HAL
open science

Classification d'images RSO polarimétriques à haute résolution spatiale sur site urbain

Maryam Soheili Majd

To cite this version:

Maryam Soheili Majd. Classification d'images RSO polarimétriques à haute résolution spatiale sur site urbain. Sciences de la Terre. Conservatoire national des arts et metiers - CNAM, 2014. Français. NNT : 2014CNAM0927 . tel-01126995

HAL Id: tel-01126995

https://theses.hal.science/tel-01126995v1

Submitted on 6 Mar 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

le cnam	CONSERVATOIRE NATIONAL DES ARTS ET METIERS

ÉCOLE DOCTORALE 415

Laboratoire de Géodésie et Géomatique (L2G)

THÈSE présentée par :

Maryam SOHEILI MAJD

Soutenue le: 28 Avril 2014

pour obtenir le grade de : **Docteur du Conservatoire National des Arts et Métiers**

Spécialité: Géomatique

Classification d'images RSO polarimétriques à haute résolution spatiale sur site urbain

"High – Resolution Polarimetric SAR image classification on urban areas"

THÈSE dirigée par :

M.POLIDORI Laurent	Professeur CNAM, Directeur de thèse
Mme SIMONETTO Elisabeth	Maître de conférences, CNAM

RAPPORTEURS :

M. NICOLAS Jean-Marie	Professeur, Télécom ParisTech
M. FERRO-FAMIL Laurent	Professeur des Universités, Université de Rennes 1

JURY :

Mme TUPIN Florence	Professeur, Télécom ParisTech
Mme ORIOT Hélène	Ingénieur, ONERA
M. RUDANT Jean-Paul	Professeur des Universités, Université Paris-Est Marne-la-Vallée

À mes parents

À mes sœurs

Remerciements

Ces travaux de doctorat ont été réalisés au sein du Laboratoire de Géodésie et Géomatique (CNAM/ESGT, Le Mans).

Je tiens ici à remercier toutes les personnes qui m'ont accompagnée, au cours de ces trois dernières années, dans cette grande aventure, scientifique et humaine.

Je remercie tout particulièrement mon directeur de thèse, M. Laurent Polidori**,** Professeur CNAM, et Mme Elisabeth Simonetto, Maître de conférences, CNAM, qui m'ont tous les deux encadrée durant la préparation de ma thèse.

J'ai pris un très grand plaisir à travailler avec eux et je les remercie très vivement pour la qualité de leur encadrement, leur investissement permanent, leurs précieux conseils, leur écoute, leur bienveillance et leur très grande disponibilité. Je leur exprime ici ma gratitude pour toute l'aide qu'ils m'ont apportée au cours des différentes étapes de cette thèse.

Je remercie très sincèrement M. Jean-Marie Nicolas**,** Professeur Télécom Paris Tech, ainsi que M. Laurent Ferro-Famil, Professeur des Universités, Université de Rennes 1, qui ont accepté de rapporter cette thèse, leurs remarques constructives et leurs avis éclairés ont beaucoup inspiré la progression de mes travaux et la finalisation du manuscrit.

Je remercie également les examinateurs de mon jury, Mme Florence Tupin, Professeur Télécom ParisTech, Mme Hélène Oriot, Ingénieur, ONERA, et M. Jean-Paul Rudant**,** Professeur des Universités, Université Paris-Est Marne-la-Vallée pour toute l'attention et l'intérêt qu'ils ont porté à ces travaux de recherche.

Enfin, je remercie avec un grand plaisir et beaucoup de gratitude mes parents, mes soeurs et toute ma famille, pour leur soutien constant et essentiel durant toutes mes années d'études, pour leur grande patience, leur vif intérêt pour mes travaux, leur confiance absolue et tout leur amour.

Résumé

Notre recherche vise à évaluer l'apport d'une seule image polarimétrique RSO (Radar à Synthèse d'Ouverture) à haute résolution spatiale pour classifier les surfaces urbaines. Pour cela, nous définissons plusieurs types de toits, de sols et d'objets.

Dans un premier temps, nous proposons un inventaire d'attributs statistiques, texturaux et polarimétriques pouvant être utilisés dans un algorithme de classification. Nous étudions les lois statistiques des descripteurs et montrons que la distribution de Fisher est bien adaptée pour la plupart d'entre eux.

Dans un second temps, plusieurs algorithmes de classification vectorielle supervisée sont testés et comparés, notamment la classification par maximum de vraisemblance basée sur une distribution gaussienne, ou celle basée sur la distribution de Wishart comme modèle statistique de la matrice de cohérence polarimétrique, ou encore l'approche SVM.

Nous proposons alors une variante de l'algorithme par maximum de vraisemblance basée sur une distribution de Fisher, dont nous avons étudié l'adéquation avec l'ensemble de nos attributs. Nous obtenons une nette amélioration de nos résultats avec ce nouvel algorithme mais une limitation apparaît pour reconnaître certains toits.

Ainsi, la forme des bâtiments rectangulaires est reconnue par opérations morphologiques à partir de l'image d'amplitude radar. Cette information spatiale est introduite dans le processus de classification comme contrainte. Nous montrons tout l'intérêt de cette information puisqu'elle empêche la confusion de classification entre pixels situés sur des toits plats et des pixels d'arbre.

De plus, nous proposons une méthode de sélection des attributs les plus pertinents pour la classification, basée sur l'information mutuelle et une recherche par algorithme génétique.

Nos expériences sont menées sur une image polarimétrique avec un pixel de 35 cm, acquise en 2006 par le capteur aéroporté RAMSES de l'ONERA.

Mots clés

Zone urbaine, classification d'images, maximum de vraisemblance, Fisher, haute résolution spatiale, polarimétrie radar, algorithme génétique, information mutuelle.

Abstract

In this research, our aim is to assess the potential of a one single look high spatial resolution polarimetric radar image for the classification of urban areas. For that purpose, we concentrate on classes corresponding to different kinds of roofs, objects and ground surfaces.

At first, we propose a uni-variate statistical analysis of polarimetric and texture attributes, that can be used in a classification algorithm. We perform a statistical analysis of descriptors and show that the Fisher distribution is suitable for most of them. We then propose a modification of the maximum likelihood algorithm based on a Fisher distribution; we train it with all of our attributes. We obtain a significant improvement in our results with the new algorithm, but a limitation appears to recognize some roofs.

Then, the shape of rectangular buildings is recognized by morphological operations from the image of radar amplitude. This spatial information is introduced in a Fisher-based classification process as a constraint term and we show that classification results are improved. In particular, it overcomes classification ambiguities between flat roof pixels and tree pixels.

In a second step, some well-known algorithms for supervised classification are used. We deal with Maximum Likelihood based on complex Gaussian distribution (uni-variate) and multivariate Complex Gaussian using coherency matrix. Meanwhile, the support vector machine, as a nonparametric method, is used as classification algorithm.

Moreover, a feature selection based on Genetic Algorithm using Mutual Information (GA-MI) is adapted to introduce optimal subset to classification method. To illustrate the efficiency of subset selection based on GA-MI, we perform a comparison experiment of optimal subset with different target decompositions based on different scattering mechanisms, including the Pauli, Krogager, Freeman, Yamaguchi, Barnes, Holm, Huynen and the Cloude decompositions.

Our experiments are based on an image of a suburban area, acquired by the airborne RAMSES SAR sensor of ONERA, in 2006, with a spatial spacing of 35 cm. The results highlight the potential of such data to discriminate some urban land cover types.

Keywords

Urban area, image classification, Maximum Likelihood, Fisher, high spatial resolution, polarimetric radar, Genetic Algorithm, Mutual Information

Contents

Remerciements **3**

Résumé **4**

Abstract **5**

Contents **7**

List of tables **11**

List of figures **15**

1. Introduction **19**

1.1. Background 19

1.2. Motivation 19

1.3. Chapter review 21

Part I: Feature extraction and feature selection

2. Feature extraction **25**

2.1. Introduction 25

2.2. Review of SAR polarimetry 26

2.2.1. Scattering matrix - Jones vector 27

2.2.2. Mueller matrix - Stokes vector 28

2.2.3. Covariance and coherency matrices 28

2.2.4. Decomposition-derived descriptors 31

2.2.4.1. Coherent decomposition 33

2.2.4.1.1. Pauli decomposition 33

2.2.4.1.2. Krogager decomposition 33

2.2.4.2. Incoherent decomposition 34

2.2.4.2.1. Huynen-based decomposition 35

2.2.4.2.2. Eigenvector-based decompositions 38

2.2.4.2.3. Model-based decompositions 44

2.2.4.3. Polarimetry parameters 46

2.2.4.4. Synthesis 50

2.3. Texture characteristic of SAR image 53

2.3.1. Extraction of texture features 54

2.4. Morphological feature extraction 55
2.4.1. Otsu method 56
2.4.2. Morphological Operations 57
2.4.2.1. Dilation morphological operation 57
2.4.2.2. Erosion morphological operation 57
2.4.2.3. Opening morphological operation 58
2.4.2.4. Closing morphological operation 58
2.5. Summary I 59
3. Feature selection 61
3.1. Feature selection 61
3.1.1. Mutual Information 62
3.1.2. Genetic Algorithm 64
3.2. Summary II 67

Part II: Classification

4. Statistics 70
4.1. Introduction 70
4.1.1. Review of SAR statistical models 70
4.2. Uni-variate Mellin transform 71
4.2.1. Definition and basic properties of Mellin transform 72
4.2.2. Mellin second kind statistics 74
4.3. Probability distribution 76
4.3.1. Generalities 76
4.3.2. Gamma, Rayleigh and Nakagami distributions 77
4.3.3. Inverse Gamma distribution 78
4.3.4. Log-normal distribution 79
4.3.5. Weibull distribution 80
4.3.6. Beta distribution (first kind Beta distribution) 81
4.3.7. Fisher distribution (second kind Beta distribution) 81
4.3.8. $\mathcal{K}$ distribution 82
4.4. Estimation methods 85
4.4.1. MLE 85
4.4.2. MoLC 86
4.5. Goodness-of-Fit 86

4.5.1. Correlation Coefficient estimation 86
4.5.2. Kolmogorov – Smirnov distance 87
4.5.3. Root Mean Square Error (RMSE) 87
4.6. Statistical analysis purpose and discussion 87
4.7. Summary III 89
5. Classification methods 91
5.1. ML classification based on uni-variate Gaussian pdf 91
5.2. Supervised ML based on multi-variate Gaussian pdf 92
5.3. Support Vector Machine (SVM) 93
5.4. ML classifier based on Fisher pdf 95
5.5. ML-Fisher classification coupled with morphological descriptor 96
5.6. Summary VI 97

Part III: Experimental results

6. Study area and data description 101
6.1. Study area 101
6.1.1. Study area location 101
6.1.2. Study area description 101
6.2. SAR data description 102
6.2.1. SAR data 102
6.3. Definition of the classes 103
6.3.1. Discussion 103
6.3.2. Ground truth data 104
7. Result on feature extraction and selection 108
7.1. Feature extraction and feature selection results 108
7.1.1. Feature extraction 108
7.1.2. GA-MI feature selection 113
7.2. Morphology process and result 130
7.2.1. Methodology 130
7.2.2. Results 130
8. Results on statistical analysis and classification 133
8.1. Statistical analysis procedure and results 133
8.1.1. Methodology 133
8.1.2. Results 134

8.2. Classification experiment I 142
8.2.1. The ML classifier based on Fisher pdf 142
8.2.2. SVM classification coupled with morphological descriptor 146
8.2.3. Comparison of results classification 148
8.3. Classification experiment II 149
8.3.1. The Complex G-ML classification coupled or not with morphological map 149
8.3.2. The SVM classification on complete descriptors 151
8.3.3. Comparison and results classification 153
8.4. Summary V 157
9. Conclusions and Future research 159
9.1. Conclusions 159
9.2. Future research 160
Bibliography 162
Appendix A 183
Résumé 192

List of tables

Table 2. 1: The positive real polarimetric descriptors considered in the statistical analysis (Chapter 3). $\langle \cdot \rangle$ is the average operator. *A,* stands for the amplitude of the descriptor... 51

Table 2. 2: Polarimetric decomposition descriptors considered in the SSM based on GA-MI (Chapters 7and 8)……..52

Table 2. 3: Original polarimetric components and discrimination parameters considered in classification (Chapter 6). $\langle \cdot \rangle$ is the average operator. S_{xy} represents the scattering coefficient of the targets, ***x*** the polarization of the incident field (H, V), ***y*** the polarization of the scattered field (H, V)..... 53

Table 2. 4: Mathematical expressions of statistical texture parameters from GLCM. $p_{d,\theta}(i,j)$, joint probability and $(i,j)^{th}$ element in the co-occurrence matrix of $\sqrt{T_{22\text{-}A}}$ and $\sqrt{T_{33\text{-}A}}$. 55

Table 4. 1: Properties of the Mellin convolution, the inverse Mellin convolution, and the Mellin correlation of two distributions defined on $\Re^+$: p_B, p_C with first and second kind characteristic functions, ϕ_B, ϕ_C, and, v^{th}-order log-cumulants, $\tilde{k}_{B,y}$, $\tilde{k}_{C,y}$. $\hat{*}$, $\hat{*}^{-1}$ and $\hat{\otimes}$ indicate convolution, inverse convolution and correlation. Adapted from (Nicolas, 2006). 77

Table 4. 2: Second kind characteristic functions of Beta and Fisher pdf obtained by inverse Mellin convolution and Mellin correlation of the G_I [1, M]) and G_I [μ, L] gamma distributions..... 83

Table 4. 3: Second kind characteristic functions of K and Fisher pdf obtained by Mellin convolution of the *IG* [1, M] Inverse gamma distribution, G_I [μ, L] and G_I [1, M] gamma distributions 83

Table 4. 4: Mellin statistics of uni-variate distribution (first second characteristic function and log-cumulants) for pdf in amplitude. Γ(.) is the Gamma function, K(.) is the second kind modified Bessel function, ψ(.) is the Digamma function, and ψ(*v*,.) is the vth-order Polygamma function. Adapted from (Nicolas, 2010; Nicolas, 2006)..... 84

Table 6. 1: Image information over covered urban area 103
Table 6. 2: Training and control samples (numbers of pixels)..... 106
Table 7. 1: positive real descriptors 109
Table 7. 2: subset-3 to 10 based on GA-MI of 127 descriptors 110
Table 7. 3: List of descriptors from 3 to 10 subset of SSM based on GA-MI through 127 complete descriptors..... 118
Table 7. 4: List of descriptors from 3 to 10 subset of SSM based on GA-MI with maximum relevancy with class 126
Table 8. 1: List of real positive descriptors in optimal subset of SSM based on GA-MI through the 58 descriptors 134
Table 8. 2: Confusion Matrix (in percent) for F-ML over optimal real positive descriptors. 143
Table 8. 3: Confusion Matrix (in percent) for G-ML over optimal real positive descriptors of 143
Table 8. 4: Confusion Matrix (in percent) for F-ML*.* stands for a classification with optimal 146
Table 8. 5: Confusion Matrix (in percent) for SVM with optimal subset of SSM of GA-MI from 147
Table 8. 6: Confusion Matrix (in percent) for SVM[1]*.[1]* stands for a classification with optimal subset of SSM of GA-MI from 58 real positive descriptors (Table 8.1).coupled with morphological map..... 147
Table 8. 7: Comparison of classification accuracies with different classifiers (in %): producer's accuracy for each class, kappa coefficient and overall accuracyover optimal subset of SSM based on GA-MI through 58 positive real descriptors (Table 8.1). [1] stands for a classification with optimal subset of SSM of GA-MI from 58 descriptors (Table 8.1).* stands for a classification with selected polarimetric descriptors and morphological map 149
Table 8. 8: Confusion Matrix (in percent) for G-ML[C]*. [C]* stands for the complex Gaussian 150
Table 8. 9: List of descriptors in optimal subset of SSM based on GA-MI through the 127 descriptors 151
Table 8. 10: Confusion Matrix (in percent) for SVM[2]. [2] stands for a classification with 152
Table 8. 11: Confusion Matrix (in percent) for SVM[2]*. [2]* stands for a classification 152

Table 8. 12: Comparison of classification accuracies with different classifiers (in %): G-ML^{C} is the complex Gaussian multi-variate ML approach based on the coherency matrix. ^{C*} stands for the complex Gaussian multi-variate ML approach based on the coherency matrix following with morphological map. ^{2} stands for a classification with optimal subset (Table 8.10) of SSM based on GA-MI through 127 descriptors (Table 7.2). 154
Table 8. 13: Comparison of SVM classification accuracies with different classifiers (in %): producer's accuracy for each class, kappa coefficient and overall accuracy of different decomposition methods based on different polarimetric matrixes anddifferent subset of SSM basedon GA-MI through 127 polarimetrydescriptors. 155
Table A. 1: Correlation Coefficients, Kolmogorov-Smirnov Distance, Root Mean Square Error between the estimated pdf and empirical data over optimal subset (Table 8.1) of real positive descriptors for **Class1-*Shadows*** respect to different estimation methods (MoLC, MLE) following with estimated parameters..... 184
Table A. 2: Correlation Coefficients, Kolmogorov-Smirnov Distance, Root Mean Square Error between the estimated pdf and empirical data over optimal subset (Table 8.1) of real positive descriptors for **Class2-*Highways*** respect to different estimation methods (MoLC, MLE) following with estimated parameters..... 185
Table A. 3: Correlation Coefficients, Kolmogorov-Smirnov Distance, Root Mean Square Error between the estimated pdf and empirical data over optimal subset (Table 8.1) of real positive descriptors for **Class3-*Lawns*** respect to different estimation methods (MoLC, MLE) following with estimated parameters..... 186
Table A. 4: Correlation Coefficients, Kolmogorov-Smirnov Distance, Root Mean Square Error between the estimated pdf and empirical data over optimal subset (Table 8.1) of real positive descriptors for **Class4-*Trees*** respect to different estimation methods (MoLC, MLE) following with estimated parameters..... 187
Table A. 5: Correlation Coefficients, Kolmogorov-Smirnov Distance, Root Mean Square Error between the estimated pdf and empirical data over optimal subset (Table 8.1) of real positive descriptors for **Class5-*Bright pixels*** respect to different estimation methods (MoLC, MLE) following with estimated parameters..... 188
Table A. 6: Correlation Coefficients, Kolmogorov-Smirnov Distance, Root Mean Square Error between the estimated pdf and empirical data over optimal subset (Table 8.1) of real positive descriptors for **Class6-*Flat roofs*** respect to different estimation methods (MoLC, MLE) following with estimated parameters..... 189

Table A. 7: Correlation Coefficients, Kolmogorov-Smirnov Distance, Root Mean Square Error between the estimated pdf and empirical data over optimal subset (Table 8.1) of real positive descriptors for **Class7-*Sloped roofs*** respect to different estimation methods (MoLC, MLE) following with estimated parameters..... 190

List of figures

Figure 1. 1: Research outline 22
Figure 2. 1: Spatial evolution of a polarized plane wave oriented with an ellipse angle, Ψ, respect to the $\hat{x}$- axis with constant amplitude, A, and ellipticity, τ. Adapted from (Lee & Pottier, 2009). 27
Figure 2. 2: Classification of the main decomposition theorems used in this work 32
Figure 2. 3: Entropy (H) and Anisotropy (A) values for four different configurations of the eigenvalues, adapted from (Cloude & Pottier, 1997) 40
Figure 3. 1: The relationship between MI and entropy 62
Figure 3. 2: Flowchart of feature selector based on GA-MI 66
Figure 5. 1: SVM optimal surface: (a) linear optimal surface and (b) nonlinear optimal surface, adapted from (Cortes & Vapnik, 1995) 94
Figure 5. 2: The proposed ML classification based on Fisher pdf 96
Figure 5. 3: Synopsis of the enriched ML classification based on Fisher-pdf and morphological tools 97
Figure 6. 1: The study area on aerial-photo (BD Ortho IGN) 101
Figure 6. 2: The study area on PoLSAR image (Pauli RGB) 102
Figure 6. 3: Urban classification based on different observations levels. 104
Figure 6. 4: (a) RAMSES image of the studied site (Toulouse, France), acquired in 2006, in Pauli colour-coded representation, (b) Aerial view from the French Geographical Institute (IGN) acquired in 2006, (c) Training samples superimposed on the RAMSES view (same colour legend as in Figure 8. 6), (d) Control samples superimposed on the view (same colour legend as in Figure 8. 6). 105
Figure 7. 1: Real Positive feature extraction for statistical analysis by different methods 109
Figure 7. 2: Feature extraction by different methods 110
Figure 7. 3: (a-c): Some of polarimetric descriptors 111
Figure 7. 4: (a) Evaluation function value vs iteration number for sets of 3 to 20 real positive descriptors. (b) Overal accuracy from uni-variate Gaussian-ML vs number of real positive descriptors. 115
Figure 7. 5: Mutual information between class and real positive descriptor in optimal subset of SSM based on GA-MI through 58 real positive descriptors. 116

Figure 7. 6:(a) Evaluation function value vs iteration number for sets of 3 to 20 descriptors of SSM based on GA-MI from 127 complete descriptors. (b) Overal accuracy from uni-variate Gaussian-ML vs number of descriptors. 117
Figure 7. 7: Mutual information between class and descriptor in subset 3 and 4 of SSM based on GA-MI through 127 descriptors. 119
Figure 7. 8:Mutual information between class and descriptor in subset 5 (optimal subset) of SSM based on GA-MI through 127 descriptors. 120
Figure 7. 9: Mutual information between class and descriptor in subset 6 of SSM based on GA-MI through 127 descriptors. 121
Figure 7. 10: Mutual information between class and descriptor in subset 7 of SSM based on GA-MI through 127 descriptors. 122
Figure 7. 11: Mutual information between class and descriptor in subset 8 of SSM based on GA-MI through 127 descriptors. 123
Figure 7. 12: Mutual information between class and descriptor in subset 9 of SSM based on GA-MI through 127 descriptors. 124
Figure 7. 13: Mutual information between class and descriptor in subset 10 of SSM based on GA-MI through 127 descriptors 125
Figure 7. 14-16:(a-c) Some of the selected descriptors in the subsets with 3 to 10 features obtained from SSM using 127 descriptors. 127
Figure 7. 17: Extracting building footprints map from S_{22-A} image through semi-automatic algorithm. 131
Figure 8. 1 (a-d): Empirical data and estimated pdfs using MLE and MoLC for different classes and optimal subset descriptors (Table 8.1) of SSM based on GA-MI through 58 real positive descriptors 136
Figure 8. 2: Estimated Fisher distribution using MoLC for seven classes and optimal subset (Table 8.1) of SSM based on GA-MI through 58 real positive descriptors. 141
Figure 8. 3: (a) G-ML and (b) F-ML. Classification over optimal subset (Table 8.1) of SSM based on GA-MI through 58 positive real descriptors (Table 7.1) 144
Figure 8. 4: (a) Building map (b) Membership of the cells to the class *trees* respect to Building map (values between 0 and 1) 145

Figure 8. 5: Classification result: (a) F-ML and (b) F-ML*.* stands for a classification with optimal subset of SSM of GA-MI from 58 descriptors (Table 8.1).coupled with morphological map..... 145
Figure 8. 6: Classification results with (a) SVM, (b) $SVM^{1}*.^{1}*$ stands for a classification with optimal subset of SSM of GA-MI from 58 real positive descriptors (Table 8.1).coupled with morphological map..... 148
Figure 8. 7: Classification results with (a) $G\text{-}ML^{C}$ and (b) $G\text{-}ML^{C}*$. ^{C} stands for the complex Gaussian multi-variate ML approach based on the coherency matrix. $^{C}*$ stands for the complex Gaussian multi-variate ML approach based on the coherency matrix following with morphological map..... 151
Figure 8. 8: Classification results with (a) SVM^{2} and (b) $SVM^{2}*.^{2}$ stands for a classification with optimal subset (Table 8.10).of SSM based on GA-MI from 127 descriptors (Table 7.2). $^{2}*$ stands for a classification with optimal subset (Table 8.10).of SSM based on GA-MI from 127 descriptors (Table 7.2).coupled with morphological map..... 153
Figure 8. 9: (a) Overal Accuracy (b) Producer Accuracy from SVM classification comparison results over different subset from SSM and different target decomposition methods...153

Chapter 1:
Introduction

1. Introduction

1.1. Background

Synthetic Aperture Radar (SAR) system is a well-known active imagery sensor system (electromagnetic pulse emitted from the sensing device) which is mounted on satellites or aircrafts (Franceschetti & Lanari, 1999; Cutrona, 1990; Curlander & McDonough, 1991). It can be used at any time of the day and under almost any weather conditions. It offers very high spatial resolution in multi-frequency (Kouskoulas, et al., 1998), multi-polarization (Ahern, et al., 1996) and multi-temporal mode (Pellizzeri, et al., 2003; Ferro-Famil & Pottier, 2007). SAR system can be able to discriminate targets according to the scattering mechanisms according to their physical properties (Soergel, 2010; Massonnet & Souyris, 2008). SAR images have been widely used in various applications: man-made structure interpretation (Lee, et al., 2006), sea ice classification (Scheuchl, et al., 2002), mining (Shekhar, et al., 2002), oil pollution monitoring (Hovland, et al., 1994; Wei, et al., 2012), oceanography (Banner, et al., 1983), navigation (Ren, et al., 2011), snow monitoring (Storvold, et al., 2006), classification of earth terrain (Kong, et al., 1990), environmental monitoring (Mesquita, et al., 2008; Koo, et al., 2012), natural disaster assessment (Brunner, et al., 2010), man-made structure surveillance (Ainsworth, et al., 2008), 3D city modeling (Soergel, et al., 2006), GIS data generation (Brenner & Roessing, 2008). They can also be used in interferometric mode, to produce some unique derivative image products that incoherent systems cannot (Zhou, et al., 2009), or to estimate scattering mechanisms from polarimetric multi-baseline interferometric SAR measurements (Sauer, et al., 2011). All these works show the advantages of using SAR data.

Moreover, such data can provide complementary properties. For instance, such complementary can be observed by means of data fusion techniques in (Sportouche & Tupin, 2011), or the complementary nature of multi sensor optical-SAR image in development of global monitoring systems are shown in (Lehmann, et al., 2012). However, with the high capabilities of SAR systems, new challenges are offered.

1.2. Motivation

The accuracy of a supervised classification in an urban context mainly depends on the SAR data characteristics such as spatial resolution, frequency, number of looks, look angle, polarization mode, etc. This accuracy is also related to the classification method, and its implementation. Moreover, an accurate result is obtained when integrating adequate ground knowledge. The high spatial resolution SAR image offers very precise information, which is related to the properties of targets e.g., surface nature, pattern, dimension, geometry structures (regular geometry). Despite this fact, the interpretation

of urban classes is still complicated according to target texture, geometry, and orientation with respect to the radar line of sight. Indeed, similar structures with different orientations can present different appearance in SAR images. In fact, tracing and recognizing the exact meaning of many details is not straightforward. Meanwhile, different targets such as streets, pavements, highways, or bare soil have low backscattering coefficients in SAR images and show the same behaviour (Dousset, 1995). Thus, some ambiguities are due to these backscattering similarities. Besides, due to elements which form dihedral and trihedral, for example between ground and building or between ground and trees, bright patterns appear in SAR images. An empirical analysis of double bounce behaviour of buildings with different aspect angles which was conducted in (Brunner, 2009; Xia & Henderson, 1997) show that strong backscattering can cause classification ambiguities between man-made structures and natural elements.

Moreover, slant range mapping causes geometrical distortions such as layover, foreshortening, and shadow effects which complicate the interpretation of scattering mechanisms. However, a larger off nadir angle (for instance 60°) causes smaller layover effects but causes larger shadow effects and consequently loss of information in shadow regions (Weydahl, 1997). For some man-made objects with regular structures strong backscattering can also occur (Lee & Pottier, 2009). These complexities bring many challenges to separate different kinds of surfaces using radar image classification and it seems that a single SAR data will not be enough to produce a good mapping of the urban surfaces. When available, hyperspectral images are preferable (Fairbarn, 2013).

However, to overcome these misclassification problems and improve the radar image classification result, many analytical and numerical approaches have been developed. In particular, full polarimetric SAR system provides additional information on the scattering mechanisms, that single or dual polarimetric mode cannot. Thus some classification methods are based on the modeling of the physical scattering mechanisms from fully polarimetric SAR image using polarimetric decomposition and descriptors (Unal & Ligthart, 1998; Cloude & Pottier, 1997; Pottier & Lee, 2000; Ferro-Famil, et al., 2001; Yang, et al., 2006; Kajimoto & Susaki, 2013; Yamaguchi, et al., 2005; Yamaguchi, et al., 2011; Simonetto & Malak, 2009) or polarimetric discriminators (Moriyama, et al., 2004 ;Iwasa, et al., 2010). Other works are based on the statistical modeling of polarimetric SAR (PoLSAR) images (Lee et al., 1999; Bombrun, et al., 2011; Bombrun & Beaulieu, 2008). Additionally, there are several works using some combinations of the above approaches (Lee et al., 1999; Pottier & Lee, 2000; Formont, et al., 2011). Moreover, some image processing techniques have been employed (Tan, et al., 2007; Ince, et al., 2012; Zhang, et al., 2009; Ince, 2010). Some authors have also used multi-frequency full polarimetric data classification approaches (Chen, et al., 1996; Kahny & Wiesbeck, 1991; Fukuda & Hirosawa, 1999). However, they all present drawbacks even if, they offer various perspectives.

In this framework, the current thesis considers the potential of a single look complex full PoLSAR image for urban mapping. We aim at producing a high spatial resolution map that shows the main

urban surfaces: different kinds of roofs, trees, bare soils, roads, etc. For that purpose, we use an X-band image. Indeed, this short wavelength is used for classification, as a higher sensitivity to surface roughness is expected.

A new supervised classification algorithm is proposed. It combines the maximum likelihood approach based on the Fisher probability function (Fisher-pdf) and morphological tools. The morphological tool allows adding spatial information as a constraint during the classification process (Figure 1.1). The classification is performed on a set of polarimetric and textural descriptors. Polarimetric descriptors are derived from various theories and some polarimetric discriminator attributes. Statistical or texture descriptors are used to enhance the different statistical behaviours. However, this algorithm assumes the statistical independency of attributes. For this reason, a feature selection based on Genetic Algorithm using Mutual Information is adapted to introduce optimal subset based on minimum redundancy and maximum relevance (mRMR) to classification method. Moreover, to illustrate the efficiency of subset selection based on GA-MI, we performed a comparison experiment of optimal subset with different target decompositions based on different scattering mechanisms, including the Pauli, Krogager, Freeman, Yamaguchi, Barnes, Holm, Huynen and the Cloude decompositions.

1.3. Chapter review

After this introductory chapter, various polarimetric descriptors, textural parameters, and a morphological feature are presented in Chapter 2. Moreover, a feature selection method to reduce the set of attributes during the classification process is presented in Chapter 3. We propose the Mutual Information (MI) and use Minimum Redundancy and Maximum Relevance (mRMR) as selection criterion and the Genetic Algorithm (GA) as iterative searching method. Chapter 4 gives a brief review of the Mellin framework for image statistical analysis. Several specific distribution functions for modeling the descriptors derived from high resolution polarimetric SAR images are given. Furthermore, for estimating their parameters, the method of maximum likelihood and the method of log-cumulant are described. In chapter 5, some widely used supervised classification methods are described. Then, we present a new algorithm that takes benefit of polarimetric, statistical, textural, and morphological contents of the SAR data. In Chapter 6, the study area and the dataset are presented. Chapter 7 shows experimental results on feature extraction and selection. Results on statistics and classification and comparison of different classifiers, are presented in Chapter 8. Finally, a conclusion and some recommendations for future research works are given in Chapter 9.

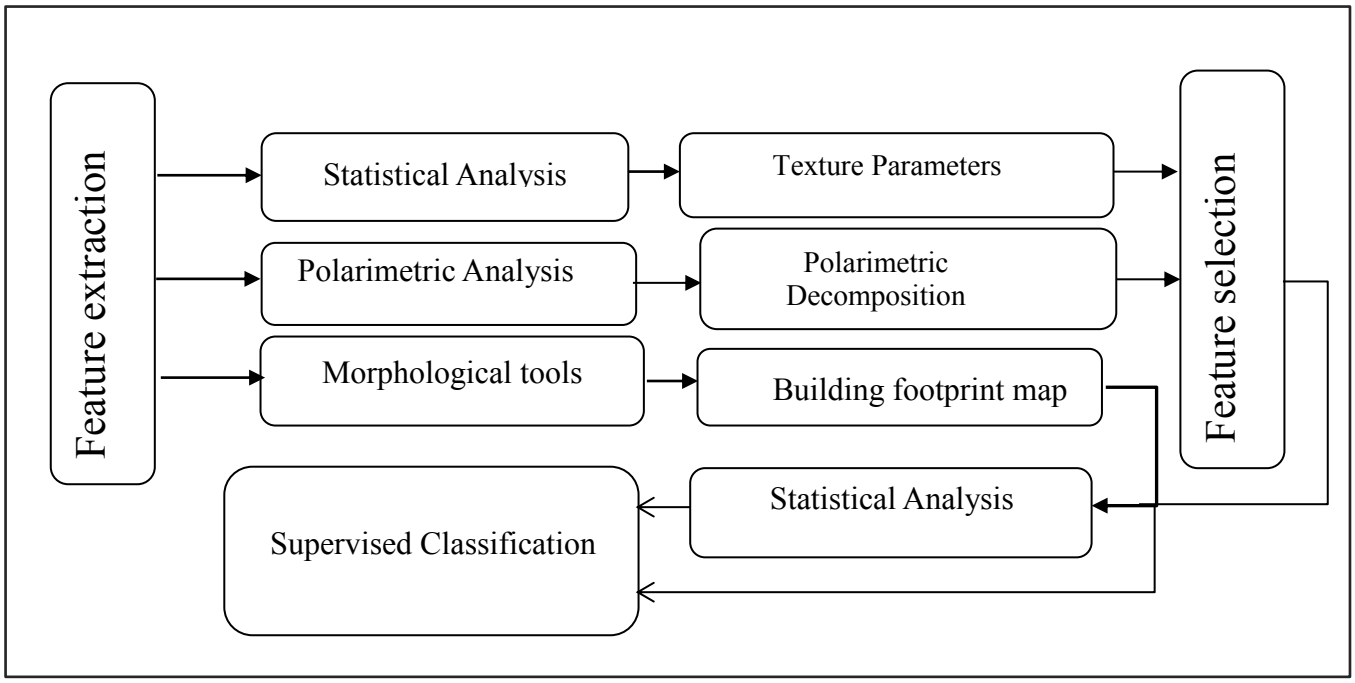


Figure 1. 1: Research outline

Part I:

Feature extraction and feature selection

Chapter 2:

Feature extraction

2. Feature extraction

2.1. Introduction

Radar backscattering depends on many factors: SAR system parameters such as wavelength, incidence angle, polarization, look direction, and target properties such as materials (wood, mud, concrete, metal, stone,…), geometry (building orientation about azimuth direction, size, shape, pattern), geophysical (roughness), and electrical characteristics.

For example, the wavelength of the radar system has a great influence on the backscattering signal from an urban feature. For instance, at X band (2.4 to 3.75 cm), the wave does not penetrate the canopy and produces more backscattering from top layer and the crowns of trees compared to system with longer wavelength such as C, L and K bands. The X band image presents more sensitivity to surface roughness compared to longer wavelengths as explained in (Xia & Henderson, 1997).

Moreover, statistical analysis and empirical observations from polarimetric data show that the co-polarization signal is dominated by surface backscattering while the cross-polarization one is dominated by volume scattering (Hussin, 1995). The cross-polarization backscattering is higher for volume targets while the co-polarization backscattering is higher for dihedral and trihedral targets. For example, the corner reflectors have strong returns in co-polarized data (Hussin, 1995). However, most man-made structures such as streets and buildings which are nonaligned in azimuthal direction produce diffuse reflection (Hussin, 1995; Xia & Henderson, 1997). These diffusions cause interpretation ambiguities between trees and roofs or between similar building structures. Moreover, low density trees can behave like dihedral and trihedral reflectors so the volume backscattering less dominates. Again it causes some interpretation ambiguities between trees and building structures.

Therefore, by varying target orientation relatively to the radar look direction, the dominating backscattering from flat, or sloped target roofs could be diffuse reflection. Simply, building not aligned along azimuth direction mainly has cross-polarized backscattering. Changing the orientation angle may cause changes in the return signals, as discussed in (Hussin, 1995; Xia & Henderson, 1997). On the other hand, the moisture influences backscattering levels since it modifies dielectric constants.

Therefore, extracting and selecting the convenient descriptors for the discrimination of different targets is very important. The main goal of this chapter is to present possible SAR descriptors from PoLSAR data for classification. We propose attributes extracted from PoLSAR data in three ways: features directly derived from the scattering matrix or using decomposition methods (Lee & Pottier, 2009), texture features from the gray-level co-occurrence matrix (GLCM) (Schistad Solberg & Jain, 1997), and morphological tools (Benediktsson, et al., 2003; Fukuda & Hirosawa, 1999).

In section 2.2, we briefly review the physical and mathematical background of radar polarimetry together with relevant equations which are essential for understanding the PoLSAR parameters. Scattering matrix, covariance and coherency matrices are described in this section and also different target decomposition methods. Then, some statistical techniques which are used for producing descriptors from GLCM can be found in section 2.3. Morphological feature extraction, its operators and a proposed method to extract building patterns, are presented in section 2.4.

2.2. Review of SAR polarimetry

Radar polarimetry is based on Radar (radio detection and ranging) technology and a fundamental property of electromagnetic waves. The history of radar polarimetry is described in (Boerner, 1987). The electric field of a plane wave can be given by the sum of two orthogonal components, i.e horizontal and vertical components. Figure 2.1 shows polarization ellipse as the property of a plane electromagnetic (EM) wave in a three dimensional space ($\hat{x}$, $\hat{y}$, and $\hat{z}$), which refers to regularity and alignment of propagation. Spatial evolution of a polarized plane wave is shown by the oriented ellipse and geometrical parameters: angle, Ψ, respect to the $\hat{x}$- axis, amplitude, A, and ellipticity, τ In SAR system ($\hat{x}$, $\hat{y}$, $\hat{z}$) represents the coordinate system centered on the transmitter or the receiver antenna. If the positive $\hat{z}$- axis is in the same direction as the travel of the wave, this convention is known as the forward scatter alignment (FSA). Otherwise, if the positive $\hat{z}$- axis is towards the target, this convention is known as backscatter alignment (BSA).

According to figure 2.1, the relationship between geometrical parameters Ψ and τ is defined by:

$$\begin{aligned} \tan(2\Psi) &= 2E_{\circ x}E_{\circ y}\cos\delta / (E_{\circ x}^2 - E_{\circ y}^2), \ (-\pi/2 \le \Psi \le \pi/2) \\ \sin(2\tau) &= 2E_{\circ x}E_{\circ y}\cos\delta / (E_{\circ x}^2 + E_{\circ y}^2), \ (-\pi/4 \le \tau \le \pi/4) \end{aligned} \tag{2. 1}$$

where, $E_{\circ x}$ and $E_{\circ y}$ are amplitudes of the wave electric field, δ is its relative phase. Transmitting and receiving of orthogonally polarized microwave pulses (horizontally (H) or vertically (V)) are obtained from PoLSAR instruments. Consequently, the fully polarimetric radar system records the complete characterization of the scattering field in all the configurations: HH, HV, VH, and VV. The target scattering behavior can be described by the scattering matrix based on the Jones formalism or by the Mueller matrix based on the Stokes formalism (Touzi, et al., 2004). The fundamental theories and applications of radar polarimetry have been presented in (Lee et al., 2004; Lee & Pottier, 2009).

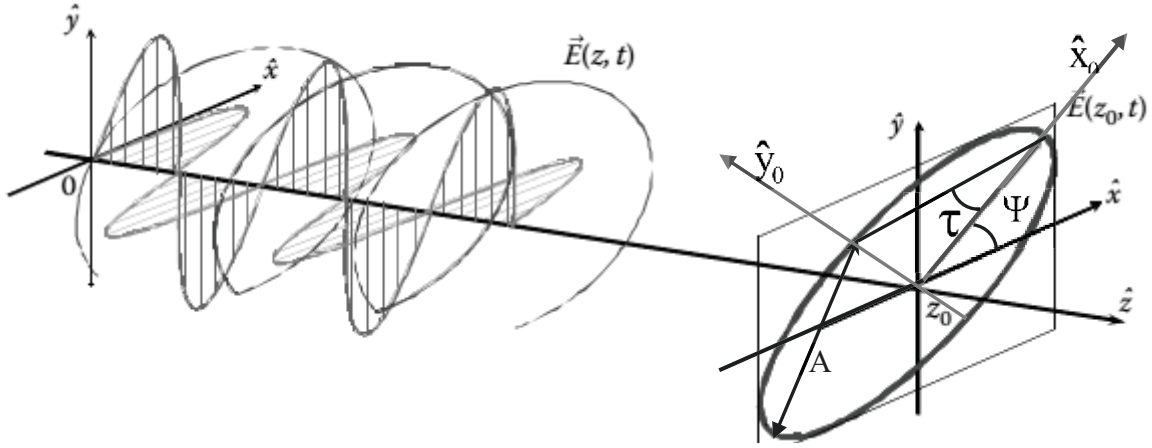


Figure 2. 1: Spatial evolution of a polarized plane wave oriented with an ellipse angle, Ψ , respect to the $\hat{x}$-axis with constant amplitude, A, and ellipticity, τ. Adapted from (Lee & Pottier, 2009).

2.2.1. Scattering matrix - Jones vector

The polarization states of both vertical and horizontal waves can be described as a vector, so that the polarization behavior of the target can be represented by the scattering matrix [S] based on Jones vector:

$$J = \begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{bmatrix} E_{\circ x}\, e^{j\delta_x} \\ E_{\circ y}\, e^{j\delta_y} \end{bmatrix} = \vec{E}^s = \frac{e^{-jkr}}{r}[S]\vec{E}^i = \frac{e^{-jkr}}{r}\begin{bmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{bmatrix}\vec{E}^i \qquad (2.\ 2)$$

$\vec{E}^i$ is the incident wave field, and $\vec{E}^s$ is the scattering wave field, *r* is the distance between target and antenna and e^{-jkr} [S]the propagation effects both, in amplitude and phase.

S_{xy} represents the scattering coefficient of the targets, ***x*** is the polarization orientation of the incident field, ***y*** the one of the scattered field. In monostatic case, Horizontal (*H*) and Vertical (*V*) polarizations coincide with *x* and *y* axes. When the *H* polarized wave is transmitted, both backscattering, related to *H* and *V* polarizations, are received in full polarimetric system. When, *V* polarized wave is transmitted, both backscattering related to *H* and *V* polarizations can be received. All four transmitting and receiving waves, $\{S_{HH}, S_{VH}, S_{HV}, S_{VV}\}$, are recorded at the same time (or alternatively). Here, we consider the monostatic case (the transmitter and receiver are collocated), and [S] is a complex symmetric matrix which holds the reciprocity property for most targets ($S_{xy} = S_{yx}$). [S] is called the Sinclair matrix. Thus, the scattering matrix is represented by:

$$[S] = e^{j\Phi_{HH}} \begin{pmatrix} |S_{HH}| & |S_{HV}|\, e^{j(\Phi_{HV}-\Phi_{HH})} \\ |S_{HV}|\, e^{j(\Phi_{HV}-\Phi_{HH})} & |S_{VV}|\, e^{j(\Phi_{VV}-\Phi_{HH})} \end{pmatrix} \qquad (2.3)$$

The scattering matrix in Equation (2.3) not only allows to obtain intensity or amplitude but also to explore the phase information provided by the phase of complex scattering coefficients. Therefore, a given target is characterized by five parameters, namely three amplitudes $\{|S_{HH}|, |S_{HV}|, |S_{VV}|\}$ and two relative phases, $\{(\Phi_{HV} - \Phi_{HH}),\ (\Phi_{VV} - \Phi_{HH})\}$, and one additional absolute phase (Lee et al., 2004; Lee & Pottier, 2009). The so-called lexicographic scattering vector or covariance vector, $\vec{K}_S$, is obtained from the vectorized writing of the scattering matrix, defined as:

$$\vec{K}_S = \begin{bmatrix} S_{HH} & \sqrt{2}\,S_{HV} & S_{VV} \end{bmatrix}^T \tag{2.4}$$

$$\vec{K}_S = \frac{1}{2} Tr([S]\{\psi_L\}) \quad , \{\psi_L\} = \left\{ 2\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, 2\sqrt{2}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, 2\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \tag{2.5}$$

$\{\psi_L\}$ is a 2×2 complex lexicographic matrix basis, here presented in the monostatic case (Cloude & Pottier, 1996). The total scattering power (span), in monostatic case and the reciprocity theorem, is defined by:

$$Span\,[S] = Tr([S]\ [S^\dagger]) = |S_{HH}|^2 + 2|S_{HV}|^2 + |S_{VV}|^2 = |\vec{K}_S|^2 \tag{2.6}$$

where † is the complex conjugate and transpose operator.

2.2.2. Mueller matrix - Stokes vector

Polarization backscattering can also be expressed in term of the Mueller matrix, [M], based on Stokes formalism (a set of values that describe the polarization state of electromagnetic radiation) (Van Zyl, 1993). This representation consists in the linear combination of the cross-products of the four basic elements of the scattering matrix. In this case, suppose that the Stokes vectors of the incident wave and backscattered wave are $\vec{S}^i$ and $\vec{S}^s$ respectively. The relationship between $\vec{S}^i$ and $\vec{S}^s$ are described through the Mueller matrix as follows:

$$\vec{S}^s = [M]\,\vec{S}^i \tag{2.7}$$

The [M]-matrix is a 4×4 real matrix and used in FSA convention (Ulaby, et al., 1992). In the same way, in BSA case, the Stokes vectors can be linked using the Kennaugh matrix.

2.2.3. Covariance and coherency matrices

In a SAR image, the pixel value results from the coherent addition of the responses of the elementary scatterers within a resolution cell. Generally, the microwave scattering is one of two types: point and complex scattering. For example, power-lines, vehicles, railways, or buildings have a dominating backscattering (strong point scattering behavior), and are known as deterministic scatterers. Natural media have complex scattering such as forest and they are known as non-deterministic scatterers.

When there is no dominant scatterer in a resolution cell, the statistic of the response is given by the complex Gaussian scattering model, given rise to the so-called speckle (spatially random multiplicative noise due to coherent superposition of multiple backscatter sources within a SAR resolution element). In fact, the resolution cell consists in more scatterers that change with time and space. The set of these targets are called "distributed target" and gives rise to a covariance or coherency matrices. The target covariance matrix (in the reciprocal and monstatic case) is generated from the lexicographic scattering vector, $\vec{K}_S$. Thus, the covariance matrix is given by:

$$[C]_{3\times3} = \left\langle \vec{K}_S \,.\vec{K}_S^{\dagger} \right\rangle = \begin{bmatrix} \left\langle |S_{HH}|^2 \right\rangle & \sqrt{2}\left\langle S_{HH} S_{HV}^* \right\rangle & \left\langle S_{HH} S_{VV}^* \right\rangle \\ \sqrt{2}\left\langle S_{HV} S_{HH}^* \right\rangle & 2\left\langle |S_{HV}|^2 \right\rangle & \sqrt{2}\left\langle S_{HV} S_{VV}^* \right\rangle \\ \left\langle S_{VV} S_{HH}^* \right\rangle & \sqrt{2}\left\langle S_{VV} S_{HV}^* \right\rangle & \left\langle |S_{VV}|^2 \right\rangle \end{bmatrix} \tag{2.8}$$

Where * is the complex conjugate operator, $\langle \ldots \rangle$ denotes the temporal or spatial averaging, and † is the conjugate and transpose operator. The diagonal represents the backscattered intensities. The off diagonal elements represent the complex covariance of the respective polarization configurations. The target coherency vector is based on linear combinations arising from the Pauli matrices (Lee & Pottier, 2009). The Pauli vector is given by:

$$\vec{K}_P = \frac{1}{2} tr([\mathrm{S}]\{\Psi_P\}) = \frac{1}{\sqrt{2}}\begin{bmatrix} S_{HH} + S_{VV} & S_{HH} - S_{VV} & 2S_{HV} \end{bmatrix}^T \tag{2.9}$$

$\{\Psi_P\}$ is a 2×2 complex Pauli matrix basis defined by:

$$\{\Psi_P\} = \left\{ \sqrt{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \sqrt{2}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \sqrt{2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\} \tag{2.10}$$

The first element, $(S_{HH} + S_{VV})/\sqrt{2}$, is assigned to odd bounce scatterer such as the sphere, the plane surface or reflectors of trihedral type. The second one, $(S_{HH} - S_{VV})/\sqrt{2}$, is related to a dihedral scatterer or double isotropic bounce and the third element, $2S_{HV}/\sqrt{2}$, is associated to the diffuse scattering or the volume scattering. In the reciprocal and monostatic case, the coherency matrix is given by:

$$[T]_{3\times3} = \left\langle \vec{K}_P \,.\, \vec{K}_P^{\dagger} \right\rangle = \frac{1}{2}\begin{bmatrix} \left\langle |S_{HH} + S_{VV}|^2 \right\rangle & \left\langle (S_{HH} + S_{VV})(S_{HH} - S_{VV})^* \right\rangle & 2\left\langle (S_{HH} + S_{VV}) S_{HV}^* \right\rangle \\ \left\langle (S_{HH} - S_{VV})(S_{HH} + S_{VV})^* \right\rangle & \left\langle |S_{HH} - S_{VV}|^2 \right\rangle & 2\left\langle (S_{HH} - S_{VV}) S_{HV}^* \right\rangle \\ 2\left\langle S_{HV} (S_{HH} + S_{VV})^* \right\rangle & 2\left\langle S_{HV} (S_{HH} - S_{VV})^* \right\rangle & 4\left\langle |S_{HV}|^2 \right\rangle \end{bmatrix} \tag{2.11}$$

Off diagonal information of coherency or covariance matrices are related to the geometry and the electrical properties of the target and can be used for physical interpretation of the scattering mechanism. The total scattering power (span), in monostatic and reciprocity case, is defined from the Pauli vector by:

$$Span(\mathrm{S}) = Tr([\mathrm{S}]\ [\mathrm{S}]^{\dagger}) = |\mathrm{S}_{HH}|^2 + 2|\mathrm{S}_{HV}|^2 + |\mathrm{S}_{VV}|^2 = |\vec{K}_p|^2 \tag{2.12}$$

Coherency and covariance matrices have the same real positive eigenvalues but their eigenvectors are different. In addition, some symmetry assumptions about the distribution of the scatterers (targets) let us simplify the scattering problems and geophysical parameter estimation. For example, if a distributed target has reflection symmetry (Nghiem, et al., 1992) in the plane normal to the line-of-sight, the averaged matrix coherency will have a simpler form:

$$[T]_{3\times3} = \frac{1}{2}\begin{bmatrix} \langle |S_{HH}+S_{VV}|^2 \rangle & \langle (S_{HH}+S_{VV})(S_{HH}-S_{VV})^* \rangle & 0 \\ \langle (S_{HH}-S_{VV})(S_{HH}+S_{VV})^* \rangle & \langle |S_{HH}-S_{VV}|^2 \rangle & 0 \\ 0 & 0 & 4\langle |S_{HV}|^2 \rangle \end{bmatrix} \tag{2.13}$$

Cross-polarization scattering coefficient is uncorrelated with the co-polarization terms. Another assumption for facilitating the calculation of decomposition is the rotation symmetry (Nghiem, et al., 1992) around the line-of-sight. For example the averaged oriented coherency matrix around the line-of-sight is given as follows:

$$[T(\theta)]_{3\times3} = R(\theta)[T]_{3\times3} R(\theta)^{-1}\ ,\ R(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(2\theta) & \sin(2\theta) \\ 0 & -\sin(2\theta) & \cos(2\theta) \end{bmatrix} \tag{2.14}$$

where $R(\theta)$ is a spatial unitary ($R(\theta)R(\theta)^{-1} = 1$) rotation matrix. The averaged oriented coherency matrix should be unchanged under the transformation, and it requires that the target vectors must be the eigenvectors of the rotation matrix, $R(\theta)$:

$$R(\theta)\vec{u} = \lambda\vec{u}, \quad \vec{u} = [u_1 \quad u_2 \quad u_3], \quad u_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \ u_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ 1 \\ j \end{bmatrix}, \ u_3 = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 \\ j \\ 1 \end{bmatrix} \tag{2.15}$$

As the coherency matrix is constructed from a linear combination of these eigenvectors, the coherency matrix under rotation symmetry assumption will be simplified in this form:

$$[T]_{3\times3} = \frac{1}{2}\begin{bmatrix} \left\langle \left|S_{HH}+S_{VV}\right|^2 \right\rangle & 0 & 0 \\ 0 & \left\langle \left|S_{HH}-S_{VV}\right|^2 \right\rangle & 2\left\langle \left(S_{HH}+S_{VV}\right)S_{HV}^* \right\rangle \\ 0 & 2\left\langle S_{HV}\left(S_{HH}-S_{VV}\right)^* \right\rangle & \left\langle \left|S_{HH}-S_{VV}\right|^2 \right\rangle \end{bmatrix} \tag{2.16}$$

Finally, a distributed target may have reflection symmetry in some special plane normal to the line-of-sight and rotation symmetry around line-of-sight which is known as azimuth symmetry (Nghiem, et al., 1992). The coherency matrix under azimuth symmetry assumption is simplified in this form (Ferro-Famil, 2011).

$$[T]_{3\times3} = \frac{1}{2}\begin{bmatrix} \left\langle \left|S_{HH}+S_{VV}\right|^2 \right\rangle & 0 & 0 \\ 0 & \left\langle \left|S_{HH}-S_{VV}\right|^2 \right\rangle & 0 \\ 0 & 0 & \left\langle \left|S_{HH}-S_{VV}\right|^2 \right\rangle \end{bmatrix} \tag{2.17}$$

The averaged covariance matrix, $[C]_{3\times3}$, corresponding to the three different scattering symmetry configurations (reflection symmetry, oriented symmetry, and azimuth symmetry) is discussed in (Lee & Pottier, 2009; Nghiem, et al., 1992). For example, the averaged covariance matrix under azimuth symmetry is simplified as follows (Ferro-Famil, 2011):

$$[C]_{3\times3} = \begin{bmatrix} \left\langle \left|S_{HH}\right|^2 \right\rangle & 0 & \left\langle S_{HH}S_{VV}^* \right\rangle \\ 0 & \left\langle 2\left|S_{HV}\right|^2 \right\rangle & 0 \\ \left\langle S_{VV}S_{HH}^* \right\rangle & 0 & \left\langle \left|S_{VV}\right|^2 \right\rangle \end{bmatrix} \tag{2.18}$$

2.2.4. Decomposition-derived descriptors

The main purpose of using different decomposition methods is to analyze the backscattering behavior of different targets in a single cell with respect to their scattering mechanisms. In fact, the recorded polarimetric data allow a better characterization of the surfaces based on decomposition theorems. The target decomposition method can decompose the mixed scattering mechanisms in the resolution cell into a sum of independent elements (each element is defining a deterministic scattering mechanism). For that purpose, some of them use the averaged coherency or covariance matrix. Indeed, polarimetric decomposition methods provide an interpretation of PoLSAR images using simple mechanisms, such as single-bounce surface scattering, double-bounce corner scattering, multiple random scattering from a volume, helix (correspond to anti-symmetric scattering mechanisms that transform the incident wave into its orthogonal circular polarization state) and wire scattering.

An expansion of target decomposition derived from the scattering matrix was first considered in (Krogager, 1993), and from the Mueller matrix in (Huynen, 1970).

Polarimetric decomposition theorems are divided into incoherent and coherent target decompositions (Lee & Pottier, 2009). Incoherent decompositions are based on the incoherently averaged covariance matrix or coherency matrix that possesses nine variables (section 2.2.3).

Incoherent decompositions mainly are of three types: those using an eigenvector analysis of the covariance matrix as used in (Ainsworth, 2002; Lopez-Martinez, et al., 2005; Vasile et al., 2010) or coherency matrix as proposed in (Cloude & Pottier, 1996; Cloude & Pottier, 1997; Touzi, 2007) those using model-based method as explained in (Freeman & Durden, 1998), and used in (Yamaguchi et al., 2011; Yamaguchi et al., 2005a; Sato, Yamaguchi, Singh, & Park, 2012; Zhang, Wang, Chen, Li, Yin, & Hong, 2013; Arii, van Zyl, & Yunjin, 2011), and finally those based on Huynen-based model as proposed in (Huynen, 1970; Huynen,1990) .

Meanwhile, coherent decompositions are based on the scattering matrix that has five independent variables (section 2.2.1), such as Krogager, Pauli, and the Cameron decompositions, explained in (Cloude & Pottier, 1996). However, the coherent decompositions are useful if the scatterers are points or pure targets.

As the incoherent decomposition method uses averaging of the scattering signal and the coherent decomposition method uses the full image resolution, these two methods complete each other. In our work, we select both coherent and incoherent decompositions and expect to recognize different types of roofs, man-made structures and clutter classes. Figure 2.2 shows the classification of the main decomposition theorems which are used in this work.

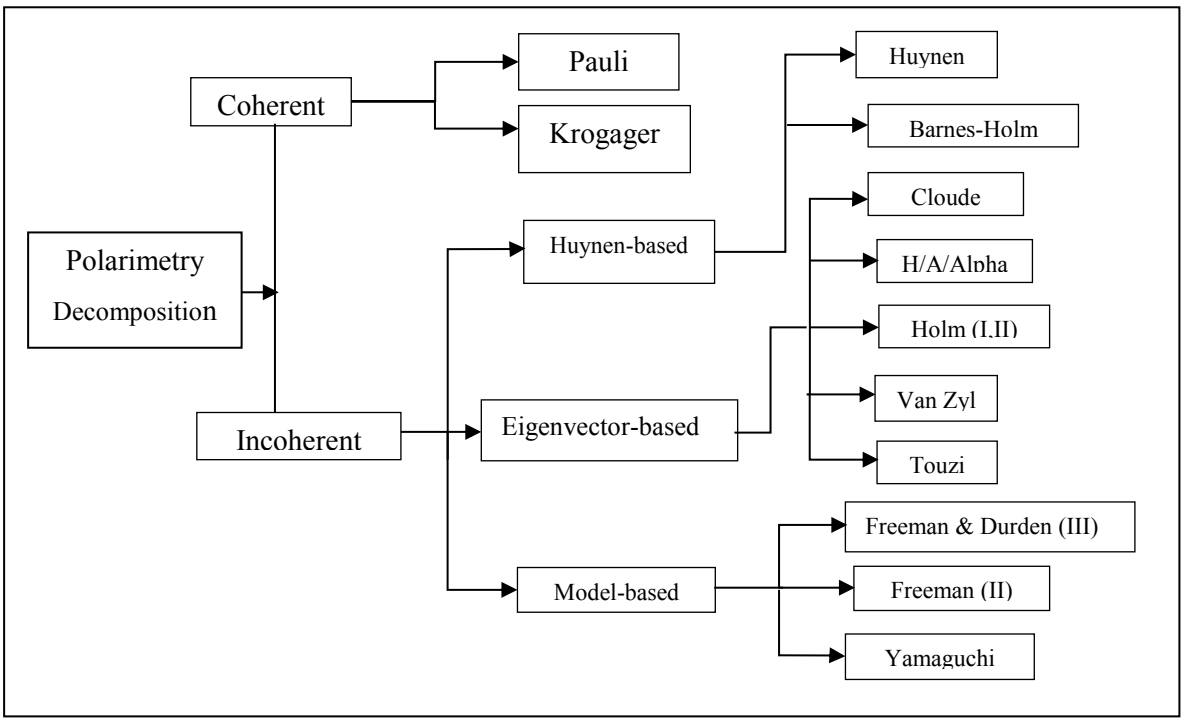


Figure 2. 2: Classification of the main decomposition theorems used in this work

2.2.4.1. Coherent decomposition

The scattering matrix describes the information of the pure or simple objects respect to their scattering mechanism. In coherent decomposition, the measured scattering matrix is expressed as the complex combination of basis matrices,$[S]_i$, corresponding to canonical scattering mechanisms (Lee & Pottier, 2009):

$$[S] = \sum_{i=1}^{k} c_i [S]_i \tag{2.19}$$

The ,$[S]_i$ indicates simple objects or canonical objects because every object under consideration is equivalent to exactly one object in canonical form, and coefficient indicates the weight of,$[S]_i$.
The coherent decompositions by Pauli and Krogager are presented in the following sections.

2.2.4.1.1. Pauli decomposition

In the Pauli decomposition, we suppose that reciprocity applies in a monostatic system configuration. Thus, the matrix, [S], is symmetric and, the Pauli basis is described by:

$$[S]_a = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ [S]_b = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},\ [S]_c = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{2. 20}$$

The scattering matrix [S] can be expressed by:

$$[S] = a[S]_a + b[S]_b + c[S]_c \tag{2. 21}$$

Where

$$a = \frac{|S_{HH} + S_{VV}|}{\sqrt{2}},\ b = \frac{|S_{HH} - S_{VV}|}{\sqrt{2}},\ c = \sqrt{2}|S_{HV}| \tag{2. 22}$$

The single bounce or odd-bounce scattering of a sphere, a plate or a trihedral is described with, $[S]_a$ and the weight coefficient, a. The, $[S]_b$ -matrix indicates a scattering mechanism characterized by double-bounce or even-bounce scattering from some corners with a *b*-weight. The, $[S]_c$ -matrix indicates a scattering mechanism from anti-symmetric or random scatterers with a *c*-weight. The,$|a|^2$ -, $|b|^2$ -and $|c|^2$ -coefficients give an easier physical interpretation from a qualitative point of view, equivalent to the brightness values in a RGB image.

2.2.4.1.2. Krogager decomposition

Another example of coherent decomposition is the Krogager one in which the complex radar target scattering matrix is decomposed into three components and provides a clearer physical interpretation picture of the physical backscattering mechanisms in three terms: diplane, sphere, and helical targets

(Krogager, 1990). The sphere and helix components present an orientation angle, θ. The Krogager decomposition is based on scattering matrix in a circular polarization basis. The transformation from the linear polarization basis, $\{S_{HH}, S_{VV}, S_{HV}\}$, to the circular polarization basis, $\{S_{RR}, S_{LL}, S_{RL}\}$, is simply given by these formulas:

$$S_{RR} = jS_{HV} + \frac{1}{2}(S_{HH} - S_{VV}),\ \ S_{LL} = jS_{HV} - \frac{1}{2}(S_{HH} - S_{VV}),\ \ S_{RL} = j\frac{1}{2}(S_{HH} + S_{VV}) \tag{2.23}$$

The Krogager decomposition presents the following formulation:

$$[\mathrm{S}_{RL}] = e^{j\varphi}\left\{e^{j\varphi_s} k_{\mathrm{S}} [\mathrm{S}]_{sphere} + k_{\mathrm{d}} [\mathrm{S}]_{diplane(\theta)} + k_{\mathrm{h}} [\mathrm{S}]_{helix(\theta)}\right\} \tag{2.24}$$

The krogager decomposition has five parameters, $\{k_S, k_{\mathrm{d}}, k_{\mathrm{h}}, \varphi_S, \theta\}$, plus the absolute phase, φ, and is written:

$$\begin{aligned}[\mathrm{S}_{RL}] &= \begin{bmatrix} S_{RR} & S_{RL} \\ S_{LR} & S_{LL} \end{bmatrix} = \begin{bmatrix} |S_{RR}|e^{j\varphi_{RR}} & |S_{RL}|e^{j\varphi_{RL}} \\ |S_{LR}|e^{j\varphi_{LR}} & -|S_{LL}|e^{j(\varphi_{LL}+\pi)} \end{bmatrix} \\ &= e^{j\varphi}\left\{e^{j\varphi_S} k_S \begin{bmatrix} o & j \\ j & 0 \end{bmatrix} + k_{\mathrm{d}} \begin{bmatrix} e^{j2\theta} & 0 \\ 0 & -e^{-j2\theta} \end{bmatrix} + k_{\mathrm{h}} \begin{bmatrix} e^{j2\theta} & 0 \\ 0 & 0 \end{bmatrix}\right\}\end{aligned} \tag{2.25}$$

The absolute phase, φ, can contain information about the scatterer. But, since its value depends on the distance between the radar and the target, it is considered as an irrelevant parameter. The phase, φ_S, represents a displacement of the sphere respect to the diplane and the helix components. The phase parameter, θ, stands for the orientation angle of the diplane and the helix components of the Krogager decomposition. Therefore, the physical interpretation of scattering target is done through the analysis of basis matrices, $[\mathrm{S}]_{sphere}$, $[\mathrm{S}]_{diplane(\theta)}$ and $[\mathrm{S}]_{helix(\theta)}$, respect to coefficients, φ_S, k_S, k_{d}, and k_{h} Target helicity is used to assess the symmetric nature of the target scattering. However, the diplane and helix scattering matrices are not orthogonal, so some even-bounce scatterers within the same resolution cell may result in diplane and helix components in the polarimetric decomposition procedure and causes some ambiguities between these two kinds of targets (Cloude & Pottier, 1996).

2.2.4.2. Incoherent decomposition

Coherent decompositions (i.e. Pauli and Krogager) can only be employed for characterizing the coherent or pure and simple scatterers (canonical targets). However in general, a pixel is more complex. It is made of distributed scatterers with the presence of speckle noise. Thus, incoherent approaches can be employed, in order to use averaging and statistical methods. In this case, incoherent decompositions based on second order statistics (coherency matrix or covariance matrix) are used for

the analysis of backscattering mechanisms. The principle is to decompose $[C]_{3\times3}$ or $[T]_{3\times3}$ as the combination of components corresponding to canonical objects. Then, these components are utilized for easier physical interpretation. Generally, incoherent decomposition theorems can be formulized as:

$$[C]_{3x3} = \sum_{i=1}^{K} p_i [C]_i \tag{2.26}$$

$$[T]_{3x3} = \sum_{i=1}^{K} q_i [T]_i \tag{2.27}$$

where p_i and q_i indicate the coefficients of $[C]_i$-and $[T]_i$-components.

According to which matrices, $[C]_{3\times3}$ or $[T]_{3\times3}$, and the index, *K*, different decompositions exist. In the following, three decomposition methods based on Huynen, eigenvector-eigenvalue and model-based are described. More detailed information can be found in the literature (Lee & Pottier, 2009).

2.2.4.2.1. Huynen-based decomposition

Huynen decomposition is used to extract the physical property of targets. Huynen type methods (i.e. Huynen/Barnes decomposition) are known as decomposition using the diagonal elements of the coherency matrix. They attempt to extract a single scattering matrix from the averaged Mueller or Kennaugh matrix. More principal mathematical information about Mueller and Kennaugh matrices can be found in (Chipman, 1995).

Huynen Decomposition

The Huynen target decomposition method factorizes matrix, $[T]_{3\times3}$, into a pure target matrix $[T_0]_{3x3}$ and a residue component *N*-target matrix, $[T_N]_{3\times3}$, which means that:

$$[T]_{3\times3} = [T_0]_{3\times3} + [T_N]_{3\times3} \tag{2.28}$$

The coherency matrix, $[T]_{3\times3}$, can also be represented by the parameters of the Kennaugh matrix (Lee & Pottier, 2009; Huynen, 1970; Unal & Ligthart, 1998; Chipman, 1995) is given by:

$$[T]_{3\times3} = \begin{bmatrix} \langle 2A_0 \rangle & \langle C \rangle - j\langle D \rangle & \langle H \rangle + j\langle G \rangle \\ \langle C \rangle + j\langle D \rangle & \langle B_0 \rangle + \langle B \rangle & \langle E \rangle + j\langle F \rangle \\ \langle H \rangle - j\langle G \rangle & \langle E \rangle - j\langle F \rangle & \langle B_0 \rangle - \langle B \rangle \end{bmatrix} \tag{2.29}$$

$$
\begin{aligned}
A_0 &= \frac{1}{4}\left|S_{HH} + S_{VV}\right|^2 \\
B_0 + B &= \frac{1}{2}\left|S_{HH} - S_{VV}\right|^2 \\
B_0 - B &= 2\left|S_{HV}\right|^2 \\
C + j\,D &= \frac{1}{2}\left|S_{HH} - S_{VV}\right|^2 + j\,Im(S_{VV}S_{HV}^*) \\
H - j\,G &= Re\left\{S_{HV}^*(S_{HH} + S_{VV})\right\} - j\,Im\left\{S_{HV}^*(S_{HH} + S_{VV})\right\} \\
E - j\,F &= Re\left\{S_{HV}^*(S_{HH} - S_{VV})\right\} - j\,Im\left\{S_{HV}^*(S_{HH} - S_{VV})\right\}
\end{aligned}
\tag{2.30}
$$

The Equation (2.30) shows the relationship between the elements of the target coherency matrix and elements of the scattering matrix. Huynen approach decomposes the vector (B_0, B, E, F) into two vectors of single target and residue target as follows:

$$B_0 = B_{0T} + B_{0N},\ B = B_T + B_N,\ E = E_T + E_N,\ F = F_T + F_N \tag{2.31}$$

where T and N denote the equivalent single target and N-target. Based on Huynen decomposition, $[T_0]_{3\times3}$ -matrix and $[T_N]_{3\times3}$ -matrix can be expressed as:

$$[T_0]_{3\times3} = \begin{bmatrix} \langle 2A_0 \rangle & \langle C \rangle - j\langle D \rangle & \langle H \rangle + j\langle G \rangle \\ \langle C \rangle - j\langle D \rangle & B_{0T} + B_T & E_T + jF_T \\ \langle H \rangle - j\langle G \rangle & E_T - jF_T & B_{0T} - B_T \end{bmatrix} \tag{2.32}$$

$$[T_N]_{3\times3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & B_{0N} + B_N & E_N + jF_N \\ 0 & E_N - jF_N & B_{0N} - B_N \end{bmatrix} \tag{2.33}$$

The basic properties of the Huynen decomposition are that targets are independent, and the average distributed N-target scattering is not changed under rotations of the antenna coordinate system about the line-of-sight (roll-invariant) (Nghiem, et al., 1992). Mathematically, this property can be expressed as:

$$T_N(\theta) = U_3(\theta)\, T_N\, U_3(\theta)^{-1} \tag{2.34}$$

where $U_3(\theta)$ is a 3 × 3 rotation matrix. Matrix, $[T_0]_{3\times3}$, has to be a 'rank 1'coherency matrix, so:

$$
\begin{aligned}
2A_0\left(B_{0T} + B_T\right) &= C^2 + D^2 \\
2A_0\left(B_{0T} - B_T\right) &= G^2 + H^2 \\
2A_0E_T &= CH - DG \\
2A_0F_T &= CG + DH
\end{aligned}
\tag{2.35}
$$

The parameters, A_0, C, D, H and G calculated from (2.30), and parameters, B_{0T}, B_T, E_T and F_T corresponding to the equivalent single target are calculated from (2.35). Thus, the three Huynen decomposition components are introduced as follows:

$$T_{11\mathrm{T}} = 2A_0$$
$$T_{22\mathrm{T}} = B_{0T} + B_T = \frac{C^2 + D^2}{2A_0} \tag{2.36}$$
$$T_{33\mathrm{T}} = B_{0T} - B_T = \frac{G^2 + H^2}{2A_0}$$

Holm and Barnes decomposition

Huynen decomposition was further extended in (Holm & Barnes, 1988) and (Yang, et al., 2006). The authors found that in addition to the Huynen decomposition components, there exist exactly two other series of decompositions components whose mixed-target-state components has the roll-invariant (rotational invariance) property. As mentioned in previous section, the basic Huyen assumptions are: 1) the single or pure target matrix, $[T_0]_{3\times3}$, and *N*-target matrix, $[T_N]_{3\times3}$, are independent and 2) the averaged distributed *N*-target is roll-invariant (Nghiem, et al., 1992). According to this, the vector space generated by $[T_N]_{3\times3}$ and the vector space generated by the pure target, $[T_0]_{3\times3}$, are mutually orthogonal. Additionally this orthogonality is maintained under rotations about the line-of-sight. Consequently, a vector, *q*, that belongs to the space generated by the pure target, $[T_0]_{3\times3}$, is orthogonal to space generated by any rotated *N*-target and:

$$T_N(\theta)\ q = 0 \Rightarrow U_3(\theta)\ T_N\ U_3(\theta)^{-1} q = 0 \tag{2.37}$$

where, $U_3(\theta)$ is a 3 × 3 rotation matrix. The condition in 2.37 is accomplished for any *q* only if:

$$U_3(\theta)^{-1} q = \lambda q \tag{2.38}$$

q is an eigenvector of the matrix $U_3(\theta)^{-1}$:

$$q_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},\ q_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ j \end{bmatrix},\ q_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ j \\ 1 \end{bmatrix} \tag{2.39}$$

So, this kind of decomposition is not unique, there are three ways to decompose the measured coherency matrix, $[T]_{3\times3}$ .The first one, which is produced from q_1, is Huynen decomposition, while the two other Huynen-like decompositions from q_2 and q_3 are known as Barnes and Holm decomposition.

