



Thèse

Présentée pour obtenir le grade de docteur
de Télécom ParisTech

Spécialité : **Signal et Images**

Daniela Espinoza Molina

Sujet :

ADVANCED METHODS FOR HIGH RESOLUTION SAR
INFORMATION EXTRACTION: DATA AND USER-DRIVEN
EVALUATION APPROACHES FOR IMAGE INFORMATION MINING

SOUTENUE LE 16 MAI 2011 DEVANT LE JURY COMPOSÉ DE:

M. Jean Philippe OVARLEZ	Rapporteur
M. Peter REINARTZ	Rapporteur
M. Eric POTTIER	Examineur
M. Emmanuel TROUVE	Examineur
M. Jean Marie NICOLAS	Examineur
M. Marin FERECATU	Examineur
M. Mihai DATCU	Directeur de thèse
M. Dušan GLEICH	Co-directeur de thèse

Abstract

We are concerned in this thesis by the problem of Image Information Mining (IIM) for exploitation and understanding of high resolution Synthetic Aperture Radar (SAR) data. Advances in this field of research contribute to the elaboration of tools for interactive exploration and extraction of the image content. In this context, analyzing and evaluating adequate image models and image information extraction methods according to user conjectures, constitute challenging issues. Our work contributes with solutions to high resolution SAR modeling and content estimation with a data-driven evaluation approach, and with the design of image mining scenarios by involving the user and his conjectures, achieved through an user-driven evaluation approach.

To represent the data and to allow extracting the information, we focus on the properties of high resolution SAR images and how the stochastic models can represent and characterize the image content through a parameter estimation step. We perform a data-driven evaluation and validation of automatic information extraction methods for high resolution SAR scenes based on Gibbs Random Field (GRF) models. Specifically, Gauss Markov Random Field (GMRF) and Auto-binomial (ABM) models are implemented in information extraction methods following the two levels of Bayesian inference: model fitting and model selection. Both methods provide as results the speckle-free image and its structure parameters. In order to assess the quality of these methods, we perform detection tests on classes such as cities, vegetation, and water bodies; using specific qualitative metrics to quantify the quality of speckle removal. The accuracy of modelling and characterization of the image content are determined using both supervised and unsupervised classifications, and confusion matrices. We conclude that both methods enhance the image during the despeckling process. The GMRF model is more suitable for natural scenes and the ABM model for man-made structures. However, evaluating the information extraction methods is not enough for a complete validation of IIM systems, as we need to adapt to the user conjectures by designing validation scenarios and assessing the user satisfaction degree as well as the effectiveness of the retrieval process. We design and generate two study cases, which reflect the user needs in solving rapid mapping applications. The end-user is included in the loop in the user-driven evaluation approach by creating the two evaluation scenarios in the framework of disaster monitoring: oil spill and flood detection. The scenarios are carried out using ScanSAR and High Resolution Spotlight TerraSAR-X products, respectively. Quantitative metrics such precision and recall are used as figures of merit. In order to have measurements about the user satisfaction degree, a group of evaluators are asked to qualitatively rank the retrieved results. We conclude that the effectiveness of the retrieval process is more than 80 percent and the degree of user satisfaction is good for both scenarios.

Contents

1	Introduction	9
1.1	Motivation	9
1.2	Goals of the thesis	10
1.3	Contribution of the thesis	10
1.4	Outline of the thesis	11
2	Synthetic Aperture Radar	13
2.1	Fundamentals of SAR	13
2.1.1	Principle of SAR	14
2.1.2	SAR image formation	15
2.1.2.1	Range resolution	15
2.1.2.2	Azimuth resolution	16
2.1.3	Basic scattering mechanisms	18
2.1.3.1	Surface and Volume Scattering	18
2.1.3.2	Single and Double Bounce	20
2.2	Geometrical and Radiometrical effects	20
2.2.1	Geometrical effects	20
2.2.1.1	Foreshortening	21
2.2.1.2	Layover	21
2.2.1.3	Shadows	22
2.2.2	Radiometric corrections	23
2.2.3	Geometric corrections	23
2.3	SAR image statistics	23
2.3.1	Speckle effect	24
2.3.2	Statistical properties of the backscattered signal	25
2.3.3	Statistical properties of Multi-Look data	26
2.3.3.1	Intensity images	26
2.3.3.2	Square-root intensity images	27
2.3.3.3	Amplitude images	28
2.3.4	The multiplicative noise model for Speckle	28
2.4	TerraSAR-X products and information content	29
2.4.1	TerraSAR-X Instrument	29
2.4.1.1	Imaging Modes	29
2.4.2	Basic TerraSAR-X products	31
2.4.2.1	Geometric resolution	31
2.4.2.2	Geometric projections and data representation	31
2.4.2.3	Product identification Scheme	32

2.4.3	Examples of TerraSAR-X image content	34
3	Image Information Mining systems	35
3.1	Content-Based Image Retrieval	36
3.2	Image Information Mining Systems	37
3.2.1	Generic Architecture	37
3.2.1.1	Image database	38
3.2.1.2	Feature extraction	39
3.2.1.3	Data reduction and content index generation	39
3.2.1.4	User interaction or human-machine communication	39
3.2.2	Examples of IIM systems	40
3.3	Evaluation in Image Information Mining systems	40
3.3.1	Data-driven evaluation approach	42
3.3.1.1	Component by component evaluation	42
3.3.1.2	General evaluation of the performance	43
3.3.2	User-driven evaluation approach	44
3.4	Conclusions	46
4	SAR image analysis using Bayesian inference and Gibbs Random Field models	47
4.1	Bayesian inference	48
4.1.1	Conditional probability and Bayes' theorem	48
4.1.2	Parameter estimation	49
4.1.3	Model Selection	49
4.2	Markov and Gibbs Random Fields	50
4.2.1	Image description, neighbourhood systems and cliques	50
4.2.2	Markov-Gibbs Random Fields equivalence	51
4.2.2.1	Hammersley-Clifford theorem	52
4.2.3	Examples of Gibbs Random Field models	52
4.2.3.1	Gauss Markov Random Field model	52
4.2.3.2	Auto-binomial model	53
4.2.3.3	Examples of realizations of GMRF and ABM models	53
4.3	Bayesian SAR image analysis	54
4.3.1	SAR despeckling filters	55
4.3.1.1	Non-Bayesian despeckling filters	55
4.3.1.2	Bayesian despeckling filters	56
4.3.2	SAR information extraction	58
4.4	Conclusion	59
5	Data-driven evaluation of SAR information extraction methods	61
5.1	Description of SAR information extraction model-based methods	62
5.1.1	Bayesian Approach	62
5.1.2	Maximum a Posteriori despeckling using Gauss Markov Random Field (MAP-GMRF) and evidence maximization	63
5.1.3	Maximum a Posteriori despeckling using Auto-binomial model (MAP-ABM) and evidence maximization	65
5.1.4	Overview of MAP-GMRF and MAP-ABM	66
5.2	Description of the test data set	69
5.2.1	Simulated SAR data	69

5.2.1.1	Synthetic textures	69
5.2.1.2	Speckled Brodatz textures	69
5.2.2	Real SAR data	72
5.2.2.1	Textures from TerraSAR-X images	72
5.2.2.2	TerraSAR-X sub-scene	72
5.3	Quantitative evaluation of despeckling	74
5.3.1	Quantitative quality measurements	74
5.3.1.1	Metrics for Simulated SAR data despeckling evaluation	74
5.3.1.2	Metrics for Real SAR data despeckling evaluation	76
5.3.2	Experimental results	77
5.3.2.1	Despeckling results based on Simulated SAR data	77
5.3.2.2	Despeckling results based on Real SAR data	80
5.3.2.3	The impact of window size and model order on despeckling performance	86
5.3.3	Conclusions	88
5.4	Quantitative evaluation of SAR information extraction	91
5.4.1	Quantitative Metrics	92
5.4.1.1	Classification and Confusion Matrix	92
5.4.2	Robustness of feature extraction evaluation in terms of classifications	92
5.4.2.1	Classification results based on Simulated SAR data	94
5.4.2.2	Classification results based on Real SAR data	98
5.4.3	The impact of window size and model order on feature extraction	104
5.4.4	Conclusions	108
5.5	Conclusions	108
6	User-driven evaluation of Image Information Mining systems	111
6.1	Knowledge-based Image Information Mining system as an Image Information Mining tool	112
6.1.1	KIM Concept	112
6.1.2	KIM Architecture	114
6.1.2.1	GUI Components	115
6.1.3	Operation of KIM	116
6.1.3.1	The Server-side	116
6.1.3.2	The Client-side	116
6.2	Effectiveness evaluation metrics	120
6.2.1	Precision and Recall	120
6.2.2	Accuracy and F-measure	121
6.2.3	Probability to over-retrieve and to forget an image	122
6.2.4	Target and misclassified images	122
6.3	Evaluation Methodology	122
6.3.1	Role of the user	122
6.3.1.1	The user as interpreter of the ground-truth	123
6.3.1.2	The user as evaluator of the system	123
6.3.2	Creation of the ground-truth	123
6.3.2.1	Definition of the relevance criterion	123
6.3.2.2	Definition of the scheme for annotating images	124
6.3.2.3	Creation of the reference database	124
6.3.3	System evaluation methods	127

6.3.3.1	Configuration of the scenario	127
6.3.3.2	Training performed by the user and definition of semantic labels	127
6.3.3.3	Assessment of the probabilistic search and interactive learning	128
6.4	Examples of validation scenarios: Disaster monitoring	128
6.4.1	Study case 1: Oil Spill in Gulf of Mexico	129
6.4.1.1	Description of data set, parameters and ground-truth	130
6.4.1.2	Training performed by the end-user	132
6.4.1.3	Evaluation results of the probabilistic search	132
6.4.1.4	Classification results	135
6.4.1.5	User satisfaction degree	136
6.4.2	Study case 2: Nepal-Embankment breach	139
6.4.2.1	Description of the study area, data set and ground truth	139
6.4.2.2	Training performed by the end-user	141
6.4.2.3	Evaluation results of the probabilistic search	141
6.4.2.4	Classification results	148
6.4.2.5	User satisfaction degree	149
6.5	Conclusions	153
7	Conclusions	155
7.1	Summary of the Thesis	155
7.2	Future work	158
A	TerraSAR-X image content	159
B	Evidence maximization approach	161
C	Validation scenarios examples	163
	Bibliography	167

Chapter 1

Introduction

1.1 Motivation

Synthetic Aperture Radar (**SAR**) images have been used in a large number of applications in different fields, such as remote sensing for mapping, automatic target recognition, search and rescue, mine detection, etc. Indeed, a SAR sensor has a number of advantages compared with an optical sensor as for example it is independent of any illumination by the sun causing that the SAR images can be acquired for any time and zone. Nowadays, the analysis of spaceborne high resolution SAR images with up to 1 meter spatial resolution has become possible with the advent of German, Italian, and Canadian missions and their subsequent data distribution. For instance, compared to previous missions with much lower resolution, the German TerraSAR-X mission provides high resolution SAR images containing an increased amount of details and information content. This kind of images offers the possibility to detect new structures (man-made), small-scale objects, etc. However, the automatic analysis of image content is more difficult due to the huge amount of information content and the nature of SAR data i.e speckle, geometry, backscattering. As a consequence, in order to fully exploit at best the automatic interpretation of high resolution SAR imagery it is needed to rely on proper approaches that automatically remove the speckle, conserving the details and extract the information content.

In parallel, the continuous growing acquisition of **Earth-Observation (EO)** imagery (i.e optical and SAR) has motivated to develop sophisticated methods, algorithms and systems in order to handle the amount of this data. Generally, the images are stored in large repositories, where accessing the image information content, in comparison with other data types, is more complicated, residing mainly in the huge volume of data and the rich information content. Hence, **Image Information Mining (IIM)** has arisen as a new field of study in order to help the users to deal with large image collections by accessing and extracting automatically their content, allowing querying and mining relevant information, and inferring knowledge about patterns hidden in the images. In general, these systems are based on a **Content-Based Image Retrieval (CBIR)** approach and incorporate components in order to extract relevant information from the images (primitive features), to generate class files by grouping the primitive features in clusters, and to interactively search for the image content through **Graphical User Interface (GUI)** that enables the human-machine communication.

For example, in our specific case, the Knowledge-based Image Information Mining

system (KIM) is integrated with the Ground-Segment of TerraSAR-X, enabling to store the EO imagery and to automatically exploit the image content by applying diverse information extraction methods and creating different application scenarios according to user interests and conjectures.

In order to exploit the EO imagery, along the years, many techniques have been proposed in feature extraction methods that can be incorporated in automatic content-based retrieval systems. However, in order to improve these techniques and systems it is needed to rely on evaluation strategies that either expose which aspects should be improved or expound the necessity of new methods. Moreover, the evaluation gives a feedback about the system performance allowing to discover the strengths and weaknesses of the system. These are the reasons for which this thesis is oriented to the evaluation of IIM and feature extraction methods for high resolution SAR images.

1.2 Goals of the thesis

The main goal of the thesis is the evaluation of Image Information Mining methods for high resolution SAR images. The evaluation must take into account objective measurements as well as user criteria. Behind this general objective lies three other important goals:

- to have an overall evaluation procedure that enables quantifying the quality of information extraction methods using high resolution SAR image.
- to have performance indices of each component of the Image Information Mining system when dealing with large databases of very high resolution SAR imagery.
- to elaborate validation scenarios involving the end-user in the creation and evaluation.

1.3 Contribution of the thesis

An IIM system using high resolution SAR data has to face large image repositories rich in volume and content. Actually, in order to deal with these issues, the IIM systems are organized as hierarchical levels starting from the low level (data base) to high level (user interaction). In the low level, the automatic extraction of the information content is performed by different methods. Here, the accuracy of the results depends on an adequate modeling and understanding of the image content by the method. In this context, the **first contribution of this thesis is the data-driven evaluation of despeckling and feature extraction methods using high resolution TerraSAR-X data**, which is achieved through a deep study of high resolution SAR data properties and how the stochastic models can represent and characterize the image content. In practice, the stochastic models: Gauss Markov Random Field (GMRF) and Auto-binomial model (ABM) are used to model the image content. The performed detection tests show that they are able to recognize different kinds of structures as for example vegetation, water bodies, and urban area. In order to assess the models, we design several experiments and propose representative metrics for qualifying (the quality of) the results. The data-driven evaluation leads us to provide recommendations of usage of the models according to the type of image content, that can further be translated by the user in a real tool for automatic information extraction.

Once the feasibility of the methods is studied, they are tested with a high number of SAR images, which really point out the robustness of the information extraction method used within the context of IIM and automatic interpretation and understanding of the image. Therefore, in the high level of an IIM system, the user accesses to the image repository using a graphical interface, which allows the interactive human-machine communication. Here, the user creates different queries reflecting his interests and conjectures. Thus, **the second contribution of this thesis, the user-driven evaluation, focuses on the creation of real application scenarios with the collaboration of the end-user in the configuration and evaluation.** In practice, the end-user is involved in the creation of the two evaluation scenarios in the framework of disaster monitoring: oil spill and flood detection. The configuration of the scenarios is performed according to recommendations provided in the data-driven evaluation. In order to have a effectiveness index of the retrieval process, quantitative metrics such as precision and recall are used. In order to have measurements about the user satisfaction degree regarding the retrieved results, a group of evaluators are asked to qualitatively rank the retrieved results.

1.4 Outline of the thesis

In **chapter 2**, we deal with the main concepts of SAR image acquisition and formation. Here, we explain the usefulness of the side-looking geometry and synthetic aperture, in increasing the image resolution. We also present the backscattering mechanisms as well as the geometrical and radiometrical properties. We also review the statistics of SAR images since they are used in the feature extraction methods. In the last part of this chapter, the study of TerraSAR-X data is of particular interest since all the experiments in this work are performed using these images.

In **chapter 3**, we give an introduction into CBIR and IIM systems by presenting the main architecture, their components, and examples of this kind of systems. In this chapter, we also explain, the problematic of the evaluation in CBIR analyzing different approaches suggested in previous works.

In **chapter 4**, we study in detail the analysis of SAR images in terms of despeckling and information extraction. Here, we explain the Bayesian inference approach and the basic concept of information extraction via stochastic modeling using Gibbs Random Field models; specifically we detail Gauss Markov Random Field and Auto-binomial model as prior models for image characterization and texture modeling. In the last part of this chapter, we also provide a review of SAR image analysis involving despeckling and information extraction.

In **chapter 5**, we contribute with the data-driven evaluation of despeckling and feature extraction methods based on Bayesian inference using TerraSAR-X images. Here, we start by describing the evaluated methods, which are based on Gauss Markov Random Field and Auto-binomial model. Both methods are adequate for modeling high resolution SAR images. In order to perform the assessment, some quantitative measurements are proposed for evaluating the quality of the feature extraction and despeckling. Moreover, we carry out the impact of model parameters such as model order and estimation window on the performance of the methods.

In **chapter 6**, we elaborate the user-driven evaluation of the KIM system. Here, we propose an evaluation methodology in order to deal with the user involvement, his conjectures and opinions, during the assessment. We introduce KIM as an IIM system by

proving a complete description of its functionalities as well as an example of operation. In this chapter, we also suggest two validation scenarios in order to carry out the evaluation of KIM functions. Here, KIM is evaluated in terms of effectiveness metrics and the user criteria. The degree of user satisfaction is taken into account as a performance metric in the evaluation.

Finally, **chapter 7** gives some conclusions summarizing the main results and suggests some perspectives for future work.

Chapter 2

Synthetic Aperture Radar

Synthetic Aperture Radar (SAR) images have been used during the past years for a large number of applications in different fields such as mapping, surface surveillance, search and rescue, target recognition, etc. In contrast to optical systems, radar technology allows us to record images through clouds, during night and day. Thus, image acquisitions can be planned for any time.

In this chapter, we start with describing the fundamentals of SAR technology, specifically the side-looking geometry, image formation, resolution in range and azimuth in section 2.1. Then, in section 2.2, the geometrical and radiometrical distortion of the SAR images, in addition to the radiometric correction are explained. The statistical properties of SAR data in a complex and multi-look domain are presented in section 2.3, and finally, the description of TerraSAR-X products and information content are detailed in section 2.4. Here, it exhibits the characteristic of typical urban-area and natural scenes in TerraSAR-X images.

2.1 Fundamentals of SAR

RADAR is an acronym used for **RA**dar **D**etection **A**nd **R**anging. It refers to a method developed for target acquisitions and distance measurements by means of radio or microwave measurements. A radar system has three primary functions: 1) It transmits microwave (radio) signals towards a scene; 2) It receives the portion of the transmitted energy backscattered from the scene; and 3) It observes the strength (detection) and the time delay (ranging) of the return signals. Radar provides its own energy source and, therefore, can be operated both during day or night and through cloud cover. Hence, images can be acquired completely independently of solar illumination, which is particularly important in high latitudes (polar night) or cloudy regions. This type of system is known as an **active remote sensing system**.

The modern uses of radar are highly diverse, including air traffic control, radar astronomy, military application such as : air-defense systems, antimissile systems, nautical radars to locate landmarks and other ships, guided-missile target-locating systems, etc. Moreover, during the last twenty years, radar has been widely used in remote sensing imagery and Earth observation.

Imaging radar is an active illumination system, which is used for image generation. It has an antenna, mounted on a platform, which transmits a radar signal in a side-looking direction towards the Earth's surface. The reflected signal, known as the echo, is backscat-

tered from the surface and received a fraction of a second later at the same antenna. Two types of imaging radar systems with side-looking view can be distinguished 1) systems with real aperture (RAR: Real Aperture Radar) and 2) systems using synthetic aperture radar (SAR: Synthetic Aperture Radar). In the radar context, aperture means the length of the antenna. In the following, we focus on the use of SAR images and point out only the most important characteristics of SAR systems. More details of SAR systems can be found on (Bamler and Schättler, 1993), (Oliver and Quegan, 1998).

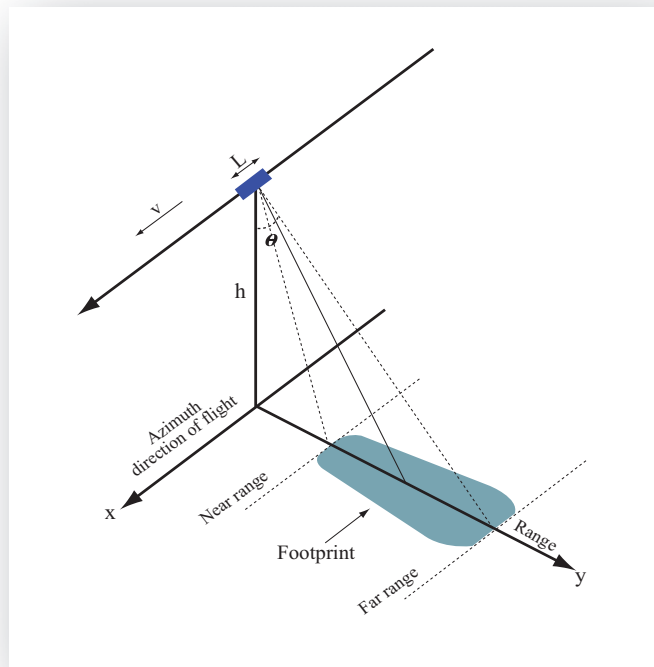


Figure 2.1: Geometry of a side-looking radar system used for surface imaging. The radar with an antenna of length L is flying at altitude h with speed v . The incidence angle between the radar beam and the ground surface is θ . The direction of the flight or azimuth is denoted by x and the range direction is denoted by y . The resolution in azimuth and range are denoted by R_x and R_y , respectively.

2.1.1 Principle of SAR

The basic geometry of a SAR is shown in Figure 2.1. A platform moving with a velocity v at altitude h carries a side-looking radar antenna that illuminates the surface of the Earth with pulses of electromagnetic radiation. The swath illuminated by the microwave beam is called **footprint**. The direction of travel of the platform is known as the **azimuth** and is denoted by x . The distance from the radar track is measured in the **range** direction and is denoted by y (Oliver and Quegan, 1998). The incidence angle θ is the angle between the radar beam and ground surface, which increases, moving across the swath from near to far range. The antenna length is denoted by L . For sake of simplicity, we deal only with

the case where the antenna is oriented parallel to the flight line.

2.1.2 SAR image formation

Since SAR is an active system, the actual sensor resolution has two dimensions: range resolution and azimuth resolution. Resolution of a SAR sensor should not be confused with pixel spacing which results from sampling done by the SAR image processor.

2.1.2.1 Range resolution

The **range resolution corresponds to the minimum distance between two distinguishable different objects**. Radar data are created in the slant range domain, but usually are projected onto the ground range plane when processed into an image. Therefore, in range, two resolutions: **slant and ground range resolutions** are considered.

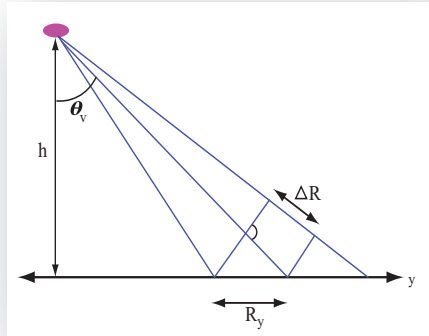


Figure 2.2: Slant (ΔR) and ground (R_y) range resolutions.

Figure 2.2 shows a radar system mounted on a platform, either airborne or spaceborne. Here, the sensor is flying at altitude h above ground, its distance from the target to be imaged is R_0 and the incidence angle is denoted by θ . Hence, an emitted pulse is received after $\Delta t = 2R_0/c$, where c represents the speed of light ($c \approx 3,0 \cdot 10^8 m/s$). We assume that the emitted waves can be described by rectangular pulses of duration τ that are repeatedly sent with the pulse repetition frequency $f_T = 1/T$, with $T \gg \tau$.

In this context, the **slant range resolution** can be calculated as

$$\Delta R = \frac{c\tau}{2}, \quad (2.1)$$

which is dependent on the length of the processed pulse. In the case of **ground range resolution**, it results in (See Figure 2.2)

$$R_y = \frac{c\tau}{2 \sin \theta}. \quad (2.2)$$

It is worth to note that R_y is independent of h and that better resolutions are achieved by larger values of θ . Hence, resolution is improved with growing range distance y . Since θ

is constrained to a certain range, e.g. due to the signal-to-noise ratio, the main parameter to increase resolution remains the pulse length τ . The emitted pulses must be as short as possible, but at the same time, they must contain enough energy to guarantee a sufficient amount of reflected intensity at the receiver.

2.1.2.2 Azimuth resolution

SAR does get its name from the azimuth processing, which can achieve an azimuth resolution of maybe hundreds of times smaller than the transmitted antenna beam width.

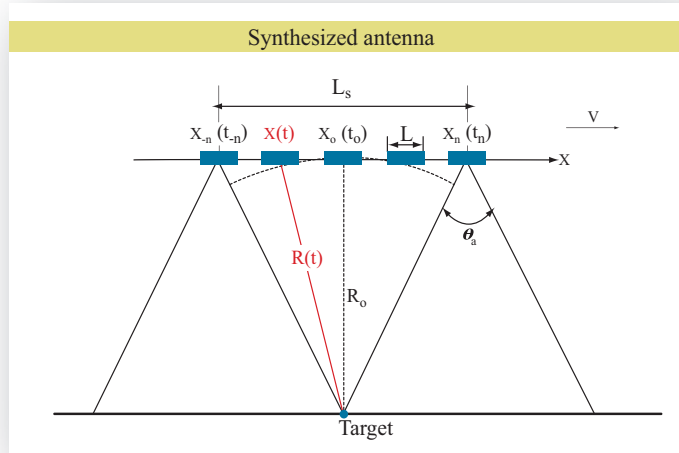


Figure 2.3: Azimuth resolution.

