicenç Torra, Yasuo Narukawa, Aïda Valls, and Josep Domingo-Ferrer, editors, *Modeling Decisions for Artificial Intelligence*, volume 3885 of *Lecture Notes in Computer Science*, pages 362–371. Springer Berlin Heidelberg, 2006.
- [103] Xizhao Wang, Yadong Wang, and Lijuan Wang. Improving fuzzy c-means clustering based on feature-weight learning. *Pattern Recognition Letters*, 25(10):1123 – 1132, 2004.

- [104] Wen-Liang Hung, Miin-Shen Yang, and De-Hua Chen. Bootstrapping approach to feature-weight selection in fuzzy c-means algorithms with an application in color image segmentation. *Pattern Recognition Letters*, 29(9):1317 – 1325, 2008.
- [105] Wen-Liang Hung, Yen-Chang Chang, and E. Stanley Lee. Weight selection in w-k-means algorithm with an application in color image segmentation. *Computers & Mathematics with Applications*, 62(2):668–676, July 2011.
- [106] Elaine Y. Chan, Wai Ki Ching, Michael K. Ng, and Joshua Z. Huang. An optimization algorithm for clustering using weighted dissimilarity measures. *Pattern Recognition*, 37(5):943 – 952, 2004.
- [107] S. Turmchokkasam and S. Mitaim. Effects of weights in weighted fuzzy c-means algorithm for room equalization at multiple locations. In *Fuzzy Systems, 2006 IEEE International Conference on*, pages 1468–1475, 2006.
- [108] W. E. Wright. Gravitational clustering. *Pattern Recognition*, 9:151–166, 1977.
- [109] Esmat Rashedi and Hossein Nezamabadi-Pour. A stochastic gravitational approach to feature based color image segmentation. *Engineering Applications of Artificial Intelligence*, 26(4):1322–1332, April 2013.
- [110] Chien-Yu Chen, Shien-Ching Hwang, and Yen-Jen Oyang. A statistics-based approach to control the quality of subclusters in incremental gravitational clustering. *Pattern Recognition*, 38(12):2256 – 2269, 2005.
- [111] H. C. Yung and H. S. Lai. Segmentation of color images based on the gravitational clustering concept. *Optical Engineering*, 37(3):989–1000, 1998.
- [112] A. Ben Said, R. Hadjidj, and S. Foufou. Gravitational weighted fuzzy c-means with application on multispectral image segmentation. In *Image Processing Theory, Tools and Applications (IPTA), 2014 4th International Conference on*, pages 1–5, Oct 2014.
- [113] I. Dokmanić and D. Petrinović. Convolution on the n -sphere with application to pdf modeling. *IEEE Transactions on Signal Processing*, 58:1157–1170, 2010.
- [114] S. Li. Concise formulas for the area and volume of a hyperspherical cap. *Asian Journal of Mathematics & Statistics*, 4:66–70, 2011.
- [115] Eric P. Xing, Andrew Y. Ng, Michael I. Jordan, and Stuart Russell. Distance metric learning, with application to clustering with side-information. In *Advances in Neural Information Processing systems 15*, pages 505–512. MIT Press, 2003.

- [116] Yang Mu, Wei Ding, and Dacheng Tao. Local discriminative distance metrics ensemble learning. *Pattern Recognition*, 46(8):2337 – 2349, 2013.
- [117] Jacob Goldberger, Geoffrey E. Hinton, Sam T. Roweis, and Ruslan Salakhutdinov. Neighbourhood components analysis. In L.K. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems 17*, pages 513–520. MIT Press, 2005.
- [118] Masashi Sugiyama. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *J. Mach. Learn. Res.*, 8:1027–1061, May 2007. ISSN 1532-4435.
- [119] Kilian Q. Weinberger, John Blitzer, and Lawrence K. Saul. Distance metric learning for large margin nearest neighbor classification. In *In NIPS*. MIT Press, 2006.
- [120] Deza, Michel Marie, Deza, and Elena. *Encyclopedia of Distances*. Springer Heidelberg New York Dordrecht London, 2009.
- [121] Hong Chang, Yi Yao, A. Koschan, B. Abidi, and M. Abidi. Improving face recognition via narrowband spectral range selection using jeffrey divergence. *Information Forensics and Security, IEEE Transactions on*, 4(1):111–122, March 2009.
- [122] J. Zheng and H. You. A new model-independent method for change detection in multitemporal sar images based on radon transform and jeffrey divergence. *Geoscience and Remote Sensing Letters, IEEE*, 10(1):91–95, Jan 2013.
- [123] J. Puzicha, T. Hofmann, and J.M. Buhmann. Non-parametric similarity measures for unsupervised texture segmentation and image retrieval. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1997*, pages 267–272, Jun 1997.
- [124] Papoulis Athanasios. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill Companies, 3rd edition, February 1991.
- [125] Ahmed Ben Said, Rachid Hadjidj, and Sebti Foufou. Cluster validity index based on jeffrey divergence. *Pattern Analysis and Applications*, pages 1–11, 2015.
- [126] H. Chang, H. Harishwaran, M. Yi, A. Koschan, B. Abidi, and M. Abidi. An indoor and outdoor, multimodal, multispectral and multi-illuminant database for face recognition. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06. Conference on*, pages 54–54, June 2006.
- [127] Pasi Fränti and Olli Virmajoki. Iterative shrinking method for clustering problems. *Pattern Recognition*, 39(5):761–775, 2006.

- [128] C.J. Veenman, M.J.T. Reinders, and E. Backer. A maximum variance cluster algorithm. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(9):1273–1280, Sep 2002.
- [129] A.S. Georghiades, P.N. Belhumeur, and D. Kriegman. From few to many: illumination cone models for face recognition under variable lighting and pose. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(6):643–660, Jun 2001.
- [130] Anil K. Jain and Richard C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988. ISBN 0-13-022278-X.
- [131] A. Ben Said, S. Foufou, and M. Abidi. A fcm and surf based algorithm for segmentation of multispectral face images. In *Signal-Image Technology Internet-Based Systems (SITIS), 2013 International Conference on*, pages 65–70, Dec 2013.
- [132] Ibai Gurrutxaga, Javier Muguerza, Olatz Arbelaitz, Jesús M. Pérez, and José I. Martín. Towards a standard methodology to evaluate internal cluster validity indices. *Pattern Recognition Letters*, 32(3):505 – 515, 2011.
- [133] Sanghamitra Bandyopadhyay, Sriparna Saha, and Witold Pedrycz. Use of a fuzzy granulation–degranulation criterion for assessing cluster validity. *Fuzzy Sets and Systems*, 170(1):22–42, 2011.
- [134] E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(11):2765–2781, Nov 2013.