By using Equations (2.37) and (2.28):

$$\left.\begin{array}{l}[T]_{3x3}\,q=[T_0]_{3x3}\,q+[T_N]_{3x3}\,q=[T_0]_{3x3}\,q=K_0K_0^{T*}q\\ q^{T*}[T]_{3x3}\,q=q^{T*}K_0K_0^{T*}q=\left|K_0^{T*}q\right|^2\end{array}\right\}\Rightarrow K_0=\frac{[T]_{3x3}\,q}{\sqrt{q^{T*}[T]_{3x3}\,q}} \tag{2.40}$$

By substitute Equations (2.29) and (2.39) in (2.40), the three normalized target vectors can be obtained as:

$$K_{01}=\frac{1}{\sqrt{\langle 2A_0\rangle}}\begin{bmatrix}\langle 2A_0\rangle\\ \langle C\rangle+j\langle D\rangle\\ \langle H\rangle-j\langle G\rangle\end{bmatrix}$$

$$K_{02}=\frac{1}{\sqrt{2(\langle B_0\rangle-\langle F\rangle)}}\begin{bmatrix}\langle C\rangle-\langle G\rangle+j\langle H\rangle-j\langle D\rangle\\ \langle B_0\rangle+\langle B\rangle-\langle F\rangle+j\langle E\rangle\\ \langle E\rangle+j\langle B_0\rangle-j\langle B\rangle-j\langle F\rangle\end{bmatrix} \tag{2.41}$$

$$K_{03}=\frac{1}{\sqrt{2(\langle B_0\rangle+\langle F\rangle)}}\begin{bmatrix}\langle H\rangle+\langle D\rangle+j\langle C\rangle+j\langle G\rangle\\ \langle E\rangle+j\langle B_0\rangle+j\langle B\rangle+j\langle F\rangle\\ \langle B_0\rangle-\langle B\rangle+\langle F\rangle+j\langle E\rangle\end{bmatrix}$$

According to the two normalized single target vector, $\vec{K}_{02}$, and $\vec{K}_{03}$, (Equation 2.41), BarnesI-matrix and BarnesII-matrix; $[Barns\ I]_{3\times3}=\langle\vec{K}_{02}\,.\,\vec{K}_{02}^{\dagger}\rangle$ and $[Barns\ II]_{3\times3}=\langle\vec{K}_{03}\,.\,\vec{K}_{03}^{\dagger}\rangle$, can be calculated recpectivelly.

2.2.4.2.2. Eigenvector-based decompositions

An eigenvector-based decomposition decomposes the incoherently averaged covariance (or coherency) matrix into three orthogonal scattering mechanisms by eigenvalues and eigenvectors. The decompositions by Cloude, Holm, Van Zyl, and Touzi are recalled here.

H/A/Alpha decomposition (Cloude – Pottier analysis method)

Cloude – Pottier incoherent decomposition model characterizes target scattering type with real descriptor from the complex coherency matrix as the incoherent sum of up to three coherency matrices, $[\mathrm{T}]_i$, representing three independent targets, each weighted by its appropriate positive real eigenvalue, λ_i, as proposed in (Cloude & Pottier, 1996):

$$[\mathrm{T}]_{3x3}=\frac{1}{N}\sum_{i=1}^{N}k_i.k_i^{\dagger}=\sum_{i=1,3}\lambda_i[\mathrm{T}]_i \tag{2. 42}$$

k_i is the target vector using the Pauli basis. As the coherency scattering matrix, $[T]_{3\times3}$, is Hermitian ($\mathrm{T}_{ij}=\mathrm{T}_{ji}^{*}$) and hence normal, and every normal matrix ($\mathrm{TT}^{\dagger}=\mathrm{T}^{\dagger}\mathrm{T}$) can be diagonalized

using a unitary matrix, U ($UU^{\dagger} = U^{\dagger}U = I$), the diagonalization of $[T]_{3\times3}$ is equivalent to finding an orthonormal polarization basis. $[T]_{3\times3}$ -matrix has three linearly independent eigenvectors. In other words, $[T]_{3\times3}$ can be decomposed as follows:

$$[T]_{3x3} = [U][\Sigma][U]^{-1} \tag{2. 43}$$

The 3×3 real diagonal matrix, $[\Sigma]_{3x3}$, contains the eigenvalues of $[T]_{3\times3}$:

$$[\Sigma]_{3x3} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}_{\lambda_1 \geq \lambda_2 \geq \lambda_3} \tag{2. 44}$$

In a simplified description, the Eigen-decomposition can be expressed as weighted sum of three matrices which are known as uncorrelated backscattering mechanisms descriptors (Cloude & Pottier, 1996; Cloude & Pottier, 1997):

$$[T]_{3x3} = \sum_{i=1}^{3} \lambda_i (u_i u_i^{\dagger}) = \lambda_1 u_1 u_1^{\dagger} + \lambda_2 u_2 u_2^{\dagger} + \lambda_3 u_3 u_3^{\dagger} \tag{2. 45}$$

where $u_i = \left[\cos\alpha_i e^{j\phi_i}, \sin\alpha_i \cos\beta_i e^{j(\delta_i+\phi_i)}, \sin\alpha_i \sin\beta_i e^{j(\gamma_i+\phi_i)}\right]$. The eigenvalues, λ_i, and eigenvectors, u_i , are known as primary parameters of the eigen decomposition of $[T]_{3x3}$. In fact, from Eigen-decomposition, the polarimetric entropy, H, the average scatter angle, α, and anisotropy, ***A***, are provided. These three secondary parameters are used to interpret physical scattering behaviors of targets. The entropy, H∈[0,1], measures the degree of the randomness of final target scattering with the probabilities, p_i . The p_i-parameters are the probabilities obtained from the λ_i -eigenvalues:

$$H = \sum_{i=1}^{3} p_i \log_3(p_i), \quad 0 \leq p_i = \frac{\lambda_i}{\sum_{i=1}^{3} \lambda_i} \leq 1 \tag{2. 46}$$

When H is low, it means that one dominant scattering mechanism happens, and $[T]_{3x3}$ has one non-zero eigenvalue. When H is increasing, depolarization is increasing. It means that the pure targets are a mixture of pure targets like distributed scattering targets.

Consequently, the anisotropy, A, gives the relative importance of the second and third eigenvalues (λ_2 and λ_3) and is described by:

$$A = \frac{p_2 - p_3}{p_2 + p_3} = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3} \tag{2. 47}$$

The anisotropy can be employed as an important criterion of discrimination and indicates the degree of polarization. The reason is that for low entropies, the second and third eigenvalues are highly affected by noise. As an example, Figure 2.3 presents four different configurations of the eigenvalues and the corresponding H and A-values. Finally, the α-parameter provides information on the dominant scattering mechanism which is computed as the weighted average of the α_i-values:

$$\alpha = \sum_{i=1}^{3} p_i \alpha_i \tag{2. 48}$$

where, p_i being the probability of α_i and α_i is obtained using Equation 2.45. When α=0, the odd or single-bounce scattering dominates. When α increases, it shows an anisotropic dihedral angle scattering. α=π/4 indicates the presence of dipole and the scattering mechanism corresponds to volume scattering. α=π/2 indicates the presence of dihedral structure and the scattering mechanism is due to double-bounce or even bounce scattering (Lee & Pottier, 2009; Lopez-Martinez et al., 2005).

However, there are ambiguities for some helix and dihedral targets, which are represented with the same α (α=π/2). The computation of α is also time-consuming because it requires eigenvalue and eigenvector calculations for each pixel of a polarimetric image as explained in (Praks, et al., 2009; Bhattacharya & Touzi, 2012). Entropy and Anisotropy are used to characterize media's scattering heterogeneity (Figure 2.3). Briefly, interpretation of the eigenvalues as probabilities led to the definition of entropy parameter, inspired by entropy from the information theory of Shannon (Shannon & Weaver, 1949). The α-angle parameter has been proposed for the eigenvector description and is interpreted as the scattering mechanism index, from surface, dipole, and double bounce (see (Lee & Pottier, 2009; Praks, et al., 2009) for more explanation).

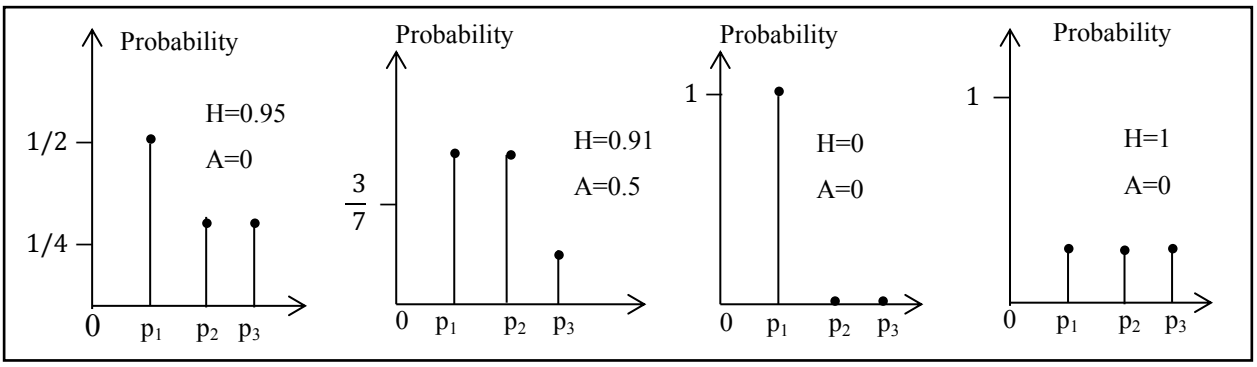


Figure 2. 3: Entropy (H) and Anisotropy (A) values for four different configurations of the eigenvalues, adapted from (Cloude & Pottier, 1997)

According to averaged parameters, α, β, δ, the unitary target vector of the mean dominant scattering mechanism ($u_0 = \lfloor \cos\alpha,\ \sin\alpha\cos\beta e^{j\delta},\ \sin\alpha\sin\beta e^{j\gamma} \rfloor$), and the target magnitude λ, the target vector K_0 is defined by:

$$K_0 = \sqrt{\lambda}\begin{bmatrix} \cos\alpha \\ \sin\alpha\cos\beta e^{j\delta} \\ \sin\alpha\sin\beta e^{j\gamma} \end{bmatrix} \tag{2. 49}$$

The *H/A/α*-matrix can be calculated using the normalized single target vector (Equation 2.49) by $[H/A/\alpha]_{3\times3} = \langle \vec{K}_0 \cdot \vec{K}_0^{\dagger} \rangle$. In addition, different descriptors corresponding to the different combinations between entropy (H) and anisotropy (A) images such as (1-H) (1-A), H (1-A), HA, and (1-H) A, are proposed in (Lee & Pottier, 2009).

Holm decomposition

Holm decomposition is an eigenvalue technique based on improving Huynen approach by combining single or pure target matrix, $[T_0]_{3\times3}$, *N*-target matrix, $[T_N]_{3\times3}$, and an unpolarized mixed state equivalent to noise, known as $[T_2]_{3\times3}$:

$$[T]_{3\times3} = \underbrace{[T_0]_{3\times3} + [T_N]_{3\times3}}_{Huynen\ decomposition(pure\ and\ mixed\ state)} + \underbrace{[T_2]_{3\times3}}_{unpolarized\ mixed\ state} \tag{2. 50}$$

When λ_1 is the largest eigenvalue, λ_2 and λ_3 the smallest ones, the diagonalized target coherency matrix, $[T]_{3\times3}$, is decomposed as follows:

$$\begin{aligned} [T]_{3\times3} &= (\lambda_1 - \lambda_2)u_1u_1^{\dagger} + (\lambda_2 - \lambda_3)(u_1u_1^{\dagger} + u_2u_2^{\dagger}) + \lambda_3 I \\ [Holm\ I]_{3\times3} &= \langle \vec{K}_1 \cdot \vec{K}_1^{\dagger} \rangle = (\lambda_1 - \lambda_2)u_1u_1^{\dagger} \\ [Holm\ II]_{3\times3} &= \langle \vec{K}_2 \cdot \vec{K}_2^{\dagger} \rangle = (\lambda_2 - \lambda_3)(u_1u_1^{\dagger} + u_2u_2^{\dagger}) \end{aligned} \tag{2. 51}$$

where K_1 and K_2 are single target vector. Therefore, the possible hybrid approach called as Holm decomposition is introduced using two normalized single target vector (Equation 2.51). More detailed explanation could be found in (Lee & Pottier, 2009; Unal & Ligthart, 1998).

Van Zyl decomposition

The Van Zyl decomposition (Van Zyl, 1993) is an incoherent eigenvector based decomposition using the covariance matrix, $[C]_{3\times3}$, with reflection symmetrical natural terrain assumption in the monostatic case. The decomposition results are mostly used for the interpretation of scattering from vegetated area (Lee & Pottier, 2009). The averaged covariance matrix in a medium with reflection symmetry has the general form:

$$[C]_{3\times3} = \sum_{i=1}^{3} \lambda_i (u_i u_i^{\dagger}) = \underbrace{\lambda_1 (u_1 u_1^{\dagger})}_{single\ backscattering} + \underbrace{\lambda_2 (u_2 u_2^{\dagger})}_{doubel\ backscattering} + \underbrace{\lambda_3 (u_2 u_2^{\dagger})}_{Volume\ backscattering} \tag{2. 52}$$

$$= \begin{bmatrix} \langle S_{HH} S_{HH}^* \rangle & 0 & \langle S_{HH} S_{VV}^* \rangle \\ 0 & \langle 2S_{HV} S_{HV}^* \rangle & 0 \\ \langle S_{VV} S_{HH}^* \rangle & 0 & \langle S_{VV} S_{VV}^* \rangle \end{bmatrix} = \alpha \begin{bmatrix} 1 & 0 & \rho \\ 0 & \eta & 0 \\ \rho * & 0 & \mu \end{bmatrix}$$

where

$$\begin{aligned} &\alpha = \langle S_{HH} S_{HH}^* \rangle, \rho = \langle S_{HH} S_{VV}^* \rangle / \langle S_{HH} S_{HH}^* \rangle \\ &\eta = \langle 2S_{HV} S_{HV}^* \rangle / \langle S_{HH} S_{HH}^* \rangle \\ &\mu = \langle S_{VV} S_{VV}^* \rangle / \langle S_{HH} S_{HH}^* \rangle \end{aligned} \tag{2. 53}$$

The eigenvalues and eigenvectors decomposition are given by:

$$\lambda_1 = \frac{\alpha}{2}\left(\mu + 1 + \sqrt{(\mu-1)^2 + 4|\rho|^2}\right), \quad u_1 = \sqrt{\frac{\mu - 1 + \sqrt{\Delta}}{(\mu - 1 + \sqrt{\Delta})^2 + 4|\rho|^2}} \begin{bmatrix} \dfrac{2\rho}{\mu - 1 + \sqrt{\Delta}} \\ 0 \\ 1 \end{bmatrix}$$

$$\lambda_2 = \frac{\alpha}{2}\left(\mu + 1 - \sqrt{(\mu-1)^2 + 4|\rho|^2}\right), \quad u_2 = \sqrt{\frac{\mu - 1 - \sqrt{\Delta}}{(\mu - 1 - \sqrt{\Delta})^2 + 4|\rho|^2}} \begin{bmatrix} \dfrac{2\rho}{\mu - 1 - \sqrt{\Delta}} \\ 0 \\ 1 \end{bmatrix} \tag{2. 54}$$

$$\lambda_3 = \alpha\ \eta, \ u_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \ with: \Delta = (\mu - 1)^2 + 4|\rho|^2$$

From Equation (2.54) and Equation (2.52), three decomposition components (odd, double and helix) are obtained.

Cloude-Pottier (α-β) model

In roll-invariant eigenvalue decomposition (H/A/α), the α-angle is used to describe target scattering type. The Cloude - Pottier (α - β) model (Cloude & Pottier, 1997) introduces two angles α and β to assign the coherent single to each target eigenvector. The target vector, $\vec{k}$, is expressed as follows:

$$\vec{k} = e^{j\phi_1} \left[\cos\alpha \quad \sin\alpha\cos\beta \cdot e^{j(\phi_2 - \phi_1)} \quad \sin\alpha\sin\beta \cdot e^{j(\phi_3 - \phi_1)} \right]^T \left|\vec{k}\right| \tag{2. 55}$$

where α, β, and ϕ_i $(i=1,3)$ are the scattering type, preferred orientation angle of the target, and target-phase parameters respectively (ϕ_1 angle denotes phase of $\left(S_{HH}+S_{VV}\right)$, ϕ_2 the difference of phase between $\left(S_{HH}-S_{VV}\right)$ and $\left(S_{HH}+S_{VV}\right)$, ϕ_3 the difference of phase between (S_{HV}) and $\left(S_{HH}+S_{VV}\right)$). The *Cloude*-matrix can be calculated using the normalized single target vector (Equation 2.55) and $\left[Cloud\,e\right]_{3\times3}=\left\langle \vec{k}\,.\,\vec{k}^{\dagger}\right\rangle$. However, for asymmetric targets, certain α-β model parameters vary with orientation angle, i.e., they are not roll-invariant (Touzi, 2007).

Touzi decomposition

The method introduced by Touzi and Charbonneau (Touzi & Charbonneau, 2002) aims at the characterization of the maximal symmetric target scattering. This decomposition is an incoherent decomposition which uses coherency matrix, $\left[T\right]_{3\times3}$, in the monostatic the reciprocal case with symmetry assumption. This method is used in some applications, such as wetlands characterization (Touzi, et al., 2009), ship identification (Touzi, et al., 2004), rice monitoring (Kun, et al., 2011), and urban classification (Bhattacharya & Touzi, 2012). Then, the Touzi decomposition proposes a roll-invariant coherent scattering model for symmetric or asymmetric targets. For each cell, the maximum roll-invariate coherent scattering model is given by (Touzi, et al., 2009; Touzi, 2007):

$$\vec{e}_T^{SV}=A.\exp(j\phi_S).\vec{V},\quad \vec{V}=\begin{bmatrix}\cos\alpha_S\cos2\tau_m\\ -j\cos\alpha_S\sin2\psi\sin2\tau_m+\cos2\psi\sin\alpha_S e^{j\phi_{\alpha S}}\\ -j\cos\alpha_S\cos2\psi\sin2\tau_m+\sin2\psi\sin\alpha_S e^{j\phi_{\alpha S}}\end{bmatrix} \tag{2. 56}$$

Thus, the coherent scatterer is uniquely characterized with five target parameters: α_S, the symmetry vector direction in the $\left(\left(S_{HH}+S_{VV}\right),\ \left(S_{HH}-S_{VV}\right)\right)$ basis, $\phi_{\alpha S}$, the phase difference between the vector components in the same basis, and the Kennaugh-Huynen maximum polarization parameters (orientation angle (ψ), helicity or ellipticity angle (τ_m), and amplitude (A)) and the absolute target phase, ϕ_S.

In the case of incoherent target decomposition, which is useful for non-coherent targets, the coherency matrix is diagonalized and each eigenvector is expressed in terms of the previous parameters:

$$\vec{k}_i=(\lambda_i,m_i,\psi_i,\tau_i,\alpha_{s\ i},\phi_{\alpha s\ i}),\ \ \lambda_i=\frac{\eta_i}{\eta_1+\eta_2+\eta_3} \tag{2. 57}$$

where η_i (i = 1; 3) is one eigenvalue of $[T]_{3\times3}$ and λ_i a normalized eigenvalue, that measures the relative energy carried by each single scattering. More detailed information can be found in literature (Touzi & Charbonneau, 2002; Touzi, et al., 2004; Touzi, 2007; Touzi, et al., 2009; Kun, et al., 2011; Bhattacharya & Touzi, 2012). In this study, we consider the effectiveness of these parameters for recognizing different roof surfaces in urban classification.

2.2.4.2.3. Model-based decompositions

The objective behind model-based decomposition is to fit physical scattering mechanisms to the polarimetric signatures in SAR data. Generally, such approach involves the fit of a combination of two or more simple scattering mechanisms (independent elements) to the polarimetric SAR observations based on covariance or coherency matrix. In this following section, Freeman-Durden three-components, Freeman two-components, and Yamaguchi three-components are reviewed.

Freeman – Durden three-component decomposition

Freeman-Durban three-component decomposition (Freeman & Durden, 1998) has been developed to fit three independent physical scattering models. The mechanisms are "volume scatters from a cloud of randomly oriented dipoles, even-bounce scatter from a pair of orthogonal surfaces with different dielectric constants and Bragg scatter from a moderately rough surface" (Lee & Pottier, 2009). This composite scattering model has been used to describe the polarimetric backscattering from natural distributed target areas in the P-L-C bands in (Yamaguchi, et al., 2005a).

Indeed, this model uses the reflection symmetry assumption $\langle S_{HH}S_{HV}^* \rangle = \langle S_{VV}S_{HV}^* \rangle = 0$ (the symmetry properties of remote sensing backscattering coefficients are presented in (Nghiem, et al., 1992)). According to this model, the measured total power can be expressed by:

$$Span(\mathrm{S}) = |S_{HH}|^2 + |S_{VV}|^2 + 2|S_{HV}|^2 = P_s + P_d + P_v \qquad (2.\ 58)$$

where the scattering powers, P_s, P_d, and P_v, correspond to surface scattering from a first-order Bragg surface scatterer, double-bounce scattering from a dihedral corner reflector, and volume scattering such as from a canopy scatterer. Typically, in this model-based decomposition, the measured covariance matrix is modelled by three scattering models, as recalled in (Yang, et al., 2001):

$$[C]_{3\times3} = f_s C_s + f_d C_d + f_v C_v \qquad (2.\ 59)$$

where f_s, f_d, and f_v are the contribution coefficients, and C_s, C_d, and C_v are the basis covariance matrices for the surface, double, and volume scattering components, respectively. C_s, C_d and C_v basis matrices are defined under the reflection symmetry assumption (see Lee & Pottier, 2009 for instance).

Respect to these basis matrices and the covariance matrix definition, the relationship between parameters and elements of covariance matrix are as follows:

$$\begin{aligned}
&\left\langle \left|S_{HH}\right|^2 \right\rangle = f_s \left|\beta\right|^2 + f_d \left|\alpha\right|^2 + \frac{3}{8} f_v \\
&\left\langle \left|S_{VV}\right|^2 \right\rangle = f_s + f_d + \frac{3}{8} f_v \\
&\left\langle S_{HH} S_{VV}^* \right\rangle = f_s \beta + f_d \alpha + \frac{f_v}{8} \\
&\left\langle \left|S_{HV}\right|^2 \right\rangle = \frac{f_v}{8}
\end{aligned} \tag{2. 60}$$

There are four Equations with five parameters. After assumption on the dominant mechanism (according to the sign of $Re(\langle S_{HH} S_{VV}^* \rangle)$, α or β is fixed, the decomposition result is the power of each scattering component, which can be derived from the f_s -, f_d - and f_v -coefficients:

$$P_s = f_s(1+\left|\beta\right|^2),\ P_d = f_d(1+\left|\alpha\right|^2),\ P_v = f_v \tag{2. 61}$$

Consequently, the scattered power, P_s, P_d and P_v, can be employed to generate a RGB image to present all the color-coded polarimetric information in a single view.

Freeman two-component decomposition

Freeman two-component simplifies the previous one and was proposed for SAR observation of forests (Freeman, 2007). This decomposition is based on two assumptions. Firstly, a canopy often has the property of reflection symmetry (Nghiem, et al., 1992) in a reciprocal medium ($S_{HV} = S_{VH}$). Secondly, a ground scattering is represented by either a double bounce (from a pair of orthogonal surfaces), or Bragg scattering from a moderately rough surface. This seems to be a reasonable approximation for a natural structure. Thus, the system solution involves the estimation of four unknown parameters from four equations which are defined as follows:

$$\begin{aligned}
&\left\langle \left|S_{HH}\right|^2 \right\rangle = f_G + f_v,\quad \left\langle \left|S_{VV}\right|^2 \right\rangle = f_G \left|\alpha\right|^2 + f_v \\
&\left\langle S_{HH} S_{VV}^* \right\rangle = f_G \alpha + \rho f_v,\quad 2\left\langle \left|S_{HV}\right|^2 \right\rangle = f_v \rho
\end{aligned} \tag{2. 62}$$

where f_G, α and ρ, f_v correspond to the double bounce or single bounce scattering component and the volume scattering component respectively.

Yamaguchi four-component decomposition

Yamaguchi (Yamaguchi, et al., 2005b) decomposition models the covariance matrix, $[C]_{3\times3}$, or the coherency matrix, $[T]_{3\times3}$, as the contribution of four scattering mechanisms under the non-reflection

symmetry condition ($\left\langle S_{HH} S_{HV}^{*} \right\rangle \neq 0$ and $\left\langle S_{VV} S_{HV}^{*} \right\rangle \neq 0$). The basic idea behind this is that the backscattering power in heterogeneous area (i.e. urban area) can be expressed as a linear sum of four different scattering power components whose categories are surface scattering power, double-bounce scattering power, volume scattering power, and helix scattering power. The fourth component is taken into account for the co-polarization and the cross-polarization correlations which regularly appear in complex urban area scattering. According to this definition, the measured power, P, can be expressed by:

$$P = \left|S_{HH}\right|^2 + \left|S_{VV}\right|^2 + 2\left|S_{HV}\right|^2 = P_s + P_d + P_v + P_h \tag{2. 63}$$

P_s, P_d, P_v, and P_h correspond to surface, double bounce, volume, and helix (non reflection symmetric cases) contributions, respectively. Hence, the Yamaguchi decomposition expresses the measured covariance matrix, $\left[C\right]_{3\times3}$, as follows:

$$\left[C\right]_{3\times3} = f_s C_s + f_d C_d + f_v C_v + f_h C_h \tag{2. 64}$$

where f_s, f_d, f_v, and f_h, and are the contribution coefficients, and C_s, and C_d are defined as in the Freeman decomposition. The C_v-matrix, depends on the value of $10log\left\langle \left|S_{VV}\right|^2 \right\rangle / \left\langle \left|S_{HH}\right|^2 \right\rangle$. The C_h-matrix is introduced for circular polarization scattering from man-made structures. The scattering powers are obtained by:

$$P_s = f_s(1+\left|\beta\right|^2),\ P_d = f_d(1+\left|\alpha\right|^2),\ P_v = f_v,\ P_h = f_h \tag{2. 65}$$

In urban area, because of the orientation angle shift, the radar cross section is affected and that causes errors in the physical interpretation of these features. Thus, the response from oblique man-made features (buildings that are oriented about the path of flight) decomposes into volume scattering component or cross-polarization components instead of double-bounce component. A more detailed presentation and comparison of the above mentioned decomposition can be found in (Lee & Pottier, 2009).

2.2.4.3. Polarimetry parameters

The eigenvalues of the coherency matrix and covariance matrix obtained from transformation by unitary matrices through different target decomposition methods have been introduced for providing a geometrical interpretation of the backscattering target behaviors. In the following, several polarization parameters based on analysis of scattering matrix or using eigenvalues will be described. Some of these parameters show the degree of polarization of a partially polarized electromagnetic field such as Degree of Polarization (*DoP*), or are based on eigenvalue interrelation, like the Pedestal Height (*PH*), the Polarization Asymmetry (*PA*), and the Polarization Phase Difference (*PPD*). Besides, some other

parameters are based on statistical expressions like Polarization Ratio (*PR*), or the Backscattering Correlation Coefficient (*BCC*). We expect them to be useful for accurate target discrimination in urban area.

1- Span: the span of the scattering matrix is defined as the sum of squares of the components of the scattering matrix. However, it has no information about how the total power is distributed among the polarimetric channels (see for instance (Lee & Pottier, 2009)):

$$Span(S) = |S_{HH}|^2 + |S_{VV}|^2 + 2|S_{HV}|^2 = \lambda_1 + \lambda_2 + \lambda_3 \tag{2. 66}$$

2- Backscattering Correlation Efficient (*BCE*): the correlation coefficients between different polarization components of scattering matrix in linear and circular bases are given by:

$$\begin{aligned} \rho_{23} &= \frac{\langle S_{HV} S_{VV}^* \rangle}{\sqrt{\langle S_{HV} S_{HV}^* \rangle \langle S_{VV} S_{VV}^* \rangle}}, \quad \rho_{RL} = \frac{\langle S_{RR} S_{LL}^* \rangle}{\sqrt{\langle S_{RR} S_{RR}^* \rangle \langle S_{LL} S_{LL}^* \rangle}} \\ \rho_{12} &= \frac{\langle S_{HH} S_{HV}^* \rangle}{\sqrt{\langle s_{HH} s_{HH}^* \rangle \langle s_{HV} s_{HV}^* \rangle}}, \quad \rho_{13} = \frac{\langle S_{HH} S_{VV}^* \rangle}{\sqrt{\langle S_{HH} S_{HH}^* \rangle \langle S_{VV} S_{VV}^* \rangle}} \end{aligned} \tag{2. 67}$$

In addition, the correlation coefficient, constructed from the off-diagonal elements of complex covariance matrix and coherency, in linear and circular bases, are given by:

$$\rho_{L,C_{xy}} = \frac{\langle C_{xy} \rangle}{\sqrt{\langle C_{yy} \rangle \langle C_{xx} \rangle}}, \quad \rho_{L,T_{xy}} = \frac{\langle T_{xy} \rangle}{\sqrt{\langle T_{xx} \rangle \langle T_{yy} \rangle}}, \quad \rho_{C,C_{ij}} = \frac{\langle C_{ij} \rangle}{\sqrt{\langle C_{jj} \rangle \langle C_{ii} \rangle}}, \quad \rho_{C,T_{ij}} = \frac{\langle T_{ij} \rangle}{\sqrt{\langle T_{jj} \rangle \langle T_{ii} \rangle}} \tag{2. 68}$$

where, $(x, y) \in \{x, y = 1, 2, 3\}$, and $\rho_{L,C_{xy}}$ and $\rho_{L,T_{xy}}$ stand for the *BCE* in linear basis while, $(i, j) \in \{i, j = 1, 2, 3\}$, and $\rho_{C,C_{ij}}$, $\rho_{C,T_{ij}}$ stand for the *BCE* in circular basis.

3- Polarization Ratio (*PR*): the different ratios are Co-polarization, Cross-polarization and De-polarization ratios which describe how much a target depolarizes the incident polarized wave.

$$\begin{aligned} &Co-polarization\ ratio \Rightarrow \frac{\langle |S_{VV}|^2 \rangle}{\langle |S_{HH}|^2 \rangle}, \quad Cross-polarization\ ratio \Rightarrow \frac{\langle |S_{HV}|^2 \rangle}{\langle |S_{HH}|^2 \rangle}, \frac{\langle |S_{VV}|^2 \rangle}{\langle |S_{HV}|^2 \rangle} \\ &De-polarization\ ratio \Rightarrow \frac{2\langle S_{HV} S_{HV}^* \rangle}{\langle S_{HH} S_{HH}^* \rangle + \langle S_{VV} S_{VV}^* \rangle} \end{aligned} \tag{2. 69}$$

The *PR* description parameter has correlation with surface roughness. In practice, ratios are transformed with log-function.

4- Differential Reflectivity (*DR*): differential reflectivity can be obtained from scattering matrix. The *DRs* are defined by:

$$DR = \langle S_{HH} - S_{HV} \rangle, \langle S_{HH} - S_{VV} \rangle, \langle S_{HV} - S_{HH} \rangle, \langle S_{VV} - S_{HH} \rangle, \langle S_{VV} - S_{HV} \rangle, \langle S_{HV} - S_{VV} \rangle \quad (2.70)$$

5- Degree of Polarization (*DoP*): the *DoP* could be as an important factor for discriminating surfaces, because it measures how much the effect of multiple reflections is contained in a target area. The *DoP* will be defined from the reflected wave's Stokes vector, Mueller matrix, or eigenvalues of the covariance matrix as proposed in (Ulaby, et al., 1992):

$$DoP = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2 + \lambda_3} \quad (2.71)$$

The eigenvalues, λ_1 and λ_2, are invariant under polarization basis transformation, thus *DoP* is an important basis-independent parameter.

6- Polarization Phase Difference (*CPX* and *XPD*): co-polarized Phase Difference (*CPD*) and Cross-polarized Phase Difference (*XPD*) are physical parameters which can be obtained directly from the covariance matrix, or the Stokes scattering matrix, without eigen analysis as proposed in (Ulaby, et al., 1992), and used in (Oh, et al., 2009). The *CPX* and *XPD* are defined by:

$$CPD \Rightarrow \varphi_{HH-VV} = tan^{-1}\left(\frac{\mathrm{Im}(S_{HH} S_{VV}^*)}{\mathrm{Re}(S_{HH} S_{VV}^*)}\right), \; XPD \Rightarrow \varphi_{HV-VV} = tan^{-1}\left(\frac{\mathrm{Im}(S_{HV} S_{VV}^*)}{\mathrm{Re}(S_{HV} S_{VV}^*)}\right) \quad (2.72)$$

7- Conformity Coefficient (*CC*): when no dominant scattering mechanism can be clearly defined, *CC*-parameter allows discriminating the three main scattering. *CC*-parameter can characterize the relative power between surface or dihedral scattering and volume scattering using the covariance matrix. For instance, it is used in (Truong-Loi, et al., 2010). It is given by:

$$CC = \frac{2(Re(\mathrm{C}_{13}) - \mathrm{C}_{33})}{\mathrm{C}_{11} + \mathrm{C}_{22} + \mathrm{C}_{33}} \approx \frac{2(Re(S_{HH} S_{VV}^*) - |S_{HV}|^2)}{|S_{HH}|^2 + 2|S_{HV}|^2 + |S_{VV}|^2} \quad (2.73)$$

8- Shannon Entropies (*SE*): the Shannon entropies (Shannon, 1948) refer to this fact: each pixel of a PoLSAR image consists in partially coherent and partially polarized backscattering with Gaussian probability distributions. In circular Gaussian distribution case, it is decomposed into intensity, SE_{I} - parameter that depends on the total backscattered power, $Tr\left([T]_{3x3}\right)$, and, polarimetric SE_p - parameter that depends on degree of polarization. It is analyzed with Gaussian probability distribution in (Refregier & Morio, 2006). These parameters are defined by:

$$SE = SE_{\mathrm{I}} + SE_{\mathrm{p}} \;, \; SE_{\mathrm{I}} = 3\log\left(\frac{\pi \mathrm{e} Tr\left([T]_{3x3}\right)}{3}\right) \;, \; SE_p = \log\left(3^3 \frac{det\left([T]_{3x3}\right)}{\left(Tr\left([T]_{3x3}\right)\right)^3}\right) \quad (2.74)$$

9- Single bounce Eigenvalue Relative Difference (*SERD*) and Double bounce Eigenvalue Relative Difference (*DERD*): in (Allain, et al., 2005), are presented the two analytical expressions which are

derived from eigenvalue and eigenvector decomposition of the coherency matrix under the reflection symmetry assumption.

10- The Kozlov Anisotropy (*KA*): the results of scattering matrix eigenvalue decomposition are two orthogonal polarization states with two eigenvalue, s_1 and s_2, which are known as pseudo eigenvalues. The *KA* shows the relative importance of these eigenvalues (Lee & Pottier, 2009):

$$KA = \frac{|s_1|^2 - |s_2|^2}{|s_1|^2 + |s_2|^2} \quad , 0 \le KA \le 1 \tag{2. 75}$$

11- *H/A/Alpha* eigenvalue parameters: several important parameters have been proposed from the eigenvalue set and are used in this work (Lee & Pottier, 2009):

- **Anisotropy12 (*A*12)**

$$A_{12} = \frac{p_1 - p_2}{p_1 + p_2} = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}, \tag{2. 76}$$

- **Pedestal Height (*PH*):** the Pedestal Height value shows the scattering randomness in the scattering process. It represents unpolarized backscattering energy. It is defined by the ratio between the minimal and the maximal eigenvalues:

$$PH = \frac{\min(\lambda_1, \lambda_2, \lambda_3)}{\max(\lambda_1, \lambda_2, \lambda_3)} = \frac{\lambda_3}{\lambda_1}, \quad 0 \le PH \le 1 \tag{2. 77}$$

- **Polarization Fraction (*PF*):** it relies on the Barnes-Holm decomposition and measures the portion of the un-polarized return, as used in (Qi, et al., 2010). *PF* is given by:

$$PF = \frac{P_{\mathbf{max}} - P_{\mathbf{min}}}{P_{\mathbf{max}} + P_{\mathbf{min}}} = 1 - \frac{3\lambda_3}{Span(\mathrm{S})} = 1 - \frac{3\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}, \quad 0 \le PF \le 1 \tag{2. 78}$$

  where, the second term shows the percentage of the total power that remains completely un-polarized. The $P_{\mathbf{min}}$ and $P_{\mathbf{max}}$ are minimal and maximal power over both co-polarized and cross-polarized signatures. When *PF*=0, the averaged return is completely un-polarized, and *PF*=1 shows that the averaged return is completely polarized.

- **Polarization Asymmetry (*PA*):** it also relies on the Barnes-Holm decomposition and shows the relative strength of the two first polarimetric scattering mechanisms. For instance, it is used in (Qi, et al., 2010), and is given by:

$$PA = \frac{(\lambda_1 - \lambda_3) - (\lambda_2 - \lambda_3)}{(\lambda_1 - \lambda_3) + (\lambda_2 - \lambda_3)} = \frac{(\lambda_1 - \lambda_2)}{(Span(\mathrm{S}) - 3\lambda_3)}, \quad 0 \le PA \le 1 \tag{2. 79}$$

- **The Lueneburg Anisotropy (*LA*):** it shows the scattering randomness. For instance, it is used in (Touzi, et al., 2004) and is defined by:

$$LA = \sqrt{\frac{3}{2}\frac{p_2^2 + p_3^2}{p_1^2 + p_2^2 + p_3^2}} \quad ,0 \le LA \le 1 \tag{2. 80}$$

2.2.4.4. Synthesis

In this work, we can use different classical decomposition methods such as model-based, Huynen-based, and eigenvalue-based ones. But, actually, none of them exclusively can satisfy our expectations, due to some symmetric assumptions, the loose of dominant polarimetric scattering properties, or predefined thresholds which cause some specific drawbacks. For example, Freeman decomposition refers to reflection symmetry assumption, introduce an independent volume scattering model, so that, it does not give any explanations about some targets, which have mixed backscattering. Because of this, the double-bounce, odd-bounce, and helix power components are negative. So these methods are extended by introducing an adaptive volume scattering model (Arii, et al., 2011), rotation of covariance matrix (Yamaguchi, et al., 2011), and adopted the latest method by double transformation on the coherency matrix (Cui, et al., 2012). However, we observe with our data that the polarimetric descriptors which are extracted from a rotation of the coherency matrix to minimize the cross-polarized component as in (Yamaguchi, et al., 2011) are very noisy for X band PoLSAR data.

Also, for example, the mathematical Cloude and Pottier eigenvalue-based decomposition, is not pertinent to interpret the results in term of known scattering, when targets have mixed backscattering. Besides, it also does not use α-complex parameter component in scattering vector explanation. However, this method is extended using hybrid scattering mechanisms (Yang, et al., 2013). Then, we propose to use the combination of coherent, incoherent decomposition descriptors, the components of backscattering matrix, coherency and covariance matrices, and some discrimination parameters (Tables 2.1, 2.2, and 2.3). The subscripts used in these tables represent the related element in corresponding category (Figure 2.2). For example, the $[Barnes\ I]_A$ matrix is related to Barnes-I matrix in amplitude, or (S_{12}) refers to the amplitude parameter in the first row and second column of Sinclair matrix, or ($T_{13\text{-}phase}$) refers to the phase parameter in the first row and third column of coherency matrix, or $[Barnes\ I]_{T_{11\text{-}I}}$ is referred to the intensity parameter in the first diagonal element in Barnes-I matrix in dB, or $T_{13\text{-}I}$ is referred to the intensity parameter in the first row and third colum of coherency matrix in dB and so on (A logarithm scale reduces a large range of values and show input quantities in especially appropriate form).

Table 2. 1: The positive real polarimetric descriptors considered in the statistical analysis (Chapter 3). $<\cdot>$ is the average operator. *A,* stands for the amplitude of the descriptor.

Polarimetric Parameters	**Expression**	**#**
H/A/Alpha element (eq. 2.49)	$\left[H / A / Alpha\right]_A$	6
Huynen elements (eq. 2.32)	$\left[Huynen\right]_A$	6
Cloude elements (eq. 2.55), Holm1,2 (eq. 2.51)	$\left[Cloude\right]_A, \left[Holm\ I\right]_A, \left[Holm\ II\right]_A$	18
Barnes1,2 (eq. 2.41)	$\left[Barnes\ I\right]_A, \left[Barnes\ II\right]_A$	12
Coherency matrix elements (eq. 2.11)	$\{T_{11\text{-}A}, T_{12\text{-}A}, T_{22\text{-}A}, T_{23\text{-}A}, T_{13\text{-}A}\}$	5
Span (eq. 2.66)	$\sqrt{Span}$	1
Sinclair matrix elements (eq. 2.3)	$S_{A,11}, S_{A,12}, S_{A,22}$	3
Sinclair ratio components (eq. 2.69)	$\{\lvert S_{VV}\rvert / \lvert S_{HH}\rvert, \lvert S_{HV}\rvert / \lvert S_{HH}\rvert, \lvert S_{VV}\rvert / \lvert S_{HV}\rvert\}$	3
Texture parameters (Table 2. 2)	$UNI - \sqrt{T_{22-A}}$, $UNI - \sqrt{T_{33-A}}$,	4
	$CON - \sqrt{T_{22-A}}$, $CON - \sqrt{T_{33-A}}$	
Total Descriptors	58	

Table 2. 2: Polarimetric decomposition descriptors considered in the SSM based on GA-MI (Chapters 7and 8).

Decomposition Technique	***Features***	#
Huynen (eq. 2.35 and eq. 2.36)	$[Huynen]_{T_{12\text{-}phase}}$, $[Huynen]_{T_{13\text{-}phase}}$, $[Huynen]_{T_{23\text{-}phase}}$ $[Huynen]_{T_{11\text{-}I}}$, $[Huynen]_{T_{22\text{-}I}}$, $[Huynen]_{T_{33\text{-}I}}$	6
Krogager (eq. 2.25)	$[Krogager]_{s,d,h}$	3
Freeman III, II (eq. 2.61 and eq. 2.62)	$[Freeman\ I]_{s,d,v}$,$[Freeman\ II]_{v,G}$	5
BarnesI,II (eq. 2.41)	$[Barnes\ I,II]_{T_{12\text{-}phase}}$, $[Barnes\ I,II]_{T_{13\text{-}phase}}$, $[Barnes\ I,II]_{T_{23\text{-}phase}}$ $[Barnes\ I,II]_{T_{11\text{-}I}}$, $[Barnes\ I,II]_{T_{22\text{-}I}}$, $[Barnes\ I,II]_{T_{33\text{-}I}}$	12
Van Zyl (eq. 2.52 and eq 2.54)	$[VanZyl]_{s,d,v}$	3
Holm I,II (eq. 2.51)	$[Holm\ I,II]_{T_{12\text{-}phase}}$, $[Holm\ I,II]_{T_{13\text{-}phase}}$, $[Holm\ I,II]_{T_{23\text{-}phase}}$ $[Holm\ I,II]_{T_{11\text{-}I}}$, $[Holm\ I,II]_{T_{22\text{-}I}}$, $[Holm\ I,II]_{T_{33\text{-}I}}$	12
H/A/Alpha (eq. 2.49)	$[H/A/\alpha]_{T_{12\text{-}phase}}$, $[H/A/\alpha]_{T_{13\text{-}phase}}$, $[H/A/\alpha]_{T_{23\text{-}phase}}$ $[H/A/\alpha]_{T_{11\text{-}I}}$, $[H/A/\alpha]_{T_{22\text{-}I}}$, $[H/A/\alpha]_{T_{33\text{-}I}}$	6
Yamaguchi (eq. 2.65)	$[Yamaguchi\ VI]_{s,d,v,h}$, $[Yamaguchi\ III]_{s,d,v}$	7
Cloude (eq. 2.45)	$[Cloude]_{T_{12\text{-}phase}}$, $[Cloude]_{T_{13\text{-}phase}}$, $[Cloude]_{T_{23\text{-}phase}}$ $[Cloude]_{T_{11\text{-}I}}$, $[Cloude]_{T_{22\text{-}I}}$, $[Cloude]_{T_{33\text{-}I}}$	6
Pauli (eq. 2.22)	$[Pauli]_{\lvert a\rvert^2,\lvert b\rvert^2,\lvert c\rvert^2}$	3
H/A/Alpha (eq. 2.45, eq. 2.47, eq. 2.46, eq. 2.48 and eq. 2.49)	*{H/A/α, H(1-A), (1-H)A, HA, (1-H)(1-A), β, γ, δ}*	10
Touzi (eq. 2.56, and eq. 2.57)	$\{\lambda_i, \psi_i, \tau_i, \alpha_{s\ i}, \phi_{\alpha s\ i}, (i = 1,3)\}$	15
Total descriptors	*88*	

Table 2. 3: Original polarimetric components and discrimination parameters considered in classification (Chapter 6). $<\cdot>$ is the average operator. S_{xy} represents the scattering coefficient of the targets, ***x*** the polarization of the incident field (H, V), ***y*** the polarization of the scattered field (H, V).

Discriminators parameters		**Features**	#
Span (eq. 2.66)		*Span*	1
Differential Reflectivity (eq. 2.70)		*DR*	6
Conformity Coefficient (eq. 2.73)		*CC*	1
Shannon Entropies (eq. 2.74)		SE_I, SE_P	2
Degree of Polarization (eq. 2.71)		*DoP*	1
Polarization Phase Difference (eq. 2.72)		*CPD, XPD*	2
Cross Correlation Coefficient (eq. 2.67 and eq. 2.68))		*CCC-A*, *CCC-phase*	2
H/A/Alpha, eigenvalue parameters	(eq. 2.78)	*PF*	8
	(eq. 2.77)	*PH*	
	(eq. 2.76)	A_{12}	
	(eq. 2.80)	*LA*	
	(eq. 2.75)	*KA*	
	(eq. 2.79)	*PA*	
		SERD and DERD	
Polarization Ratio (eq .2.69)	$\left\langle \lvert S_{xy} \rvert^2 \right\rangle / \left\langle \lvert S_{x'y'} \rvert^2 \right\rangle \quad (x, y, x', y' = (H, V))$		3
De-polarisation Ratio Ratio (eq .2.69)	$2\left\langle S_{HV} S_{HV}^* \right\rangle / (\left\langle S_{HH} S_{HH}^* \right\rangle + \left\langle S_{VV} S_{VV}^* \right\rangle)$		1
Sinclair matrix elements (eq. 2.3)	$S_{A,11}, S_{A,12}, S_{A,22}$		3
Coherency matrix elements	$\left\{ \begin{matrix} T_{12-1}, T_{13-1}, T_{23-1}, T_{22-1}, T_{11-1}, T_{33-1}, (T_{12-Phase}), \\ (T_{13-Phase}), (T_{23-Phase}) \end{matrix} \right\}$		9
Total descriptors	39		

2.3. Texture characteristic of SAR image

A texture feature describes the spatial distribution of grey values in one descriptor. There are four different mathematical procedures to generate a texture feature: statistical, geometrical, model-based,

and image processing, which consists in Fourier transform, convolution filters, co-occurrence matrix, spatial autocorrelation, or fractals as explained in (Tuceryan & Jain, 1998). The texture information can be used for improving the classification results (Zhang, et al., 2009). This part aims at providing various texture parameters that might improve urban classification.

2.3.1. Extraction of texture features

The widely common second order statistical tool used in texture analysis is the one based on the definition of the joint probability of pairs of pixels. This joint probability, $p_{d,\theta}(i,j)$, is derived from the so-called gray level co-occurrence matrix (GLCM) as proposed by (Haralick, 1979). Let an image be denoted by: $\left\{I(x,y), 0 \le x \le N_x - 1, 0 \le y \le N_y - 1\right\}$. $p_{d,\theta}(i,j)$,of the L×L GLCM for a distance, $d = (dx, dy)$, and direction, θ, is defined by:

$$p_{d,\theta}(i,j) \Rightarrow \left\{((r,s),\ (t,v))\ :\ I(r,s)=i,\ I(t,v)=j\right\} \tag{2. 81}$$

where $(r,s) \in N_x \times N_y$ and $(t,v) = (r + dx, s + dy)$

L is the number of gray levels in the image.

According to the definition (2.81), each element of this matrix shows the relative frequencies $p_{d,\theta}(i,j)$, of two neighboring values, one with gray level, i ,and the other with gray level, j, that are separated with the distance, d, and angle, θ. GLCM is:

$$\mathrm{GLCM}_{d,\theta}^{L\times L} = \begin{bmatrix} p_{d,\theta}(0,0) & p_{d,\theta}(0,1) & \cdots & p_{d,\theta}(0,L-1) \\ p_{d,\theta}(1,0) & p_{d,\theta}(1,1) & \cdots & p_{d,\theta}(1,L-1) \\ \cdots & \cdots & p_{d,\theta}(i,j) & \cdots \\ p_{d,\theta}(L-1,0) & p_{d,\theta}(L-1,1) & \cdots & p_{d,\theta}(L-1,L-1) \end{bmatrix} \tag{2. 82}$$

The direction of θ can be either horizontally (θ=0°), vertically (θ=90°), or diagonally (θ= 45°, or 135°).

As all texture information is present in such matrix, various types of statistical descriptors (e.g., energy, entropy, contrast, homogeneity,…) can be derived as proposed in (Schistad Solberg & Jain, 1997).

We describe some of these parameters, which are used in the current study (see Chapter 7, Table 7.1 and Table 7.2):

1- Uniformity (*UNI*) gives a measurement of local textural uniformity.

2- Contrast (*CON*) explains the local variation of grey levels.

As the Sinclair matrix can only characterize coherent or canonical scatterers, we prefer to use second order polarimetry representation like coherency or covariance matrix to analyze the distributed scatterer texture. Thus, textural parameters are computed on the root square of two images: T_{22} and

T_{33}, which are two diagonal components of coherency matrix $[T]_{3\times3}$. Moreover, many textural descriptors can be derived from the co-occurrence matrix (Schistad Solberg & Jain, 1997) but we select only Uniformity and Contrast.

Table 2. 4: Mathematical expressions of statistical texture parameters from GLCM. $p_{d,\theta}(i,j)$, joint probability and $(i,j)^{th}$ element in the co-occurrence matrix of $\sqrt{T_{22\text{-}A}}$ and $\sqrt{T_{33\text{-}A}}$.

Texture parameter	**Expression**
UNI (uniformity)	$\sqrt{\sum_{i,j} p^2_{d,\theta}(i,j)}$
CON (contrast)	$\sum_{i,j} \lvert i-j \rvert^2 \, p_{d,\theta}(i,j)$

2.4. Morphological feature extraction

The morphological feature extraction method focuses on structural information (shape and structure) in the image domain. A good overview of historical note of mathematical morphology is given in (Matheron & Serra, 2000). The fundamental theories concern set theory for binary image (Gonzales & Woods, 2001), fuzzy set theory for gray images and color images (Burillo et al., 2003; Burillo et al., 2001) and operator definitions in mathematical morphology (Soille, 2003; Serra, 1982).

The morphology operators are translating in an image with a set of a known shape, called a structuring element (SE). The center pixel of the structuring element defines the pixel which is being processed. The elementary morphological operations are Erosion and Dilation. For binary images, the result of applying Erosion operator to an image is an output image, which shows where the SE fits the objects in the image. The result of applying Dilation operator shows, where the SE hits the objects in the image. All other morphological operators such as opening and closing can be expressed in terms of Erosion and Dilation. The idea behind applying morphological operation is to recover objects according to the structuring element (SE) (Benediktsson et al., 2003). An early study about mathematical morphology operations to extract objects with special shapes such as roads, rivers and villages can be found in (Vincent & Beuche, 1989). Then, the composition of geodesic opening (in the geodesic case, transformations are defined conditionally to a reference set) (Fauvel, et al., 2005) and using closing operations of different sizes (also known as morphological reconstruction) in order to build a morphological profile (DMP) are presented in (Vincent, 1993; Fauvel, et al., 2005) respectively. Moreover, DMP gives information about the size of an object and its contrast with its neighboring. This approach is explained in (Pesaresi & Benediktsson, 2001). A multi-scale DMP

reconstruction approach with different structuring elements is adopted to explore a large range of potential object sizes which is used for urban land cover classification (Benediktsson et al., 2003) and 3D building extraction (Sportouche & Tupin, 2011). Here, we mainly consider the elementary morphological operations like opening and closing, associated with a structural element with the expected size of the desired buildings, and Otsu threshold (Otsu, 1979) to produce binary image, and from which we propose a semi-automatic building extraction.

In the following sections, the Otsu threshold technique and basic morphological operations like Dilation, Erosion, opening and closing are presented. Then, a morphological building extraction approach is proposed.

2.4.1. Otsu method

An exhaustive survey of thresholding methods can be found in (Sezgin & Sankur, 2004; Sahoo, et al., 1988). The main idea behind Otsu threshold is to minimize the within-class variance that turns out to be the same as maximizing the between-class variance. Otsu's thresholding technique divides the gray image into two classes C_1 (object interest) and C_2 (background) at gray level t such that $C_1 = \{1, 2, ..., t\}$ and $C_2 = \{t+1, t+2, ..., N\}$ where N is the total number of gray levels in the image. Let the number of pixels at the ith gray level is n_i, then, the probability of occurrence of gray level i is defined by:

$$p_i = \frac{n_i}{\mathrm{N}_t} \tag{2. 83}$$

N_t : total number of pixels in the image. The probabilities of C_1 and C_2 classes are w_1 and w_2 which are defined by:

$$w_1(t) = \sum_{i=1}^{t} p_i, \qquad w_2(t) = \sum_{i=t+1}^{\mathrm{N}} p_i = 1 - w_1(t) \tag{2. 84}$$

Thus, the means and variances for the two classes can be calculated by:

$$\begin{aligned} \mu_1(t) &= \sum_{i=1}^{\mathrm{t}} i p_i / w_1(t), \qquad \sigma_1^2(t) = \sum_{i=1}^{\mathrm{t}} (i - \mu_1(t))^2 p_i / w_1(t) \\ \mu_2(t) &= \sum_{i=t+1}^{\mathrm{N}} i p_i / w_2(t), \qquad \sigma_2^2(t) = \sum_{i=t+1}^{\mathrm{N}} (i - \mu_2(t))^2 p_i / w_2(t) \end{aligned} \tag{2. 85}$$

The σ_{in}^2, μ_T, and σ_{T}^2 are in-class variance, total mean and variance of the whole image respectively and are defined by:

$$\sigma_{in}^2(t) = w_1(t)\sigma_1^2(t) + w_2(t)\sigma_2^2(t), \qquad \mu_T = \sum_{i=1}^{N} ip_i, \qquad \sigma_T^2 = \sum_{i=1}^{N}(i-\mu_T)^2 p_i \tag{2. 86}$$

An optimal threshold, t*, is obtained by minimizing in-class variance, σ_{in}^2, or maximizing the between-class variance σ_B^2:

$$t^* = \underset{1\le t\le N}{\text{Argmax}}\left(\frac{\sigma_B^2(t)}{\sigma_T^2}\right), \quad \text{where,} \quad \sigma_B^2(t) = \sigma_T^2 - \sigma_{in}^2(t) \tag{2. 87}$$

However, Otsu's method is efficient for separating an image into two classes where two types of fairly distinct classes exist in the image (Hongzhi & Ying, 2008). Some limitations of Otsu method like uni-scale, or global restriction are discussed in (Cheriet, et al., 1998). To overcome this problem, some modified algorithms, like adapted model to multi classe are presented in (Moghaddam & Cheriet, 2011).

2.4.2. Morphological Operations

Morphology is based on set theory. In a morphological operation, the value of each point, (x, y), of output image is defined according to a membership function, *f* (x, y), of digital input image based on a structuring element (SE).

2.4.2.1. Dilation morphological operation

The dilation of A by B is defined as the set of all points x such that B_X hits A. Simply, the result is equivalent to a union of all image translated with respect to structural element, *B*. The dilation of a binary image, $A \subset Z^2$, by structural element, $B \subset Z^2$ is defined by:

$$A \oplus B = \left\{x \mid (B)_x \cap A \neq \varnothing\right\}, \text{or} \quad A \oplus B = \left\{x \mid \left[(B)_x \cap A\right] \subset A\right\} \tag{2. 88}$$

B_X denotes the translation of B. The dilation operation measures how well the shape represented by *B* fits above *A*. The dilation definition of gray image *f* by *b* is given by:

$$(f \oplus b)(s,t) = \max\left\{\begin{matrix} f(s-x, t-y) + b(x,y) \mid \\ (s-x), (t-y) \in D_f; (x,y) \in D_b \end{matrix}\right\} \tag{2. 89}$$

D_f and D_b are the domain of *f* and *b,* respectively. This operation measures maximum difference between *f* and structuring element *b*. In general, dilation operation causes to expand an original image, and hence, the small hole can be eliminated.

2.4.2.2. Erosion morphological operation

The erosion of A with B is the translation in which B is completely covered by A, and defined by:

$$A\Theta B = \left\{ x \mid (B)_x \subset A \right\} \tag{2. 90}$$

This operation measures the minimum difference between A and structuring element, B. The erosion measures how well the shape represented by B fits under A. The erosion definition of gray image *f* by *b* is given by the expression:

$$(f\ \Theta b)(s,t) = \min \left\{ \begin{matrix} f(s+x, t+y) - b(x,y) \mid \\ (s+x), (t+y) \in D_f; (x,y) \in D_b \end{matrix} \right\} \tag{2. 91}$$

This operation measures the minimum difference between *f* and structuring element *b*. The erosion measures how well the shape represented by *b* fits under *f*. In general, erosion operation causes to shrink the original image, and hence, small peak can be eliminated.

2.4.2.3. Opening morphological operation

The definition of a morphological opening of an image is erosion followed by dilation, using the same structuring element for both operations, and is defined by:

$$AoB = (A\Theta B) \oplus B \tag{2. 92}$$

The opening result is generally smooth and the spot-like noise is removed (Gonzales & Woods, 2001).

2.4.2.4. Closing morphological operation

The morphological closing of an image consists in a dilation followed by erosion with the same structuring element, and is defined by:

$$A \bullet B = (A \oplus B)\Theta B \tag{2. 93}$$

The closing operation fills the holes. In other words, it has the effect of 'clustering' each spatial point to a connected set (Soille, 2003; Serra, 1982; Gonzales & Woods, 2001).

Here, the main purpose is to access a map building without any needs to use auxiliary data such as DEM. The semi-automatic solution, which is proposed here, can produce a map that is expected to represent the building footprints. To produce binary image we can use the Otsu threshold method (Otsu, 1979). Moreover, we can also convert gray image to binary image by introducing suitable threshold manually. Then, we remove regions with less than 'option' pixels with morphologically open binary image (remove all small objects) over previous binary image. The effect of this operation is to remove all structures that are both smaller than the 'option' pixels and brighter than the surroundings such as spot point and noise. Thus, all streets, railways following with spot points are eliminated. Then, applying 'closing minus opening' using the square SE which maintains the actuall size of objects and removes all remain noises and also fill the holes in our interesting area..

2.5. Summary I

This chapter has provided short accounts of some various methods of feature extraction techniques used for obtaining PoLSAR attributes in order to achieve effectively information of interest. At the beginning, an overview about SAR polarimetry and different methods for producing backscattering formulation (Stokes, Jones formulation, Scattering matrix, Mueller matrix, Covariance and Coherency matrices) was presented. Further, the chapter reviewed different target decomposition approaches i.e. coherent and incoherent methods and several polarimetry discriminators. Afterwards, feature texture extraction from matrix (GLCM) based on gray-level spatial dependencies was described. The list of SAR attributes that are used in this study were detailed in tables: SAR parameters from coherent and incoherent target decompositions are summarized in Table 2.1, discrimination parameters which are extracted from backscattering analysis are listed in Table 2.2., the texture parameters based on (GLCM) are listed in Table 2.3.

In addition, a method for extracting building in high resolution imagery is described. The building descriptor result might improve urban area classification which is described in Chapter 8.

Chapter 3:
Feature selection

3. Feature selection

3.1. Feature selection

In classification domain, the main idea behind a subset selection among M descriptors, $\mathrm{X} = \{x_i, i = 1,...,M\}$, is to find a subspace of m features that "optimally" characterizes the classes, $C = \{c_1, c_2, ..., c_k\}$ with k the number of classes (Peng, et al., 2005). It aims at choosing the optimal descriptors by eliminating features with little or no interesting information (irrelevant information) and also features which may be redundant. The subset selection not only reduces the dimensional space of the descriptors, and as consequence, increases the algorithm speed performance (reduced computational effort), but also simplifies classifier complexity and improves the classification accuracy through removing irrelevant or noisy data. A good overview of feature selection methods along with different perspectives has been presented in (Pudil, et al., 1998; Dash & Liu, 1997).

Generally, feature selection process involves four basic parts: subset generation (specifying a candidate subset for evaluation), subset evaluation (to introduce an appropriate evaluation criterions), stopping iteration based on many constrains which are introduced in initialization part (for example, number of features or iteration) or during an execution algorithm, and finally result validation (based on prior knowledge or classification results from different selected feature subsets) (Liu & Yu, 2005).

A subset generation may include two basic decisions. First decision is the search starting point assumptions, like *forward* (start with an empty set and successively add features), *backward* (start with a full set and successively remove features) or *bi-directional* (start with both ends and add and remove features simultaneously). Second decision is the search strategy, such as complete or exponential (Somol, et al., 2004), sequential (Pernkopf & O'Leary, 2001), and finally random search (Liu & Yu, 2005).

Some basic algorithms are categorized into: *Filter* (use of criteria; independent from classification algorithm) (Lanzi, 1997), *Wrappe*r and also *Embedded* (use of measurement of classification accuracy to evaluate each attribute; dependent on the classification algorithm) (Guyon & Elisseeff, 2003), and *Hybrid* models (Camps-Valls, et al., 2010). The criteria (evaluation function) in filter approaches may be based on distance measurement, information measurements, and dependency or correlation (Arauzo-Azofra, et al., 2008; Dash & Liu, 1997). The dependent criteria (in wrapper algorithm) may be based on classifier error rate measures (Dash & Liu, 1997).

A widely conceptual base of feature selection, taxonomy of feature selection algorithms, and categorization of feature selection algorithms are described in (Jain & Zongker, 1997; Liu & Yu,

2005). In addition, the comparison of different algorithms, their evaluation, and some of the main problems of these methods are presented in (Guyon & Elisseeff, 2003; Pudil, et al., 1994).

Some limitations (for instance, computational cost of wrapper approaches) of wrapper and filter approaches are discussed in (Vinh, et al., 2012). To overcome these addressed limitations, some others integrate advanced measurements such as Mutual Information (MI) with advanced searching strategies algorithm like as Genetic Algorithm (GA) are presented in (Bhanu & Lin, 2003; Li & Zeng, 2010; Vinh, et al., 2012; Ludwig, et al., 2009). In this research, we focus on the discussion of Max-Relevance and Min-Redundancy (mRMR) which is extracted from MI and use GA in order to solve optimization problem and iterative searching algorithm.

In the following sections, we first present a brief overview of MI in order to describe the mRMR as validation criterion. Secondly, we describe the genetic algorithm (GA) conception (Siedlecki & Sklansky, 1989) which is used as iterative searching algorithm. Finally, we describe the feature selection algorithm using GA based on MI. This selection will be done before the classification using train samples (ground truth).

3.1.1. Mutual Information

MI is a measurement of dependence for nonlinear correlation, without assuming the distribution of variables and meanwhile, the random variable can be numerical or categorical. Generally, MI describes quantitatively the relationship between two features or between a feature and a class (Vinh, et al., 2012). Simply, it describes how much information two random variables share with each other. MI has been used to measure the dependency in many literatures (Monahan & DelSole, 2009).

The following notations describe the description of MI which are taken from Shannon's information theory (Shannon & Weaver, 1949) and based on probability theory and statistics. According to Shannon's information theory, the relations between MI and the entropy are easily expressed in venn diagram (Figure 3. 1). The notation for the description of MI and proposed GA algorithm is taken from (Ludwig & Nunes, 2010; Peng, et al., 2005; Ludwig, et al., 2009).

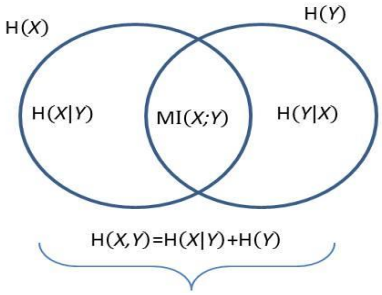


Figure 3. 1: The relationship between MI and entropy

The entropy, *H*, is a measure of the uncertainty of a random variable with the probability, p_x , defined by:

$$H(X) = -\sum_i p_x(\mathrm{x}_i)\log p_x(\mathrm{x}_i) \tag{3. 1}$$

The Shannon's entropy represents the average amount of information contained in random variable, *X*. Therefore, the conditional entropy quantifies the amount of information needed to describe a random variable, *Y,* given the value of another random variable, *X*, and is given by:

$$H(X\,|Y) = -\sum_i \sum_j p_{\mathrm{x,y}}(\mathrm{x}_i,\mathrm{y}_j)\log p_{\mathrm{x|y}}(\mathrm{x}_i|\mathrm{y}_j) \tag{3. 2}$$

Then, in the same way, $H(Y\,|X)$ is the result of averaging of $H(Y\,|X = x_i)$ over all possible values, x_i, that *X* may take.

The joint entropy of *X* and *Y* with joint pdf, $p_{\mathrm{x,y}}(\mathrm{x}_i,\mathrm{y}_j)$, is defined by:

$$H(X,Y) = -\sum_i \sum_j p_{\mathrm{x,y}}(\mathrm{x}_i,\mathrm{y}_j)\log p_{\mathrm{x,y}}(\mathrm{x}_i,\mathrm{y}_j) = H(X\,|Y) + H(Y) = H(Y\,|X) + H(X) \tag{3. 3}$$

Joint entropy is the measurement how much entropy is contained in a joint system of two random variables. According to Figure 3.1, MI is a measurement of mutual dependence of two variables based on the entropy and defined by:

$$MI(X;Y) = H(X) - H(X\,|\,Y)$$

$$= H(X) + H(Y) - H(X,Y) = \sum_i \sum_j \left[p_{\mathrm{x,y}}(\mathrm{x}_i,\mathrm{y}_j)\log\left(\frac{p_{\mathrm{x,y}}(\mathrm{x}_i,\mathrm{y}_j)}{p_{\mathrm{x}}(\mathrm{x}_i).p_{\mathrm{y}}(\mathrm{y}_j)}\right)\right] \tag{3. 4}$$

Let us note, X_i, the feature variables and C is the class variable. $S_m = \{X_1,...,X_m\}$ is a feature set variable, and $c = \{c_1, c_2, ..., c_k\}$ is a class label set with *k* classes. The MI between class label and feature set is defined by:

$$MI(S_m;C) = \sum_{\mathrm{c}\in\Omega_C} \sum_{\mathrm{S}_i \in \Omega_{S_m}} \left[p_{\mathrm{S}_m,C}(\mathrm{s}_i,c)\log\left(\frac{p_{\mathrm{S}_m,C}(s_i,\mathrm{c})}{p_{\mathrm{S}_m}(s_i).p_C(\mathrm{c})}\right)\right] \tag{3. 5}$$

Ω_C is the set of all possible class labels, and Ω_{S_m} is the domain of variable S_m-variable.

The main goal of feature selection description is to achieve the maximum relevance between feature and class and the minimum redundancy between features. The Relevance and Redundancy are defined by:

$$D(S_m,C) = \frac{1}{|S_m|}\sum_{x_i\in\Omega_{S_m}} MI(X_i;C), \quad |S_m| = m$$

$$R(S_m) = \frac{1}{|S_m|^2}\sum_{\substack{x_i,x_j\in\Omega_{S_m}\\ i\neq j}} MI(X_i;X_j) \tag{3. 6}$$

Therefore, the final subset of features is obtained by optimizing the following equation:

$$\hat{S}_m = \arg\max_{S_m} \left[U(S_m) = D(S_m) - R(S_m) \right] \quad (3.7)$$

The joint probability distribution, $p_{\mathrm{x,y}}(\mathrm{x}_i, \mathrm{y}_j)$, are calculated in term of a joint cumulative distribution function (joint probability mass function) for each of two dicrite random variable.

3.1.2. Genetic Algorithm

The total number of possible combinations of individual is huge. GA is one solution to the previous optimization problem. GA is one of widely common iterative searching algorithm that aims at optimizing an evaluation function. It is inspired from natural evolution derived from the Darwinian theory of survival of the fittest individuals. The GA was initially introduced by John Holland in the early 1970s. A Genetic feature subset selection process consists in four basic steps (Dash & Liu, 1997): a population generation (each population consists in some individuals usually known as set of chromosomes), an evaluation criterion (each individual is selected based on evaluation criterion or evaluation function), termination (it may be based on iteration number, feature number, threshold, or population convergence), and a result validation to verify the validity of the selected solution (Togan & Daloglu, 2008).

Typically, a GA uses three basic probability operators to generate a new population from the previous one: selection which operates on each individual, crossover (recombination) which operates on pair of features and mutation (random modifications) which operates on features. These last operators introduce new individuals, which may be better than the previous ones. Generally, there are different potential solutions to the selection problem such as traditional selection (roulette wheel, deterministic sampling, etc.), ranking, tournament selection and alternative selection mechanisms. For crossover there are different methods like crossover from one point, two points or more, uniform crossover or crossover based on non-uniform distribution. For mutation, also different types exist: flip bit, boundary, uniform, non-uniform, and Gaussian. Moreover, to use the elitism method after creating new generation by crossover and /or mutation, it copies the best features (or a few best ones) to new population. In other words, elitism can increase very rapidly performance of GA, because it prevents from losing the best found solutions. More detailed information about GA, taxonomy of GA operators, and some examples of applications can be found in (Weisstein, 2013).

The proposed algorithm for GA feature selection using MI description which is noted by GA-MI consists in the following steps:

Initialization: the initial parameters are:

1- Population size or number of individuals in one generation (N). If there are too few individuals, GA has a few possibilities to perform crossover and only a small part of search space is explored and there will not be enough chances to find the global optimum. Besides, if there are too many individuals, GA

slows down. However, some research experiment shows that after some limits (like: problem difficulty, pressure factor, evaluation function, and number of selected features) which depend mainly on search space, it is not useful to increase the population size, because it does not make solving the problem better (Alander, 1992).

2- Maximum number of generations (iteration).

3- Pressure factor. It defines the strength of selection method and focus on the utilization of the search space. In fact, it controls the speed of the search performance during GA procedure.

4- Number of interested features (*m*)

5- Evaluation threshold. The evaluation function stops when the evaluation value in the current individual becomes less than the threshold (our-specified) and this threshold is set to minimize the evaluation value between [0, 0. 5].

Initial Generation:

Generally, the GA starts with a population of random individuals

Evaluation:

The evaluation function is based on Max-Relevance and Min-Redundancy (mRMR) (Equation (3.7)). For this purpose, the evaluation function for each individual of current population is calculated. Then, the algorithm sorts the individuals according to their fitness values, ranking the indexes of the individuals from high to low. Each individual in higher positions after ranking will have more chance for participating in the crossover operations. For example, the first individual has a maximal probability for selection.

Selection: to choose an individual from current generation for entering to next generation,…, the selection is going according to ranking value of previous procedure. Simply, each individual takes a score based on mRMR evaluation function, then according to this ranking, the population is sorted (rearranged) and only those with a score above validation threshold are used in crossover.