Azimuth resolution describes the ability of an imaging radar to separate two closely spaced scatterers in the flight direction. The resolution in azimuth denoted by R_x depends on the slant range distance R_0 to the target (T) and on the antenna aperture in azimuth θ_a . The width of the main antenna lobe at -3dB, conventionally computed at -3dB, is given by

$$\theta_a \approx \frac{\lambda}{L}, \quad (2.3)$$

where λ is the wavelength of the emitted signal and L denotes the antenna length in azimuth. Then, the resolution in azimuth direction can be shown as

$$R_x = R_0 \theta_a \approx \frac{R_0 \lambda}{L}. \quad (2.4)$$

In (2.4)¹ it can be observed that the resolution depends mainly on the antenna length L and, moreover on the wavelength and range, so that finer azimuth resolution can be achieved with longer antenna, and therefore, the resolution is limited by the antenna length. However, in practice, it is impossible to build an antenna large enough to produce high resolution data.

¹Here, R_x is referred for azimuth resolution in RAR systems, where two targets in the azimuth or along-track direction can be separated only if the distance between them is larger than the radar beam width.

In order to improve the resolution in azimuth and to overcome the limitation of antenna length, **SAR** synthesizes electronically an extremely long antenna or aperture by using the forward motion of a small antenna (with length L) and a special recording and processing of the backscattered echoes. Figure 2.3 illustrates how the long synthesized antenna is achieved. As a target T first enters the radar beam (t_{-n}), the backscattered echoes from each transmitted pulse begin to be recorded. Then, while the radar continues to move forward, all echoes from target T for each pulse are recorded during the entire illumination time. The point at which the target leaves the view of the radar beam (t_n), determines the length of the synthesized antenna L_s . Hence, L_s is equivalent to the beam width

$$L_s = R_0 \theta_a \approx \frac{R_0 \lambda}{L}. \quad (2.5)$$

From Figure 2.3 we could also observe that during the illumination time of the target T , the range distance R , from the radar to the target will be different for every received pulse. The change in range from pulse to pulse may only be a few millimeters, but that is enough to give the signal, which is received from the target, a different phase at each pulse. This change in phase is analogous to the **Doppler effect**. Flying at speed v , the radar travels the distance $\Delta x = v\Delta t$ within the time Δt . The radial distance R_0 obtained at time t_0 when the radar is at x_0 , is minimal. Since $\Delta x \ll R_0$, the distance $R(t)$ to the target as a function of the time $t = x/v$ can be approximated by

$$R(t) \approx R_0 + \frac{v^2 t^2}{2R_0}, \quad (2.6)$$

which results in a phase (φ) of

$$\varphi(t) = -\frac{4\pi}{\lambda} R(t) \approx -\frac{4\pi}{\lambda} \left(R_0 + \frac{v^2 t^2}{2R_0} \right). \quad (2.7)$$

The doppler frequency f_D at a distance x can later be written as a linear function of the time t

$$f_D(t) = -\frac{1}{2\pi} \frac{d}{dt} \varphi(t) \approx \frac{2v^2}{\lambda R_0 t}, \quad (2.8)$$

with $t \in [t_{-n}, t_n]$. The formula (2.8) can be interpreted as another linear modulation of the frequency as a function of the time t , and could thus be processed in the same way as the range direction. The bandwidth of the doppler frequency equals

$$Bd = \frac{2v^2 |t_n - t_{-n}|}{\lambda R_0}. \quad (2.9)$$

Since $|t_n - t_{-n}| = L_s/v \approx R_0 \lambda / vL$, then, (2.9) is given by

$$Bd = 2 \frac{v}{L}. \quad (2.10)$$

Similarly to (2.4), the resolution in azimuth equals

$$R_x = \frac{v}{Bd} = \frac{L}{2}. \quad (2.11)$$

Therefore, the azimuth resolution depends only on the length of the small antenna L and no more on the wavelength and range.

2.1.3 Basic scattering mechanisms

The principle of radar system is to emit electromagnetic waves with wavelengths ranging from a few centimeters up to one decimeter and to receive the backscattered reflection from the ground surface. Thus, SAR images represent an estimate of the radar backscatterer to a specific area in the ground. For example, darker areas in the image represent low backscatter, while brighter areas represent high backscatter. Figure 2.4 illustrates examples of different levels of backscattering according to the target area. Urban areas appear as bright features meaning that a large fraction of the radar energy was reflected back to the radar, while dark features (asphalt in a road) imply that very little energy was reflected. In the case of vegetation, it appears as medium bright since not all the signal is backscattered



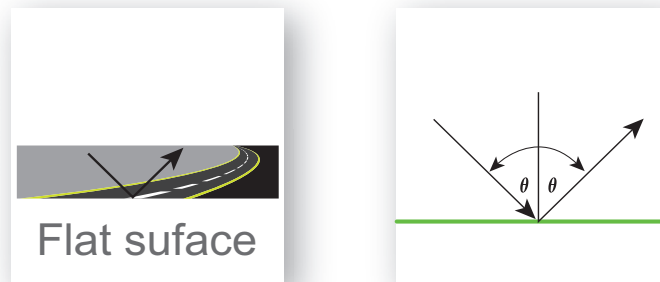
Figure 2.4: Examples of different levels of backscattering. Flat surface will appear as dark area in a SAR image. Forest areas, reflects medium backscattering. Backscattering over cities will appear as bright area in a SAR image.

Generally, backscatter for a target area at a particular wavelength will vary for a variety of conditions, such as the physical size of the scatterers in the target area, the target's electrical properties and the moisture content, with wetter objects appearing bright, and drier targets appearing dark. However, the exception to this is a smooth body of water, which will act as a flat surface and reflect incoming pulses away from the sensor. These bodies will appear dark. The wavelength and polarization of the SAR pulses, and the observation angles will also affect backscatter.

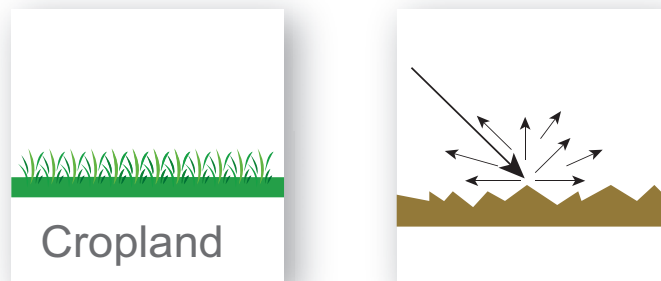
2.1.3.1 Surface and Volume Scattering

Flat and dry surfaces (smooth surfaces) reflect little or no radio or microwave energy back toward the radar, they will appear dark in radar images. However, **rough surfaces** reflect the wavefront in multiple directions, then, only a part of the emitted energy is received by the sensor. This effect is called diffuse scattering. Figures 2.5(a)-(b) show examples of smooth and rough surfaces, respectively. Smooth surfaces reflect very few of the incoming energy back to the emitting antenna, when the incidence angle θ between sensor and surface is zero and if no double bouncing occurs.

In **volume scattering**, the same effect could be seen, e.g. vegetation is usually moderately rough on the scale of most radar wavelengths, then the wavefront partly penetrates the scatterers. It is illustrated in Figure 2.6.



(a) Smooth surface.



(b) Rough surface.

Figure 2.5: Smooth and Rough surfaces. (a) Flat surfaces reflect little or no energy back toward the radar, while (b) rough surfaces reflect in multiple directions.

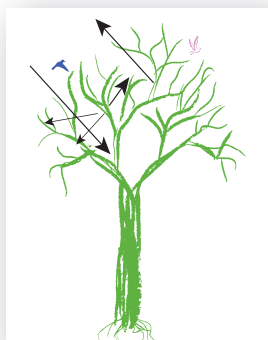


Figure 2.6: Volume scattering.

2.1.3.2 Single and Double Bounce

Figure 2.7 depicts examples of single and double bounce. **Single bounce** occurs when the entire power is reflected back to the sensor like a mirror. **Double bounce** occurs when there are two surfaces, one flat on the ground (horizontal) and the other upright (vertical), and the reflected pulse hits both surfaces one after the other as shown in Figure 2.7, i.e. the grass (horizontal) hits the tree's stump (vertical). This type of scattering will appear light gray to white in the radar image.



Figure 2.7: Single and double bounce.

Another example is, when city streets or buildings are lined up in such a way that the incoming radar pulses are able to bounce off the streets and then bounce again off the buildings (called a **double-bounce**) and directly back towards the radar. They will also appear very bright (white) in radar images. Roads and freeways are flat surfaces and will therefore appear dark. Buildings which do not line up so that the radar pulses are reflected straight, back will appear light grey, like very rough surfaces.

2.2 Geometrical and Radiometrical effects

SAR systems detect the distance between the sensor emitting a microwave pulse and the target reflecting the energy back to the receiving antenna. This range measurement principle leads to specific geometric distortions in the processed SAR image, which makes SAR images more difficult to interpret than optical images.

In the following, the most common geometric distortions presented in the specific topography are described in section 2.2.1. A review of the radiometric and geometric corrections are detailed in section 2.2.2 and 2.2.3, respectively.

2.2.1 Geometrical effects

Figures 2.8-2.10 sketch the principal geometrical effects in SAR images. Here, the points a , b , and c are imaged as a' , b' , and c' in the slant range plane. This shows how minor differences in elevation can cause considerable range distortions. These induced effects are called foreshortening, layover and shadow and are described in the following.

2.2.1.1 Foreshortening

Foreshortening is a geometrical effect that causes slant range differences between two points located on foreslopes of mountains to be smaller than they would be in flat areas. Foreshortening occurs specially in mountainous areas, where the mountains seem to lean towards the sensor. An example is shown in Figure 2.8. Here, the point a , b and c are equally spaced when vertically projected on the ground. However, the distance between a' and b' is considerable shortened compared to b' and c' , because the top of the mountain is relatively closer to the SAR sensor.

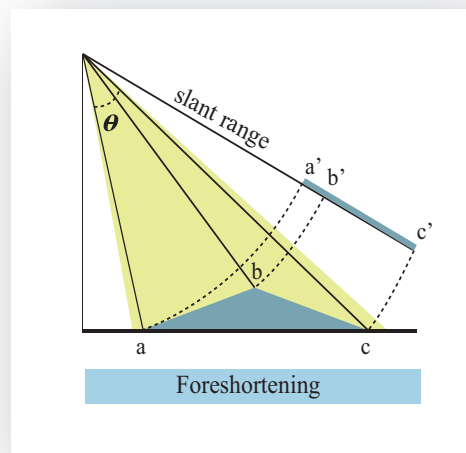


Figure 2.8: Foreshortening effect. The point a , b and c are equally spaced. However, the distance between a' and b' is considerable shortened compared to b' and c' , because the top of the mountain is relatively closer to the SAR sensor.

The foreshortening situation has two consequences: 1) Due to the change in ground resolution, foreshortening areas appear compressed in SAR images, i.e. their extension in range direction is reduced; and 2) foreshortening areas are characterized by brighter image gray values. The received energy within a resolution cell is higher due to the larger imaged area.

2.2.1.2 Layover

Layover is an extreme case of foreshortening and occurs in the case of very steep terrain slopes, when the top of a mountain is closer to the sensor than its bottom. Layover areas appear as particularly bright regions in the image with an inverted geometrical order. In Figure 2.9, this is depicted by points b' and a' . Note that the ordering of surface elements on the radar image is the reverse of the ordering on the ground.

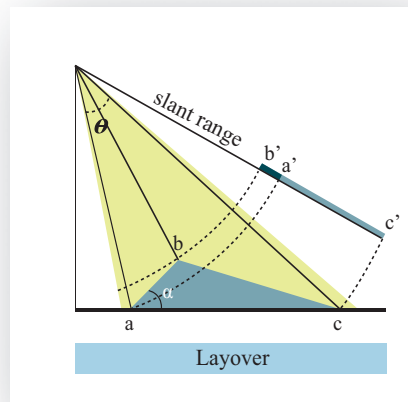


Figure 2.9: Layover effects. The ordering of surface elements on the radar image is the reverse of the ordering on the ground. Points b' and a' appear inverted.

2.2.1.3 Shadows

Shadows occur in the areas, which are not illuminated by the radar beam. Points b and c in Figure 2.10 illustrate the shadow effect.

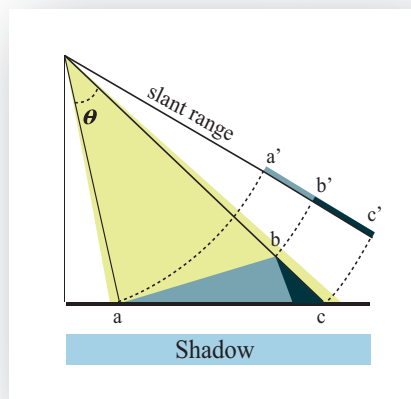


Figure 2.10: Shadow effect is caused by objects, which cover part of the terrain behind them.

2.2.2 Radiometric corrections

For radiometric correction of SAR layover areas, the energy accumulated in one image pixel has to be redistributed among those ground resolution cells, which are mapped onto that particular pixel. This problem of energy redistribution is also encountered in foreshortening areas, where each image pixel contains the energy from a larger area, comprising generally more than one ground resolution cell.

Radiometric correction of SAR data are performed by using Digital Elevation Model (DEM) information. Basically, the radiometric correction can therefore be performed by a division by the size of the imaged area, which, as a first approximation, is a function of the local incidence angle. In general, the local incidence angle is calculated by using DEMs of similar or, if available, higher resolution.

2.2.3 Geometric corrections

Two different range measures are commonly used for SAR Images: **slant range** and **ground range**. The first one refers to the actual SAR measurement subject to any calibration corrections made in the processing. However, the geometry is distorted to a map projection. Alternatively, the data can be geocoded to ground range so that the geometry is correct in the map projection. The geometric corrections or geocoding (Schreier, 1993) is the procedure of minimizing geometrical distortions and resampling the image to a homogeneous, predefined map grid, e.g. to the Universal Transversal Mercator (UTM) grid.

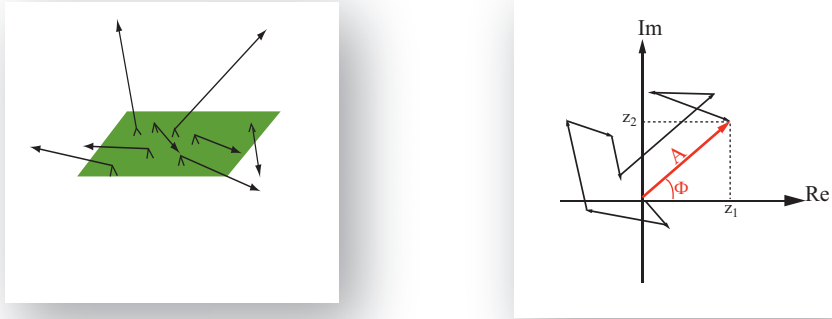
In order to perform this operation, the exact elevation of each pixel must be known which is usually taken from a DEM. It is clear that the success of accurate geocoding, as for radiometric correction, depends on the DEM quality and its resolution. The purpose of geocoding is to generate a map-like representation of the satellite image, where the SAR image is aligned with a Cartesian map projection grid. Otherwise, uncorrected data can hardly be interpreted due to effects in slant range images, such as mountains which appear to be leaning towards the sensor.

2.3 SAR image statistics

SAR images are affected by a multiplicative process called **speckle**, which is produced by coherent systems such as SAR or laser systems (Goodman, 1975). It is caused by random interferences, either constructive or destructive, between the electromagnetic waves which are reflected from different scatterers present in the imaged area as shown in Figure 2.11(a). When illuminated by a SAR instrument, each sub-pixel scatterer contributes to backscatter energy which, along with phase and power changes, is then coherently summed for all local scatterers (Oliver and Quegan, 1998). It becomes evident as a point-like random structure affecting the amplitudes of the image pixels as shown in Figure 2.11(b).

Speckle becomes visible only in the detected amplitude or intensity signal. The complex signal by itself is distorted by thermal noise and signal processing induced effects only.

Since the speckle presence in SAR images reduces the detectability of objects in the image and also the capability to separate and classify distributed targets, as a consequence, the interpretation of detected SAR images is highly disturbed. The speckle hides



(a) Speckle effect inside of a resolution cell. (b) Speckle effect seen in the complex plane.

Figure 2.11: Speckle is represented as a random walk in the complex plane.

this information contained in the image and makes the extraction of important information within the image difficult. Therefore, speckle reduction has been the focus of many works, which are detailed in section 4.3.

In the following, we start by presenting in section 2.3.1, the speckle effect, which causes a SAR image to be represented by a complex number. The statistics of the backscattered signal (complex image) and their derived probability density functions (pdfs) are presented in section 2.3.2. In the case of multilooked SAR images, their statistics are presented in section 2.3.3.

2.3.1 Speckle effect

The SAR data are known as the **complex image** since the estimation of the local reflectivity at each pixel can also be represented by the complex number $Ae^{i\Phi}$. From the complex images a variety of other products can be formed, for example, images of the real part $A \cos \Phi$, the imaginary part $A \sin \Phi$, the amplitude A , the phase Φ , the intensity $I = A^2$, the log intensity $\log I$. For a properly calibrated system, these are all true measurements of the scattering properties of the Earth's surface, but usually they produce quiet different representations. From all these image types, the **amplitude and log images are often preferred**.

An explanation of why the speckle effect arise is given by how the the world might appear to the radar nature since the SAR is making a true measurement of the Earth's scattering properties. In distributed targets we can think of each resolution cell as containing a number N of discrete scatterers as shown in Figure 2.11(a). As the wave interacts with the target, each scatterer k contributes a backscattered wave with a phase Φ_k and amplitude A_k . Therefore the total returned modulation of the incident wave is given by

$$Ae^{i\Phi} = \sum_{k=1}^N A_k e^{i\Phi_k}. \quad (2.12)$$

This summation (2.12) is over the number of scatterers illuminated by the beam. The

individual scattering amplitudes A_k and phases Φ_k are unobservable because the individual scatterers are on much smaller scales than the resolution of the SAR, and there are normally many such scatterers per resolution cell.

2.3.2 Statistical properties of the backscattered signal

An immediate conclusion from (2.12) is that the observed signals will be affected by interference effects as consequence of the phase differences between scatterers. Actually, speckle can be understood as an interference phenomenon in which the principal source of the noise-like quality of the observed data is the distribution of the phase terms Φ_k . To infer this distribution, note that the slant range resolution is typically many wavelength across. Hence, scatterers at different parts of the resolution cell will contribute very different phases to the return (2.12) even if their scattering behavior is identical. In practice, we can think of the phase Φ_k as being uniformly distributed in $[-\pi, \pi]$ and independent of the amplitude A_k . The sum (2.12) then looks like a **random walk** (See Figure 2.11(b)) in the complex plane, where each step of length A_k is in a completely random direction. For a large number of statistically identical scatterers (N), several properties of the received signal can be derived (Goodman, 1975; Oliver and Quegan, 1998):

1. Joint probability density function

The observed in-phase and quadrature components, $z_1 = A \cos \Phi$ and $z_2 = A \sin \Phi$, are independent identically distributed Gaussian random variables with zero-mean and whose variance $\sigma/2$ is determined by the scattering amplitudes A_k . Their joint probability density function (pdf) is given by:

$$P_{z_1, z_2}(z_1, z_2) = \frac{1}{\pi\sigma} \exp\left(-\frac{z_1^2 + z_2^2}{\sigma}\right). \quad (2.13)$$

2. Phase pdf

The observed phase Φ is uniformly distributed over range $[-\pi, \pi]$.

3. Amplitude pdf

The amplitude $A = |z| = \sqrt{z_1^2 + z_2^2}$ follows a Rayleigh distribution:

$$P_A(A) = \frac{2A}{\sigma} \exp\left(-\frac{A^2}{\sigma}\right) \quad A \geq 0 \quad (2.14)$$

with mean value $E\{A\} = \frac{\sqrt{\pi\sigma}}{2}$ and standard deviation $\sqrt{(1 - \frac{\pi}{4})\sigma}$. Another widely used statistic in the SAR community is the **Coefficient of Variation** (CV) defined as the ratio between the standard deviation and the mean. Hence, for amplitude data, this takes the value $\sqrt{4/\pi - 1} = 0.52$.

4. Intensity pdf

The observed intensity or power $I = A^2$ does have a negative exponential distribution:

$$P_I(I) = \frac{1}{\sigma} \exp\left(-\frac{I}{\sigma}\right) \quad I \geq 0 \quad (2.15)$$

with mean value and standard deviation equal to σ . Hence, in this case the CV is 1, showing the high statistical uncertainty of the measured intensity. This uncertainty is directly reflected in the noisy appearance of detected SAR images and is interpreted as speckle noise.

The distribution 1 to 4 are of fundamental importance in handling SAR data. It is worth noting that with exception of the phase distribution, they are completely characterized by a single parameter σ , which carries all the available information about the target and corresponds to the **average intensity**.

The implications of the presented speckle model are important. In fact, if our interest is on single SAR scenes of distributed targets, and hence phase provides no information (since its distribution is target independent), we can get rid of it and use only amplitude or intensity data. Actually, in general, the phase information results to be important only in the case of polarimetric and interferometric data, which is not the scope of this thesis. It is worth to mention that all the experiments developed in this work are based on amplitude data and the despeckled methods were adapted for this kind of SAR images.

2.3.3 Statistical properties of Multi-Look data

Speckle can not be completely removed by filtering using the multi-looking technique. Multi-look processing is applied during image formation, and this procedure averages several statistically independent looks of the same scene to reduce speckle (Porcello et al., 1976).

The multi-looking can either be performed in the frequency domain by extraction of complex looks and averaging them or, in the spatial domain, by averaging neighboring pixels. The variance of speckle is reduced by L , the number of incoherently summed looks or pixels. Hence, $Var(P)/1 = L$. This increase in radiometric resolution is gained at the expense of spatial resolution, which is reduced by the same factor. So, a compromise between spatial and radiometric resolution has to be found.

2.3.3.1 Intensity images

The separate images are referred to as looks, so that this process of averaging in intensity is known as multilooking and the resultant image is known as L-look. The L-look average intensity is given by

$$I = \frac{1}{L} \sum_{k=1}^L I_k \quad L \geq 0 \quad (2.16)$$

where the I_k are independent exponentially distributed variables with a mean σ , known to obey a Gamma distribution with order parameter L

$$P_I(I) = \frac{1}{\Gamma(L)} \left(\frac{L}{\sigma}\right)^L I^{L-1} \exp^{-LI/\sigma} \quad I \geq 0 \quad (2.17)$$

The average intensity has moments

$$\langle I^m \rangle = \frac{\Gamma(m+L)}{\Gamma(L)} \left(\frac{\sigma}{L}\right)^m \quad (2.18)$$

with special cases $\langle I \rangle = \sigma$ (mean) and $\text{var}(I) = \sigma^2/L$ (variance). The latter relation induces the definition of the **Equivalent Number of Looks (ENL)** as

$$ENL = \frac{(\text{mean})^2}{\text{variance}} \quad (2.19)$$

where the averages are carried out in intensity over a uniformly distributed target. The ENL is equivalent to the number of independent intensity values averaged per pixel. It is often applied not just to describe the original data but also to characterize the smoothing effects of post-processing operations such as speckle filtering (Oliver and Quegan, 1998). ENL should be computed in homogenous areas being recommended a window of 35×35 pixel size for its computation.

The conditional pdf also known as the **likelihood function** of the L-look average intensity I given its mean values u_I are obtained by changing the variables $E\{I\} = \sigma = u_I$ and this is given by

$$P_I(I|u_I) = \frac{1}{\Gamma(L)} \left(\frac{L}{u_I} \right)^L I^{L-1} \exp^{-LI/u_I}, \quad (2.20)$$

with conditional mean $E\{I|u_I\} = u_I$ and conditional variance $\text{var}(I|u_I) = (u_I)^2/L$.

2.3.3.2 Square-root intensity images

The multi-look square-root intensity is calculated by averaging L uncorrelated pixel with intensity I_k and taking the square root of a multi-looking intensity image. Therefore, the effect of reducing the variance is equivalent to

$$S = \sqrt{\frac{1}{L} \sum_{k=1}^L I_k} \quad (2.21)$$

The equation (2.21) is used for display purposes since this has been a reduced dynamic range. The conditional square root Gamma distribution is obtained by changing the variables $S = \sqrt{I}$ and $u_s = \sqrt{u_i}$, resulting in

$$P_S(S|u_s) = \frac{2}{\Gamma(L)} \left(\frac{L}{u_s^2} \right)^L S^{2L-1} \exp^{-LS^2/u_s^2}, \quad (2.22)$$

which is the Rayleigh distribution for L=1. The condition mean and variance are found to be

$$E\{S|u_s\} = \frac{\Gamma(L+0.5)}{\Gamma(L)\sqrt{L}} u_s \neq u_s \quad (2.23)$$

and

$$\text{var}(S|u_s) = u_s^2 - E\{S|u_s\}^2 = u_s^2 \left(1 - \frac{\Gamma(L+0.5)}{(\Gamma(L)\sqrt{L})} \right)^2. \quad (2.24)$$

This yields to the coefficient of variation

$$CV_S = \sqrt{\frac{\Gamma(L)^2 L}{\Gamma(L+0.5)^2} - 1}. \quad (2.25)$$

2.3.3.3 Amplitude images

In the case of multi-look amplitude images, the distribution cannot be derived since the multi-looked amplitude is computed by

$$A = \frac{1}{L} \sum_{k=1}^L \sqrt{I_k}. \quad (2.26)$$

The pdf of A is determined by a convolution of L Rayleigh distributions.

As a consequence, images of this kind are seldom used, statistical conclusions are difficult to be drawn, and accurate interpretation cannot be easily done. However, amplitude images exhibit one nice property: In contrast to square-root images, they easily permit the computation of L out of the measured coefficient of variation CV_A , which is given by

$$CV_A = \frac{\sigma_A^2}{u_A^2} = \frac{0.5227^2}{L}, \quad (2.27)$$

where σ_A denotes the standard deviation and u_A represents the mean of A. Since the coefficients of variation for amplitude and square-root images are almost identical; this allows for square-root intensity images to analytically calculate L from an analysis of mean and variance without having to deal with the Gamma functions of (2.25). Of course, it is also possible to directly estimate L from the intensity image.