Crossover:

The aim behind crossover is to produce new individuals which could be better than their two parents. Simply, each feature (gene) of a new individual (chromosome) is taken from parents in feature level selection procedure. The proposed crossover probability is based on an asymmetric random distribution which is used by (Ludwig & Nunes, 2010; Ludwig, et al., 2009):

$$f(r,a,i) = round\left[(i-1)\frac{e^{a\,r}-1}{e^{a}-1}\right], i \in [1\ N] \qquad (3.8)$$

where $r \in [0, 1]$ is a random variable with uniform distribution, *i*, is the index of individual, and *a* is a positive constant which is known as pressure factor. The pressure factor controls the search performance during GA procedure that must be initialized (Ludwig, et al., 2009; Ludwig & Nunes, 2010). The $f(r,a,N)$ gives the individual index with high ranking indexes, meanwhile the chances to

select features with small ranking still exist. It is used to select two parents. Then, the crossover consists in selecting the best parent.

Elitism: One or more individuals with higher ranking are repeated in the next population. The number of these individuals is the elitism value. Here, we select only the individual with rank 1, to increase performance of GA procedure.

Termination: A new generation is iteratively processed through three steps: 1) evaluation based on mRMR evaluation function, which leads to rearrangement generation according to ranking values (individual-level selection), 2) crossover (attribute-level selection) and 3) termination criterion by which the algorithm decides whether to continue searching or stop. Simply, the search procedure continues (evaluation, selection, crossover) until the termination criterion, is met.

Here, the termination criterion is a pre-determined number of generations.

Validation Result: The result from GA feature selection may be compared with other feature selection methods.

Here, we validate the result through evaluation of pre-determined parameters. First for selecting a convenient number of generations, we consider the relationship between evaluation function and number of generation. Secondly, for selecting a convenient feature number (m), we consider the relationship between classification overall accuracy and number of features. More details will be discussed in Chapter 7 (section 7.1.2).

In our case, Figure 3.2 represents the flowchart of feature selection based on GA-MI.

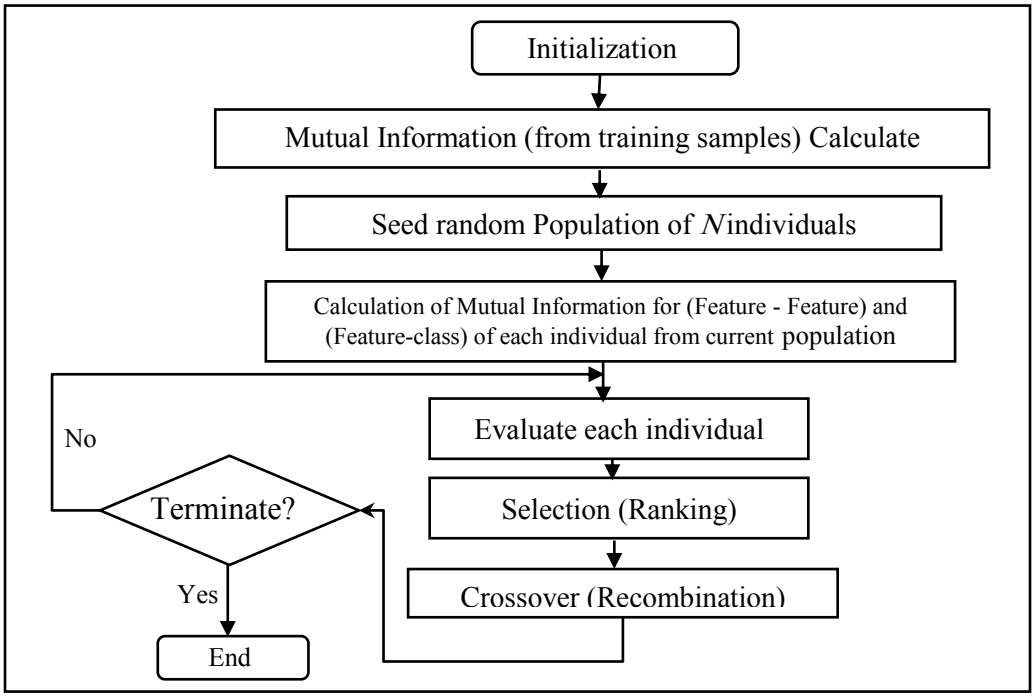


Figure 3. 2: Flowchart of feature selector based on GA-MI

3.2. Summary II

This chapter described feature selection methods and used genetic algorithm as an iterative searching method to replace complex classification (all descriptors) to the simple one (optimum descriptors) with higher accuracy.

The results of Subset Selection Method (SSM) based on GA-MI might improve urban area classification which is described in Chapter 7 and Chapter 8.

Part II:
Classification

Chapter 4:

Statistics

4. Statistics

4.1. Introduction

Statistical modeling describes the statistical characteristics of a variable. It includes the mathematical expression such as the probability density function (pdf), the statistical moments and an estimation of the pdf parameters. They are applied in several works such as statistical modeling in (Moser, et al., 2006b; Bombrun & Beaulieu, 2008; Salaza, et al., 1998; Anfinsen & Eltoft, 2011), image classification (Simonetto, et al., 2002; Tison, et al., 2004), and image segmentation (Benboudjema, et al., 2007; Galland, et al., 2009), speckle noise filtering (Chitroub, et al., 2002). A review of litterature published on statistical modeling of SAR images in the last years are described in (Gao, 2010).

This chapter gives a brief review and recalls the SAR models and the Mellin transform framework which consists in Mellin transform definition, and its properties. Then, several specific uni-variate pdfs which are adapted for polarimetric SAR images are described. The Mellin second kind statistics of these pdfs are given. Moreover, two estimation parameters methods, based on method of log-cumulant (MoLC) and Maximum Likelihood estimation (MLE), following with several Goodness-of-Fit (GoF) tests, are described.

4.1.1. Review of SAR statistical models

Generally, SAR image are characterized by the pattern of speckle. It is an inherent phenomenom in SAR image. It refers to coherent processing of backscattered signals from multiple distributed targets (predominantly multiplicative noise) (Goodman, 1963). However it can be reduced by multilook processing or using spatial filtering.

The product model suggests that the complex signal, is received by the SAR sensor from target, derives from two statistically independent random variables; the square root of a positive random variable (backscattering intensity texture), X, and the speckle noise, Y_p: $z_p = \sqrt{X}\, Y_p$, where, p=3 in monostatic case and reciprocal condition. We have: $Z_3 = \left[S_{HH}, \sqrt{2} S_{HV}, S_{HH} \right]^T$.

Moreover, some assumptions have been considered for modeling speckle, such as the fully developed speckle. In this case, the scatterers are statistically independent in the resolution cell, the number of scatterers is large, the instantaneous phase and the scattering amplitude are independent, the phase has Gaussian distribution over the range [0,2π], and there is not any dominant scattering (Goodman, 1963). The SAR intensity can be modeled by an uni-variate exponential distribution:

$p(v) = \lambda \exp(-\lambda v)$, where λ^{-1}, presents the first order moment of the backscattering signal, while the SAR amplitude can be modeled by a Rayleigh distribution: $p(r) = 2r\lambda \exp(-\lambda r^2)$.

However, in heterogeneous scene, the SAR amplitude cannot be fitted with a Rayleigh model. Different types of model have been proposed like Lognormal (Goldstein, 1973), K (Jakeman & Pusey,1976), Weibull (Oliver, 1993), Inverse Gamma. Moreover, the models with heavy tail and more flexibility with three parameters such as Fisher pdf are also proposed in (Tison, et al., 2003; Bombrun & Beaulieu, 2008).

In the case of single look and multi-variate data, the speckle part may be modeled by a zero mean complex Gaussian distribution whose covariance matrix is derived from the different polarimetric channels. This information refers to their second order moments (Khan, 2012; Krylov, et al., 2011).

For texture modeling, the Generalized Inverse Gaussian (GIG) is discussed (more information are given in (Krylov, et al., 2011; Khan, 2012). We will not address such models in this study.

Recently, these models and more complex and multi parameter distributions have been analyzed through the Mellin transform.

When the moments and cumulants are generated from first and second characteristics functions based on the Fourier transform, it is known as first kind statistic. Recently, log-moments and log-cumulants are generated from first and second characteristics functions based on the Mellin transform, which is known as second kind statistics (Nicolas, 2002). This method is very flexible and convenient to generate second kind statistics for different type of pdfs.

Here, we use the Mellin transform and its second kind statistics to model non-Gaussian distributions such as Fisher, Gamma, K, Weibull, and Lognormal distribution function for uni-variate PoLSAR descriptors like polarimetry ratio, coherency components (Table 2.1), and texture parameters (Table 2.4).

4.2. Uni-variate Mellin transform

The Mellin Transform is an analytical tool for studying the distribution of independent positive random variables (Epstein, 1948), which was pointed out by Robert Hjalmar Mellin (1854-1933). The uni-variate Mellin transform was first developed for statistical modeling of radar images and was introduced as second kind statistics by (Nicolas, 2002). A good overview of historical note of Mellin transform is given in (Anfinsen, 2010). More details on mathematical definitions including a list of fundamental properties, some examples, and Mellin transforms of common functions, can be found in (Poularikas, 1999; Debnath & Bhatta, 2007; Bertrand, et al., 2000; Nicolas, 2006).

Furthermore, the extension of the Mellin transform for multi-variate functions is given in (Mathur & Krishna, 1977). The statistical analysis of multi-look polarimetric radar data using the multi-variate Mellin transform is given in (Anfinsen & Eltoft, 2011). Some special functions such as the Meijer's

functions (a generalisation of hypergeometrical functions) are defined from inverse Mellin transforms in (Nicolas, 2011) and could be used for SAR modelling.

4.2.1. Definition and basic properties of Mellin transform

The Mellin's integral maps the function f into the function φ which is defined on the complex plane with positive variable u. The integral does exist for complex values, $s = a + jb$, such that a-value verifies $a_1 < a < a_2$, where a_1 and a_2 depend on the function f(u) to transform (Epstein, 1948; Bertrand, et al., 2000). The Mellin transform and the inverse Mellin transform are defined by:

$$\mathcal{M}\{f(u)\}(s) = \int_0^{+\infty} u^{s-1} f(u)du = \varphi(s)$$

$$\mathcal{M}^{-1}\{\varphi(s)\}(u) = f(u) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} u^{-s}\varphi(s)ds$$

where f(u) is a real valued function defined on $\Re$ and s-variable is a complex number.

To introduce second kind statistics, some of the basic operational properties of uni-variate Mellin transform are described as follows (the proofs of some operational properties can be found in (Debnath & Bhatta, 2007)):

1- Scaling property

$$\mathcal{M}\{f(au)\}(s) = \int_0^{+\infty} u^{s-1} f(au)du = a^{-s}\varphi(s), \quad a > 0 \tag{4. 1}$$

2- The Multiplication by u^a:

$$\mathcal{M}\{u^a f(u)\}(s) = \int_0^{\infty} u^{(s+a)-1} f(u)du = \varphi(s+a) \tag{4. 2}$$

3- Raising the independent variable to a real power:

$$\mathcal{M}\{f(u^a)\}(s) = \int_0^{+\infty} u^{s-1} f(u^a)du = \int_0^{+\infty} x^{(\frac{s-1}{a})} f(x)(\frac{1}{a}x^{\frac{1}{a}-1}dx) = a^{-1}\int_0^{+\infty} x^{(\frac{s}{a}-1)} f(x)(dx) = \frac{1}{a}\varphi(\frac{s}{a}) \tag{4. 3}$$

4- Inverse and multiple of variable:

$$\mathcal{M}\{\frac{1}{u} f(\frac{1}{u})\}(s) = \varphi(1-s) \tag{4. 4}$$

5- Multiplication by *logu*:

$$\mathcal{M}\{(\log u) f(u)\}(s) = \frac{d}{ds}\varphi(s) \tag{4. 5}$$

The multiplication property can be proved by using: $u^{(s-1)} = e^{(s-1)\log u}$.

$$\frac{d}{ds}u^{s-1} = \frac{d}{ds}e^{(s-1)\log u} = u^{s-1}\log u$$

$$\mathcal{M}\{(\log u)f(u)\}(s) = \int_0^{+\infty} u^{s-1}\log u f(u)du = \int_0^{+\infty}\frac{d}{ds}u^{s-1}f(u)du = \frac{d}{ds}\varphi(s)$$

6- Multiplication by a power of *logu*:

$$\mathcal{M}\{(\log u)^k f(u)\}(s) = \frac{d^k\mathcal{M}\left[f(u)(\log u)^k\right](s)}{ds^k} = \frac{d^k}{ds^k}\varphi(s) \tag{4. 6}$$

Equation (4.6) is the result of:

$$\frac{d^k}{ds^k}u^{s-1} = \frac{d^k}{ds^k}e^{(s-1)\log u} = u^{s-1}(\log u)^k$$

7- Mellin transform of derivatives (Debnath & Bhatta, 2007):

$$\mathcal{M}\{f^{(n)}(u)\}(s) = \int_0^{+\infty} u^{s-1}f^{(n)}(u)du = (-1)^n\frac{\Gamma(s)}{\Gamma(s-n)}\mathcal{M}\{f(u)\}(s-n) \tag{4. 7}$$

8- Mellin convolution (Debnath & Bhatta, 2007; Epstein, 1948):

$$h = f\,\hat{*}\,g \Leftrightarrow \mathcal{M}\{h\}(s) = \mathcal{M}\{f\}(s)\mathcal{M}\{g\}(s) \tag{4. 8}$$

If *f* and g are pdfs, then *h* is also a pdf (i.e., *h*(u) ≥ 0 ∀u ∈ $\Re^+$, and $\mathcal{M}\{h\}(1) = 1$

9- Mellin inverse convolution

$$f = h\,\hat{*}^{-1}\,g \Leftrightarrow h = f\,\hat{*}\,g$$

10- Mellin correlation:

$$f = h\,\hat{\otimes}\,g \Leftrightarrow \mathcal{M}\{h\}(s) = \mathcal{M}\{f\}(s)\mathcal{M}\{g\}(2-s) \tag{4. 9}$$

where $\left(f\,\hat{\otimes}\,g\right)(u) = \int_0^{+\infty} f(u\,y)g(y)ydy$

By introducing the new function g_I , $g_I(u) = \frac{1}{u^2}g(\frac{1}{u})$, the Mellin transform of g_I is defined by:

$$\mathcal{M}\{g_I(u)\}(s) = \mathcal{M}\{g(u)\}(2-s) \tag{4. 10}$$

And finally the relationship between correlation and convolution according to Equation (4.8) and Equation (4.9) is defined by:

$$f\,\hat{\otimes}\,g = f\,\hat{*}\,g_I = \int_0^{\infty} f\left(\frac{u}{y'}\right)\frac{u}{y'^2}g\left(\frac{1}{y'}\right)\frac{dy'}{y'}, \quad where\ y = \frac{1}{y'} \tag{4. 11}$$

4.2.2. Mellin second kind statistics

The first second kind characteristic function is obtained from Mellin's integral definition. From that the log-moments can be given.

Let x be a positive random variable with pdf, $p_x(u)$, for $u \in \Re^+$. According to Mellin transform definition, the first second characteristic function based on Mellin transform of p_x is defined by:

$$\phi_x(s) = E\left\{x^{s-1}\right\} = \mathcal{M}\left\{p_x(u)\right\}(s) = \int_0^{+\infty} u^{s-1} p_x(u) du, \quad \phi_x(s)|_{s=1} = 1 \tag{4. 12}$$

Besides, using multiplication property, Equation (4.6), exponential function expansion, and with the setting, $x^{(s-1)} = e^{(s-1)\log x}$, the relationship between the first second kind characteristic function, $\phi_x(s)$, and log-moments, $\tilde{m}_v$, are given. We have:

$$\phi_x(s) = \int_0^{+\infty} e^{(s-1)\log u} p_x(u)\, du$$

$$\phi_x(s) = E\left\{e^{(s-1)\log x}\right\}$$

$$= E\left\{1 + (s-1)\log x + (\log x)^2 (s-1)^2 / 2 + \cdots + (\log x)^v (s-1)^v / v! + \cdots\right\}$$

From log-moment definition:

$$\tilde{m}_v = E\left\{(\log x)^v\right\} = \int_0^{+\infty} (\log u)^v p_x(u) du, \quad where\ v \in \mathrm{N} \tag{4. 13}$$

We obtain:

$$\phi_x(s) = 1 + (s-1)\tilde{m}_1 + \tilde{m}_2 (s-1)^2 / 2 + \cdots + \tilde{m}_v (s-1)^v / v! + \cdots$$

$$= 1 + \sum_{v=1}^{+\infty} \frac{(s-1)^v \tilde{m}_v}{v!} \tag{4. 14}$$

In fact, Equation (4.13) shows that, the v times derivative of the first characteristic function, and evaluates at s = 1, defines log-moments of order v (Stephens, 1993):

$$\tilde{m}_v = \left.\frac{d^v \phi_x(s)}{ds^v}\right|_{s=1} \tag{4. 15}$$

Let $\tilde{\tilde{m}}$ be log-mean, it is defined by: $log(\tilde{\tilde{m}}) = \tilde{m}_1 \Leftrightarrow \tilde{\tilde{m}} = e^{\tilde{m}_1}$. Then, the vth-order of central log-moment is given by:

$$\tilde{M}_v = \int_0^{+\infty} (\log u - \tilde{m}_1)^v p_x(u) du \tag{4. 16}$$

According to log-moments definition (Equation 4.13), the relation between log-moments and central log-moments are given by:

$$\begin{aligned}\tilde{M}_2 &= \tilde{m}_2 - \tilde{m}_1^{\,2}\\ \tilde{M}_3 &= \tilde{m}_3 - 3\tilde{m}_2\tilde{m}_1 + 2\tilde{m}_1^{\,3}\\ \tilde{M}_4 &= \tilde{m}_4 - 4\tilde{m}_3\tilde{m}_1 + 6\tilde{m}_1^{\,2}\tilde{m}_2 - 3\tilde{m}_1^{\,4}\end{aligned} \tag{4. 17}$$

The second second kind characteristic function, $\Psi_x(s)$, is defined by the natural logarithm of the first second kind characteristic function, $\phi_x(s)$, which is introduced by the following definition:

$$\Psi_x(s) = \log(\phi_x(s)) \tag{4. 18}$$

Besides, using Equation 4.14 and logarithm function expansion, $\log u = -\sum_{n=1}^{+\infty}\frac{1}{n}(1-u)^n$, we have:

$$\begin{aligned}\Psi_x(s) = \log(E\left\{e^{(s-1)\log x}\right\}) &= -\sum_{n=1}^{+\infty}\frac{1}{n}(1-E\left\{e^{(s-1)\log x}\right\})^n\\ &= -\sum_{n=1}^{+\infty}\frac{1}{n}\left(-\sum_{v=1}^{+\infty}\tilde{m}_v\frac{(s-1)^v}{v!}\right)^n\\ &= \tilde{m}_1(s-1) + (\tilde{m}_2 - \tilde{m}_1^{\,2})\frac{(s-1)^2}{2!} + (\tilde{m}_3 - 3\tilde{m}_2\tilde{m}_1 + 2\tilde{m}_1^{\,3})\frac{(s-1)^3}{3!} + \cdots\end{aligned}$$

With the setting, $\tilde{k}_1 = \tilde{m}_1,\ \tilde{k}_2 = \tilde{m}_2 - \tilde{m}_1^2,\ \tilde{k}_3 = \tilde{m}_3 - 3\tilde{m}_1\tilde{m}_2 + 2\tilde{m}_1^3,$ and so on, we obtain:

$$\begin{aligned}\Psi_x(s) &= \tilde{k}_1(s-1) + \tilde{k}_2\frac{(s-1)^2}{2!} + \tilde{k}_3\frac{(s-1)^3}{3!} + \cdots\\ &= \sum_{v=1}^{+\infty}\frac{(s-1)^v}{v!}\tilde{k}_v\end{aligned} \tag{4. 19}$$

$\tilde{k}_v$ are log-cumulants (second kind cumulants).

Finally, the relationships between the second characteristic function, $\psi_x(s)$, and the log-cumulant of order v is given by (Abramowitz & Stegun, 1964; Stephens, 1993):

$$\tilde{k}_v = \frac{\mathrm{d}^v\Psi_x(s)}{\mathrm{d}s^v}\Big|_{s=1} = \frac{d^v\log(\phi_x(s))}{\mathrm{d}s^v}\Big|_{s=1} \tag{4. 20}$$

From Equation (4.13) and Equation (4.20), the first three log-cumulants are estimated by (Nicolas, 2002):

$$\hat{\tilde{k}}_1 = \frac{\sum_{i=1}^{N}\log x_i}{N}$$

$$\hat{\tilde{k}}_2 = \frac{\sum_{i=1}^{N}(\log x_i)^2}{N} - \left(\frac{\sum_{i=1}^{N}\log x_i}{N}\right)^2 \tag{4. 21}$$

$$\hat{\tilde{k}}_3 = \frac{\sum_{i=1}^{N}(\log x_i)^3}{N} - 3\frac{(\sum_{i=1}^{N}\log x_i)\sum_{i=1}^{N}(\log x_i)^2}{N^2} + 2\left(\frac{\sum_{i=1}^{N}\log x_i}{N}\right)^3$$

where N is the sample size and x_i an element of the sample. Moreover, according to Equation (4.10), the log-moment and log-cumulant of vth order of the inverse function are given by:

$$\begin{aligned} \tilde{m}_{inverse,v} &= (-1)^v \tilde{m}_v \\ \tilde{k}_{inverse,v} &= (-1)^v \tilde{k}_v \end{aligned} \tag{4. 22}$$

Simply, if we can calculate Mellin's integral for the function *f*, then, by differentiating the result with respect to s and evaluating at s = 1, and by differentiating from the logarithm of that result with respect to s and evaluating at s = 1, log-moment and log-cumulant will be obtained.

4.3. Probability distribution

As already mentioned, part of our work is to model different positive real descriptors (Table 2.1 and Table 2.4), as random variables with widely common pdfs. For this purpose, we review the possible pdfs and their second kind properties (Nicolas, 2006).

4.3.1. Generalities

Suppose that *u* is a positive random variable, where p_A , and ϕ_A are probability amplitude function and first characteristic probability amplitude function respectively. Setting $v = u^2$, the probability intensity function and first characteristic probability intensity function, p_I and ϕ_I respectively verify:

$$p_A(u) = 2u\, p_I(u^2), \quad \phi_A(s) = \phi_I\left(\frac{s+1}{2}\right) \tag{4. 23}$$

From Equations (4.23) and (4.20), the log-cumulants for uni-variate probability amplitude function can be deduced by:

$$\tilde{k}_{A,v} == \left(\frac{1}{2}\right)^v \tilde{k}_{I,v} \tag{4. 24}$$

Furthermore, the probability inverse function, p_Y , of the random variable Y=1/X, where the X-random variable follows the distribution function, p_X , is given by:

$$p_Y(u) = \frac{1}{u^2} p_X(\frac{1}{u}) \tag{4. 25}$$

From Equation (4.10):

$$\phi_Y(s) = \phi_X(2-s) \tag{4. 26}$$

From Equation (4.22):

$$\tilde{k}_{Y,v} = (-1)^v \tilde{k}_{X,v} \tag{4. 27}$$

Moreover, the second kind statistics for combination of two positive variables are summarized in Table (4.1).

Table 4. 1: Properties of the Mellin convolution, the inverse Mellin convolution, and the Mellin correlation of two distributions defined on $\Re^+$: p_B, p_C with first and second kind characteristic functions, ϕ_B, ϕ_C, and, v^{th}-order log-cumulants, $\tilde{k}_{B,y}, \tilde{k}_{C,y}$. $\hat{*}$, $\hat{*}^{-1}$ and $\hat{\otimes}$ indicate convolution, inverse convolution and correlation. Adapted from (Nicolas, 2006).

	First characteristic function	Second characteristic function	Log-cumulants
$p_B \hat{*} p_C$	$\phi_B(s)\phi_C(s)$	$\psi_B(s)+\psi_C(s)$	$\tilde{k}_{B,y}(s)+\tilde{k}_{C,y}(s)$
$p_B \hat{*}^{-1} p_C$	$\phi_B(s)\phi_C(s)$	$\psi_B(s)-\psi_C(s)$	$\tilde{k}_{B,y}(s)-\tilde{k}_{C,y}(s)$
$p_B \hat{\otimes} p_C$	$\phi_B(s)\phi_C(2-s)$	$\psi_B(s)+\psi_C(2-s)$	$\tilde{k}_{B,y}(s)+(-1)^y \tilde{k}_{C,y}(s)$

4.3.2. Gamma, Rayleigh and Nakagami distributions

Typically, a Gamma distribution with location *a>0*, and shape *b>0* parameters, is defined by (Walck, 2007):

$$f[a,b](u) = \frac{a(a\,u)^{b-1} e^{-au}}{\Gamma(b)}, \quad for\ u \geq 0 \tag{4. 28}$$

The Gamma distribution function which is adapted for SAR data intensity with two parameters μ and L, so that a= $\frac{L}{\mu}$, and *b*=*L*, is defined by:

$$G_I[\mu,L](u) = \frac{1}{\mu}\frac{L}{\Gamma(L)}\left(\frac{Lu}{\mu}\right)^{L-1} e^{-\frac{Lu}{\mu}}, \tag{4. 29}$$

μ and L are scale and shape parameters. If L =1, the Gamma pdf becomes an exponential distribution. The first characteristic function based on Mellin second kind statistics (Equation (4.1)) is:

$$\phi_{G,I}(s) = \mu^{(s-1)} \frac{\Gamma(L+s-1)}{L^{s-1}\Gamma(L)} \tag{4. 30}$$

To model SAR amplitude data, the Gamma distribution (Equation (4.29)) is transformed into the Nakagami distribution, $\mathcal{RN}\ [\mu, L]$, using Equation (4.23), which is defined by:

$$\mathcal{RN}[\mu', L](u) = \frac{2}{\mu'} \frac{\sqrt{L}}{\Gamma(L)} \left(\frac{\sqrt{L}u}{\mu'} \right)^{2L-1} e^{-\left(\frac{\sqrt{L}u}{\mu'}\right)^2} \text{,where} \quad \mu' = \sqrt{\mu} \tag{4. 31}$$

For L=1, the Rayleigh distribution is obtained:

$$\mathcal{RN}[\mu', L = 1](u) = \frac{2}{\mu'} \left(\frac{u}{\mu'} \right) e^{-\left(\frac{u}{\mu'}\right)^2} \tag{4. 32}$$

However, the Rayleigh distribution does not have enough flexibility and the ability to model data that has heavy-tailed distributions (Kuruoglu & Zerubia, 2004).

According to Equation (4.23) and Equation (4.30), the first characteristic amplitude distribution function is defined by:

$$\phi_{G,A}(s) = \mu'^{(s-1)} \frac{\Gamma(L + \frac{s-1}{2})}{L^{\frac{s-1}{2}} \Gamma(L)} \tag{4. 33}$$

According to Equations (4.20) and (4.33) and using definitions of digamma ($\psi^{(0)}(z) = \Gamma'(z) / \Gamma(z)$) and polygamma ($\psi^{(v)}(z) == \frac{d^v \psi^{(0)}}{dz^v} = \frac{d^{v+1}}{dz^{v+1}} ln\Gamma(z)$) (Bateman, 1953; Nicolas, 2006), the three first log-cumulants are obtained:

$$\begin{aligned} \tilde{k}_1 &= \log(\mu') + \frac{1}{2}\psi(L) - \frac{1}{2}\log(L), \\ \tilde{k}_2 &= \frac{1}{4}\psi(1, L), \\ \tilde{k}_3 &= \frac{1}{8}\psi(2, L) \end{aligned} \tag{4. 34}$$

These log-cumulants (Equation (4.34)) can also be obtained directly from Equation (4.24).

4.3.3. Inverse Gamma distribution

According to Equations (4.25) and (4.29), the Inverse Gamma intensity distribution with scale parameter, μ, and shape parameter, L, is defined by:

$$IG_I[\mu, L](u) = \frac{1}{\Gamma(L)} \frac{1}{L\mu} \left(\frac{L\mu}{u} \right)^{L+1} e^{-\frac{L\mu}{u}} \tag{4. 35}$$

For L=1, the Inverse Gamma pdf becomes the Inverse exponential pdf. For $L \geq 0$ and μ>0, the first characteristic function for the Inverse Gamma distribution in intensity from Equation (4.26) and Equation (4.30) is defined by:

$$\phi_{IG,I}(s) = \mu^{s-1} \frac{\Gamma(L+1-s)}{L^{1-s}\Gamma(L)} \tag{4. 36}$$

According to Equation (4.27) the log-cumulant for the Inverse Gamma pdf in intensity is defined by:

$$\begin{aligned} \tilde{k}_1 &= \log(\mu) - (\psi(L) - \log(L)), \\ \tilde{k}_2 &= \psi(1,L), \\ \tilde{k}_3 &= -\psi(2,L) \end{aligned} \tag{4. 37}$$

According to Equation (4.23), the Inverse Gamma distribution function in amplitude is defined by:

$$IG_A\left[\mu', L\right](u) = \frac{1}{\Gamma(L)} \frac{2}{\sqrt{L}\mu'} \left(\frac{\sqrt{L}\mu'}{u} \right)^{2L+1} e^{-\left(\frac{\sqrt{L}\mu'}{u}\right)^2} \tag{4. 38}$$

According to Equation (4.23), and Equation (4.36), the first characteristic function for Inverse Gamma distribution in amplitude is defined by:

$$\phi_{IG,A}(s) = \mu'^{\left(\frac{s-1}{2}\right)} \frac{\Gamma\left(L + \frac{1-s}{2}\right)}{L^{\frac{1-s}{2}}\Gamma(L)} \tag{4. 39}$$

According to Equations (4.23) and (4.39) and using the definitions of digamma and polygamma (Bateman, 1953; Nicolas, 2006), the three first log-cumulants are obtained:

$$\begin{aligned} \tilde{k}_1 &= \log(\mu') - \frac{1}{2}(\psi(L) - \log(L)), \\ \tilde{k}_2 &= \frac{1}{4}\psi(1,L), \\ \tilde{k}_3 &= -\frac{1}{8}\psi(2,L) \end{aligned} \tag{4. 40}$$

These log-cumulants (Equation (4.40)) can also be obtained directly from Equation (4.24).

4.3.4. Log-normal distribution

The Log-normal distribution is given by:

$$p(u) = \frac{1}{u\sqrt{2\pi v}} e^{\left(-\frac{(\log u - \beta)^2}{2v}\right)}, u > 0$$

where β and ν are the mean and variance of the variable, that is normally distributed, *logu*, respectively. In single look SAR data modeling context, a Log-normal distribution, $L\left[\mu,\sigma^2\right](u)$ with location of peak parameter β=μ>0 (mean) and shape parameter, $\sigma^2=\nu$, is defined by:

$$L\left[\mu,\sigma^2\right](u)=\frac{1}{u\sqrt{2\pi}}e^{\left(-\frac{(\log u-\mu)^2}{2\sigma^2}\right)},u>0 \tag{4. 41}$$

The Mellin second kind characteristic function for the Log- normal distribution is as follows:

$$\phi(s)=e^{\mu(s-1)}e^{\left(\sigma^2\frac{(s-1)^2}{2}\right)} \tag{4. 42}$$

The expectation value and the variance of the distribution are given by

$$E(u)=e^{\mu+\frac{\sigma^2}{2}},\qquad V(u)=e^{2\mu+\sigma^2}(e^{\sigma^2}-1) \tag{4. 43}$$

According to Equation (4.23), the characteristic function for amplitude from the intensity one is derived. Consequently, we obtain log-cumulants, using Equation (4.24). However, for order *v*>2, the log-cumulants are zero, so, this model does not have flexibility and the ability to model data that has heavy-tailed distributions. The results are summarized in Table 4.4.

4.3.5. Weibull distribution

Generally the Weibull distribution is given by: $p(u)=\frac{cu^{c-1}}{b^c}e^{-\left(\frac{u}{b}\right)^c}$

In the context of the modelling of SAR data, the Weibull pdf is defined by (Nicolas, 2002):

$$W\left[\mu,\eta\right](u)=\frac{|\eta|}{\mu}\left(\frac{u}{\mu}\right)^{\eta-1}e^{-\left(\frac{u}{\mu}\right)^{\eta}} \tag{4. 44}$$

where, b= μ and c= η are scale and shape parameters respectively. Thus, for η =1, the Weibull pdf becomes an exponential distribution, or a Gamma one ($\mathcal{G}[\mu,1]$). For η =2, the Weibull distribution becomes a Rayleigh distribution ($\mathcal{RN}[\mu,1]$). More details can be found in (Nicolas 2002; Nicolas, 2010).

According to the Mellin transform definition (Poularikas, 1999), the first second kind characteristic function for Weibull distribution is defined by:

$$\phi_w(u)=\mu^{(s-1)}\Gamma(L+\frac{s-1}{\eta}) \tag{4. 45}$$

From Equations (4.20), (4.45) and definitions of digamma and polygamma (Bateman, 1953; Nicolas, 2006), the three first orders and the νth log-cumulants are:

$$\begin{aligned}
\tilde{k}_1 &= \log(\mu) + \frac{\psi(1)}{\eta} \\
\tilde{k}_2 &= \frac{1}{\eta^2}\psi(1,1) \\
\tilde{k}_3 &= \frac{1}{\eta^3}\psi(2,1)
\end{aligned} \tag{4. 46}$$

The characteristic function of the Weibull distribution in amplitude is obtained from expression (4.23). Moreover, according to Equation (4.24), log-cumulants can be retrieved (Nicolas, 2002; Anfinsen, 2010). The results are listed in Table 4.4.

4.3.6. Beta distribution (first kind Beta distribution)

The Beta intensity distribution is defined by:

$$B[\mu, L, M](u) = \frac{\Gamma(M)}{\Gamma(L)\Gamma(M-L)} \frac{L}{M\mu}\left(\frac{Lu}{M\mu}\right)^{L-1}\left(1-\frac{Lu}{M\mu}\right)^{M-L-1} \tag{4. 47}$$

where $\mu>0$ is location parameter, $L>0$, and $M>0$, are shape parameters (see (Nicolas, 2006) for more details). The first characteristic function of the Beta distribution is obtained by inverse convolution of the Gamma distribution $G_I[\mu, L]$ with the Gamma normalized distribution $G_I[1, M]$ (Table 4.2):

$$\phi_{B,I}(s) = \mu^{s-1}\frac{\Gamma(L+s-1)}{L^{s-1}\Gamma(L)}\frac{M^{s-1}\Gamma(M)}{\Gamma(M+s-1)}, \quad M > L \tag{4. 48}$$

According to Equation (4.20), the log-cumulants for Beta intensity pdf can be retrieved:

$$\begin{aligned}
\tilde{k}_1 &= \log(\mu) + \psi(L) - \log(L) - (\psi(M) - \log(M)) \\
\tilde{k}_2 &= (\psi(1,L) - \psi(1,M)) \\
\tilde{k}_3 &= (\psi(2,L) - \psi(2,M))
\end{aligned} \tag{4. 49}$$

From Equation (4.23), the Beta distribution in amplitude and the first characteristic function in amplitude are given. From Equations (4.24) and (4.49), the log-cumulants can directly be retrieved. The results are listed in Table 4.4.

4.3.7. Fisher distribution (second kind Beta distribution)

The fisher pdf in intensity is defined by:

$$F[\mu,L,M](u)=\frac{\Gamma(L+M)}{\Gamma(L)\Gamma(M)}\frac{L}{M\mu}\frac{(\frac{Lu}{M\mu})^{L-1}}{\left(1+\frac{Lu}{M\mu}\right)^{L+M}} \tag{4. 50}$$

where μ>0 is the location parameter and L>0 and M>0 are shape parameters. The first characteristic functions of Fisher intensity distribution is obtained by convolution of the Gamma distribution, G_I[μ, L], with the Inverse Gamma normalized distribution IG[1, M] (Table 4.2). Thus, the first characteristic function is written as follows:

$$\phi_{F,I}(s)=\mu^{s-1}\frac{\Gamma(L+s-1)}{L^{s-1}\Gamma(L)}\frac{\Gamma(M+1-s)}{M^{1-s}\Gamma(M)} \tag{4. 51}$$

According to Mellin transform properties which are listed in Table 4.1, the log-cumulants can be retrieved (Nicolas, 2002; Anfinsen, 2010; Nicolas, 2006), or according to Equation (4.20), the log-cumulants for Fisher intensity distribution can be directly calculated:

$$\begin{aligned}\tilde{k}_1&=\log(\mu)+\psi(L)-\log(L)-(\psi(M)-\log(M))\\ \tilde{k}_2&=\psi(1,L)+\psi(1,M)\\ \tilde{k}_3&=\psi(2,L)-\psi(2,M)\end{aligned} \tag{4. 52}$$

From Equation (4.23), the Fisher distribution in amplitude and the first characteristic function in amplitude are given. From Equations (4.24) and (4.52), the log-cumulants can directly be retrieved. The results are listed in Table 4.4.

4.3.8. $\mathcal{K}$ distribution

The $\mathcal{K}$ pdf in intensity is defined by:

$$\mathcal{K}[\mu,L,M](u)=\frac{1}{\Gamma(L)\Gamma(M)}\frac{2LM}{\mu}\left(\frac{LMu}{\mu}\right)^{\frac{M+L}{2}-1}K_{M-L}\left(2\left(\frac{LMu}{\mu}\right)^{\frac{1}{2}}\right) \tag{4. 53}$$

where, the μ, L, M parameters are scale and shape. K is second kind modified Bessel function. The first characteristic function is obtained from the convolution of a Gamma distribution, G[μ, L] with a Gamma normalized distribution, G[1, M] (Table 4.2). Thus, the first characteristic function is written as follows:

$$\phi_{\mathcal{K},I}(s)=\mu^{s-1}\frac{\Gamma(L+s-1)}{L^{s-1}\Gamma(L)}\frac{\Gamma(M+s-1)}{M^{s-1}\Gamma(M)} \tag{4. 54}$$

According to Mellin transform properties which are listed in Table 4.1, the log-cumulants can be retrieved (Nicolas, 2002; Anfinsen, 2010; Nicolas, 2006), or according to Equation (4.20), the log-cumulants for $\mathcal{K}$ intensity distribution can be obtained:

$$\tilde{k}_1 = \log(\mu) + \psi(L) - \log(L) + \psi(M) - \log(M)$$
$$\tilde{k}_2 = \psi(1,L) + \psi(1,M) \qquad (4.\ 55)$$
$$\tilde{k}_3 = \psi(2,L) + \psi(2,M)$$

From Equation (4.23) the $\mathcal{K}$ amplitude distribution in amplitude and the first characteristic function in amplitude are given. From Equations (4.24) and (4.55), the log-cumulants can directly be retrieved. The results are listed in Table 4.4.

Table 4. 2: Second kind characteristic functions of Beta and Fisher pdf obtained by inverse Mellin convolution and Mellin correlation of the G_I [1, M]) and G_I [μ, L] gamma distributions

$\hat{*}^{-1}$ ↗	G_I [1, M]	$\hat{\oplus}$ ↗	G_I [1, M]
G_I [μ, L]	Beta distribution, $\mu^{(s-1)} \frac{\Gamma(L+s-1)}{L^{s-1}\Gamma(L)} \frac{M^{s-1}\Gamma(M)}{\Gamma(M+s-1)}$	G_I [μ, L]	Fisher distribution, $\mu^{(s-1)} \frac{\Gamma(L+s-1)}{L^{s-1}\Gamma(L)} \frac{\Gamma(M+1-s)}{M^{1-s}\Gamma(M)}$

Table 4. 3: Second kind characteristic functions of $\mathcal{K}$ and Fisher pdf obtained by Mellin convolution of the *IG* [1, M] Inverse gamma distribution, G_I [μ, L] and G_I [1, M] gamma distributions

$\hat{*}$ ↗	G_I [1, M]	*IG* [1, M]
G_I [μ, L]	$\mathcal{K}$ distribution, $\mu^{(s-1)} \frac{\Gamma(L+s-1)}{L^{s-1}\Gamma(L)} \frac{\Gamma(M+s-1)}{M^{s-1}\Gamma(M)}$	Fisher distribution, $\mu^{(s-1)} \frac{\Gamma(L+s-1)}{L^{s-1}\Gamma(L)} \frac{\Gamma(M+1-s)}{M^{1-s}\Gamma(M)}$

Table 4. 4: Mellin statistics of uni-variate distribution (first second characteristic function and log-cumulants) for pdf in amplitude. Γ(.) is the Gamma function, K(.) is the second kind modified Bessel function, ψ(.) is the Digamma function, and ψ(*v*,.) is the vth-order Polygamma function. Adapted from (Nicolas, 2010; Nicolas, 2006).

	Pdf	**First Characteristic function**	**Log-Cumulants** $\tilde{k}_V, v = 1,2,3$
Log-Normal $\mathcal{L}_A[\mu,\sigma](\mathrm{u})$	$\frac{2}{\sigma\sqrt{2\pi}u} e^{\left(-2\frac{(\log u-\frac{\mu}{2})^2}{\sigma^2}\right)}$	$e^{\mu(\frac{s-1}{2})} e^{\left(\sigma^2(\frac{s-1}{4})^2\right)}$	$\tilde{k}_1 = \mu/2, \tilde{k}_2 = \sigma^2/4,$ $\tilde{k}_3 = 0$
Gamma $G_A[\mu', L](\mathrm{u})$	$\frac{2}{\mu'}\frac{\sqrt{L}}{\Gamma(L)}\left(\frac{\sqrt{L}u}{\mu'}\right)^{2L-1} e^{-\left(\frac{\sqrt{L}u}{\mu'}\right)^2},$ $\mu' = \sqrt{\mu}$	$\mu'^{(s-1)}\frac{\Gamma(L+\frac{s-1}{2})}{L^{\frac{s-1}{2}}\Gamma(L)}$	$\tilde{k}_1 = \log(\mu') + \frac{1}{2}\psi(L) - \frac{1}{2}\log(L),$ $\tilde{k}_2 = \frac{1}{4}\psi(1,L), \quad \tilde{k}_3 = \frac{1}{8}\psi(2,L)$
Weilbull $\mathcal{W}_A[\mu', \eta](\mathrm{u})$	$\frac{\lvert 2\eta \rvert}{\mu'}\left(\frac{u}{\mu'}\right)^{2\eta-1} e^{-\left(\frac{u}{\mu'}\right)^{2\eta}}$	$\mu'^{(s-1)}\Gamma(1+\frac{s-1}{2\eta})$	$\tilde{k}_1 = \log(\mu') + \frac{1}{2\eta}\psi(1), \quad \tilde{k}_2 = \frac{1}{4\eta^2}\psi(1,1)$ $\tilde{k}_3 = \frac{1}{8\eta^3}\psi(2,1)$
Beta $\mathcal{B}_A[\mu', L, M](\mathrm{u})$	$\frac{\Gamma(M)}{\Gamma(L)\Gamma(M-L)}\sqrt{\frac{L}{M}}\frac{2}{\mu'}$ $\left(\sqrt{\frac{L}{M}}\frac{u}{\mu'}\right)^{2L-1}\left(1-(\sqrt{\frac{L}{M}}\frac{u}{\mu'})^2\right)^{M-L-1}$	$\mu'^{(s-1)}\frac{\Gamma(L+\frac{s-1}{2})}{L^{\frac{s-1}{2}}\Gamma(L)}\frac{M^{\frac{s-1}{2}}\Gamma(M)}{\Gamma(M+\frac{s-1}{2})}$	$\tilde{k}_1 = \log(\mu') + \frac{1}{2}\psi(L) - \frac{1}{2}\log(L) - \frac{1}{2}(\psi(M) - \log(M)),$ $\tilde{k}_2 = \frac{1}{4}(\psi(1,L) - \psi(1,M)), \quad \tilde{k}_3 = \frac{1}{8}(\psi(2,L) - \psi(2,M))$
$\mathcal{K}$ $\mathcal{K}_A[\mu', L, M](\mathrm{u})$	$\frac{1}{\Gamma(L)\Gamma(M)}\frac{4\sqrt{LM}}{\mu'}\left(\frac{\sqrt{LM}u}{\mu'}\right)^{M+L-1}$ $K_{M-L}\left(\frac{2\sqrt{LM}u}{\mu'}\right)$	$\mu'^{(s-1)}\frac{\Gamma(L+\frac{s-1}{2})}{L^{\frac{s-1}{2}}\Gamma(L)}\frac{\Gamma(M+\frac{s-1}{2})}{M^{\frac{s-1}{2}}\Gamma(M)}$	$\tilde{k}_1 = \log(\mu') + \frac{1}{2}\psi(L) - \frac{1}{2}\log(L) + \frac{1}{2}(\psi(M) - \log(M)),$ $\tilde{k}_2 = \frac{1}{4}(\psi(1,L) + \psi(1,M)), \quad \tilde{k}_3 = \frac{1}{8}(\psi(2,L) + \psi(2,M))$
Fisher $\mathcal{F}_A[\mu', L\, M](\mathrm{u})$	$\frac{\Gamma(L+M)}{\Gamma(L)\Gamma(M)}\sqrt{\frac{L}{M}}\frac{2}{\mu'}\frac{(\sqrt{\frac{L}{M}}\frac{u}{\mu'})^{2L-1}}{\left(1+(\sqrt{\frac{L}{M}}\frac{u}{\mu'})^2\right)^{L+M}}$	$\mu'^{(s-1)}\frac{\Gamma(L+\frac{s-1}{2})}{L^{\frac{s-1}{2}}\Gamma(L)}\frac{\Gamma(M+\frac{1-s}{2})}{M^{\frac{1-s}{2}}\Gamma(M)}$	$\tilde{k}_1 = \log(\mu') + \frac{1}{2}\psi(L) - \frac{1}{2}\log(L) - \frac{1}{2}(\psi(M) - \log(M)),$ $\tilde{k}_2 = \frac{1}{4}\psi(1,L) + \psi(1,M)), \quad \tilde{k}_3 = \frac{1}{8}(\psi(2,L) - \psi(2,M))$

4.4. Estimation methods

The main part of modeling is to estimate the parameters which can be adjusted by minimizing the distance between model and data. The parameters of well-known radar distributions mostly can be estimated through the Maximum Likelihood Estimation (MLE), Method of Moments (MoM), and Method of Log-Cumulants (MoLC). We will not address the MoM in this study.

4.4.1. MLE

One of popular and traditional method of estimation for a statistical modeling is the Maximum Likelihood Estimation. Let us begin with a sample $\mathrm{X} = (\mathrm{x}_1, \ldots\ldots, \mathrm{x}_n)$ of independent and identically distributed random variables chosen according to one of a family of probabilities $p_\theta(x)$ which is described by c parameters, $\theta_1, \ldots\ldots, \theta_c$. The probability of observing given data as a function of θ is defined by:

$$p_\theta(x_1, \ldots, x_n) = p(x_1, \ldots, x_n | \theta) = \prod_{k=1}^{N} p(x_k \mid \theta) \tag{4. 56}$$

This expression represents the likelihood. By taking the logarithm, we obtain an additive expression, the log-likelihood:

$$L(\theta) = \log(p(x_1, \ldots, x_n \mid \theta) = \sum_{k=1}^{N} \log(p(x_k \mid \theta)) \tag{4. 57}$$

The maximum likelihood is described by:

$$\frac{\partial \log(p(x_1, \ldots, x_n \mid \theta)}{\partial \theta_j} = \sum_{k=1}^{N} \frac{\partial \log(p(x_k \mid \theta))}{\partial \theta_j} = 0, \quad \forall_j \in [1, c] \tag{4. 58}$$

In other words, for estimating parameters, the differentiate of the log-likelihood function should be calculated with respect to each parameter and set the results (Equation 4.58) to zero, then, the system of equations should be solved to find extrema. Consequently, an essential part consists in applying a numerical optimization algorithm. The search algorithm should satisfy maximization of the maximum likelihood function (analytical expressions of the derivatives, (Equation 4.58)). There are different search algorithms, for example Newton and quasi-Newton are used in (Antoniou & Lu, 2007), expectation and maximization (EM) in (Todd, 1996), least-squares (L-S) in (Marquardt, 1963), gradient method, or Nelder-Mead Simplex direct search algorithm in (Nelder & Mead, 1965; Lagarias, et al., 2012; Swann, 1969). However, each of these methods have limitations, some of them are given in (Baudin, 2010).

We use the Nelder-Mead Simplex algorithm (NM) through the Statistics Toolbox in MATLAB to solve this optimization problem. The NM algorithm is a direct search method for multi-dimensional nonlinear

unconstrained minimization function. The NM minimizes the object function without using any derivatived functions (Equation (4.58)) (see (Kelley, 1999) for more details).

4.4.2. MoLC

As for the widely well-known method of moments, the method of the cumulant consists in calculating the log-cumulants (Equation 4.21) of 1 to v^{th} order from training sample data (Nicolas, 2006). Then v equations with c unknown parameters for the proposed distribution are obtained. It is possible, if all v log-moments exist, according to existence theorem of log-cumulants. An essential part of parameter estimation is to apply the proper numerical optimization algorithm which can satisfy the equation system. We used the numerical algorithm, that have been suggested in (Nicolas, 2006), and the optimization method based on direct search Simplex Neder-Mead algorithm (Nelder & Mead, 1965; Lagarias et al., 2012).

Moreover, to introduce initial parameter values, the golden-section direct search has been used. In fact, the Golden-section direct search let us find the initial values very fast. Otherwise, it may take a long time or even lead to divergence case.

4.5. Goodness-of-Fit

The Goodness-of-Fit (GoF) measures the difference between the empirical and fitted theoretical probability distributions (Baglivo et al., 1992; Stephens, 1974). In a GoF, a null hypothesis is tested:

$$H_0 : F(x) = F_0(x) \ \forall x, vs.\ H_1 : F(x) \neq F_0(x) \tag{4. 59}$$

where $F_0(x)$ is a known distribution function (empirical pdf) and $F(x)$ the theoretical one. The main idea is to measure the deviation from the hypothesis. In this work, to assess whether a given distribution is suited to empirical data, we consider the following measurements: estimation of Correlation Coefficient (Filliben, 1975; Vogel, 1986), Kolmogorov–Smirnov distance (Stephens, 1974; Conover, 1972), and Root Mean Square Error (Stephens, 1974).

4.5.1. Correlation Coefficient estimation

The correlation coefficient (cc) measures the linear relationship between empirical data and distributed variables. The cc is estimated by:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y^2) - (\sum y)^2}} \tag{4. 60}$$

where n is the number of pairs of data, x and y are empirical data and distribution variables respectively. The correlation coefficient takes values ranging from -1 to 1. A value close to 1 shows the positive linear relationship, values close to -1 shows a negative linear relationship, and 0 suggest there is no

linear relationship between data. A perfect correlation of ± 1 occurs only when the data points all lie exactly on a straight line. If r = +1, the slope of this line is positive. If r = -1, the slope of this line is negative. More information may be found in (Filliben, 1975).

4.5.2. Kolmogorov – Smirnov distance

This measurement compares an empirical distribution function (EDF) with fitted theoretical distribution, and finding the maximum vertical distance between them (Stephens, 1974). The K-S measurement is given by:

$$D = \sup_{x} \left[F(x) - F_{\circ}(x) \right] \tag{4. 61}$$

where the fitted cumulative distribution function and empirical distribution are $F(x)$ and $F_{\circ}(x)$ respectively. More information may be found in (Slakter, 1965; Stephens, 1974; Conover, 1972). However, it is difficult to draw any robust fitness conclusions between empirical data and statistical distribution model through K-S test due to its sensitivity to the size of sample and it turns out to be less convenient on small samples according to our experiments. This is the reason why we do not perform the test here. Furthermore, it calculates only the maximal difference between sample data and statistical model.

4.5.3. Root Mean Square Error (RMSE)

RMSE may be used to assess the overall fit of a model. The root mean square deviation presents the measurement of the differences between values which are predicted by a model or an estimator and the values actually observed. These individual differences are also called residuals. The RMSE also presents the square root of the variance, and is given by:

$$RMSE = \sqrt{\frac{\sum (x - y)^2}{n}} \tag{4. 62}$$

Where x is the observed value and y is the predicted value from the fitted model and n is the number of points. More information is given in (Browne & Cudeck, 1992).

4.6. Statistical analysis purpose and discussion

The aim of the proposed statistical analysis is to investigate the modeling of different descriptors such as amplitude ratio, and some coherency amplitude components (Table 2.1, Table 2.4), using Mellin second kind statistics.

Moreover, we plan to introduce the statistical analysis information results in an adapted supervised classification scheme based on maximum likelihood. Therefore, the precision of the statistical properties

are essential. Besides, this analysis allows us to check the possibility of discriminating each type of land cover according to its statistical behavior.

There are many reasons for using the Mellin transforms: logarithmic statistics are easier to handle compared to classical methods (Fourier transform and method of moment). The log-cumulants can be easily expressed. The Mellin kind statistics also lead to simple expressions for moment and log-moments and log-cumulants, so they are easy to exploit (Epstein, 1948; Nicolas, 2010). Mellin kind statistics are commonly used in both single and multi-look radar image data under the product model (Tison, et al., 2003; Krylov, et al., 2011; Tupin, 2011). As some of Meijer's functions (Bateman, 1953) on $\Re^+$ well match with multiplicative noise in coherent imagery and can be defined as inverse Mellin transform, it is possible to deal with functions in the Mellin statistics framework (Nicolas, 2011) and easily use them to model heterogeneous media. Moreover, the complicated distributions and also new ones are deduced by using some of basic properties of the Mellin transform (Nicolas, 2010). Besides, the Mellin kind statistics turn out to lead to solutions and is fast (Bateman, 1953; Nicolas, 2011).

We select MoLC estimation method for estimating the parameters of distributions defined on $\Re^+$ due to its feasibility (Mellin expressions of log-cumulants do exist; Table 4.1) and the existence of simple form for higher order moments (Table 4.4). It has been applied to a large number of parametric distribution function families employed for SAR amplitude and intensity statistical modeling (Tison, et al., 2003; Nicolas & Tupin, 2002; Krylov, et al., 2011).

Moreover, as Meijer's functions generalize almost all the probability density functions which are used in coherent imagery (Nicolas, 2011), using the MoLC estimation method for estimating their parameters seems possible.

In comparison with MLE, the MLE variance decreases as the sample size increases but MLE is often not a good estimator for small samples. Moreover, as the rate of convergence is depending on the number of variables to optimize, the rate of convergence in MLE is slower than in MoLC method. Indeed, for calculation of Fisher distribution parameters through MoLC approach, first two shape parameters (L, M) are estimated using the Simplex Neder-Mead method (Han & Neumann, 2006) and then by using Equation (4.54), the μ-scale parameter is directly calculated. Therefore, it is faster than the MLE method that estimates three parameters (μ, L, and M) simultaneously. Moreover, MLE needs more specialized mathematical solution for solving complex non-linear equations such as for $\mathcal{K}$ distribution (Bessel function), so that finding optimization function seems difficult.

In the litterature, several different theoretical and empirical pdfs were selected for modelling noise or texture in SAR data context. For example, the generalized Rayleigh pdf in amplitude which turns out to be a symmetric α-stable distribution has been proposed for SAR image. It has demonstrated its flexibility in urban SAR image noise modelling when compared to Rayleigh, Weibull, and $\mathcal{K}$-distributions (Kuruoglu & Zerubia, 2004).

Furthermore, the generalized Gaussian Rayleigh (GGR) distribution has been proposed on ERS-1, E-SAR, X-SAR, and NASA/JPL airborne SAR data respect to MoLC estimation and has provided more

flexibility in SAR image land cover modelling than the symmetric α-stable distribution, Nakagami-distribution, $\mathcal{K}$-distribution, G° (square root of the generalized inverse Gaussian distribution) (Frery, et al., 1997), and generalized Gamma (Moser, et al., 2006a).

Moreover, several theoretical pdf models have been estimated in order to adapt a finite mixture model on SAR image to descriminate different land-cover classes (Moser, et al., 2006b). The heterogeneous pdf model adopting by a finite mixture model has been recently proposed on very high resolution SAR imagery by (Krylov et al., 2011). In this work, we focus on the Fisher pdf.

The texture may have an extended tail under certain conditions (heterogeneous texture areas), such that the Weibull and Log-normal models cannot describe and model the data. Thus these distributions are not appropriate. Furthermore, a complex model with many parameters better fits data than a simple model with few parameters (Myung, 2002).

We want to consider the analysis of several statistical models which are adapted for SAR image, over different PoLSAR positive real descriptors. We expect Fisher distribution to be the more appropriate as it has already been validated for high resolution SAR statistics over urban regions in (Tison et al., 2003; Nicolas, 2002; Krylov et al., 2009; Tupin, 2011; Tison et al., 2004; Bombrun & Beaulieu, 2008). The result of this statistical analysis is used to introduce a new classification algorithm based on Fisher-ML classifier.

4.7. Summary III

This chapter begins with a short overview of the Mellin transformation expression, its fundamental properties, and the Mellin second kind statistics (the first and the second characteristic functions and log-cumulants). A number of well-known distribution functions and their second kind statistics are presented. Moreover, two estimation parameters methods, MoLC and MLE, are discussed. Three measurements of GoF (RMSE, K-S and correlation coefficient) are described. Moreover, we note some important points in a discussion section.

Chapter 5:
Classification

5. Classification methods

The image classification is the task in which a set of given data elements (pixels or objects) is assigned to some classes such that the cost for assigning a pixel or object to a related class is minimal. Here we deal with a pixel-based approach. Typically there are two approaches for the classification, whether using training samples or not: supervised or unsupervised respectively. Besides, whether the classifier assumes that the data fit to a specific distribution or not, it is divided into parametric or nonparametric classifiers respectively.

This chapter starts by presenting some well-known algorithms for supervised classification of PoLSAR data and our new supervised classification algorithm based on the Fisher pdf. Subsequently, in Chapter 8, we perform a comprehensive evaluation of those classifications in urban areas using a single PoLSAR X-band image.

5.1. ML classification based on uni-variate Gaussian pdf

Maximum Likelihood Classification (MLC) is one of the most common classification algorithms in remote sensing. The uni-variate complex Gaussian distribution is used in Maximum Likelihood Classification by (Goodman, 1963). It assumes that the statistics for each class in each band is normally distributed and calculates the probability that a given pixel belongs to a specific class (Canty, 2006). Let us note x, the attribute vector of a pixel, x is a vector with N observations Thus, the conditional probability to observe x from class r, or probability density function based on Bayes Theory (Fitelson, 2001) is described as follows:

$$p(r \mid x) = \frac{p(x \mid r)\mathrm{P}(r)}{p(x)} \tag{5. 1}$$

where the $\mathrm{P}(r)$, r=1, 2,…, N_C, are prior probabilities and N_C is the class number, $p(x \mid r)$ is a class-specific probability density function, and where $p(x)$ is given by

$$p(x) = \sum_{i=1}^{\mathrm{N}_C} p(x \mid i)\mathrm{P}(i) \tag{5. 2}$$

Since $p(x)$ is independent from r, we can write the decision rule as x is in class r:

$$p(x \mid r)\mathrm{P}(r) \geq p(x \mid j)\mathrm{P}(j) \ \textit{for all} \ j=1,\ldots,N_C, \text{ and } r \neq j \tag{5. 3}$$

Suppose that the observations from class r, are sampled from a multi-variate normal distribution. Then the density function is given by:

$$p(x \mid r) = \frac{1}{(2\pi)^{N/2} \left|\Sigma_r\right|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_r)^T \Sigma_r^{-1} (x - \mu_r) \right) \tag{5. 4}$$

where μ_r is the mean vector of class *r*, Σ_r is variance-covariance matrix of class *r* and $\left|\Sigma_r\right|$ is determinant of Σ_r, which may be estimated from training data and N the variable number. The logarithm form of Equation (5.4) gives:

$$\log(p(x \mid r)) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_r| - \frac{1}{2}(x - \mu_r)^T \Sigma_r^{-1} (x - \mu_r) \tag{5. 5}$$

The first term may be ignored, as it is independent from *r*. Together with Equation (5. 3) and Equation (5. 5) a discrimination function is defined:

$$d_r(x) = \log(\mathrm{P}(r)) - \frac{1}{2}\log|\Sigma_r| - \frac{1}{2}(x - \mu_r)^T \Sigma_r^{-1} (x - \mu_r) \tag{5. 6}$$

We obtain the minimum likelihood classifier; x is in class *r* if: $d_r(x) \geq d_j(x) \quad \forall r \neq j$

As often there is no information about prior class probabilities $\mathrm{P}(r)$, in which case they can be set equal and ignored in the classification. Thus, the discrimination function of Equation (5.6) becomes:

$$d_r(x) = -\log|\Sigma_r| - (x - \mu_r)^T \Sigma_r^{-1} (x - \mu_r) \tag{5. 7}$$

More details can be found in (Morrison, 1976; Canty, 2006)

5.2. Supervised ML based on multi-variate Gaussian pdf

After multi-looking, the multi-variate polarimetric SAR covariance or coherency matrix is generally modeled by a complex Wishart distribution (Lee & Pottier, 2009). In supervised maximum likelihood classification based on the coherency matrix $[T]_{3\times3}$, the averaged observation vector of class *r* is replaced by $[\Sigma_r]$ (expectation of coherency matrix of class *r*) which is used in (Lee et al., 2001; Lee et al., 1994). Thus, in the case of reciprocal backscattering, this coherency matrix $[T]_{3\times3}$ from an area of many random canonical scatters follows a complex Wishart pdf (Srivastava, 1965), $W(0,[\Sigma_r])$, given by:

$$P_w([T]_{3\times3}) = \frac{n^{q\,n} \left|[\Sigma_r]\right|^{n-q} \exp\left(-n\ tr([\Sigma_r]^{-1}[T]_{3\times3})\right)}{K(n,q)\left|[\Sigma_r]\right|^n}, \text{ with } K(n,q) = \pi^{q(q-1)/2} \prod_{j=1}^{q} \Gamma(n-j+1) \tag{5. 8}$$

An inherent limitation (Muirhead, 1982; Tracy, 2006; Anderson, 2003; Giri, 2004) appears in the scaling function $K(n,q)$ that verifies ($K(n,q) \to \infty$ if $n \leq q-1$, $\Rightarrow P([T]_{3\times3}) \to 0$). *n* is number of look and *q* the dimension of the polarimetric target vector k, with *q*=3 in the reciprocal case $(\mathrm{S_{HV}} = \mathrm{S_{VH}})$, $\Gamma(.)$

represents the Gamma function, *tr* denotes matrix trace, $[T]_{3\times3} = \frac{1}{q}\sum_{i=1}^{q} k_i.k_i^{\dagger}$ is the coherency matrix, and $k = \frac{1}{\sqrt{2}}[S_{HH}+S_{VV} \quad S_{HH}-S_{VV} \quad 2S_{HV}]^{T}$ is the complex target vector (Equation (2.9)).

However, according to Equation (5.8), for single-look SAR data (n=1) and a mono-static SAR configuration (q = 3), n < q − 1, which limits the use of the classification based on Wishart pdf. Therefore, the single look multi-variate PoLSAR data coherency matrix is modeled by the zero-mean multi-variate complex Gaussian distribution (Goodman, 1963):

$$P_G([T]_{3\times3}) = \frac{\exp\left(-k^{\dagger}[\Sigma_r]^{-1}k)\right)}{\pi^q \|[\Sigma_r]\|} \tag{5. 9}$$

where k is the complex target vector or target coherency vector (Pauli vector). The log-likelihood function is given by:

$$log(P_G([T]_{3\times3})) = -\log(\|[\Sigma_r]\|) - tr([\Sigma_r]^{-1}[T]_{3\times3}) + q(1-q)\log\left|[T]_{3\times3}\right| \tag{5. 10}$$

After removing the last terms that do not depend on the class *r*, the distance used in the decision rule becomes:

$$d([T]_{3\times3} \mid [\Sigma_r])) = log(\|[\Sigma_r]\|) + tr([\Sigma_r]^{-1}[T]_{3\times3}) \tag{5. 11}$$

During the classification process, a pixel is then assigned according to the minimum distance criterion. This classification is noted by G-ML^{C}.

5.3. Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a one of powerful kernel-based classification method based on statistical learning theory (Cortes & Vapnik, 1995). In the following, a brief description of the SVM is given. More details, some widely applications and extensions of support vector machines can be found in (Burges, 1998). The SVM is widely used for PoLSAR image classification. In addition, few studies have also been conducted with SAR data (Mercier & Girard-Ardhuin, 2005). Fuzzy support vector machine (FSVM), which is a variante of the SVM algorithm, has been used to classify PoLSAR image data (Zhou, et al., 2013). Besides, composite kernels in SVM over multisource information, both spectral and spatial, have been presented in (Tuia, et al., 2010).

Moreover, the integrated SVM classification with Markov random field models for taking into account the spatial context was proposed using multichannel SAR and multispectral high-resolution images in (Moser & Serpico, 2010).

The basic approach of SVM is first explained in a linear case and originally designed for binary classification (Schölkopf, et al., 1998), and then the extension to nonlinear classifier through the use of nonlinear kernels is described. Generally, the task of SVM classification is to find a rule (a decision

surface), based on external observations that maximizes the margin between the classes (Müller, et al., 2001). In fact, input vectors are mapped to a very high-dimension feature space. The feature surface that separates two samples is often called the optimal surface, and the data points closest to the optimal surface are called support vectors (Figure 5.1).

Let define a training set of instance-label pairs by: $\{(x_i, y_i),\ i = 1,...,l\}$, where $x_i \in R^n$ is a *n*-dimension training vector and $y_j \in \{-1,1\}$. The aim of SVM classification is to assign x_j to one of the two classes, $y_j \in \{-1,1\}$, respect to the optimal surface ($H: w^T . x + b = 0$), which is defined by:

$$f(x_j) : Sign\{w^T . x_j + b\} \tag{5. 12}$$

where, w is the normal vector to the optimal surface and b is the offset (Figure 5.1).

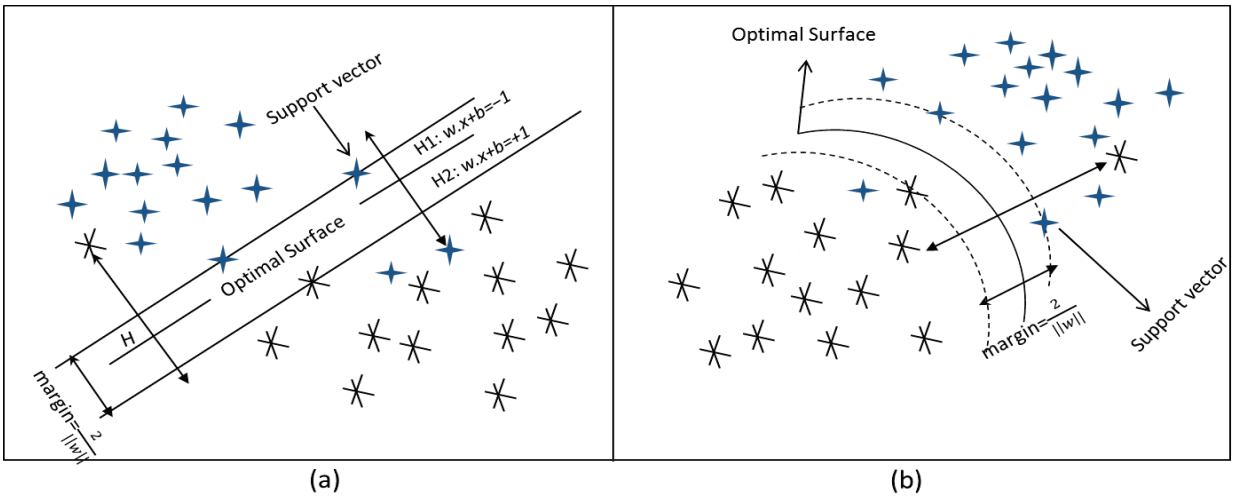


Figure 5. 1: SVM optimal surface: (a) linear optimal surface and (b) nonlinear optimal surface, adapted from (Cortes & Vapnik, 1995)

Thus, the SVM should find a solution for optimization problem. According to the Lagrangian formalism, the optimal decision function can be obtained by:

$$f(x_j) : Sign\{\sum_{i=1}^{n} y_i \alpha_i^* x_j^T . x_i + b^*\} \tag{5. 13}$$

where α_i^* are the estimated Lagrange multipliers, and b^* is estimated using the Karush Kuhn Tucker condition (Kuhn & Tucker, 1951). Indeed, the goal of SVM is to compute the optimal surface or equivalently, to maximize the margin by finding min $\left\{\frac{1}{2}\|w\|^2\right\}$ under the inequality condition, $y_i(w^T \phi(x_i) + b) - 1 \geq 0$ (Figure 5.1).

In fact, the SVM classifier maps the training vectors x_i into hyper dimension space by the function ϕ. In the nonlinear case, to find the optimal surface, SVM uses some kernel methods to map the data with a non-linear transformation to a feature space of higher dimension where the data separation may become linear. Thus, the classification function *f* is:

$$f(x_j) : Sign\{\sum_{i=1}^{r} y_i \alpha_i^* K(x_j, x_i) + b^*\} \tag{5. 14}$$

In general, three common kernel functions are used:

- Polynomial: $K(x_j, x_i) = (\gamma x_j^T . x_i + \rho)^r$, with $\rho > 0$
- Sigmoid: $K(x_j, x_i) = \tanh(x_j^T . x_i + \rho)$
- Radial basis function: $K(x_j, x_i) = \exp(-\frac{\|x_j - x_i\|^2}{2\sigma^2})$

where $K(x_j, x_i) = \phi(x_j)^T \phi(x_j)$ is known as the kernel function, ρ, γ, and r, are kernel parameters. The SVM approach may be expanded for multi-classes. Two main methods exist: one against all or one against one strategy (Hsu & Lin, 2002).

5.4. ML classifier based on Fisher pdf

In this work, we propose a new Maximum Likelihood Classification, based on Fisher pdf. Indeed, we show in Chapter 7 that our attributes have a Fisher distribution.

With the assumption that the distribution of a class sample is a Fisher one, every class can be characterized by Fisher parameters ($\mu_{k,b}, L_{k,b}, M_{k,b}$), and $b \in \{1, 2, ..., N\}$ where N is space dimensions (number of descriptors) and k the class index. Given these three characteristics for each class, the statistical probability is computed for each class over the N-multidimensional space, to determine the membership of the pixels to one class. Prior class probabilities can be set optional. Each object is classified to the class to which it has the highest probability of being a member.

Suppose that the observations from class k are sampled from a uni-variate Fisher distribution, the membership of the pixel (i, j) to the class k is:

$$P(k, i, j) = \sum_{b=1}^{N} \log(p\left[\mu_{k,b}, L_{k,b}, M_{k,b}\right](u_b(i, j))) \tag{5. 15}$$

where $p\left[...\right](u_b(i, j))$ is the Fisher pdf for pixel u of the bth feature, whose value is $u_b(i, j)$.

According to the decision rule (Equation (5.3)), the discriminant function becomes:

$$C(i, j) = \underset{k \in \{1, ..., N_c\}}{\arg\max}\left[P(k, i, j)\right] \tag{5. 16}$$

where N_c is the number of classes and $C(i,j)$ is the decided class of the pixel (i,j). Then, the algorithm follows the classical steps of ML-classification but using the Fisher pdf (Figure 5.2).