2.3.4 The multiplicative noise model for Speckle

In (Touzi, 2002) the multiplicative and product speckle models were presented in order to approximate the speckle noise. However, the difference between them is very vague (Oliver and Quegan, 1998) being the most widely used model for speckle the multiplicative model, which can be modeled following the Gamma distribution (Goodman, 1975),(Oliver and Quegan, 1998). In the case of full developed speckle, it is characterized by a negative exponential distribution in a singlelook intensity SAR image or equivalently a Rayleigh distribution in an amplitude image (Goodman, 1975),(Ulaby et al., 1981). The probability density function of the intensity follows a Gamma distribution in multi-look images.

From the previous statistics, an interesting property of pdf (2.20) around a given mean value u_I is its interpretation as a **multiplicative noise** (Oliver and Quegan, 1998). It can be seen that the distribution of I with mean u_I is identical to the one obtained by multiplying a fixed cross-section u_I with a noise process n_I that is distributed according to $P_I(I|u_I) = 1$:

$$I = u_I \times n_I = u_I \times I_{u_I=1}. \quad (2.28)$$

Due to this property, speckle is considered to be a multiplicative noise with

$$P_{n_I}(I) = P_I(I|u_I = 1), \quad (2.29)$$

which manifests itself only in the amplitude or intensity signal.

By analogy to the intensity case considered before, in the case of square root intensity images, the speckle can be interpreted again as a multiplicative noise n_s affecting the mean value u_s with

$$S = \sqrt{I} = \sqrt{u_I \cdot n_I} = u_s \cdot n_s = u_s \cdot S_{u_s=1}. \quad (2.30)$$

As was previously pointed out, speckle found in SAR intensity or amplitude images can be regarded as a multiplicative process. However, many image processing techniques assume that an image is affected only by **additive noise** and perform poorly otherwise. This problem, which hinders the analysis of SAR images containing multiplicative speckle, can be partly overcome by taking the logarithm of the image prior to interpretation. The **log-operation transforms the speckle to being additive**

$$Z = \log I = \log u_I + \log n_I = u_z + n_z. \quad (2.31)$$

Speckle models can be incorporated in filtering techniques, and the most frequently assumed model is the multiplicative model. In most of the despeckling filters, the speckle noise is often assumed to follow a Gaussian or Gamma distribution. However, some despeckling techniques such as using wavelet decomposition transforms the speckle multiplicative model to additive model by using logarithms.

2.4 TerraSAR-X products and information content

TerraSAR-X is the German radar satellite launched on June 2007. It operates in the X-band and is a side-looking Synthetic Aperture Radar (SAR) based on active phased array antenna technology. It does supply high quality radar data for purposes of scientific observation of the Earth for a period of at least five years.

In the following sections, the description of the TerraSAR-X instrument and its orbit and altitude parameters are presented in section 2.4.1. The main TerraSAR-X available products are described in section 2.4.2. Finally, examples of typical image content are detailed in section 2.4.3. The following sections are based on the TerraSAR-X products description published in (DLR, 2009).

2.4.1 TerraSAR-X Instrument

The TerraSAR-X characteristic values of the platform and the SAR instrument are presented in Table 2.1 (DLR, 2009).

2.4.1.1 Imaging Modes

In the case of TerraSAR-X, four imaging modes have been designed to support a variety of applications ranging from medium resolution polarimetric imaging to high resolution mapping. The standard imaging modes defined for the generation of basic product are:

- **Stripmap** mode ("SM") in single or dual polarization
- **ScanSAR** mode ("SC") in single polarization
- **Spotlight** mode ("SL") in single or dual polarization
- **High Resolution Spotlight** mode ("HS") in single or dual polarization

The configuration of the imaging modes offers products with different resolution in azimuth and range, which are summarized in Table 2.2.

Table 2.1: TERRASAR-X System, Orbit and Attitude Parameters.

System Parameters	
Radar carrier frequency	9.65 GHz
Radiated RF Peak Power	2 kW
Incidence angle range for stripmap / Scan- SAR	20° –45° full performance (15-60 accessible)
Polarizations	HH, VH, HV, VV
Antenna length	4.8 m
Nominal look direction right Antenna width	0.7 m
Number of stripmap / ScanSAR elevation beams	12 (full performance range) 27 (access range)
Number of spotlight elevation beams	91 (full performance range) 122 (access range)
Number spotlight azimuth beams	229
Incidence angle range for spotlight modes	20° –55° full performance (15° – 60° accessible)
Pulse Repetition Frequency	(PRF) 2.0 kHz –6.5 kHz
Range Bandwidth	max. 150 MHz (300 MHz experimental)
Orbit and Attitude Parameters	
Nominal orbit height at the equator	514 km
Orbits / day	$15\frac{2}{11}$
Revisit time (orbit repeat cycle)	11 days
Inclination	97.44°
Ascending node equatorial crossing time	18 : 00 ± 0.25 h (local time)
Attitude steering	Total Zero Doppler Steering

Table 2.2: Basic TerraSAR-X Imaging Mode parameters versus Azimuth and Range resolution.

Imaging Mode	Incidence Ang. [deg.]	Azimuth Res. [m]	Ground Range Res. [m]
Stripmap (SM)	20 - 45		
Single Polarization		3.3	3.49 - 1.70
Dual Polarization		6.6	3.49 - 1.70
ScanSAR (SC)	20 - 45	18.5	3.49 - 1.70
Spotlight (SL)	20 - 55		
Single Polarization		1.7	3.49- 1.48
Dual Polarization		3.4	3.49- 1.48
High Resolution Spotlight (HS)	20 - 55		
Single Polarization		1.1	3.49 - 1.48
Dual Polarization		2.2	3.49 - 1.48

2.4.2 Basic TerraSAR-X products

In the context of TerraSAR-X ground segment, the SAR raw data are processed to basic products by the TerraSAR-X Multi Mode SAR Processor (TMSP), which generates a set of products depending on the appropriate algorithms and parameters as follows

- geometric and radiometric resolution
- geometric projection
- auxiliary information and annotation

2.4.2.1 Geometric resolution

The theoretical maximum slant range resolution of TerraSAR-X in single polarization is 0.89 meter based on the range bandwidth of 150 MHz if no spectral weighting is applied. However, for all products the maximum resolution is deliberately reduced by weighting the range and the azimuth spectrum with a Hamming window (α coefficient 0.6) in order to suppress the sidelobes of the point target response (PTR) function (DLR, 2009). This better side-lobe suppression is specifically important in imaging urban and industrial areas where the high spatial resolution of the system exposes high numbers of extremely strong scatterers leading to high image contrasts. As well, the level of azimuth ambiguities is decreased.

The **geometric reduction offers two possible configurations: Spatially Enhanced Products (SE) or Radiometrically Enhanced Products (RE)**; which are described in the following.

1. Spatially Enhanced Products (SE)

The spatially enhanced product is designed for the highest possible square ground resolution. Depending on the imaging mode, polarization and incidence angle the larger resolution value of azimuth or ground range determines the square pixel size. The smaller resolution value is adjusted to this size and the corresponding reduction of the bandwidth is used for speckle reduction.

2. Radiometrically Enhanced Products (RE)

The radiometrically enhanced product is optimized with respect to radiometry. The range and azimuth resolution are intentionally decreased to significantly reduce speckle by averaging approximately 6 (5 to 7) looks to obtain a radiometric resolution of about 1.5 dB. Therefore, the speckle is reduced at expense of spatial resolution reduction.

2.4.2.2 Geometric projections and data representation

Besides the different imaging modes, TerraSAR-X also allows a selection among different geometrical projections and data representations. These are described in the following:

1. Single look Slant range (complex-valued) (SSC)

This product is the basic single look product of the focused radar signal. The data are represented as complex numbers and the pixels are spaced equidistant in azimuth and in slant range. The SSC product is intended for scientific applications that require the

full bandwidth and the phase information, e.g. SAR interferometry and interferometric polarimetry. For this product, the geometric projection is **Azimuth - Slant Range** (time domain).

2. Multi-look Ground range Detected (MGD)

This product is a detected multi look product with reduced speckle and approximately square resolution cells on ground. The image coordinates are oriented along flight direction and along ground range. The pixel spacing is equidistant in azimuth and in ground range. A simple polynomial slant to ground projection is performed in range using a WGS84 ellipsoid and an average, constant terrain height parameter. The advantage of this product is that no image rotation to a map coordinate system is performed and interpolation artifacts are thus avoided. Consequently, the pixel localization accuracy is lower than in geo-coded products. The geometric projection is **Azimuth - Ground Range** (without terrain correction).

3. Geo-coded Ellipsoid Corrected (GEC)

This product is a multi look detected product. It is projected and re-sampled to the WGS84 reference ellipsoid assuming one average terrain height. Available grid formats are UTM (Universal Transverse Mercator) and UPS (Universal Polar Stereographic). The pixel location accuracy varies according to the terrain since the ellipsoid correction does not consider a DEM (Digital Elevation Model). The geometric projection is a **map geometry with ellipsoidal corrections only** (no terrain correction performed).

4. Enhanced Ellipsoid Corrected (EEC)

This product is a multi look detected product. It is projected and re-sampled to the WGS84 reference ellipsoid. The image distortions caused by varying terrain height are corrected using an external DEM. Available grid formats will be either UTM or UPS. The pixel localization in these products is highly accurate. However, the accuracy still depends on the type of terrain as well as the quality and resolution of the DEM and on the incidence angle. The geometric projection is **map geometry with terrain correction**, using a digital elevation model (DEM).

2.4.2.3 Product identification Scheme

The different basic products for TerraSAR-X are identified and classified by a scheme, which is split into 4 sub-identifiers. Thus, the **global product name** is composed as:

< **projection** > _ < **resolution class** > _ < **imaging mode** > _ < **polarization mode** >

The purpose of this scheme is to enable the identification of products that are compatible with respect to the imaging mode or polarization mode. Since the compatibility requirements depend on the intended application, a minimal set of radar and product parameters are used to compose a product identifier. Table 2.3 shows the values used in the identification scheme for TerraSAR-X basic products.

For example, a TerraSAR-X product with a specification of **MGD.SE.SM.S** stands for a Spatially Enhanced (**SE**) Single Polarization (**S**) Stripmap product (**SM**) in Multi Look Ground range projection (**MGD**). If a sub-identifier is not applicable, like the resolution class for complex products, it is omitted.

Table 2.3: TerraSAR-X Product identification scheme. The scheme is composed by **projection + resolution + imaging mode + polarization mode**.

Projection Identifier	
Projection, data representation	Identifier
Single Look Slant Range, Complex representation	SSC
Multi Look Ground Range, Detected representation	MGD
Geocoded Ellipsoid Corrected, detected representation	GEC
Enhanced Ellipsoid Corrected, detected representation	EEC
Resolution	
Geometric resolution	Identifier
Spatially Enhanced Products	SE
Radiometrically Enhanced Products	RE
Radar Imaging Mode Identifier	
Imaging Mode	Identifier
High Resolution Spotlight	HS
Spotlight	SL
Stripmap	SM
ScanSAR	SC
Polarization Mode Identifier	
Polarization Mode	Identifier
Single	S
Dual	D
Quad	Q
Twin	T

2.4.3 Examples of TerraSAR-X image content

Figure 2.12 shows typical examples of TerraSAR-X image content. Figure 2.12(a) and 2.12(b) present scenes with low backscatterers, which appear in **natural areas** such as rivers and lakes; while a medium brightness pattern appears in vegetation. The **man-made structures** are presented in Figure 2.12(c) and 2.12(d), where some industrial area and the Egypt pyramid are depicted. More detailed examples are presented in Appendix A.

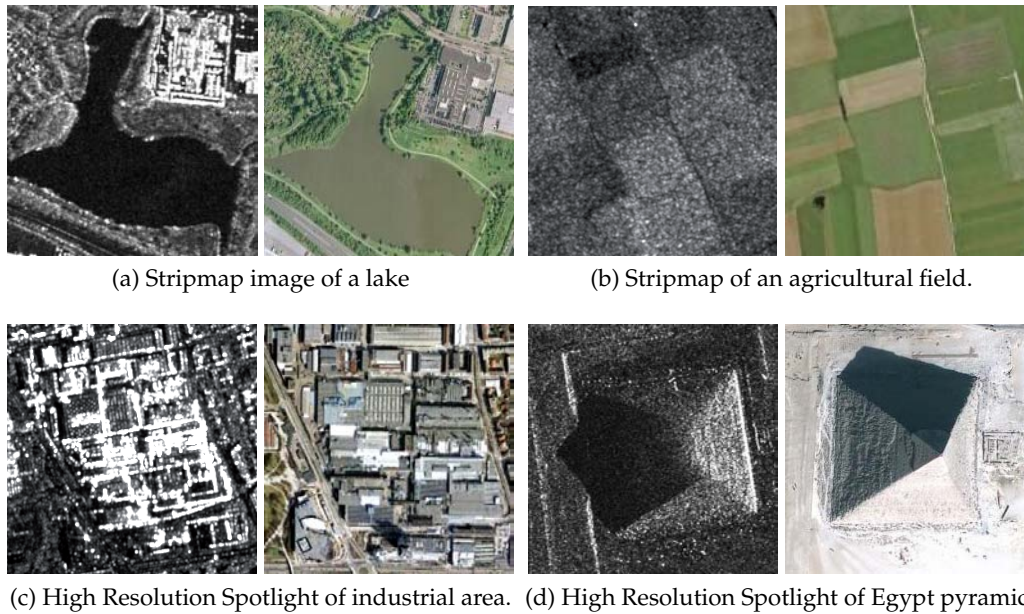


Figure 2.12: (left) TerraSAR-X images and (right) corresponding Google Earth. (a) The **water bodies** are the dark gray values in SAR image. (b) The image corresponds to an **agricultural field**, which is represented as flat homogenous areas with different levels of darkness. (c) The places with **industrial areas** appears with very high brightness. (d) The pyramid presents a 3D effect thanks to the high resolution of the image.

Chapter 3

Image Information Mining systems

Content-Based Image Retrieval (CBIR) systems allow to search for an image in a large database by giving as query the image content. The term **content** in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. This kind of systems have been widely used in many fields of applications like medicine, face recognition, military applications, remote sensing, etc, because they allow flexibility in the query without restriction to labels, text, coordinates, etc. Generally, CBIR systems are composed by several modules such as feature extraction, clustering, catalogue generation, semantic definition, and relevance feedback.

Later in early 2000 arose a new field of study, which integrates concepts of CBIR, Knowledge Database Discovery (KDD) and data mining. This new science is known as **Image Information Mining (IIM)**. An IIM system allows the user to access a large image repository, to extract and to infer knowledge about the patterns hidden in the images retrieving dynamically a collection of relevant images, where the user provides to the system high-levels of abstraction by defining semantics (Datcu et al., 2007). Thus **it enables the communication between heterogenous sources of information and the user with diverse interests at semantic abstraction.**

Many CBIR and IIM systems have been developed along the last two decades following different approaches and using a wide range of images. Although many algorithms, techniques and solutions have been proposed for successful information retrieval, only little attention has been paid to system verification and evaluation. Most of the evaluations in CBIR systems were carried out using their own image databases and with performance effectiveness measures that were arbitrarily selected to obtain good results.

Research groups like Text REtrieval Conference (TREC, 1999) or Evaluation framework for interactive Multimedia Information Retrieval Applications (MIRA, 1999) have been working on establishing a common benchmark test environment for testing Information Retrieval (IR) systems. For example, TREC addressed methodologies for evaluating document retrieval systems, while that MIRA work group stressed the user importance and user context in the evaluation and work with multimedia data. However, there is not a common research group in evaluating CBIR and IIM systems.

Evaluation of CBIR systems includes in general comparison of image retrieval methods, selection of adequate image features that best represent the content, automatic evaluation of the generated content-index and verification of the user relevance (Smith and Li, 1998). However, evaluation in CBIR systems becomes nowadays a complicated issue due to the growing complexity of the systems, the lack of standard performance measures, the subjective user interaction, the increasing resolution in both optical and radar

images, etc. Moreover, development and improvement of these systems are highly limited to an appropriate evaluation of their components and the overall performance. Thus an evaluation methodology for an IIM system, when it comes to deal with large databases of very high resolution imagery, and the user interaction is clearly justified.

This chapter presents a chronological review of CBIR systems showing the different systems and their components developed along the years in section 3.1. Since the main goal of this dissertation is the evaluation of IIM system, section 3.2 introduces IIM systems, their general architecture, main components and examples of IIM system implementations. The main part of this chapter is described in section 3.3, which focuses on the evaluation aspects in IIM presenting different approaches: data-driven and user-driven. Finally, section 3.4 gives some conclusions.

3.1 Content-Based Image Retrieval

CBIR systems became very important during the last two decades due to the large coverage of applications in different fields such as remote sensing systems, personal photos, medical imaging, geographical information, education and training, etc. CBIR was first proposed in the early 1990s mainly to overcome the problems of text-based image retrieval. (Kato, 1991) presented the **query by visual example** (QVE) and **query by subjective description** (QBD) as mechanisms of visual interaction for image database systems. Later, (Kato, 1992) formulated the term **Content-Based Image Retrieval** to describe his experiments into **automatic retrieval of images from a database** by color and shape features. The term has since then been widely used to describe the process of retrieving desired images from a large collection on the basis of features (such as color, texture and shape) that can be automatically extracted from the images themselves.

Along those years many systems have been developed like Query By Image Content system (QBIC) (Flickner et al., 1995) by IBM that retrieves images based on visual image content such as color percentages, color layout, and textures occurring in the images. Chabot (Ogle and Stonebraker, 1995) is a simple system which computes the color histogram and provides search by a combination of text and color. Excalibur system by Visual RetrievalWare (Dowe, 1993) used in Yahoo Image Surfer, allows several visual query features like color, shape, texture whose weights are specified by the user. A system used for image and video content based retrieval is VIRAGE (Gupta and Jain, 1997). VisualSEEK (Smith and Chang, 1996) and WebSeek (Smith and Chang, 1997) use diagramming spatial arrangements of color regions to form a query. PicToSeek (Gevers and Smeulders, 2000) combines the color and shape invariants for indexing and retrieving images, resulting in an excellent retrieval accuracy. However, all these systems hardly allow flexibility in the user requirements.

Later, CBIR systems incorporated **relevance feedback** to improve query formulation by changing the original query. More specifically the user is included in the retrieval loop, where the retrieval session is divided into few consecutive feedbacks, which are provided by the user about the retrieval result and where he specifies whether images are relevant or irrelevant (Kowalski and Maybury, 2000). Implementations of the relevance feedback were presented for instance in the Photobook (Pentland et al., 1996) developed by MIT, which has an interactive learning agent for retrieval. Later, MIT introduced the FourEyes (Minka and Picard, 1996) system, which is a more recent version of Photobook that combines multiple features. **FourEyes looks at the user's interaction** with the data

and learns the best features by letting the user give positive and negative example images instead of choosing the relative weight between the features. **This system is based on a "society of model"** for learning the interaction between the user and a selection of the image content. PicHunter (Cox et al., 2000) presented a general Bayesian framework for using a relevance feedback to direct a search. A high level resume of this phase is summarized in (Smeulders et al., 2000), which presents a review of 200 references in CBIR. More recent reviews are also provided by (Lew et al., 2006) and (Datta et al., 2008).

Although the involvement of the user in the CBIR evaluation scheme has improved its efficiency, it raises the problem of the so-called **semantic gap**. In (Smeulders et al., 2000) the semantic gap is defined as the lack of coincidence between the information that one can extract from visual data and the interpretation that the same has for a user in a given situation. Recently, the problem of the semantic gap between the user and the machines has become of a significant interest in **semantic image retrieval** (Vasconcelos, 2007). A semantic image retrieval system aims at two complementary goals: **image annotation and search**. A semantic retrieval system is based on the automatic annotation of images: the starting point for such systems is a training database of images, each annotated with a natural language caption. From this database, the system learns to create a mapping between words and visual features, and subsequently allows the user to perform queries also with the aid of keywords.

3.2 Image Information Mining Systems

Image Information Mining (IIM or I2M) is an interdisciplinary approach to automatize remote sensing analysis that draws upon expertise in computer vision, image understanding, data mining, machine learning, databases, artificial intelligence, and software design (Burl et al., 1999). The **main goals** of this field are to **seek solutions to automatize the extraction of information from Earth Observation (EO) archives** that can lead to image understanding and knowledge discovery (Datcu et al., 2007). An IIM system offers a user the opportunity to deal with a large collection of images by accessing in a large image database and allows the user to extract and infer knowledge about patterns hidden in the images, so that the set of relevant images is dynamic, subjective and unknown (Hsu et al., 2002).

In section 3.2.1 we present a description of the main components of IIM systems. All of these components are grouped into a generic architecture depicted in Figure 3.1. Section 3.2.2 introduces examples of IIM systems.

3.2.1 Generic Architecture

Figure 3.1 shows the typical architecture of IIM systems. Generally, an IIM system is composed of two fundamental parts: off-line part and online part. The tasks of storing the image into the database, to extract the main features, to perform the clustering algorithms, to create the index catalogue, and to save the catalogue in the database are executed in the off-line part, whilst the user interaction by given queries as well as the interactive training and the probabilistic search are performed in the online part.

A generic architecture of an IIM system has the following components: image database, feature extraction, data reduction and content index generation, user iteration or human-machine communication, which are described in detail in the next sections.

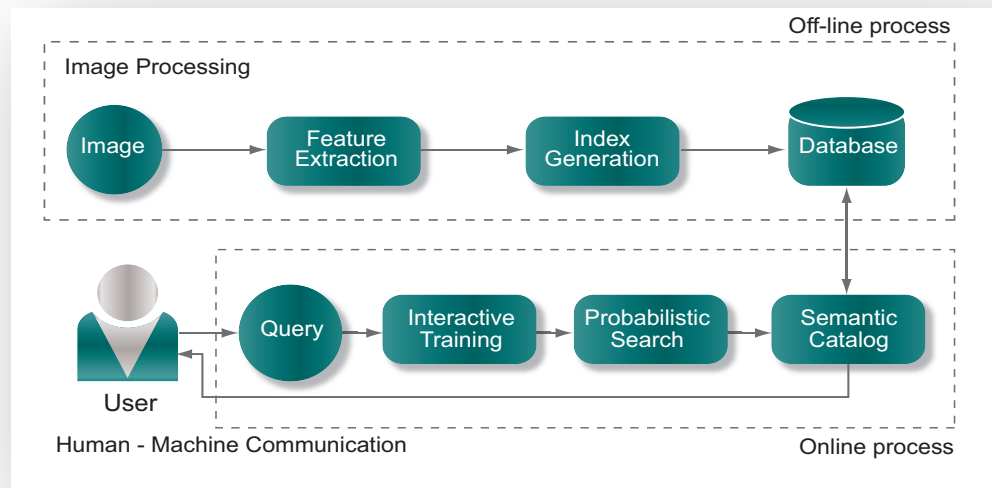


Figure 3.1: Main components of Image Information Mining systems. The off-line processes involve feature extraction, clustering and catalog generation, while the online processes enable the human-machine communication.

3.2.1.1 Image database

Today, Earth images from space are acquired continuously and have become powerful scientific tools to enable better understanding and improved management of the Earth and its environment. An EO image is an invaluable source of information since it provides the human being with perception and understanding about the scene that it represents.

Unlike image processing which focuses on understanding and extracting specific features from a single image, the focus of image mining is the automatic extraction of patterns from a large collection of images. The image database is formed by EO data provided by multiple sensor both optical and SAR as for example IKONOS, Landsat, TerraSAR-X, etc. Compared to another database, an image database presents some special characteristics (Hsu et al., 2002) like:

- The implicit spatial information is critical for the interpretation of the image content.
- Image characteristics could have multiple interpretation for the same visual patterns.
- The values themselves may not be significant unless the context support them.

3.2.1.2 Feature extraction

An image is characterized by its primitive features such as color, shape (Jain and Vailaya, 1996), texture (Tuceryan and Jain, 1998). Thus an image will be represented as a multi-dimensional feature vector acting as a signature. The amount of details in an image is increasing with the resolution. As a consequence more classes can be extracted from the image and more features/signatures are needed to characterize accurately an image content.

The **primitive feature extraction methods** aim at obtaining the best features for characterizing the image content, either using optical or radar images. When the IIM system uses radar images, the optical human capabilities to distinguish the image content become more complicated, therefore the feature extraction methods and their evaluation turn into an important issue.

3.2.1.3 Data reduction and content index generation

In IIM, after the multidimensional feature vector is obtained, images are assigned with a *suitable content based* description extracted from these features. The processes involved in this step are:

- **Clustering:** Similar extracted features are grouped together using a classification algorithm. Thus, pixels containing similar features belong to the same class. An example of a clustering algorithm is the k-means (MacQueen, 1967) method, which assigns each data point to the cluster whose center is nearest.
- **Catalogue Generation:** Once a clustering algorithm is performed, a data structure for indexing is created and stored in a database.

3.2.1.4 User interaction or human-machine communication

The interaction between the user and IIM systems is performed through the graphical interface, which enables the human-machine communication. The user defines the **semantic label** using words from natural language and associates it with a generic class from the catalogue. Thus, an image is represented by its objects and a semantic label is assigned to each of them allowing the system to retrieve the image content using predefined labels. The semantic labels correspond to a high level of abstraction of the image.

However, since the user creates the association between the semantic labels and signal classes there is always a **semantic gap**, which limits the use of IIM systems. Many methods have been presented trying to fill this gap as for instance the work of (Li and Bretschneider, 2007) which proposed a Bayesian network to infer the semantic concept of segmented scenes. A common technique to provide regions with semantic meaning consists in manual annotation as it was published in (Comaniciu and Meer, 2002). Schröder et al. (2000) presented a concept of interactive learning and probabilistic retrieval, where the user gives positive and negative examples redefining the query and probabilistic map.