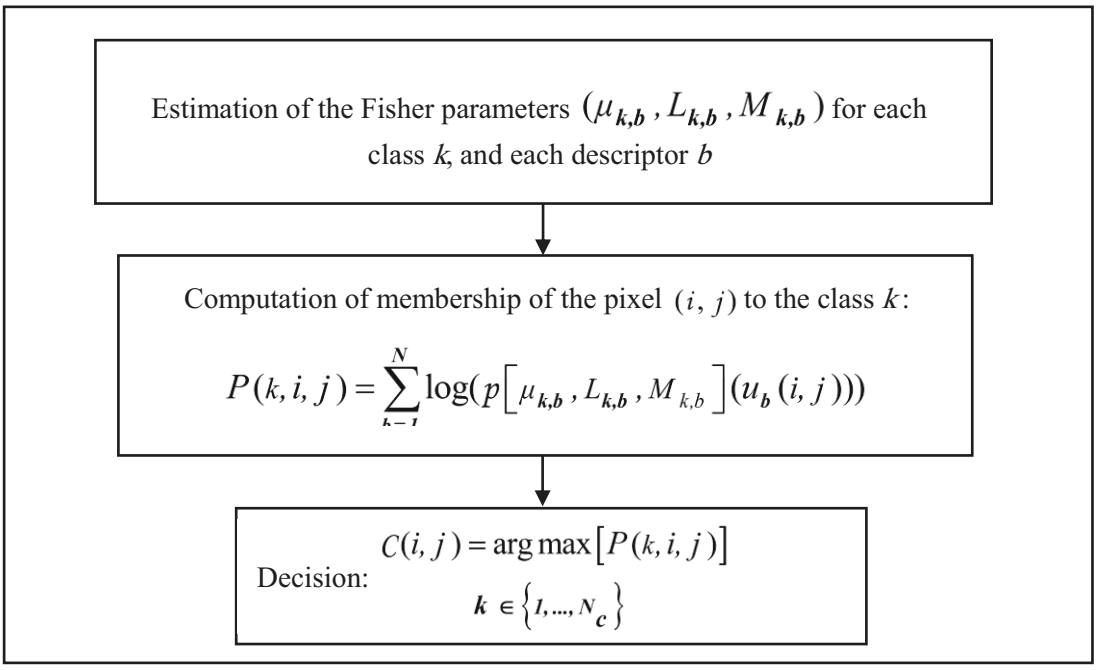


Figure 5. 2: The proposed ML classification based on Fisher pdf

5.5. ML-Fisher classification coupled with morphological descriptor

All these classification frameworks, considered in previous sections, are concerned by the drawbacks due to complex scattering mechanisms in the urban environment. For instance, after first experiences, we notice that building not aligned with the azimuthal direction appear as volume scattering, many rough surfaces occur as volume scattering and some vegetated areas appear as surface scattering. To improve the classification results and reduce the ambiguities, we propose an enriched Maximum Likelihood classification based on Fisher pdf (ML-F), using morphological tools (Chapter 2, section 2.4.2).

The morphological information is defined by:

$$f(i,j) = \begin{cases} 0, & \forall i,j \in C_{\text{background}} \\ 1, & \forall i,j \in C_{\text{object}} \end{cases}$$

Let, C_1 and C_2 be sets of pixels in two classes form, $C \in \{1,...,C_k\}$, where, $k \in \{1,...,N_c\}$, N_C is the number of classes. Suppose that ambiguities are observed in the classification result between these two classes. Suppose that $C_1 \subset C_{object}$ and we want that $C_2 \cap C_{object} = \varnothing$. The membership of each pixel (i,j) for each class, k, is noted $P_k(i,j)$. To introduce constraint to Maximum Likelihood Classification procedure, we have:

Step1: calculate constrained value:

Constraint= $\min\left\{P_{C_1}(i,j), \quad \forall i,j \in C_2\right\}$

Step2: $P_{C_2}(i,j) = \text{Constraint}, \quad \forall i,j \in \underset{object}{C}$

In fact, the minimum value of $P_{C_1}(i,j)$ is assigned to pixel membership $P_{C_2}(i,j)$ when related to object morphology information. The enriched supervised classification algorithm is summarized in Figure 5.3.

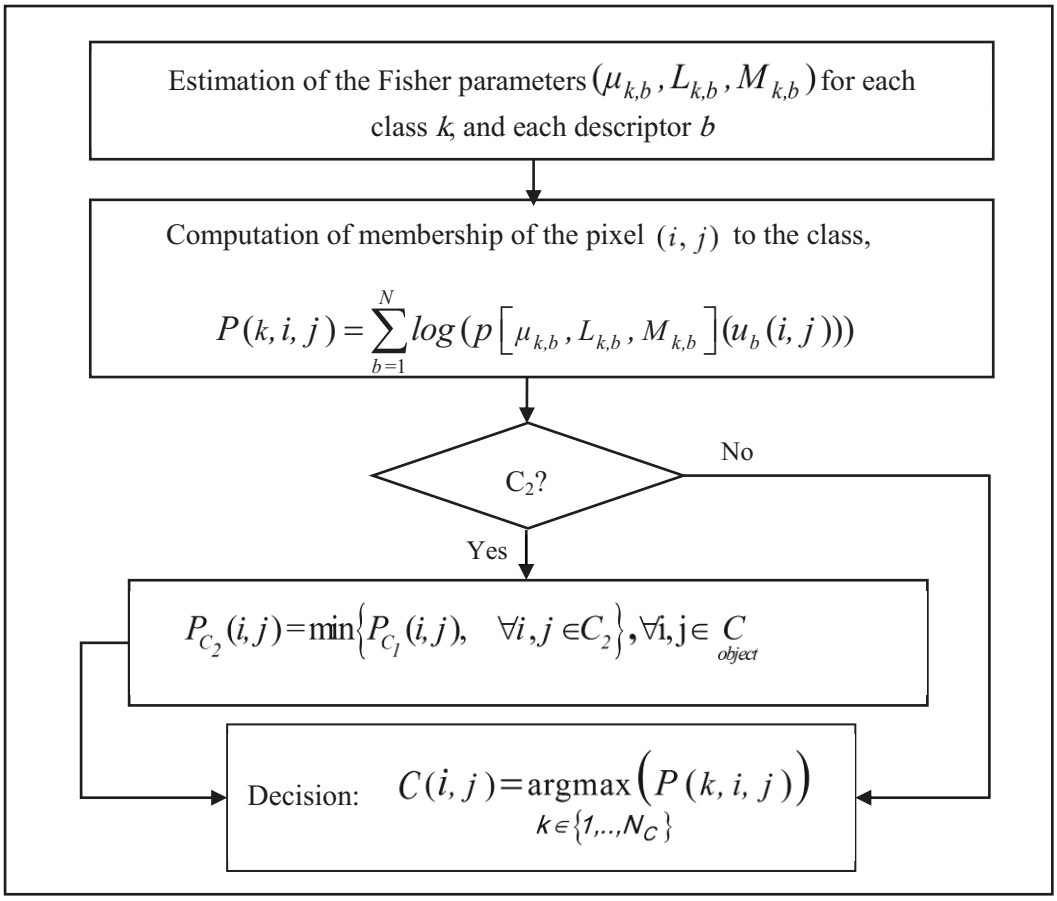


Figure 5. 3: Synopsis of the enriched ML classification based on Fisher-pdf and morphological tools

5.6. Summary VI

This chapter reviewed some of the common algorithms for classification of single look PoLSAR image. As the reference data have already been collected from the study area, the proposed methods are supervised. First, we have reminded the ML-classification algorithm based on Gaussian model in uni-variate case using vectorial descriptors and the multi-variate case using the coherency matrix descriptors. In this chapter, we explain the ML-classification based on Fisher pdf that might provide new considerations. Using the morphological tools for extracting building feature, in the previous Fisher ML classification algorithm might provide better results. The ML-classification with Fisher pdf assumes here that descriptors are independent. However, this may not be the case and to overcome this, we propose the selection (see Chapter 3.1) of the features to keep non-redundant ones. Moreover, in this

work study, support vector machine, which is a nonparametric method, is used as reference classification algorithm.

Part III:

Experimental results

Chapter 6:
Study area and data description

6. Study area and data description

6.1. Study area

6.1.1. Study area location

The study area falls between 43°33' N to 43°32' N latitudes and 1°28' E to 1°29' E longitudes near the south-east of Toulouse. Figure 6.1 and Figure 6.2 show the study area on aerial photo and PoLSAR image (the polarization color composite image is based on the Pauli vector by assigning $|S_{HH}- S_{VV}|/\sqrt{2}$, $|S_{HH}+ S_{VV}|/\sqrt{2}$, and $\sqrt{2}|S_{HV}|$ as red, blue and green of a color image) respectively.

Figure 6. 1: The study area on aerial-photo (BD Ortho IGN)

6.1.2. Study area description

The study area is characterized by a mixture of urban attribute types and natural objects. Typically, from the observation, urban attributes can be classified into two categories: man-made and natural attributes. The natural one is characterized by four sub-categories, trees, lawns, water bodies and the bare-soil. Meanwhile, two major types of roads and roofs are assigned to man-made category (Nasarudin & Shafri, 2011). According to this observation the roof sub-category consists in seven types of roofs (Figure 6.3).

However, the common roof styles (geometry shape) used for homes and small buildings, are gable, hip, flat and shed. Meanwhile, for some special building architectures use gambrel, mansard or Dutch roofs (Barnow, 1986). Moreover, there are different roof materials, such as rubber, fiberglass, copper,

aluminum, composite, shingle slate, Bamboo, asphalt, metal, slate, wood, wood-shake, cement, tile, and asphalt roofing shingles (Aarti, 2011).

6.2. SAR data description

6.2.1. SAR data

The corresponding radar image has been obtained from RAMSES (Radar Aéroporté Multi-spectral d'Etude des Signatures) system, the ONERA airborne multispectral SAR sensor (Cantalloube, et al., 2006). It works over a wide range of frequency bands L (1.3GHZ), S (3.2GHZ), C (4.3GHZ), X (9.5GHZ), Ku (14.3GHZ) , Ka (35 GHZ), and W (95GHZ) (Boutry, 1993). The incidence angles can be set from 30 to 85 degrees.

The data was generated in stripmap mode SAR which takes advantage of illuminating the ground swath with a continuous sequence of pulses while the antenna beam is pointed to a fixed angle in elevation and azimuth (Krieger, et al., 2010). The data was delivered as Single-Look Complex (SLC), with a resolution of around 50 cm and was acquired in 2006 with an off-nadir viewing angle of 60° and a pixel size of 35 cm in both azimuth and range directions in X-band fully polarimetric mode (Figure 6.2). Some of general image information are summarized in Table 6.1. Because of the low depression angle, roofs are well represented but not the facades. Besides, the resolution gives access to many urban objects and texture on roofs (Simonetto & Malak, 2009). An image subset of 1000×1000 pixels was selected for testing the developed classification algorithm. We used Hamming filter (Nuttall, 1981) with α =0.54 and β= 0.46 parameters to minimize the maximum peak side-lobe level in (SLC) images.

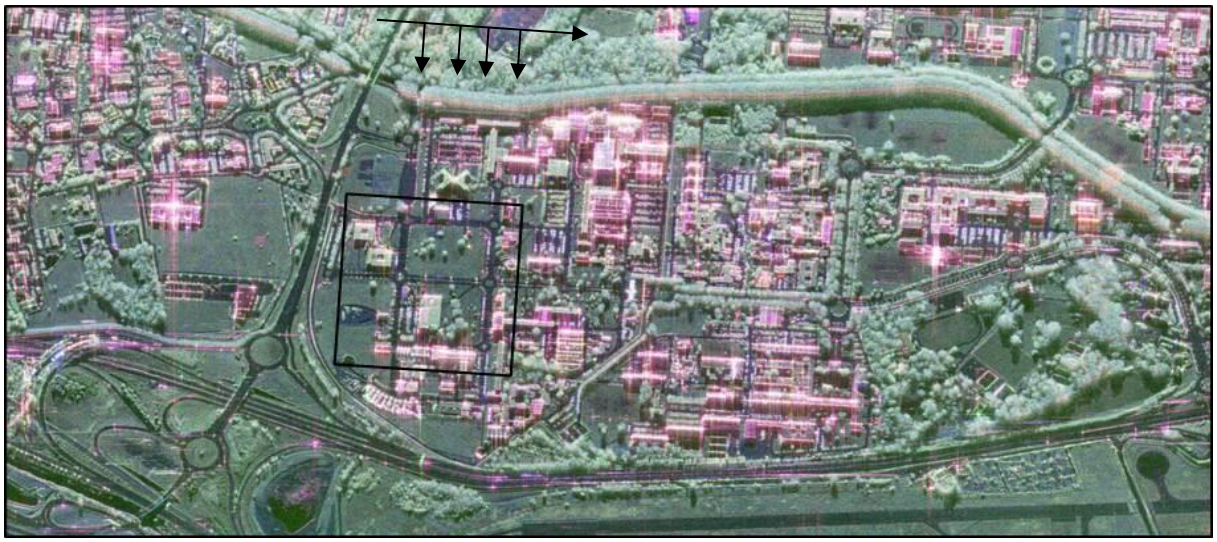


Figure 6. 2: The study area on PoLSAR image (Pauli RGB), the trajectory and viewing directions are displayed with black arrows.

Table 6. 1: Image information over covered urban area

Image	SAR
Platform	Airborne
Sensor	RAMSES
Spectral band	X (9.5GHz)
Data Format	Single Look Complex (SLC)
Image Mode	Full Polarimetric
Radar pixel size (m)	0.35
Acquisition date	March, 2006
Angle(°)	60°

6.3. Definition of the classes

6.3.1. Discussion

From the aerial photo (Figure 6.1) observations, three roof types are distinguished according to their shapes, such as sloped (red color), flat (different gray colors), and materials such as tile, concrete, gravel, and slate with different gray, red, and blue colors, or metallic with white color. The tile roofs mainly represent sloped roofs (Figure 6.1). From the PoLSAR image (Figure 6.2), we distinguish two types of roofs, according to shape and materials. Concrete, gravel and slate roofs represent flat roofs (Figure 6.3). According to our classification experiences with and without considering the metallic roofs as one class, this attribute cannot be recognized as an individual class roofs.The classification results show that mostly metllaic-pixels are assigened to *flat roof* class.

In addition, in SAR image, the bright pixel effects exist due to specular phenomena because these surfaces are smooth in comparison to the radar wavelength. They appear from direct and multiple bounces, for instance between the building walls and the ground or between some objects which are located on a roof and with the roof itself. For example, the smooth surfaces can be brighter than rough surfaces at small incidence angles. Another effect that can be seen in PoLSAR image is the shadow effect from objects not visible in the resulting radar image. Thus, according to visual SAR interpretation associated with aerial photo observations, there are two types of roofs, two types of natural attributes, one type of roads and two special types due to SAR effects. Figure 6.3 shows the result of observation from urban area, aerial photo and PoLSAR image.

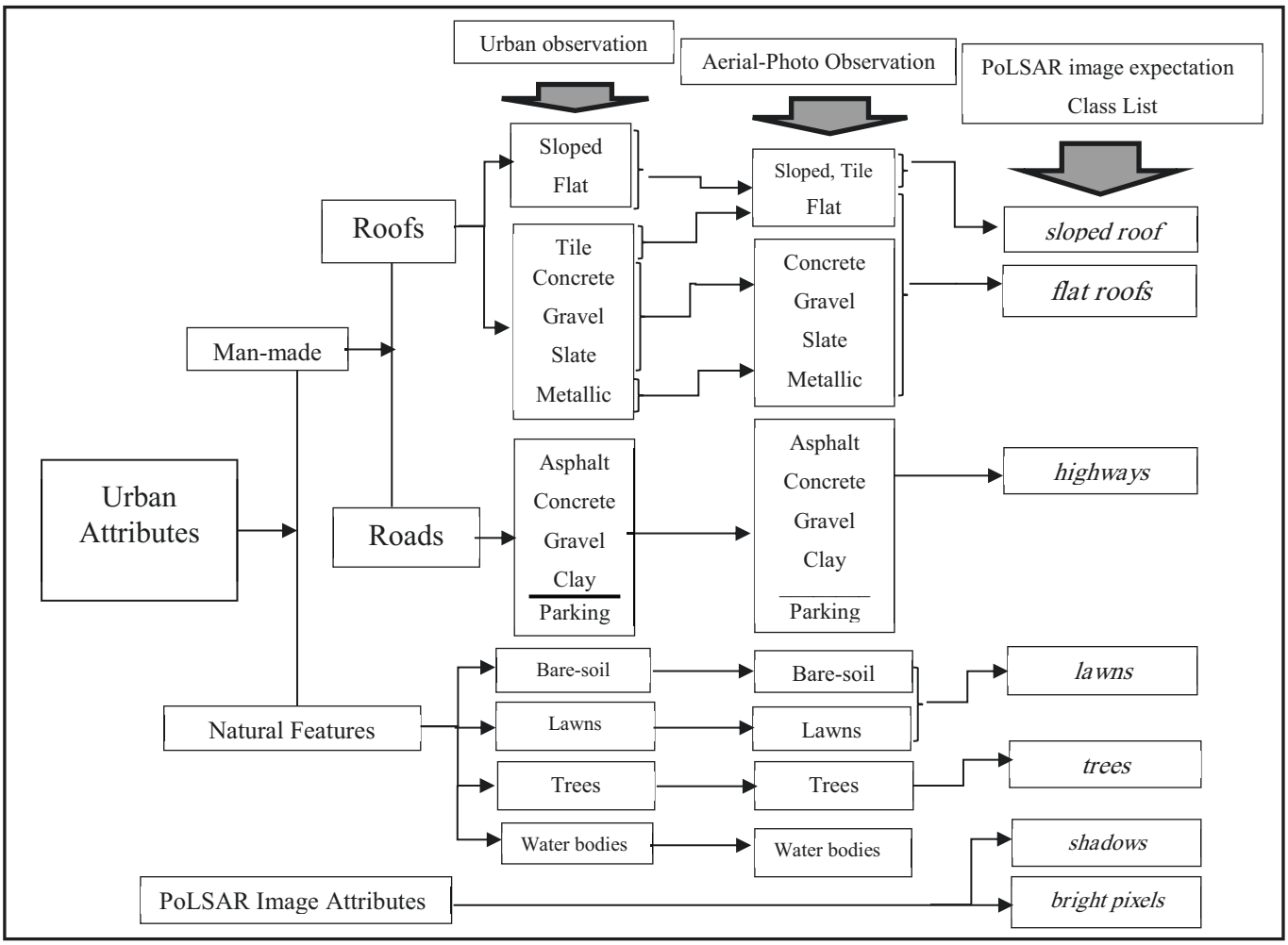


Figure 6. 3: Urban classification based on different observations levels.

6.3.2. Ground truth data

We define two sets of knowledge: training samples as input in the supervised classification processes and control samples to assess the performance. Training and control data have been extracted manually by visual interpretation of an aerial view from the French Geographical Institute (IGN) acquired in 2006 (Figure 6.1) and the radar image (Figure. 6.2). Our main objective is to estimate different urban classes; in particular we are interested in discriminating two types of building roofs: *flat roofs*, and *sloped roofs*. The *flat roofs* class consists in different types of roofs (gravel, concrete and slate roofs), but not metallic roofs. Indeed, we do not separate different kinds of *flat roofs* based on roof material, because it is not possible to distinguish them by image interpretation using both optical and radar views and so it is not possible to build distinct reliable training samples for supervised classification. In fact, some roofs with the same appearance in the radar image can appear with different colours in the optical view and inversely. In the optical view, they appear in different colours (red, brown and different gray levels) and in the radar data they also show a different appearance. One could work with more roof classes if a ground truth were available to build the training samples. The *sloped roofs* are more likely to be covered by tiles, which are confirmed by their red color observed in the optical image. The other classes are labeled *lawns*, including lawn and bare soil, *trees* including trees and shrubs, *highways* including roads

and car parks, and also two classes, (radar) *shadows* and *bright pixels*. Table 6.2 shows the set of classes with their corresponding number of training and test samples. These two last classes do not correspond to any kind of urban objects but have turned out to be useful to limit misclassification. Figure 6.4 (c and d), shows the training samples over the study area.

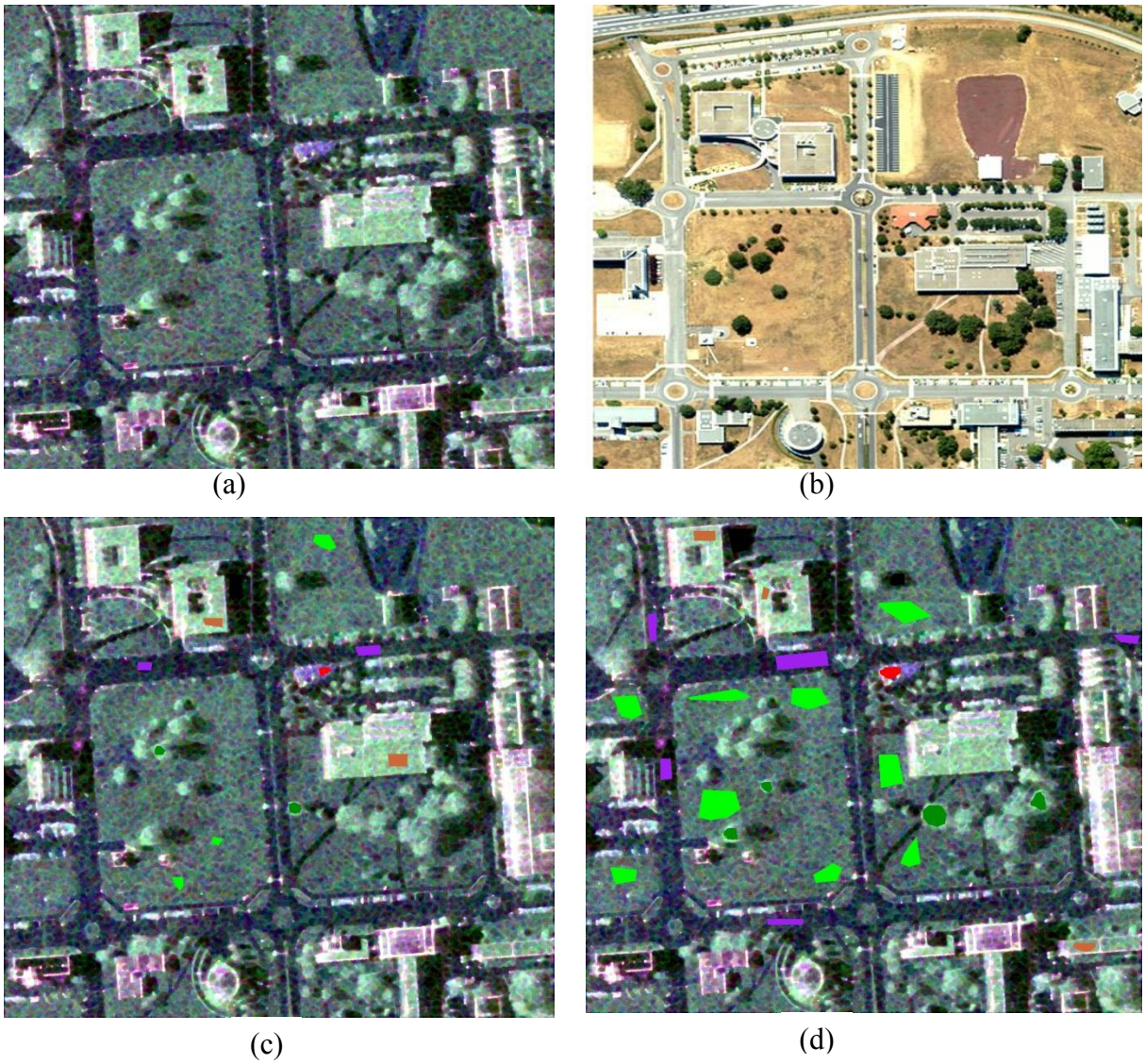


Figure 6. 4: (a) RAMSES image of the studied site (Toulouse, France), acquired in 2006, in Pauli colour-coded representation, (b) Aerial view from the French Geographical Institute (IGN) acquired in 2006, (c) Training samples superimposed on the RAMSES view (same colour legend as in Figure 8. 6), (d) Control samples superimposed on the view (same colour legend as in Figure 8. 6).

Table 6. 2: Training and control samples (numbers of pixels).

Classes	Training data	Control data
(Radar) *shadows*	1313	2275
highways (roads, parks…)	1216	5714
lawns (lawns, bare soils)	1518	17929
Trees (trees, small shrubs)	674	2963
bright pixels	838	829
flat roofs	1388	1643
sloped roofs	269	665

Chapter 7:

Result on feature extraction and selection

7. Result on feature extraction and selection

7.1. Feature extraction and feature selection results

This chapter focuses on two experimental results, over real positive descriptors and complete descriptors. After extracting and gathering all SAR descriptors from single look complex full PoLSAR X-band image which were described in Chapter 2, the performance of feature subset selection method (SSM) using GA based on MI is applied on descriptors.

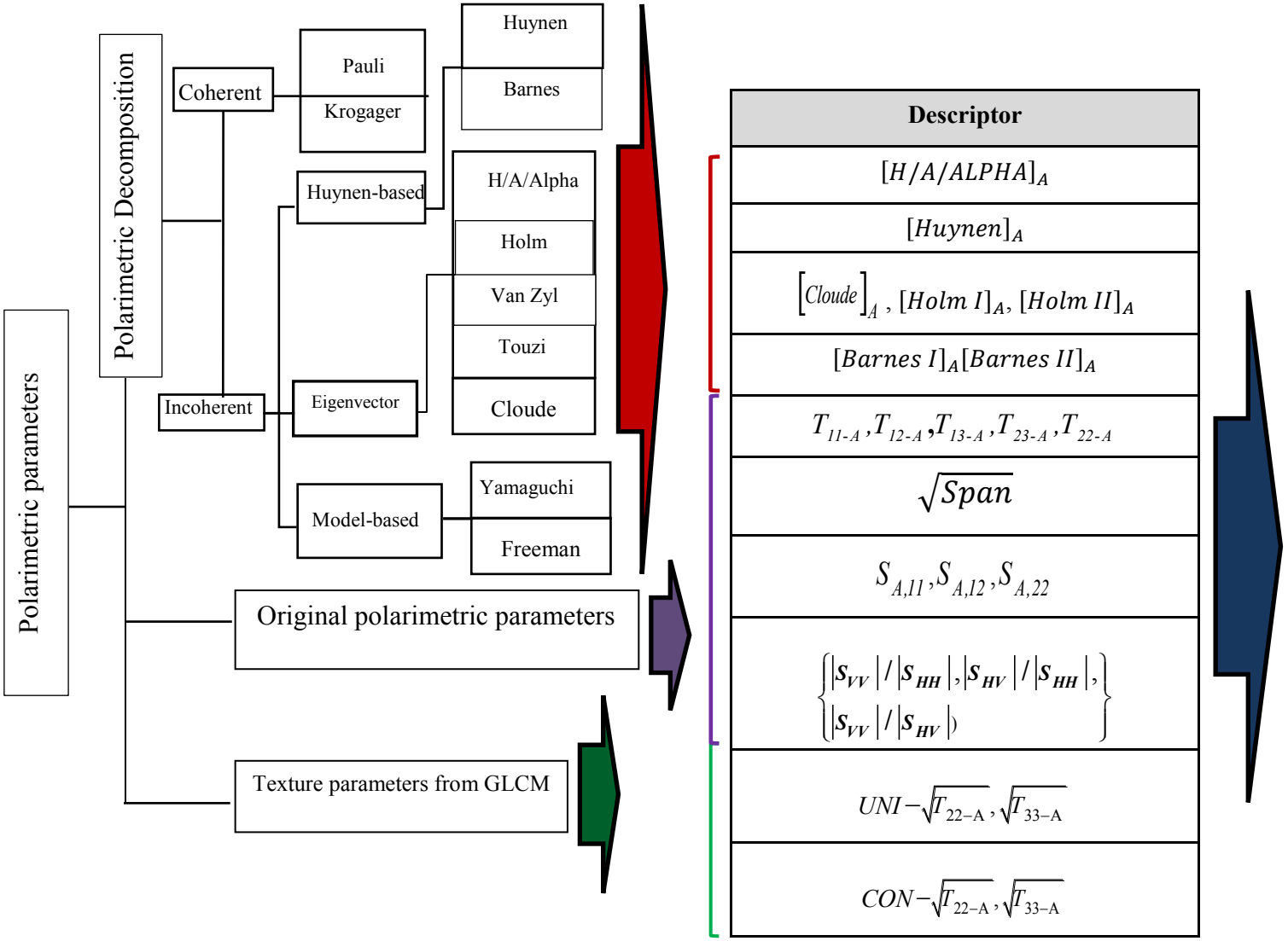
7.1.1. Feature extraction

The steps of feature extraction include:

- Calculate the texture descriptors from PoLSAR images based on GLCM, with horizontal direction, d=1, and θ=0°. It derives from that nine features (Figure 7.1).
- Obtain the features from PoLSAR data:
    - Compute the 58 previously described features and to settle the form of feature vector (Figure 7.1).
    - Compute the 127 descriptors preciously described features and to settle the form of feature vectors (Figure 7.2) (see comment in Chapter 2, subsection 2.2.4.4).

Several descriptors are shown in Figure 7.3. All computations were done through the PoLSARpro v4.0 (Pottier, et al., 2005) and ENVI 4.8/IDL (RSI, 2003; RSI-IDL, 2005). In this study a simple averaging filter with a processing window size of 5 is used. In fact the coherent descriptors are extracted from unfiltered Sinclair scattering matrix, this causes the coherent technique maintain the full resolution. Moreover, meaningful high-frequency information refers to detailed features, such as edges and point targets can be preserved. Incoherent parameters are extracted from filtered coherency matrix, it causes the incoherent parameters to have less speckle noise.

The normalization technique is applied on all decomposition components, so that incoherent parameters are independent from total power (*Span*).



Descriptor
$\lvert S_{HV}\rvert / \lvert S_{HH}\rvert$
$\sqrt{Span}$
$CON-\sqrt{T_{22-A}}$
$CON-\sqrt{T_{33-A}}$
$\lvert S_{VV}\rvert / \lvert S_{VH}\rvert$
Total: 5/58

Table 7. 1: positive real descriptors selected through GA (Table2.1).

Figure 7. 1: Real Positive feature extraction for statistical analysis by different methods (Table2.1).

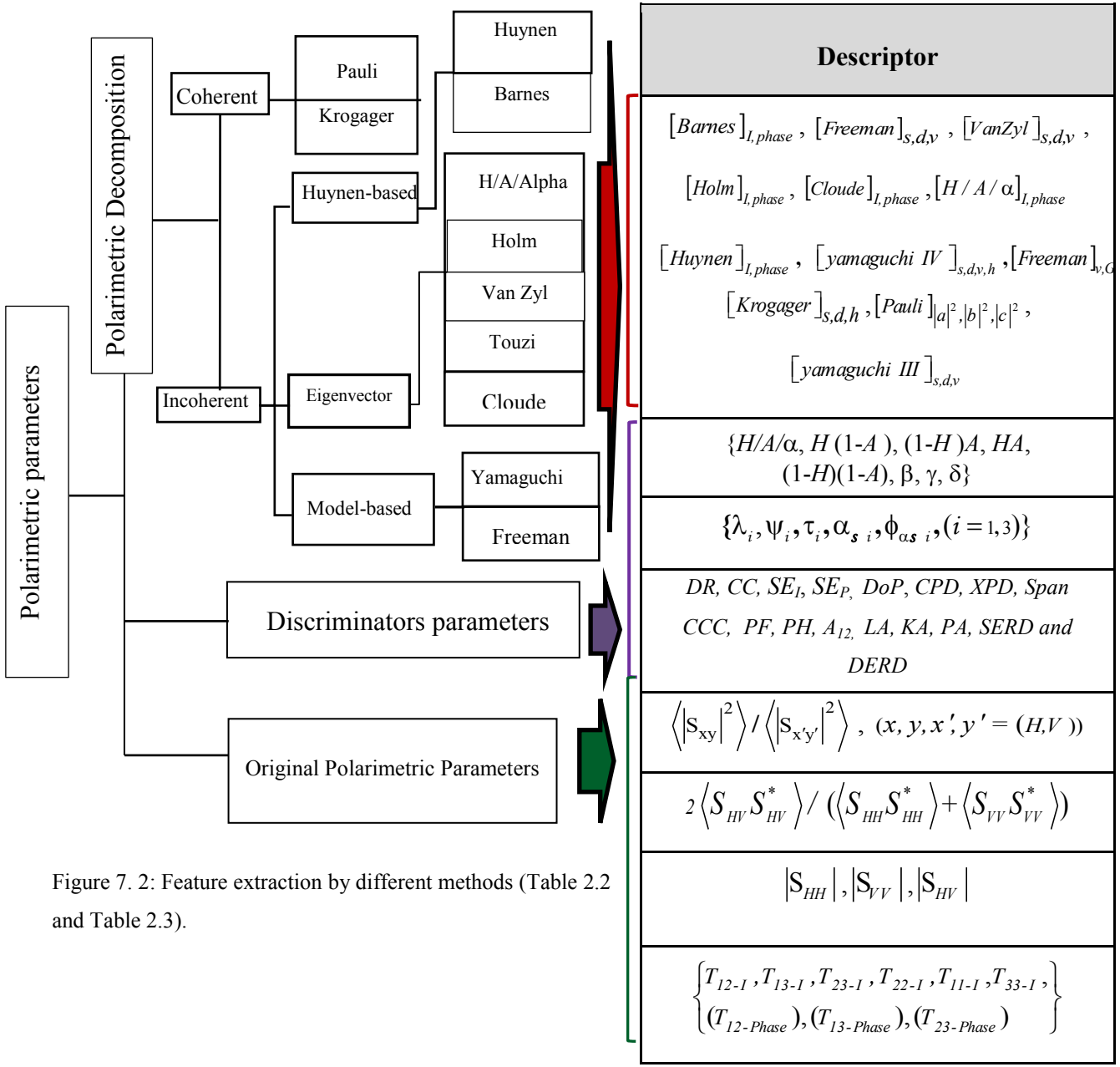


Figure 7. 2: Feature extraction by different methods (Table 2.2 and Table 2.3).

#	Descriptor
3	*Barnes I-$T_{12\text{-}phase}$, Polarization-Fraction (PF), **Span***
4	*Shanon-Entropy-I (SE_I), Anisotropy-Lueneburg (AL), $T_{33\text{-}I}$, Cloude_$T_{12\text{-}phase}$*
5	*Shanon-Entropy-I (SE_I), H (1-A), $T_{33\text{-}I}$, $T_{12\text{-}phase}$, H/A/Alpha-$T_{11\text{-}I}$*
6	*Span, $T_{33\text{-}I}$, Anisotropy-12, $T_{11\text{-}I}$, Holm I-$T_{23\text{-}phase}$, Barnes I-$T_{12\text{-}phase}$*
7	*Span, Anisotropy-Lueneburg (AL), $T_{33\text{-}I}$, $T_{23\text{-}phase}$, $T_{11\text{-}I}$, $T_{22\text{-}I}$, Holm I-$T_{13\text{-}phase}$*
8	*Shanon-Entropy-I (SE_I), $T_{33\text{-}I}$, $T_{11\text{-}I}$, $T_{22\text{-}I}$, Holm I-$T_{12\text{-}phase}$, Krogager-d, Barnes I-$T_{11\text{-}I}$, Cross-Correlation Coefficient*
9	*Span, S_{12}-$_{phase}$, Shanon-Entropy-I (SE_I), Anisotropy-Lueneburg (AL), $T_{33\text{-}I}$, $T_{11\text{-}I}$, $T_{22\text{-}I}$, HolmII-$T_{23\text{-}phase}$, BarnesI-$T_{12\text{-}phase}$*
10	*Span, Shanon-Entropy-I (SE_I), $\|S_{HH}\|^2/\|S_{HV}\|^2$, $T_{33\text{-}I}$, $T_{23\text{-}phase}$, $T_{11\text{-}I}$, $T_{22\text{-}I}$, Cloude-$T_{11\text{-}I}$, Barnes I-$T_{13\text{-}phase}$, $\|S_{HV}\|^2/\|S_{VV}\|^2$*

Table 7.2: subset-3 to 10 basedon GA-MI of 127 descriptors.

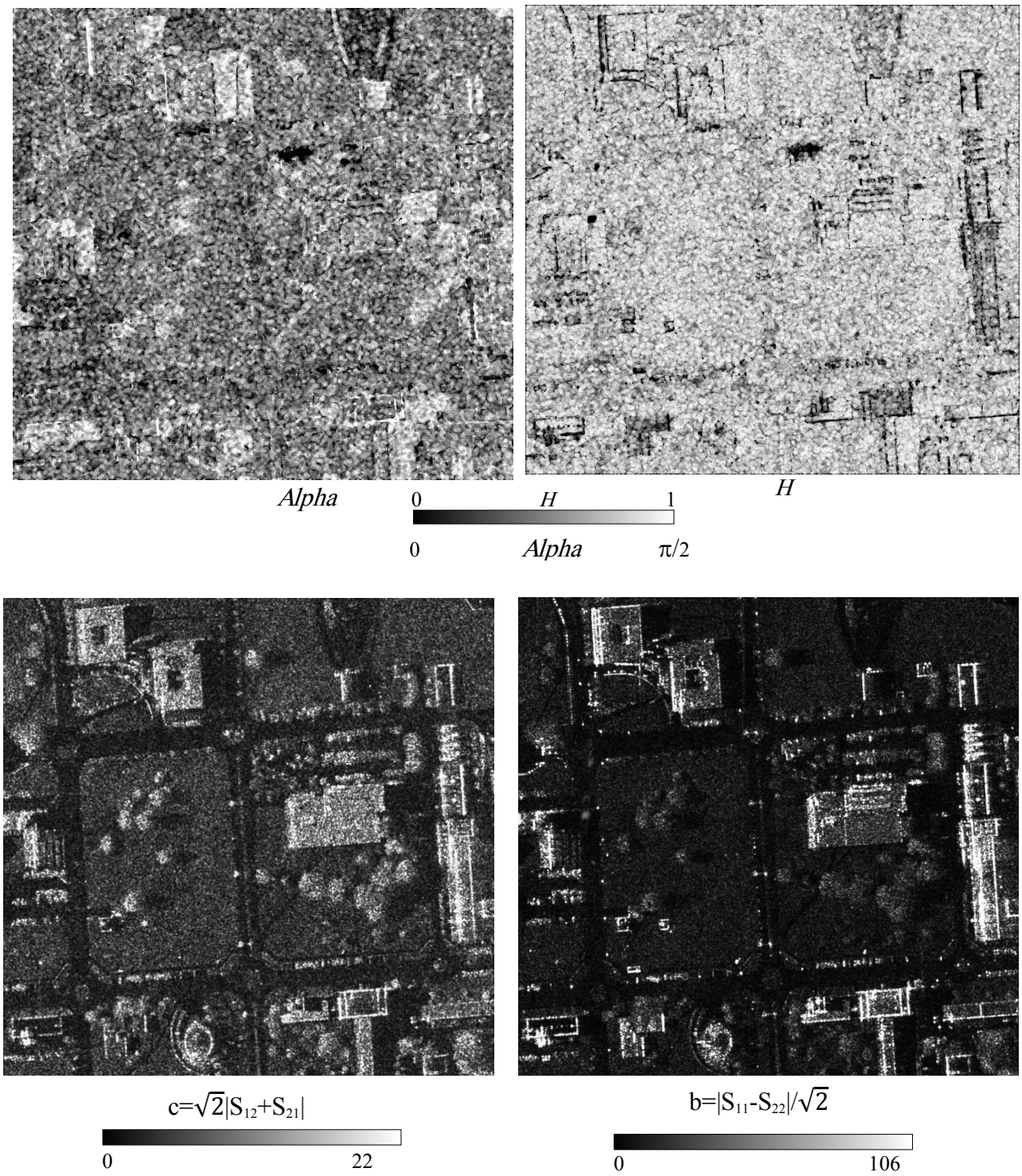


Figure 7. 3: (a): Some of polarimetric descriptors

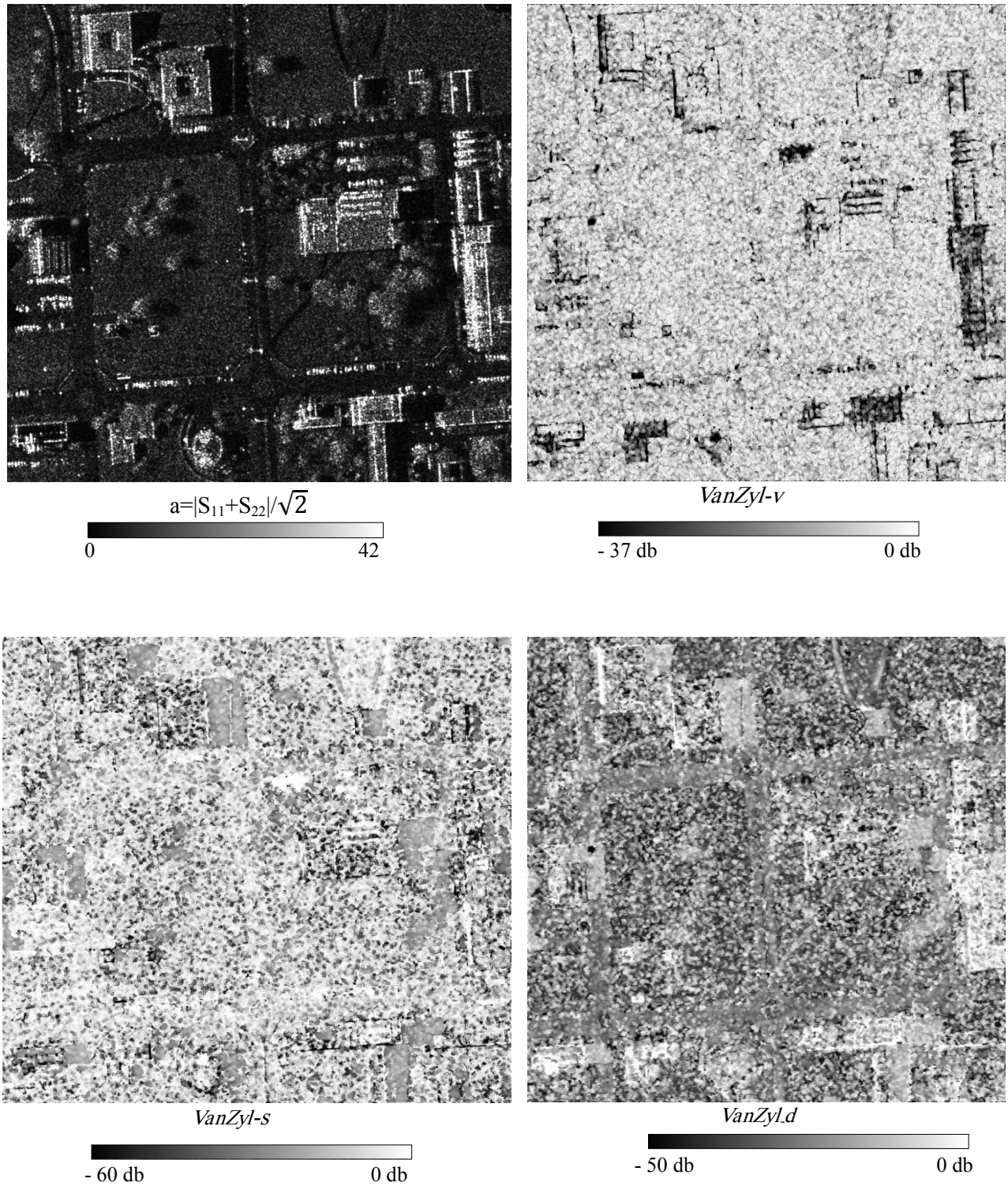


Figure 7. 3 (b): Some of polarimetric descriptors

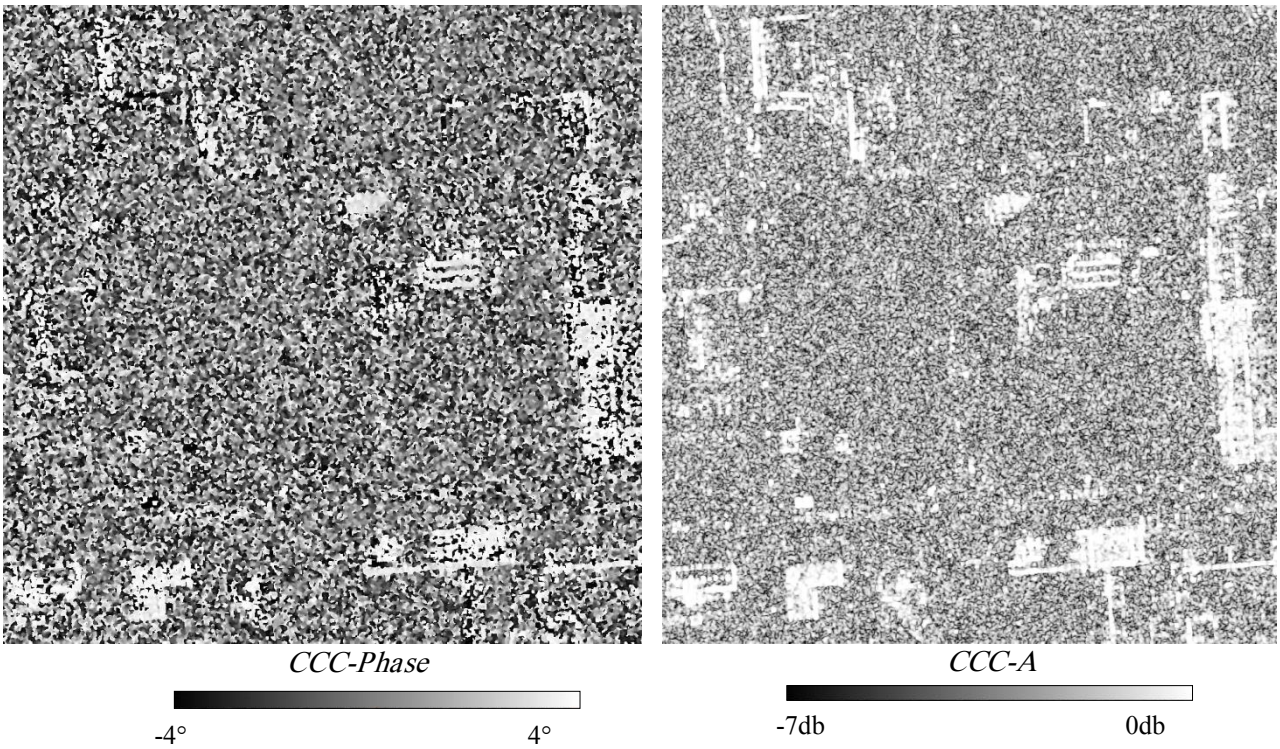


Figure 7. 3 (c): Some of polarimetric descriptors

7.1.2. GA-MI feature selection

The feature subset selection method (SSM) consists in using GA-MI which is detailed in section 3.1.2. In GA, the elitism parameter and the validation threshold are set to 1 and positive value 0.002 respectively. The main questions are: what are the optimal number of generation (iteration) and number of descriptors to consider during the GA-MI process? To answer these questions, first for known population and 58 real positive descriptors, we consider the relationship between iteration number and maximum evalution value for different sizes of feature set (3,4,5,….,20) through GA and mRMR criteria (Equation 3.7). Figure 7.4 (a) shows that after several iterations, the evaluation value does not vary a lot from one set size to the other. In fact, the relation between generation and evaluation is close to an exponential shape. To select the number of descriptors, the classification accuracies (overall accuracy) for different set sizes (3, 4, 5,…., 20) are calculated. Moreover, for this purpose, a uni-variate Gaussian-ML classifier is used. Figure 7.4 (b) shows that the optimal feature number could be 5, with an overall accuracy around 83.5% and an iteration number of 20. This optimal subset selection (Table 7.1) is used in the statistical analysis, and classification performance. In addition, Figure 7.5 shows the MI between each seven classes and real positive descriptors of SSM based on GA-MI through 58 real positive descriptors. The descriptors of optimal subset are shown in Figure 7.5.

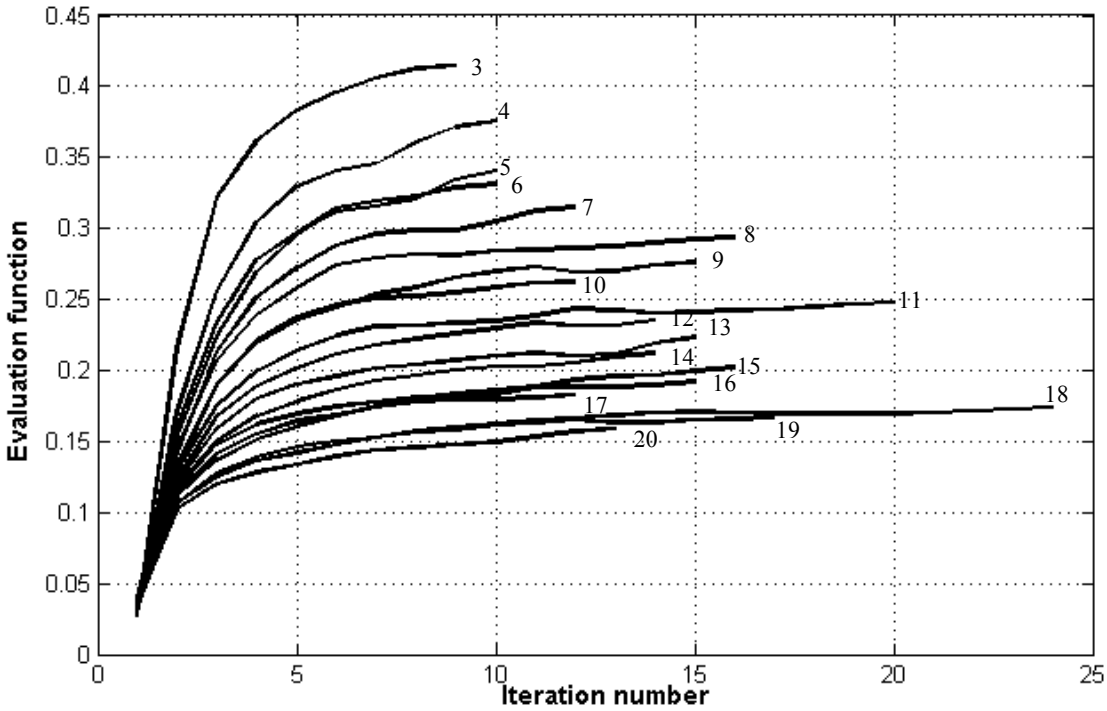
Similarly, these procedures for complete descriptors consist in using 127 descriptors from different types of polarimetric parameters (Figure 7.2), for different subsets from 3 to 20. Figure 7.6 (a) shows the evaluation function value in term of iteration number for sets of 3 to 20 descriptors of SSM based on GA-MI from 127 complete descriptors. According to Figure 7.6 (a) the iteration number of 16 is selected because after this iteration the value of evalution function does not change with more iteration. Figure 7.6 (b) shows the overal accuracy from uni-variate Gaussian-ML in term to number of descriptors. The list of descriptors in each subset (subset-3 to subset-10) is summerized in Table 7.3.

Moreover, MI between each class and descriptor of different subsets (subset-3 to subset-10 descriptors of SSM based on GA-MI through 127 descriptors) are calculated. The results are given in Figure 7.7 to Figure 7.13. Some of selected descriptors of these different subsets are shown in Figure 7.14 to Figure 7.16. The results for five real positive descriptors and optimal subset of 127 complete descriptors are given in Figure 7.5 and Figure 7.8 respectively. Moreover, the list of descriptors that have the maximum relevancy with referred classes, are given in Table 7.4. However, they are not the best discriminators for related class.

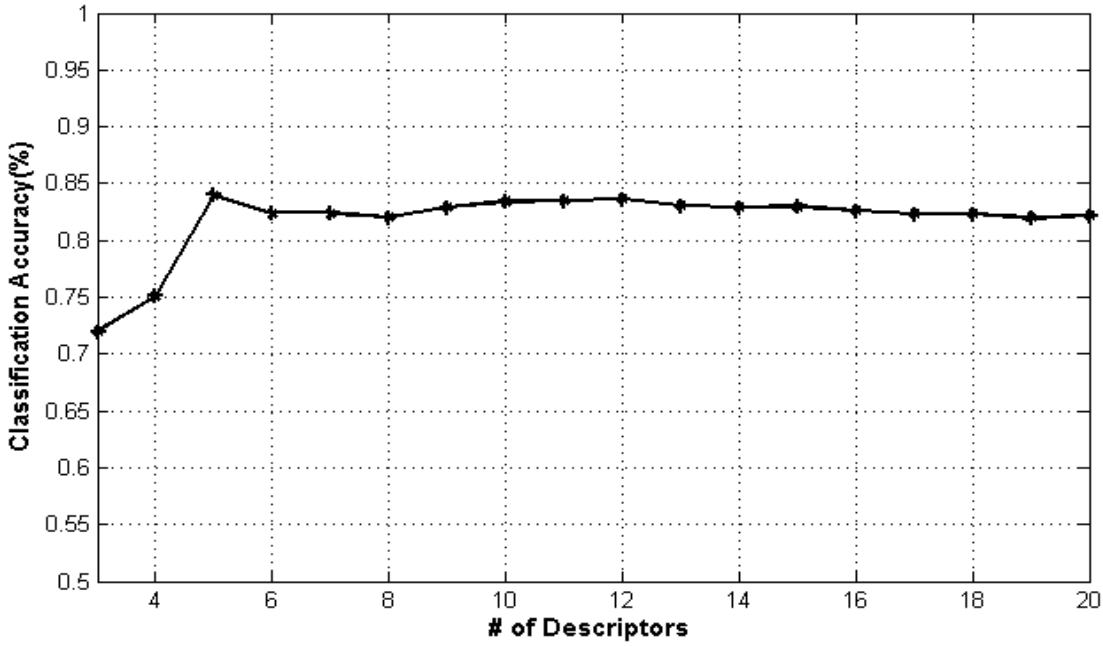
Furthermore, two remarks on the selected features in subset-3 to subset-10 (Table 7.3) are: firstly, at least one phase corresponding to difference of phase of the Sinclair components is always retained. Secondly, we can recognize in the selected features, some features more adapted for a particular kind of scattering mechanism such as *H/A/Alpha-*$T_{11\text{-}I}$, *Barnes I-*$T_{11\text{-}I}$, *Krogager-d*, *Cloude-*$T_{11\text{-}I}$.

To calculate feature selection based on Genetic algorithm using mutual information, we use MATLAB source code (Ludwig & Nunes, 2010) (updated 11 Sep 2012, available at[1]). Moreover, we use this code for processing SAR image data and adapted it for evaluating the GA-MI initial parameters, and calculate the mutual information between each class and descriptors.

[1] http://www.mathworks.com/matlabcentral/fileexchange/29553-feature-selector-based-on-genetic-algorithms-and-information-theory.



(a)



(b)

Figure 7. 4: (a) Evaluation function value vs iteration number for sets of 3 to 20 real positive descriptors. (b) Overal accuracy from uni-variate Gaussian-ML vs number of real positive descriptors.

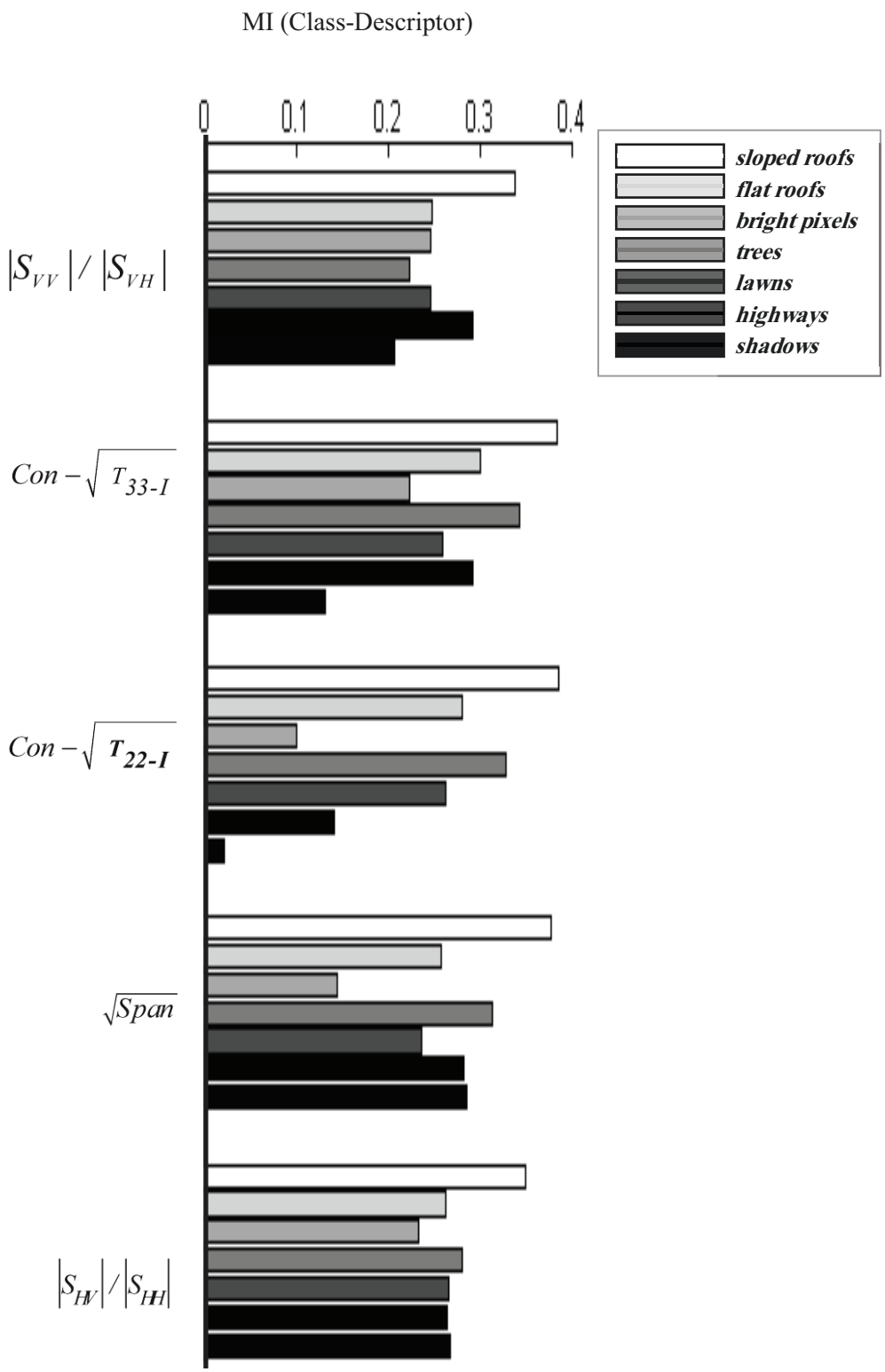


Figure 7. 5: Mutual information between class and real positive descriptor in optimal subset of SSM based on GA-MI through 58 real positive descriptors.

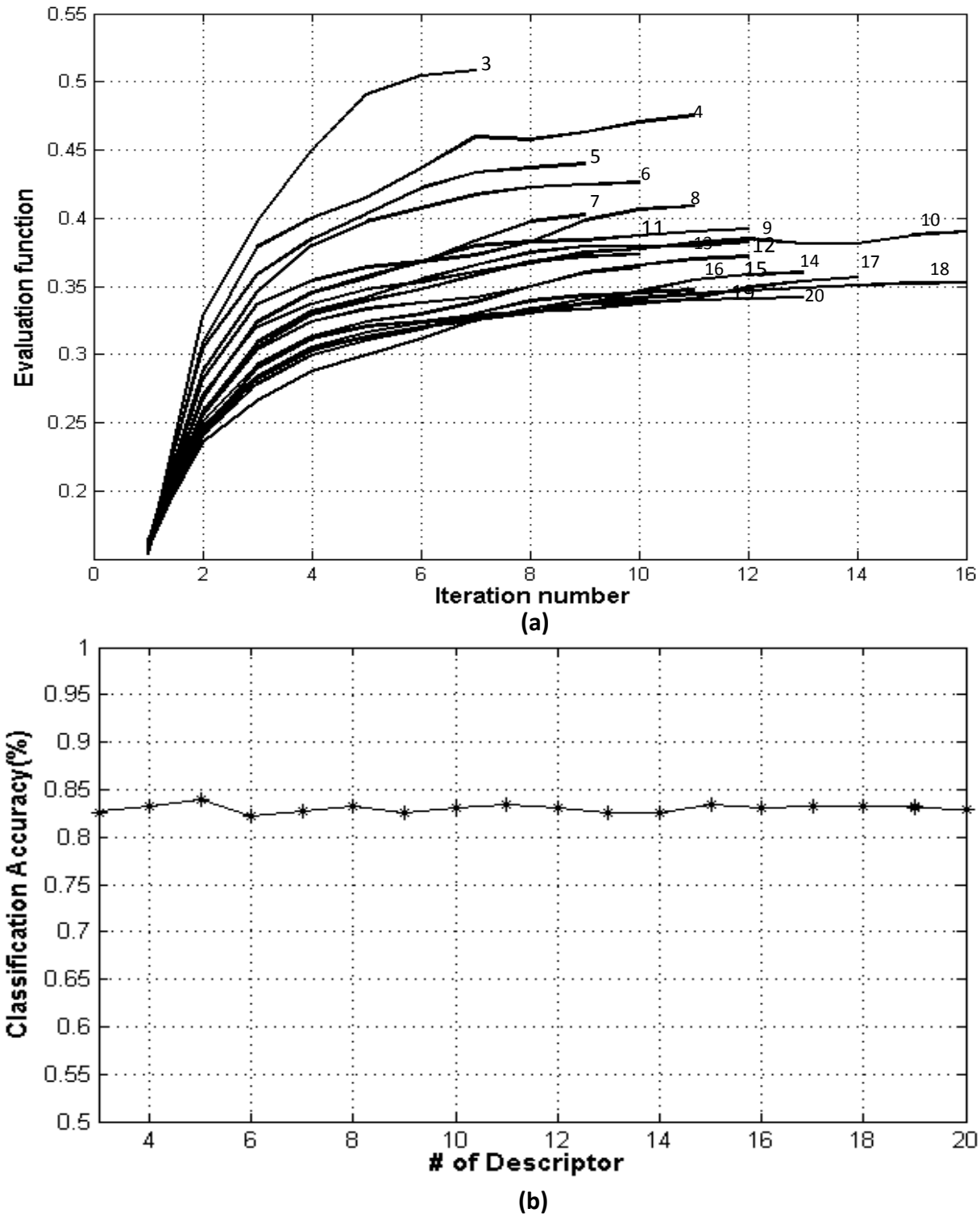


Figure 7. 6:(a) Evaluation function value vs iteration number for sets of 3 to 20 descriptors of SSM based on GA-MI from 127 complete descriptors. (b) Overall accuracy from uni-variate Gaussian-ML vs number of descriptors.

Table 7. 3: List of descriptors from 3 to 10 subset of SSM based on GA-MI through 127 complete descriptors

Subset.	**Selected Subset Descriptors**
3	*Span, Polarization-Fraction* (*PF*), *Barnes I-*$T_{12\text{-}phase}$
4	*Shanon-Entropy-I* (SE_I), *Anisotropy-Lueneburg* (*AL*), T_{33}, *Cloude*.$T_{12\text{-}phase}$
5	*Shanon-Entropy-I* (SE_I), *H* (*1-A*), $T_{33\text{-}I}$, $T_{12\text{-}phase}$, *H/A/Alpha-*$T_{11\text{-}I}$
6	*Span,* $T_{33\text{-}I}$, *Anisotropy-12,* $T_{11\text{-}I}$, *Holm I-*$T_{23\text{-}phase}$, *Barnes I-*$T_{12\text{-}phase}$
7	*Span, Anisotropy-Lueneburg* (*AL*), $T_{33\text{-}I}$, $T_{23\text{-}phase}$, $T_{11\text{-}I}$, $T_{22\text{-}I}$, *Holm I-*$T_{13\text{-}phase}$
8	*Shanon-Entropy-I* (SE_I), $T_{33\text{-}I}$, $T_{11\text{-}I}$, $T_{22\text{-}I}$, *Holm I-*$T_{12\text{-}phase}$, *Krogager-d, Barnes I-*$T_{11\text{-}I}$, *Cross-Correlation Coefficient* (*CCC-A*)
9	*Span,* $S_{12\text{-}phase}$, *Shanon-Entropy-I* (SE_I), *Anisotropy-Lueneburg* (*AL*), $T_{33\text{-}I}$, $T_{11\text{-}I}$, $T_{22\text{-}I}$, *HolmII-*$T_{23\text{-}phase}$, *BarnesI-*$T_{12\text{-}phase}$
10	*Span, Shanon-Entropy-I* (SE_I), $\lvert S_{HH}\rvert^2/\lvert S_{HV}\rvert^2$, $T_{33\text{-}I}$, $T_{23\text{-}phase}$, $T_{11\text{-}I}$, $T_{22\text{-}I}$, *Cloude-*$T_{11\text{-}I}$, *Barnes I-*$T_{13\text{-}phase}$, $\lvert S_{HV}\rvert^2/\lvert S_{VV}\rvert^2$

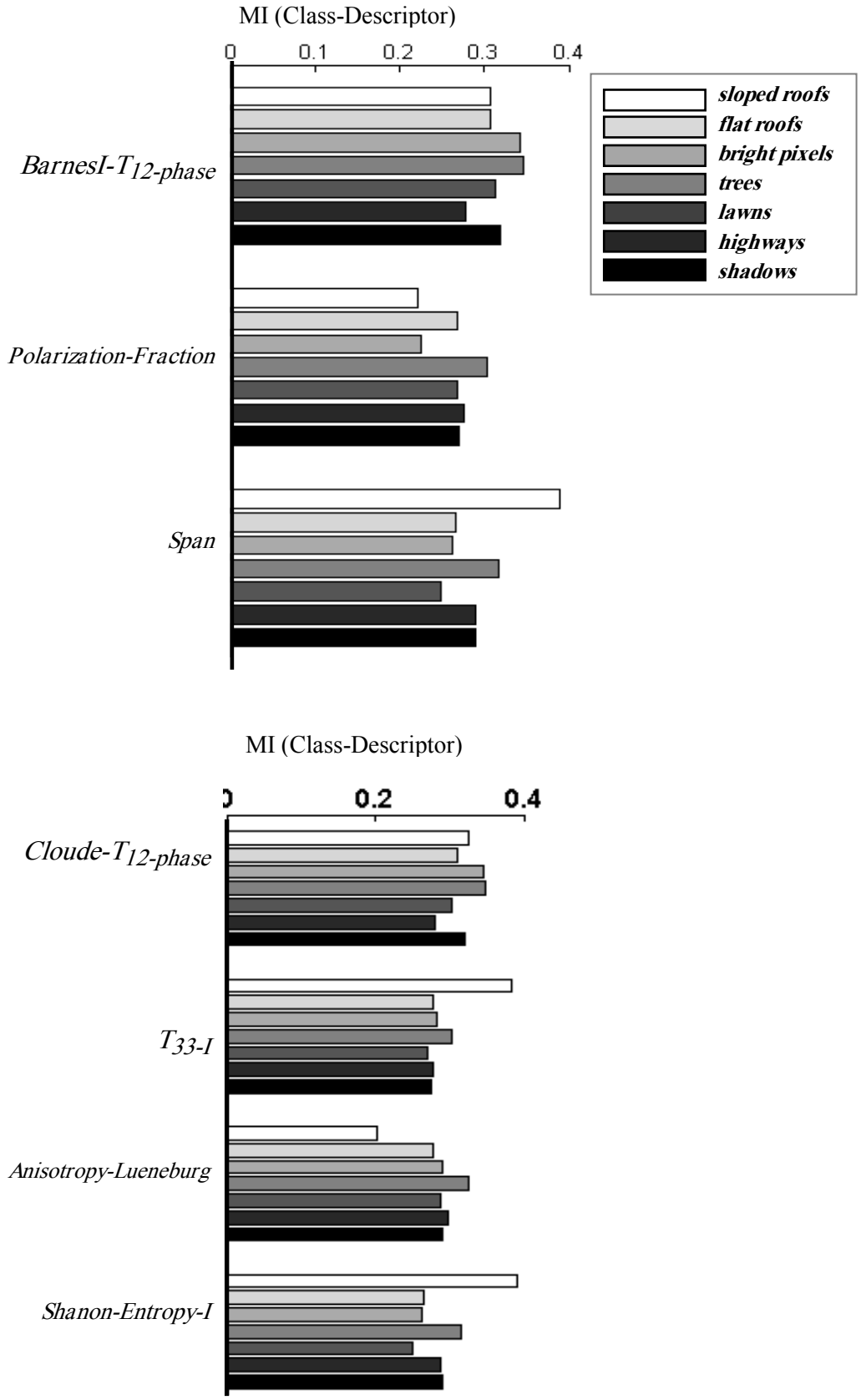


Figure 7. 7: Mutual information between class and descriptor in subset 3 and 4 of SSM based on GA-MI through 127 descriptors.

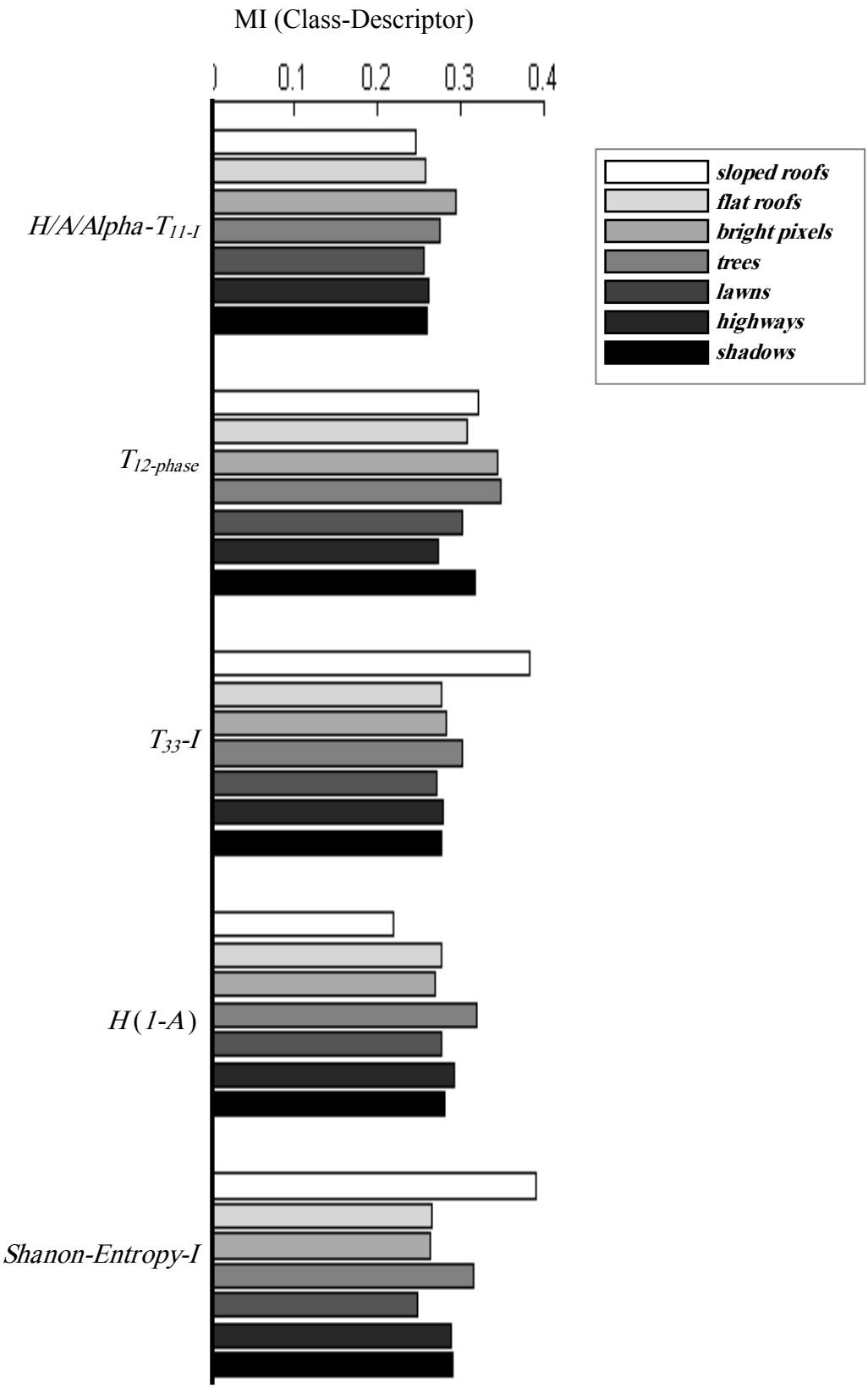


Figure 7. 8: Mutual information between class and descriptor in subset 5 (optimal subset) of SSM based on GA-MI through 127 descriptors.

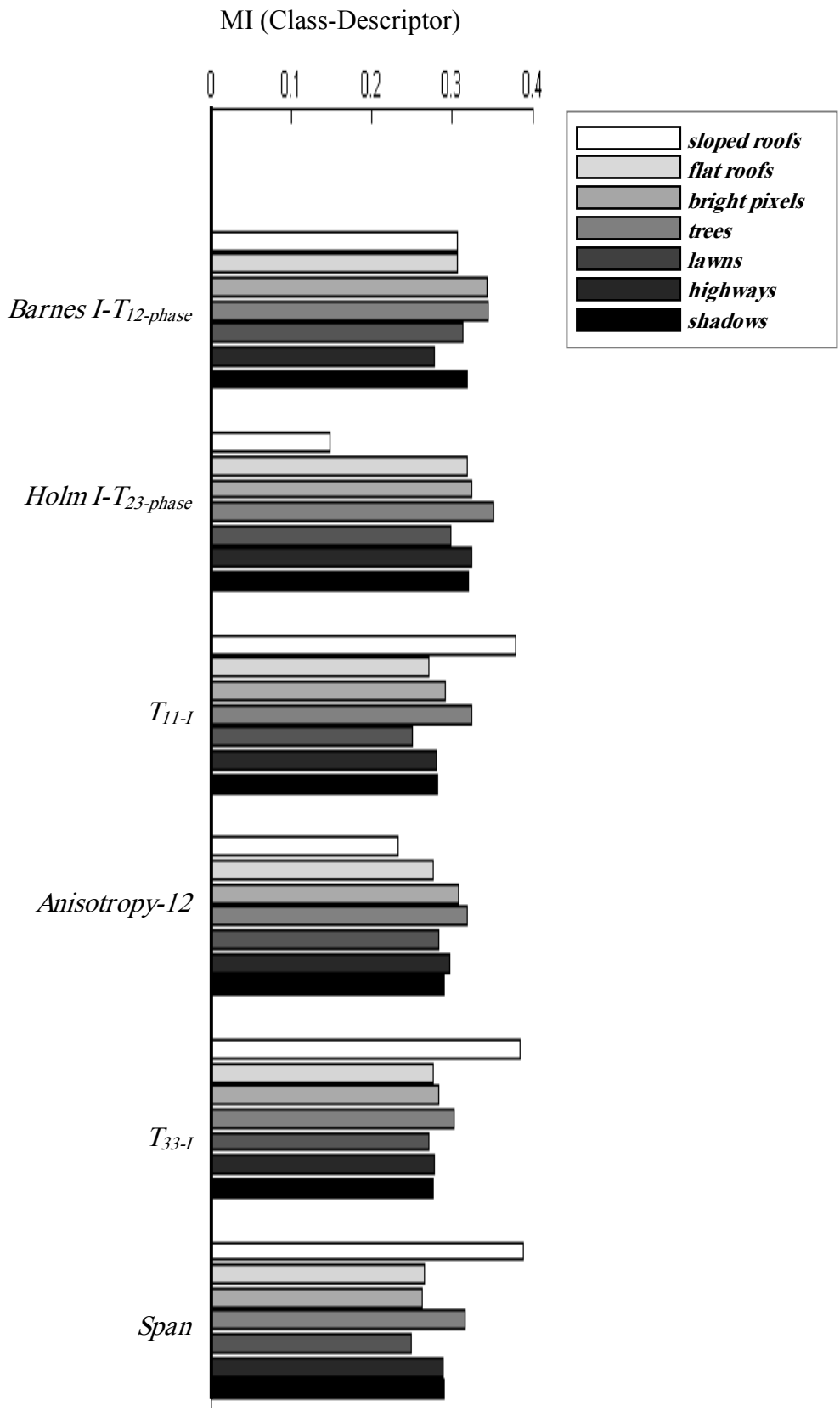


Figure 7. 9: Mutual information between class and descriptor in subset 6 of SSM based on GA-MI through 127 descriptors.

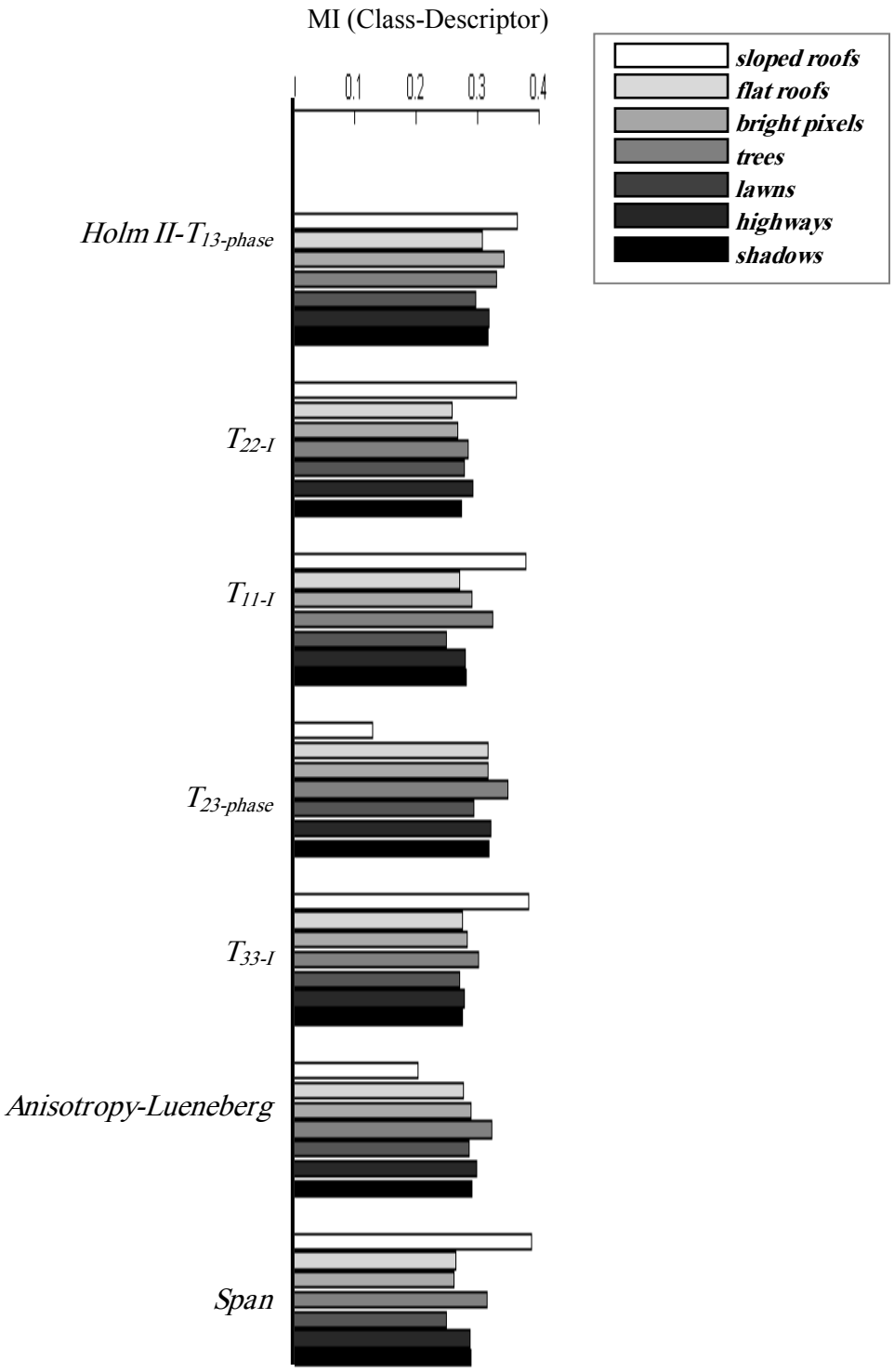


Figure 7. 10: Mutual information between class and descriptor in subset 7 of SSM based on GA-MI through 127 descriptors.

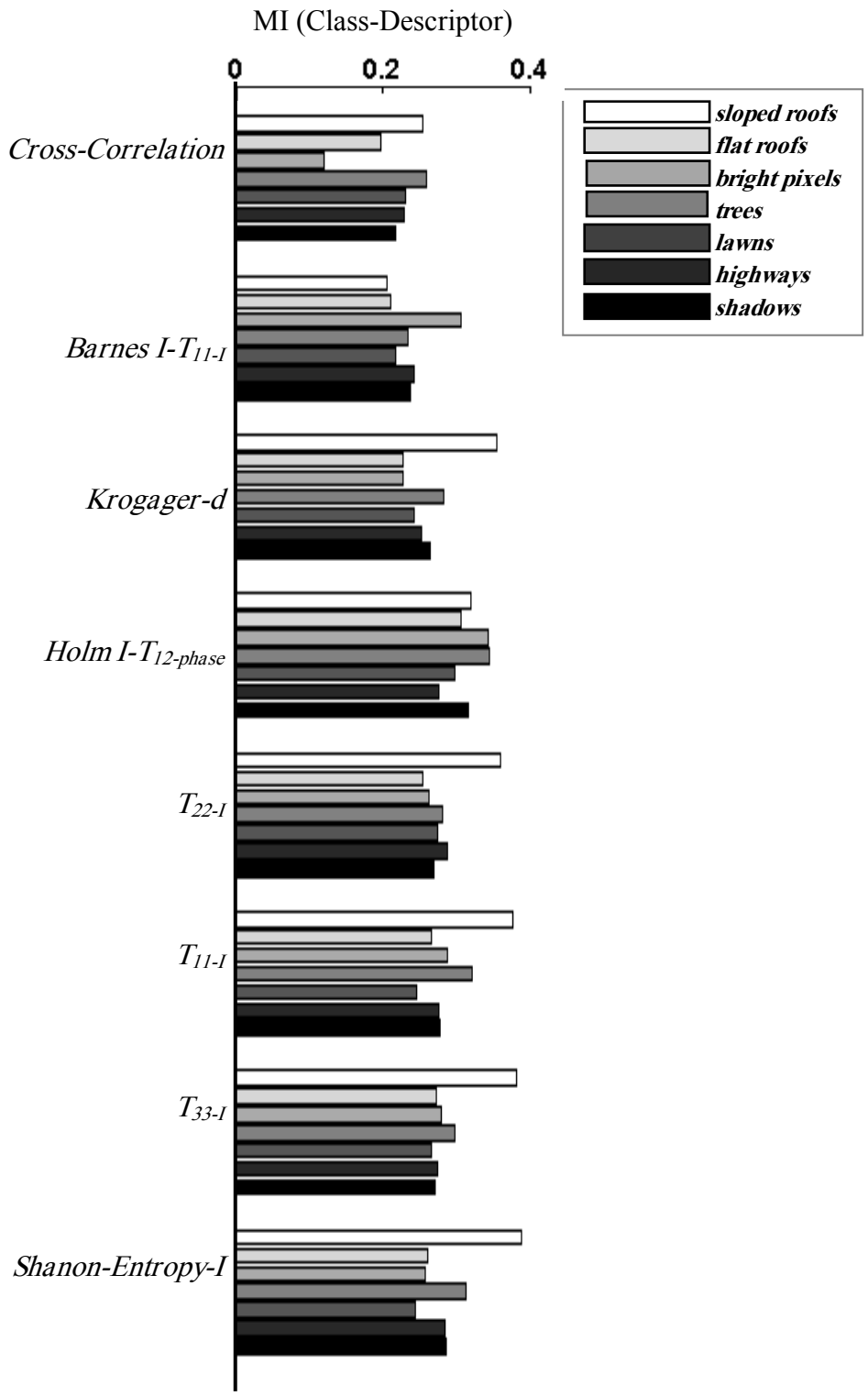


Figure 7. 11: Mutual information between class and descriptor in subset 8 of SSM based on GA-MI through 127 descriptors.

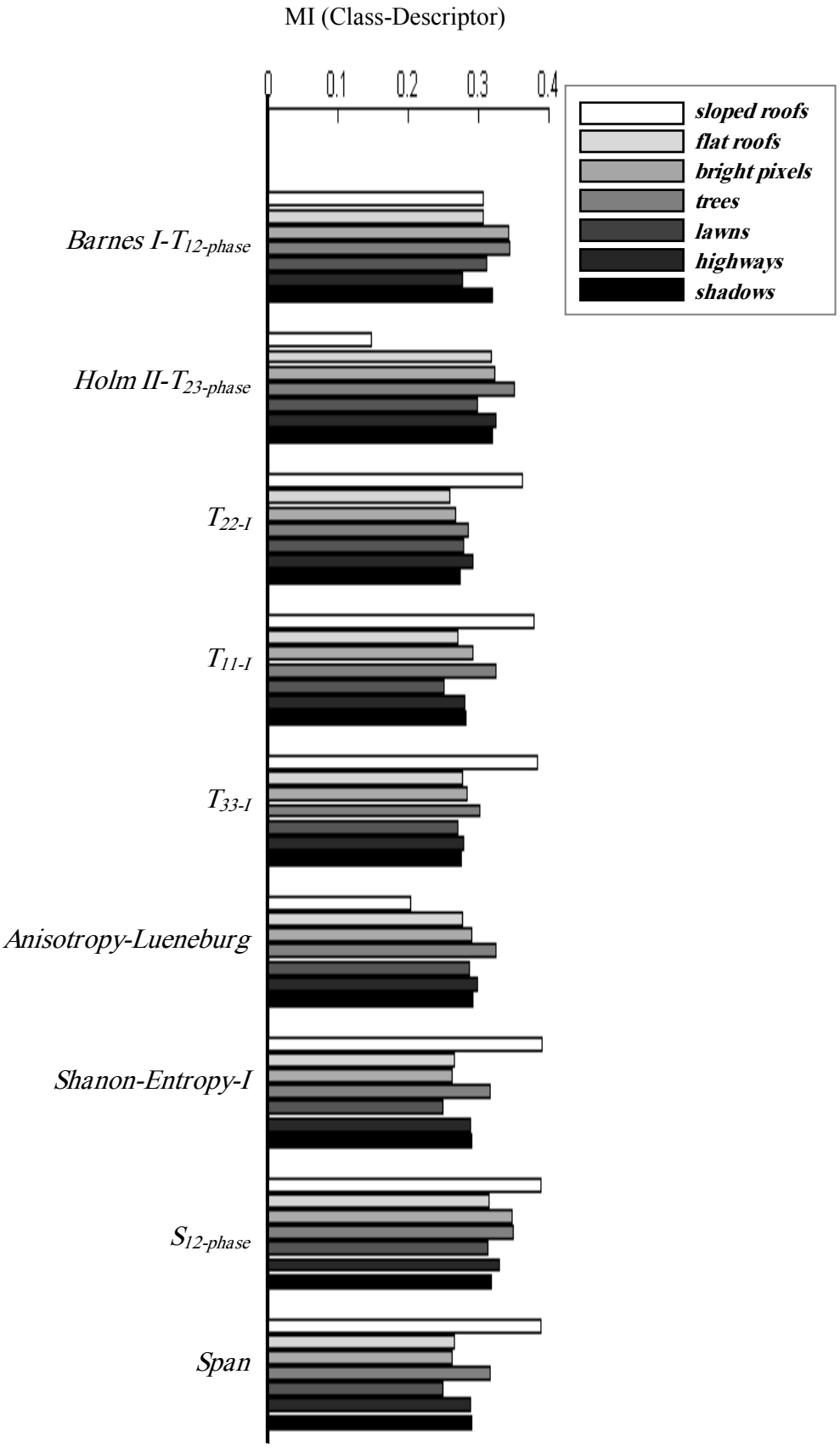


Figure 7. 12: Mutual information between class and descriptor in subset 9 of SSM based on GA-MI through 127 descriptors.

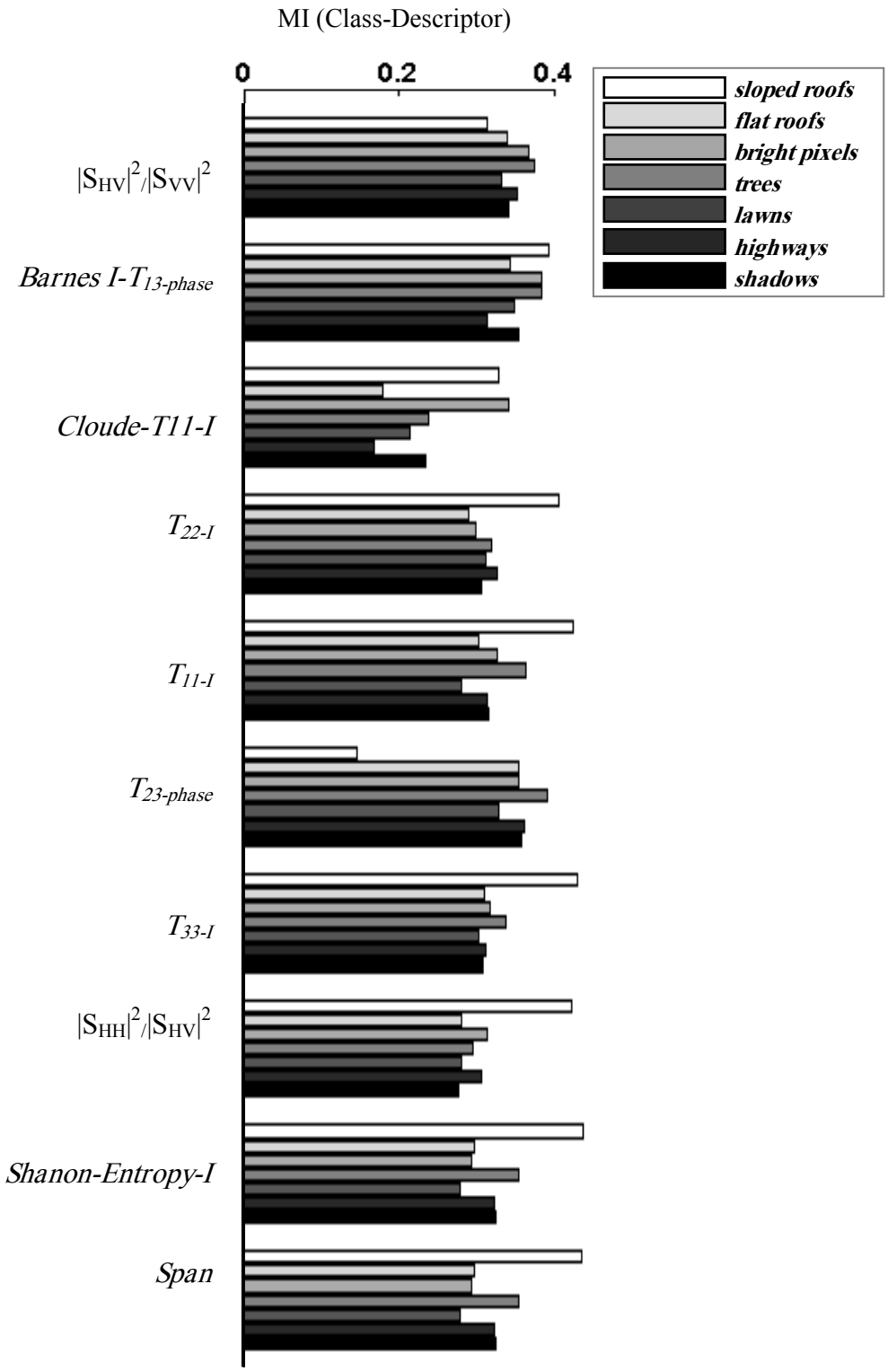


Figure 7. 13: Mutual information between class and descriptor in subset 10 of SSM based on GA-MI through 127 descriptors

Table 7. 4: List of descriptors from 3 to 10 subset of SSM based on GA-MI with maximum relevancy with class

Class	Descriptor
Shadows	*Barnes I-$T_{12\text{-}phase}$, Cloude-$T_{12\text{-}phase}$, $T_{12\text{-}phase}$, Holm I-$T_{13\text{-}phase}$, $T_{23\text{-}phase}$, Holm I-$T_{12\text{-}phase}$*
highways	*Span, Anisotropy-Lueneburg (AL), H (1-A), $T_{23\text{-}phase}$, Holm I-$T_{13\text{-}phase}$, T_{22}-I, $\lvert S_{HV}\rvert^2/\lvert S_{VV}\rvert^2$*
lawns	*Barnes I-$T_{12\text{-}phase}$, Cloude-$T_{12\text{-}phase}$, $T_{12\text{-}phase}$, Holm I-$T_{13\text{-}phase}$, Holm I-$T_{12\text{-}phase}$*
trees	*Barnes I-$T_{12\text{-}phase}$, Cloude-$T_{12\text{-}phase}$, $T_{12\text{-}phase}$, $T_{23\text{-}phase}$, Holm I-$T_{23\text{-}phase}$, Barnes I-$T_{12\text{-}phase}$, Holm I-$T_{12\text{-}phase}$, Cross-Correlation Coefficient (CCC-A), Barnes I-$T_{13\text{-}phase}$*
Bright pixels	*Barnes I-$T_{12\text{-}phase}$, Cloude-$T_{12\text{-}phase}$, $T_{12\text{-}phase}$, Holm I-$T_{13\text{-}phase}$, Holm I-$T_{12\text{-}phase}$, Barnes I-$T_{13\text{-}phase}$*
Flat roofs	*Barnes I-$T_{12\text{-}phase}$, $T_{12\text{-}phase}$, $T_{23\text{-}phase}$, Holm I-$T_{12\text{-}phase}$*
Sloped roofs	*Span, Shanon-Entropy-I (SE_I), T_{33}-I, T_{11},-I, Krogager-d*

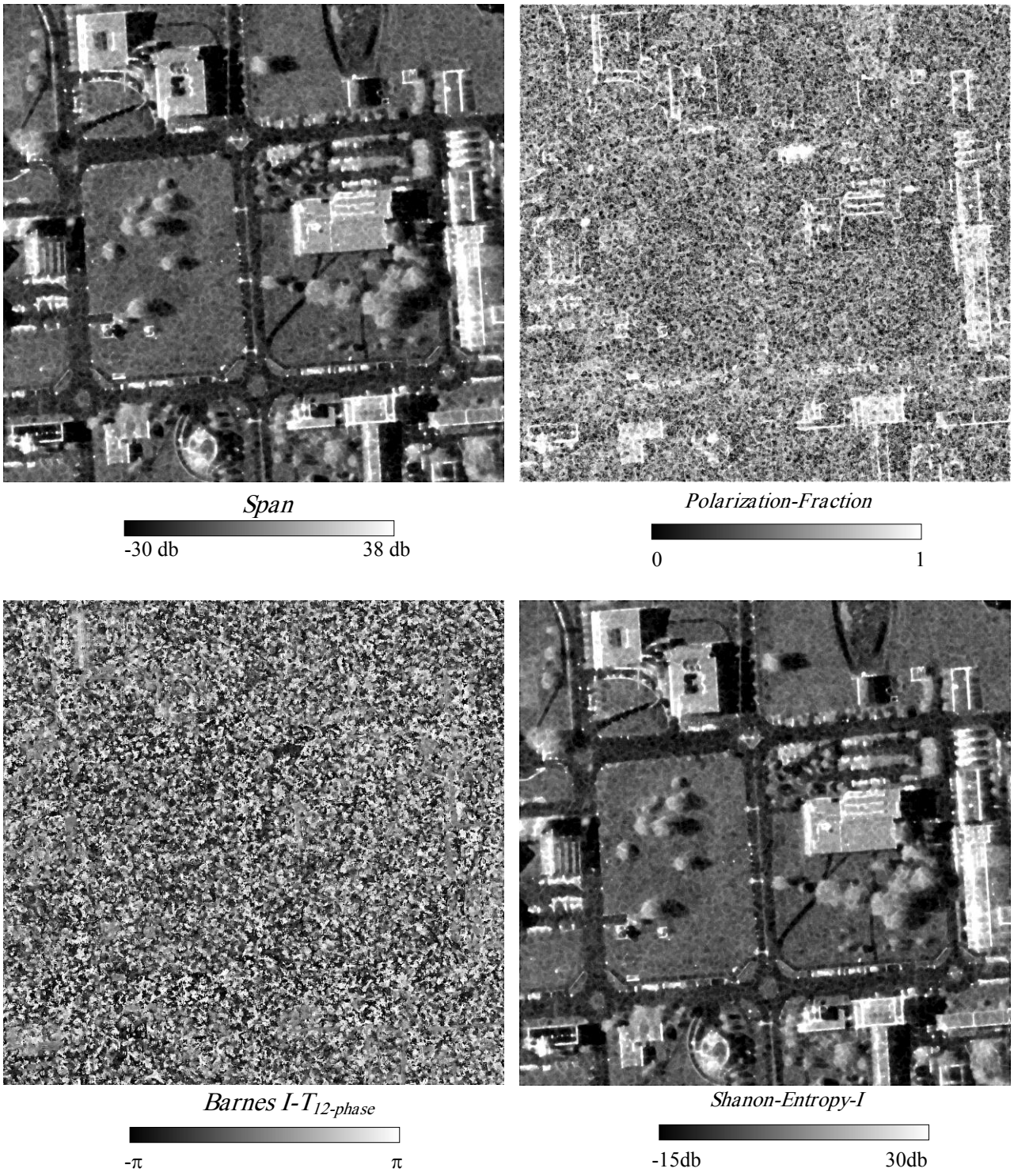


Figure 7. 14:(a) Some of the selected descriptors in the subsets with 3 to 10 features obtained from SSM using 127 descriptors.

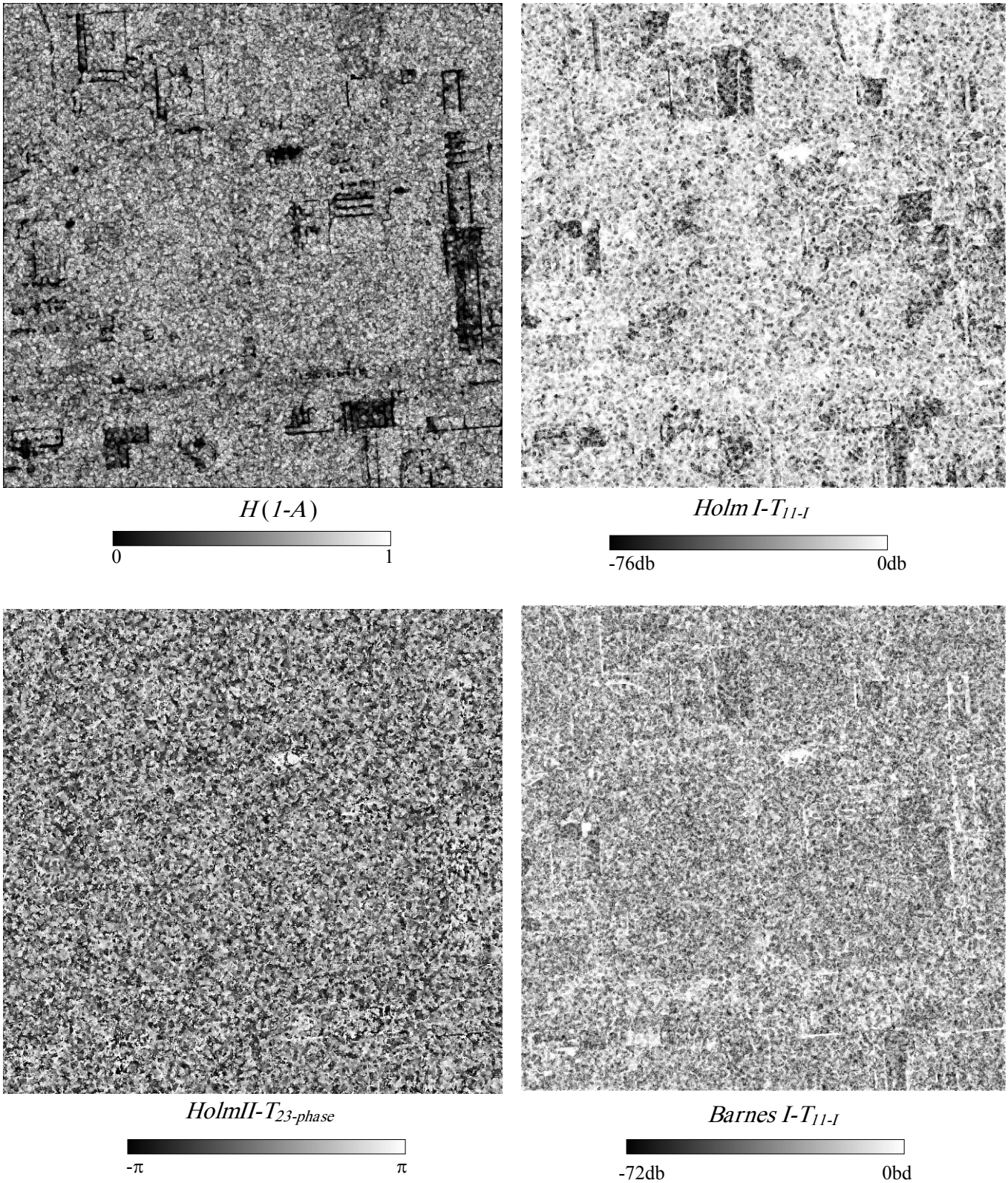


Figure 7. 15:(b) Some of the selected descriptors in the subsets with 3 to 10 features obtained from SSM using 127 descriptors.

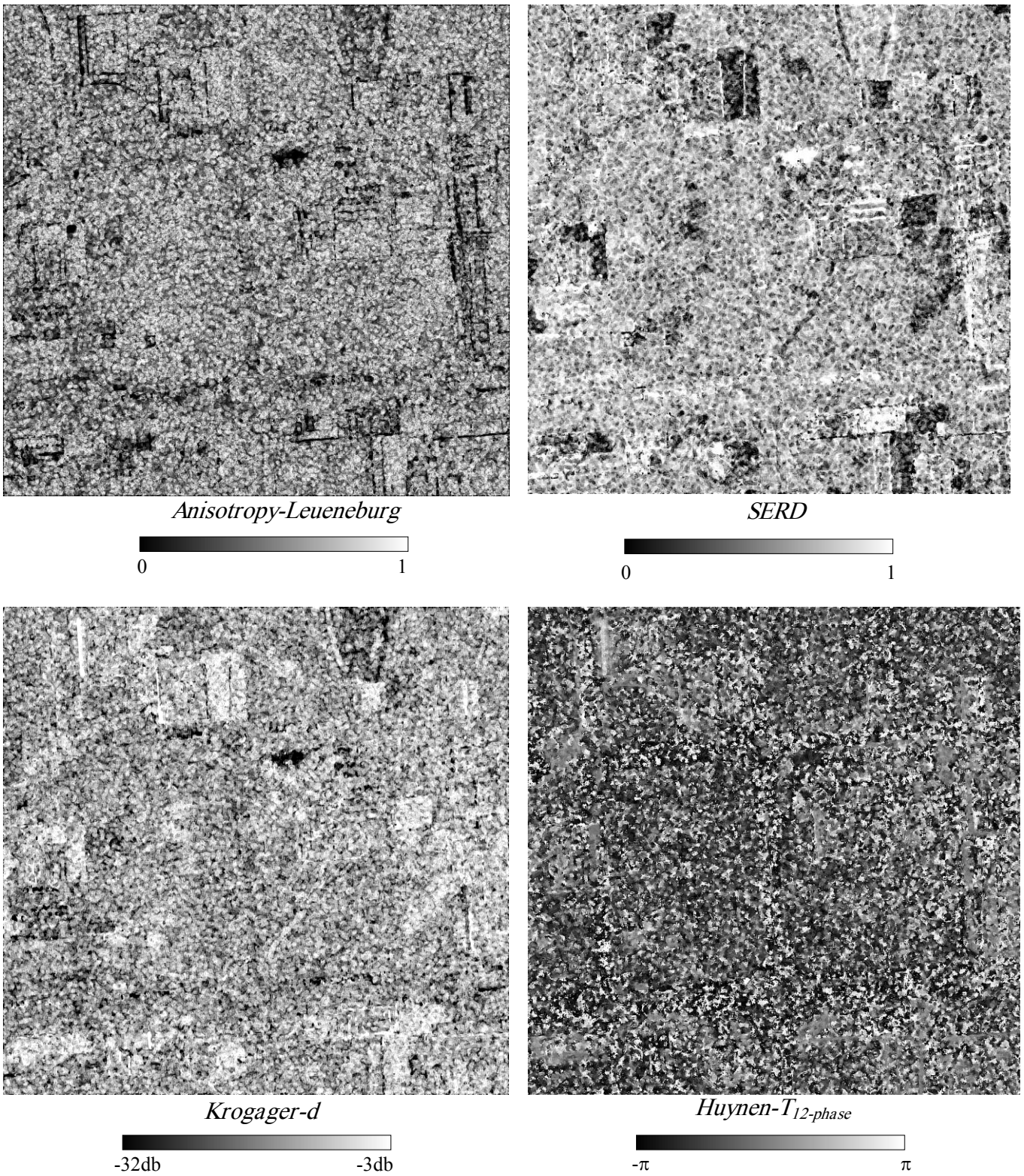


Figure 7. 16:(c) Some of the selected descriptors in the subsets features obtained from SSM using 127 descriptors.

7.2. Morphology process and result

7.2.1. Methodology

Here, we use the proposed morphological approach to produce the building footprints, previously described in chapter 2, section 2.4. The building map is extracted from semi-automatic method using PoLSAR amplitude (S_{VV}) image through the following steps:

– Morphological basic operation: first, the morhological map is obtained by Otsu threshold. We can also produce binary image by introducing manually threshold using semi-automatic morphology tools. Then, we remove region with less than 'option' pixels with morphologically open operator (remove all small objects) over binary image. The effect of this operation is to remove all structures that are both smaller than the 'option' pixels and brighter than the surroundings such as spot point and noise. Thus, all streets, railways following with spot points are eliminated. Then, we apply 'closing minus opening' using the square SE which maintains the actual size of the objects and removes all remain noises and also fill the holes in our interesting area.

To provide a semi-automatic method to extract building map using morphological tools, we used from MATLAB source code which is available at[2]. We reviewed and developed this source code in a convenient manner for our purpose.

7.2.2. Results

The 'option' pixels size relates to the building interested size, which is selected from semi-automatic tools and apply to binary image. For example, if 'option' pixels have small size, the result will consist in trees, highways and noise. On the other hand, a too large size may miss certain building footprints. To remove any residual noise and to maintain actual building size, the 'opening minus closing' operator is used. The building footprints based on morphological tool results over S_{VV} amplitude from semi automatic morphological tool is shown in Figure 7.17, where, the 'option' pixel size is 1505. For 'closing minus opening' performance the specified neighborhood parameter is 7. This result shows that most of buildings in the image can be extracted. As the real ground truth is not available, the result morphology performance is only validated by a comparison of classification result, with and without this information (see Chapter 8).

[2] http://www.mathworks.com/matlabcentral/fileexchange/11242-interactive-binarization-by-thresholding

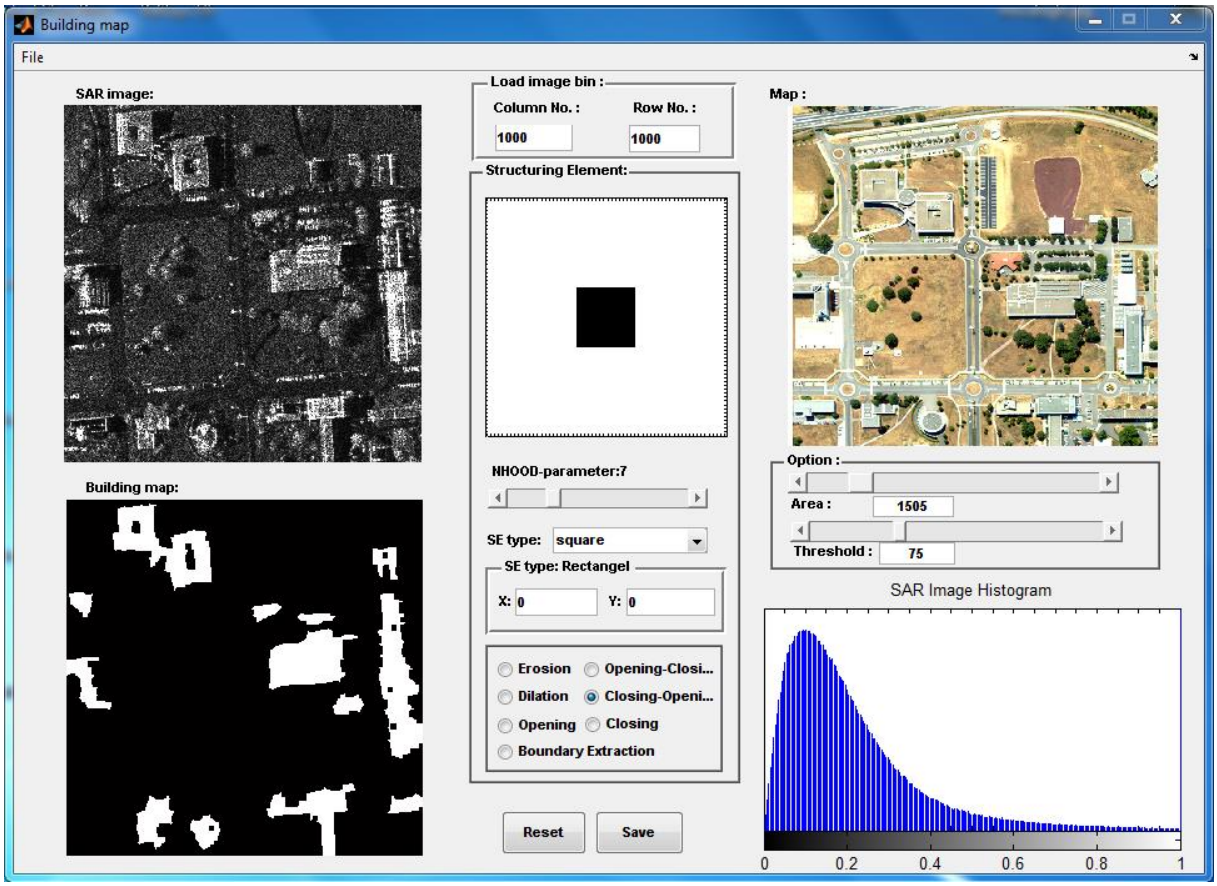


Figure 7.17: Extracting building footprints map from S_{22-A} image through semi-automatic algorithm.

Chapter 8:
Results on statistical analysis and classification

8. Results on statistical analysis and classification

The main goal of this statistical analysis is to determinate whether the differences between statistical behaviors are sufficient for recognizing the different classes or not.

In the first experiment, the optimal real positive subset from SSM is used for statistical analysis. This performance consists in extracting training samples from the representative area of each class (*shadows*, *highways*, *lawns, trees*, *bright pixels, flat roofs* and *sloped roofs*) in the SAR image, train statistical modeling, and evaluation of fitness between empirical data and theoretical models. Then, we compare Fisher-ML classifier with other classifiers (SVM and G-ML). The performance of supervised Fisher-ML and SVM cassifiers following with morphological map (building map) which is extracted from semi-automatic algorithm based on morphological operations are assessed on descriptors (optimal subset of real positive descriptors).

In the second experiment, the performance of the SVM classifier is computed on the different subsets of SSM from all descriptors and compare with G-ML^{C} (maximum likelihood classification based on muti-variate complex Gaussian pdf). In the following, the performance of the SVM classifier over the optimal subset from all descriptors and G-ML^{C} following with morphological map are assessed.

Then, the classification results are compared with SVM classifier from different types of decomposition components. In fact, to illustrate the efficiency of each subset of SSM based on GA-MI, we make a comparison using the SAR target decompositions based on different scattering mechanisms, including the Pauli, Krogager, based on back scattering matrix (full resolution coherent technique) Freeman, VanZyl, Yamaguchi three and four-Component, Barnes, Holm, Huynen,Touzi and the Cloude decomposition. In the comparison experiment, the decomposition coefficients of each polarimetric matrix with different subset of SSM are processed with SVM classifier to generate the classification results described in Chapter 8.

8.1. Statistical analysis procedure and results

8.1.1. Methodology

After obtaining the selected features (Table 7.1), statistical analyses are applied on each of them over all classes. The purpose of this step is to select the best statistical model for each train data and each feature using models and methods explained in Chapter 4. The statistical experiment is performed on optimal subset of SSM based on GA-MI through 58 real positive descriptors which are summarized in Table 8.1.

This procedure consists in:

- Estimating the Fisher, $\mathcal{K}$, Gamma, Beta I, Log-normal and Weilbull distributions parameters using MLE, and MoCL.
- Calculating the GoF using K-S (test), correlation coefficients and RMSE.

To estimate pdf parameters, we have written the source code according to (Nicolas, 2006) suggestion method and developed it for different type of pdf (Gamma, Fisher, Weibull, $\mathcal{K}$, Beta I and Log-normal) estimation parameters.

Table 8. 1: List of real positive descriptors in optimal subset of SSM based on GA-MI through the 58 descriptors

# Of Des.	1	2	3	4	5
Des.	$\lvert S_{HV}\rvert / \lvert S_{HH}\rvert$	$\sqrt{Span}$	$CON\text{-}\sqrt{T_{22\text{-}A}}$	$CON\text{-}\sqrt{T_{33\text{-}A}}$	$\lvert S_{VV}\rvert / \lvert S_{HV}\rvert$

8.1.2. Results

The experimental results from statistical analysis show that among some of well-known statistical models such as $\mathcal{K}$, Weibull, Log-normal, Beta I and Gamma, the Fisher pdf is able to model very different media (such as different surfaces in urban area as well as other radar attributes like shadows and bright pixels). The MoLC estimation method is particularly appropriate to take into account the three parameters (scale and shape parameters) of the Fisher pdf. According to the results of GoF (results are shown in Appendix A, Table A.1-7), this method can successfully lead to better fitness values with empirical data compared to ML estimation method.

The experimental results confirm that MoLC can mostly provide accurate parameters estimates for all sample data over all descriptors, both from the view point of visual comparisons (Figure 8.1 (a-c)) and GoF values (Appendix A, Table A.1-7). For example, in some cases, the estimator Fisher–MLE does not exist, because the size of taining samples is small. Besides, the correlation coefficients are higher than 90% and GoF-RMSE are minimum for Fisher pdf using MoLC for all classes and all descriptors (Appendix A, Table A.1-7). For example, the correlation coefficients related to the De-polarization $\lvert S_{HV}\rvert / \lvert S_{HH}\rvert$-ratio for *flat roofs*, are 0.9882, 0.9881, 0.9866, 0.9863, 0.9890, and 0.9891, respectively, for the following models: Log-Normal (MoLC), Gamma (MLE), Fisher (MLE), Gamma (MoLC), $\mathcal{K}$ (MoLC), and Fisher (MoLC). Besides, in most cases, the RMSE value relative to Fisher (MoLC) is smaller compared to other models (Appendix A, Table A.1-7).

Although texture modeling based on Gamma and $\mathcal{K}$ distributions for slightly heterogeneous (middle density) areas gives good results, this experimental research confirms that they fail to model these areas over different textures (for *shadows* over, polarimetry.parameter-$\sqrt{Span}$). The visual comparison results between these pdfs and the empirical data confirms this conclusion, as shown in Figure 8.1(a-d)

and Figure 8.2. The texture distribution can have an extended tail under certain conditions (heterogeneity texture areas with high density of different types of roofs like as gravel, concrete, slate and flat roofs), that the Weibull and Log-normal cannot describe and model, thus these models are not appropriate. Furthermore, a complex model with many parameters better fits data than a simple model with few ones (Weilbull and Log-normal) (Myung, 2002).

Mostly for texture parameters, only Fisher pdf manage to model train samples. This is the case for all train samples over $CON\text{-}\sqrt{T_{22\text{-}A}}$, and $CON\text{-}\sqrt{T_{33\text{-}A}}$-texture parameters. Our statistical analysis experiment confirms the existence of clear different statistical behaviors such as:

- *bright pixels* and *highways* using a depolarization $|S_{HV}|/|S_{HH}|$ -ratio,
- *highways* and *shadows* using the depolarization $|S_{VV}|/|S_{HV}|$-ratio ,
- *lawn* and *shadows* using $\sqrt{Span}$ -descriptor,
- *flat roofs* and *sloped roofs* using depolarization ratio- $|S_{VV}|/|S_{HV}|$.

Besides, $CON\text{-}\sqrt{T_{33\text{-}A}}$ for *sloped roofs*, and $CON\text{-}\sqrt{T_{22\text{-}A}}$ for *lawns* have different statistical behavior that requires further considerations as shown in Figure 8.1 and Figure 8.2. In general, these experimental results prove the pertinence of the Fisher pdf that can be retained for all features and all classes using MoLC. However, it is difficult to draw any robust fitness conclusions between empirical data and statistical distribution model through K-S distance due to its sensitivity to the size of sample. Furthermore, it calculates only the maximum difference between to sample data and statistical model.

To show the efficiency of Fisher pdf modeling over all real positive descriptors in optimal subset, we use Fisher pdf in the proposed Fisher-ML classifier which is detailed in Chapter 5. Moreover, to illustrate the efficiency of Fisher–ML classifier, we performed a comparison experiment of different classifiers such as SVM, G-ML.

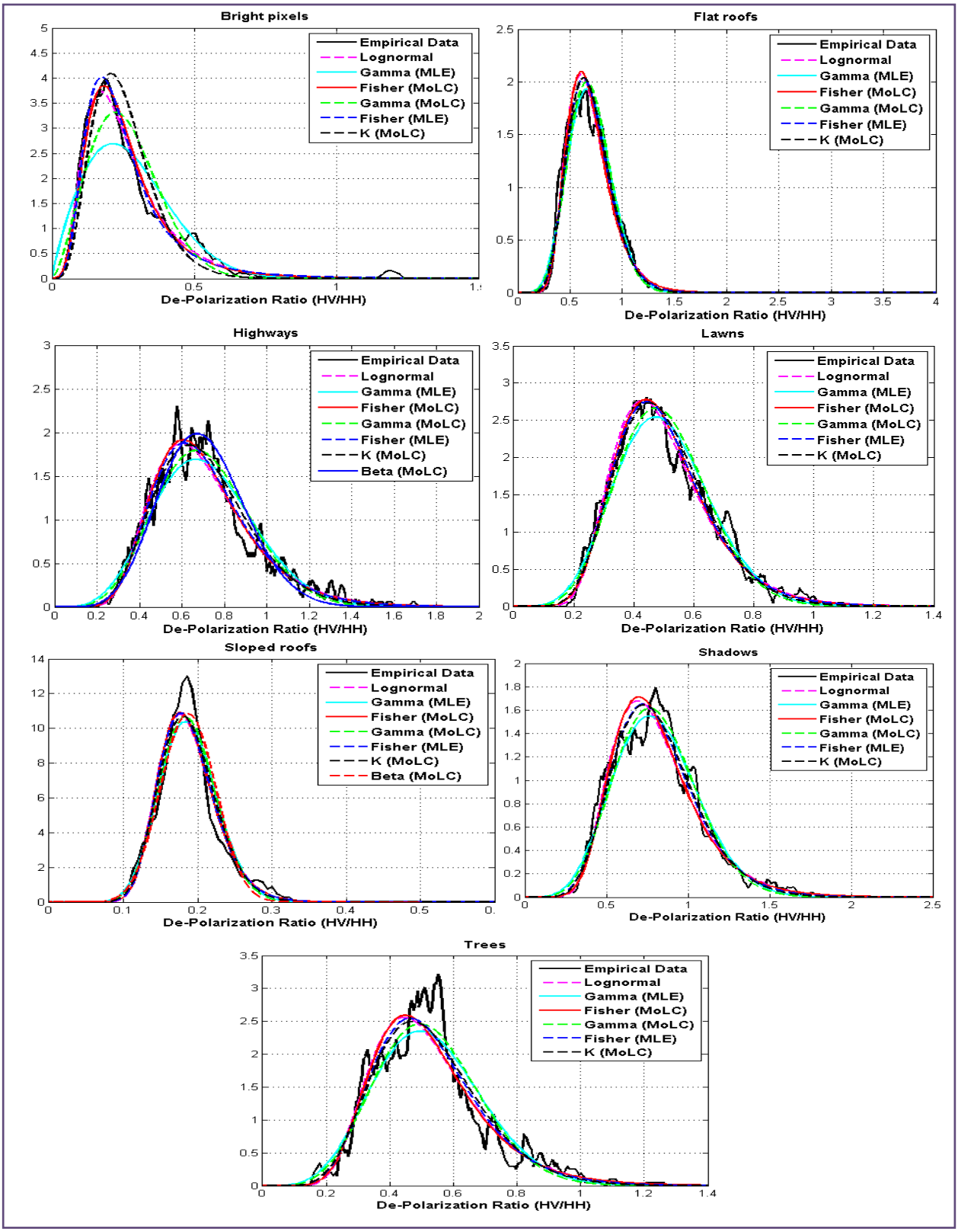


Figure 8. 1 (a): Empirical data and estimated pdfs using MLE and MoLC for different classes over De-Polarisation ratio (HV/HH) descriptor of SSM based on GA-MI through 58 real positive descriptors.

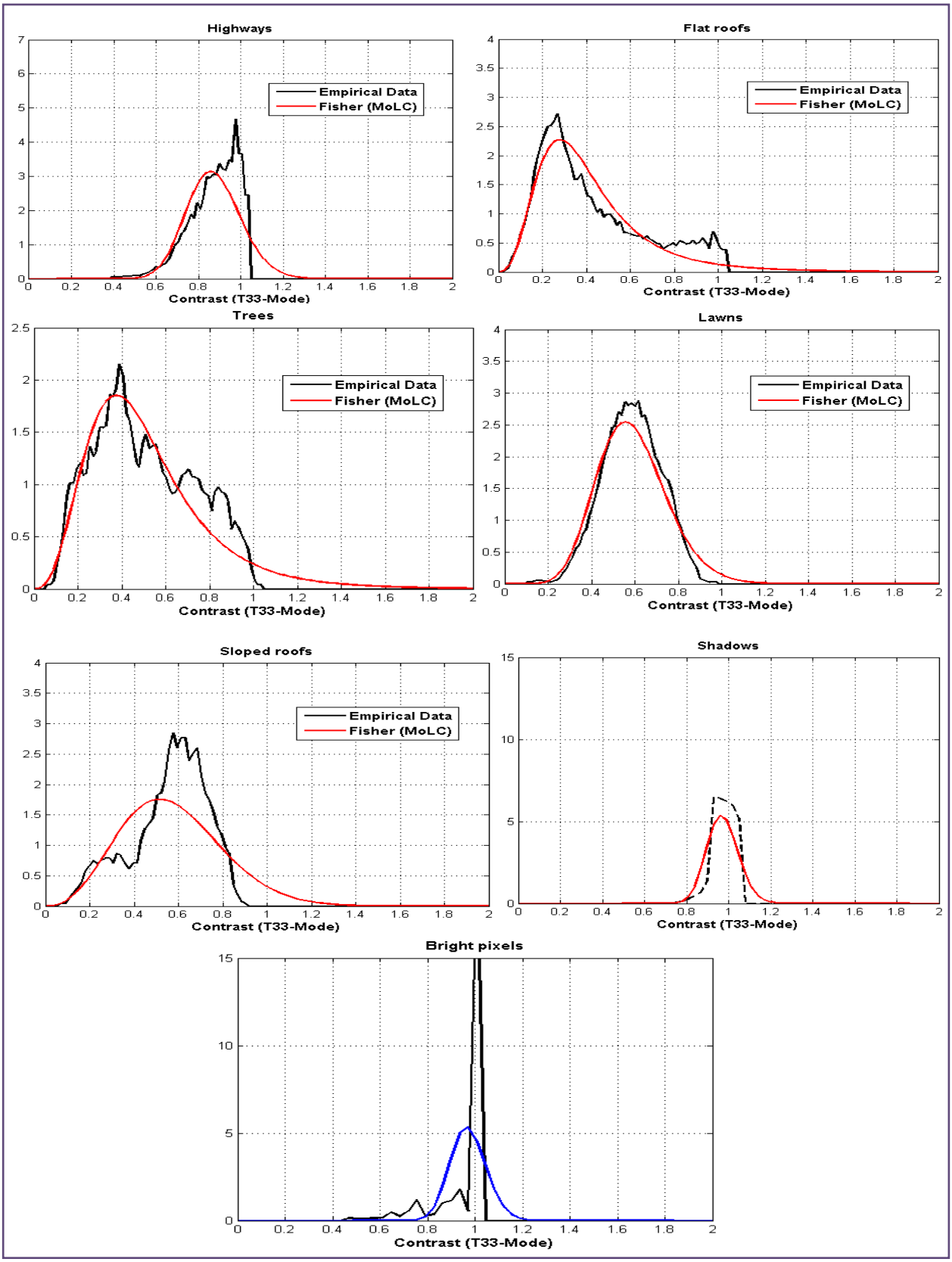


Figure 8.1 (b): Empirical data and estimated pdfs using MLE and MoLC for different classes over Contrast (T_{33}-Mode) descriptor of SSM based on GA-MI through 58 real positive descriptors.

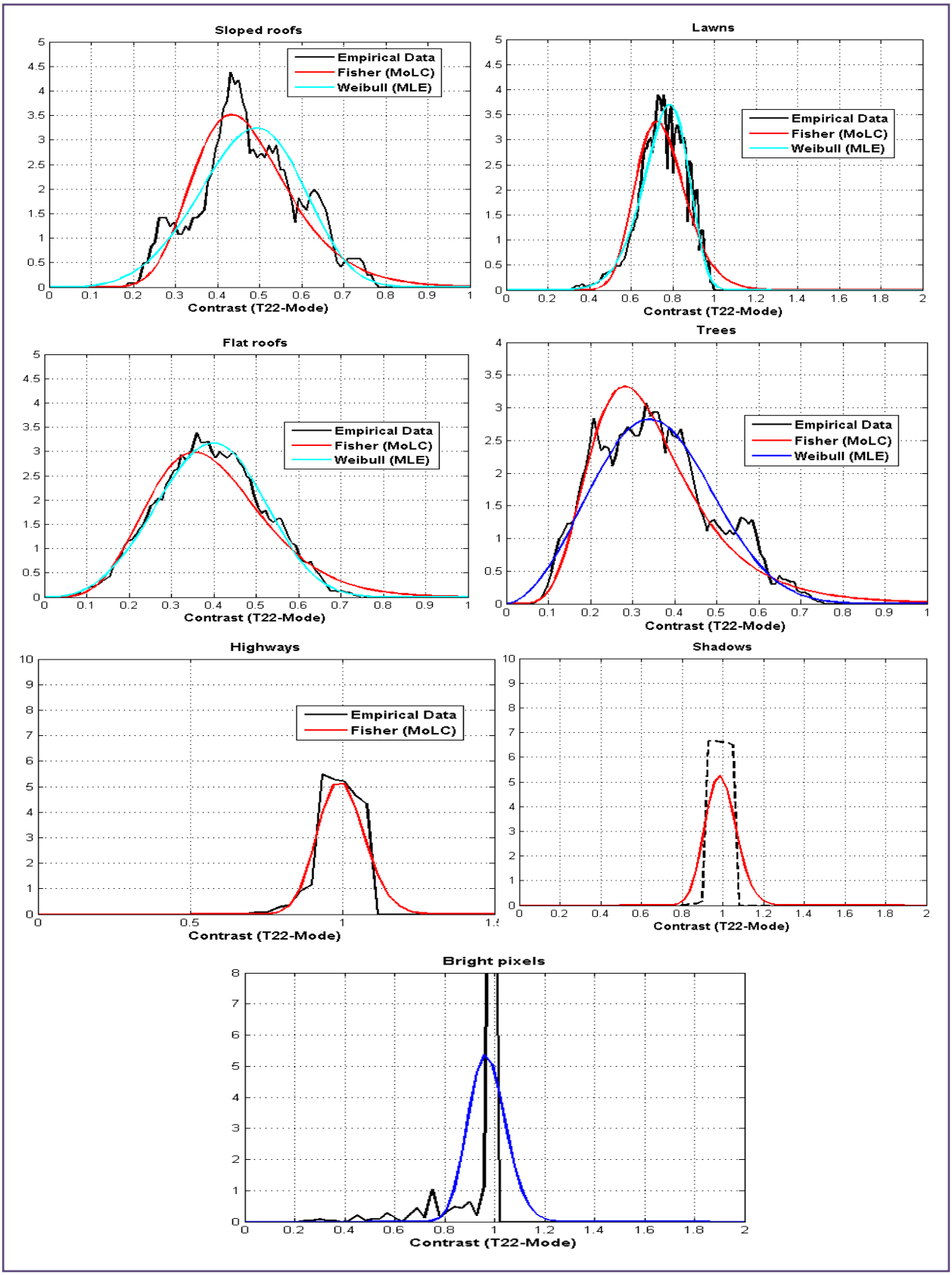


Figure 8.1 (c): Empirical data and estimated pdfs using MLE and MoLC for different classes over Contrast (T_{22}-Mode) descriptor of SSM based on GA-MI through 58 real positive descriptors.

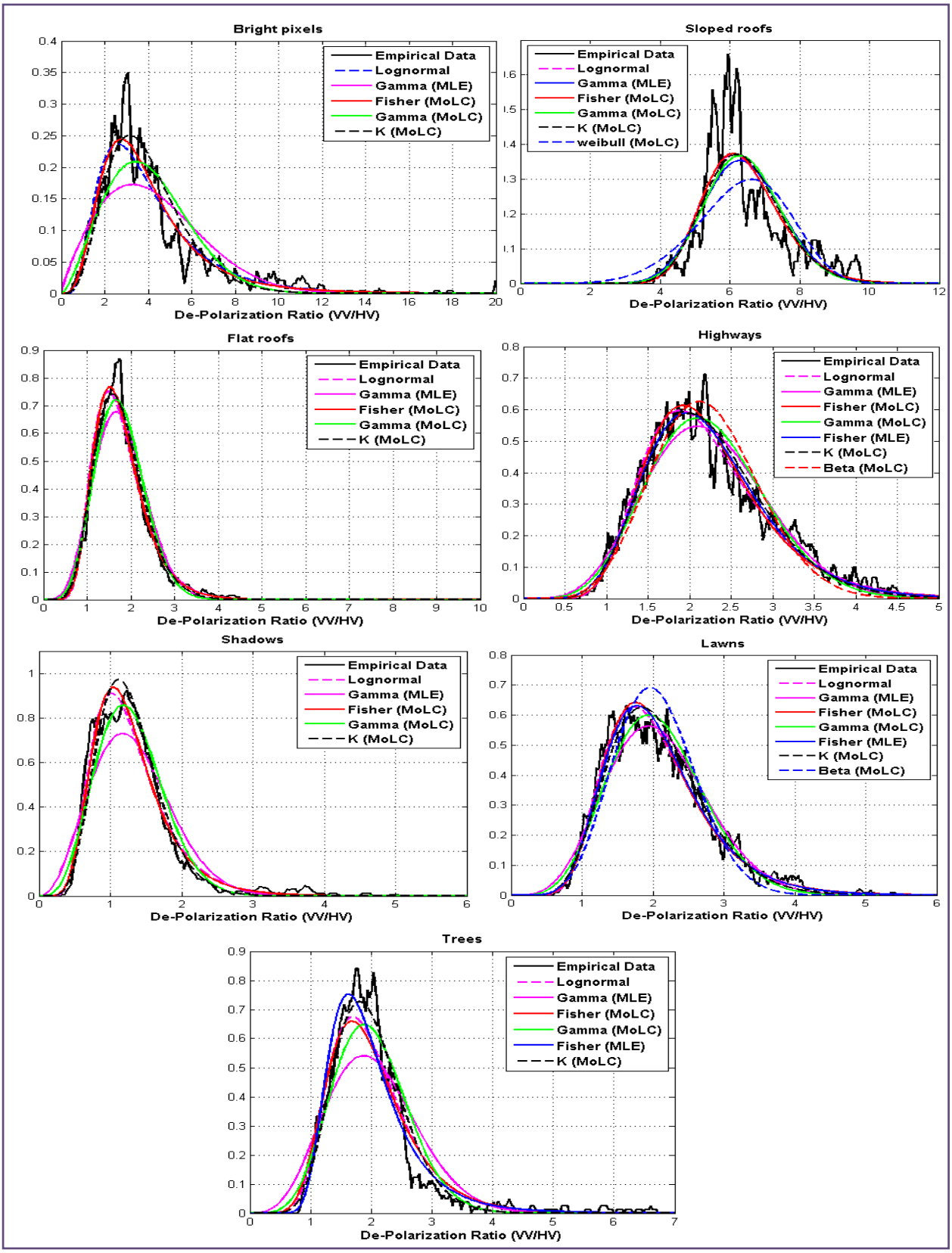


Figure 8.1 (d): Empirical data and estimated pdfs using MLE and MoLC for different classes over De-Polarisation ratio (VV/HV) descriptor of SSM based on GA-MI through 58 real positive descriptors.

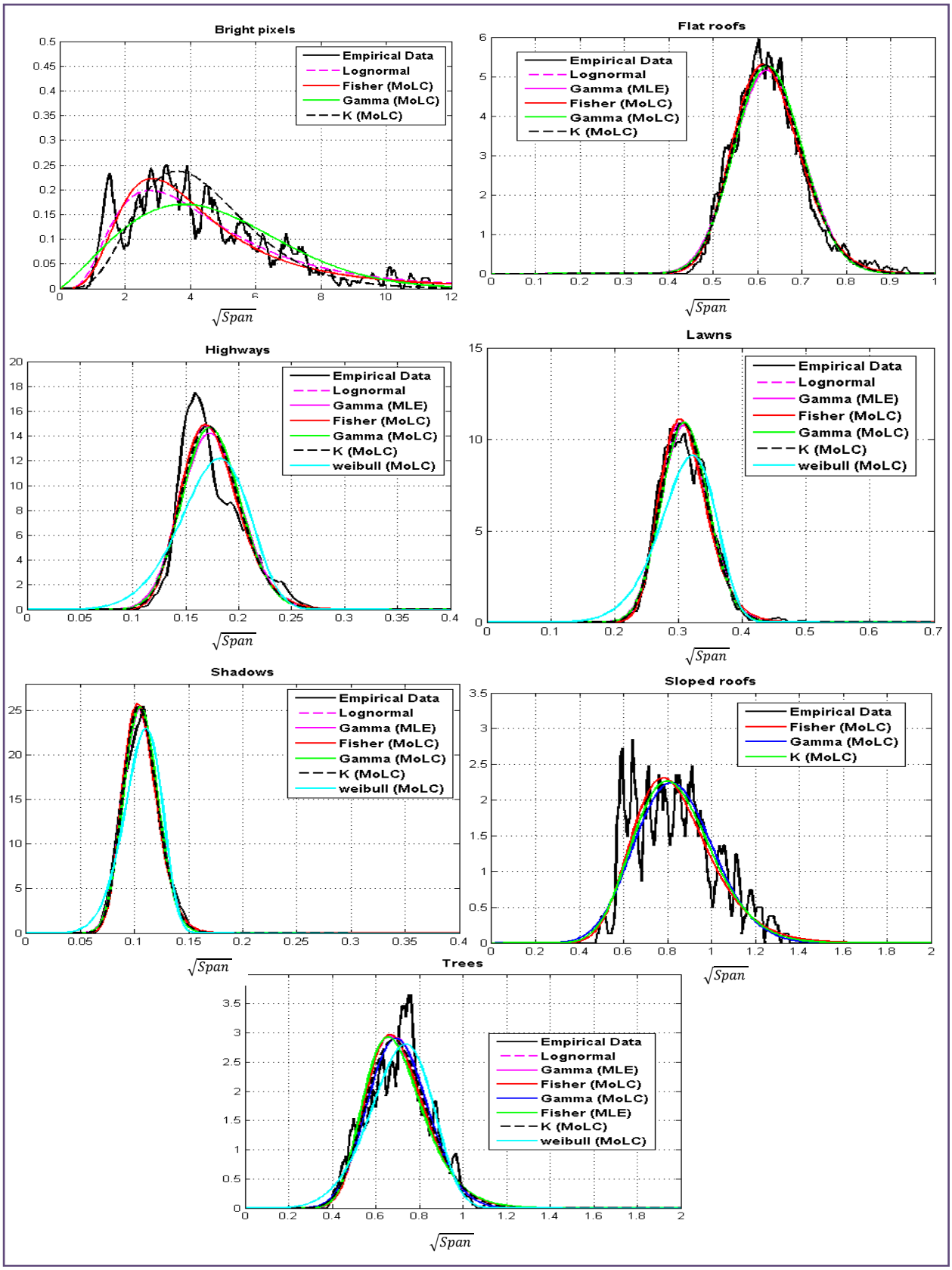


Figure 8.1 (e): Empirical data and estimated pdfs using MLE and MoLC for different classes over $\sqrt{Span}$ - descriptor of SSM based on GA-MI through 58 real positive descriptors.

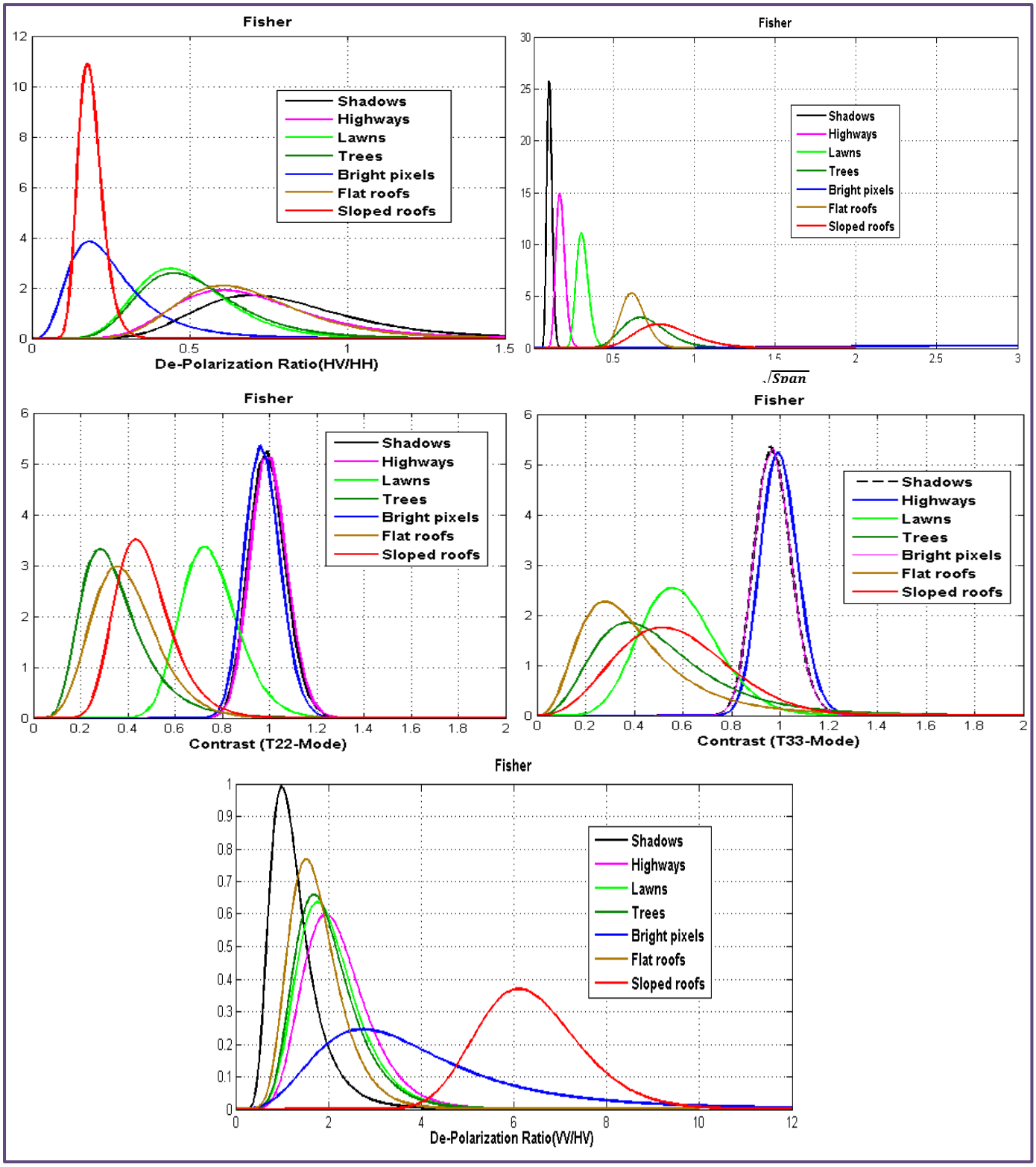


Figure 8. 2: Estimated Fisher distribution using MoLC for seven classes and optimal subset (Table 8.1) of SSM based on GA-MI through 58 real positive descriptors.

8.2. Classification experiment I

This section first describes the experimental result of the Fisher-ML classifier (F-ML) which is computed on optimal subset of real positive descriptors (Table 8.1). To illustrate the efficiency of F-ML classifier we made a comparison experiment of classification results based on three widely used classifiers, i.e., SVM, and G-ML (uni-variate). Moreover, to show the efficiency of building footprint map to improve classification result, an experiment of classification results based on these classifiers associated with the morphological component and real positive descriptors (Table 8.1) are presented.
The source code for Fisher-ML, uni-variate Gaussian-ML, and complex multi-variate Gaussian-ML classification have been written in MATLAB.

8.2.1. The ML classifier based on Fisher pdf

After obtaining the most appropriate features (Table 7.1) based on GA using MI (Chapter 3), Fisher-ML classifier is trained on selected features (Table 8.1). This classification method was detailed in Chapter 5, and Figure 4. 2. The procedure of classification consists in:

- Calculating Fisher pdf parameters ($\mu_{k,b}, L_{k,b}, M_{k,b}$) for each train data (*shadows*, *highways*, *lawns*, *trees*, *bright pixels*, *flat roofs*, and *sloped roofs*) over all real positive descriptors listed in Table 8.1, through Equation 4.23 and Equation 4.54.
- Calculating the ML for each pixel and class. Then, assigning each pixel to the class with maximum likelihood.
- Assessing the quality of classification using control samples.

The result of a classification is evaluated using confusion matrix, an overall accuracy, the Producer's Accuracy (PA), and the User's Accuracy (UA), and Kappa coefficent (Kappa Co.) as described in (Congalton, 1991). The confusion matrix shows the classification accuracy result by comparing the classification result (location and class in the classification image) with ground truth information or ROI (Region Of Interest; location and class of each ground truth pixel), errors of commission (percentage of extra pixels in class) and omission (percentage of pixels left out of class). The columns in Table 8.3 of the confusion matrix represent ground truth (actual class), while each row represents the obtained class. The number of pixels that are correctly classified is found on the diagonal of the matrix, known as Producer Accuracy: actually, the tables are showing the "percentage of pixels". The Overall Accuracy is calculated by the sum of correctly classified pixels divided by the total number of all ground truth pixels. The User Accuracy is a measurement of the probability that a pixel is class A given that the classifier has labeled the pixel into class A. Errors of commission represent pixels that belong to another class that are labeled as belonging to the class of interest (the rows of the confusion matrix). However, when the classification technique has failed to classify ground truth pixels into the proper class, the result presents errors of omission (the columns of the confusion

matrix) (RSI, 2003). The confusion matrix in percent for F-ML method over 5 real positive descriptors (Table 8.1) is given in Table 8.2. The overall accuracy of F-ML is 84.1%. Morover, Figure 8.3 shows the result of classification for G-ML.

Table 8. 2: Confusion Matrix (in percent) for F-ML over optimal real positive descriptors of SSM based on GA-MI from 58 descriptors.

G.truth Class	*shadows*	*highways*	*lawns*	*trees*	*bright pixels*	*flat roofs*	*sloped roofs*
Shadows	**86.2**	**22.2**	**0.0**	**0.0**	**0**	**0**	**0**
highways	**13.8**	**75.2**	**4.3**	**0.0**	**0**	**0**	**0**
lawns	**0.0**	**2.6**	**94.5**	**3.5**	**0**	**8.2**	**0**
trees	**0**	**0**	**0.8**	**56.4**	**0**	**41.3**	**17.7**
Bright pixels	**0**	**0**	**0**	**0.0**	**96.0**	**0**	**0**
Flat roofs	**0**	**0**	**0.4**	**39.3**	**3.9**	**50.5**	**0.6**
Sloped	**0**	**0**	**0**	**0.8**	**0**	**0**	**81.7**
PA	**86.2**	**75.2**	**94.5**	**56.4**	**96.0**	**50.5**	**81.7**
UA	**56.7**	**82.4**	**97.6**	**63.8**	**100**	**40.2**	**91.1**

Table 8. 3: Confusion Matrix (in percent) for G-ML over optimal real positive descriptors of SSM based on GA-MI from 58 descriptors.