A common technique used to human-machine communication is the **relevance feedback**, which is an automatic process used to improve the retrieval effectiveness. In the work of (Cox et al., 2000) the implementation of relevance feedback is based on Bayesian networks. This method models the user reaction to a certain target image and infers the probability of the target image based on the previously performed actions.

3.2.2 Examples of IIM systems

Several systems have been developed following the IIM approach for remote sensing imagery such as (Agouris et al., 1999), (Datcu et al., 2003), (Li and Narayanan, 2004), (Shyu et al., 2007). These systems have made significant contributions in the design of system frameworks for geo-spatial CBIR techniques with large-scale spatial databases.

- Agouris et al. (1999) presented an environment for CBIR from large spatial databases. This system enabled queries on shapes and topology. The authors presented the results obtained with the system but did not present an evaluation of the system.
- The integration of spectral and spatial information mining in remote sensing imagery was presented in (Li and Narayanan, 2004). The system is used in a specific application for land cover and land use classifications. The features are extracted using Gabor wavelets and the classification is performed through Super Vector Machine (SVM). The authors posed the necessity of clustering and retrieval performance analysis from a large image database. A set of quantitative experiments were performed in order to test the accuracy of the retrieval results provided by the system. Their experiments did not consider the user interaction and the subjective factor in defining the semantic labels.
- Datcu et al. (2003) introduced the **Knowledge-driven content based image information mining system called KIM**. It is based on a Bayesian approach, and its architecture is defined in terms of stochastic problem modelling of probabilities. The system is composed of two main parts: off-line process and on-line process. The features extraction, data reduction and catalog generation and ingestion in the image database are performed in the off-line process whilst in the on-line section the system presents a Graphical User Interface (GUI), which allows the interaction between the user and the system. The user is a main actor since he is the provider of the semantic interpretation of the image contents. This system has an application-free orientation using remote sensing imagery with optical as well as radar data. The evaluation of KIM was presented in (Daschiel and Datcu, 2005) using optical images. A complete description of KIM is provided in chapter 6 since it has been used during the subjective evaluation of this dissertation.
- The **Geospatial Information Retrieval and Indexing (GeoIRIS)** system (Shyu et al., 2007) was introduced in 2007. It enables scalable processing and retrieval of a large volume of data by automatically preprocessing and indexing satellite images. This system incorporates automatic feature extraction methods, visual content mining from a large-scale image database, and high dimensional indexing for fast retrieval. In addition, GeoIRIS includes techniques for complex queries that merge the information from heterogeneous geospatial databases, retrieves the objects based on shape and visual characteristics and a semantic model to link low level image features with high-level visual descriptors. However, the performance evaluation of this system has not yet been carried out.

3.3 Evaluation in Image Information Mining systems

Along the years several efforts have been made in improving techniques in the different components of the CBIR and IIM systems, for instance new techniques in feature extrac-

tion (Ma and Shan, 2008), clustering and relevance feedback. However, no methodologies have been proposed for evaluating and validating the systems taking into account the user satisfaction. Thus, the evaluation issue is almost non-existent in most of the literature. Most of the evaluations in those systems consist only in few examples of results. A single example result does not reveal a great deal about the real performance of the systems or of the methods. In addition the query could be chosen by the author so that it provides the best possible results.

Nowadays with the numerous competitive techniques in CBIR, the evaluation has become a very important issue. **Actually, the evaluation methods proposed in Information Retrieval (IR) could still be used in CBIR since the two systems are highly linked.** Therefore, within the framework of IR during the past 20 years the common evaluating methods were the Cranfield (Cleverdon, 1991; Cleverdon et al., 1966) model and Precision and Recall curves (Harman, 1993).

Saracevic (1995) reviewed the history and nature of evaluation in IR and describes six different levels of IR evaluation from system to user levels. However, most IR evaluations are only based on the system level(s) and lack user-centered evaluation. To achieve a more comprehensive picture of IR performance and users needs, both system and user-centered evaluations are needed.

Jermyn et al. (2002) mentioned that the development of evaluation methodologies for image retrieval systems becomes a matter of priority. Here, the analysis of possible evaluation methodologies, indicating the pros and cons and some constructive critics were presented. The analysis is based on knowledge scenarios with different degrees of structure.

Datta et al. (2008) proposed that an evaluation strategy should consider the following aspects:

- An appropriate dataset for evaluation, which should be general enough to cover a large range of semantics from a human point-of-view and also should be large enough for the evaluation to be statistically significant.
- A ground-truth for relevance. Ground-truth is a very subjective issue, especially for multimedia. Since people usually associate a given picture with a wide range of high-level semantics.
- An appropriate metric and criteria for evaluating competitive approaches. The evaluation criteria should try to model human requirements from a population perspective.

The review of the literature and the state of the art in the evaluation of CBIR systems reveals the importance of the user involvement in the evaluation process. In this context, the evaluation of CBIR and IIM systems can be performed in terms of the performances of the individual modules as objective approach and taking into account the user criteria as subjective approach (Daschiel and Datcu, 2005).

In this thesis, Figure 3.2 depicts how the evaluation of CBIR and IIM can be organized. The overall evaluation of the system is divided in the data-driven (objective) evaluation presented in section 3.3.1, and the user-driven (subjective) evaluation approach detailed in section 3.3.2.

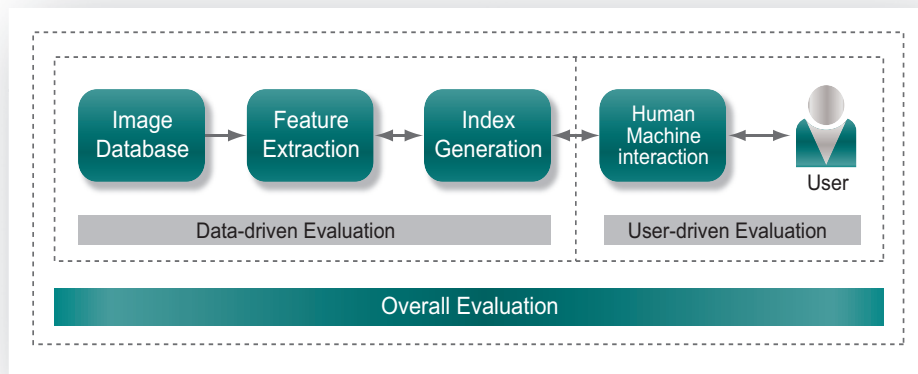


Figure 3.2: Evaluation in Content-Based Image Retrieval systems. The overall evaluation is composed of the objective and the subjective evaluation.

3.3.1 Data-driven evaluation approach

Smeulders et al. (2000) posed that with the growing complexity of the CBIR systems composed of several modules, it is necessary to evaluate system parts individually as well as their mutual dependencies. Therefore, the objective evaluation can be analyzed considering two approaches: 1) Component by component evaluation, and 2) general evaluation of the performance.

3.3.1.1 Component by component evaluation

Figure 3.1 shows the generic IIM system architecture, where the components in the off-line part are image database, feature extraction and index generation, so that the evaluation of each component consists of the evaluation of feature extraction methods and clustering methods.

Several works have been presented in the **evaluation of feature extraction methods** using optical and SAR images and their different features like color, shape, texture. As for instance works **using optical images** and texture as feature were presented in (Sharma and Singh, 2001), (Razniewski and Strzelecki, 2005). The authors of (Sharma and Singh, 2001) compared five feature extraction methods (autocorrelation, edge frequency, primitive-length, Law's method, and co-occurrence matrices) for image analysis using artificial and natural textures. (Razniewski and Strzelecki, 2005) described a study on feature selection methods for classification purposes. The authors compared texture features obtained using mutual information with texture features obtained with Fisher coefficient in terms of classifications. The experiments were done using texture images from the Brodatz album.

Kachouri et al. (2008) presented a **hierarchical feature extraction model** and the relevance evaluation of several features for an heterogeneous optical image database. The authors used multiple primitive features (color, shape, texture) to describe an image. The

evaluation consisted in making classifications using an adjusted version of SVM that supports multiple classes. The authors compared different features employed separately, different combinations of the same kind of features, aggregated features and the proposed hierarchical feature model.

A study to exploit morphological features from multispectral Ikonos imagery was presented in (Huang et al., 2009). The extracted features were compared using the object-based analysis and the Gray-Level Co-occurrence Matrix (GLCM).

Li and Shawe-Taylor (2005) experimented the texture classification using **multiresolution features extracted** from dyadic wavelet, wavelet frame, Gabor wavelet, and steerable pyramid. The classifications were made using SVM as classifier. The experimental results show that the steerable pyramid and Gabor wavelet classify the texture images with the highest accuracy. However, experimental results on fused features demonstrated the combination of two feature sets always outperformed each method individually.

Karkanis et al. (2001) presented an **evaluation of textural feature extraction using medical images**. The authors compared four texture extraction methods (GLCM, run length encoding, fractal dimension and discrete wavelet transform descriptor) by means of classifications.

Selection of **feature extraction methods with SAR imagery** was published in (Solberg et al., 1997). The authors evaluate the performance of texture features derived from 1) the GLCM, 2) local image statistics, 3) fractal features, and 4) lognormal field models. In order to test the performance of the texture features the authors made urban area classifications using a K-Nearest Neighbor algorithm as classifier.

In (Clausi and Jernigan, 1998) and (Clausi and Yue, 2004) the performance of GLCM, Markov Random Fields (MRF) and Gabor features in classifying sea-ice imagery was compared. The authors found that the GLCM produced the overall best results in terms of classification accuracy, followed by the Gabor features. However, the GLCM features were found to be more sensitive to texture boundaries as compared to MRF.

Kandaswamy et al. (2005) proposed the use of approximate textural features for fast image texture analysis. Rather than using the entire image, approximate features are derived from a carefully selected subset of the original image, based on the notion of patch reoccurrence. Later, the proposed approximated features can be extracted for two texture analysis methods 1) the GLCM, and 2) Gabor wavelets. The results are expressed in terms of classifications.

As conclusion the component by component evaluation is predominantly based on classification results, which reflect the accuracy of the feature extraction methods.

3.3.1.2 General evaluation of the performance

The mainly used measures in image retrieval evaluation are the **Precision and Recall (PR)** measures, which are widely used in the evaluation of text document retrieval (Harman, 1993). However, these measures are not appropriate for image retrieval (Dyson and Box, 1997) and they are only of limited use for image collections (Müller et al., 2001). In fact, RP techniques have two drawbacks: 1) Selection of a *relevant* set in an image database is more complicated than in a text database since the definition of the image meaning can have a long number of interpretations. 2) In an image database the application of a selected query returns a ranking list of results instead of an undifferentiated set of relevant images (Smeulders et al., 2000). However, in spite of these shortcomings, **PR are useful** in special

circumstances or with special considerations as **the addition of strong semantics to the images database provided by labeling or by textual description** (Smith and Li, 1998).

Müller et al. (2001) presented an overview and proposal in performance evaluation in CBIR. Here, the authors identified the basic problems in CBIR performance evaluation as for example defining common image collections, obtaining the relevance judgments and making comparisons with textual information retrieval. In addition, the authors gave a summary of the most commonly used evaluation methods such as PR graphs, the rank of first retrieved relevant image (Rank1), the average normalized rank (Rank). The authors recommended the use of the normalized average rank and highlighted the necessity of standard performance measures, a standard image database, and the integration of the user in the evaluation process.

The work of Aksoy (2001) posed the retrieval problem in a probabilistic framework where the aim is to minimize the error in a setting of two classes: the relevance and the irrelevance classes of the query. The author proposed new methods in different components of the CBIR like feature extraction, image matching, feature combination and relevance feedback and presented a validation of those methods in terms of comparisons between state of the art methods and the proposed ones. The performance evaluation was done using extensive experiments on three different manually ground-truthed databases, including aerial satellite, texture and stock images. The used databases are: ISL Database (MIT, 2001) which contains EO images from Texas, VisTeX composed by texture images and COREL Photo Stock Library. The metrics used in the evaluation were PR curves, number of retrievals that have a specific target image among the set of retrieved images, and classifications.

Deselaers et al. (2004) analyzed the different evaluation metrics proposed by (Müller et al., 2001) and complemented this work proposing a classification error rate (ER) as performance evaluation metric assuming a connection between the CBIR and image classification.

The work of Huijsmans and Sebe (2005) presented the shortcomings of PR graphs, mentioning the fact that they provide the user an incomplete information about how well the IR system will perform for various relevant class sizes and various irrelevant class sizes. Here, the authors introduced the term "generality" to describe the influence of the relevant items in the database and proposed a 3D graphic, which shows the PR values as a function of the generality. Also, the importance of normalizing the performance measures with respect to the class size was highlighted. The authors proposed a well-normalized description of the ranking performance compared to the performance of an ideal retrieval system defined by ground-truth for a large number of predefined queries.

3.3.2 User-driven evaluation approach

One of the most important criticisms has concerned the validity of deriving quantitative performance measures from subjective relevance judgments. This has led to the rise of more user-centered approaches, emphasizing the use of multiple evaluation methods which take into account the interactive nature of modern systems, the user's work and task context, and the variability of information over time (Eakins and Graham, 1999). (Tague-Sutcliffe, 1996) specifies user satisfaction as the main concern of an IR system evaluation. Moreover, the relevance feedback and involvement of the user in the loop of searching the most adequate image according to its content makes clear the necessity of user-driven evaluation, which takes into account the user requirements and expectations.

Joemon et al. (1998) presented a methodology for evaluating two versions of the Epic system. The first version enables spatial queries and the second one allows textual queries only. The authors trained the user to work with two versions of Epic and later, the user was asked to responde a questionnaire by scoring both systems according to his level of satisfaction. The authors remark that the user often had a "mental image" of a potentially satisfying picture in mind.

PicHunter system (Cox et al., 2000) introduced two experimental paradigms to quantitatively evaluate the performance of the system and presented psychophysical experiments that allowed the involvement of the user. PicHunter proposes a user model which is supported by psychophysical experiments. These experiments helped to choose the distance measure as well as the form of the probability function. The experiments made using PicHunter consisted in testing the two versions of the system. One version implemented semantics definition whilst the older version did not include it. Thus, they involved the user in searching a specific target in an image database using the two versions of the system, with and without semantics. The results demonstrated that the best performance could be reached with semantics in the queries. The authors believes that data under the target search paradigm offers an objective measure of the system performance.

The evaluation of KIM system (Datcu et al., 2003) was presented in (Daschiel and Datcu, 2005). Here the authors proposed a methodology which aims at determining the objective technical quality of the system and includes the subjective factor as well. The evaluation was performed using: 1)the classification error, 2)the information transfer for the quality of the learning process, and 3)the complexity for the human-computer dialogue. Those tasks were accomplished with the help of the user-trace, which monitored and saved all the user movements. The evaluation was done in terms of validation of individual system modules. The aim was to merge objective measurements with subjective concepts as for example user satisfaction and analyzed how well they fit together. The experiments were performed using EO optical images.

Shirahatti and Barnard (2005) developed a reference optical image database by the user assessment. The authors presented a strategy for evaluating image retrieval algorithms provided by both a query based on text and a query based on content. The authors developed a software tool, which presents one query image and four result images. The users were asked to score the retrieved image on a scale from 1 to 5. Thus, the authors collected a large data set of human evaluations of retrieved results as ground-truth data. The authors stressed that semantics play an important role in what users consider a match or a correct retrieval result.

The work of Tsai et al. (2006b) presented a qualitative evaluation methodology for systems which automatically assign keyword to images. It is composed of two evaluation methods: the first is based on a human assessment of annotation accuracy and the second on the construction of comparable ground-truth. As conclusion the authors show that the human-centered evaluation methodology provides deeper understanding about not only the performance of an image annotation system but also the consistency of human judgements about the relevance of high-level concept terms to images. The evaluation was done using a CLAssifying Images for REtrieval (CLAIRE) (Tsai et al., 2006a) system.

3.4 Conclusions

In this chapter, the following items were presented:

- The evaluation of CBIR and IIM systems is still an open question. In most of the CBIR literature, researchers presented examples of queries to visually evaluate the performance of their systems. No standard evaluation has been included.
 - The best way to evaluate a complex CBIR system consists in dividing the system in its components and testing the individual performance of the methods, using adequate metrics according to what has to be evaluated.
 - The overall evaluation of the system is divided in two parts: data-driven evaluation, considering mathematical and statistical methods in order to assess the performance of the feature extraction module, and user-driven evaluation taking into account the user criterion and involving the user in the evaluation process.
 - Evaluation of feature extraction methods accordingly to the type of image ensure a proper performance of IIM systems, since the feature extraction methods are the main component of these systems.
 - For a complete evaluation of CBIR the user has to be involved to judge the results provided by the CBIR system. In fact the user with his expectations is a vital part of the system and his opinions could be incorporated in the interactive loop. Thus, in this thesis we involve the subjective factor given by the user criteria in the evaluation methodology.
 - The subjective evaluation results obtained with the IIM systems which included semantics show that semantics play an important role in understanding the image and user satisfaction.
 - The most common metric used in the evaluation of performance of the CBIR systems is the Precision and Recall curves. However, in order to apply this metric it is needed to have a ground-truth (reference) database for evaluation.
 - The use of SAR images in EO mining is a new topic and it presents some difficulties due to the nature of the SAR sensor as for example the speckle noise-like, which makes the visual interpretation more difficult, and the new hidden classes that appear with the high-resolution. Therefore, the evaluation of IIM systems working with this kind of images is a matter of research nowadays.
-

Chapter 4

SAR image analysis using Bayesian inference and Gibbs Random Field models

Nowadays, the increasing resolution of the Synthetic Aperture Radar (SAR) satellites provides us with very high resolution SAR images, which contain an increased amount of details and information. As a consequence, a robust detection and recognition of small scale man-made structures representing buildings, roads, harbors, bridges, etc has become a new challenging task.

In parallel, the automatic interpretation of SAR images is an ongoing research, where the goal is to provide methods, which can deal with the large amount of information contained in the image in a fast automatic way. However, an important property of SAR data is the presence of speckle which, in most cases, complicates an automatic interpretation of SAR images. In general, a robust analysis of SAR images is a two step approach using first despeckling filters and then information extraction methods.

The Bayesian approach relying on models and their parameters to fit the data could be an automated method being able to extract and interpret the genuine information contained in high resolution SAR images. The image information content is extracted using model-based methods based on Gibbs Random Fields (GRFs) combined with a Bayesian inference approach. The approach enhances the local adaptation by using a prior model, which learns the image structure; it enables despeckling with minimum loss of resolution and simultaneously estimates the local description of the structures. From these we may obtain detection, classification, and recognition of the image content.

This chapter starts with presenting the theoretical concepts of the Bayesian approach and its two inference levels, the parameter estimation and the model selection, in section 4.1. Later, the Markov and Gibbs Random Field models, and examples of their realization in image modelling are explained in section 4.2. After that, the application of the Bayesian inference to the SAR image analysis is investigated considering the despeckling process and information extraction methods in section 4.3. Some conclusions are presented in section 4.4.

4.1 Bayesian inference

Bayesian inference is an approach to statistics in which all forms of uncertainty are expressed in terms of probability. In order to address the Bayesian inference, the basic definition of probability is presented.

There are two approaches to define a probability: the frequentist and the Bayesian. In the frequentist approach the probability $P(A)$ of an event A is given by

$$P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n}, \quad (4.1)$$

where n_A is the number of occurrences of A and n is the number of trials. This requires repeatable experiments and it is seemingly objective. However, within the Bayesian approach $P(A)$ represents the measurement of confidence in proposition A . It uses a prior knowledge and is seemingly subjective.

In the following, we start with presenting the conditional probability and the famous Bayes theorem, which are the basis for understanding the two levels of inference. Later, the two levels of Bayesian inference, the parameter estimation and the model selection are presented.

4.1.1 Conditional probability and Bayes' theorem

The conditional probability of two events A and B denoted by $P(A|B)$, is by definition the ratio

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad (4.2)$$

where $P(B)$ is assumed not to be zero (Papoulis, 1991). The sum and product rules from the basic algebra of probability theory guide us to formulate the **Bayes' theorem**, which is derived from the product rule and expressed as follows:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}. \quad (4.3)$$

The importance of this theorem to data analysis becomes more intuitive if we replace A and B by *hypothesis* and *data*, thus

$$P(\text{hypothesis}|\text{data}) = \frac{P(\text{data}|\text{hypothesis}) \cdot P(\text{hypothesis})}{P(\text{data})}, \quad (4.4)$$

which describes the probability of the hypothesis A given some data B . Since A does not depend on $P(B)$ then the conditional probability $P(A|B)$ is proportional to the probability of the observation given the hypothesis $P(B|A)$ multiplied by a term containing the prior information $P(A)$ (Silva, 1996).

The terms in Bayes equation have formal names: $P(\text{data}|\text{hypothesis})$ is known as **likelihood** function, it describes the likelihood of the *data* given a *hypothesis*. The quantity $P(\text{hypothesis})$ is called the **prior** probability, which represents our state of knowledge about the truth of the hypothesis before we have analyzed the current data. The term $P(\text{data})$ is known as **evidence** (Silva, 1996). The combination of the likelihood function with the prior probability and the evidence, yields to find the **posterior** probability $P(\text{hypothesis}|\text{data})$. In other words, the Bayes theorem states as

$$\text{posterior} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}. \quad (4.5)$$

In many data analysis problems, such as those involving **parameter estimation**, the evidence term is omitted because it acts as normalization constant since it does not depend explicitly on *hypothesis*. However, this term plays a crucial role for **model selection**. The parameter estimation and model selection correspond to first and second levels of Bayesian inference. The Bayesian inference allows us to select the most probable model, thus to reject highly biased, unprovable or subjective prior assumptions that have no relevance, as it is called, evidence to the data.

4.1.2 Parameter estimation

The parameter estimation is a common data analysis problem, which is a part of the statistical decision theory. We start with the definition of x and \mathbf{y} . Let x be the parameter to be estimated e.g. "true" pixel value in a noisy image, \mathbf{y} is the vector of (measurement) data used for the estimation (observation). Then, the task of parameter estimation is to infer a parameter value \hat{x} (estimated of x) from an observation \mathbf{y} .

There exists a variety of different estimators. However, here, we restrict ourselves to the discussion of the Maximum Likelihood (MLE) and Maximum *a posteriori* (MAP) estimators for finding the value of \hat{x} .

- The MLE approach maximizes the likelihood ignoring the prior knowledge. This is given by

$$\hat{\mathbf{x}}_{MLE} = \arg \max_x p(\mathbf{y}|\mathbf{x}). \quad (4.6)$$

- While MAP approach finds the most probable value of x that maximizes the posterior probability $p(\mathbf{x}|\mathbf{y})$, so MAP searches for

$$\hat{\mathbf{x}}_{MAP} = \arg \max_x p(\mathbf{x}|\mathbf{y}) = \arg \max_x p(\mathbf{y}|\mathbf{x}) \cdot p(\mathbf{x}) \quad (4.7)$$

In the case of SAR images, the prior pdf introduces some knowledge about the scene allowing us to model the data. The problem of MAP estimation is also known as the first level of Bayesian inference and it has been criticized because of the choice of the correct prior distribution, which is compensated with the second level of inference (the model selection) (Walessa, 2001).

4.1.3 Model Selection

The model selection process or second level of Bayesian inference is one of the major advantages of the Bayesian estimation theory. The model selection searches for the most suitable model from the different alternative models. In here, the evidence term plays a fundamental role. The evidence is obtained by **marginalization** (Papoulis, 1991), for example by integration over x , and results in

$$p(\mathbf{y}|M) = \int_{-\infty}^{+\infty} p(\mathbf{y}|\mathbf{x}, M)p(\mathbf{x}|M)dx, \quad (4.8)$$

where M represents a model. Considering a set of different models ($M1, M2, \dots, Mn$), this equation enables us to calculate the likelihood of each model, and, thereby selecting the most probable model relies on the measure of the random variable \mathbf{y} .

4.2 Markov and Gibbs Random Fields

The use of a Bayesian framework in image restoration and reconstruction requires the specification of a prior distribution for the image (Besag, 1989). To construct prior statistical models of images two approaches were followed. While in the first one images were modeled as **Markov Random Fields** (e.g. Chellappa and Jain (1993); Cross and Jain (1983)), the second one was based on Gibbs models (e.g. Besag (1974); Hassner and Sklansky (1980); Descombes and Zhizhina (2008)). Markov and **Gibbs Random Fields** (MRFs and GRFs) models play an important role, since they are able to statistically describe correlations, or even more generally, any kind of statistical dependence between neighbouring pixels. A MRF is a lattice of random variables (for example the pixel values in images), where the conditional distribution of each variable depends only on the neighbouring variables and the joint probability distribution is a Gibbs distribution.

This section starts with the description of an image, the neighbourhood systems and cliques. Later we present a basic definition of the Gibbs distribution and GRFs. Finally, two models of the GRFs family: Gauss Markov Random Field (GMRF) and the Auto-binomial model (ABM) are presented.

4.2.1 Image description, neighbourhood systems and cliques

The **modeled image** is considered as a realization of a random field X that is composed of random variables x_s at different sites s representing the individual pixels. The set $\{x_s\}$ states for of all pixels in the image. If the pdf or the pixel x_s of this random field is determined only by values of pixels in the local neighbourhood, this field is called Markovian (Schröder et al., 1998).

Then, the **neighbourhood system**, represented by Ω , around the single pixel x_s at site s is defined by (Li, 1995)

$$\Omega_s^m = \{t\} \text{ with } \begin{cases} s \notin \Omega_s \\ t \in \Omega_s \end{cases} \text{ then } s \in \Omega_t \quad (4.9)$$

where the term $\{t\}$ is a set of all pixels in the vicinity of site s , excluding s . The neighbourhood systems have different sizes, which are determined by the **model order**. The size defines the maximum limits of the structural complexity of the model. The m^{th} **order neighbourhood system** is represented by the symbol Ω^m .