G. truth ---------- Class	*shadows*	*highways*	*lawns*	*trees*	*bright pixels*	*flat roofs*	*sloped roofs*
shadows	**80.8**	**22.0**	**0.0**	**0**	**0**	**0**	**0**
highways	**19.2**	**72.8**	**3.2**	**0**	**0**	**0**	**0**
lawns	**0**	**4.3**	**94.8**	**2.5**	**0**	**4.9**	**0**
trees	**0**	**0**	**1.1**	**58.5**	**0**	**44.4**	**16.1**
bright pixels	**0**	**0**	**0.5**	**0**	**100**	**0**	**1.2**
flat roofs	**0**	**0**	**0.4**	**38.7**	**0**	**50.4**	**1.2**
sloped	**0**	**0.9**	**0**	**0.3**	**0**	**0.3**	**81.5**
PA	**80.8**	**72.8**	**94.8**	**58.5**	**100**	**50.4**	**81.5**
UA	**55.4**	**82.9**	**97.4**	**62.2**	**100**	**38.3**	**88.4**

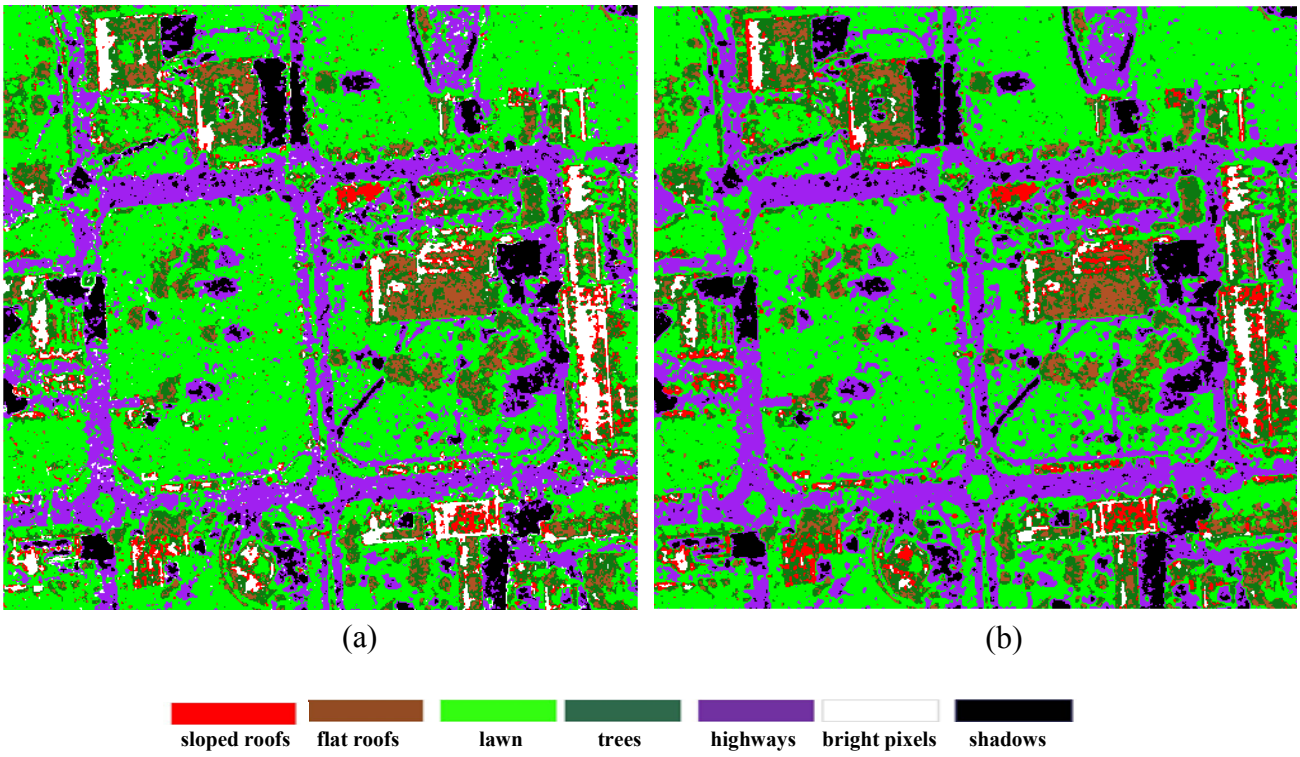


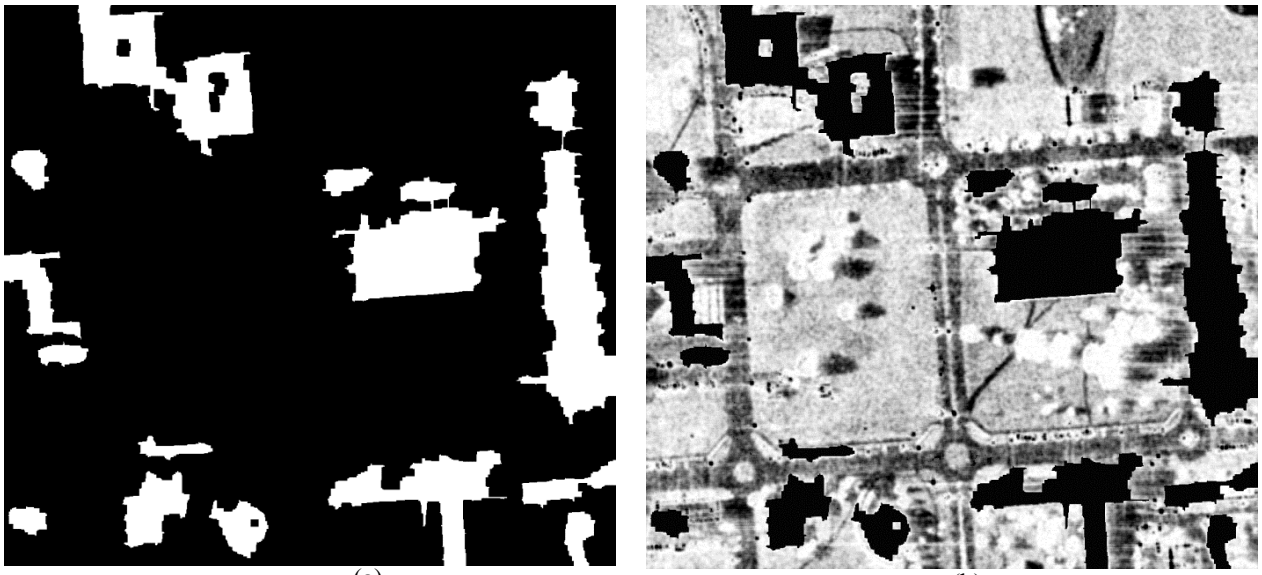
Figure 8. 3: (a) G-ML and (b) F-ML. Classification over optimal subset (Table 8.1) of SSM based on GA-MI through 58 positive real descriptors (Table 7.1)

Fisher-ML classifier coupled with building map as denoted by F-ML* is trained on optimal subset (Table8.1) of MSS based on GA-MI of 58real positive descriptors and buiding map. This classification method was detailed in Chapter 5, Figure 5. 3. The procedure of classification based on F-ML* consits in:

- Calculating Fisher pdf parameters ($\mu_{k,b}, L_{k,b}, M_{k,b}$) for each train data over descriptors through Equation 4.23 and Equation 4.54.
- Calculating the ML for each pixel and class.
- Producing building map from semi-automatic interface framework (section 5.5, Figure 5.3)
- Using this morphological information as a constraint during F-ML classification procedure. Here, C_1 and C_2 are *flat roof* and *trees* classes
- Assigning each pixel to class with maximum likelihood.
- Assessing the quality of classification result using control samples.

The building map and the membership of the pixels to the class *trees* respect to building map are shown in Figure 8.4 (a) and Figure 8.4 (b) respectively. The F-ML* accuracy classification result is summarized in the corresponding confusion matrix (Table 8.4). Finally, the result of F-ML* classification is shown in Figure 8.5. This approach provides a good classification of the two surface types as well as for the other classes and is able to further improve the classification result from 84.1% to 86.20% overall accuracy. The errors of commission between *trees* and *flat roofs*

have completely disappeared (Table 8.4, 5th row). However, the errors of omission between *trees* and *flat roofs* still exist (Table 8.4, 5th column).



(a) (b)

Figure 8. 4: (a) Building map (b) Membership of the cells to the class *trees* respect to Building map (values between 0 and 1)

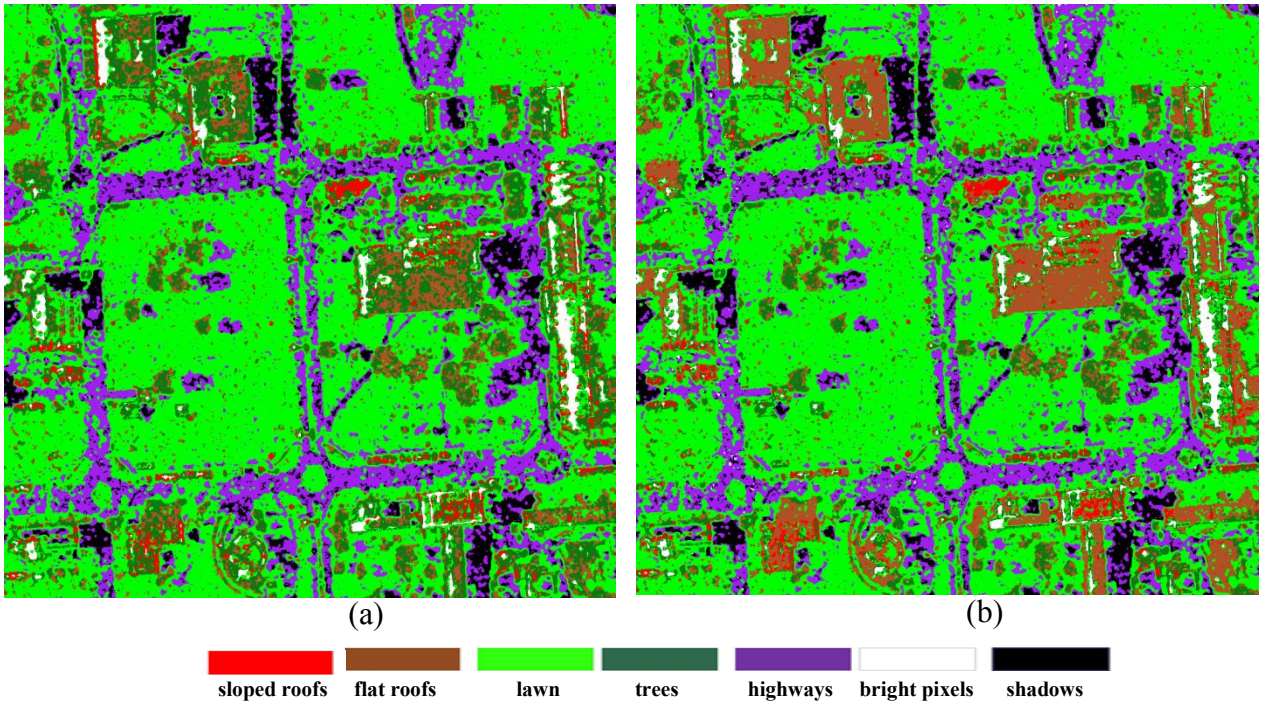


Figure 8. 5: Classification result: (a) F-ML and (b) F-ML*.* stands for a classification with optimal subset of SSM of GA-MI from 58 descriptors (Table 8.1) coupled with morphological map.

Table 8. 4: Confusion Matrix (in percent) for F-ML*.* stands for a classification with optimal subset of SSM of GA-MI from 58 descriptors (Table 8.1) coupled with morphological map.

G. truth ---------- Class	*shadows*	*highways*	*lawns*	*trees*	*bright pixels*	*flat roofs*	*sloped*
shadows	86	22.2	0	0	0	0	0
highways	13.8	75.2	4.3	0	0	0	0
lawns	0	2.5	94.5	3.5	0	9.1	0
trees	0	0	0.8	56.4	3.0	0	0
bright pixels	0	0	0	0	94.6	0	0
flat roofs	0	0	0.4	39.2	0	90.1	11.0
sloped	0	0	0	0.9	2.4	0.8	89.0
PA	86	75.2	94.5	56.4	94.6	90.1	89.0
UA	56.7	82.4	97.5	90.7	100	53.3	91.8

8.2.2. SVM classification coupled with morphological descriptor

The SVM classifier is trained on feature selection subset (Table 8.1) coupled or not with buiding footprint, which are defined as SVM^{1}, and SVM^{1*} respectively. The confusion matrices are given in Table 8.5 and Table 8.6 respectively.

The SVM^{1}* classifier provides a good classification of the two surface types as well as for the other classes and is able to further improve the classification result from 80.3% to 88.5% overall accuracy The errors of commission between *trees* and different types of roofs have completely disappeared (Table 8.6, 5th row). The errors of omission between *trees* and other classes is less than 0.12%.

The classification result in Figure 8.6 shows that the class *trees* is completely discriminated from two types of roofs. Moreover, the Producer's Accuracy of class *trees* has increased from 35.6% to 87.9% (Table 8.5 and Table 8.6). However, the errors of commission between *flat roofs* and *sloped roofs* still exist (Table 8.6, 7th row).

We used from ENVI -SVM algorithm, and LIBSVM package (Hsu & Lin, 2002), which supports both 2-class and multiclass classification with SVM and is available at[3].

[3] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

Table 8. 5: Confusion Matrix (in percent) for SVM with optimal subset of SSM of GA-MI from 58 descriptors (Table 8.1)

G. truth ---------- Class	***shadows***	***highways***	***lawns***	***trees***	***bright pixels***	***flat roofs***	***sloped roofs***
shadows	**76.3**	**22.4**	**0**	**0**	**0**	**0**	**0**
highways	**23.7**	**71.0**	**3.4**	**0**	**0.6**	**0**	**0**
lawns	**0**	**6.6**	**92.2**	**9.5**	**2**	**10**	**2.6**
trees	**0**	**0**	**0.6**	**35.6**	**0**	**24.4**	**25.1**
bright pixels	**0**	**0**	**0**	**0**	**96.1**	**0**	**0.3**
flat roofs	**0**	**0**	**3.5**	**54.6**	**0**	**65.6**	**3.6**
sloped	**0**	**0**	**0.3**	**0.3**	**1.3**	**0**	**68.4**
PA	**76.3**	**71.0**	**92.2**	**35.6**	**96.1**	**65.6**	**68.4**
UA	**57.5**	**77.9**	**95.1**	**60.8**	**99.8**	**32.2**	**87.2**

Table 8. 6: Confusion Matrix (in percent) for SVM[1]*.[1]* stands for a classification with optimal subset of SSM of GA-MI from 58 real positive descriptors (Table 8.1).coupled with morphological map.

G. truth ---------- Class	***shadows***	***highways***	***lawns***	***trees***	***bright pixels***	***flat roofs***	***sloped roofs***
shadows	**75.8**	**21.3**	**0**	**0**	**0**	**0**	**0**
highways	**24.2**	**71.8**	**3.3**	**0**	**3.2**	**0**	**0**
lawns	**0**	**6.7**	**94.5**	**12.1**	**2.3**	**0**	**0**
trees	**0**	**0.1**	**2.2**	**87.9**	**0.1**	**0**	**0**
bright pixels	**0**	**0**	**0**	**0**	**94.1**	**0**	**0.3**
flat roofs	**0**	**0**	**0**	**0**	**0**	**99.8**	**16.5**
sloped	**0**	**0.1**	**0**	**0**	**0.3**	**0.2**	**83.2**
PA	**75.8**	**71.8**	**94.5**	**87.9**	**94.1**	**99.8**	**83.2**
UA	**58.3**	**77.8**	**95.7**	**87.0.**	**99.7**	**93.6**	**98.8**

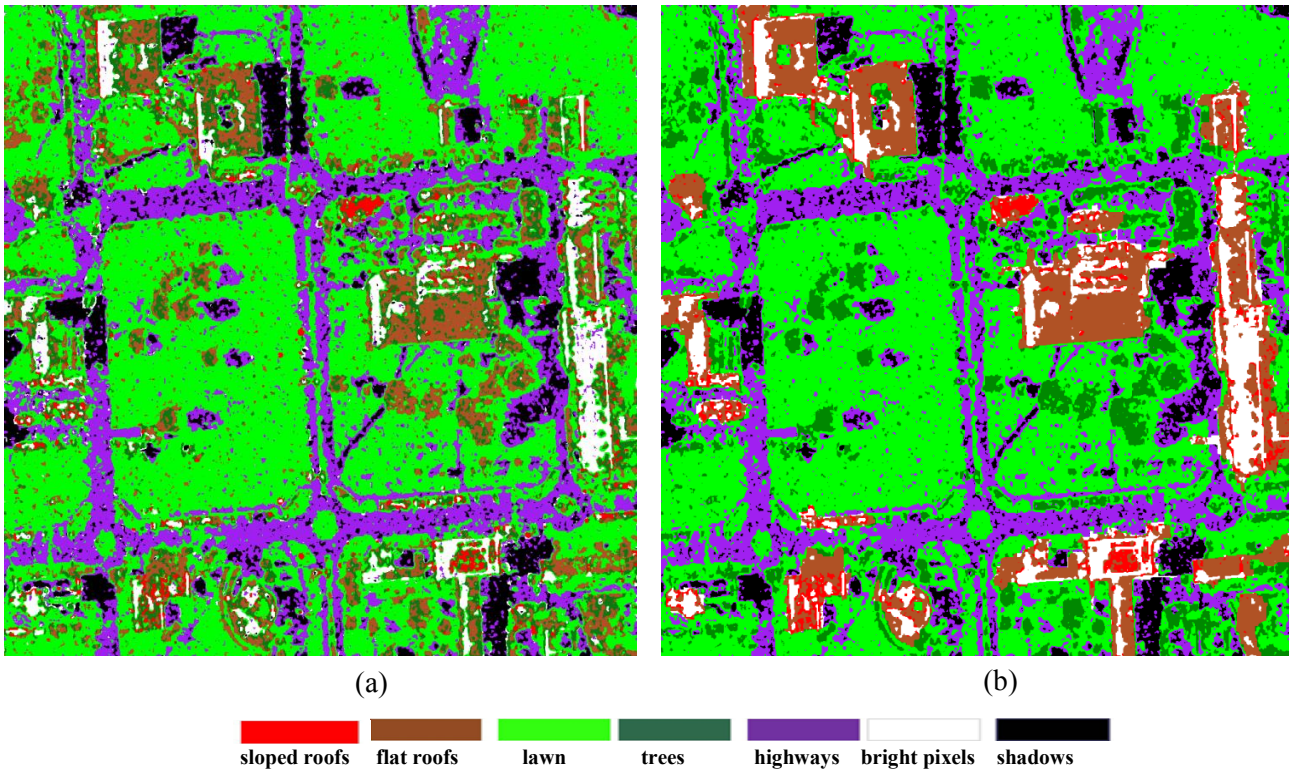


Figure 8. 6: Classification results with (a) SVM[1], (b) SVM[1]*.[1]* stands for a classification with optimal subset of SSM of GA-MI from 58 real positive descriptors (Table 8.1) coupled with morphological map.

8.2.3. Comparison of results classification

To show the efficiency of F-ML classification algorithm, we compare the classification on optimal subset (consist in 5 real positive descriptors) with widely common classification methods such as G-ML (uni-variant Gauusian maximum likelihood classifier), SVM Table 8.7).

Table 8.7 gives the detailed comparison results (Table 8.7, 2th row, 3th row, and 4^{h} row). As shown in Table 8.7, the F-ML overall accuracy is better than the other classifiers and Producer's Accuracy (PA) of *shadows* and *highways* classes shows more improvement than in the two other common classifiers.

Table 8. 7: Comparison of classification accuracies with different classifiers (in %): producer's accuracy for each class, Kappa coefficient and overall accuracyover optimal subset of SSM of 58 positive real descriptors (Table 8.1). [1] stands for a classification with optimal subset of SSM from 58 descriptors (Table 8.1).* stands for a classification with selected polarimetric descriptors and morphological map

Method	*Shadows*	*highways*	*lawns*	*trees*	*bright pixels*	*flat roofs*	*sloped roofs*	Kappa Co.	OA
G-ML	**80.8**	**72.8**	**94.8**	**58.5**	**100**	**50.4**	**81.5**	**0.75**	**83.3**
SVM[1]	**76.3**	**71.0**	**92.2**	**35.6**	**96.1**	**65.6**	**68.4**	**0.70**	**80.3**
F-ML	**86.0**	**75.2**	**94.5**	**56.4**	**96**	**50.5**	**81.7**	**0.76**	**84.1**
F-ML*	**86.0**	**72.2**	**94.5**	**56.4**	**94.6**	**90.1**	**89.0**	**0.79**	**86.2**
SVM[1]*	**76.0**	**71.8**	**94.5**	**88.0**	**94.1**	**99.8**	**83.2**	**0.82**	**88.5**

8.3. Classification experiment II

In the second experiment, the efficiency of SSM (subset selection method) based on GA-MI through 127 descriptors is considered. For this purpose, first, the SVM classification results on the optimal subset (5 descriptors, Table 7.2) of SSM through 127 descriptors are compared with the G-ML^C classification results, without and with the morphological component.

Secondly, we make a comparison experiment of SVM classifier over 3 to 10 features subset (Table 7.2), with SVM classifier over different SAR target decomposition components, including the Pauli, Krogager, Freeman, Four-Component, Barnes, Holm, Huynen and Cloude decompositions.

8.3.1. The Complex G-ML classification coupled or not with morphological map

In this classification, we use the complex multi-variate Gaussian distribution (Equation 5.9) to model the coherency matrix (Equation 2.11). The procedure of muti-variate complex G-ML classification coupled with building footprints which is notified by G-ML^C* is also discussed.

The procedure of classification consists in:

- Calculating the membership for each class and each pixel using Equation 4.10.
- Producing building map from semi-automatic interface framework (section 7.2.2)
- Using this morphological information as a constraint during G-ML^C classification procedure: Here, C_1 and C_2 are *flat roof* and *trees* classes respectively.
- Assigning each pixel to class with maximum likelihood.

- Calculating the accuracy of classification result using control samples. The G-ML^{C}* accuracy classification result is summarized in the corresponding confusion matrix (Table 8.8). Figure 8.7 shows the classification results between G-ML^{C} and G-ML^{C}*. According to Table 8.8, this approach with overall accuracy 84.51% provides good approximations for *shadows*, *lawns*, *bright pixels*, *flat roofs*, and *sloped roofs.* The errors of commission between *trees* and *flat roofs* (Table 8.9, 5th row). However, the errors of omission between *trees* and *flat roofs* still exist (Table 8.8, 5th column). So the ambiguties between *flat roofs* and *trees* are considerable.

The source code for Fisher-ML, uni-variate Gaussian-ML, and complex multi-variate Gaussian-ML classification have been written in MATLAB.

Table 8. 8: Confusion Matrix (in percent) for G-ML^{C}*. ^{C}* stands for the complex Gaussian multi-variate ML approach based on the coherency matrix following with morphological map.

G. truth ---------- Class	***shadows***	***highways***	***lawns***	***trees***	***bright pixels***	***flat roofs***	***sloped roofs***
shadows	**83.1**	19	0	0	0	0	0
highways	16.9	**79**	3	0	0	0	0
lawns	0	2	**95.8**	3.4	0	8.2	0.2
trees	0	0	0	**27**	0.2	0	0
bright pixels	0	0	0	0	**99.8**	0	2.6
flat roofs	0	0	1.2	69.6	0	**91.8**	28.9
sloped	0	0	0	0	0	0	**68.3**
PA	**83.1**	**79**	**95.8**	**27**	**99.8**	**91.8**	**68.3**
UA	40.4	85	97.8	100	97.4	38	96.4

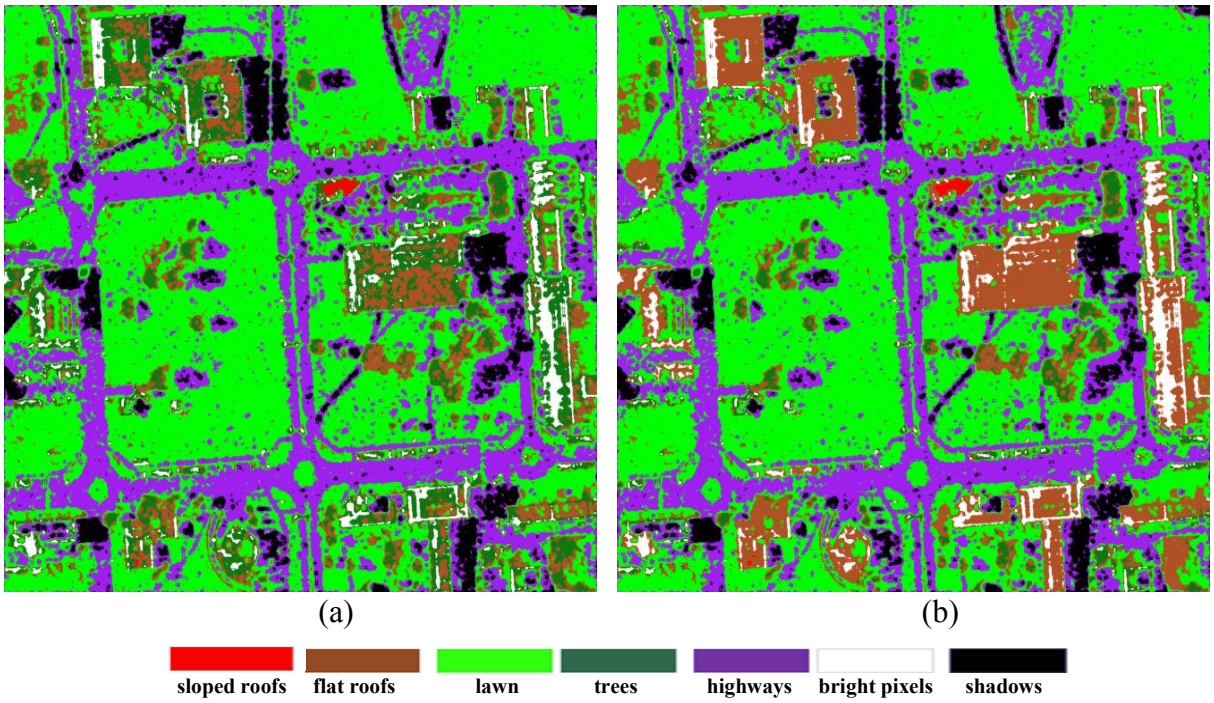


Figure 8. 7: Classification results with (a) G-ML^{C} and (b) G-ML^{C}*. ^{C} stands for the complex Gaussian multi-variate ML approach based on the coherency matrix. ^{C}* stands for the complex Gaussian multi-variate ML approach based on the coherency matrix following with morphological map.

8.3.2. The SVM classification on complete descriptors

To ilustrate the efficiency of SSM based on GA-MI, we follow the common experiment setup for a SVM classification on optimal subset (Table 8.9) of different types of polarimetry descriptors (Table 7.2). Table 8.10, and Figure 8.8 show the confusion matrix and classification result respectively. Moreover, the confusion matrix of SVM classifier is obtained on this optimal subset (Table 8.9) coupled with buiding footprint, which are defined as SVM^{2*}and results of classification are given in Table 8.11 and Figure 8.8 respectively.

Table 8. 9: List of descriptors in optimal subset of SSM based on GA-MI through the 127 descriptors

# Of Des.	**1**	**2**	**3**	**4**	**5**
Des.	*Shanon-Entropy-I*	$H(1\text{-}A)$	$T_{33\text{-}1}$	$T_{12\text{-phase}}$	$H/A/Alpha\text{-}T_{11}$

The SVM^{2}* classifier provides a good classification of the two surface types as well as for the other classes and is able to further improve the classification result from 83.8% to 90.93% overall accuracy. The errors of commission between *trees* and different types of roofs completely disappeared (Table 8.11, 5th row). The errors of omission between *trees* and other classes is less than 3.1%.

Moreover, the Producer's Accuracy of class *trees* has increased from 31.2% to 96.91% (Table 8.10 and Table 8.11). The error of commission between *flat roofs* and *sloped roofs* has decreased (Table 8.11, 7th row). The classification result in Figure 8.8 shows that the class *trees* is completely discriminated from different types of roofs.

Table 8. 10: Confusion Matrix (in percent) for SVM^2. ^{2} stands for a classification with optimal subset (Table 8.10) of SSM based on GA-MI from 127 descriptors (Table 7.2).

G. truth ---------- Class	***shadows***	***highways***	***lawns***	***trees***	***bright pixels***	***flat roofs***	***sloped roofs***
shadows	**88.5**	17.4	0	0	0	0	0
highways	11.5	**81.3**	5.2	0	0	0	0
lawns	0	1.3	**93.7**	4.3	0	8.8	1.2
trees	0	0	0.1	**31**	0	21.2	19.3
bright pixels	0	0	0	1.0	**99.8**	0.2	3.6
flat roofs	0	0	1.0	63.7	0.2	69.8	3.9
sloped	0	0	0	0	0	0	**72.0**
PA	**88.5**	**81.3**	**93.7**	**31**	**99.8**	**69.8**	**72.0**
UA	67	79.5	98	65	93.8	35.5	96.8

Table 8. 11: Confusion Matrix (in percent) for SVM^2*. ^{2}* stands for a classification with optimal subset (Table 8.9)of SSM based on GA-MI from 127 descriptors (Table 7.2) coupled with morphological map.

G. truth ---------- Class	***shadows***	***highways***	***lawns***	***trees***	***bright pixels***	***flat roofs***	***sloped roofs***
shadows	**87.8**	17.9	0	0	0	0	0
highways	12.2	**80.8**	5.4	0	0	0	0
lawns	0	1.3	**93.1**	3.1	0	0.0	0.0
trees	0	0	1.5	**96.9**	5.3	0	0
bright pixels	0	0	0	0	**93.6**	0.6	6.9
flat roofs	0	0	1.2	0.0	1.1	**99.3**	14.3
sloped	0	0	0	0	0	0.1	**78.8**
PA	**87.8**	**80.8**	**93.1**	**96.9**	93.6	99.3	**78.8**
UA	66.2	78.9	99.0	90.1	97.4	94.0	99.6

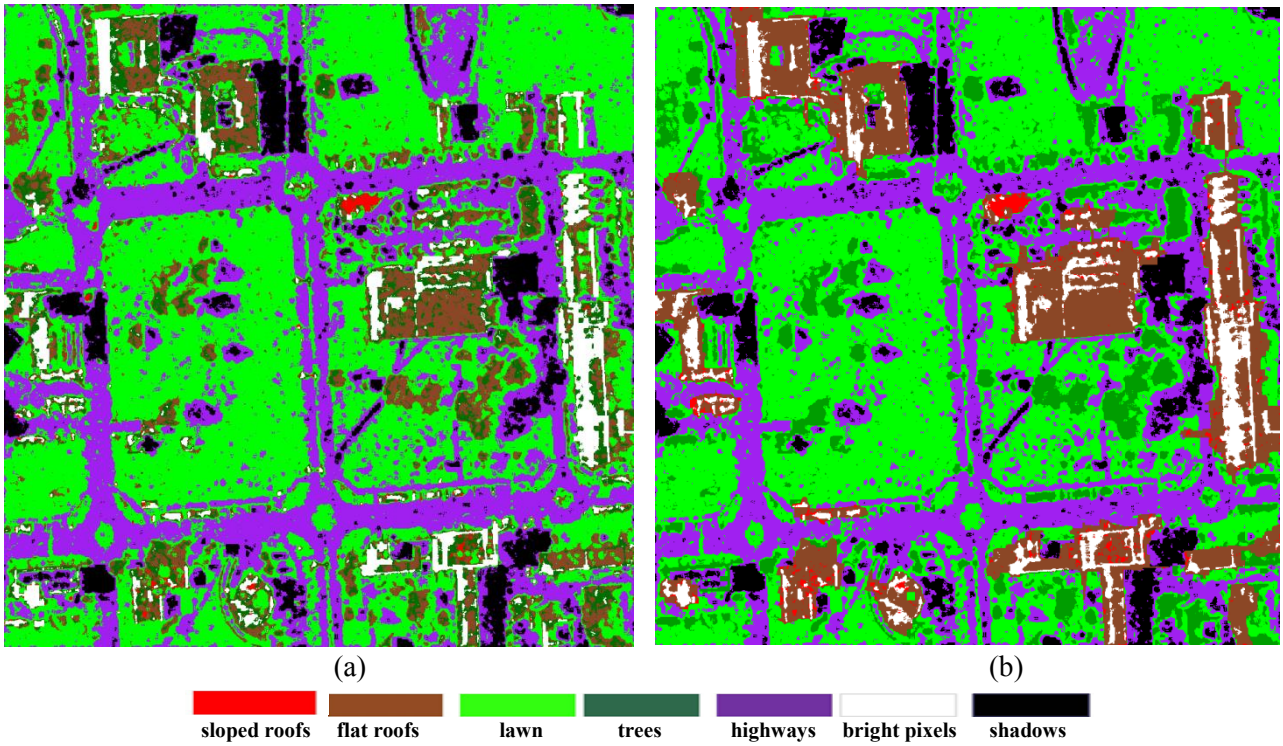


Figure 8. 8: Classification results with (a) SVM^2 and (b) SVM^2*.^{2} stands for a classification with optimal subset (Table 8.10) of SSM based on GA-MI from 127 descriptors (Table 7.2). ^{2}* stands for a classification with optimal subset (Table 8.10) of SSM based on GA-MI from 127 descriptors (Table 7.2) coupled with morphological map.

8.3.3. Comparison and results classification

The comparison experiment between $SVM^{2,}$ SVM ^{2}*, G-ML^C, and G-ML^C* is summarized in Table 8.12.

The result of SVM^2* classifier (over subset-5 and using the morphological map) and G-ML^C* shows a decrease of the ambiguities between roofs and *trees*. SVM^2*, with overall accuracy of 90.9%, is the best classifier.

Besides, the class *trees* with producer's accuracy of 96.91% is completely discriminated from *flat roofs*, *sloped roofs* with 99.3%, and 78.8% producer's accuracies respectively (Table 8.12). In fact, SVM^2* result shows the efficiency of GA-MI feature selection performance, morphological building footprints and the power of non parametric and non linear SVM classification. However, the classification accuracy of G-ML^C with 9 descriptors (coherency matrix) is 84% and 0.1% higher than SVM classification of optimal subset (5 descriptors) of SSM based on GA-MI.

To illustrate the efficiency of each subset of SSM based on GA-MI, we made a comparison experiment of SAR target decompositions based on different scattering mechanisms, including the

Pauli, Krogager, (based on the backscattering matrix and full resolution coherent technique), Freeman, VanZyl, Yamaguchi three and four-Component, Barnes, Holm, Huynen and the Cloude decompositions (following with Span component). In this comparison experiment, number of decomposition components from different decomposition methods are equal with the number of descriptors in different subsets from SSM based on GA-MI which are processed by SVM classifier. Table 8.13 gives the detailed comparison results of target decomposition and different subsets of SSM based on GA-MI.

From Table 8.13, the classification results over three descriptors show that, the Yamaguchi III, Cloude and subset-3 of SSM with overall accuracy 83.2%, 82.8%, and 82.6% are better than other target decompositions such as the Freeman, Holm, Barnes, Krogager, Pauli, Touzi and Van Zyl.

In addition, the SVM classification result from subset-3 of SSM leads to the best producer's accuracy for *flat roofs* and *sloped roofs* compared to the other decompositions with 3 components. However, HolmII decomposition component is preferable for *highways*.

In comparison, the classification results over four descriptors show the subset-4 of SSM has the same performance compare with Yamaguchi III (three decomposition components+ Span) with overall accuracy 83.2%.

In addition, we can also find that the SVM classification result of subset-5 of SSM shows better results than the Freeman II+III. Moreover, from Table 8.13, it is obvious that the optimal subset of SSM refers to subset-5 with an overall accuracy of 83.9%. In addition, the classification accuracies for all subset of SSM are higher than 82.5%. Finally, from Table 8.13, the main drawback is around class *trees* which is easily misclassified with two types of roof in urban area.

Table 8. 12: Comparison of classification accuracies with different classifiers (in %): G-ML^{C} is the complex Gaussian multi-variate ML approach based on the coherency matrix. ^{C}* stands for the complex Gaussian multi-variate ML approach based on the coherency matrix following with morphological map. ^{2} stands for a classification with optimal subset (Table 8.10) of SSM based on GA-MI through 127 descriptors (Table 7.2).

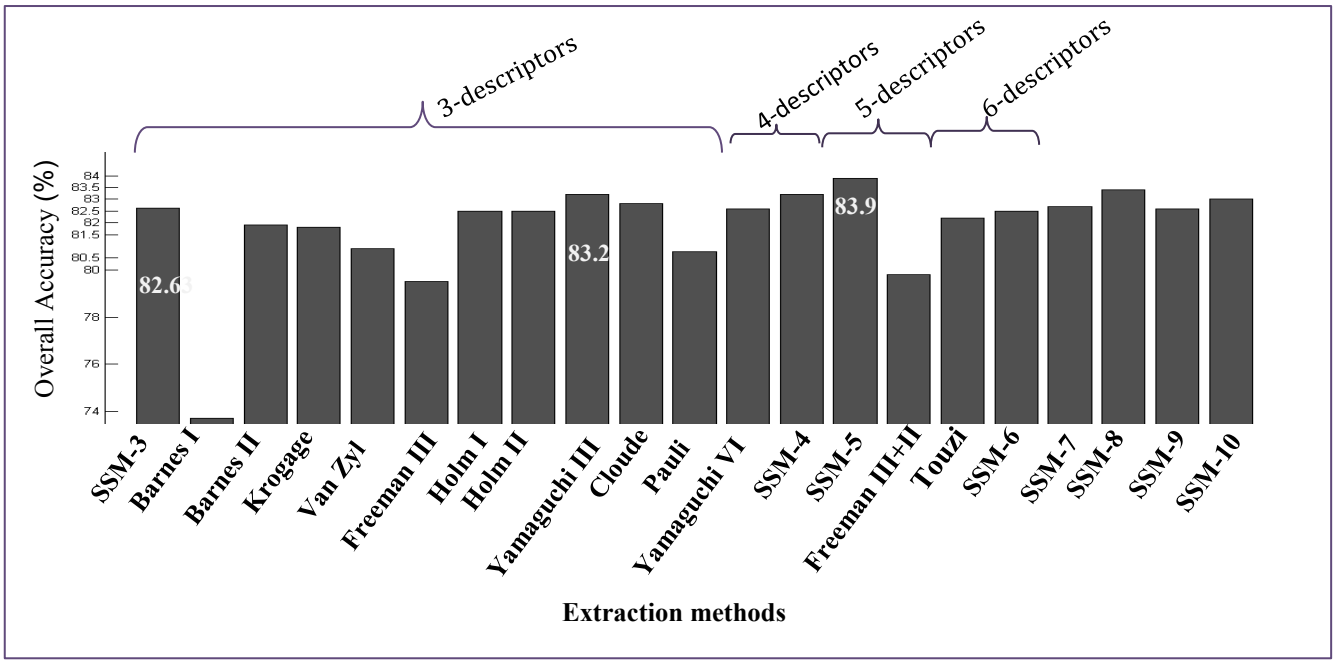
* stands for a classification with selected polarimetric descriptors and morphological map 127

Method	***Shadows***	***highways***	***lawns***	***trees***	***bright pixels***	***flat roofs***	***sloped roofs***	Kappa Co.	OA
G-ML^{C}	**83.1**	**78.0**	**96.0**	**43.0**	**99.7**	**59**	**65.3**	**0.75**	**84.0**
SVM^{2}	**87.4**	**82.4**	**93.3**	**32.0**	**100**	**70.5**	**73.5**	**0.75**	**83.9**
G-ML^{C} *	**83.1**	**79**	**95.8**	**27**	**99.8**	**91.8**	**68.3**	**0.76**	**84.5**
SVM^{2}*	**87.8**	**80.8**	**93.1**	**96.9**	**93.6**	**99.3**	**78.8**	**0.86**	**90.9**

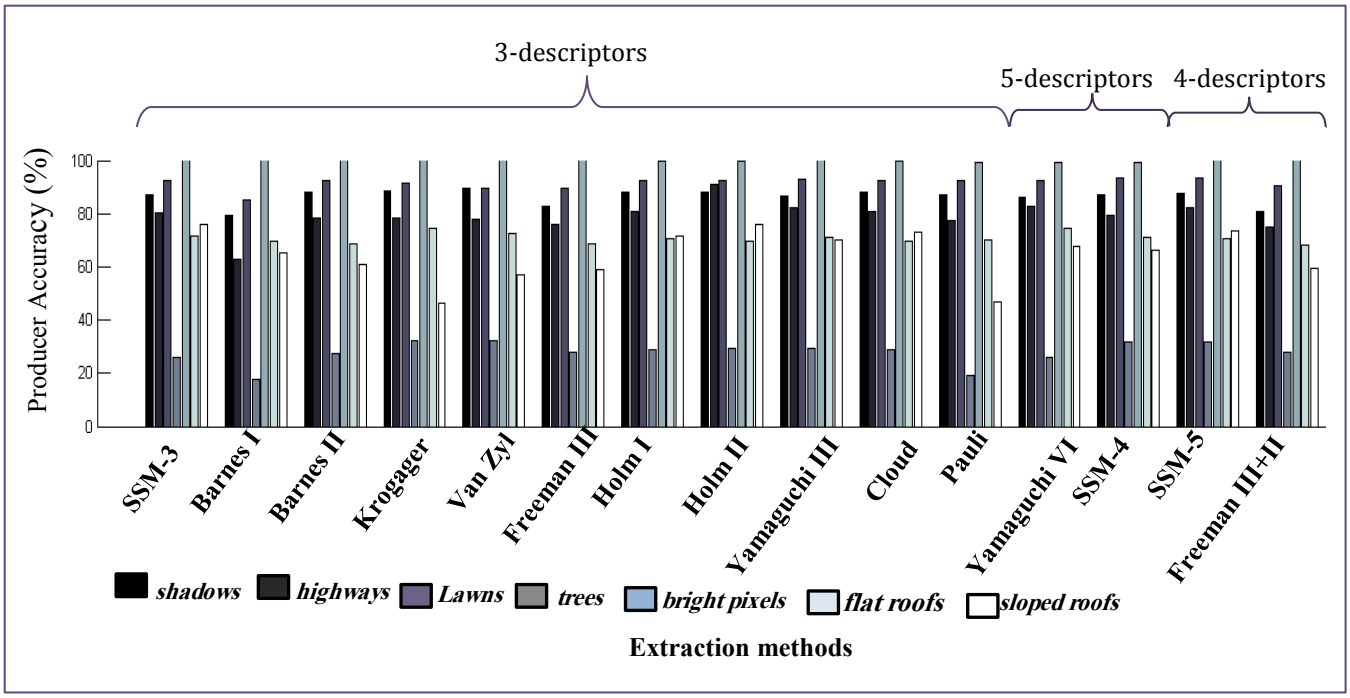
Table 8. 13: Comparison of SVM classification accuracies with different classifiers (in %): producer's accuracy for each class, Kappa coefficient and overall accuracy of different decomposition methods based on different polarimetric matrixes and different subset of SSM based on GA-MI through 127 polarimetrydescriptors.* Each target decomposition subset is made of the normalized components and Span descriptor.

#of Des.	***Extraction Method****	***Shadows***	***highways***	***lawns***	***trees***	***bright pixels***	***flat roofs***	***sloped roofs***	**Kappa Co.**	**OA**
3	SSM	87.2	80.3	92.7	25.8	99.8	71.7	76.0	0.73	82.63
	Barnes I	79.3	63	85.2	17.6	99.9	69.9	65.1	0.60	73.7
	Barnes II	88.3	78.6	92.4	27.3	99.9	68.5	61.1	0.72	81.9
	Krogager	88.5	78.4	91.4	32.3	100	74.5	46.3	0.72	81.8
	Van Zyl	89.5	77.8	89.6	32.5	99.9	72.4	57.1	0.71	80.9
	Freeman III	82.6	76.1	89.5	27.8	100	68.5	59	0.69	79.5
	Holm I	87.91	80.80	92.5	28.7	99.6	70.4	71.6	0.74	82.5
	Holm II	88.04	91.0	92.4	29.3	99.6	69.8	76.2	0.74	82.5
	Yamaguchi III	86.8	82.1	93.0	29.3	99.9	71.2	70.2	0.74	83.2
	Cloude	88.3	80.9	92.5	28.7	99.6	69.5	72.9	0.74	82.83
	Pauli	87.1	77.6	92.5	19.1	99.5	70.0	46.8	0.71	80.77
4	Yamaguchi VI	86.1	82.9	92.4	25.9	99.5	74.4	67.5	0.74	82.6
	SSM	87.3	79.5	93.4	31.7	99.4	71.3	66.3	0.74	83.2
5	SSM	87.4	82.4	93.3	32.0	100	70.5	73.5	0.75	83.9
	Freeman III+II	80.8	75.2	90.5	28.1	100	68.3	59.4	0.69	79.8
6	SSM	87.2	81.0	91.0	31.4	98.6	71.9	68.9	0.73	82.2
	Touzi	87.4	81.5	92.4	26.8	99.9	69.4	67.4	0.73	82.1
7	SSM	87.8	80.4	92.0	31.1	99.6	72.3	69.8	0.74	82.7
8		88.0	82.3	93.2	29.8	99.8	67.1	73.8	0.75	83.4
9	SSM	88.5	80.3	92.1	28.8	99.5	72.0	70.4	0.73	82.6
10	SSM	87.6	80.0	92.8	29.1	99.6	71.5	76.2	0.74	83.0

SVM results from different subsets from SSM based on GA-MI from 127 descriptors with different subsets of different target decomposition methods are described through Figure 8.9. Figure 8.9 (a) shows the overall accuracy of SVM classification. Figure 8.9 (b) shows the producer accuracies of of the seven classes.



(a)



(b)

Figure 8. 9: (a) Overal Accuracy (b) Producer Accuracy of SVM classification results over different subsets from SSM and different target decomposition methods.

8.4. Summary V

In part III, Chapter 7, the result of feature extraction and feature selection is presented. These results consist in the optimal subset (5 real positive descriptors) of 58 initial one and 3 to 10 feature subset of 127 possible descriptors. Then, the MI between each class and descriptor for optimal subset of real positive descriptor (Table 7.1) of SSM based on GA-MI of 58 positive variables and different subset from 3 to 10 of SSM based on GA-MI from 127 possible features (Table 7.2) are calculated. In Chapter 8, the statistical modeling, the performance of their estimations and evaluation using different GoF for seven classes over the optimal real positive subset (Table 8.1) are given. Based upon these results, the MoLC is found to be better than MLE. These experimental studies also reveal that all the train samples (descriptors and texture parameters) can be modeled by Fisher pdf using the MoLC method estimation. According to this statistical analysis, we calculate F-ML classification over the optimal subset from SSM based on GA-MI from 58 real positive descriptors.

To illustrate the efficiency of F-ML classifier, we made a comparison experiment of classification based on two widely used classifiers, i.e., SVM, G-ML (uni-variate). To show the efficiency of building footprint map to improve classification result, a classification based on F-ML and SVM classifiers associated with the morphological component and real positive descriptors (Table 8.1) is presented (Table 8.7). From Table 8.7, we can see that the ML approach based on the Fisher pdf achieves much better performance than other classifiers with an overall accuracy of 84.1%.

Then, in second experiment, the efficiency of SSM (subset selection method) based on GA-MI through 127 descriptors is considered. For this purpose, we made a comparison experiment of SVM classifier over 3 to 10 subset of SSM (Table 7.2), with SVM classifier over different SAR target decomposition components, including the Pauli, Krogager, Freeman, Four-Component, Barnes, Holm, Touzi, Huynen and the Cloude decompositions (Table 8.13). From Table 8.13, the classification results over three descriptors show that, the classification accuracies for all subsets of SSM are higher than 82.5%. Moreover, the Yamaguchi III with an overall accuracy of 83.2% and subset-3 of SSM with 82.63% show better performance than other target decompositions. However, the Yamaguchi III result must be compared with the SVM classification (overall accuracy 83.2%). Moreover, Table 8.13 indicates that the optimal subset of SSM refers to subset-5 with an overall accuracy of 83.9% from all target decomposition components.

However, the classification accuracy of complex Gaussian multi-variate ML approach based on the coherency matrix (nine descriptors) is 84% and 0.1% higher than the one from the optimal subset (5 descriptors). According to all experiences, the main drawback is around class *trees*, which is easily misclassified with two types of roof in urban area.

Chapter 9:

Conclusions and Future research

9. Conclusions and Future research

9.1. Conclusions

In this thesis, we studied the capacity of single look PoLSAR image for discriminating different surfaces in urban area. First a basic frame-work was set up in which statistical polarimetry descriptors were considered. Then, we used a maximum likelihood classifier method based on Fisher distribution. It was shown the improvement of classification accuracy using real positive polarimetry descriptors along with texture parameters. Next, we evaluated the probability to discriminate trees from roofs using morphological tools. Then, we produced building footprints. This has demonstrated to be a good constraint for this purpose and the classification results have shown that it can mostly overcome the lack of additional information and prior knowledge.

We proposed a F-ML classification method and to use it, one must assume the independency of the descriptors and the suitability of Fisher distribution.

For these purposes, first, features were extracted from different methods like decomposition methods, GLCM and polarimetric discriminators. Then, the optimal descriptors based on Genetic algorithm using mRMR evaluation function were selected. Descriptors according to maximum relevance to classification variables and minimum redundancy values (unsimilarity) were evaluated and selected.

Furthermore, to find a general statistical model which is able to model backscattering in heterogonous texture area, the most direct way is to consider different distribution functions. The statistical results showed that the Fisher distribution may be preferred because it allows a better modeling over different train classes than other well-known statistical models such as $\mathcal{K}$, Weibull, Beta I, Log-normal, and Gamma according to GoF results. In each case, it was found that the MoLC is better than the MLE method to estimate parameters.

Moreover, to illustrate the efficiency of F-ML classifier, we made a comparison experiment when using the optimal subset consisting in five real positive descriptors with different classification methods. The results reveal that Fisher-ML achieves much better performances than others. Fisher-ML classification result shows that it discriminates *shadows* from *highways* and *sloped roof* from *trees*. However, the selection of the statistical model should be based on its generalizability; here it can be fitted with our current data but the prediction of its fitness with other data needs deeper consideration. Furthermore, more experiments using other available real data sets should be done for a general conclusion.

We followed the common experiment setup with other polarimetry descriptors. To illustrate the efficiency of subset selection based on GA-MI, we performed a comparison of optimal subset (five

descriptors) with different target decompositions based on different scattering mechanisms, including the Pauli, Krogager, Freeman, Yamaguchi, Barnes, Holm, Huynen and Cloude decompositions. Experimental results on the optimal subset of SSM based on GA-MI of polarimetry SAR data indicate that the proposed method gives a good selection of different scattering mechanisms and gives better performance in the SAR images classification.

The advantage of the proposed SSM based on GA-MI algorithm is largely: first, it is easy to implement, fast, the chances of getting an optimal solution exist, and it does not have limited space of search; secondly, MI has capability to handle non linearity correlation between feature and class as well as anticorrelation between a pair of features; third, GA-MI is used to capture different types of polarimetric SAR images and it is independent of the values and the shape of distribution; finally, this method can be applied to different types of land cover and different sensors as a future work.

However, the selection of the evaluation function, the population size, and the method of selection during the cross over procedure are very important. In other words, a difficulty with genetic algorithm based on mutual information using mRMR evaluation function, is that it requires the number of variables to be selected as the input parameter. Moreover, this approach should be compared to other methods.

9.2. Future research

In this section we discuss the directions for future research. There are some limitations of the current work but solutions can already be foreseen:

The approaches and methods in this research are directly applicable to this study area; however, it can be probably applied to other areas using images with other resolution, or using different sensors (wavelength). While the proposed Fisher classifier has already focused on the use of PoLSAR image for the urban mapping, this method is general and may be applied to other kind of feature sets, which consist in discriminating different classes.

Moreover, it would be a benefit to evaluate the experimental results if the class definition and ground truth could be based on the microwave library of the objects commonly found in the urban area.

Obviously, it can lead to more improvements if the urban map was calculated with high precision DEM or LIDAR data. Consequently, the results could be probably a good tool to improve the quality of high precision DSM data.

As the morphological informations are very useful in classification procedure, investigating ways of expanding the proposed semi-automatic building footprint map extraction to an automatic one remains a direction for future work.

One of future work is to consider and calculate Fisher–ML classification using GA-MI method to take into account possible correlation existing on the different descriptors (classification follows with feature selection).

As the Meijer's functions are defined from inverse Mellin transform, the parameter estimation of these hypergeometrical distribution function using cumulative function can be defined (Nicolas, 2011). Thus, the classification based on Meijer's distribution functions for heterogeneous area remains a possible area for future research.

Bibliography

Aarti, R. (2011). *Roofing Material Types*. Retrieved from Buzzle,Visited at 13.Dec.2011: http://www.buzzle.com/articles/roofing-material-types.html

Abramowitz, M., & Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York.

Ahern, F. J., Mckirdy, I., & Brown, J. (1996). Boreal forest information content of multi-season, multi-polarization C-band SAR data. *Canadian journal of remote sensing , 22(4)*, 349-472.

Ainsworth, T. L., & Schuler, D. (2000). Multi-frequency polarimetric SAR data analysis of ocean surface features. *Proceedings of IGARSS 00*, 24-28.

Ainsworth, T. L., Schuler, D. L., & Lee, J.-S. (2008). Polarimetric SAR characterization of man-made structures in urban areas using normalized circular-pol correlation coefficients. *RemoteSensing of Environment, 112(6)*, 2876-2885.

Ainsworth, T., Cloude, S., & Lee, J. (2002). Eigenvector analysis of polarimetric SAR data. *IEEE Geoscience and Remote Sensing Symposium, 1*, 626-628.

Alander, J. (1992). On optimal population size of genetic algorithms. *IEEE Xplore*, 65-70.

Alberga, V., & Chandra, M. (2000). Analysis of Amplitude Ratios in SAR Polarimetry. *In Kleinheubacher Berichte*, 527-534.

Allain, S., Lopez-Martinez, C., Ferro-Famil, L., & Pottier, E. (2005). New Eigenvalue-based Parameters for Natural Media Characterization . *IEEE Geoscience and Remote Sensing Symposium, 1*.

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis* (third ed.). John Wiley & Sons, Inc.

Anfinsen, S. N. (2010). *Statistical analysis of multilook polarimetric radar images with the Mellin transform. PhD thesis.* Norway: University of Tromso.

Anfinsen, S. N., & Eltoft, T. (2011). Application of the Matrix-Variate Mellin Transform to Analysis of Polarimetric Radar Images. *IEEE Transactions on Geoscience and Remote Sensing*.

Antoniou, A., & Lu, W.-S. (2007). *Practical Optimization: Algorithms and Engineering Application* (2 ed.). New York: Springer Science.

Arauzo-Azofra, A., Benitez, J.-M., & Castro, J.-L. (2008). Consistency measures for feature selection. *J Intell Inf Syst,DOI 10.1007/s10844-007-0037-0, 30*, 273-292.

Arii, M., Van Zyl, J. J., & Yunjin, K. (2011). Adaptive model-based decomposition of polarimetric SAR covariance matrices. *IEEE transactions on Geoscience and Remote Sensing, 49(3)*, 1104-1113.

Baglivo, J. B., Olivier, D., & Pagano, M. (1992). Methods for Exact Goodness-of-Fit Tests. *Journal of the American Statistical Association, 87(418)*, 464-469.

Bamler, R. (1992). A Comparison of Range-Doppler and Wavenumber Domain SAR Focusing Algorithm. *IEEE Transactions on Geoscience and Remote Sensing, 30(4)*.

Banner, M., Jain, A., & Stromberg, W. (1983). A comparative study on synthetic aperture radar and optical imagery of ocean waves. *Reide Publishing Company, 27*, 115-128.

Barnow, B. (1986). Basic Roof. In C. R. Framing. Inc., Blue Ridge Summit, Pa.

Bateman, L. (1953). *Higher transcendental functions.* McGRAW-HILL BOOK COMPANY ,INC.

Baudin, M. (2010). *Nelder-Mead User's Manual.*

Beckman, P., & Spizzichino, A. (1987). *The scattering of electromagnetic waves from rough surfaces.* Norwood, Massachusetts, USA: Artech House Radar Library.

Benboudjema, D., Tupin, F., Pieczynski, W., Sigelle, M., & Nicolas, J.-M. (2007). Unsupervised segmentation of SAR images using triplet Markov fields and Fisher noise distributions. *IEEE International Geoscience and Remote Sensing Symposium, 1*, 3891–3894.

Benediktsson, J., Pesaresi, M., & Arnason, K. (2003). Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Transactions on Geoscience and Remote Sensing, 41(9)*, 1940–1949.

Bertrand, J., Bertrand, P., & Ovarlez, J.-P. (2000). *The Transforms and Applications Handbook. Chapter 12: The Mellin Transform.* Boca Raton, USA: CRC Press Inc.

Bhanu, B., & Lin, Y. (2003). Genetic algorithm based feature selection for target detection in SAR images. *Image and Vision Computing, 21*, 591–608.

Bhattacharya, B., & Touzi, R. (2012). Polarimetric SAR urban classification using the Touzi target scattering decomposition. *Journal canadien de télédétection, 37(4)*, 323-332.

Boerner, W. M. (1987). Recent advances in radar polarimetry - Assessment of the historical development. *International Conference on Antennas and Propagation*, 359-363.

Boerner, W., Läuneburg, E., & Danklmayer, A. (2007). Principal Component Analysis (PCA) in the Context of Radar Polarimetry. *Piers online, 3(5)*, 633-636.

Bombrun, L., & Beaulieu, J.-M. (2008). Fisher distribution for texture modeling of polarimetric SAR data. *IEEE Geoscience and Remote Sensing Letters, 5(3)*, 512–516.

Bombrun, L., Vasile, G., Gay, M., & Totir, F. (2011). Hierarchical Segmentation of Polarimetric SAR Images Using Heterogeneous Clutter Models. *IEEE Transactions on Geoscience And Remote Sensing , 49(2)*, 726-737.

Boutry, J. M. (1993). *ONERA airborne SAR facilities.* Châtillon, France: Office national d'études et de recherches aérospatiales, Châtillon, France.

Brenner, A., & Roessing, L. (2008). Radar Imaging of Urban Areas by Means of Very High-Resolution SAR and Interferometric SAR. *IEEE Transactions on Geoscience and Remote Sensing, 46(10)*, 2971-2982.

Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research, 21(2)*, 230-258.

Brunner, D. (2009). *Advanced Methods for Building Information Extraction from Very High Resolution SAR Data to Support Emergency Response. PhD thesis.* Trento: DISI-University of Trento.

Brunner, D., Lemoine, G., & Bruzzone, L. (2010). Earthquake Damage Assessment of Buildings Using VHR Optical and SAR Imagery. *IEEE Transactions on Geoscience and Remote Sensing, 48(5)*, 2403-2420.

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 121–167.

Burillo, P., Frago, N., & Fuentes, R. (2001). Generation of Fuzzy Mathematical Morphologies. *Mathware & Soft Computing, 8*, 31-46.

Burillo, P., Frago, N., & Fuentes, R. (2003). Fuzzy Morphological Operators in Image Processing. *Marthware & Soft Computing, 10*, 85-100.

Camps-Valls, G., Mooij, J., & Schölkopf, B. (2010). Remote Sensing Feature Selection by Kernel Dependence Measures . *IEEE Geoscience and Remote Sensing Letters, 7(3)*, 587-591 .

Cantalloube, H., Colin, E., Dubois-Fernandez, P., Dupuis, X., Fromage, P., Garestier, F. et al. (2006). The ONERA RAMSES SAR: latest signification results and future developments. *IEEE Conference on Radar*.

Canty, M. (2006). *Image Analysis, Classification and Change Detection in Remote Sensing: With Algorithms for ENVI/IDL.* CRC Press Taylor & Francis Group.

Cao, F., Hong, W., Wu, Y., & Pottier, E. (2007). An unsupervised segmentation with an adaptive number of clusters using the SPAN/H/Alpha/A space and the complex Wishart clustering for fully polarimetric SAR data analysis. *IEEE Transactions on Geoscience and Remote Sensing, 45(11)*, 3454-3467.

Chen, K. S., Huang, W. P., Tsay, D. H., & Amar, F. (1996). Classification of multifrequency polarimetric SAR imagery using a dynamic learning neural network. *IEEE Transactions on Geoscience and Remote Sensing, 34(3)*, 814-820.

Cheriet, M., Said, J. N., & Suen, C. Y. (1998). A Recursive Thresholding Technique for Image Segmentation. *IEEE Transactions on Image Processing, 7(6)*, 918-921.

Chipman, R. (1995). *POLARIMETRY,Chapter 22 in Handbook of Optics II* (2 ed.). New york: McGraw-Hill.

Chitroub, S., Houacine, A., & Sansal, B. (2002). Statistical characterisation and modelling of SAR images. *Signal Processing, 82*, 69-92.

Clausi, D. A. (2002). An analysis of co-occurrance texture statistics as a function of grey level quantization. *Canadian J. Remote Sensing, 28(1)*, 45-62.

Cloude, S. R. (1986). Group theory and polarization algebra. *Optik, 75(1)*, 26–36.

Cloude, S. R., & Pottier, E. (1996). A Review of Target Decomposition Theorems in Radar Polarimetry. *IEEE Transactions on Geoscience and Remote Sensing, 34(2)*, 498-518.

Cloude, S. R., & Pottier, E. (1997). An entropy based classification scheme for land application of polarimetric SAR. *IEEE Transactions on Geoscience and Remote Sensing, 35(1)*, 68-78.

Congalton, R. G. (1991). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing Eviron , 37*, 35-46.

Conover, W. J. (1972). A Kolmogorov Goodness-of-Fit Test for Discontinuous Distributions. *American Statistical Association, 67(339)*, 591-596.

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning, 20*, 273-297.

Cui, Y., Yamaguchi, Y., Yang, J., Park, S., Kobayashi, H., & Singh, G. (2012). Three-Component Power Decomposition for Polarimetric SAR Data Based on Adaptive Volume Scatter Modeling. *Remote Sens., 4*, 1559-1572.

Curlander, J. C., & McDonough, R. N. (1991). *Synthetic Aperture Radar: System and Signal Processing.* A Wiley Interscience Publication.

Cutrona, L. J. (1990). *Synthetic Aperture Radar: in Radar Handbook.* NewYork: McGraw-Hill.

Dash, M., & Liu, H. (1997). Feature Selection for Classification. *Intelligent Data Analysis, 1(3)*, 131–156.

Debnath, L., & Bhatta, D. (2007). *Integral Transforms and their Applications. Chapter 8.* Boca Raton, USA: Chapman & Hall/CRC.

Dousset, B. (1995). Synthetic aperture radar imaging of urban surfaces: a case study. *IEEE International Geoscience and Remote Sensing Symposium, 3*, 2092-2096.

Epstein, B. (1948). Some applications of the Mellin transform in statistics. *Ann.Math.Statist, 19(3)*, 370–379.

Fairbarn Jr, K. G. (2013). *Visible-near infared (VNIR)and shortwave infrared (SWIR) spectral variability of urban materials.Thesis.* Monterey,California.

Fauvel, M., Palmason, J. A., Benediktsson, J. A., Chanussot, J., & Sveinsson, J. R. (2005). Classification of remote sensing imagery with high spatial resolution. *Image and Signal Processing for Remote Sensing XI*, 598201-15.

Ferro-Famil, L. (2011). SAR POLARIMETRY Basics Concepts.(Advanced course on radar Polarimetry). Frascati (Rome), Italy.

Ferro-Famil, L., & Pottier, E. (2007). Urban area remote sensing from L-band PoLSAR data using time-frequency techniques. *IEEE Joint Event Urban Remote Sensing*, 1-6.

Ferro-Famil, L., Pottier, E., & Lee, J. S. (2001). Unsupervised classification of multifrequency and fully polarimetric SAR images based on the H/A/Alpha Wishart classifier. *IEEE Transactions on Geoscience and Remote Sensing, 39(11)*, 2332-2342.

Filliben, J. J. (1975). The Probability Plot Correlation Coefficient Test for Normality. *Technometrics, 17(1)*, 111-117.

Fitelson, B. (2001). *Studies in Bayesian Confirmation Theory.* PhD thesis. University of Wisconsin- Madison.

Formont, P., Pascal, F., Vasile, G., Ovarlez, J.-P., & Ferro-Famil, L. (2011). Statistical Classification for Heterogeneous Polarimetric SAR Images. *IEEE Journal of Selected Topics in Signal Processing, 5(3)*, 567-576.

Franceschetti, G., & Lanari, R. (1999). *Synthetic Aperture Radar Processing.* New York: CRC Press.

Freeman, A. (2007). Fitting a Two-Component Scattering Model to Polarimetric SAR Data From Forests. *IEEE Transactions on Geoscience and Remote Sensing, 42(8)*, 583-2592 .

Freeman, A., & Durden, S. (1998). A three-component scattering model for polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing, 36(3)*, 963-973.

Frery, A. C., Muller, H.-J., Yanasse, C. C., & Sant'Anna, S. J. (1997). A model for extremely heterogeneous clutter. *IEEE Transactions on Geoscience and Remote Sensing, 35(3)*, 648-659.

Frery, A., Correia, A., & Da Freitas, C. (2007). Classifying multifrequency fully polarimetric imagery with multiple sources of statistical evidence and contextual information. *IEEE Transactions on Geoscience and Remote Sensing, 45(10)*, 3098–3109.

Fukuda, S., & Hirosawa, H. (1999). A wavelet-based texture feature set applied to classifcation of multifrequency polarimetric SAR images. *IEEE Transactions on Geoscience and Remote Sensing, 37(5)*, 2282-2286.

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition* (2nd ed.). New York: Academic.

Gaetano, R., Moser, G., Poggi, G., Scarpa, G., & Serpico, S. (2008). Region-Based Classification of Multisensor Optical-SAR Images. *IEEE International Geoscience & Remote Sensing Symposium, IGARSS*, 81-84.

Galland, F., Nicolas, J.-M., Sportouche, H., Roche, M., Tupin, F., & Réfrégier, P. (2009). Unsupervised synthetic aperture radar image segmentation using Fisher distributions. *IEEE Transactions on Geoscience and Remote Sensing, 47(8)*, 2966–2972.

Gao, G. (2010). Statistical Modeling of SAR Images: A Survey. *Sensors,doi:10.3390/s100100775, 10(1)*, 775–795.

Garrara, W. G., Goodman, R. S., Majewski, R. M., & Williams, B. (1995). *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms.* Boston: Artech House.

Giri, N. C. (2004). *Multivariate Statistical Analysis: Revised And Expanded.* New York: Marcel Dekker, Inc.

Gonzales, C., & Woods, E. (2001). *Digital Image Processing. Chapter 9.* New Jersey: Prentice Hall Publications.

Goodman, N. R. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution: an introduction. *in Ann. Math. Statist, 34*, 152–177.

Gulgezen, G., Yu, L., & Cataltepe, Z. (2008). Fast Correlation Based Filter (FCBF) with a different search strategy. *International Symposium on Computer and Information Sciences*, 1-4.

Guyon, I., & Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research, 3(3)*, 1157-1182.

Han, L., & Neumann, M. (2006). Effect of dimensionality on the nelder-mead simplex method. *Optimization Methods and Software, 21(1)*, 1-16.

Haralick, R. (1979). Statistical and Structural Approaches to Texture. *Proceedings of the IEEE, 67(5)*, 786-804.

Haralick, R., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics, 3*, 610-621.

Holm, W. A., & Barnes, R. (1988). On radar polarization mixed target state decomposition techniques. *IEEE National radar Conf*, 249-254.

Hongzhi, W., & Ying, D. (2008). An Improved Image Segmentation Algorithm Based on Otsu Method. *SPIE Digital Library, 6625*, 662501-8.

Hovland, H. A., Johannessen, J. A., & Digranes, G. (1994). Slick detection in SAR images. *IEEE International Geoscience and Remmote Sensing Symposium*, 2038–2040.

Hsu, C. W., & Lin, C.-J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks, 15(2)*, 415-425.

Hussin, Y. A. (1995). Effect of polarization and incidence angle on radar return from urban features using L-band aircraft radar data . *IEEE Int.Geoscience and Remote Sensing Symposium, IGARSS, 1*, 178-180.

Huynen, J. R. (1970). *Phonomenological Theory of Radar Targets.Ph.D.thesis.* Netherlands: University of Technology Delf.

Huynen, J. R. (1990). The Stokes matrix parameters and their interpretation in terms of physicaltarget properties. *SPIE 1317, 195*.

Ince, T. (2010). Polarimetric SAR image classification using a radial basis function neural network. *PIERS 2010, 41(4)*, 636–646.

Ince, T., Kiranyaz, S., & Gabbouj, M. (2012). Evolutionary RBF classifier for polarimetric SAR images. *Expert Systems with Applications, 39*, 4710–4717.

Iwasa, S., Susaki, J., & Tamura, M. (2010). Classification of Building Area using Slant Angle and Density Indices Derived from Polarimetric SAR Data. *International Archives of*

the Photogrammetry, Remote Sensing and Spatial Information Science, XXXVIII,Part 8, 219-224.

Jakeman, E., & Pusey, P. N. (1976). A Model for Non-Rayleigh Sea Echo. *IEEE Transactions on Antennas and Propagation*, *24(6),* 806-814. doi:10.1109/TAP.1976.1141451

Jain, A., & Zongker, D. (1997). Feature Selection: Evaluation, Application, and Small Sample Performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(2)*, 153-158.

Johnson, W. (2002). The Curious History of Faàdi Bruno's Formula. *Am. Math. Monthly, 109(3)*, 217–234.

Kahny, D., & Wiesbeck, W. (1991). Optimum Input Parameters For Classification Of Multifrequency Polarimetric SAR data Using Neural Networks. *IGARSS Geoscience and Remote Sensing Symposium, 4*, 2157-2160.

Kailath, T. (1967). The Divergence and Bhattacharyya Distance Measures in Signal Selection. *IEEE Transactions on Communication Technology, 15(1)*, 52-60 .

Kajimoto, M., & Susaki, J. (2013). Urban-Area Extraction From Polarimetric SAR Images Using Polarization Orientation Angle . *IEEE Geoscience and Remote Sensing Letters, 10(2)*, 337-341.

Kelley, C. T. (1999). *Iterative Methods for Optimization.* Raleigh, North Carolina: Society for Industrial and Applied Mathematics.

Khan, S. (2012). On Single-Look Multivariate G Distribution for PoLSAR Data. *IEEE Applied Earth Observations and Remote Sensing, 5(4)*, 1149-1163.

Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection . *Artificial Intelligence, 97*, 273-324.

Kong, J. A., Lim, H. H., Shin, R., & Van Zyl, J. (1990). Classification of earth terrain using polarimetric synthetic aperture radar images. *Progress In Electromagnetics Research, 3*, 327–370.

Koo, V. C., Chan, Y. K., Vetharatnam, G., Chua, M. Y., Lim, C. H., Lim, C.-S. et al. (2012). New unmanned aerial vehicle synthetic aperture radar for environmental monitoring. *In Electromagnetics Research, 122*, 245-268.

Kouskoulas, Y., Ulaby, F., & Dobson, M. C. (1998). Classification of short vegetation using multifrequency SAR . *IEEE International Geoscience and Remote Sensing Symposium Proceedings , 1*, 103-105.

Krieger, G., Younis, M., Gebert, N., Huber, S., Bordoni, F., Patyuchenko, A. et al. (2010). Advanced Concepts for High-Resolution Wide-Swath SAR Imaging . *EUSAR*, 524-527.

Krogager, E. (1990). A new decomposition of the radar target scattering matrix. *Electron. Lett, 26(18)*, 1525-1526.

Krogager, E. (1993). *Aspects of Polarimetric Radar Imaging.PhD thesis.* Copenhagen: Technical University of Denmark.

Krylov, V. A., Moser, G., Serpico, S. B., & Zerubia, J. (2009). Modeling the statistics of high resolution SAR images. *inria-00342681, 1*.

Krylov, V. A., Moser, G., Serpico, S. B., & Zerubia, J. (2011). Enhanced dictionary-based SAR amplitude distribution estimation and its validation with very high-resolution data. *IEEE Geoscience and Remote Sensing Symposium Letters, 8(1)*, 148-152.

Kuhn, H., & Tucker, A. (1951). Nonlinear programming, in (J. Neyman, el). *University of California Press, Berkeley*, 481-492.

Kun, L., Shao, Y., Touzi, R., Brisco, B., & Zhang, F. (2011). Rice Monitoring Using Touzi Decomposition Based on Polarimetric SAR Data in Southwestern China. *Progress In Electromagnetics Research Symposium Proceedings*, 827-830.