The **cliques**, denoted by c , associated with a neighbourhood Ω specify the possible topology of the random structures in different realization of the stochastic process (Derin and Elliott, 1987). A **clique** consists of a single pixel or a set of pixels that are neighbors to each others. The collection of all cliques of Ω is expressed as C .

Figure 4.1 shows examples of neighbourhood systems (Ω_s^m) around the central pixel x_s with different model orders (m). Here, first, second, third and sixth order neighbourhoods are depicted. The neighbourhood order of a pixel x_{ij} around a center pixel $x_s = x_{01}$ is denoted by the index i , i.e., x_{62} corresponds to a sixth order neighbourhood system. The index j is a simple counting variable. It is worth noting that the indexes i and j are not expressed in cartesian coordinates. In the right part of Figure 4.1, the cliques associated with the respective m order neighbourhood system and the model parameter vector represented by θ are depicted. The parameter vector θ is composed of b_{ij} elements, where the total number of elements is determined by the model order, for example, θ has two

elements when the model order is one, ten when the model order is four. The b_{ij} components weight the interaction between the central pixel x_s and its neighbouring pixels (x_{ij} and x'_{ij}), i.e., in the third neighbourhood system, b_{11} weights the interaction between x_s and the horizontal pixels x_{11} and x'_{11} , while b_{21} weights the interaction between x_s and the diagonal pixels x_{21} and x'_{21} .

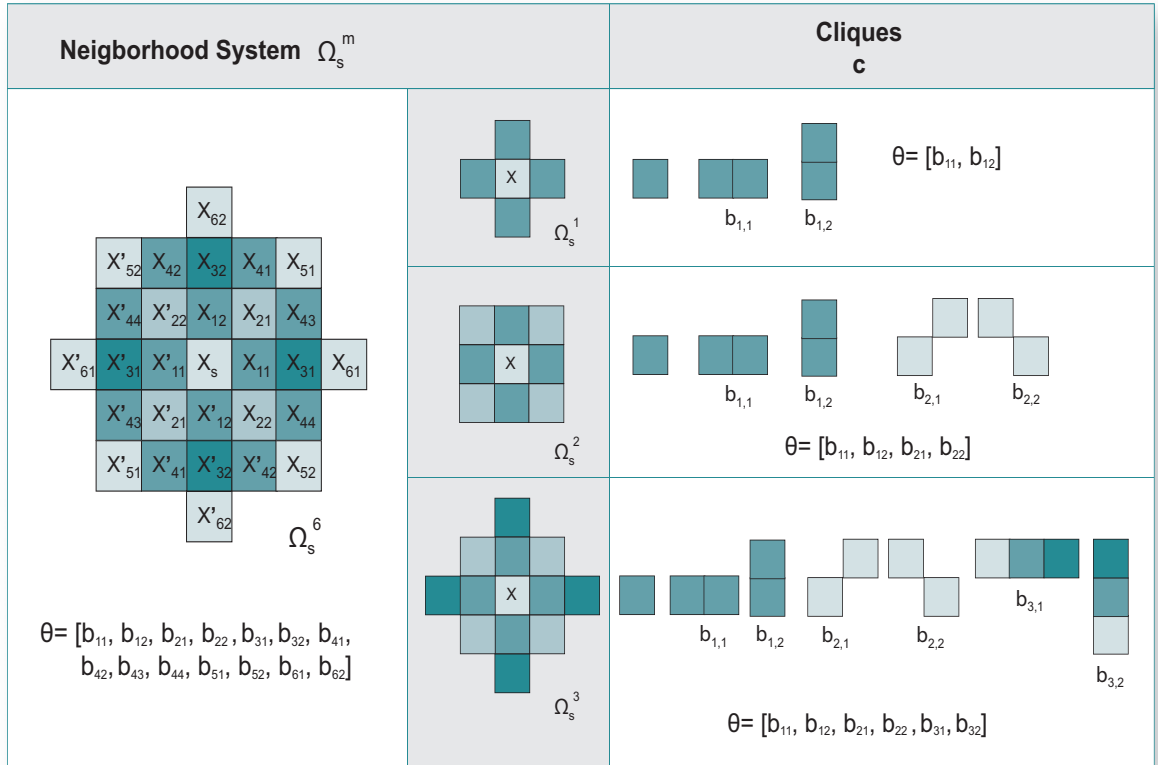


Figure 4.1: Neighbourhood systems Ω_s^m with order m , Cliques c and model parameter model θ . (Left) The neighbourhood system Ω_s^m is assumed to interact with the central pixel x_s . The interaction of x_s with the neighbourhood pair x_{ij} and x'_{ij} is weighted by the component b_{ij} in the parameter vector θ . The neighbourhood order of a given pixel x_{ij} is denoted by i . (Center) Examples of first, second and third neighbourhood orders. (Right) Cliques associated with a first, second and third order neighbourhood system. The model parameter vector θ and its elements.

4.2.2 Markov-Gibbs Random Fields equivalence

Let Ω be a neighbourhood system defined over the finite lattice L (rectangular grid). A random field $X = X_s$ defined on L has Gibbs Distribution (GD) or equivalently is a Gibbs Random Field (GRF) with respect to Ω if and only if its joint distribution is of the form of an exponential distribution given by

$$p(X = x) = \frac{1}{Z} \exp^{-U(x, \theta)} \quad (4.10)$$

where x is a realization of X and the term $U(x, \theta)$ is the **energy function** given by

$$U(x, \theta) = \sum_{c \in C} V_c(x, \theta) \quad (4.11)$$

where θ **characterizes the model parameter vector** and it is expressed as $\theta = [b_{11}, b_{12}, b_{21}, b_{22}, \dots]$. The term $V_c(x, \theta)$ represents the potential associated with the clique c , so the **energy function** (U) is defined by the sum of the potential function attached to each clique. The quantity Z stands for the **partition function** and is given by

$$Z = \sum_X \exp^{-U(x, \theta)}. \quad (4.12)$$

The partition function Z is simply a normalizing constant. The only condition on the otherwise totally arbitrary clique potential $V_c(x)$ is that it depends only on the pixel values in clique c (Derin and Elliott, 1987).

A stochastic modeling of the data X using a set of parameters θ is equivalent to assigning to each state of X a joint probability $p(X|\theta)$.

4.2.2.1 Hammersley-Clifford theorem

The Hammersley-Clifford theorem (Hammersley and Clifford, 1971) asserts that, for a limited neighbourhood and discrete state of X_i , a Markov Random Field X characterized by a neighbourhood system (Ω_s^m) is equivalent to a Gibbs Random Field and a certain potential associated to each clique within this neighbourhood. Hence, each Markov random field can be represented in the following form

$$p(X = x) = \frac{1}{Z} \exp \left(\sum_{c \in C} V_c(x, \theta) \right). \quad (4.13)$$

In other words, the Hammerley-Clifford Theorem says that, every Markov Random Field has an equivalent Gibbs Random Field.

4.2.3 Examples of Gibbs Random Field models

In the literature there are several Gibbs Random Field models (Li, 1995) as for instance the auto-exponential and the auto-binomial models (Besag, 1974), the multi-level logistic model (Derin and Elliott, 1987) and the Gauss Markov Random Field (GMRF) model (Chellappa and Chatterjee, 1985), which were used for modeling image properties such as regions and textures. The GMRF and Auto-binomial (ABM) models are presented in detail in this section since they are used in the feature extraction methods, that will be evaluated in this thesis. GMRF and ABM belong to the GRFs family, since they follow the Gibbs distribution (4.10), where the energy function (U) changes according to the model as described in the following.

4.2.3.1 Gauss Markov Random Field model

In this model, the energy function (4.11) is defined as

$$U(x_s | \Omega_s, \theta) = \frac{(x_s - \eta_s)^2}{2\sigma^2} \quad (4.14)$$

where

$$\eta_s = \sum_{ij} b_{ij} \cdot (x_{ij} + x'_{ij}) \quad (4.15)$$

is the interaction between the pair of neighbouring pixels, x_{ij} and x'_{ij} , and the parameter vector is $\boldsymbol{\theta} = [b_{11}, b_{12}, \dots]$. This model corresponds to an autoregressive process of the form

$$x_s = \sum_{ij} b_{ij}(x_{ij} + x'_{ij}) + n_s \quad (4.16)$$

where some distortion has been assumed in the prediction of η_s by zero mean white Gaussian noise n_s with variance σ^2 . The corresponding conditional pdf for the Gauss Markov Random Field model can be expressed as (Walesa, 2001)

$$p(x_s|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x_s - \eta_s)^2}{2\sigma^2} \right]. \quad (4.17)$$

4.2.3.2 Auto-binomial model

In the ABM model, the energy function (4.11) is defined as

$$U(x_s|\Omega_s, \boldsymbol{\theta}) = -\ln \binom{G}{x_s} - x_s \cdot \eta_s \quad (4.18)$$

where G is the maximum gray value, $\binom{G}{x_s}$ represents the binomial coefficients, and η is defined as follows

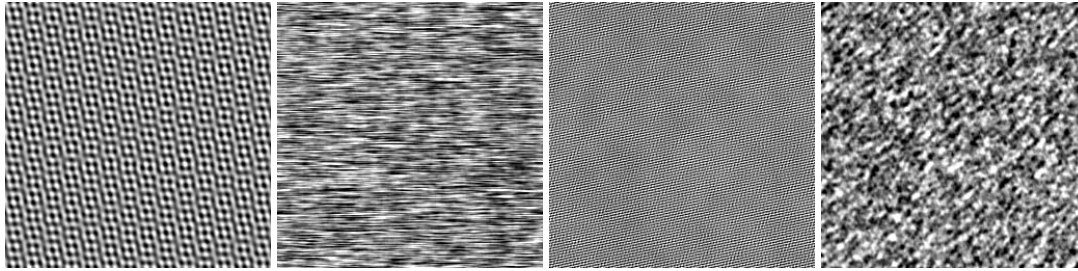
$$\eta_s = a + \sum_{ij} b_{ij} \frac{(x_{ij} + x'_{ij})}{G} \quad (4.19)$$

which reflects the joint influence between the neighbors x_{ij} and x'_{ij} and the parameter vector $\boldsymbol{\theta} = [a, b_{11}, b_{12}, b_{21}, \dots]$. Note that a parameter in the vector represents a self interaction and this parameter tends to be given by approximately the negative sum of all b_{ij} parameters. From (4.18) we can note that, the energy function consists of two terms, where the first one just depends on x_s and the second is a product of x_s and η . Then, $x_s \approx \frac{G}{2}$ for $\eta \approx 0$ (Schröder, 2000).

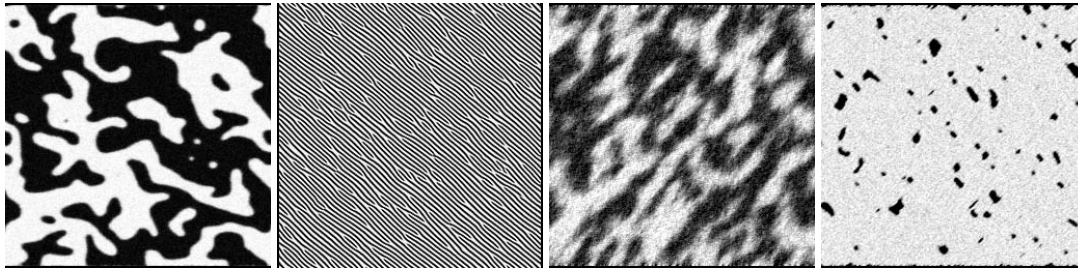
4.2.3.3 Examples of realizations of GMRF and ABM models

Figure 4.2 shows examples of realizations of both prior models, GMRF and ABM, with different model orders and model parameter vectors $\boldsymbol{\theta}$. The values of $\boldsymbol{\theta}$ are presented in Table 4.1. These images show characteristic structures that have a signature in the parameter vector $\boldsymbol{\theta}$ since the b_{ij} elements weighted the interaction between the central analyzed pixel and its neighbouring pixels. A positive value of b_{ij} means that the central pixel and its neighbouring pixels are attracting each other, while a negative value of b_{ij} denotes a repulsive interaction between the pixels. The realizations of GMRF present flat textures, while the realizations of ABM show more variety of textures.

Both prior models GMRF and ABM are implemented in robust algorithms for SAR images despeckling and information extraction. Since they are based on the Bayesian inference framework, they allow us to extract texture parameters as main primitive features from SAR images and perform despeckling simultaneously.



(a) Textures generated using Gauss Markov Random Field model.



(b) Textures generated using Auto-binomial model.

Figure 4.2: Example of realization of GMRF and ABM models. From left to right, first, second, third and fourth order used in the generation of the texture images. The realizations of GMRF present flat textures, while the realizations of ABM show more variety of textures. The ABM model is able to generate blob-like textures, which are often present in SAR images.

Table 4.1: Values of θ parameters used in the generation of the examples of textures depicted in Figure 4.2.

GMRF											
Order		b_{11}	b_{12}	b_{21}	b_{22}	b_{31}	b_{32}	b_{41}	b_{42}	b_{43}	b_{44}
1st		0.33	0.27								
2nd		0.3	0.5	-0.15	-0.15						
3rd		0.408	0.125	-0.456	-0.36	0.409	0.374				
4th		0.593	0.415	-0.235	-0.216	-0.183	-0.003	0.080	-0.007	-0.018	0.074
ABM											
Order	a	b_{11}	b_{12}	b_{21}	b_{22}	b_{31}	b_{32}	b_{41}	b_{42}	b_{43}	b_{44}
1st	-2.73	1.73	1.00								
2nd	-2.01	1.73	1.91	-2.37	1.00						
3rd	-2.083	1.02	0.75	-0.52	0.53	0.056	0.247				
4th	-2.32	2.206	1.23	-0.64	0.166	-0.75	-0.33	-0.013	0.009	0.138	0.32

4.3 Bayesian SAR image analysis

Synthetic Aperture Radar (SAR) images are used in a large number of applications over different fields, such as remote sensing for mapping, surface surveillance, search and rescue, mine detection, automatic target recognition, etc. However, they are affected by a multiplicative speckle process, considered as noise-like, which reduces the radiometric quality and makes difficult the extraction of important information within the image.

The speckle presence in SAR images reduce the detectability of objects in the image and also the capability to separate and classify distributed targets (Oliver and Quegan, 1998). Therefore, the analysis of SAR images involves both despeckling and information extraction methods. In the following, the most known speckle filters and how they were improved are explained in section 4.3.1. Later, the feature extraction methods for SAR images, with a special emphasis on methods which are based on stochastic models and their parameters, are presented in section 4.3.2.

4.3.1 SAR despeckling filters

Along the years many methods have been proposed to remove the speckle of SAR imagery. The main goal of despeckling filters is to remove the speckle with preservation of small details within the image such as: texture, edges, geometry (Fjortoft et al., 2000), (Xiao et al., 2003).

Speckle can be removed by filtering using the multi-looking techniques. Multi-look processing is applied during image formation, and this procedure averages several statistically independent looks of the same scene to reduce speckle (Porcello et al., 1976). It is considered as a pre-image formation filter. A major disadvantage of this technique is that the resulting images suffer from a reduction of the ground resolution that is proportional to the number of looks L (Martin and Turner, 1993). Improvement of the radiometric resolution is obtained at the cost of the geometric resolution (De Vries, 1998).

Later many despeckling techniques have been emerged, which perform the despeckling after the image formation. This section presents a review of post-image formation despeckling techniques, which are grouped into non-Bayesian (or non-Parametric) and Bayesian (or Parametric) filters.

4.3.1.1 Non-Bayesian despeckling filters

The non-Bayesian or non-parametric filters do not use parameters about the scene to be modeled, they only consider the speckle model and are generally based on local statistics, which are computed in sliding windows. This kind of filters uses the Minimum Mean Squared Error (MMSE) as estimator. The adaptive filters, which are based on local statistics belongs to this category. The most used adaptive filters are described in the following:

1. Adaptive filters

Lee (1980) suggested a method to enhance a digital image and noise filtering by using the local statistics considering additive and multiplicative noise, which follows the Gaussian distribution. This filter uses the sample mean and variance of pixels in a sliding window. In here, the *a priori* mean and variance of each pixel is derived from its local mean and variance, then it uses the MMSE estimator to obtain the noise filtered image.

An improvement of Lee filter was proposed by (Kuan et al., 1985). Here, the authors presented a generalization of the local statistics of Lee filter using a different algorithm to compute the local variance. The author considered only a multiplicative speckle model and designed a linear filter based on the MMSE criterion that has an optimal performance when both the scene and the detected intensity follow a Gaussian distribution. However, a common characteristics of local statistic filters is that the noise in the edge areas is not

smoothed. A solution for this problem was proposed by (Lee, 1981). The author introduced the Refined MMSE filter known as **Refined Lee filter**, which gives better results than the original filter presented in (Lee, 1980), at the presences of edges. In order to deal with the common problem of noisy edge boundaries, Lee revised the algorithm using edge directed windows. Eight non-square windows for eight edge orientations are created within 7×7 windows. The local mean and local variance are computed using only those pixels in the edge directed windows and then, the local statistics filter is applied. Later (Frost et al., 1982) introduced the Frost filter, which is based on the local statistics. It models the speckle as multiplicative noise. It differs from the Lee filter in the fact that the scene reflectivity is estimated by convolving the observed image with the impulse response of the SAR system and it averages less over the edged areas to preserve edges.

As conclusion all the adaptive filters operate in spatial domain and they use the MMSE as estimator. Most of them assume that the speckle follows a Gaussian distribution, which is not the case of SAR images (Oliver and Quegan, 1998) and most of them model only the speckle noise without considering a model for the reflectivity of the scene (data). In the following, we introduce the Bayesian filters, which take the SAR properties into consideration. These filters enable to use statistical schemes to model the data, thereby, they are considered as advanced speckle filtering techniques (Lopes et al., 1993).

4.3.1.2 Bayesian despeckling filters

The Bayesian or parametric approach is based on models and estimation of model's parameters, which determine how well the estimated data corresponds to the image data. It enhances local adaptation by using a *prior* model, which learns the image structure, enabling thus despeckling without resolution loss, whilst estimating local descriptions of the structures. Therefore, this approach is considered as Bayesian approach since it relies on models and parameters to fit the data. An example of this approach is the Gamma-Gamma MAP filter (Lopes et al., 1993), which is based on the **Maximum a posteriori (MAP) estimation**. The MAP approach has several advantages compared to the previous filters. In fact, it can simultaneously take into account realistic first and second order statistical models for both speckle and underlying scene reflectivity, and combine them through Bayes' rule. The Bayesian filters can operate in an original spatial (image) domain or in a frequency domain, using wavelets. They are presented in the following.

1. Bayesian filters operating in the spatial domain

Kuan et al. (1987) were the first to propose the Maximum *a posteriori* (MAP) approach for filtering SAR images by the assumption that the speckle follows a multiplicative model. This filter is called Gamma-MAP. In here, a prior knowledge of the probability density function of the scene is required and it was assumed to be Gaussian distributed while the speckle was assumed to have a Gamma distributed intensity. Later, in the work of (Lopes et al., 1993, 1990), they modified the Gamma-MAP filter by assuming a Gamma distributed scene and setting up two thresholds, resulting in the Gamma-Gamma MAP filter, which nowadays is one of the most prominent and broadly used despeckling filter. Gamma-Gamma MAP is based on a Bayesian analysis of the statistics considering that both the signal and the speckle noise follow a Gamma distribution. This filter achieved at the same time a good preservation of the textural properties and a good enhancement of the structural features present in the image, as well as a very high degree of speckle

smoothing. Even though the modeling of Gamma-Gamma MAP is more elaborated than the approaches previously presented, the visual impression and quantitative despeckling results are similar to Kuan and Lee filters (Medeiros et al., 2003). In fact, it uses a simple data model (Gamma distribution) described by only two parameters (mean and variance). This means that Gamma prior seems not to be sufficient to model a SAR scene.

Geman and Geman (1984) introduced the use of Gibbs Random Fields (GRFs) models as prior model of the scene for despeckling SAR images. In here, these filters use stochastic models in order to describe the spatial structures in the images. They encapsulate more information than only mean and variance by using more complex parametric models. For example, the use of Gauss Markov Random Field (GMRF), which belongs to GRFs family was proposed in (Walessa and Datcu, 2000; Walessa, 2001). This filter is based on Bayesian inference, using the Gamma distribution as speckle noise model (likelihood) and the GMRF as prior to represent the knowledge of the scene. GMRF uses a neighbourhood system around a central pixel in order to give the value of the central analyzed pixel considering its relations with its neighbors. This filter estimates the MAP of the noise-free image through multiplying the likelihood by the prior. Recently, in (Hebar et al., 2009), the authors proposed the use of an Auto-binomial model, which also belongs to GRFs family, as prior model and the Gamma distribution as speckle model. This technique is based on Bayesian inference and MAP estimation. The main advantage of both filters is the fact that they are able to extract information from the SAR image. Both despeckling techniques operate in spatial domain and they are fully described in chapter 5.1.

2. Bayesian filters operating in the frequency domain

It is worth noting that in the last 10 years an important improvement in the field of despeckling images was given by multi-resolution processing, using wavelets. Wavelets have been recognized as a powerful tool for the analysis of non-stationary signals and images (Mallat, 1989). Wavelet despeckling approaches are based on modifying the (log-transformed) speckle noisy wavelet coefficients according to some rules (shrinkage) and reconstructing the filtered image from them (Achim et al., 2003).

Examples of wavelet based despeckling algorithms were proposed in (Foucher et al., 2001), (Xie et al., 2002), (Dai et al., 2004), (Argenti and Alparone, 2002), (Argenti et al., 2006), (Gleich and Datcu, 2007), (Gleich and Datcu, 2009), etc. The work of (Xie et al., 2002) proposed a despeckling algorithm that fuses Bayesian wavelet denoising with a regularizing prior.

The wavelet-based denoising using un-decimated wavelet transform and MAP estimation were performed in (Dai et al., 2004) (Argenti and Alparone, 2002); (Argenti et al., 2006) and (Foucher et al., 2001). Specifically, the work of (Dai et al., 2004) discussed a Bayesian shrinkage which relies on edge information. (Argenti et al., 2006) proposed despeckling in the undecimated wavelet decomposition and MAP using a space-varying generalized Gaussian distribution for the wavelet coefficients. The solution of the MAP equation yields to the MAP estimate of the wavelet coefficients of the noise-free image.

A MAP estimator with an alpha-stable prior within the wavelet framework is proposed in (Achim et al., 2003). Later, an adaptive MAP estimator with a heavy-tailed Rayleigh signal model has been suggested in (Achim et al., 2006). Here, it uses log-transform to convert multiplicative noise to additive noise. The work of (Gleich and Datcu, 2007) proposed a despeckling filter based on the combination of wavelet transform and GMRF. Here, noise in the wavelet domain is modeled as an additive signal-dependent noise

with Gaussian distribution and the distribution of the noise-free scene is Generalized GMRF (GGMRF). In (Gleich and Datcu, 2009) the denoising of SAR images within the dyadic wavelet domain using a particle filter was introduced. The objective was to model the SAR image statistics and perform denoising and information extraction within the decimated wavelet domain. The dyadic transform is a shift variant; therefore, many authors use stationary or non-decimated wavelet transform, which is a shift invariant.

The second-generation wavelets Chirplet (Cui and Wong, 2006), Contourlet (da Cunha et al., 2006), Bandelet (Pennec and Mallat, 2005) have appeared over the past few years. Despeckling using Contourlet transform (da Cunha et al., 2006) and Bandelet (Pennec and Mallat, 2005) transforms show superior despeckling results for SAR images compared with the wavelet-based methods (Gleich et al., 2010).

4.3.2 SAR information extraction

Information extraction is an important issue when designing automatic image interpretation systems, as it provides intrinsic features in few parameters. After that, a classifier can use these parameters as features for categorizing and indexing the image content. An image is characterized by its primitive features such as color, shape (Jain and Vailaya, 1996), texture (Tuceryan and Jain, 1998). Thus an image will be represented as a multidimensional feature vector acting as a signature.

Texture is considered as an important feature in an image since it provides a valuable clue in identifying homogeneous regions and characteristics in the images. We recognize texture when we see it but it is very difficult to define. In (Tuceryan and Jain, 1998), the authors divided the principal approaches used to describe texture into statistical, geometrical, and model-based. The statistical approaches define texture as a spatial distribution of the gray values and analyze their relations. An example of this approach is the gray level co-occurrence matrix (GLCM) proposed by (Haralick et al., 1973; Haralick, 1979). Here, the authors used gray level co-occurrence features to analyze remote sense images. They computed GLCM for a distance of one with four directions, using seven classes in the classification problem, they obtained approximately 80 percent classification accuracy using texture features. The work of (Rignot and Kwok, 1990) used texture features computed from GLCM for classifications of SAR images. In addition, despeckling algorithms were used to remove speckle in SAR images in order to improve classification results.

In geometrical approaches, texture is considered to be composed of texture elements or primitives. Then, the methods which follow these approaches are based on the geometrical properties of these texture elements. Analysis according to this paradigm includes extraction of texture elements, their shape analysis, and the estimation of their placement rule. An example of this approach was presented in (Tuceryan and Jain, 1990). Here, the authors proposed the extraction of texture tokens by using the properties of the Voronoi tessellation of a given image.

The model-based approaches assume that the image can be modeled using texture and that the parameters of the texture represent information about the scene. The modeled parameters capture the essential qualities of the texture being able to characterize the image content. An example of this approach was presented in (Datcu et al., 1998). Here, the authors aim at characterizing the image-content using stochastic models within a Bayesian framework. In particular, they draw the potential usage of Gibbs Random Fields (GRFs) models for the extraction of spatial information from remote sensing im-

ages. The authors demonstrated that it is possible to characterize texture through the estimation of the parameters of the model.

In fact, the most common models used for texture modeling are Markov Random Fields (MRFs) (Hassner and Sklansky, 1980) and GRFs models (Besag, 1974), (Geman and Geman, 1984). The justification for using GRFs for image modeling is given by the Hammersly - Clifford Theorem (Besag, 1974), (Hammersley and Clifford, 1971), which asserts that there is one-to-one correspondence between MRF and GRF. Since any random field can be modeled by a MRF with an enough large neighbourhood system, and any random field can be modeled by a GRF with an enough large clique system (Schröder, 2000). Since then, many model-based methods have emerged for information extraction based on the GRFs. For example, originally, the auto-binomial model was introduced for information extraction in (Cross and Jain, 1983). Later, in (Schröder et al., 1998) had highlighted the potential of Gibbs-Markov Random Fields models as a powerful and robust descriptor of spatial information of remote sensing image data. Here, the authors demonstrated the usefulness of GRF for spatial information extraction using both optical and SAR images. Specifically, the authors also introduce the Auto-binomial model (ABM) as major contribution in their work (Schröder, 2000). Recently, in (Hebar et al., 2009), the authors used the ABM model for information extraction and despeckling simultaneously using SAR images.

Gauss Markov Random Field (GMRF) also belongs to the Gibbs family and constitutes a quite simple image model. It is widely used for texture description and segmentation. For example (Chellappa and Chatterjee, 1985) used GMRF for image synthesization proving the generation of texture using this model. In (Derin and Elliott, 1987) GMRF was also used for segmentation. In (Jeng and Woods, 1991), the GMRF model was used to extract information from remote sensing images. The authors consider the MAP and MMSE estimate for both image estimation and image restoration. The work of (Walessa and Datcu, 2000) proposed a method that combines despeckling and feature extraction from remote sensing SAR images using GMRF.

As a conclusion of this review we remark that the methods, which perform despeckling and information extraction from SAR images (e.g. (Walessa and Datcu, 2000) and (Hebar et al., 2009) cover a quite complete analysis of SAR images. These methods are involved in the data-driven evaluation approach of this thesis. Their workflows are fully discussed in section 5.1.

4.4 Conclusion

- As a conclusion we remark that the use of MAP in the estimation of noise-free images offers many advantages since it allows the introduction of a prior knowledge about the scene through stochastic models, which approximate the scene. Then, the methods for SAR image analysis could extract information from images depending on the used prior model.
 - The Bayesian inference approach through the use of likelihood and prior models presented satisfactory results in the analysis of SAR images, considering the despeckling process and information extraction.
 - The Markov-Gibbs Random Fields models have been used for many decades in texture modeling, showing a good adaptation of the model to the texture image.
-

Chapter 5

Data-driven evaluation of SAR information extraction methods

In this chapter, we suggest automated methods being able to extract and interpret the genuine information contained in high resolution SAR images. The image information content is extracted using model-based methods based on Gibbs Random Fields models combined with a Bayesian inference approach. The approach enhances the local adaptation by using a prior model, which learns the image structure; it enables despeckling with minimum loss of resolution and simultaneously estimates the local description of the structures. From these we may obtain detection, classification, and recognition of the image content.

In this context, this chapter evaluates Bayesian despeckling and information extraction methods based on Gauss Markov and Auto-binomial Gibbs Random Fields with application to TerraSAR-X data. The first evaluated method, here named *MAP-GMRF*, uses a Gauss Markov Random Field (GMRF) as prior model in order to preserve texture (Walessa and Datcu, 2000). The second method uses an Auto-binomial model (ABM) as prior (Hebar et al., 2009). It is referred as *MAP-ABM* in this work. Both methods are based on Maximum *a posteriori* estimation, and evidence maximization framework. The overall objective of this chapter is to carry out a comparative evaluation of both despeckling and information extraction of SAR images methods. Several quantities that can be considered as objective metrics are proposed for assessing the quality of the despeckling process. The information extraction is expressed as parameter estimation problem (Datcu et al., 1998), where the evaluation focuses on determining the robustness of texture extraction using unsupervised and supervised classifications, which give us the clues to decide which model to be used according to the kind of structure in the SAR image to be extracted. The evaluation is performed using two types of data 1) Simulated or synthetic data generated by a known prior model (GMRF or ABM), and its respective parameters for the data, and a likelihood model for speckle (Gamma distribution). 2) TerraSAR-X images as real SAR data, which follow a known speckle distribution, thus allowing us to prove how well the prior model adapts to the data. The evaluation aims at answering the question: How well could the model explain the nature of the data?

In the following sections, we start with describing the evaluated methods in section 5.1. Here, the Bayesian approach applied to information extraction and despeckling of SAR images, and the overview of *MAP-GMRF* and *MAP-ABM* methods are explained. Section 5.2 presents the used data set, which is formed by simulated and real SAR data.

The quantitative evaluation of despeckling is shown in section 5.3, here, we propose several objective measurements that can be used as metrics to qualify the quality of the despeckled image and these metrics have a semantic meaning. Section 5.4 details the quantitative evaluation of information extraction as well as the used metrics for the evaluation. Finally, the conclusion is presented in section 5.4.4.

5.1 Description of SAR information extraction model-based methods

The speckle within the SAR images is modelled as multiplicative noise (Touzi, 2002) representing a SAR images as follows

$$\mathbf{y} = \mathbf{x} \cdot \mathbf{z}, \quad (5.1)$$

where \mathbf{y} represents a SAR image, \mathbf{x} denotes an ideal noise-free image, and \mathbf{z} stands for the speckle, which is modelled as unit mean Gamma distribution (Goodman, 1975; Oliver and Quegan, 1998).

The probability density function (pdf) for the likelihood of the square root intensity of (5.1) is well-approximated by a Gamma distribution (Oliver and Quegan, 1998) given by

$$p(y_s|x_s) = 2 \left(\frac{y_s}{x_s} \right)^{2L-1} \frac{L^L}{x_s \Gamma(L)} \exp \left(-L \left(\frac{y_s}{x_s} \right)^2 \right), \quad (5.2)$$

where y_s is the SAR image pixel, x_s represents the noise free pixel, $\Gamma(\cdot)$ denotes the Gamma function, and L is the number of looks of the SAR image. When the speckle is fully developed within the SAR image, L is equal to 1. Multiple looks may be generated by averaging the resolution over the azimuth or range of the image, which degrades spatial resolution but reduces the speckle (De Vries, 1998). The term *Equivalent Number of Looks* (ENL) is used to estimate the number of looks L , and is defined as the number of independent intensity values averaged per pixel (Oliver and Quegan, 1998), which is expressed as

$$ENL = \frac{(\text{mean})^2}{\text{variance}}, \quad (5.3)$$

over a homogeneous area.

5.1.1 Bayesian Approach

The Bayesian inference approach within the image processing field was presented in (Datcu et al., 1998). The idea of a Bayesian-based despeckling and information extraction method is to find a *prior* model, which well approximates the distribution of the image, and a likelihood model, which estimates the speckle within the SAR image, in order to obtain the *posterior* pdf. The two levels of inference are model fitting (parameter estimation) and model selection. The evidence maximization framework and the Maximum *a Posteriori* (MAP) estimator are used in order to find the best model parameters. The Bayesian inference is given through the Bayes theorem as follows

$$p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}) = \frac{p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{\theta})}{p(\mathbf{y}|\boldsymbol{\theta})}, \quad (5.4)$$

where \mathbf{x} is the noise-free image, \mathbf{y} represents the SAR image, and $\boldsymbol{\theta}$ stands for the unknown model parameters. The conditional pdf denoted by $p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})$ indicates the *posterior* pdf, which stands for an ideal noise-free image, $p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$ represents the likelihood pdf, $p(\mathbf{x}|\boldsymbol{\theta})$ denotes the prior pdf, and $p(\mathbf{y}|\boldsymbol{\theta})$ is the evidence term, which acts as a normalization constant in the first level of inference.

According to Bayesian inference the first level is used to model at best SAR data, estimating the MAP of the noiseless image, which is expressed by the *posterior* pdf and is obtained using the likelihood and prior functions together, as follows

$$\hat{\mathbf{x}}(\mathbf{y}) = \arg \max_x p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{\theta}), \quad (5.5)$$

where $\hat{\mathbf{x}}$ represents the despeckled image.

The second order Bayesian inference enables the estimation of the parameter $\boldsymbol{\theta}$ using evidence maximization.

$$p(\mathbf{y}|\boldsymbol{\theta}) = \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\boldsymbol{\theta})d\mathbf{x}. \quad (5.6)$$

This evidence characterizes how well the estimated data fit to the original one. By selecting the best model, and taking into account the speckle likelihood, the solution can be found by maximizing the evidence as a function of $\boldsymbol{\theta}$ presented in (5.6).

5.1.2 Maximum a Posteriori despeckling using Gauss Markov Random Field (MAP-GMRF) and evidence maximization

A Maximum *a posteriori* (MAP) estimation and evidence maximization framework was proposed in (Walessa and Datcu, 2000) by using model-based despeckling and information extraction by means of a Gauss Markov Random Field (GMRF) model. It uses the Bayesian approach by taking advantage from the first level of Bayesian inference to obtain a MAP estimate of the despeckled image. This method uses as likelihood (5.2) and as prior, a GMRF model, which allows us to preserve and to characterize the texture. It is given by

$$p(x_s|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_s - \mu_s)^2}{2\sigma^2}\right), \quad (5.7)$$

where x_s represents the pixel to be evaluated at site s , σ^2 states for the variance, and μ_s is given by

$$\mu_s = \sum_{k \in \Omega_s} b_k \cdot (x_k + x'_k), \quad (5.8)$$

where $\boldsymbol{\theta} = [b_k]$ with $k = 1, 2..p$ can be denoted as the model vector parameter describing texture information, p is the total elements in $\boldsymbol{\theta}$, and Ω_s is the neighbourhood system around the central pixel x_s . The complexity of the method is determined by the size of the neighbourhood system, given by the model order (Chellappa and Chatterjee, 1985). Figure 4.1 depicts examples of different order neighborhood system. Here, note that the subindex k denotes the position of b element in the vector parameter $\boldsymbol{\theta}$.

MAP estimation (5.5) of the noise-free image is obtained by finding the product of the

likelihood (5.2) and prior (5.7) as follows

$$\begin{aligned} \frac{\partial}{\partial x_s} \log p(x_s|y_s) &= \frac{\partial}{\partial x_s} \log \left[2 \left(\frac{y_s}{x_s} \right)^{2L-1} \frac{L^L}{x_s \Gamma(L)} \exp \left(-L \left(\frac{y_s}{x_s} \right)^2 \right) \right] \\ &+ \frac{\partial}{\partial x_s} \log \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(x_s - \mu_s)^2}{2\sigma^2} \right) \right], \end{aligned} \quad (5.9)$$

where the solution for (5.9) is provided by setting the first derivative of the *posterior* logarithm to zero. Therefore, a fourth-order polynomial is obtained with four solutions, as follows

$$x_s^4 - x_s^3 \mu_s + 2L\sigma^2 x_s^2 - 2L\sigma^2 y_s^2 = 0. \quad (5.10)$$

A valid solution for x_s^{MAP} must be real-valued and positive and can be found by a case study of the four possible complex-valued roots, using the ICM (Iterative Conditional Modes) (Besag, 1986), (Winkler, 1995) algorithm for reaching a convergence of the method (Walessa, 2001). This convergence is reached after 5 iterations on average, therefore, the computation of the ICM to is limited to 10 iterations.

The estimation of parameter θ is determined using the evidence maximization as a function of θ , as given by (5.6), in the second level of Bayesian inference. Due to the complexity of the integral it can not be computed analytically, therefore, several approximations must be made (Walessa and Datcu, 2000):

1. The integral (5.6) consists of mutually-independent random variables, breaking the joint pdf into the products of its components.
2. The multidimensional pdf is approximated by a multivariate Gaussian pdf with Hessian matrix (\mathbf{H}), which is centered on the MAP pdf.

Approximation of the evidence $p(\mathbf{y}|\theta)$ is given by

$$\begin{aligned} p(\mathbf{y}|\theta) &= \int p(\mathbf{y}|\mathbf{x}, \theta) p(\mathbf{x}|\theta) dx, \\ &\approx \int \prod_{s=1}^S p(y_s|x_s^{MAP}) p(x_s^{MAP}|\theta) \cdot \exp -\frac{1}{2}(\mathbf{x} - \mathbf{x}^{MAP})^T \mathbf{H}(\mathbf{x} - \mathbf{x}^{MAP}) dx, \\ &\approx \frac{(2\pi)^{\frac{|S|}{2}}}{\sqrt{\det \mathbf{H}}} \prod_{s=1}^{|S|} p(y_s|x_s^{MAP}) p(x_s^{MAP}|\theta), \end{aligned} \quad (5.11)$$

where x_s^{MAP} is the MAP estimate of x_s obtained using the fixed parameter vector θ and (5.10). S is the whole set of sites (pixels) of the image and $|S|$ denotes its cardinal. \mathbf{H} is the Hessian matrix, which is the square matrix of the second-order partial derivatives of an univariate function given by

$$\mathbf{H} = -\nabla\nabla \sum_{s=1}^S \log (p(y_s|x_s^{MAP}) p(x_s^{MAP}|\theta)). \quad (5.12)$$

The determinant of matrix \mathbf{H} is given by

$$\det \mathbf{H} \approx \prod_{s=1}^{|S|} h_{ss}, \quad (5.13)$$

where h_{ss} are the components on the main diagonal of the matrix \mathbf{H} , which are found by (Walessa, 2001)

$$h_{ss} = \frac{6Ly_s^2}{x_s^{4MAP}} - \frac{2L}{x_s^{2MAP}} + \frac{1}{\sigma^2} \left(1 + \sum_{k \in \Omega_s} \theta_k^2 \right). \quad (5.14)$$

In order to obtain the highest evidence, the set of chosen parameter θ changes iteratively by using an evidence maximization algorithm (Datcu et al., 1998). The details of MAP computation and the evidence maximization algorithm, can be found in section 5.1.4.

5.1.3 Maximum a Posteriori despeckling using Auto-binomial model (MAP-ABM) and evidence maximization

The ABM model belongs to the group of Markovian models. The ABM is a discrete model, which is able to generate larger sets of textures, when compared to the GMRF. The GMRF is often used in the approximation of natural scenes and is able to generate textures, as for example forests, fields, etc. The ABM model is able to generate blob-like textures, which are often present in SAR images. The ABM has been successfully used for SAR texture estimation and despeckling (Hebar et al., 2009) and shows superior results over the GMRF for the despeckling of SAR images.

A Maximum *a posteriori* estimation using the Auto-binomial model method (Hebar et al., 2009) is also based on the Bayesian approach. The likelihood probability is modelled by a Gamma distribution (5.2). This method uses an Auto-binomial model as a prior pdf expressed as

$$p(x_s | \theta) = \binom{G}{x_s} \rho_s^{x_s} (1 - \rho_s)^{G - x_s}, \quad (5.15)$$

where x_s is the observed pixel at site s , G is the maximum gray value of the analyzed image, ρ_s is a function of $\theta = [a_s, \mathbf{b}_k]$ given by $\rho_s = \frac{1}{1 + \exp(-\eta_s)}$ and

$$\eta_s = a_s + \sum_{k \in \Omega_s} b_k \cdot \left(\frac{x_k + x'_k}{G} \right). \quad (5.16)$$

The value of a parameterizes the probability distribution of x_s without spatial interaction. The number of parameters b_k depends on the model order and x_k and x'_k are neighboring pixels around the analyzed pixel (Chellappa and Chatterjee, 1985). Fig 4.1 summarizes the neighborhood systems and the model order.

The MAP (5.5) is found by using as likelihood (5.2) and ABM prior (5.15).

$$\begin{aligned} \frac{\Delta}{\Delta x_s} \log p(x_s | y_s) &= \frac{\Delta}{\Delta x_s} \log \left[2 \left(\frac{y_s}{x_s} \right)^{2L-1} \frac{L^L}{x_s \Gamma(L)} \exp \left(-L \left(\frac{y_s}{x_s} \right)^2 \right) \right] \\ &+ \frac{\Delta}{\Delta x_s} \log \left[\binom{G}{x_s} \rho_s^{x_s} (1 - \rho_s)^{G - x_s} \right]. \end{aligned} \quad (5.17)$$

In (5.17), the first term can be solved analytically. However, the second one, which represents the Auto-binomial model, is solved by subtracting $\log p((x_s + 1) | \theta_s) - \log p(x_s | \theta_s)$, numerically. The finite difference (Δ) is introduced in order to solve the MAP estimation numerically since the analytical solution being difficult to derive at (Hebar et al., 2009). The result is given by the zeros of

$$-\frac{2L}{x_s} + 2L \frac{y_s^2}{x_s^3} + \log \left(\frac{G - x_s}{x_s + 1} \right) + \log \left(\frac{\rho_s}{1 - \rho_s} \right). \quad (5.18)$$

The MAP estimate (5.18) was found using the Brent's method algorithm (Brent, 1973) for the numerical solution of roots (5.18).

At the second level of Bayesian inference, the evidence maximization (5.6) was adapted into the logarithm form using the Auto-binomial model (5.15), as follows

$$\begin{aligned} \log p(\mathbf{y}|\boldsymbol{\theta}) = \sum_{s=1}^S & \left[\frac{1}{2} (S \log 2\pi - \log h_{ss}) \right. \\ & + \log \left(2 \left(\frac{y_s}{x_s} \right)^{2L-1} \frac{L^L}{x_s \Gamma(L)} \exp \left(-L \left(\frac{y_s}{x_s} \right)^2 \right) \right) \\ & \left. + \log \left(\binom{G}{x_s} \rho_s^{x_s} (1 - \rho_s)^{G-x_s} \right) \right], \end{aligned} \quad (5.19)$$

where h_{ss} is the approximated component of the Hessian Matrix \mathbf{H} given by (Hebar et al., 2009)

$$h_{ss} \approx \sum_{s=1}^S \left(\frac{6Ly_s^2}{x_s^{4MAP}} - \frac{2L}{x_s^{2MAP}} + \frac{1}{G - x_s^{MAP}} + \frac{1}{x_s^{MAP} + 1} \right). \quad (5.20)$$

The final approximation of the determinant of \mathbf{H} is given by

$$\det \mathbf{H} \approx \prod_{s=1}^{|S|} h_{ss}, \quad (5.21)$$

where only the diagonal elements from Hessian matrix are considered.

5.1.4 Overview of MAP-GMRF and MAP-ABM

The flowcharts of MAP-GMRF and MAP-ABM algorithms are given in Figure 5.1. Here, the main steps of both methods are detailed. The Evidence Maximization approach (EVM) is presented by Algorithm 1 in appendix (B).

We start by describing the mandatory input parameters of the methods since the results depend on them, and can be improved by their tuning. Further, in the next sections, the evaluation will show the relevance of the parameters on both despeckling and feature extraction performance.

1. Input parameters

Both methods need as input parameters:

- SAR image to be processed.
- The size of the analyzed image in columns and rows.
- The equivalent number of looks (ENL).
- The model order (MO).
- The size of the estimation window (M).
- The size of the step (N).

The MAP and evidence maximization procedures are performed inside the estimation window, whose size is $M \times M$. The window $N \times N$ determines how many pixels is

moved the estimation windows to continue with the next estimation. The window $M \times M$ is usually set by default at 15×15 , and $N \times N$ is set at 1. A model order (MO) determines the size of the model vector parameter θ and it can be chosen from 1 to 6. Then, the number of parameters differs according to the selected model order as follows (Model Order, Total parameters) = $\{(1, 2); (2, 4); (3, 6); (4, 10); (5, 12); (6, 12)\}$.

2. Steps of the Methods

The despeckling and information extraction from SAR images are performed by MAP-GMRF and MAP-ABM following the steps described in Figure 5.1. The two methods differ in term of equations in the MAP and EVM computations (inside the blue box in Figure 5.1). The different steps of Figure 5.1 are explained in the following:

1. Set initial parameter: The window sizes $N \times N$ and $M \times M$ and initial values of θ are set.
2. The equivalent number of look (ENL) is computed using (5.32) inside window with size 35×35 pixels over homogeneous area.
3. Strong scatterers are removed from the SAR image.
4. **MAP estimation:** The MAP estimation is performed using (5.10) in the case of MAP-GMRF and (5.18) in the case of MAP-ABM for all the pixels inside the estimation window ($M \times M$). It is computed using the initial values of θ and the parameter θ is changed iteratively until the maximum evidence is reached. Later, the estimated θ parameter is used for final MAP estimation.
5. **Evidence maximization.** The parameters of the models are estimated by means of evidence maximization approach. The evidence is computed according to (5.11) for GMRF and (5.19) for ABM using the MAP estimate (x_s^{MAP}) found in the previous step, for all pixels inside the $M \times M$ window.¹ The pseudo code for evidence maximization is detailed by algorithm 1 in appendix B.
6. Estimate the MAP using the new estimated parameters ($\hat{\theta}$).
7. Re-insert the strong scatterers in the image.
8. Write the results: the despeckled image (\hat{x}) and the estimated vector parameter $\hat{\theta}$.

¹Observe that step 4 and 5 are repeated until the maximum evidence is reached.

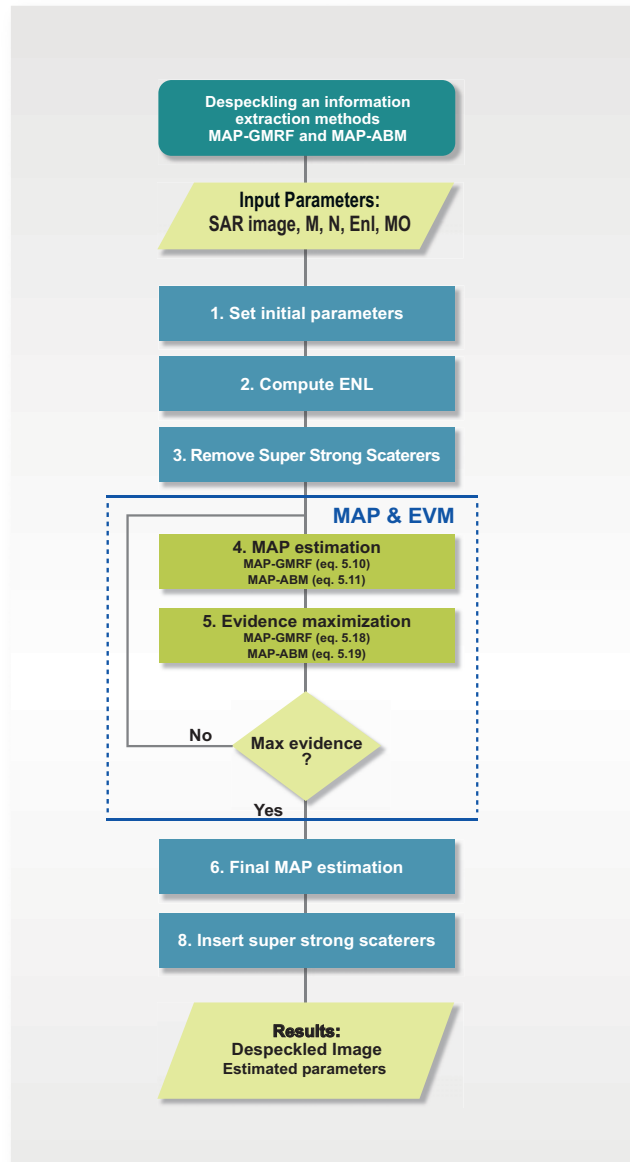


Figure 5.1: Flowchart of MAP-GMRF and MAP-ABM. The main steps are performed inside the blue box, which shows the MAP and EVM approaches. They differ in the equation. MAP-GMRF uses (5.10) and (5.11) and MAP-ABM uses (5.18) and (5.19). Both methods provide as results the despeckled image (\hat{x}) and the estimated model vector parameter $\hat{\theta}$, which corresponds to the estimation of texture.

5.2 Description of the test data set

The methods were assessed using simulated-speckled data and real SAR images. Simulated images were used in order to evaluate the performances of the despeckling methods in terms of quantitative measures such as Fidelity, Signal-to-Noise Ratio (SNR), Peak SNR (PSNR), Structural SIMilarity (SSIM) index, where the noise-free image is required.

TerraSAR-X images (DLR, 2009) were selected as real SAR data. TerraSAR-X is a new generation, high resolution German radar satellite operating within the X-Band. It was launched in June 2007 and will operate for a period of at least 5 years. A complete description of TerraSAR-X products and their performance are presented in (Breit et al., 2006) and chapter 2.

5.2.1 Simulated SAR data

Two simulated data sets were created in order to evaluate despeckling and the texture estimations of the MAP-GMRF and MAP-ABM methods. The simulated data sets were degraded with speckle. Simulation of the speckle was performed according to (Raney and Wessels, 1988) with the number of looks equals to three in most of the cases.

5.2.1.1 Synthetic textures

The first data set presents the synthetic textures generated using GMRF and ABM models. Images with two kind of structures were created; firstly one with large-scale structures and secondly one with fine-scale structures. The images presented in Figure 5.2 were created using GMRF distribution according to (Chellappa and Chatterjee, 1985) with a given θ . The Gibbs sampler (Casella and George, 1992) was used to generate ABM textures, as presented in Figure 5.3. The values of the texture parameters θ are shown in Table 5.1 for the GMRF and ABM models, respectively. The generated textures were degraded with multiplicative speckle-noise and are presented in Figures. 5.2(b) and 5.2(d) in the case of GMRF, and Figures. 5.3(d) and 5.3(b) in the case of the ABM model. Each image has 256×256 pixels. The images are named *gmrf-t.l* (large-scale structures) and *gmrf-t.s* (small-scale structures) in the case of GMRF, and *abm-t.l* (large-scale structures) and *abm-t.s* (small-scale structures) in the case of Auto-binomial model.