Kuruoglu, E. E., & Zerubia, J. (2004). Modeling SAR Images With a Generalization of the Rayleigh Distribution. *IEEE Transactions on Image Processing, 13(4)*, 527-533.

Lagarias, J. C., Poonen, B., & Wright, M. H. (2012). Convergence of the Restricted Nelder-Mead Algorithm in Two Dimensions. *SIAM Journal on Optimization, 22(2)*, 501–532.

Lagarias, J., Reeds, J., Wright, M., & Wright, P. (1998). Convergence properties of the Nelder-Mead Simplex method in low dimensions. *SIAM Journal of Optimization, 9(1)*, 112–147.

Lee, J. S., & Ainsworth, T. (2011). The Effect of Orientation Angle Compensation on Coherency Matrix and Polarimetric Target Decompositions. *IEEE Transactions on Geoscience and Remote Sensing, 49(1)*, 53-64.

Lee, J. S., & Pottier, E. (2009). *Polarimetric Radar Imaging: From Basics to Applications.* CRC Press,Taylor & Francis Group.

Lee, J. S., Boerner, W., Schuler, D., Ainsworth, T., Hajnsek, I., Papathanassiou, K. et al. (2004). A Review of Polarimetric SAR Algorithms and Their Applications. *Photogrammetry and Remote Sensing, 9(3)*, 31-80.

Lee, J. S., Grunes, M. R., & Pottier, E. (2001). Quantitative Comparison of Classification Capability: Fully Polarimetric Versus Dual and Single-Polarization SAR. *IEEE Transaction on Geoscience and Remote Sensing, 39(11)*, 2343-2351.

Lee, J. S., Grunes, M. R., Ainsworth, T., Du, L.-J., Schuler, D., & Cloude, S. R. (1999). Unsupervised classification using polarimetric decomposition and the complex Wishart classifier. *IEEE Transactions on Geoscience and Remote Sensing, 37(5)*, 2249–2257.

Lee, J. S., Grunes, M., & Kwok, R. (1994). Classifcation of multi-look polarimetric SAR imagery based on complex Wishart distribution. *Int. J. Remote Sensing, 15(11)*, 2299-2311.

Lee, J. S., Krogager, E., Ainsworth, T., & Boerner, W.-M. (2006). Polarimetric Analysis of Radar Signature of a Manmade Structure. *IEEE Geoscience and Remote Sensing Letters, 3(4)*, 555-559.

Lee, K.-Y., Oh, Y., & Kim, Y. (2012). Evaluation of DoP-CPD Classification Technique and Multi Looking Effects for RADARSAT-2 Images. *Korean Journal of Remote Sensing, 28(3)*, 329-336.

Lee, L., Schuler, D., Ainsworth, T., & Boerner, W. (2003). Polarization orientation estimation and applications: a review. *IEEE International Geoscience and Remote Sensing Symposium, 1*, 428-430.

Li, Y., & Zeng, X. (2010). Sequential multi-criteria feature selection algorithm based on agent genetic algorithm. *Appl Intell,DOI 10.1007/s10489-008-0153-8, 33*, 117-131.

Liu, H., & Yu, L. (2005). Towards integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering, 17(4)*, 491–502.

Liu, B., Tupin, F. Liu Y.and Yu, W. (2013). Extraction and characterization of buildings in urban areas via joint use of high-resolution SAR and optical imagery. IGARSS, Melbourne, Australia, 2013.

Long, N., Simonetto, E., & Bocher, E. (2010). A combined approach to detect urban features from multi-spectral and radar data. *IEEE Transaction on Geoscience and Remote Sensing*, 1469–1472.

Lopez-Martinez, C., Pottier, E., & Cloude, S. R. (2005). Statistical Assessment of Eigenvector-Based Target Decomposition Theorems in Radar Polarimetry. *IEEE Transactions on Geoscience and Remote Sensing, 43(9)*, 2058-2074.

Ludwig, O., & Nunes, U. (2010). Novel Maximum-Margin Training Algorithms for Supervised Neural Networks. *IEEE Transactions on Neural Networks, 21(6)*, 972-984.

Ludwig, O., Nunes, U., Araujo, R., Schnitman, L., & Lepikson, H. A. (2009). Applications of information theory, genetic algoritms, and neural models to predict oil flow. *Communications in Nonlinear Science and Numerical Simulation, 14(7)*, 2870-2885.

Marquardt, D. W. (1963). An agorithm for least squares estimation of nonlinear parameters. *Journal of the Society for Industerial and Application Mathematics , 11(2)*, 431-441.

Massonnet, D., & Souyris, J.-C. (2008). *Imaging with Synthetic Aperture Radar.* Boca Roton: EPFL Press, CRC Press.

Materka, A., & Strzelecki, M. (1998). *Texture Analysis Methods – A Review.* Brussels : Technical University of Lodz, Institute of Electronics, COST B11 report.

Matheron, G., & Serra, J. (2000). The Brith of Mathematical Morphology . *ISMM.*

Mathur, B., & Krishna, S. (1977). On multivariate fractional integration operators. *Indian J. Pure Appl. Math, 8*, 1078–1082.

Mercier, G., & Girard-Ardhuin, F. (2005). Unsupervised Oil Slick Detection by SAR Imagery using Kernel Expansion. *IEEE International Geoscience and Remote Sensing Symposium, 1(1)*, 494 - 497.

Mesquita, H., Dupas, C., Silva, M., & Valeriano, D. M. (2008). Amazon Deforestation Monitoring System with ALOS SAR complementary Data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVII*, 1067-1070.

Moghaddam, F. R., & Cheriet, M. (2011). AdOtsu: An adaptive and parameterless generalization of Otsu's method for document image binarization. *The Journal of the Pattern Recognition Society. doi:10.1016/j.patcog.2011.12.013* .

Monahan, A. H., & DelSole, T. (2009). Information theoretic measures of dependence, compactness, and non-gaussianity for multivariate probability distributions. *Nonlinear Processes in Geophysics, 15*, 57–64.

Moriyama, T., Yamaguchi, Y., Uratsuka, S., Umehara, T., Satake, M., Nadai, A. et al. (2004). A Study on Extraction of Urban Areas from Polarimetric Synthetic Aperture Radar image. *IEEE, 1*, 703-706.

Morrison, D. F. (1976). *Mulltivariate Statistical Methods* (Third ed.). Pennylvania: McGraw-Hill, Inc.

Moser, G., & Serpico, S. B. (2010). Contextual remote-sensing image classification by support vector machines and Markov random fields. *IEEE Geoscience and Remote Sensing Symposium, 50(3)*, 3728-3731.

Moser, G., Zerubia, J., & Serpico, S. (2006). SAR Amplitude Probability Density Function Estimation Based on a Generalized Gaussian Model. *IEEE Transactions on Image Processing , 15(6)*, 1429-1442.

Moser, G., Zerubia, J., & Serpico, S. B. (2006). Dictionary-based stochastic expectation-maximization for SAR amplitude probability density function estimation. *IEEE Transactions on Geoscience and Remote Sensing, 44(1)*, 188-200.

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory.* Hoboken, Now Jercy: John Wiley & Sons, Inc.

Müller, K.-R., Mika, S., Rätsch, G., Tsuda, K., & Schölkopf, B. (2001). An Introduction to Kernel-Based Learning Algorithms. *IEEE Transactions on Neural Networks , 12(2)*, 181-201.

Myung, I. J. (2002). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology, 47*, 90-100.

Nasarudin, N. E., & Shafri, H. Z. (2011). Development and Utilization of Urban spectal Library for Remote Sensing of Urban Environment. *Journal of Urban and Environmental Engineering, 5(1)*, 44-56.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal, 7(4)*, 308–313.

Nghiem, S. V., Yueh, S. H., Kwok, R., & Li, F. K. (1992). Symmetry properties in polarimetric remote sensing. *Radio Science, 27(5)*, 693-711.

Nicolas, J. -M., & Tupin, F. (2002). Gamma mixture modeled with "second kind statistics": application to SAR image processing. *IEEE Geoscience and Remote Sensing Symposium, 4*, 2489-2491.

Nicolas, J.-M. (2002). Introduction aux statistiques de deuxième espèce: Application des logs-moments et des logs-cumulants à l'analyse des lois d'images radar. *Traitement du Signal, 19(3)*, 139–167.

Nicolas, J.-M. (2006). *Application de la transformée de Mellin: Étude des lois statistiques de l'imagerie cohérente.* (T. R. 2006D010, Ed.) Paris, France: Ecole Nationale Supérieure des Télécommunications.

Nicolas, J.-M. (2010). *Projet EFIDIR : aspects théoriques en imagerie RSO Lois statistiques des images" produits" et des images " rapports "*. Paris: TélécomParisTech.

Nicolas, J.-M. (2011). *Les distributions de Meijer et leurs propriétés en statistiques de Mellin.* Paris: Télécom ParisTech.

Novey, M., Adali, T., & Roy, A. (2010). A Complex Generalized Gaussian Distribution—Characterization, Generation, and Estimation. *IEEE Transactions on Signal Processing, 58(3)*, 1427-1433.

Oh, Y., Lee, K.-Y., & Jang, G. (2009). New Unsupervised Classification Technique for Polarimetric SAR Images. *Received June 10, 2009; Revised J, 25(3)*, 255-261.

Okan Sakar, C., & Kursun, O. (2010). A Hybrid Method for Feature Selection Based on Mutual Information and Canonical Correlation Analysis . *International Conference on Pattern Recognition (ICPR)*, 4360 - 4363 .

Oliver, C. (1993). Optimum texture estimators for SAR clutter. *Journal of Physics D: Applied Physics, 26*, 1824–1835.

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics, 9(1)*, 62-66.

Pellizzeri, T. M., Gamba, P., Lombardo, P., & Dell'Acqua, F. (2003). Multitemporal/Multiband SAR Classification of Urban Areas Using Spatial Analysis: Statistical Versus Neural Kernel-Based Approach. *IEEE Transaction on Geoscience and Remote Sensing , 41(10)*, 2338-2353.

Peng, H., Long, F., & Ding, C. (2005). Feature Selection Based on Mutual Information:Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence , 27(8)*, 1226-1238.

Pernkopf, F., & O'Leary, P. (2001). Feature Selection for Classification Using Genetic Algorithms with a Novel Encoding. *Computer Analysis of Images and Patterns, 2124*, 161-168 .

Pesaresi, M., & Benediktsson, J. A. (2001). A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing , 39(2)*, 309-320.

Pottier, E., Ferro-Famil, L., Cloude, S., Hajnsek, I., Papathanassiou, L., Moreira, A. et al. (2005). PoLSARpro v2.0 Software A Versatile Polarimetric SAR Data Processing and Educatinal Toolbox.

Poularikas, A. D. (1999). *The Handbook of Formulas and Tables for Signal Processing,Chapter 18: The Mellin Transform.* Boca Raton, U.S: CRC Press & IEEE Press.

Praks, J., Koeniguer, E. C., & Hallikainen, M. (2009). Alternatives to Target Entropy and Alpha Angle in SAR Polarimetry. *IEEE Transactions on Geoscience and Remote Sensing , 47(7)*, 2262-2274.

Pudil, P., Novovičová, J., & Kittler, J. (1994). Floating search methods in feature selection. *Pattern Recognition Letters, 15*, 1119-1125.

Pudil, P., Novovičová, J., Somol, P., & Vrňata, R. (1998). Conceptual base of feature selection consulting system. *Kybernetika, 34(4)*, 451-460.

Qi, Z. Y. (2010). Land use and land cover classification using RADARSAT-2 polarimetric SAR image. *IAPRS, Volume XXXVIII, Part 7A*, 198-203.

Refregier, P., & Morio, J. (2006). Shannon entropy of partially polarized and partially coherent light with Gaussian fluctuations. *JOSA A, 23(12)*, 3036-3044.

Reigber, A., Jager, M., He, W., Ferro-Famil, L., & Hellwich, O. (2007). Detection and classification of urban structures based on high-resolution SAR imagery. *IEEE In Urban Remote Sensing Joint Event*, 1-6.

Ren, S. C., Jin, T., & Wang, Z. (2011). Automated SAR reference image preparation for navigation. *Progress In Electromagnetics Research, 121*, 535-555.

Richards, J. A. (1999). *Remote Sensing Digital Image Analysis* (3rd ed.). Canberra, Australia: Springer.

Rizzo, M. L. (2008). New goodness-of-fit test for pareto distributions. *Astin Bulletin, 39(2)*, 691-715.

RSI. (2003). *ENVI User's Guide.* Research Systems Inc.

RSI-IDL. (2005). Version 6.2 Refrence Guide. Research Systems Inc.

Sahoo, P. K., Soltani, S., Wong, A. K., & Chan, Y. C. (1988). A Survey of Thresholding Techniques. *Computer Vision,Graphics, and Image Processing, 41*, 233-260.

Salaza, J. S., Hush, D. R., Koch, M. W., Fogler, R. J., & Hostetler, L. D. (1998). Statistical modeling of target and clutter in single-look non-polarimetric SAR imagery. *Signal and Image Processing*, 281-158.

Sato, A., Yamaguchi, Y., Singh, G., & Park, S.-E. (2012). Four-component scattering power decomposition with extended volume scattering mode. *IEEE Geoscience and Remote Sensing Letters, 9(2)*, 166-170.

Sauer, S., Ferro-Famil, L., Reigber, A., & Pottier, E. (2011). Three-Dimensional Imaging and Scattering Mechanism Estimation Over Urban Scenes Using Dual-Baseline Polarimetric InSAR Observations at L-Band. *IEEE Transactions on Geoscience and Remote Sensing, 49(11)*, 4616-4629.

Scheuchl, B., Hajnsek, I., & Cumming, I. (2002). Sea ice classification using multi-frequency polarimetric SAR data. *IEEE International Geoscience and Remote Sensing Symposium, 3*, 1914-1916.

Schistad Solberg, A. H., & Jain, A. K. (1997). Texture Fusion and Feature Selection Applied to SAR Imagery. *IEEE Transactions on Geoscience and Remote Sensing, 35(2)*, 475-479.

Schölkopf, B., Burges, C. J., & Smola, A. J. (1998). *Advances in Kernel Methods—Support Vector Learning.* „London, England : The MIT Press.

Serra, J. A. (1982). *Image Analysis and Mathematical Morphology.* New York: Academic Press Inc.

Sezgin, M., & Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging, 13(1)*, 146-165.

Shannon, C. E. (1948). The mathematical theory of communication. *Reprinted with corrections from The Bell System Technical Journal, 27*, 379–423, 623–656.

Shekhar, S., Schrater, P. R., Vatsavai, R. R., Wu, W., & Chawla, S. (2002). Spatial Contextual Classification and Prediction Models for Mining Geospatial Data. *IEEE Transactions on Mutimedia, 4(2)*, 174-188.

Siedlecki, W., & Sklansky, J. (1989). A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters, 10*, 335-347.

Simonetto, E., & Malak, C. (2009). Urban area structuring mapping using an airborne polarimetric SAR image. *SPIE 7477 Image and Signal Processing for Remote Sensing XV, 7477*.

Simonetto, E., Oriot, H., & Garello, R. (2002). Extraction of industrial buildings from stereoscopic airborne radar images. *Proc. SPIE, 4543*, 121-129.

Simonetto, E., Oriot, H., & Garello, R. (2005). Rectangular building extraction from stereoscopic airborne radar images. *IEEE Transactions on Geoscience and Remote Sensing, 43(10)*, 2386-2395.

Slakter, M. J. (1965). A comparison of the Pearson chi-square and Kolmogorov goodness-of-fit tests with respect to validity. *Journal of the American Statistical Association, 60(311)*, 854-858.

Soergel, U. (2010). *Radar Remote Sensing of Urban Areas.* Hannover: Springer.

Soergel, U., Schulz, K., Thoennessen, U., & Stilla, U. (2006). Integration of 3D data in SAR mission planning and image interpretation in urban areas. *Journal of Information Fusion, 6*, 301-310.

Soille, P. (2003). *Morphological Image Analysis—Principles and Applications* (2nd ed., Vol. 2). Berlin,Germany: Springer Verlag.

Soille, P., & Pesaresi, M. (2002). Advances in mathematical morphology applied to geoscience and remote sensing. *IEEE Transactions on Geoscience and Remote Sensing, 40(9)*, 2042-2055.

Somol, P., Pudil, P., & Kittler, J. (2004). Fast branch & bound algorithms for optimal feature selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(7)*, 900-912.

Sportouche, H., & Tupin, F. (2011). Extraction and Three-Dimensional Reconstruction of Isolated Buildings in Urban Scenes From SAR Spaceborne Images. *IEEE International Geoscience and Remote Sensing , 49(10)*, 3932-3946.

Srivastava, M. S. (1965). On the complex Wishart distribution. *Annals of Mathematical Statistics, 36(1)*, 313–315.

Stephens, M. -A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association, 69(347)*, 730-737.

Stephens, M. -A. (1993). *Moments in Statistics: Approximations to Densities and Goodness-of-Fit, Technical Report 469.* Stanford, California: Department of Statistics.

Storvold, R., Malnes, E., Larsen, Y., Hogda, K., Hamran, S., Mueller, K. et al. (2006). SAR remote sensing of snow parameters in norwegian areas. *Journal of Electromagnetic Waves and Applications— Current status and future perspective, 20(13)*, 1751–1759.

Swann, W. H. (1969). A survey of Non-Linear Optimization Techniques. *FEBS Letters, 2(1)*, 39-55.

Szajnowski, W. (1977). Estimators of log-normal distribution parameters. *IEEE Transactions on Aerospace and Electronic Systems, 13(5)*, 533–536.

Tan, C. P., Lim, K. S., & Ewe, H. T. (2007). Image processing in polarimetric SAR images using a hybrid entropy decomposition and maximum likelihood (EDML). *In*

Proceedings of the International Symposium on Image and Signal Processing and Analysis (ISPA) , 418–422.

Tison, C., Nicolas, J.-M., & Tupin, F. (2003). Accuracy of Fisher distributions and log-moment estimation to describe histograms of high-resolution SAR images over urban areas. *IEEE Geoscience and Remote Sensing Symposium, 3*.

Tison, C., Nicolas, J.-M., Tupin, F., & Maître, H. (2004). A new statistical model for Markovian classification of urban areas in high-resolution SAR images. *IEEE Transactions on Geoscience and Remote Sensing, 42(10)*, 2046–2057.

Todd, K. M. (1996). The Expectation-Maximization Algorithm. *IEEE Signal Processing Magazine*, 47-60.

Togan, V., & Daloglu, A. T. (2008). An improved genetic algorithm with initial population strategy and self-adaptive member grouping. *Computers and Structures, 86*, 1204-1218.

Touzi, R. (2007). Target Scattering Decomposition in Terms of Roll Invariant Target Parameters. *IEEE Transactions on Geoscience and Remote Sensing, 45(1)*, 73–84.

Touzi, R., & Charbonneau, F. (2002). Characterization of target symmetric scattering using polarimetric SARs. *IEEE Transactions on Geoscience and Remote Sensing, 40(11)*, 2507-24-516.

Touzi, R., Boerner, W., & Lueneburg, E. (2004). A review of polarimetry in the context of synthetic aperture radar: concepts and information extraction. *Canadian Journal of Remote Sensing, 30(3)*, 380-407.

Touzi, R., Deschamps, A., & Rother, G. (2009). Phase of target scattering for wetland characterization using polametric C-band SAR. *IEEE Trans. Geosci. Remote Sensing, 47(9)*, 3241-3261.

Touzi, R., Raney, R. K., & Charbonneau, F. (2004). On the use of symmetric scatterers for ship characterization. *IEEE Trans. Geosci. Remote Sensing, 42(10)*, 2039-2045.

Tracy, C. A. (2006). A Tutorial on Multivariate Statistical Analysis. *SAMSI*.

Truong-Loi, M.-L., Dubois-Fernandez, P., Freeman, A., & Pottier, P. (2010). Compact Polarimetry Mode at Low Frequency for Vegetation Applications . *IEEE Xplore*, 1-4.

Tuceryan, M., & Jain, A. K. (1998). Texture Analysis.The Handbook of Pattern Recognition and Computer Vision (2nd Edition) by C. H. Chen, L. F. Pau, P. S. P. Wang (eds.). *World Scientific Publishing Co*, 207-248.

Tuia, D., Ratle, F., Pozdnoukhov, A., & Camps-Valls, G. (2010). Multisource Composite Kernels for Urban-Image Classification. *IEEE Geoscience and Remote Sensing Letters, 7(1)*, 88-92.

Tupin, F. (2011). How advanced image processing helps for SAR image restoration and analysis. *IEEE Geoscience and Remote Sensing Society Newsletter*, 10-17.

Ulaby, F. T., Sarebandi, K., & Nashashibi, A. (1992). Statistical properties of the Mueller matrix of distributed targets. *IEEE Radar and Signal Processing, 139(2)*, 136-146 .

Unal, C. M., & Ligthart, L. P. (1998). Decomposition Theorems Applied to Random and Stationary Radar Targets. *PIER 18, Progress In Electromagnetics Research*, 45–66.

Unler, A., Murat, A., & Chinnam, R. B. (2011). mr2PSO:A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vectormachine classification. *Information Sciences, 181*, 4625-4641.

Van Zyl, J. J. (1993). Application of Cloude's target decomposition theorem to polarimetric imaging radar data. *in Proceedings SPIE Conference on Radar Polarimetry,SanDiego,CA, 1748*, 184–212.

Van Zyl, J. J., Arii, M., & Yunjin, K. (2011). Model-Based Decomposition of Polarimetric SAR Covariance Matrices Constrained for Nonnegative Eigenvalues. *IEEE Transactions on Geoscience and Remote Sensing, 49(9)*, 3452-3459.

Van, A., & Bos, D. (1995). The Mutivariate complex Normal Distribution -A Generalization . *IEEE Transactions on Information Theory , 41(2)*, 537-539.

Vapnik, V. (1998). *Statistical Learning Theory.* New York: John Wiley & Sons.

Vasile, G., Ovarlez, J.-P., Pascal, F., & Tison, C. (2010). Coherency matrix estimation of heterogeneous clutter in high resolution polarimetric SAR images. *IEEE Transactions on Geoscience and Remote Sensing, 48(4)*, 1809–1826.

Vasile, G., Pascal, F., Ovarlez, J.-P., Formont, P., & Gay, M. (2011). Optimal Parameter Estimation in Heterogeneous Clutter for High Resolution Polarimetric SAR Data. *IEEE Geoscience and Remote Sensing Letters, 8(6)*, 1046-1050.

Vincent, L. (1993). Morphological Grayscale Reconstruction in Image Analysis : Application and Efficient Algorithm. *IEEE Transaction on Image Processing, 2(2)*, 176-201.

Vincent, L., & Beuche, S. (1989). *The morphological approach to segmentation: an introduction. Technical report, Ecole des Mines.* Paris: CMM.

Vinh, L. T., Lee, S., Park, Y.-T., & D'Auriol, B. J. (2012). A novel feature selection method based on normalized mutual information. *Appl Intell, DOI 10.1007/s10489-011-0315-y, 37*, 100-120.

Vladimir, N. V. (1999). An Overview of Statistical Learning Theory. *IEEE Transactions on Neural Networks, 10(5)*, 988-999.

Vogel, R. M. (1986). The Probability Plot Correlation Coefficient Test for the Normal Lognormal, and Gumbel Distributional Hypothesis. *Water Resources Research , 22(4)*, 587-590.

Walck, C. (2007). *Hand-book on Statistical Distributions for experimentalists.* Stockholm: Internal Report SUF–PFY/96–01.

Wei, L., Hu, Z., Guo, M., Jiang, M., & Zhang, S. (2012). Texture feature analysis in oil spill monitoring by SAR image. *IEEE International on Geoinformatics*, 1-6.

Weisstein, E. -W. (2012b). *Existence Theorem*. Retrieved from Wolfram MathWorld,Visited at 10.Dec.2012: http://mathworld.wolfram.com/ExistenceTheorem.html

Weisstein, E. -W. (2012c). *Correlation Coefficient,visited at 20 Desember 2012*. Retrieved from Wolfram MathWorld:
http://mathworld.wolfram.com/CorrelationCoefficient.html

Weisstein, E. -W. (2013). *Genetic Algorithm*. Retrieved from Walfram MathWorld,Visited at 1.Mars.2013: http://mathworld.wolfram.com/GeneticAlgorithm.html

Weisstein, E.-W. (2012a). *Bell Polynomial*. Retrieved from Wolfram MathWorld,Visited at 12.Nov.2012: http://mathworld.wolfram.com/BellPolynomial.html

Weydahl, D. (1997). Identifying urban features using RADARSAT images taken at multiple incidence angles. *In IEEE International Geoscience and Remote Sensing Symposium, 1*, 287-289.

Xia, Z., & Henderson, F. (1997). Understanding the relationships between radar response patterns and the bio- and geophysical parameters of urban areas. *IEEE Transactions on Geoscience and Remote Sensing, 35(1)*, 93-101.

Yamaguchi, Y., Ishido, M., Moriyama, T., Yang, J., & Yamada, H. (2005). Polarimetric SAR Image Decomposition for Non-Reflection Symmetry Condition. *Proceeding of ISAP,ISBN: 89-86522-78-0*, 431-434.

Yamaguchi, Y., Moriyama, M., Ishido, M., & Yamada, H. (2005). Four-Component Scattering Model for Polarimetric SAR Image Decomposition. *IEEE Transactions on Geoscience and Remote Sensing, 43(8)*, 1699-1706.

Yamaguchi, Y., Sato, A., Boerner, W.-M., Sato, R., & Yamada, H. (2011). Four Components Scattering Power Decomposition with Rotation of Coherency Matrix. *IEEE Transactions on Geoscience and Remote Sensing, 49(6)*, 2251-2258.

Yang, J., Peng, Y. -N., Yamaguchi, Y., & Yamada, H. (2006). On Huynen's Decomposition of a Kennaugh Matrix. *IEEE Transactions on Geoscience and Remote Sensing, 3(3)*, 369-372.

Yang, J., Yamaguchi, Y., Yamada, H., Boerner, W.-M., Mott, H., & Peng, Y. (2001). Development of Target Null Theroy . *IEEE Transactions on Geoscience and Remote Sensing , 39(2)*, 330-338.

Yang, W., Song, H., Xia, G.-S., & Xu, X. (2013). On the Mixed Scattering Mechanism Analysis of Model -Based Decomposition for Polarimetric SAR data. *Progress In Electromagnetics Research B, 52*, 327-345.

Zebker, F., & Van Zyl, J. (1991). Imaging radar polarimetry: A review. *In Proceedings of the IEEE, 79*, 1583–1606.

Zhang, J. J., Wang, P., Chen, L., Li, Y., Yin, Q., & Hong, W. (2013). Seven-category model-based segmentation for polarimetric SAR data. *PolInSAR'13,ESA*.

Zhang, L., Zou, B., Jia, Q., & Zhang, Y. (2009). Polarimetric SAR image classification using Multiple-Component Scattering Model and Support Vector Machine. *IEEE Synthetic Aperture Radar*, 805-808.

Zhou, X., Chang, N. B., & Li, S. (2009). Application of SAR Interferometry in Earth and Environment Sceince Research. *Sensors, 9*, 1876-1912.

Zhou, X., He, G., Fan, Y., Xiao, Y., Kunnath, S. K., & Monti, G. (2013). Fuzzy Support Vector Machine for PoLSAR Image Classification. *Advanced Materials Research, 639-640*, 1162-1167.

Appendix A

Table A. 1: Correlation Coefficients, Kolmogorov-Smirnov Distance, Root Mean Square Error between the estimated pdf and empirical data over optimal subset (Table 8.1) of real positive descriptors for **Class1-*Shadows*** respect to different estimation methods (MoLC, MLE) following with estimated parameters.

Pdf	Method		Des.	CE	K-S(test)	RMSE	Par1	Par2	Par3
Log-Normal	MoLC	(mu,sigma)	1	0.9782	0.5817	0.1325	-0.3996	0.3367	
Gamma	MLE	(mu,L)	1	0.9671	0.5587	0.1638	0.7490	2.4142	
Fisher	MLE	(mu,L,M)	1	0.9804	0.5847	0.1256	0.6817	4.2517	5.7502
Gamma	MoLC	(mu,L)	1	0.9699	0.5754	0.1555	0.7407	2.6709	
Fisher	MoLC	(mu,L,M)	1	0.9796	0.5889	0.1285	0.6720	4.8393	5.0349
K	MoLC	(mu,L,M)	1	0.9789	0.5847	0.1306	0.7410	5.1705	5.1705
Log-Normal	MoLC	(mu,sigma)	2	0.9711	0.8299	0.7498	-1.7563	0.1578	
Fisher	MoLC	(mu,L,M)	2	0.9707	0.8269	0.7547	0.1727	20.6060	20.5941
Weibull	MLE	(mu,eta)	2	0.9117	0.6452	1.2939	0.1873	3.0580	
Gamma	MLE	(mu,L)	2	0.9546	0.8299	0.9355	0.1772	9.8815	
Gamma	MoLC	(mu,L)	2	0.9540	0.8328	0.9439	0.1769	10.5408	
K	MoLC	(mu,L,M)	2	0.9629	0.8328	0.8489	0.1769	20.9143	20.9143
Fisher	MoLC	(mu,L,M)	3	0.9631	0.9655	0.1702	0.9963	80.6809	90.8045
Fisher	MoLC	(mu,L,M)	4	0.9214	0.8985	0.2790	0.8771	14.1210	70.5050
Log-Normal	MoLC	(mu,sigma)	5	0.9812	1.0000	0.0398	0.7455	0.3354	
Gamma	MLE	(mu,L)	5	0.9646	1.0000	0.0535	2.3463	2.4821	
Fisher	MLE	(mu,L,M)	5	0.9798	1.0000	0.0410	2.1804	3.7492	7.1855
Gamma	MoLC	(mu,L)	5	0.9642	1.0000	0.0545	2.3261	2.6873	
Fisher	MoLC	(mu,L,M)	5	0.9801	1.0000	0.0419	2.1252	4.7032	5.5197
K	MoLC	(mu,L,M)	5	0.9773	1.0000	0.0439	2.3303	5.1356	5.1356

Table A. 2: Correlation Coefficients, Kolmogorov-Smirnov Distance, Root Mean Square Error between the estimated pdf and empirical data over optimal subset (Table 8.1) of real positive descriptors for **Class2-*Highways*** respect to different estimation methods (MoLC, MLE) following with estimated parameters.

Pdf	Method	Des.	CE	K-S(test)	RMSE	Par1	Par2	Par3
Log-Normal	MoLC (mu,sigma)	1	0.9782	0.5817	0.1325	-0.3996	0.3367	
Gamma	MLE (mu,L)	1	0.9671	0.5587	0.1638	0.7490	2.4142	
Fisher	MLE (mu,L,M)	1	0.9804	0.5847	0.1256	0.6817	4.2517	5.7502
Gamma	MoLC (mu,L)	1	0.9699	0.5754	0.1555	0.7407	2.6709	
Fisher	MoLC (mu,L,M)	1	0.9796	0.5889	0.1285	0.6720	4.8393	5.0349
K	MoLC (mu,L,M)	1	0.9789	0.5847	0.1306	0.7410	5.1705	5.1705
BetaI	MoLC (mu,L,M)	1	0.9697	0.6061	0.1664	0.7280	3.1332	132.9833
Log-Normal	MoLC (mu,sigma)	2	0.9711	0.8299	0.7498	-1.7563	0.1578	
Fisher	MoLC (mu,L,M)	2	0.9707	0.8269	0.7547	0.1727	20.6060	20.5941
Weibull	MLE (mu,eta)	2	0.9117	0.6452	1.2939	0.1873	3.0580	
Gamma	MLE (mu,L)	2	0.9546	0.8299	0.9355	0.1772	9.8815	
Gamma	MoLC (mu,L)	2	0.9540	0.8328	0.9439	0.1769	10.5408	
K	MoLC (mu,L,M)	2	0.9629	0.8328	0.8489	0.1769	20.9143	20.9143
Fisher	MoLC (mu,L,M)	3	0.9631	0.9655	0.1702	0.9963	80.6809	90.8045
Fisher	MoLC (mu,L,M)	4	0.9214	0.8985	0.2790	0.8771	14.1210	70.5050
Log-Normal	MoLC (mu,sigma)	5	0.9812	1.0000	0.0398	0.7455	0.3354	
Gamma	MLE (mu,L)	5	0.9646	1.0000	0.0535	2.3463	2.4821	
Fisher	MLE (mu,L,M)	5	0.9798	1.0000	0.0410	2.1804	3.7492	7.1855
Gamma	MoLC (mu,L)	5	0.9642	1.0000	0.0545	2.3261	2.6873	
Fisher	MoLC (mu,L,M)	5	0.9801	1.0000	0.0419	2.1252	4.7032	5.5197
K	MoLC (mu,L,M)	5	0.9773	1.0000	0.0439	2.3303	5.1356	5.1356
BetaI	MoLC (mu,L,M)	5	0.9577	1.0000	0.0647	2.2902	3.0458	74.0341

Table A. 3: Correlation Coefficients, Kolmogorov-Smirnov Distance, Root Mean Square Error between the estimated pdf and empirical data over optimal subset (Table 8.1) of real positive descriptors for **Class3-*Lawns*** respect to different estimation methods (MoLC, MLE) following with estimated parameters.

Pdf	Method	Des.	CE	K-S(test)	RMSE	Par1	Par2	Par3
Log-Normal	MoLC (mu,sigma)	1	0.9920	0.6619	0.1056	-0.7447	0.3199	
Gamma	MLE (mu,L)	1	0.9845	0.6560	0.1449	0.5242	2.6859	
Fisher	MLE (mu,L,M)	1	0.9919	0.6619	0.1064	0.4873	4.2538	7.2417
Gamma	MoLC (mu,L)	1	0.9837	0.6657	0.1502	0.5200	2.9107	
Fisher	MoLC (mu,L,M)	1	0.9917	0.6619	0.1082	0.4834	4.5720	6.5912
K	MoLC mu,L,M)	1	0.9906	0.6679	0.1160	0.5206	5.5969	5.5969
Log-Normal	MoLC (mu,sigma)	2	0.9924	0.7934	0.3195	-1.1760	0.1189	
Fisher	MoLC (mu,L,M)	2	0.9922	0.7940	0.3250	0.3063	76.4942	23.5866
Weibull	MLE (mu,eta)	2	0.9678	0.6760	0.6702	0.3276	4.0250	
Gamma	MLE(mu,L)	2	0.9904	0.7940	0.3568	0.3130	17.6953	
Gamma	MoLC (mu,L)	2	0.9898	0.7970	0.3685	0.3128	18.2047	
K	MoLC (mu,L,M)	2	0.9914	0.7970	0.3385	0.3129	36.0328	36.0328
Fisher	MLE (mu,L,M)	2	0.9918	0.7970	0.3310	0.3101	23.5278	74.3155
Fisher	MoLC (mu,L,M)	3	0.9643	0.8269	0.2317	0.7428	19.3549	19.3549
Weibull	MLE (mu,eta)	3	0.9814	0.7366	0.1677	0.7990	3.9802	
Fisher	MoLC (mu,L,M)	4	0.9906	0.8496	0.0923	0.5988	4.1109	25.0209
Log-Normal	MoLC (mu,sigma)	5	0.9878	1.0000	0.0326	0.6692	0.3460	
Gamma	MLE (mu,L)	5	0.9753	1.0000	0.0459	2.1949	2.2905	
Fisher	MLE (mu,L,M)	5	0.9862	1.0000	0.0348	1.9783	4.1634	5.2198
Gamma	MoLC (mu,L)	5	0.9748	1.0000	0.0467	2.1673	2.5517	
Fisher	MoLC mu,L,M)	5	0.9856	1.0000	0.0359	1.9528	4.6677	4.6726
K	MoLC (mu,L,M)	5	0.9836	1.0000	0.0386	2.1655	4.9941	4.9941
BetaI	MoLC (mu,L,M)	5	0.9631	0.9986	0.0650	2.1153	3.1772	89.9025

Table A. 4: Correlation Coefficients, Kolmogorov-Smirnov Distance, Root Mean Square Error between the estimated pdf and empirical data over optimal subset (Table 8.1) of real positive descriptors for **Class4-*Trees*** respect to different estimation methods (MoLC, MLE) following with estimated parameters.

Pdf	Method	Des.	CE	K-S(test)	RMSE	Par1	Par2	Par3
Log-Normal	MoLC (mu,sigma)	1	0.9664	0.6347	0.2123	-0.6991	0.3315	
Gamma	MLE (mu,L)	1	0.9633	0.6228	0.2234	0.5521	2.5350	
Fisher	MLE (mu,L,M)	1	0.9709	0.6347	0.1979	0.5133	3.8933	7.3682
Gamma	MoLC (mu,L)	1	0.9661	0.6377	0.2130	0.5474	2.7434	
Fisher	MoLC (mu,L,M)	1	0.9678	0.6345	0.2078	0.4973	4.9340	4.9852
K	MoLC (mu,L,M)	1	0.9702	0.6347	0.2002	0.5496	5.1383	5.1383
Log-Normal	MoLC (mu,sigma)	2	0.9570	0.7284	0.2600	-0.3670	0.1999	
Fisher	MoLC (mu,L,M)	2	0.9568	0.7284	0.2609	0.6928	13.0161	13.0188
Weibull	MLE (mu,eta)	2	0.9704	0.7194	0.2149	0.7626	2.8618	
Gamma	MLE (mu,L)	2	0.9707	0.7313	0.2149	0.7194	6.7999	
Gamma	MoLC (mu,L)	2	0.9708	0.7265	0.2143	0.7196	6.7549	
K	MoLC (mu,L,M)	2	0.9666	0.7254	0.2284	0.7206	9.3544	20.7664
Fisher	MLE (mu,L,M)	2	0.9503	0.7254	0.2785	0.6836	14.6044	10.6647
Fisher	MoLC (mu,L,M)	3	0.9729	0.7574	0.1840	0.3263	3.4066	3.4568
Weibull	MLE (mu,eta)	3	0.9822	0.7731	0.1437	0.3969	1.4166	
Fisher	MoLC (mu,L,M)	4	0.9690	0.7812	0.1142	0.4791	1.7889	2.8024
Log-Normal	MoLC (mu,sigma)	5	0.9788	1.0000	0.0488	0.6372	0.3307	
Gamma	MLE (mu,L)	5	0.9378	1.0000	0.0843	2.1601	2.0309	
Fisher	MLE (mu,L,M)	5	0.9763	1.0000	0.0491	1.7735	15.4850	3.2611
Gamma	MoLC (mu,L)	5	0.9668	1.0000	0.0592	2.0821	2.7547	
Fisher	MoLC (mu,L,M)	5	0.9770	1.0000	0.0523	1.8787	4.8381	4.3207
K	MoLC (mu,L,M)	5	0.9838	1.0000	0.0406	2.0532	6.2450	6.2450

Table A. 5: Correlation Coefficients, Kolmogorov-Smirnov Distance, Root Mean Square Error between the estimated pdf and empirical data over optimal subset (Table 8.1) of real positive descriptors for **Class5-*Bright pixels*** respect to different estimation methods (MoLC, MLE) following with estimated parameters.

Pdf	Method		Des.	CE	K-S(test)	RMSE	Par1	Par2	Par3
Log-Normal	MoLC	(mu,sigma)	1	0.9954	0.8102	0.0713	-1.4727	0.5391	
Gamma	MLE	(mu,L)	1	0.9509	0.8149	0.2345	0.3120	0.9435	
Fisher	MLE	(mu,L,M)	1	0.9943	0.8055	0.0826	0.2177	2.6800	1.8046
Gamma	MoLC	(mu,L)	1	0.9592	0.8418	0.2108	0.2857	1.2779	
Fisher	MoLC	(mu,L,M)	1	0.9928	0.8085	0.0900	0.2293	2.1788	2.1787
K	MoLC	(mu,L,M)	1	0.9675	0.8532	0.2085	0.2724	3.0601	3.0601
Log-Normal	MoLC	(mu,sigma)	2	0.9097	1.0000	0.0296	1.3850	0.6053	
Fisher	MoLC	(mu,L,M)	2	0.9027	1.0000	0.0309	3.6684	2.5132	1.4519
Gamma	MoLC	(mu,L)	2	0.8813	1.0000	0.0337	5.2005	1.0840	
K	MoLC	(mu,L,M)	2	0.8836	1.0000	0.0367	4.7321	3.1082	3.1082
Fisher	MLE	(mu,L,M)	2	0.9074	1.0000	0.0300	3.7984	2.1672	1.5793
Fisher	MoLC	(mu,L,M)	3	0.9079	0.9559	1.7972	0.9700	80.8045	89.8045
Weibull	MLE	(mu,eta)	3	0.9382	0.7754	1.6991	0.9905	9.5469	
Fisher	MoLC	(mu,L,M)	4	0.9214	0.9604	0.2174	0.9700	80.8045	89.8045
Log-Normal	MoLC	(mu,sigma)	5	0.9591	1.0000	0.0221	1.2619	0.5570	
Gamma	MLE	(mu,L)	5	0.8929	1.0000	0.0357	4.8345	0.9279	
Fisher	MLE	(mu,L,M)	5	0.9663	1.0000	0.0200	3.4650	2.2076	1.9246
Gamma	MoLC	(mu,L)	5	0.9171	1.0000	0.0309	4.4526	1.2194	
Fisher	MoLC	(mu,L,M)	5	0.9658	1.0000	0.0201	3.5323	2.0650	2.0650
K	MoLC	(mu,L,M)	5	0.9505	1.0000	0.0247	4.2865	2.7389	2.7389

Table A. 6: Correlation Coefficients, Kolmogorov-Smirnov Distance, Root Mean Square Error between the estimated pdf and empirical data over optimal subset (Table 8.1) of real positive descriptors for **Class6-*Flat roofs*** respect to different estimation methods (MoLC, MLE) following with estimated parameters.

Pdf	Method	Des.	CE	K-S(test)	RMSE	Par1	Par2	Par3
Log-Normal	MoLC (mu,sigma)	1	0.9882	0.8072	0.0834	-0.4142	0.3056	
Gamma	MLE (mu,L)	1	0.9881	0.8005	0.0819	0.7226	2.9550	
Fisher	MLE (mu,L,M)	1	0.9866	0.8057	0.0797	0.6868	4.1043	10.2224
Gamma	MoLC (mu,L)	1	0.9863	0.8064	0.0896	0.7185	3.1481	
Fisher	MoLC (mu,L,M)	1	0.9891	0.8088	0.0893	0.6611	5.7602	5.8146
K	MoLC (mu,L,M)	1	0.9890	0.8072	0.0803	0.7199	6.0069	6.0069
Log-Normal	MoLC (mu,sigma)	2	0.9934	0.6647	0.1948	-0.4751	0.1220	
Fisher	MoLC (mu,L,M)	2	0.9934	0.6657	0.1939	0.6217	34.3137	33.6253
Gamma	MLE (mu,L)	2	0.9894	0.6627	0.2480	0.6313	16.5701	
Gamma	MoLC (mu,L)	2	0.9900	0.6668	0.2377	0.6309	17.3056	
K	MoLC (mu,L,M)	2	0.9922	0.6668	0.2101	0.6310	34.3200	34.3200
Fisher	MoLC (mu,L,M)	3	0.9905	0.7623	0.1143	0.4024	2.6401	11.0442
Weibull	MLE (mu,eta)	3	0.9968	0.7689	0.0644	0.4363	1.8016	
Fisher	MoLC (mu,L,M)	4	0.9707	0.7804	0.1220	0.3676	1.7815	2.0401
Log-Normal	MoLC (mu,sigma)	5	0.9906	0.9807	0.0286	0.5122	0.3368	
Gamma	MLE (mu,L)	5	0.9835	0.9896	0.0387	1.8656	2.3985	
Fisher	MLE (mu,L,M)	5	0.9936	0.9807	0.0236	1.6994	4.2180	5.9062
Gamma	MoLC (mu,L)	5	0.9876	0.9852	0.0328	1.8435	2.6692	
Fisher	MoLC (mu,L,M)	5	0.9931	0.9807	0.0244	1.6791	4.6904	5.2476
K	MoLC (mu,L,M)	5	0.9931	0.9822	0.0244	1.8462	5.1153	5.1153

Table A. 7: Correlation Coefficients, Kolmogorov-Smirnov Distance, Root Mean Square Error between the estimated pdf and empirical data over optimal subset (Table 8.1) of real positive descriptors for **Class7-*Sloped roofs*** respect to different estimation methods (MoLC, MLE) following with estimated parameters.

Pdf	Method		Des.	CE	K-S(test)	RMSE	Par1	Par2	Par3
Log-Normal	MoLC	(mu,sigma)	1	0.9832	0.8806	0.3518	-1.6945	0.2052	
Gamma	MLE	(mu,L)	1	0.9813	0.8866	0.3723	0.1915	6.1376	
Fisher	MLE	(mu,L,M)	1	0.9849	0.8836	0.3343	0.1845	11.3011	13.8339
Gamma	MoLC	(mu,L)	1	0.9830	0.8866	0.3538	0.1911	6.4451	
Fisher	MoLC	(mu,L,M)	1	0.9844	0.8806	0.3389	0.1837	12.4106	12.4108
K	MoLC	(mu,L,M)	1	0.9846	0.8836	0.3375	0.1913	12.4008	12.6349
Fisher	MoLC	(mu,L,M)	2	0.9207	0.6955	0.3003	0.8167	11.2141	10.6436
Gamma	MoLC	(mu,L)	2	0.9156	0.6896	0.3090	0.8556	5.6782	
K	MoLC	(mu,L,M)	2	0.9200	0.6955	0.3011	0.8560	11.0840	11.0840
Fisher	MoLC	(mu,L,M)	3	0.9619	0.8149	0.2311	0.4616	7.6380	8.4277
Weibull	MLE	(mu,eta)	3	0.9651	0.7539	0.2199	0.5211	2.2351	
Fisher	MoLC	(mu,L,M)	4	0.9050	0.8269	0.2259	0.6137	1.7234	116.9334
Log-Normal	MoLC	(mu,sigma)	5	0.8997	1.0000	0.0621	1.8392	0.1758	
Gamma	MLE	(mu,L)	5	0.8708	1.0000	0.0698	6.4976	7.9178	
Gamma	MoLC	(mu,L)	5	0.8750	1.0000	0.0685	6.4804	8.6098	
Fisher	MoLC	(mu,L,M)	5	0.9009	1.0000	0.0617	6.2914	16.7349	16.7349
K	MoLC	(mu,L,M)	5	0.8893	1.0000	0.0648	6.4797	17.1240	17.1240
Weibull	MLE	(mu,eta)	5	0.7990	1.0000	0.0859	6.8915	2.6608	

Classification d'images RSO polarimétriques à haute résolution spatiale sur site urbain

Maryam Soheili Majd

Résumé

Notre recherche vise à évaluer l'apport d'une seule image polarimétrique RSO (Radar à Synthèse d'Ouverture) à haute résolution spatiale pour classifier les surfaces urbaines. Pour cela, nous définissons plusieurs types de toits, de sols et d'objets.

Dans un premier temps, nous proposons un inventaire d'attributs statistiques, texturaux et polarimétriques pouvant être utilisés dans un algorithme de classification. Nous étudions les lois statistiques des descripteurs et montrons que la distribution de Fisher est bien adaptée pour la plupart d'entre eux.

Dans un second temps, plusieurs algorithmes de classification vectorielle supervisée sont testés et comparés, notamment la classification par maximum de vraisemblance basée sur une distribution gaussienne, ou celle basée sur une distribution gaussienne complexe comme modèle statistique de la matrice de cohérence polarimétrique, ou encore l'approche SVM.

Nous proposons alors une variante de l'algorithme par maximum de vraisemblance basée sur une distribution de Fisher, dont nous avons étudié l'adéquation avec l'ensemble de nos attributs. Nous obtenons une nette amélioration de nos résultats avec ce nouvel algorithme mais une limitation apparaît pour reconnaître certains toits.

Ainsi, la forme rectangulaire des bâtiments est reconnue par des opérations morphologiques à partir de l'image d'amplitude radar. Cette information spatiale est introduite dans le processus de classification comme contrainte. Nous montrons tout l'intérêt de cette information puisqu'elle empêche la confusion de classification entre pixels situés sur des toits plats et sur des arbres.

Une méthode de sélection basée sur l'algorithme génétique utilisant l'Information mutuelle (AG-IM) est adaptée pour déterminer l'ensemble optimal des attributs pour la méthode de classification. Pour illustrer l'efficacité de la sélection, nous avons comparé son utilisation avec celle des attributs issus de décompositions polarimétriques basées sur des hypothèses différentes de mécanismes de diffusion.

Les résultats mettent en évidence le potentiel de ces données pour discriminer certains types de couverture de l'espace urbain.

Mots clés

Zone urbaine, classification d'images, maximum de vraisemblance, Fisher, haute résolution spatiale, polarimétrie radar, algorithme génétique, information mutuelle.

Introduction

La précision d'une classification supervisée dans le contexte urbain dépend principalement des caractéristiques des données RSO telles que la résolution spatiale, fréquence, nombre de vues, angle de visée, mode polarimétrique, etc. Elle est également liée au type de descripteurs polarimétriques utilisés, à la méthode de classification et à sa mise en œuvre. En outre, un résultat plus précis est obtenu lorsque la connaissance du terrain est prise en compte.

L'image à haute résolution spatiale RSO offre des informations très précises sur les propriétés des cibles. Cependant, l'interprétation des classes urbaines dépend de la texture des cibles, de leur géométrie et de leur orientation par rapport au radar. En effet, des structures similaires avec des orientations différentes peuvent présenter une apparence différente dans l'image RSO. Par ailleurs, différentes cibles telles que rues, trottoirs, routes ou sol nu ont des coefficients de rétrodiffusion faibles dans l'image RSO et présentent le même comportement ce qui engendre des ambigüités (Dousset, 1995). De plus, la cartographie selon la visée oblique provoque des distorsions géométriques comme les effets de raccourcissement et de repliement, ainsi que des ombres, ce qui complique l'interprétation des mécanismes de diffusion.

Ces complexités constituent de nombreux défis pour séparer les différents types de surfaces par classification d'images radar. Toutefois, afin de surmonter ces problèmes de classification et d'améliorer le résultat de classification d'images radar, plusieurs approches analytiques et numériques ont été développées.

Ainsi, certaines méthodes de classification sont basées sur la modélisation physique des mécanismes de diffusion de l'image de RSO polarimétrique en utilisant une décomposition polarimétrique et des descripteurs (Unal & Ligthart, 1998 ; Cloude & Pottier, 1997; Ferro-Famil, et al., 2001 ; Yang, et al., 2006 ; Kajimoto & Susaki, 2013 ; Yamaguchi, et al., 2005 ; Yamaguchi, et al., 2011 ; Simonetto & Malak, 2009) ou des discriminateurs polarimétriques (Moriyama, et al., 2004 ; Iwasa, et al., 2010). D'autres travaux sont basés sur la modélisation statistique des images RSO polarimétriques (Lee et al., 1999 ; Bombrun, et al., 2011 ; Bombrun & Beaulieu, 2008). Il existe en outre plusieurs travaux combinant les approches ci-dessus (Lee et al., 1999 ; Formont, et al., 2011). De plus, certaines techniques de traitement d'image ont été employées (Tan et al., 2007 ; Ince, et al., 2012. Zhang, et al., 2009 ; Ince, 2010). Certains auteurs ont également utilisé la classification de données polarimétriques et multifréquences (Chen et coll., 1996 ; Kahny & Wiesbeck, 1991 ; Fukuda & Hirasawa, 1999). Cependant, toutes ces approches présentent des inconvénients même si elles offrent des perspectives. En particulier, le système RSO polarimétrique complet fournit des informations supplémentaires sur les mécanismes de diffusion, non accessibles avec une seule ou deux polarisations.

Dans ce cadre, cette thèse considère le potentiel d'une image RSO polarimétrique (POLSAR) complexe monovue pour la cartographie urbaine. Nous visons à produire une carte à haute résolution spatiale qui montre les principales surfaces urbaines: différents types de toitures, arbres, pelouses,

routes, etc. À cette fin, nous utilisons une image en bande X. En effet, cette courte longueur d'onde est utilisée pour la classification, car elle est supposée présenter une plus grande sensibilité à la rugosité de surface.

Un nouvel algorithme de classification supervisée est proposé. Il combine l'approche du maximum de vraisemblance basée sur la fonction de probabilité Fisher (F-pdf) et les outils morphologiques (Figure 1). L'outil morphologique permet d'ajouter des informations spatiales comme une contrainte au cours du processus de classification. La classification s'appuie sur un ensemble de descripteurs polarimétriques et texturaux. Les descripteurs polarimétriques sont tirés selon diverses théories. Les descripteurs statistiques ou texturaux servent à préciser les différents comportements statistiques. Cependant, cet algorithme suppose l'indépendance statistique des attributs. Pour cette raison, nous proposons ensuite d'utiliser le critère de « redondance minimale et pertinence maximale » pour déterminer un ensemble optimal d'attributs (Ludwig & Nunes, 2010 ; Peng et al., 2005). Nous utilisons un algorithme génétique pour sélectionner cet ensemble.

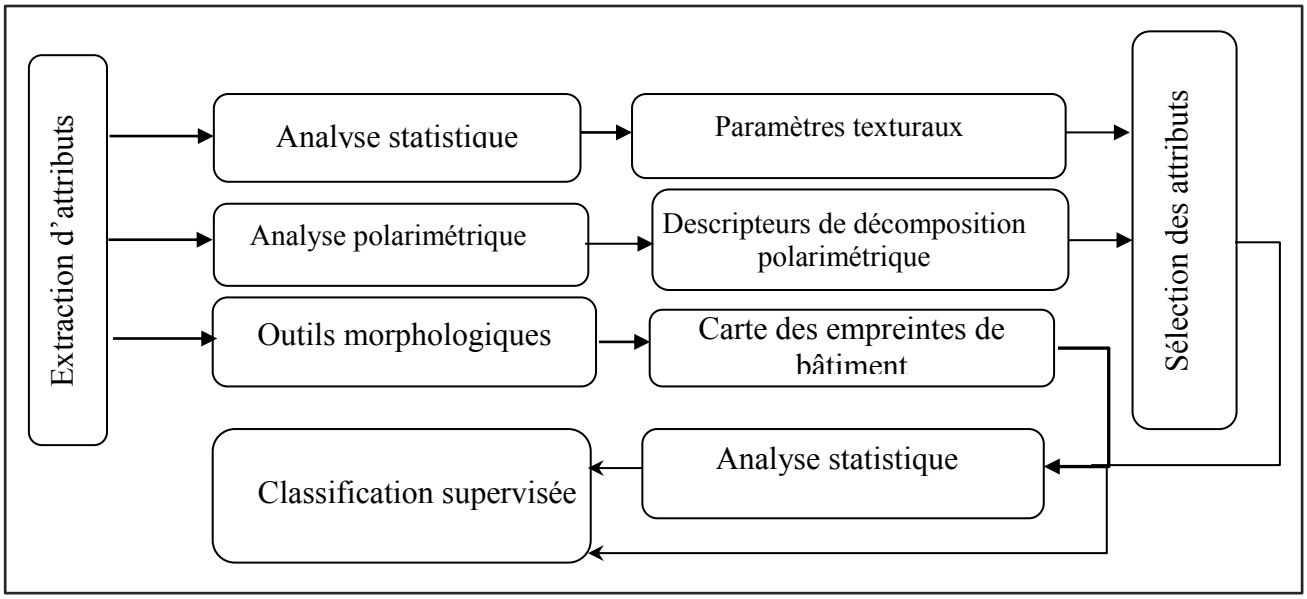


Figure 1 : Description de l'approche suivie

Méthodes

Présentation de la donnée

Dans notre travail, nous utilisons une image polarimétrique monovue en bande X sur Toulouse, acquise en 2006 par le capteur aéroporté RAMSES de l'ONERA, avec 60° d'incidence et 35 cm de taille de pixel (Figure 2). La résolution spatiale est d'environ 50 cm.

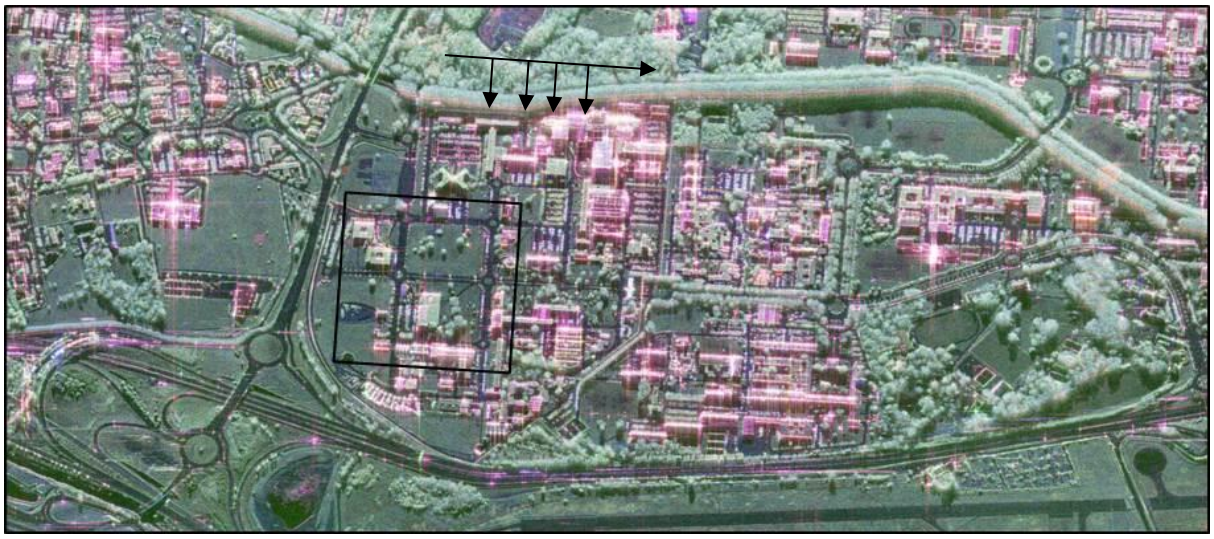


Figure 2: Image POLSAR sur la zone d'étude (Pauli RVB), la trajectoire et le sens d'éclairement sont représentés par des flèches noires.

Nos classes sont la voirie, les pelouses, les arbres, les toits plats et les toits pentus. Nous considérons également deux classes résultant des effets de l'imagerie radar: ombres et pixels brillants. Les échantillons d'apprentissage et les échantillons de contrôle sont définis à partir de la photographie aérienne et de l'image POLSAR.

Extraction des attributs

Dans ce travail, un cadre est mis en place pour définir les différentes décompositions à extraire. Comme la méthode de décomposition incohérente utilise une moyenne du signal rétrodiffusé et la méthode de décomposition cohérente utilise l'image en pleine résolution, ces deux méthodes se complètent (Figure 3). Dans notre travail, nous choisissons comme descripteurs les matrices de Sinclair et de cohérence, les descripteurs de décompositions cohérentes et incohérentes, des discriminateurs basés sur une définition statistique selon ces deux méthodes et des descripteurs texturaux.

Tout d'abord, plusieurs attributs à valeurs réelles et positives sont extraites de données POLSAR de deux façons: (1) extraction directe à partir de la matrice de rétrodiffusion ou à l'aide de méthodes de décomposition en mode amplitude ; (2) descripteurs texturaux de l'uniformité et du contraste basés sur les matrices de co-occurrence (Schistad Solberg & Jain, 1997).

Ces paramètres texturaux sont extraits de matrices de co-occurrence normalisées qui sont calculées avec une distance inter-pixel d = 1, une orientation $\theta = 0$ °, 32 niveaux de gris et une taille de la fenêtre de 5x5, sur les images des racines carrées de T_{22} et T_{33}. La liste complète des descripteurs réels positifs est présentée dans le Tableau 1.

Ensuite, nous avons extrait plusieurs attributs à valeurs réelles, positives ou non, des données POLSAR, de deux façons: directement dérivés de la matrice de rétrodiffusion ou à l'aide de méthodes

de décomposition (Lee & Pottier, 2009). La liste complète des descripteurs est présentée dans le Tableau 2.

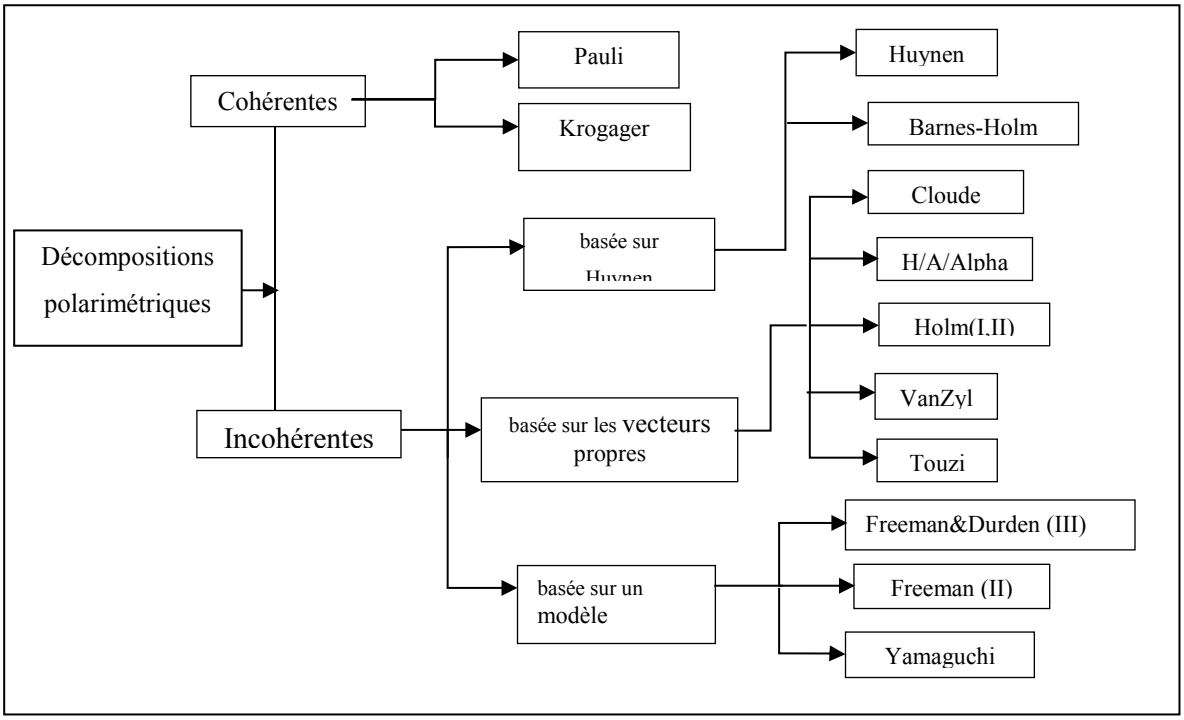


Figure 3: Classification des principales décompositions utilisées dans ce travail

Les indices utilisés dans ce tableau représentent l'élément associé à la catégorie correspondante. Par exemple, la matrice $[Barnes\ I]_A$ est liée à la matrice Barnes-I en amplitude, (S_{12}) fait référence au terme d'amplitude de la première ligne et la deuxième colonne de la matrice de Sinclair, ($T_{13\text{-Phase}}$) se rapporte au terme de phase de la première ligne et la troisième colonne de la matrice de cohérence, $[\text{Barnes I}]_{\text{I, phase}}$, se réfère à la phase des différents éléments dans la matrice de Barnes en intensité (en dB) et ainsi de suite.

Ces descripteurs ont été calculés en utilisant POLSARpro (Pottier, et al., 2005) et des codes en langage MATLAB.

Tableau 1: Descripteurs polarimétriques à valeurs réelles positives pris en considération dans l'analyse statistique. « A » correspond à l'amplitude du descripteur.

Technique d'extraction des descripteurs	Descripteur	#
Méthode de décomposition	$[H/A/Alpha]_A$, $[Huynen]_A$, $[Cloude]_A$, $[Holm\ I]_A$ $[Holm\ II]_A$, $[Barnes\ I]_A$, $[Barnes\ II]_A$	42
Descripteurs polarimétriques originaux	$\{T_{11\text{-}A}, T_{12\text{-}A}, T_{22\text{-}A}, T_{23\text{-}A}, T_{13\text{-}A}\}$	5
	$\sqrt{Span}$	1
	$S_{A,11}, S_{A,12}, S_{A,22}$	3
	$\{\lvert S_{VV}\rvert / \lvert S_{HH}\rvert, \lvert S_{HV}\rvert / \lvert S_{HH}\rvert, \lvert S_{VV}\rvert / \lvert S_{HV}\rvert\}$	3
Paramètres de texture	$UNI-\sqrt{T_{22-A}}$, $UNI-\sqrt{T_{33-A}}$,	4
	$CON-\sqrt{T_{22-A}}$, $CON-\sqrt{T_{33-A}}$	
Nombre total de descripteurs	58	

Tableau 2: Descripteurs polarimétriques considérés. < > désigne la moyenne. S_{xy}représente le coefficient de rétrodiffusion de la cible, *x* la polarisation du champ incident (H ou V), *y* la polarisation du champ rétrodiffusé (H ou V) (Lee &Pottier, 2009; Soheili Majd, 2014)

Technique d'extraction des descripteurs	Descripteur	#
Méthode de décomposition	$[Huynen]_{T_{I,\ phase}}$, $[H/A/\alpha]_{T_{I,\ phase}}$, $[Freeman\ I]_{s,d,v}$, $[Freeman\ II]_{v,G}$, $[Yamaguchi\ VI]_{s,d,v,h}$ $[Barnes\ I,II]_{T_{I,\ phase}}$, $[Krogager]_{s,d,h}$, $[Yamaguchi\ III]_{s,d,v}$, $[VanZyl]_{s,d,v}$, $[Pauli]_{\lvert a\rvert^2,\lvert b\rvert^2,\lvert c\rvert^2}$ $[Huynen]_{T_{I,\ phase}}$, $[Barnes\ I,II]_{T_{I,\ phase}}$, $[Holm\ I,II]_{T_{I,\ phase}}$, $[Cloude]_{T_{I,\ phase}}$	63
	$\{H/A/\alpha, H(1\text{-}A), (1\text{-}H)A, HA, (1\text{-}H)(1\text{-}A), \beta, \gamma, \delta\}$	10
	$\{\lambda_i, \psi_i, \tau_i, \alpha_{s\ i}, \phi_{\alpha s\ i}, (i=1,3)\}$	15
Descripteurs polarimétriques originaux	$\{T_{12\text{-}I}, T_{13\text{-}I}, T_{23\text{-}I}, T_{22\text{-}I}, T_{11\text{-}I}, T_{33\text{-}I}, (T_{12\text{-}Phase}), (T_{13\text{-}Phase}), (T_{23\text{-}Phase})\}$	9
	$2\langle S_{HV} S_{HV}^* \rangle / (\langle S_{HH} S_{HH}^* \rangle + \langle S_{VV} S_{VV}^* \rangle)$	1
	$\langle \lvert S_{xy}\rvert^2 \rangle / \langle \lvert S_{x'y'}\rvert^2 \rangle \quad (x, y, x', y' = (H,V))$	3
	$S_{A,11}, S_{A,12}, S_{A,22}$	3
	$DR = \langle S_{HH} - S_{HV} \rangle, \langle S_{HH} - S_{VV} \rangle, \langle S_{HV} - S_{HH} \rangle, \langle S_{VV} - S_{HH} \rangle, \langle S_{VV} - S_{HV} \rangle, \langle S_{HV} - S_{VV} \rangle$	6
Discriminateurs	*CC, SE_I, SE_P, DoP, CPD, XPD, Span, CCC_{phase}, CCC_A, PF, PH, A_{12}, LA, KA, PA, SERD et DERD*	17
Nombre total de descripteurs	127	

Extraction de caractéristiques morphologiques

En outre, une méthode d'extraction de bâtiments dans l'image haute résolution est utilisée. Un tel descripteur de bâtiment peut améliorer la classification en milieu urbain. L'objectif principal est d'accéder à une carte du bâti sans utiliser de données auxiliaires telles qu'un MNE (Modèle Numérique d'Elévation). La solution semi-automatique qui est proposée ici peut produire une carte qui est censée représenter les empreintes de bâtiment. Pour produire une image binaire, nous pouvons utiliser la méthode de seuillage d'Otsu (Otsu, 1979). Nous pouvons aussi convertir l'image de niveaux de gris en image binaire en définissant manuellement le seuil approprié. Ensuite, nous supprimons les zones dont le nombre de pixels est inférieur à un seuil avec une ouverture morphologique sur l'image binaire. L'effet de cette opération est de supprimer toutes les structures qui sont trop petites et plus lumineuses que leur environnement, comme les points brillants ou le bruit. Ainsi les rues et chemins de fer présentant des points brillants sont éliminés. Puis, en appliquant 'fermeture moins ouverture' avec un élément structurant carré, la taille réelle des objets est maintenue, le bruit est supprimé et les trous sont comblés dans les zones d'intérêt.

Sélection des attributs

Information Mutuelle

L'information mutuelle (IM) est une mesure de dépendance entre variables aléatoires dans le cas de la corrélation non linéaire, sans faire d'hypothèse sur leurs distributions. En général, l'IM décrit quantitativement la relation existant entre deux descripteurs ou un descripteur et une classe. La notation pour la description de l'IM basée sur la théorie de l'information de Shannon (Shanon, 1948) provient de (Ludwig & Nunes, 2010).

Les mesures de l'entropie H d'une variable aléatoire X avec la fonction de densité de probabilité (pdf) et de l'Entropie conjointe de X et Y avec une fonction de densité de probabilité, $p_{\mathrm{x,y}}(\mathrm{x}_i,\mathrm{y}_j)$, sont définies par:

$$H(X) = -\sum_i p_x(\mathrm{x}_i)\log p_x(\mathrm{x}_i),\ H(X,Y) = -\sum_i\sum_j p_{\mathrm{x,y}}(\mathrm{x}_i,\mathrm{y}_j)\log p_{\mathrm{x,y}}(\mathrm{x}_i,\mathrm{y}_j) \quad (1)$$

L'IM est une mesure de la dépendance mutuelle de deux variables, basée sur l'entropie et définie par:

$$MI(X;Y) = H(X)+H(Y)-H(X,Y) = \sum_i\sum_j\left[p_{\mathrm{x,y}}(\mathrm{x}_i,\mathrm{y}_j)\log\left(\frac{p_{\mathrm{x,y}}(\mathrm{x}_i,\mathrm{y}_j)}{p_{\mathrm{x}}(\mathrm{x}_i).p_{\mathrm{y}}(\mathrm{y}_j)}\right)\right] \quad (2)$$

L'objectif principal de la sélection d'un ensemble d'attributs est d'atteindre la pertinence maximale entre un attribut sélectionné et une classe, notée, $D(S_m,C)$, et la redondance minimale de l'attribut sélectionné notée $R(S_m)$.

Selon l'équation (2), la pertinence, la redondance et la fonction d'évaluation mRMP sont définies par :

$$D(S_m, C) = \frac{1}{|S_m|} \sum_{x_i \in \Omega_{S_m}} MI(X_i; C), \quad |S_m| = m$$
$$R(S_m) = \frac{1}{|S_m|^2} \sum_{\substack{x_i, x_j \in \Omega_{S_m} \\ i \neq j}} MI(X_i; X_j) \quad \Rightarrow \hat{S}_m = \arg\max_{S_m} \left[U(S_m) = D(S_m) - R(S_m) \right] \tag{3}$$

où la variable X_i est un attribut et la variable C est une classe. $S_m = \{X_1, ..., X_m\}$ est l'ensemble des attributs, et $c = \{c_1, c_2, ..., c_k\}$ est l'ensemble des étiquettes de classes avec k classes. Ω_{S_m} est le domaine de la variable S_m.

Algorithme Génétique

L'algorithme génétique (AG) est l'un des algorithmes itératifs de recherche les plus usuels pour optimiser une fonction d'évaluation. L'AG est inspiré de l'évolution naturelle dérivée de la théorie darwinienne de la survie du plus adapté et a été présenté par John Holland dans les années 1970. Un processus génétique de sélection d'un ensemble d'attributs consiste en quatre étapes fondamentales: une génération de population, un processus de sélection des individus, un arrêt (il peut être fondé sur le nombre d'itérations, le nombre d'attributs, un seuil de convergence), ainsi qu'une validation du résultat pour vérifier la validité de la solution retenue (Vinh et al., 2012). Notre méthode de sélection (notée MSS) suit ce schéma classique. Un individu est un ensemble comprenant un certain nombre d'attributs choisi parmi tous les attributs définis précédemment. Ici, la fonction d'évaluation est le mRMP (minimum de redondance et maximum de pertinence) calculé à partir de la théorie de l'IM. La mise en œuvre de l'AG suit ici les étapes suivantes:

Initialisation des paramètres: cinq paramètres sont initialisés: la taille de la population ou le nombre d'individus dans une génération, le nombre maximal de générations (itérations), le facteur de pression qui définit la force de la méthode de sélection et l'influence de l'espace de recherche, la taille de l'ensemble d'attributs à rechercher et le seuil d'évaluation.