Table 5.1: Values of model vector parameter θ used in the generation of the four examples of textures depicted in Figure 5.2 and Figure 5.3.

GMRF											
Parameter θ		b_{11}	b_{12}	b_{21}	b_{22}	b_{31}	b_{32}	b_{41}	b_{42}	b_{43}	b_{44}
gmrf-t.l (Fig. 5.2(a))		0.3800	0.4500	-0.1600	-0.2400	-0.0500	-0.4400	-0.2100	0.0900	0.4700	0.2100
gmrf-t.s (Fig. 5.2(c))		0.3012	0.4451	-0.1330	-0.1117	-0.0178	-0.1180	0.0386	0.0354	0.0484	0.0118
ABM											
	a	b_{11}	b_{12}	b_{21}	b_{22}	b_{31}	b_{32}	b_{41}	b_{42}	b_{43}	b_{44}
abm-t.l (Fig. 5.3(a))	-2.658	1.75	0.24	-0.18	0.09	-0.46	-0.195	0.59	-0.017	-0.15	0.99
abm-t.s (Fig. 5.3(c))	-2.559	3.42	1.21	-0.83	-0.041	-1.67	-0.33	0.05	0.04	0.31	0.4

5.2.1.2 Speckled Brodatz textures

The second data set was created using nine different textures taken from the Brodatz data base (Brodatz, 1966). Figure 5.4(a) shows a mosaic of nine different textures, where each

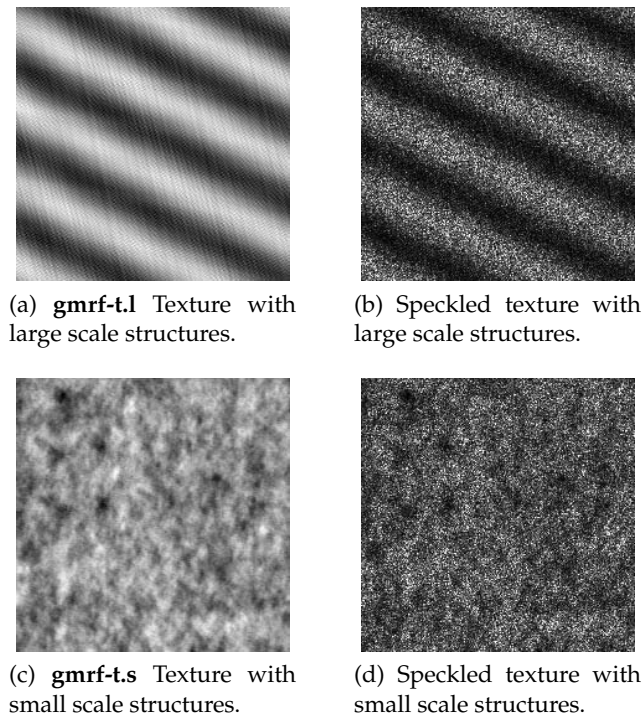


Figure 5.2: (a,c) Synthetic textures generated using GMRF distribution. (a) *gmrf-t.l* represents large scale structures, and (c) *gmrf-t.s* represents fine scale structures. (b,d) GMRF synthetic textures degraded with artificial speckle, ENL=3.

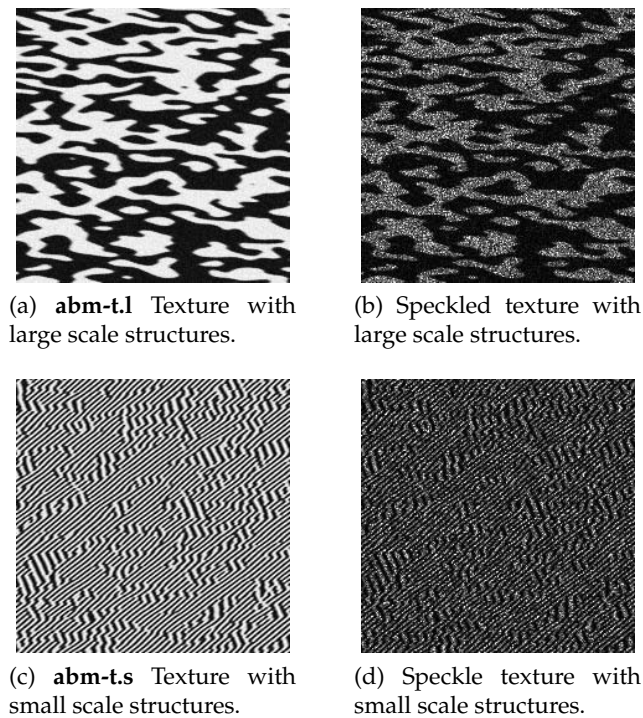
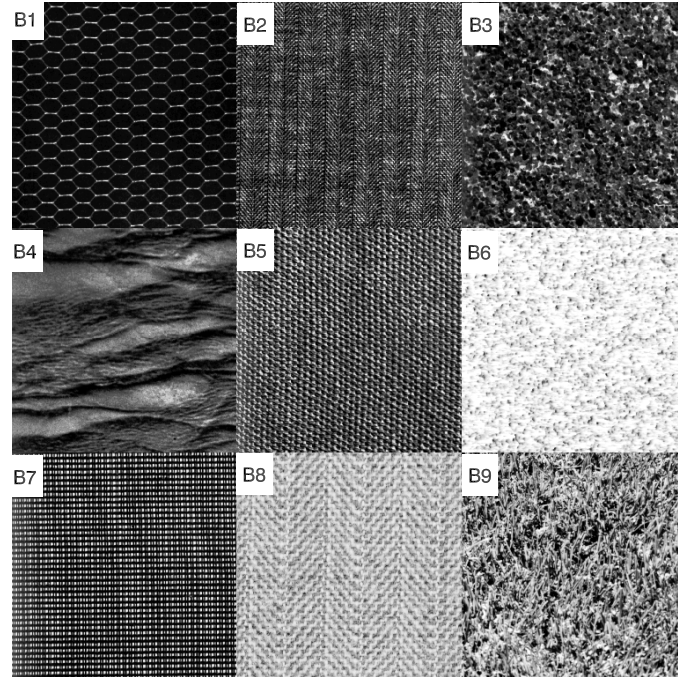
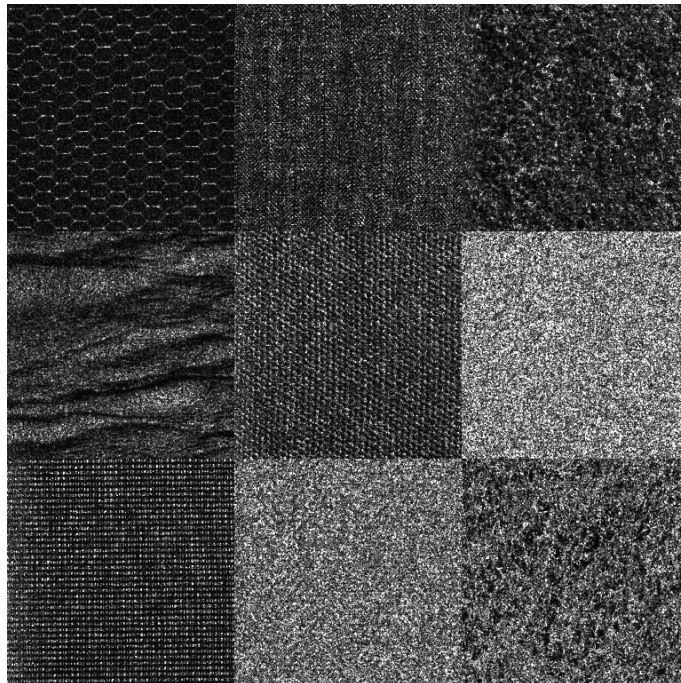


Figure 5.3: (a,c) Synthetic textures generated using Gibbs sampler and ABM prior. (a) *abm-t.l* contains large scale structures, and (c) *abm-t.s* represents fine scale structures. (b,d) ABM synthetic textures degraded with artificial speckle, ENL=3.

tile has 256×256 pixels. Here, the images are labelled using abbreviations B1 to B9. The mosaic was degraded with speckle and is depicted in Figure 5.4(b).



(a) Original Brodatz textures



(b) Speckled Brodatz textures

Figure 5.4: (a) Mosaic formed by 9 Brodatz textures. (b) Mosaic degraded with speckle, whose ENL=3.

5.2.2 Real SAR data

TerraSAR-X images (DLR, 2009) were chosen as real SAR data. The high resolution of the satellite allowed us to work with a higher variety of structures. The methods were assessed using a mosaic of real TerraSAR-X data and an application oriented TerraSAR-X sub-scene.

5.2.2.1 Textures from TerraSAR-X images

A mosaic of nine TerraSAR-X sub-scenes was created. The size of each image was 200×200 pixels. It is depicted in Figure 5.5. The different images correspond to radiometrically-enhanced single polarization High Resolution Spotlight products with a Multi-look Ground range Detected, whose pixel spacing = 1.25 m, range resolution = 2.89 m, azimuth resolution = 2.90 m, and ENL = 8.23 (DLR, 2009). The data are labelled from T1 to T9, starting from the upper-left part of the image. The different kind of structures contained in each tile of the mosaic is described in Table 5.2.

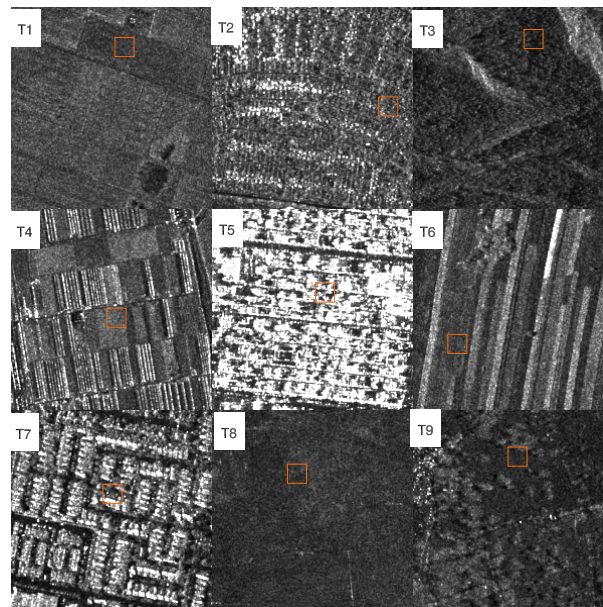


Figure 5.5: Mosaic of 9 TerraSAR-X sub-scenes. Texture T1 represents agricultural fields with some trees in the lower-right, in T2 there are several small houses with gardens, however it has in the upper part strong scatterers which represents medium size houses, in T3 a deep forest is shown, T4 represents agricultural fields and some greenhouses, in T5 there are high buildings, T6 correspond to dry fields, T7 has medium size houses, T8 represents a homogeneous forest and finally T9 is a park with several trees. Small squares in the images represent the areas where some quantitative measures were computed.

5.2.2.2 TerraSAR-X sub-scene

A 2000×2000 TerraSAR-X sub-scene over Hamburg in Germany was used. The image corresponds to radiometrically-enhanced single polarization Stripmap products with, Multi-look Ground range Detected, whose pixel spacing = 2.25 m, range resolution =

Table 5.2: Different kind of structures in the TerraSAR-X mosaic depicted in Figure 5.5.

TerraSAR-X tile	Image content
T1	Fields with some trees in the lower part.
T2	Several small houses with gardens and a small part with urban area.
T3	Deep forest.
T4	Agricultural fields and greenhouses.
T5	High buildings.
T6	Dry fields.
T7	Medium size houses.
T8	Homogeneous forest.
T9	Park with several trees inside.

5.77 m, azimuth resolution = 5.77, and ENL = 6.03 (DLR, 2009). Figure 5.6 shows the SAR image.



Figure 5.6: Example of TerraSAR-X sub-scene (2000×2000 pixels) over Hamburg in Germany. Small orange square in the upper part represents the area where some quantitative measures were computed.

5.3 Quantitative evaluation of despeckling

SAR images have been used in a large number of applications in different fields such as remote sensing for mapping, surface surveillance, search and rescue, mine detection, automatic target recognition. However, they are affected by a multiplicative speckle process, considered as noise, which reduces the radiometric quality and makes the tasks of feature extraction and scene recognition more difficult.

Speckle can be suppressed by filtering. Despeckling filters are tools, which help us to obtain more convenient images for visual interpretation because they have been enhanced, but it often generates a loss of information. Most commonly used speckle reduction filters have a good speckle smoothing capabilities. However, the resulting images are subject to the degradation of spatial and radiometric resolution, which can result in a loss of image information (Fjortoft et al., 2000; Xiao et al., 2003). Therefore, despeckling techniques should put emphasis on speckle removal (radiometric enhancement) with texture and structure preservation, in order to conserve spatial resolution. An accurate despeckling means that the textural and structural features contained in the image were properly modelled and this could allow an optimal information extraction.

A common way to evaluate the despeckling quality is using visual inspection. The visual observation provides us with our first indication of quality of the despeckled image. However, it is based on user interpretation and how the user perceives the image, bringing us different criteria when compared to the opinions of the users, thus introducing a subjective factor. In this thesis, we propose quantitative measurements that can be used as metrics to qualify the quality of the despeckled image and these metrics have a semantic meaning. The objective measurements consolidate the assessment of the despeckling quality. Previous works have been presented in (Espinoza-Molina et al., 2010b, 2011).

5.3.1 Quantitative quality measurements

While some of the quantitative measures presented below are calculated for both simulated and real data and some others are only used to evaluate the despeckling of the simulated data, since the noise-free image is required.

In the following, the *measurements* that are used to quantify the despeckling quality of both methods using simulated and real SAR data are presented.

5.3.1.1 Metrics for Simulated SAR data despeckling evaluation

It is worth noting that the following criteria were only used to evaluate the despeckling of the synthetic data, due to the noise-free image being available.

- **Fidelity**

Fidelity (5.22) is expressed as the Mean Squared Error (MSE) (Papoulis, 1991) between the despeckled image $\hat{x}_{h,l}$ and the original synthetic (noise free) image $g_{h,l}$ according to

$$MSE = \frac{1}{AB} \sum_{h=1}^A \sum_{l=1}^B (g_{h,l} - \hat{x}_{h,l})^2, \quad (5.22)$$

where A represents the columns and B stand for the rows in the image.

- **Signal-to-noise ratio (SNR)**

SNR (5.23) is given by

$$SNR = 10 \cdot \log_{10} \left(\frac{\sigma_g^2}{\sigma_e^2} \right), \quad (5.23)$$

where σ_g^2 is the variance of the noise-free image and σ_e^2 stands for the variance of the error between the original and the despeckled images. Higher SNR values correspond to a well despeckled image.

- **Peak SNR (PSNR)**

PSNR (5.24) is given by

$$PSNR = -10 \cdot \log_{10} \left(\frac{MSE}{d_{max}^2} \right), \quad (5.24)$$

where d_{max}^2 is the maximum intensity of the noisy image.

- **Structural SIMilarity Index (SSIM)**

The structural information in an image is defined as those attributes representing objects' structures within the scene, independent of the average luminance and contrast (Wang et al., 2004). Similarity measurement has three independent components: *luminance*, *contrast*, and *structure*, that combined to yield an overall similarity measure between images \mathbf{x} and \mathbf{y} :

$$S(\mathbf{x}, \mathbf{y}) = f(l(\mathbf{x}, \mathbf{y}), c(\mathbf{x}, \mathbf{y}), s(\mathbf{x}, \mathbf{y})), \quad (5.25)$$

where $l(\mathbf{x}, \mathbf{y})$ represents luminance, $c(\mathbf{x}, \mathbf{y})$ is the contrast, and $s(\mathbf{x}, \mathbf{y})$ denotes structure. The luminance comparison function $l(x, y)$ is function of mean values (μ_x) and (μ_y). It is defined as follows

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}. \quad (5.26)$$

The contrast comparison function takes a similar form

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}. \quad (5.27)$$

Finally, the structure comparison function is defined as follows

$$c(\mathbf{x}, \mathbf{y}) = \frac{(\sigma_{xy} + C_3)}{\sigma_x + \sigma_y + C_3}. \quad (5.28)$$

In order to simplify the expression, is assumed that $C_3 = \frac{C_2}{2}$. The result of combining (5.26), (5.27) and (5.28) is the similarity measure expressed in a specific form of the SSIM index (Wang et al., 2004) between images \mathbf{g} and $\hat{\mathbf{x}}$:

$$SSIM(\mathbf{g}, \hat{\mathbf{x}}) = \frac{(2\mu_{\mathbf{g}}\mu_{\hat{\mathbf{x}}} + C_1)(2\sigma_{\mathbf{g}\hat{\mathbf{x}}} + C_2)}{(\mu_{\mathbf{g}}^2 + \mu_{\hat{\mathbf{x}}}^2 + C_1)(\sigma_{\mathbf{g}}^2 + \sigma_{\hat{\mathbf{x}}}^2 + C_2)}, \quad (5.29)$$

where $\mu_{\mathbf{g}}$ represents the mean of the \mathbf{g} noise-free image, $\mu_{\hat{\mathbf{x}}}$ represents the mean of the $\hat{\mathbf{x}}$ despeckled image, C_1 is an included constant to avoid instability when

$\mu_g^2 + \mu_x^2$ is very close to zero. C_1 can be expressed as $C_1 = (K_1 F)^2$, where F is the dynamic range of the pixel values (255), and $K_1 \ll 1$ is a small constant. Similar considerations are applied to constant C_2 , which is included to avoid instability when $\sigma_g^2 + \sigma_x^2$ is very close to zero. The range of values for the *SSIM* lies between +1 to -1. The best value corresponds to +1 and the worst is -1.

5.3.1.2 Metrics for Real SAR data despeckling evaluation

The following criteria can be used for despeckling evaluation of real and simulated SAR images.

- **Bias**

Bias (5.30) is calculated as the absolute value of the difference between the estimation of the noisy image and the estimation of the filtered image.

$$Bias = |E\{\mathbf{y}\} - E\{\hat{\mathbf{x}}\}|. \quad (5.30)$$

The mean value of the filtered image should be equal to the mean value of the SAR image, thus indicating that the pixel values of the SAR images have been preserved. The best result is obtained when the difference is zero (VanTrees, 1968).

- **Ratio**

Ratio (5.31) is given by estimating the ratio between the noisy and filtered images. Since the speckle is considered as multiplicative noise, the best case is reached when the expectation of the ratio equals one (Oliver and Quegan, 1998).

$$Ratio = E\left\{\frac{\mathbf{y}}{\hat{\mathbf{x}}}\right\}. \quad (5.31)$$

- **Smoothness**

Smoothness (5.31) is also called ENL, it measures how well the speckle was removed without affecting the homogeneous areas and inserting distortions. It is given by the ENL of the despeckled image $\hat{\mathbf{x}}$ (Oliver and Quegan, 1998).

$$Smoothness = ENL = \frac{\mu_{\hat{\mathbf{x}}}^2}{\sigma_{\hat{\mathbf{x}}}^2}, \quad (5.32)$$

where $\mu_{\hat{\mathbf{x}}}$ is the mean and $\sigma_{\hat{\mathbf{x}}}^2$ represents the variance of the despeckled image. It must be computed over small homogenous areas. The higher the ENL is, the smoother the image is.

- **Sharpness**

Sharpness (5.33) represents the level of preserving the details and structures in an image. It is calculated by the ENL of the ratio between the noisy and filtered images. Sharpness is considered as a resolution preservation criterion. Indeed, it measures the quality of despeckling in terms of small details and structure preservation (Wallea, 2001).

$$Sharpness = ENL\left(\frac{\mathbf{y}}{\hat{\mathbf{x}}}\right). \quad (5.33)$$

Good despeckling is obtained when the value of the sharpness is closer to the ENL of the noisy image before filtering.

5.3.2 Experimental results

This section provides the experimental results obtained with the MAP-GMRF and MAP-ABM methods. The quality of the despeckling is evaluated in terms of the quantitative measurements previously presented in section 5.3.1 and visual inspection, using all data sets described in section 5.2. We note that apart from numerical values the visual impression of an image is still one the best way to evaluate. In order to ease the visual inspection, the original images are presented with the despeckled ones.

The experimental results are organized as follows 1) the despeckled images and the quantitative measures obtained with both methods using simulated and real SAR data are presented in section 5.3.2.1 and 5.3.2.2, respectively. 2) The impact of the estimation window size and the model order on despeckling performance is presented in section 5.3.2.3.

5.3.2.1 Despeckling results based on Simulated SAR data

All experiments were performed using an estimation window of 31×31 pixels size, a step size of 1×1 , and a model order of 4, which were chosen experimentally (Hebar et al., 2009).

1. Simulated SAR synthetic textures

The synthetic textures generated using a GMRF model, *gmrf-t.l* and *gmrf-t.s*, (see Figures 5.2(b) and 5.2(d)) and ABM model, *abm-t.l* and *abm-t.s*, (see Figures 5.3(b) and 5.3(d)), which correspond to textures with large and small scale structures, were despeckled using MAP-GMRF and MAP-ABM methods in order to evaluate the quality of despeckling. In the following, we start by presenting the results obtained with MAP-GMRF method.

- **MAP-GMRF results**

The despeckling results for MAP-GMRF method are shown in Figures 5.7(e)-(h). With visual inspection it can be seen that the speckle was slightly removed from all textures. However, better impressions are obtained in the case of textures with small scale structures depicted in Figures 5.7(f) and Figure 5.7(h) and the worse cases in texture with large scales (Figure 5.7(e) and Figure 5.7(g)).

Table 5.3 shows the objective measures of the despeckling using all textures. Here, the results show that the fidelity measures have high values demonstrating that the despeckling filter had introduced distortions during the reconstruction process. However, the higher fidelity values are in the textures generated with ABM model. This could be explained by the fact that an estimation window of 31×31 pixels used for estimating texture parameters, which is not large enough to estimate spatial interactions, since the image has large scale structures. The other possibility is that the algorithms for estimation of texture did not converge, therefore, the speckle distribution presents a texture, which was also estimated, as shown in Figures 5.7(e) and 5.7(g). The speckle removal and the values of the objective measures, in the case of Figure 5.7(f) are acceptable, showing that the used estimation window is enough large to estimate this kind of texture (small scale structure). The SNR and the PSNR values are very low. The negative values of the SNR can be thought of as by how many dBs of signal is beneath the noise floor. The SSIM index indicates that the despeckled images are not reliable, when GMRF is used for synthetic-generated

textures. This method introduces quite a large bias, the mean of the ratio image is, therefore, not close to 1, the smoothness is badly estimated because the images do not contain homogenous areas. The experimental results show that the MAP-GMRF method estimates better both synthetically-generated textures with GMRF.

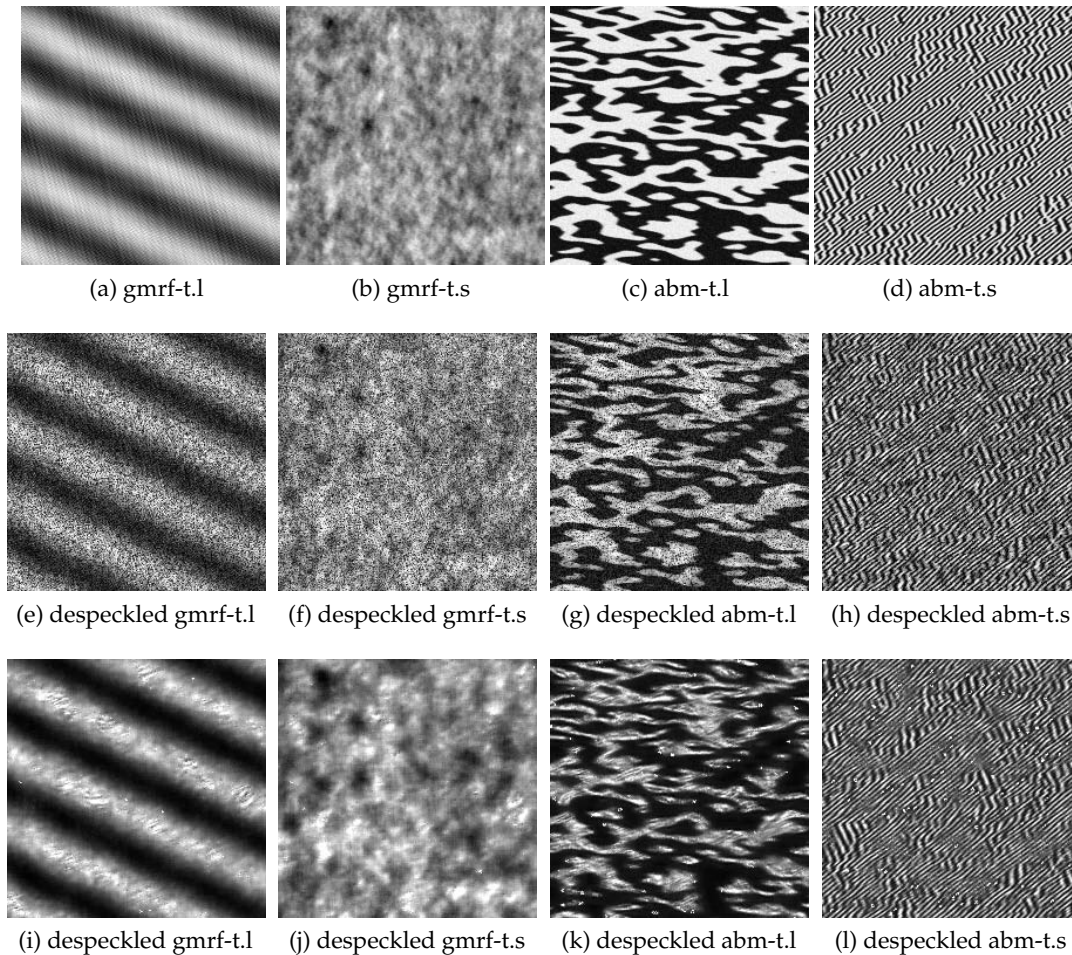


Figure 5.7: (First row) Noise-free synthetic textures generated with GMRF and ABM models. Despeckled images obtained with (second row) MAP-GMRF, (third row) MAP-ABM.