Évaluation: la fonction d'évaluation repose sur le mRMP (Ludwig & Nunes, 2010), Equation 3. À cet effet, la fonction d'évaluation pour chaque individu de la population actuelle est calculée. Ensuite, l'algorithme trie les individus selon leur niveau d'adéquation (classement des indices des individus de haut en bas).

Sélection: les meilleurs individus sont sélectionnés pour le croisement ou la recombinaison, si leur valeur est supérieure au seuil d'évaluation.

Croisement ou Recombinaison: les individus sélectionnés sont introduits de manière aléatoire dans la nouvelle génération par l'opérateur de probabilité de croisement. Celui-ci est défini par une distribution asymétrique (Equation 4):

$$f(r, a, i) = round\left[(i-1) \frac{e^{ar} - 1}{e^{a} - 1} \right], i \in [1 \; N] \tag{4}$$

où $r \in [0,1]$ est une variable aléatoire avec une distribution uniforme, N est le nombre d'individus, i est l'indice individuel, et a est une constante positive qui est connue comme facteur de pression (Ludwig & Nunes, 2010).

Processus itératif : Une nouvelle génération est traitée de manière itérative selon trois étapes: 1) évaluation basée sur la fonction d'évaluation mRMP, qui conduit à la génération de réarrangement selon le classement des valeurs (sélection des individus) ; 2) Croisement (sélection de nouveaux attributs) ; 3) poursuite de la recherche ou arrêt. Simplement, la procédure de recherche continue (évaluation, sélection, croisement) jusqu'à ce que le critère d'arrêt soit rempli. Ici, le critère d'arrêt est un nombre prédéterminé de générations.

Validation du résultat: Le résultat de la sélection par l'AG peut être comparé à d'autres méthodes de sélection d'attributs. Plus de détails peuvent être trouvés dans (SoheiliMajd, 2014).

La Figure 4 représente le diagramme de sélection d'attributs par MSS. L'algorithme est mis en œuvre dans Matlab et adapté à partir des codes source fournis par (Ludwig & Nunes, 2010).

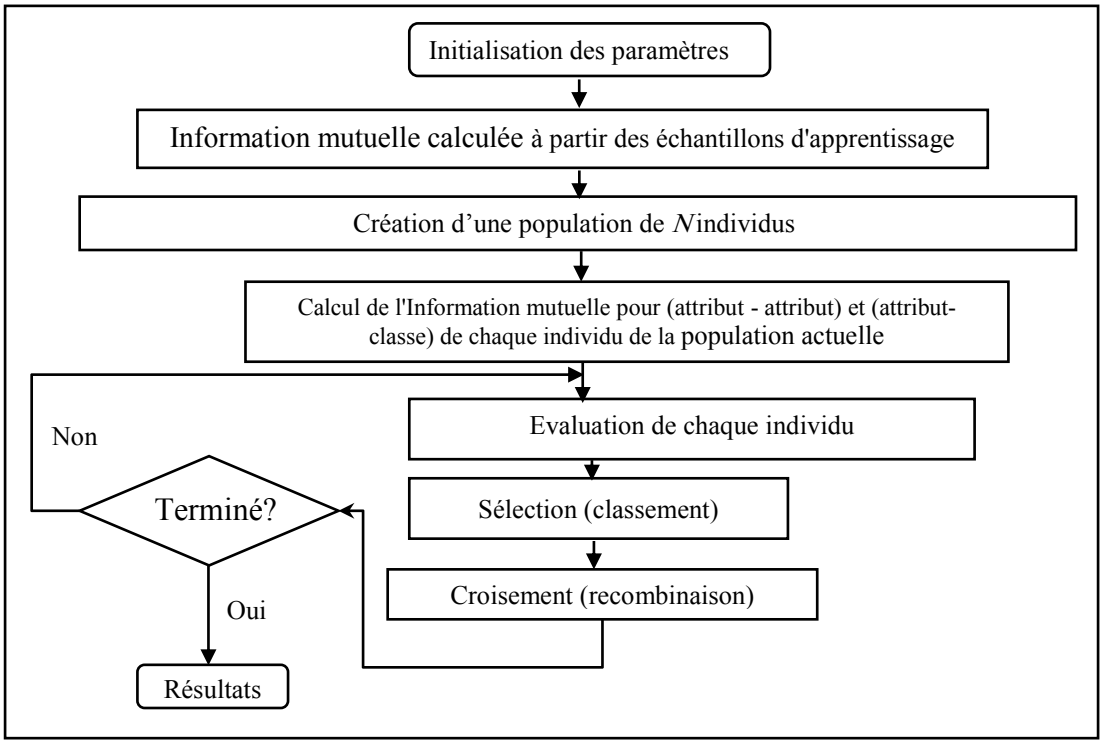


Figure4: Diagramme du sélecteur d'attributs MSS par AG-IM

Analyse statistique

La modélisation statistique décrit les caractéristiques statistiques d'une variable à partir d'expressions mathématiques telles que la fonction de densité de probabilité (pdf), les moments statistiques et une estimation des paramètres de la pdf. Généralement, le processus de modélisation statistique consiste en la sélection de plusieurs modèles de distribution statistique connue, l'estimation des paramètres de la distribution et l'évaluation de la qualité de l'ajustement des modèles.

Pour ce faire, nous utilisons le cadre de transformation de Mellin qui consiste en la définition de la transformation de Mellin, de ses propriétés et des statistiques de deuxième type de Mellin (log-moments et log-cumulants).

La transformation de Mellin est un outil d'analyse pour l'étude de la distribution de variables positives (Epstein, 1948), qui a été soulignée par Robert Hjalmar Mellin (1854-1933). La transformation de Mellin uni-variable aléatoire a été d'abord développée pour la modélisation statistique des images radar et a été présentée comme statistique de deuxième type par (Nicolas, 2002 ; Nicolas, 2006). Un bon aperçu de la note historique de transformation de Mellin est donné par (Anfinsen, 2010). On trouvera plus de détails sur les définitions mathématiques, y compris une liste de propriétés fondamentales, quelques exemples et les transformations de Mellin des fonctions courantes de distribution dans (Poularikas, 1999 ; Delphine & Bhatta, 2007 ; Bertrand, et al., 2000 ; Nicolas, 2006).

Ce modèle statistique sert ensuite dans le processus de classification. En outre, cette analyse permet de vérifier la possibilité d'une discrimination de chaque type de couverture terrestre selon son comportement statistique.

Méthodes de classification

Nous considérons certains algorithmes classiques pour la classification d'une image POLSAR monovue. Comme des données de référence ont déjà été acquises sur la zone d'étude, les méthodes proposées sont supervisées. Nous utilisons l'algorithme de classification par maximum de vraisemblance (MV) basé sur un modèle gaussien dans le cas de variables aléatoires univariées (G-ML) et multivariées (G-ML^{C}) en utilisant les descripteurs issus de la matrice de cohérence. Dans ce qui suit, nous proposons la classification MV basée sur la pdf de Fisher (F-ML). Grâce aux outils morphologiques permettant d'extraire les attributs des bâtiments, l'algorithme de classification F-ML est amélioré. La classification F-ML suppose cependant que les descripteurs soient indépendants. Toutefois, cela peut ne pas être le cas, et pour y remédier, nous utilisons la MSS pour garder ceux qui sont non redondants. En outre, dans cette étude, la SVM (Support Vector Machine) qui est une méthode non paramétrique, est utilisée comme algorithme de classification de référence.

Classification MV basée sur la pdf de Fisher

Dans ce travail, nous proposons une nouvelle méthode de classification par maximum de vraisemblance, basée sur la pdf de Fisher. En effet, nous montrons dans ce qui suit que nos attributs ont une distribution de Fisher.

En partant du principe que chaque classe a une distribution de Fisher, chaque classe peut être caractérisée par les paramètres de Fisher ($\mu_{k,b}, L_{k,b}, M_{k,b}$), où $b \in \{1, 2, ..., N\}$, N est le nombre de dimensions de l'espace (nombre de descripteurs) et k l'index de la classe.

Compte tenu de ces trois caractéristiques pour chaque classe, la probabilité statistique est calculée pour chaque classe dans l'espace à N dimensions pour déterminer l'appartenance des pixels à une classe. Les

probabilités a priori des classes peuvent être définies en option. Chaque objet est rangé dans la classe pour laquelle il a la plus forte probabilité d'appartenir.

Selon cette hypothèse, l'adhésion du pixel (i, j) à la classe k est :

$$P(k, i, j) = \sum_{b=1}^{N} \log(p\left[\mu_{k,b}, L_{k,b}, M_{k,b}\right](u_b(i, j))) \tag{5}$$

où, $p\left[...\right](u_b(i, j))$ est la pdf de Fisher pour la b-ième caractéristique du pixel (i,j) dont la valeur est $u_b(i, j)$. Selon la règle de décision (Equation 5), la fonction discriminante devient :

$$C(i, j) = \underset{k \in \{1, ..., N_c\}}{\arg\max}\left[P(k, i, j)\right] \tag{6}$$

où N_c est le nombre de classes et $C(i, j)$ est la classe attribuée au pixel (i, j). La Figure 5 montre la classification MV basée sur la pdf de Fisher.

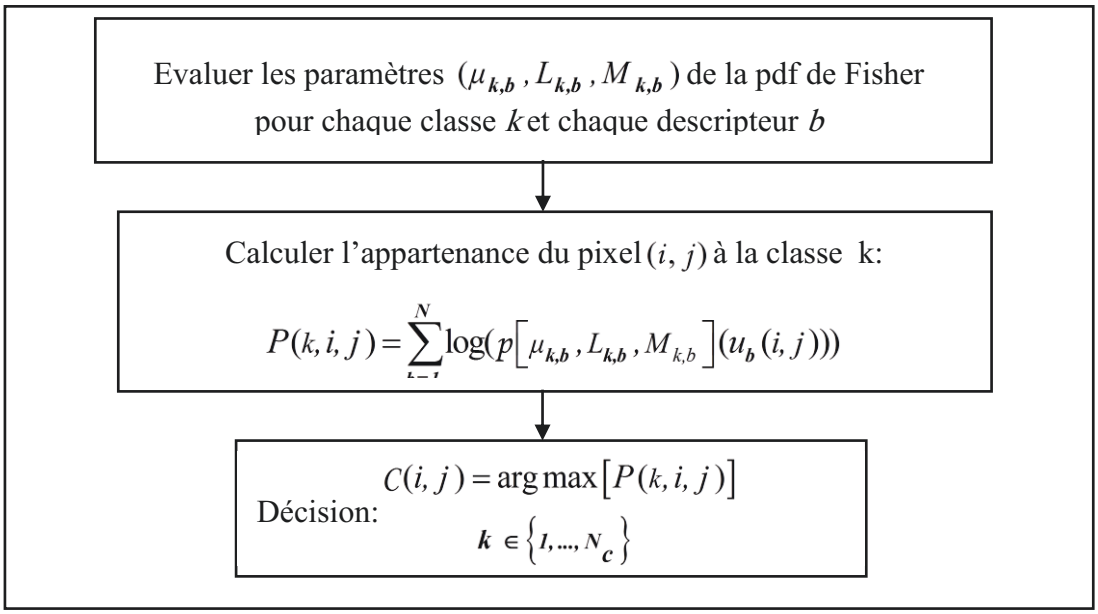


Figure 5 : Classification MV basée sur la pdf de Fisher

Classification F-MVcouplée à une description morphologique

Tous les cadres de classification considérés dans notre travail présentent des inconvénients en raison des mécanismes de rétrodiffusion complexes dans l'environnement urbain. Par exemple, d'après les premières expériences, nous remarquons qu'un bâtiment non aligné avec la direction azimutale produit une diffusion volumique, beaucoup de surfaces rugueuses produisent une diffusion volumique et quelques zones végétalisées donnent lieu à une rétrodiffusion de surface. Pour améliorer les résultats de la classification et réduire les ambiguïtés, nous proposons une classification par maximum de

vraisemblance enrichieà l'aide d'outils morphologiques. L'information morphologique est définie par :

$$f(i,j) = \begin{cases} 0, & \forall i,j \in C_{\text{background}} \\ 1, & \forall i,j \in C_{\text{object}} \end{cases}$$

Soient C_1 et C_2 deux ensembles de pixels formant deux classes, $C \in \{1,\ldots,C_k\}$, où, $k \in \{1,\ldots,N_c\}$, N_C est le nombre de classes. Supposons que ces deux classes aient des ambiguïtés. Supposons que $C_1 \subset C_{object}$ et nous voulons que $C_2 \cap C_{object} = \varnothing$. L'appartenance de chaque pixel (i,j) à la classe k est notée $P_k(i,j)$. Pour introduire la contrainte dans la procédure de classification par maximum de vraisemblance, nous avons:

Etape 1: calculer la valeur de contrainte :

Contrainte= $\min\{P_{C_1}(i,j), \quad \forall i,j \in C_2\}$

Etape 2: $P_{C_2}(i,j) = \text{Constrainte}, \quad \forall i,j \in C_{object}$

En fait, la valeur minimale de $P_{C_1}(i,j)$ est affectée à l'appartenance $P_{C_2}(i,j)$ et liée à des informations de la morphologie des objets. L'algorithme de classification supervisée enrichi est résumé dans la Figure 6.

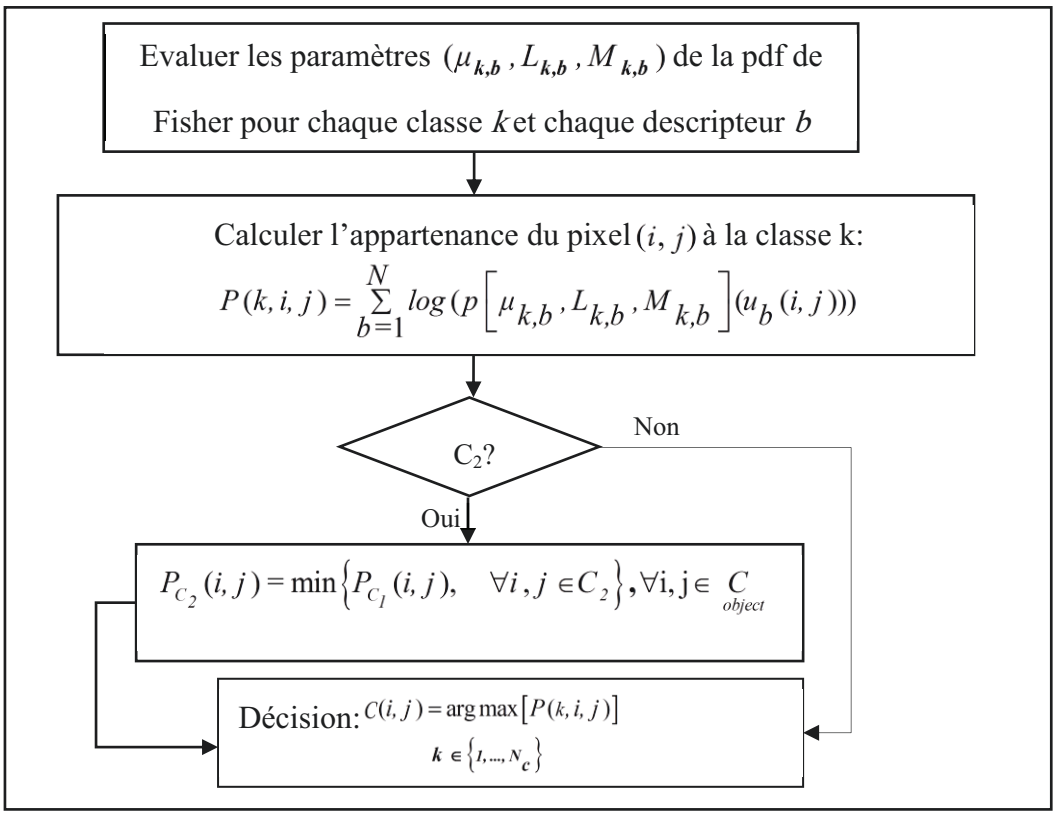


Figure 6: Classification F-MV couplée avec un descripteur morphologique

Résultats Expérimentaux

Initialisation des paramètres de l'AG

Pour obtenir le nombre optimal de générations (itérations), nous considérons la relation entre le nombre d'itérations et la valeur maximale obtenue par la fonction évaluation pour différentes tailles d'ensembles d'attributs (3, 4, 5,..., jusqu'à 20). La figure 7 (a) montre qu'après plusieurs itérations pour toute taille définie, la valeur de la fonction d'évaluation ne varie plus. Pour sélectionner le nombre de descripteurs (c'est la taille d'un ensemble), la précision de la classification (précision globale) pour différentes tailles de l'ensemble (3, 4, 5,..., jusqu'à 20) est estimée. La figure 7 (c) montre que le nombre optimal d'attributs pourrait être cinq, conduisant à une précision globale de 83,9 % environ et un nombre d'itérations de 16.

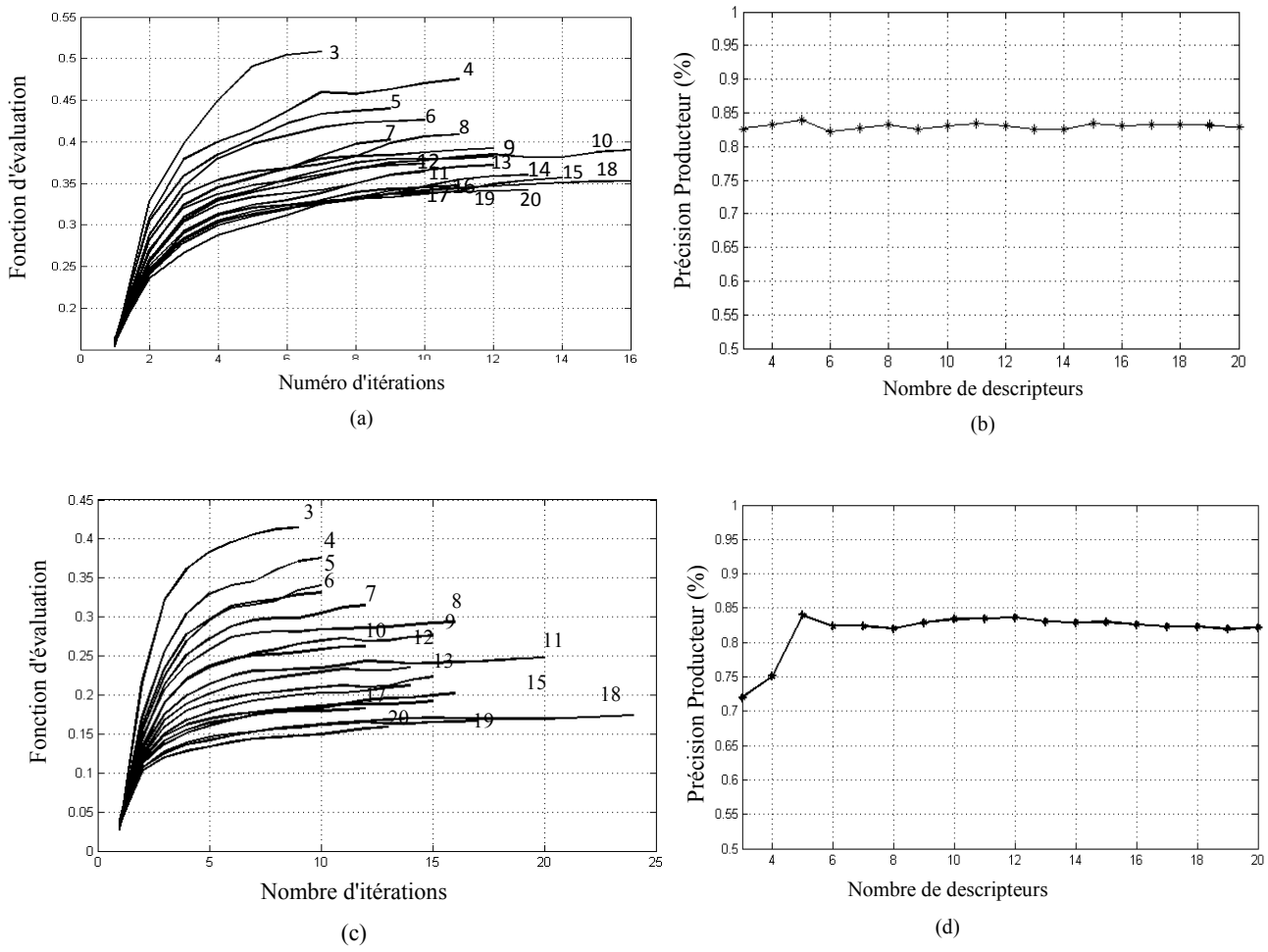


Figure 7: (a) valeurs de la fonction d'évaluation vs nombre d'itérations pour des ensembles de 3 à 20 descripteurs réels positifs. (b) précision globale d'une classification MV pour une loi gaussienne univariée vs nombre de descripteurs réels positifs. (c) Nombre d'itérations vs valeur de la fonction d'évaluation pour des ensembles de 3 à 20 descripteurs, obtenus par le MSS à partir de 127 descripteurs. (d) précision globale de la classification par SVM vs nombre de descripteurs par ensemble.

Résultat de la sélection par AG-IM

La sélection de l'ensemble optimal parmi les 58 descripteurs positifs réels donne cinq descripteurs, donnés au Tableau 3. Le résultat de la sélection de l'ensemble de 3 à 6 des 127descripteurs est indiqué dans le Tableau 4. Selon les remarques précédentes, l'ensemble optimal peut désigner l'ensemble avec 5 descripteurs. En outre, la Figure 8 et la Figure 9 indiquent les valeurs d'IM entre chaque classe et chaque descripteur retenu dans les ensembles optimaux.

Tableau 3: Liste des descripteurs positifs réels de l'ensemble optimal par la MSS à partir de 58descripteurs.

# Of Des.	1	2	3	4	5
Descripteur	$\lvert S_{HV}\rvert / \lvert S_{HH}\rvert$	$\sqrt{Span}$	$CON\text{-}\sqrt{T_{22\text{-}A}}$	$CON\text{-}\sqrt{T_{33\text{-}A}}$	$\lvert S_{VV}\rvert / \lvert S_{VH}\rvert$

Tableau 4: Liste des descripteurs sélectionnés par MSS pour des ensembles de 3, 4, 5 ou 6 descripteurs à partir de 127 descripteurs.

#	Descripteur
3	[*Barnes I*]-$T_{12\text{-}phase}$, *Polarization-Fraction* (*PF*) ,*Span*
4	*Shanon-Entropy-I* (SE_I), *Anisotropy-Lueneburg* (*AL*), $T_{33\text{-}I}$,[*Cloude*]$T_{12\text{-}phase}$
5	*Shanon-Entropy-I* (SE_I), *H*(*1-A*), $T_{33\text{-}I}$, $T_{12\text{-}phase}$, [*H/A/α*]-$T_{11\text{-}I}$
6	*Span*, $T_{33\text{-}I}$, A_{12}, $T_{11\text{-}I}$,[*Holm I*]-$T_{23\text{-}phase}$,[*Barnes I*]-$T_{12\text{-}phase}$

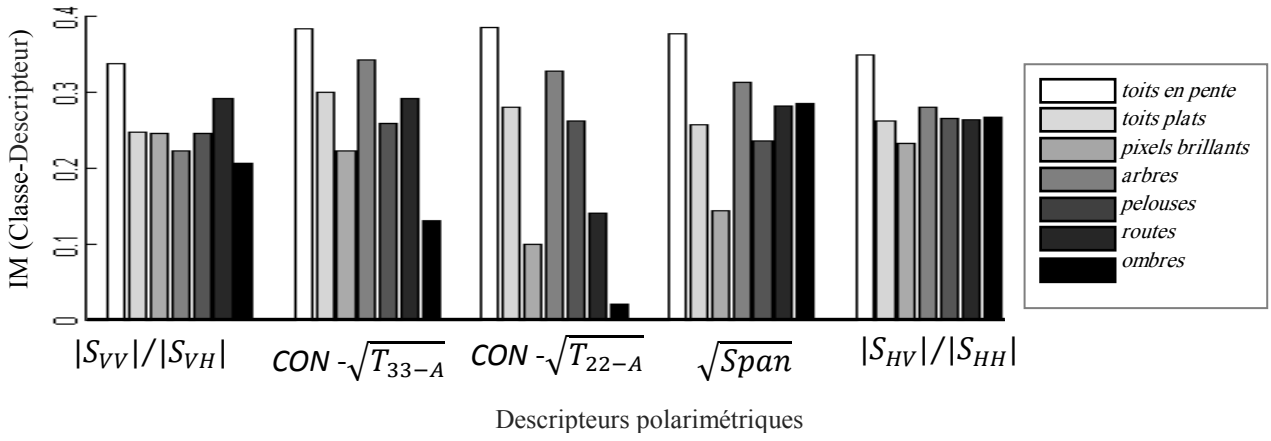


Figure 8: Valeurs d'information mutuelle entre la classe et le descripteur de l'ensemble optimal de 5 descripteurs (Tableau 3) par MSS à partir de de 58 descripteurs positifs réels (Tableau 1).

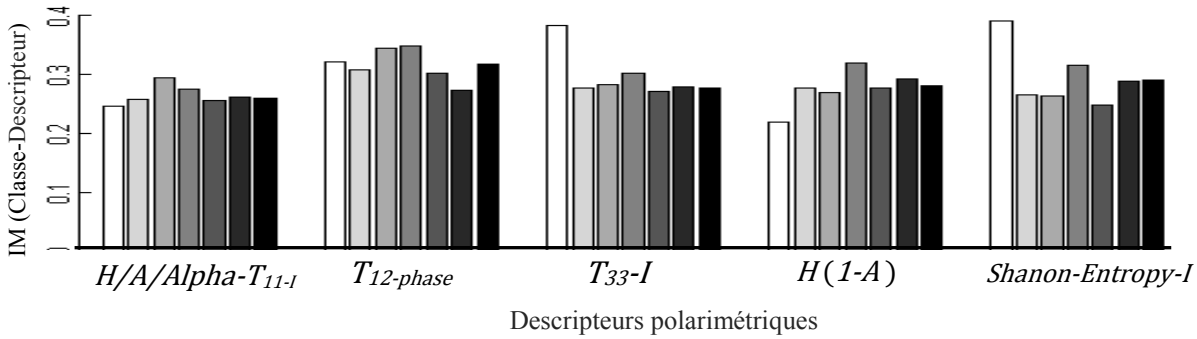


Figure 9: Valeurs d'information mutuelle entre la classe et le descripteur de l'ensemble optimal de 5 descripteurs (Tableau 4) par MSS à partir de 127 descripteurs (Tableau 2).

Résultats de l'analyse statistique

L'expérience statistique est effectuée sur l'ensemble optimal (Tableau 3) sélectionné par à partir de 58 descripteurs positifs réels (Tableau 1). Elle consiste en trois étapes :

- Extraction des échantillons d'apprentissage dans une région représentative de chaque classe (*ombres*, *routes*, *pelouses, arbres*, *pixels brillants, toits plats* et *toits en pente*), dans l'image RSO.
- Estimation des paramètres des distributions de Fisher, $\mathcal{K}$, Gamma, Beta I, Log-normale et Weilbull par MV et MoCL.
- Evaluation du GoF (adéquation du modèle) par distance de K-S , coefficients de corrélation et erreur RMS.

Pour estimer les paramètres des pdf, nous avons écrit un code source selon la méthode suggérée par (Nicolas, 2006) et l'avons appliqué à différents types de pdf (Fisher, Weibull, $\mathcal{K}$, Gamma, Beta I, Log-normale).

Les résultats expérimentaux de l'analyse statistique montrent que pour certains modèles bien connus tels que $\mathcal{K}$, Weibull, Log-normal, Beta I et Gamma, la pdf de Fisher est capable de modéliser différentes surfaces (par exemple, différentes surfaces en zone urbaine ainsi que des artefacts du radar comme les ombres et les pixels brillants). La méthode d'estimation MoLC est particulièrement appropriée pour tenir compte des trois paramètres de la pdf de Fisher (échelle et forme) . Selon les résultats de GoF, cette méthode peut conduire à de meilleurs résultats avec des données empiriques par rapport à la méthode d'estimation par MV (SoheiliMajd, 2014).

Les résultats de la comparaison visuelle entre la pdf de Fisher et les données empiriques confirment cette conclusion, comme illustré à la Figure 10 et la Figure 11.

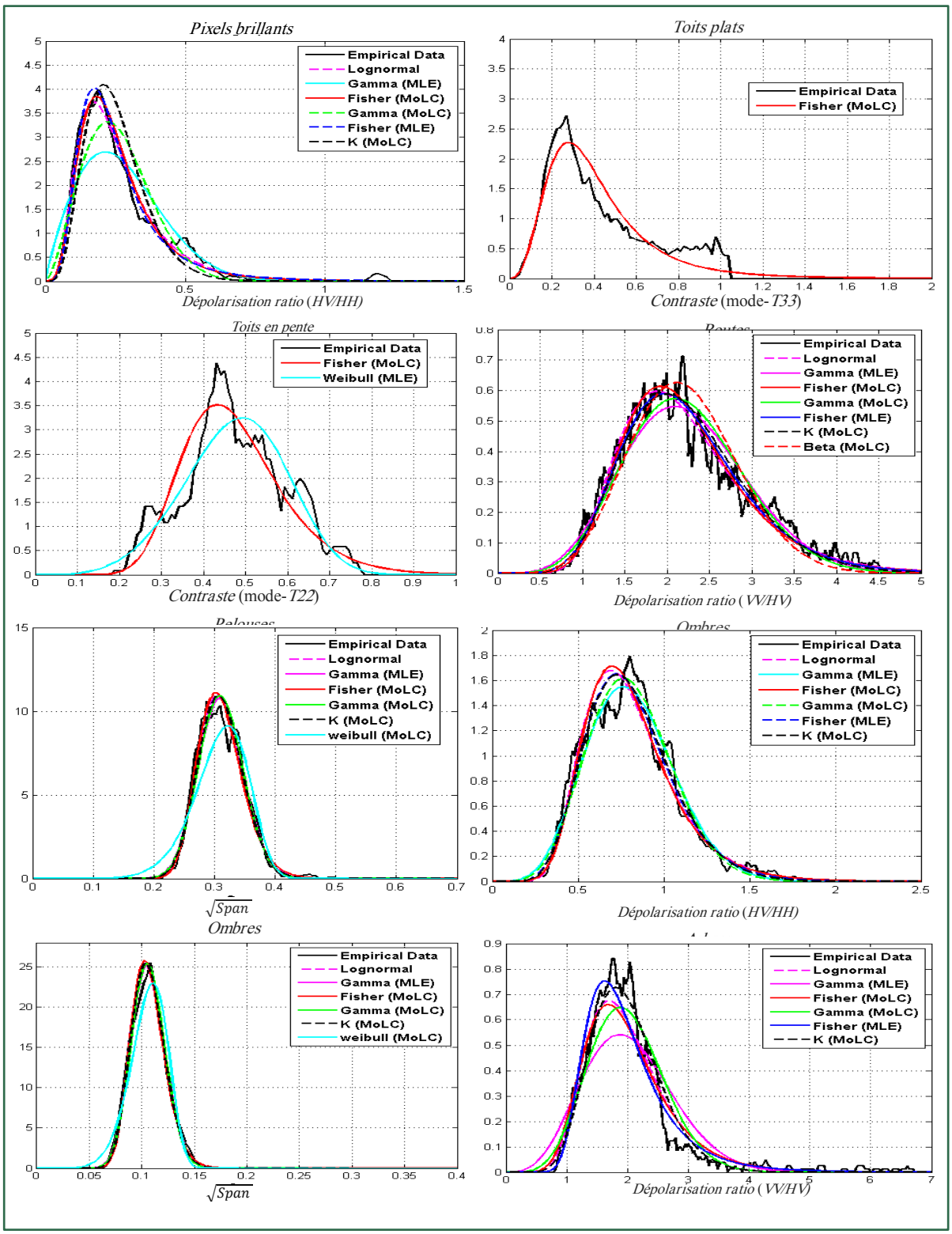


Figure10:Données empiriques et pdf estimées par MV et MoLC pour différentes classes sur différents ensembles optimaux de descripteurs (Tableau 3) sélectionnés par MSS à partir de 58 descripteurs positifs réels.

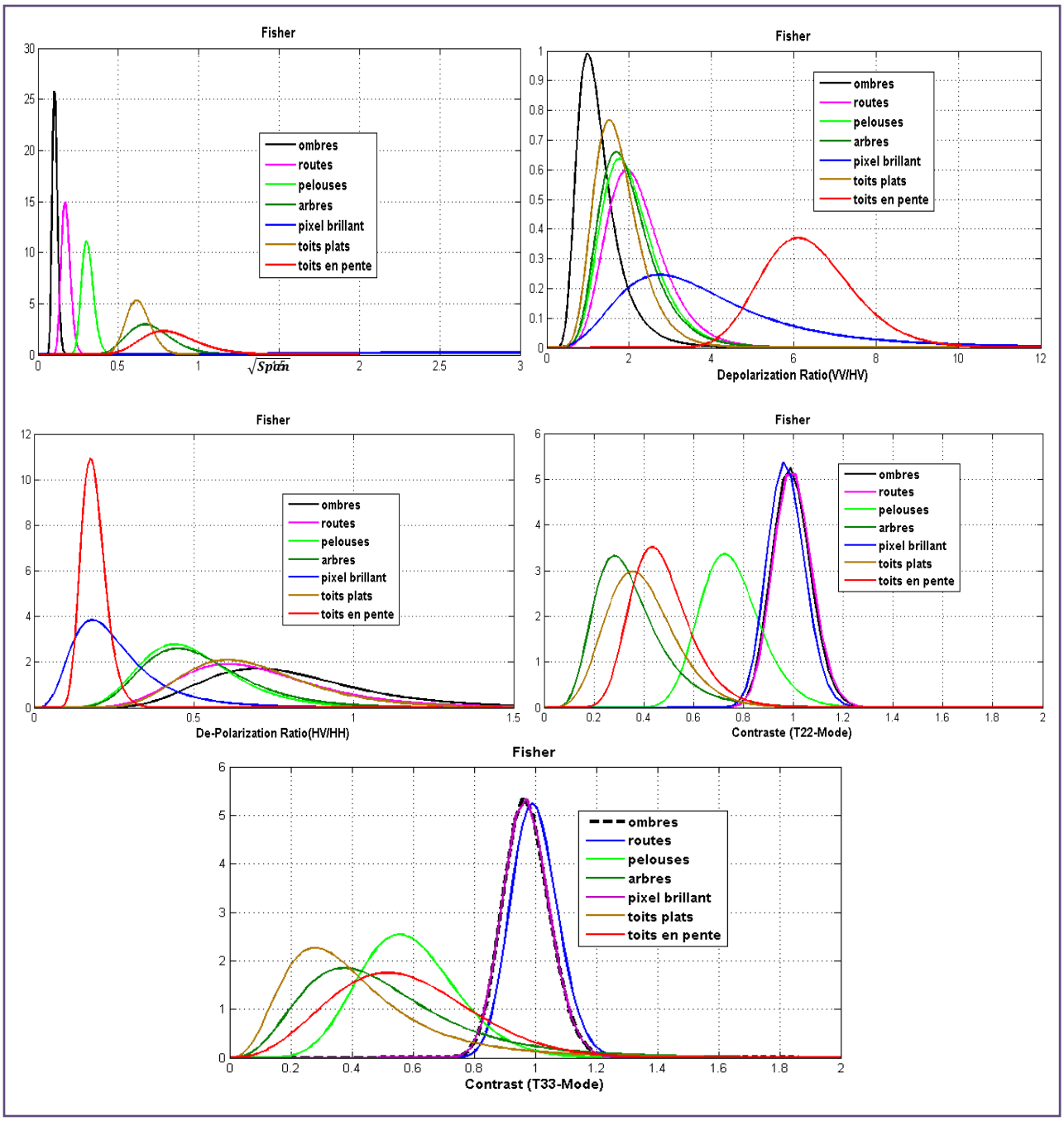


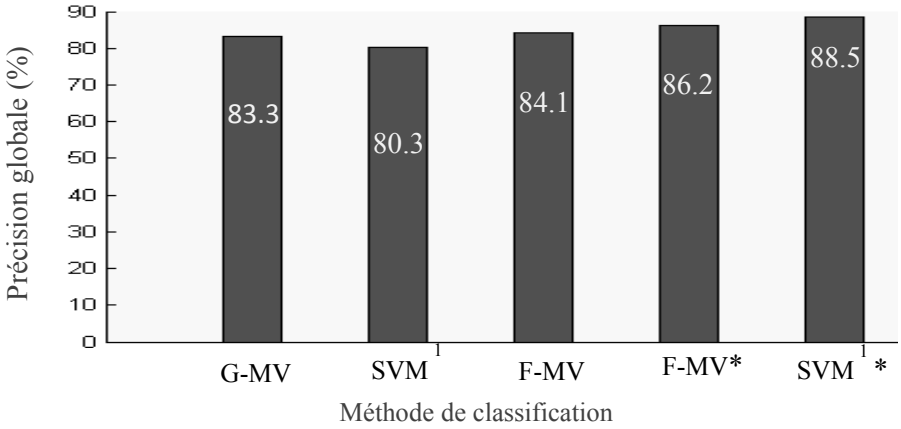
Figure11. Pdf de Fisher estimée par MoLC pour sept classes et un ensemble optimal de 5 descripteurs issu de la MSS (Tableau3) à partir de 58 descripteurs positifs réels (Tableau 1).

Expérience de classification I

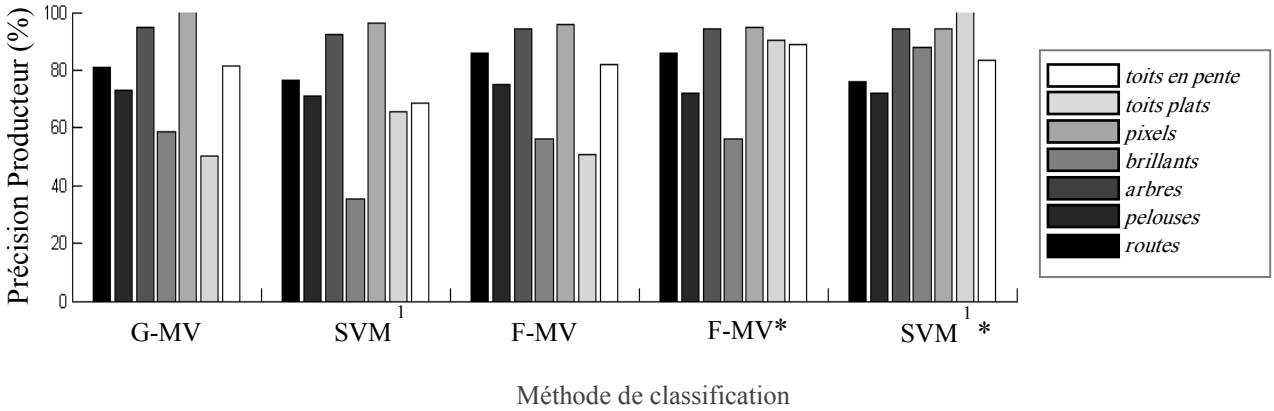
Pour montrer l'efficacité de la pdf de Fisher sur tous les descripteurs positifs réels dans un ensemble optimal, nous utilisons la pdf de Fisher dans le classifieur F-MV proposé. Le résultat est illustré par la Figure 13. Ce résultat est comparé avec ceux de différents classifieurs tels que SVM, G-MV (Figure 12). Par ailleurs, pour montrer l'intérêt de construire une carte de l'emprise des bâtiments pour améliorer le résultat de la classification, une expérience de classification basée sur les caractéristiques

morphologiques et les descripteurs positifs réels (Tableau 3) est présentée (Figure 12).

Les résultats indiquent que la méthode de classification F-MV a la plus grande précision globale (84,1 %) par rapport aux méthodes de classification G-MV et SVM. Cependant le résultat de SVM[1 *] avec un ensemble optimal et la carte du bâti est meilleure que celui d'autres méthodes (Figure 13).



(a)



(b

`

Figure 12: Précision globale (a) et précision par classe (b) des différents classifieurs (en %): 1 correspond à une classification avec un ensemble optimal de 5 descripteurs issu de la MSS (Tableau 3) à partir de 58 descripteurs (Tableau 1) et * correspond à une classification avec de descripteur morphologique.

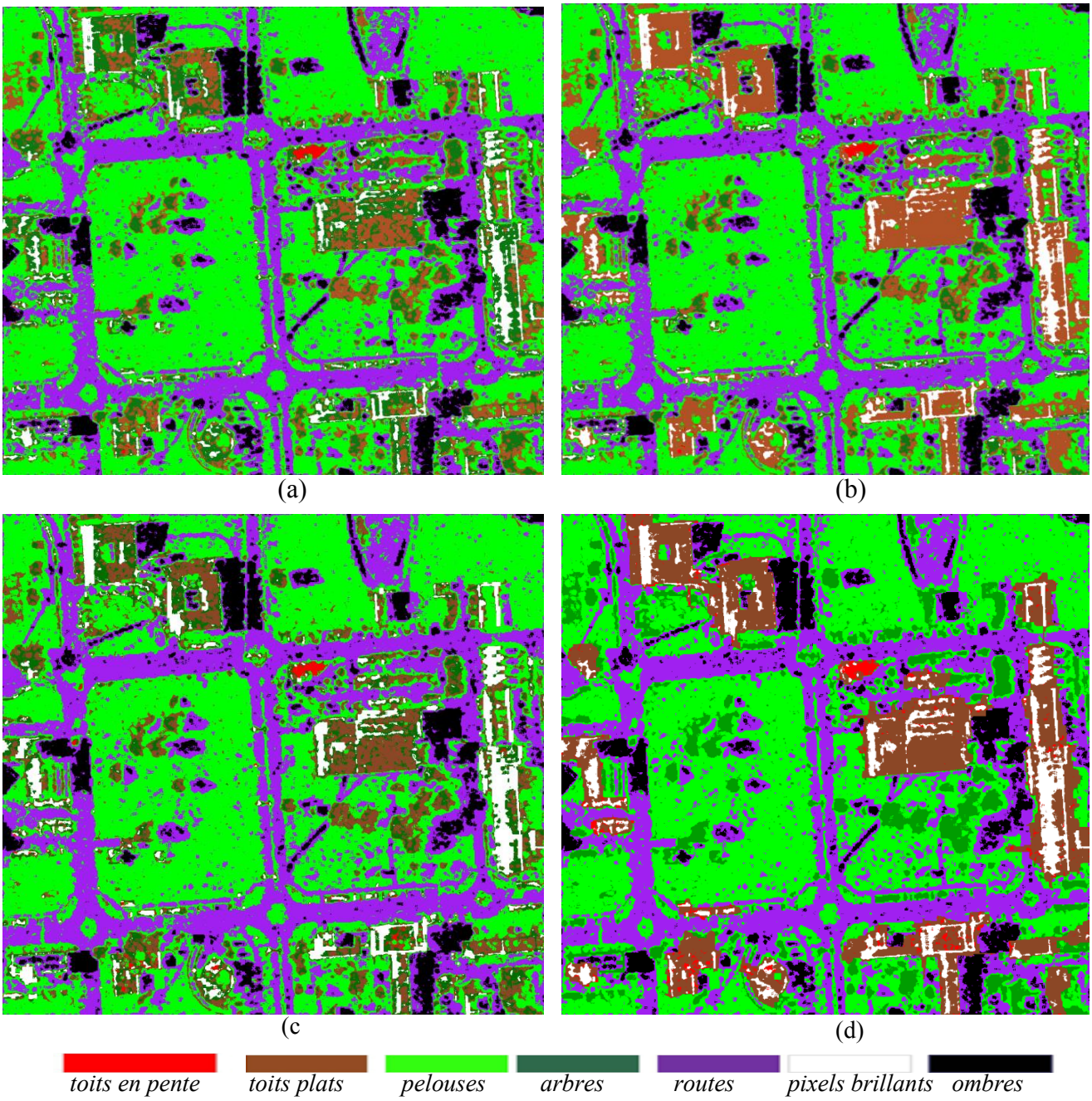


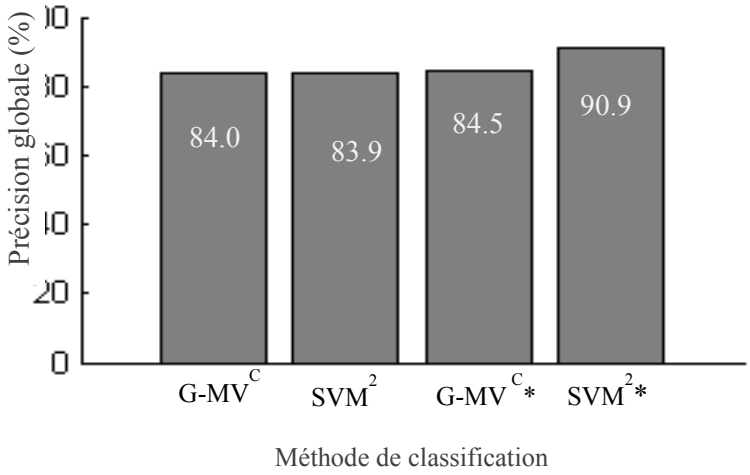
Figure 13: Résultat de Classification: (a) F-MV, (b) F-MV*,(c) SVM^1 et (d) SVM^{1*}. ^{1*} signifie une classification sur un ensemble optimal (Tableau 3) de à partir de 58 descripteurs positifs réels (Tableau 1) couplés avec la carte morphologique.

Expérience de classification II

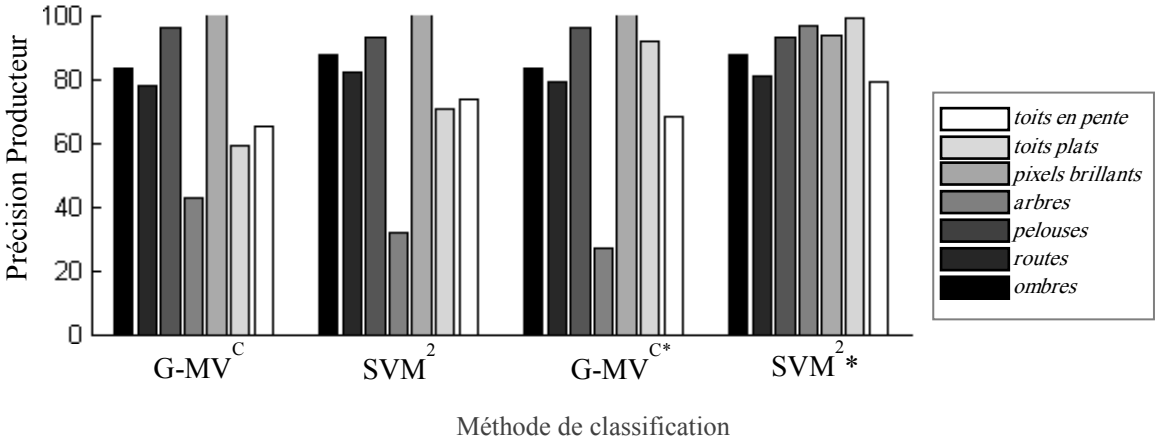
Dans la seconde expérience, l'efficacité de la MSS avec les 127 descripteurs est considérée. À cet effet, les résultats de classification par SVM sur l'ensemble optimal de 5 descripteurs (Tableau 4) sont comparés avec les résultats $G\text{-}MV^C$ (classification par maximum de vraisemblance pour une distribution gaussienne multi-variée complexe), sans et avec la composante morphologique. Les résultats de l'expérience de comparaison entre SVM^2, SVM^2*, $G\text{-}MV^C$et $G\text{-}MV^C$* sont indiqués sur la Figure 14.

Les résultats de SVM^{2}* et G-MV^{C}* montrent une diminution des ambiguïtés entre les *arbres* et les toits. SVM^{2}* donne le meilleur résultat (Figure 15).

Finalement, les résultats de SVM^{2}* montrent l'efficacité de la méthode de sélection d'attributs couplée avec les empreintes de bâtiments et le potentiel de la classification par SVM non paramétrique et non linéaire. La précision de la classification G-MV^{C} avec 9 descripteurs (de la matrice de cohérence) est de 84 %, soit 0,1 % plus élevé que la classification par SVM avec un ensemble optimal de 5 descripteurs.



(a)



(b)

Figure 14: Précision globale (a) et précision par classe (en %) des classifieurs G-MV^{C} et SVM^{2} : ^{2} correspond à un classification avec un ensemble optimal de 5 descripteurs (Tableau 4) issu de la MSS depuis 127 descripteurs (Tableau 2) et * correspond à une classification avec descripteur morphologique.

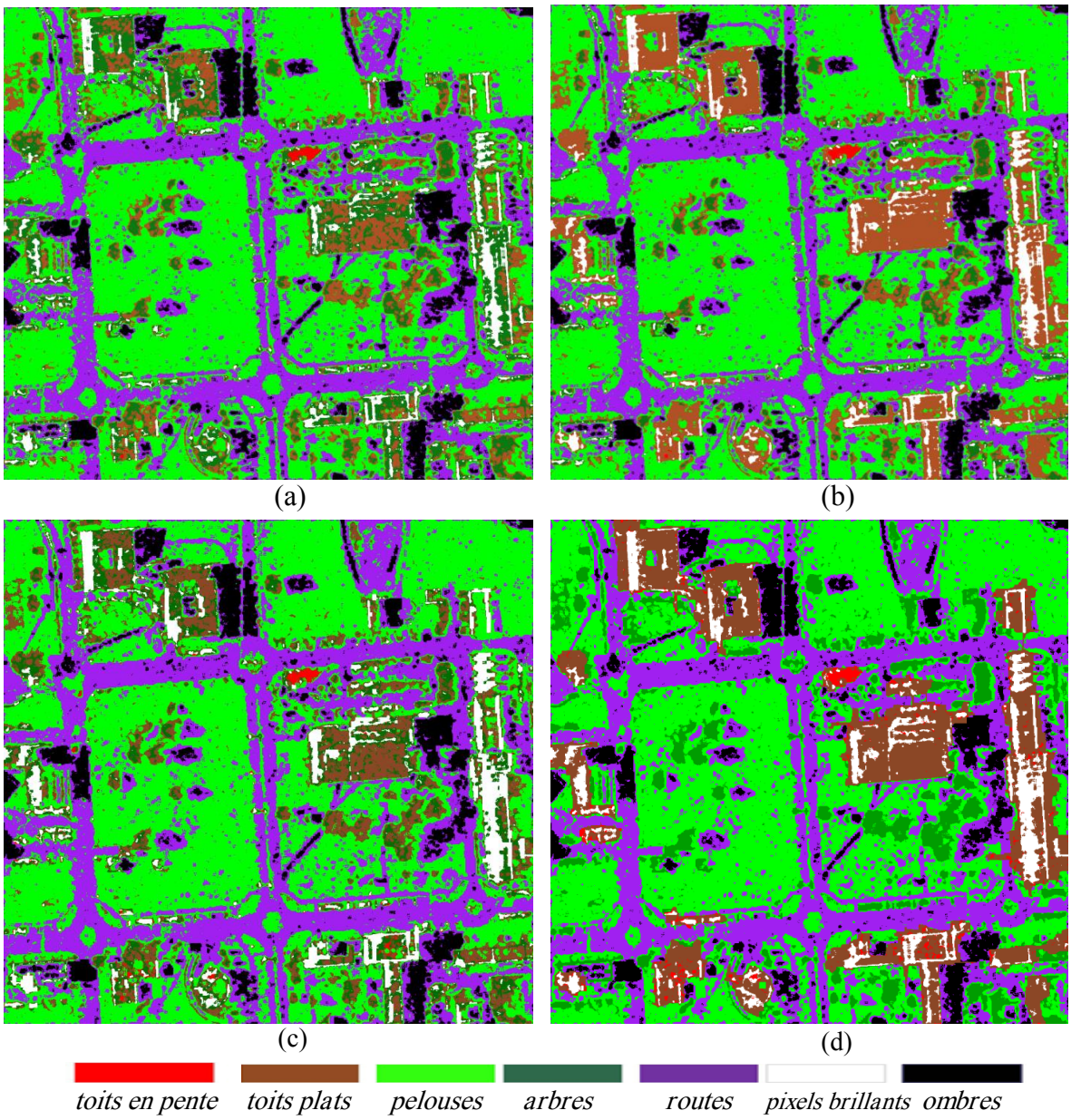
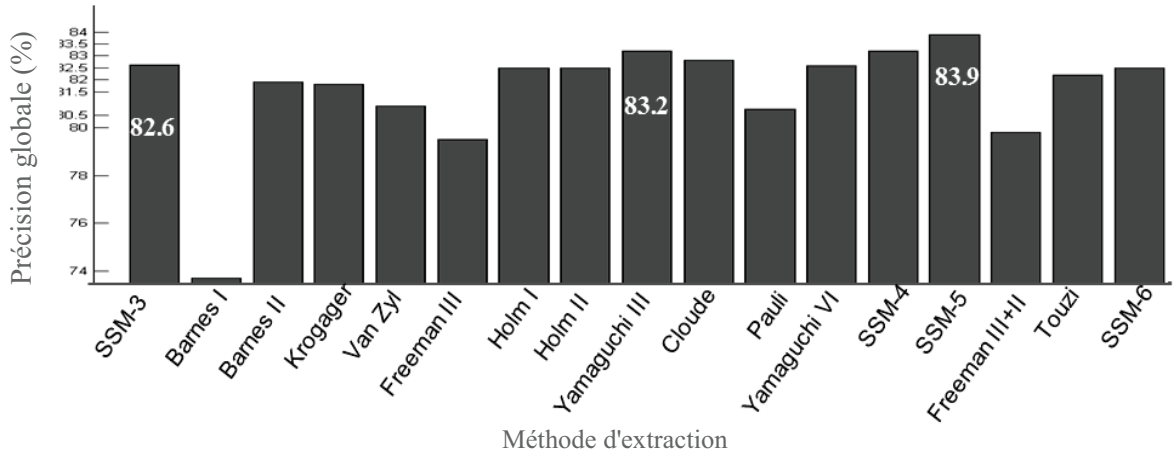


Figure 15: Résultat de Classification: (a) G- MV^{C},(b) G-MV^{C*}, (c) SVM^{2} et (d) SVM^{2*}.^{2*} correspond à une classification avec sous-ensemble optimal (Tableau 4) de MSS par GA-MI de 127 descripteurs (Tableau 2) couplé avec la carte morphologique.

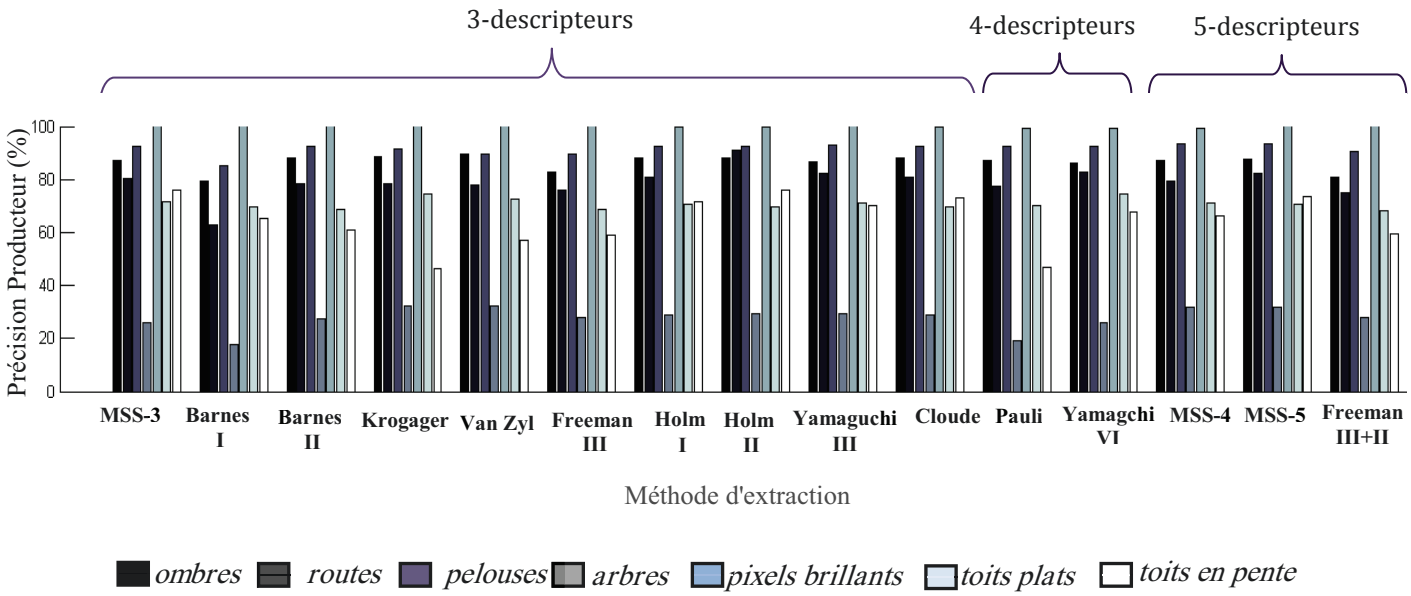
Afin de démontrer l'efficacité de cette approche, nous devons effectuer une comparaison entre les résultats de la classification SVM obtenue à partir de l'ensemble optimal d'attributs et pour les attributs seuls issus de décompositions de cibles basés sur différents mécanismes de rétrodiffusion : Pauli (Cloude & Pottier, 1996), Huynen (Huynen, 1970), Cloude (Cloude &Pottier, 1997), Holm-Barnes (Holm & Barnes, 1988), H/A/Alpha (Cloude & Pottier, 1996), Freeman (Freeman, 2007 ; Freeman & Durden, 1998), VanZyl (Van Zyl, 1993), Krogager (Krogager, 1990), Yamaguchi (Yamaguchi et al., 2005), ou Touzi (Bhattacharya & Touzi, 2012). Les résultats expérimentaux

montrent que la méthode proposée permet une sélection pertinente d'attributs, relatifs à différents mécanismes de rétrodiffusion, et mène à une classification plus précise.

La précision globale est présentée sur la Figure 16 (a). Avec 83,9 %, elle est la meilleure par rapport à l'utilisation de méthodes classiques de décomposition de cibles. La Figure 16 (b) montre la précision obtenue pour les sept classes.



(a)



(b)

Figure 16: Précision globale (a) et la précision par classe (b) de la classification par SVM pour différentes méthodes d'extraction de descripteurs.

Conclusions et Perspectives

Dans cette thèse, nous avons étudié la capacité de l'imagerie POLSAR monovue pour discriminer différentes surfaces en zone urbaine. Une première base de travail a été mise en place à partir de descripteurs statistiques et polarimétriques. Ensuite, nous avons utilisé une classification par

maximum de vraisemblance basée sur la distribution de Fisher. Il a été montré l'amélioration de la précision de classification par l'utilisation de descripteurs polarimétriques réels positifs avec des paramètres texturaux. Ensuite, nous avons évalué la probabilité de discriminer les arbres par rapport aux toits plats en utilisant des outils morphologiques.

Nous avons proposé une méthode de classification F-ML qui exige l'indépendance des descripteurs et l'adéquation de la distribution de Fisher.

C'est pourquoi, nous avons proposé une méthode de sélection des attributs basée sur un algorithme génétique et une fonction d'évaluation mRMP : critère de maximum de pertinence et de redondance minimale.

Pour trouver un modèle statistique qui est en mesure de modéliser la rétrodiffusion dans des zones de textures hétérogènes, la façon la plus directe est d'examiner les différentes fonctions de distribution. Les résultats statistiques montrent que la distribution de Fisher peut être préférable, car elle permet une meilleure modélisation de différentes classes que d'autres modèles statistiques bien connus tels que $\mathcal{K}$, Weibull, Beta I, log-normale, et Gamma selon les résultats du GoF. Dans chaque cas, il a été constaté que le MoLC est préférable à la méthode MV pour estimer les paramètres.

En outre, pour illustrer l'efficacité de la classification par F-MV, nous avons fait une expérience comparative en utilisant l'ensemble optimal composé de cinq descripteurs réels positifs avec différentes méthodes de classification. Les résultats révèlent que F-MV aboutit aux meilleures performances. La classification F-MV permet de discriminer les ombres des autoroutes et les toits inclinés des arbres. Toutefois, le choix du modèle statistique devrait être fondé sur son caractère généralisable; dans le cas présent, il peut être ajusté à nos données, mais son adéquation à d'autres données nécessite un examen plus approfondi. En outre, plus d'expériences utilisant d'autres données réelles disponibles devraient être faites pour une conclusion générale.

Nous avons appliqué la même base expérimentale à d'autres descripteurs polarimétriques. Pour illustrer l'efficacité de la méthode de sélection AG-IM, nous avons effectué plusieurs comparaisons avec l'utilisation seule de différentes décompositions de cibles à savoir Pauli, Krogager, Freeman, Yamaguchi, Barnes, Holm, Huynen et Cloude. Les résultats expérimentaux indiquent que la méthode proposée donne une bonne sélection d'attributs, montrant différents mécanismes de rétrodiffusion, et de meilleures performances pour la classification des images SAR.

La MSS proposée, basée sur l'algorithme AG-IM, présente plusieurs avantages : d'abord, il est facile à mettre en œuvre, rapide, les chances d'obtenir une solution optimale existent, et l'espace de recherche n'est pas limité; deuxièmement, l'IM peut gérer une corrélation non linéaire entre attributs et classes ainsi qu'une anti corrélation entre une paire d'attributs; troisièmement, il est adapté à différents types d'images radar polarimétriques et il est indépendant des valeurs et de la forme des distributions; enfin, cette méthode peut être appliquée à différents types d'occupation du sol et différents capteurs, ce qui pourra faire l'objet de travaux futurs.

Toutefois, la sélection de la fonction d'évaluation, la taille de la population, et la méthode de sélection au cours de la procédure de cross-over sont très importantes et peuvent influencer les résultats. En outre, cette approche devrait être comparée à d'autres méthodes.

Comme les informations morphologiques sont très utiles pour la classification, l'automatisation de la production de la carte des empreintes de bâtiments demeure une orientation des travaux futurs.

Références

Anfinsen, S. N. (2010). *Statistical analysis of multilook polarimetric radar images with the Mellin transform. PhD thesis.* Norway: University of Tromso.

Bertrand, J., Bertrand, P., & Ovarlez, J.-P. (2000). *The Transforms and Applications Handbook. Chapter 12: The Mellin Transform.* Boca Raton, USA: CRC Press Inc.

Bhattacharya, B., & Touzi, R. (2012). Polarimetric SAR urban classification using the Touzi target scattering decomposition. *Journal canadien de télédétection, 37(4)*, 323-332.

Bombrun, L., Vasile, G., Gay, M., & Totir, F. (2011). Hierarchical Segmentation of Polarimetric SAR Images Using Heterogeneous Clutter Models. *IEEE Transactions on Geoscience And Remote Sensing , 49(2)*, 726-737.

Bombrun, L., & Beaulieu, J.-M. (2008). Fisher distribution for texture modeling of polarimetric SAR data. *IEEE Geoscience and Remote Sensing Letters, 5(3)*, 512–516.

Chen, K. S., Huang, W. P., Tsay, D. H., & Amar, F. (1996). Classification of multifrequency polarimetric SAR imagery using a dynamic learning neural network. *IEEE Transactions on Geoscience and Remote Sensing, 34(3)*, 814-820.

Cloude, S. R., & Pottier, E. (1997). An entropy based classification scheme for land application of polarimetric SAR. *IEEE Transactions on Geoscience and Remote Sensing, 35(1)*, 68-78.

Cloude, S. R., & Pottier, E. (1996). A Review of Target Decomposition Theorems in Radar Polarimetry. *IEEE Transactions on Geoscience and Remote Sensing, 34(2)*, 498-518.

Debnath, L., & Bhatta, D. (2007). *Integral Transforms and their Applications. Chapter 8.* Boca Raton, USA: Chapman & Hall/CRC.

Dousset, B. (1995). Synthetic aperture radar imaging of urban surfaces: a case study. *IEEE International Geoscience and Remote Sensing Symposium, 3*, 2092-2096.

Epstein, B. (1948). Some applications of the Mellin transform in statistics. *Ann.Math.Statist, 19(3)*, 370–379.

Ferro-Famil, L., Pottier, E., & Lee, J. S. (2001). Unsupervised classification of multifrequency and fully polarimetric SAR images based on the H/A/Alpha Wishart.

Freeman, A. (2007). Fitting a Two-Component Scattering Model to Polarimetric SAR Data From Forests. *IEEE Transactions on Geoscience and Remote Sensing, 42(8)*, 583-2592 .

Freeman, A., & Durden, S. (1998). A three-component scattering model for polarimetric SAR data. *IEEE Transactions on Geoscience and Remote Sensing, 36(3)*, 963-973.

Formont, P., Pascal, F., Vasile, G., Ovarlez, J.-P., & Ferro-Famil, L. (2011). Statistical Classification for Heterogeneous Polarimetric SAR Images. *IEEE Journal of Selected Topics in Signal Processing, 5(3)*, 567-576.

Fukuda, S., & Hirosawa, H. (1999). A wavelet-based texture feature set applied to classifcation of multifrequency polarimetric SAR images. *IEEE Transactions on Geoscience and Remote Sensing, 37(5)*, 2282-2286.

Holm, W. A., & Barnes, R. (1988). On radar polarization mixed target state decomposition techniques. *IEEE National radar Conf*, 249-254.

Huynen, J. R. (1970). *Phonomenological Theory of Radar Targets.Ph.D.thesis.* Netherlands: University of Technology Delf.

Ince, T., Kiranyaz, S., & Gabbouj, M. (2012). Evolutionary RBF classifier for polarimetric SAR images. *Expert Systems with Applications, 39*, 4710–4717.

Ince, T. (2010). Polarimetric SAR image classification using a radial basis function neural network. *PIERS 2010, 41(4)*, 636–646.

Iwasa, S., Susaki, J., & Tamura, M. (2010). Classification of Building Area using Slant Angle and Density Indices Derived from Polarimetric SAR Data. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science, XXXVIII,Part 8*, 219-224.

Kahny, D., & Wiesbeck, W. (1991). Optimum Input Parameters For Classification Of Multifrequency Polarimetric SAR data Using Neural Networks. *IGARSS Geoscience and Remote Sensing Symposium, 4*, 2157-2160.

Kajimoto, M., & Susaki, J. (2013). Urban-Area Extraction From Polarimetric SAR Images Using Polarization Orientation Angle . *IEEE Geoscience and Remote Sensing Letters, 10(2)*, 337-341.

Krogager, E. (1990). A new decomposition of the radar target scattering matrix. *Electron. Lett, 26(18)*, 1525-1526.

Lee, J. S., & Pottier, E. (2009). *Polarimetric Radar Imaging: From Basics to Applications.* CRC Press,Taylor & Francis Group.

Lee, J. S., Grunes, M. R., Ainsworth, T., Du, L.-J., Schuler, D., & Cloude, S. R. (1999). Unsupervised classification using polarimetric decomposition and the complex Wishart classifier. *IEEE Transactions on Geoscience and Remote Sensing, 37(5)*, 2249–2257.

Ludwig, O., & Nunes, U. (2010). Novel Maximum-Margin Training Algorithms for Supervised Neural Networks. *IEEE Transactions on Neural Networks, 21(6)*, 972-984.

Ludwig, O., Nunes, U., Araujo, R., Schnitman, L., & Lepikson, H. A. (2009). Applications of information theory, genetic algoritms, and neural models to predict oil flow. *Communications in Nonlinear Science and Numerical Simulation, 14(7)*, 2870-2885.

Moriyama, T., Yamaguchi, Y., Uratsuka, S., Umehara, T., Satake, M., Nadai, A. et al. (2004). A Study on Extraction of Urban Areas from Polarimetric Synthetic Aperture Radar image. *IEEE, 1*, 703-706.

Nicolas, J. -M., & Tupin, F. (2002). Gamma mixture modeled with "second kind statistics": application to SAR image processing. *IEEE Geoscience and Remote Sensing Symposium, 4*, 2489-2491.

Nicolas, J.-M. (2002). Introduction aux statistiques de deuxième espèce: Application des logs-moments et des logs-cumulants à l'analyse des lois d'images radar. *Traitement du Signal, 19(3)*, 139–167.

Nicolas, J.-M. (2006). *Application de la transformée de Mellin: Étude des lois statistiques de l'imagerie cohérente.* (T. R. 2006D010, Ed.) Paris, France: Ecole Nationale Supérieure des Télécommunications.

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics, 9(1)*, 62-66.

Peng, H., Long, F., & Ding, C. (2005). Feature Selection Based on Mutual Information:Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence , 27(8)*, 1226-1238.

Pottier, E., Ferro-Famil, L., Cloude, S., Hajnsek, I., Papathanassiou, L., Moreira, A. et al. (2005). PoLSARpro v2.0 Software A Versatile Polarimetric SAR Data Processing and Educatinal Toolbox.

Poularikas, A. D. (1999). *The Handbook of Formulas and Tables for Signal Processing,Chapter 18: The Mellin Transform.* Boca Raton, U.S: CRC Press & IEEE Press.

Schistad Solberg, A. H., & Jain, A. K. (1997). Texture Fusion and Feature Selection Applied to SAR Imagery. *IEEE Transactions on Geoscience and Remote Sensing, 35(2)*, 475-479.

Shannon, C. E. (1948). The mathematical theory of communication. *Reprinted with corrections from The Bell System Technical Journal, 27*, 379–423, 623–656.

Simonetto, E., & Malak, C. (2009). Urban area structuring mapping using an airborne polarimetric SAR image. SPIE 7477 Image and Signal Processing for Remote Sensing XV, 7477.

Soheili Majd, M, (2014). High-resolution polarimetric SAR image classification on urban areas. Ph.D, thesis. France: CNAM.

Tan, C. P., Lim, K. S., & Ewe, H. T. (2007). Image processing in polarimetric SAR images using a hybrid entropy decomposition and maximum likelihood (EDML). In Proceedings of the International Symposium on Image and Signal Processing and *Analysis (ISPA)* , 418–422.

Unal, C. M., & Ligthart, L. P. (1998). Decomposition Theorems Applied to Random and Stationary Radar Targets. *PIER 18, Progress In Electromagnetics Research*, 45–66.

Van Zyl, J. J. (1993). Application of Cloude's target decomposition theorem to polarimetric imaging radar data. *in Proceedings SPIE Conference on Radar Polarimetry,SanDiego,CA, 1748*, 184–212.

Vinh, L. T., Lee, S., Park, Y.-T., & D'Auriol, B. J. (2012). A novel feature selection method based on normalized mutual information. *Appl Intell, DOI 10.1007/s10489-011-0315-y, 37*, 100-120.

Yamaguchi, Y., Moriyama, M., Ishido, M., & Yamada, H. (2005). Four-Component Scattering Model for Polarimetric SAR Image Decomposition. *IEEE Transactions on Geoscience and Remote Sensing, 43(8)*, 1699-1706. Simonetto, E., & Malak, C. (2009). Urban area structuring mapping using an airborne polarimetric SAR image. *SPIE 7477 Image and Signal Processing for Remote Sensing XV, 7477*.

Yamaguchi, Y., Moriyama, M., Ishido, M., & Yamada, H. (2005). Four-Component Scattering Model for Polarimetric SAR Image Decomposition. *IEEE Transactions on Geoscience and Remote Sensing, 43(8)*, 1699-1706.

Yang, J., Peng, Y. -N., Yamaguchi, Y., & Yamada, H. (2006). On Huynen's Decomposition of a Kennaugh Matrix. *IEEE Transactions on Geoscience and Remote Sensing, 3(3)*, 369-372.

Zhang, L., Zou, B., Jia, Q., & Zhang, Y. (2009). Polarimetric SAR image classification using Multiple-Component Scattering Model and Support Vector Machine. *IEEE Synthetic Aperture Radar*, 805-808.