- **MAP-ABM results**

The despeckling results obtained using the MAP-ABM method, are shown in Figures 5.7(i)-5.7(l) and Table 5.3. The visual inspection shows that the speckle was well removed in the four images.

However, in Table 5.3 the quantitative measurements for simulated data (fidelity, SNR, PSNR, SSIM) show that the best results are obtained with textures (gmrf-t.l and gmrf-t.s.) generated with GMRF model. Specifically, the SSIM measure indicates that the despeckled images were enhanced. Moreover, the other objective measures in Table 5.3 show that the method well estimates speckle, because the ratio and bias measures are satisfactory. However, the sharpness is not close to 3, showing that distortions were introduced.

When comparing both methods, we notice that the MAP-ABM method estimates well the synthetically-generated textures with ABM as well as the ones with GMRF models, degraded by the speckle. The MAP-GMRF method has difficulties estimating the speckled texture with large scale structures, because the selected window size is too small, therefore, the MAP-GMRF method is unable to estimate texture and estimates random texture, which represents speckle.

Table 5.3: **Despeckling quality:** Quantitative evaluation criteria computed for **synthetic SAR data**. Best values in each category are displayed in bold. ENL** = 2.89 in all cases.

Data	Mean*	Bias	Ratio	Smoothness	Sharpness	Fidelity	SNR	PSNR	SSIM
MAP-GMRF									
Fig. 5.7(e) gmrf-t.l	128.3	25.8	0.81	2.74	5.9	2712.8	2.37	13.83	0.71
Fig. 5.7(f) gmrf-t.s	140.18	33.38	0.77	5.95	7.17	2855.8	-2.67	13.61	0.45
Fig. 5.7(g) abm-t.l	126.417	42.3	0.78	1.79	4.39	5406.1	5.29	11.62	0.74
Fig. 5.7(h) abm-t.s	125.46	52.2	0.71	2.12	6.8	5414.1	4.7	10.83	0.66
MAP-ABM									
Fig. 5.7(i) gmrf-t.l	128.29	3.07	1.11	5.88	4.34	492.63	8.64	21.24	0.93
Fig. 5.7(j) gmrf-t.s	140.18	3.78	1.05	35.7	29.52	432.8	3.51	21.8	0.85
Fig. 5.7(k) abm-t.l	126.41	3.15	1.39	2.48	4.44	2805.95	5.15	13.68	0.83
Fig. 5.7(l) abm-t.s	125.46	3.34	1.01	3.52	0.79	2960.9	4.9	13.45	0.79

* Mean refers to mean of original SAR image. ** ENL refers to Equivalent Number of Looks of the SAR image before the despeckling.

2. Simulated SAR Brodatz textures

Figures 5.8(b) and 5.8(c) depict the despeckled mosaic obtained with MAP-GMRF and MAP-ABM, respectively. Table 5.4 shows the quantitative measures for both methods.

• MAP-GMRF results

According to the visual inspection of Figure 5.8(b), all the images show good despeckling results and the textures were preserved, with exception of textures B6, B7 and B8. MAP-GMRF over-smoothes texture B6, where it becomes difficult to distinguish the texture in the reconstructed image, and destroys the texture in B8 and B7.

From the quantitative criteria presented in Table. 5.4, we note that in most of the cases the fidelity shows high values, specially in texture B6, which is the worst case. The SNR, PSNR and SSIM measures present acceptable values. However, the quantitative values obtained with texture B7 are not satisfactory. This method does not introduce bias in the reconstructed Brodatz images, as shown in Table 5.4. In here, all the textures have small bias values, showing that the images were well reconstructed during the despeckling. This is also confirmed with the ratio criterion, whose values are very close to 1. The smoothness measure shows good results in all the cases, but we note that texture B6, which has the highest value, is over-smoothed. The sharpness shows a loss of resolution in most of the cases. B7 presents the worst despeckling results in terms of bias, ratio, sharpeners, SSIM and smoothness.

• MAP-ABM results

With this method the visual inspection shows that all the images were well despeckled and the textures were preserved. Only, texture B2 is over-smoothed in its upper part.

Looking at Table 5.4, we observe that the fidelity and SSIM measures present poor performance, showing not accurately reconstructed images. The SNR and PSNR are acceptable. However, the values of sharpness show that the resolution was preserved. The ratio and bias measures present small values demonstrating that the method does not introduce bias in the despeckling process.

Table 5.4: **Despeckling quality:** Quantitative evaluation criteria computed for **simulated SAR data**. Best values in each category are displayed in bold. ENL** = 2.89 in all cases.

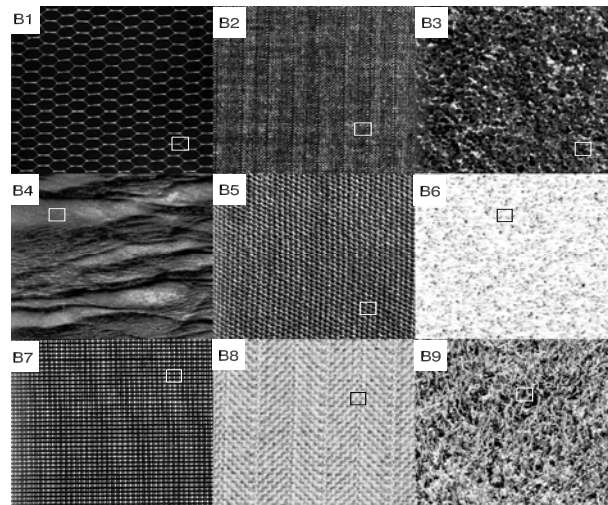
Data	Mean*	Bias	Ratio	Smoothness	Sharpness	Fidelity	SNR	PSNR	SSIM
MAP-GMRF									
Fig. 5.8(b)-B1	40.41	0.17	1.06	2.85	5.48	526.61	3.81	19.10	0.78
Fig. 5.8(b)-B2	77.69	0.40	1.08	2.66	5.00	1461.24	6.10	17.06	0.84
Fig. 5.8(b)-B3	71.14	0.09	1.02	3.46	10.57	984.01	6.11	18.13	0.85
Fig. 5.8(b)-B4	80.85	0.13	1.00	4.82	16.97	994.50	6.88	19.07	0.87
Fig. 5.8(b)-B5	88.72	0.09	1.06	2.69	6.30	1724.63	5.73	15.82	0.83
Fig. 5.8(b)-B6	215.95	0.32	1.00	9.18	12.83	5540.99	5.50	11.20	0.80
Fig. 5.8(b)-B7	73.97	0.57	1.75	2.07	0.74	3793.31	1.78	10.63	0.44
Fig. 5.8(b)-B8	173.97	0.16	1.01	7.87	16.25	4003.36	6.25	13.63	0.83
Fig. 5.8(b)-B9	127.85	0.18	1.02	4.61	13.01	2833.35	6.99	15.20	0.87
MAP-ABM									
Fig. 5.8(c)-B1	40.41	0.17	0.86	5.43	3.60	515.54	2.65	18.06	0.70
Fig. 5.8(c)-B2	77.69	0.40	0.87	17.12	2.37	1499.62	1.99	13.26	0.52
Fig. 5.8(c)-B3	71.14	0.09	0.81	8.74	3.81	933.87	3.99	15.91	0.74
Fig. 5.8(c)-B4	80.85	0.13	0.91	11.67	4.35	595.63	3.22	15.52	0.66
Fig. 5.8(c)-B5	88.72	0.09	0.83	5.84	3.47	1442.90	3.47	13.68	0.70
Fig. 5.8(c)-B6	215.95	0.32	0.99	33.24	4.18	2418.83	1.55	7.41	0.43
Fig. 5.8(c)-B7	73.97	0.57	0.73	5.31	1.94	2794.35	2.78	12.08	0.63
Fig. 5.8(c)-B8	173.97	0.16	0.99	25.22	4.20	1686.56	1.65	9.26	0.44
Fig. 5.8(c)-B9	127.85	0.18	0.90	12.11	3.67	1949.35	2.96	11.36	0.64

* Mean refers to mean of original SAR image. ** ENL refers to Equivalent Number of Looks of the SAR image before the despeckling.

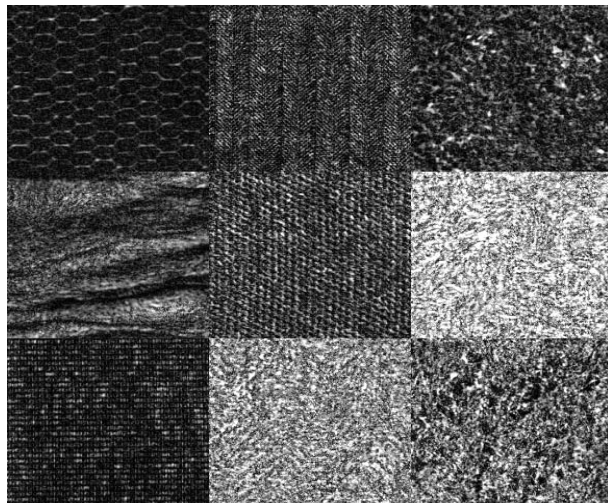
When comparing both methods, it can be noted that the best smoothness, sharpness, and fidelity criteria are provided by MAP-ABM method, for despeckled textures in general, although the ratio is not very close to 1. Also the MAP-GMRF method gives better ratio, SNR, PSNR and SSIM results, when it is compared with the MAP-ABM. The values for bias are the same for the two methods. **We can conclude that the MAP-GMRF method provides better results for the Brodatz textures, because they are natural textures and GMRF can model them better than the ABM.** We also observe that, with this kind of data, the smoothness criterion is not very relevant since the mosaic is composed of texture.

5.3.2.2 Despeckling results based on Real SAR data

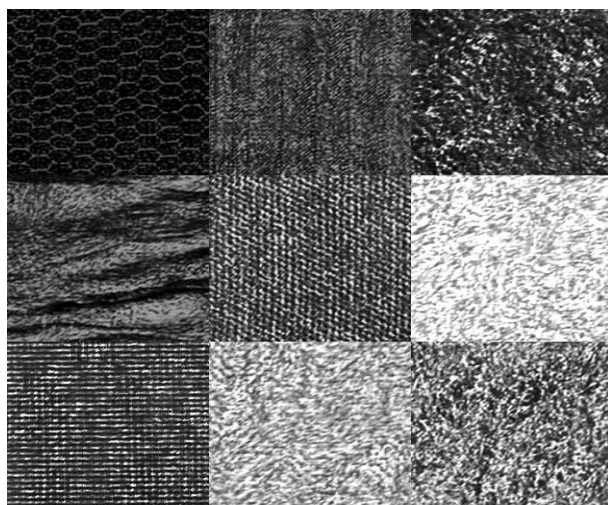
All experiments were performed using an estimation window of 31×31 pixels size, the step size of 1×1 , and a model order of 4, which were chosen experimentally (Hebar et al., 2009).



(a) Noise-free Brodatz Mosaic



(b) Despeckled Brodatz mosaic by MAP-GMRF



(c) Despeckled Brodatz mosaic by MAP-ABM

Figure 5.8: (a) Noise-free Brodatz mosaic. Despeckled images obtained with (b) MAP-GMRF, (c) MAP-ABM.

1. Mosaic of textures from TerraSAR-X images

Figures 5.9(b) and 5.9(c) show despeckled mosaic images obtained with MAP-GMRF and MAP-ABM, respectively. The quantitative results of this mosaic, are shown in Table 5.5 for both methods.

- **MAP-GMRF results**

In Figure 5.9(b) the despeckled images T2, and T6 were well-despeckled, here, the textures and structures were well-preserved. However, the despeckled images T1, T3, T8 and T9 were over-smoothed. Texture T5 is a controversial image because it presents very strong scatterers that should be preserved by despeckling methods.

According to the quantitative measurements from Table 5.5, the real SAR images were well despeckled using the MAP-GMRF method, since the mean is well estimated and the bias is quite low indicating a good preservation of the structures and textures within the scenes. The best bias results were obtained for scene T3 and T6. The results of the ratio images were very good, while, the best value belonged to texture T2, and the best smoothness criteria to texture T1. However, the values of sharpness are too high, demonstrating that some distortions were introduced.

- **MAP-ABM results**

Good visual impressions could be done from Figure 5.9(c). Here almost all images are well despeckled. However, texture T8 is over smoothed in its right part.

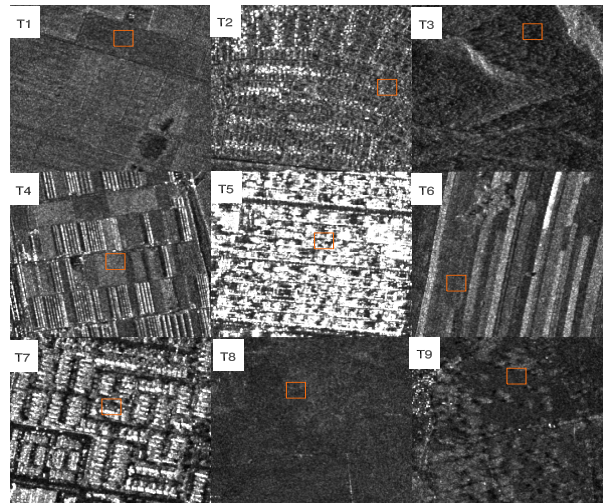
The MAP-ABM method introduced bias, as shown in Table 5.5, and was not able to estimate speckle noise, because the mean of the ratio images was not very close to 1. The quantitative measurements tell us that the best bias was obtained for scene T2, the best ratio and sharpness for scene T7, and finally, T1 presented the best smoothness criterion. The shifting of the mean of the SAR image is connected with the properties of the ABM model. The maximum value G (5.15) of the image, when is modelled with the ABM, should correspond to twice the mean of the image. The SAR scenes did not correspond to these values, therefore the MAP-ABM method introduced a bias, which occurs when the maximum value is not equal twice the mean of the image. In all experiments, the value of G was set at eight times the image mean for all scenes. The bias could be eliminated using adaptive estimation of the value G by using local statistics of the analyzing windows.

Both methods could not estimate the textures in scene T5, therefore the results were bad for this scene, because this scene had strong scatterers and the presence of speckle was minimal. Both methods removed strong scatterers at the beginning of despeckling and re-inserted strong scatterers at the end of despeckling.

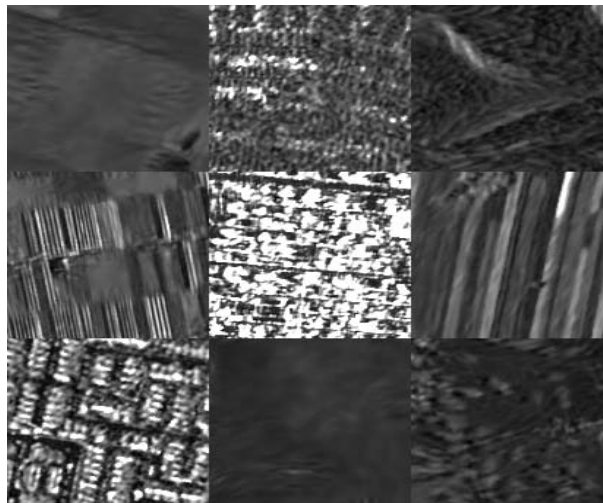
Comparing both methods, we note that the ABM model was able to estimate blob-like textures much better than the GMRF model. However, the GMRF model modelled natural scenes well. MAP-ABM gives better results in terms of smoothness and sharpness, showing a compromise between despeckling homogeneous areas and detail preservation; while MAP-GMRF provides better result in terms of bias and ratio criteria, showing the mean value of the original images is preserved.

2. TerraSAR-X scene over Hamburg

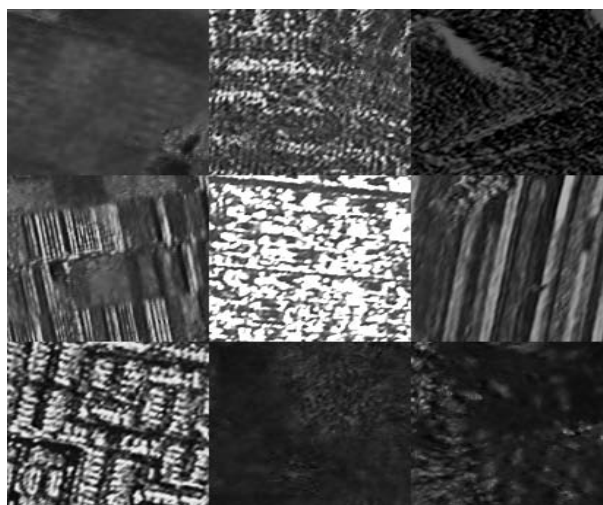
The TerraSAR-X over Hamburg shown in Figure 5.6 was despeckled using both methods. In order to notice the level of preserving details and speckle removal in homogeneous areas, a small part of the despeckled images is zoomed and presented.



(a) Original TerraSAR-X mosaic



(b) Despeckled TerraSAR-X mosaic using MAP-GMRF



(c) Despeckled TerraSAR-X mosaic using MAP-ABM

Figure 5.9: (a) TerraSAR-X mosaic. Despeckled TerraSAR-X mosaic obtained with (b) MAP-GMRF, (c) MAP-ABM.

Table 5.5: **Despeckling quality**: Quantitative evaluation criteria computed for TerraSAR-X data. The best values are in bold. ENL** = 8.23 in all cases.

Data	Mean*	Bias	Ratio	Smoothness	Sharpness
MAP-GMRF					
Fig. 5.9(b)-T1	116.99	0.33	0.99	496.07	33.23
Fig. 5.9(b)-T2	149.72	3.30	1.00	13.22	12.74
Fig. 5.9(b)-T3	86.55	0.15	0.99	20.47	27.85
Fig. 5.9(b)-T4	141.72	0.64	0.99	30.59	28.37
Fig. 5.9(b)-T5	416.02	40.50	1.10	2.22	1.80
Fig. 5.9(b)-T6	129.52	0.16	0.99	55.02	34.50
Fig. 5.9(b)-T7	86.64	0.38	0.99	439.20	30.55
Fig. 5.9(b)-T8	84.68	0.29	0.99	28.43	27.40
Fig. 5.9(b)-T9	85.84	0.29	0.99	10.61	27.21
MAP-ABM					
Fig. 5.9(c)-T1	116.99	7.95	0.93	540.69	31.72
Fig. 5.9(c)-T2	149.72	1.76	0.95	19.17	10.45
Fig. 5.9(c)-T3	86.55	3.95	0.94	16.39	30.70
Fig. 5.9(c)-T4	141.72	5.84	0.94	42.55	20.06
Fig. 5.9(c)-T5	416.02	83.67	1.08	5.32	1.97
Fig. 5.9(c)-T6	129.52	4.43	0.95	58.16	30.57
Fig. 5.9(c)-T7	172.23	8.51	0.97	6.17	8.17
Fig. 5.9(c)-T8	86.64	7.36	0.92	490.01	36.72
Fig. 5.9(c)-T9	84.68	7.49	0.90	112.41	19.42

* Mean refers to mean of original SAR image. ** ENL refers to Equivalent Number of Looks of the SAR image before the despeckling.

- **MAP-GMRF results**

The despeckled image obtained with MAP-GMRF is shown in Figure 5.10(a). Fig.5.10(b) depicts the zoomed part of the image marked with the white square. Here, it can be noted that MAP-GMRF over-smoothed the homogenous areas. The very strong scatterers were well-preserved and some small details were removed during the despeckling process.

The quantitative results are presented in Table 5.6 for MAP-GMRF. Looking at the ratio criterion, it was noticed that the best value was obtained by MAP-GMRF showing that the method did not introduce distortions during the despeckling. However, this method did not estimate the mean of this image well.

- **MAP-ABM results**

MAP-ABM showed better performance since the speckle was well-removed within the whole image, thus conserving the small details and strong structures as is shown in Figure 5.10(c) and Figure 5.10(d). The quantitative results are presented in Table 5.6 for MAP-ABM. The bias measure showed the superiority of the MAP-ABM method as well as smoothness and sharpness criteria, demonstrating that a better despeckling was performed with MAP-ABM, because it conserved the details and did not blurry the image. In general, the visual impressions of the results showed good performances by both methods.

When comparing both methods, this image had the presence of super-strong scatterers in the urban areas, which are preserved by both methods. **However, according to visual inspection and quantitative measurements MAP-ABM outperforms MAP-GMRF, in this case.**

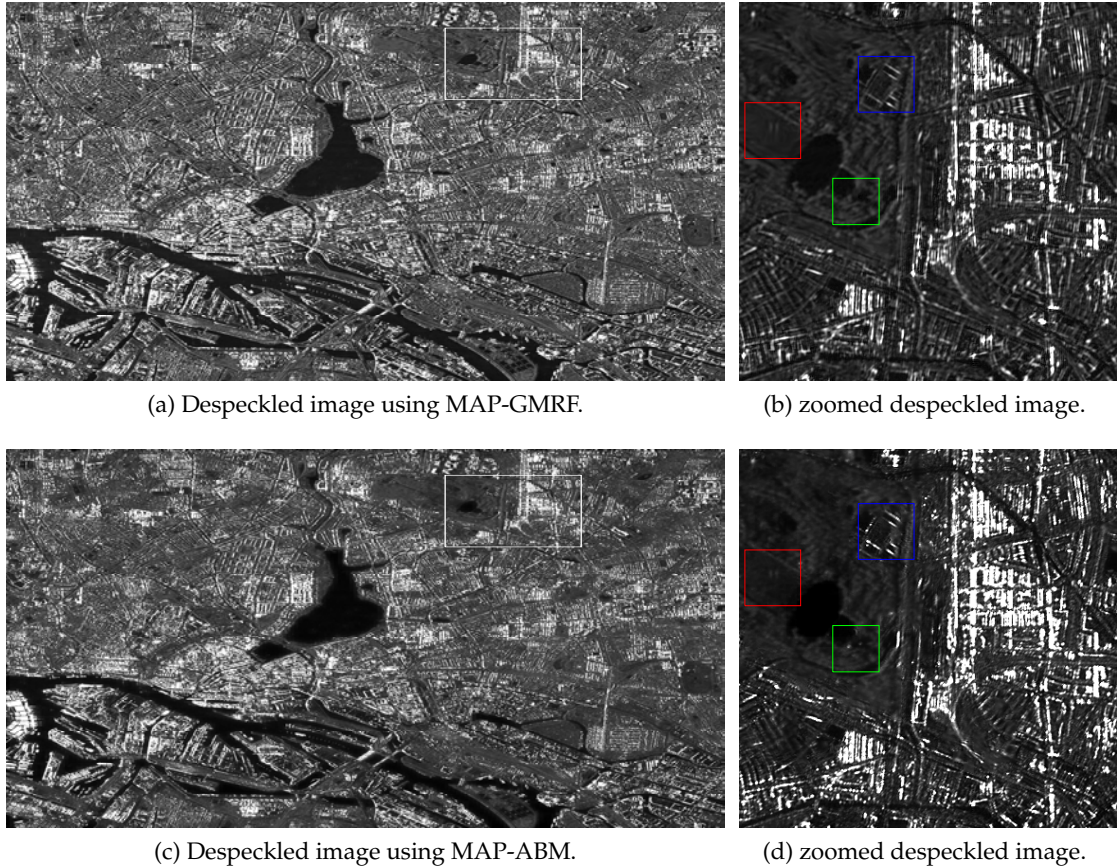


Figure 5.10: Despeckled TerraSAR-X sub-scene over Hamburg obtained using (a) MAP-GMRF and (c) MAP-ABM. Zoomed part within the despeckled image, in the case of (b) MAP-GMRF, and (d) MAP-ABM. In the zoomed images can be observed the level of preserving details (blue square), speckle removal in homogeneous areas (green square), and introduced distortion (red square).

Table 5.6: **Despeckling quality:** Quantitative evaluation criteria computed for **TerraSAR-X data over Hamburg** (see Figure 5.10). The best values are in bold. ENL** = 6.03.

Method	Mean*	Bias	Ratio	Smoothness	Sharpness
MAP-GMRF	264.74	20.84	1.05	116.01	1.86
MAP-ABM	264.74	6.73	0.89	462.19	5.17

* Mean refers to mean of original SAR image. ** ENL refers to Equivalent Number of Looks of the SAR image before the despeckling.

5.3.2.3 The impact of window size and model order on despeckling performance

The MAP-GMRF and MAP-ABM methods need the estimation window to perform the despeckling and to estimate texture parameters. **The size of the estimation window should be determined by the complexity of the scene or texture**, for example, we can guess that a larger estimation window could better estimate larger structures, being a consequence of using more data for the computations. However, the use of larger estimation window increases the computation time.

The model order is also determined according to the structural complexity of the scene since a higher model orders uses larger neighborhoods, which can capture high structural complexity in the images. To use a higher model order implies to use a larger analyzing window. Then, **we need to find a good compromise between the estimation window size, model order and computation time**.

In the next experiments, the impact of the estimation window and the model order on despeckling performance is evaluated using model order from two to four and an estimation window from (2×2) to (32×32) pixel size. We start by presenting the results of Figure 5.11, where the 1) visual inspection is used to evaluate the despeckling and gives an idea about how the estimation window size and the model order affect the despeckling quality. Later, 2) the quantitative measurements versus the computation time are detailed in Table 5.7 and the results are commented. Finally, 3) the level of preserving the details is shown in Figure 5.12.

1. Visual inspection

One of the textures (T1) from TerraSAR-X mosaic (See Figure 5.5) was despeckled using both methods.

The despeckled images are shown in Figures 5.11(a)-(b) in the case of MAP-GMRF method with model order 2 and 4, and Figures 5.11(c)-(d) show despeckled images using model order 2 and 4 with MAP-ABM method. The window size for each model was changed from 2×2 to 32×32 .

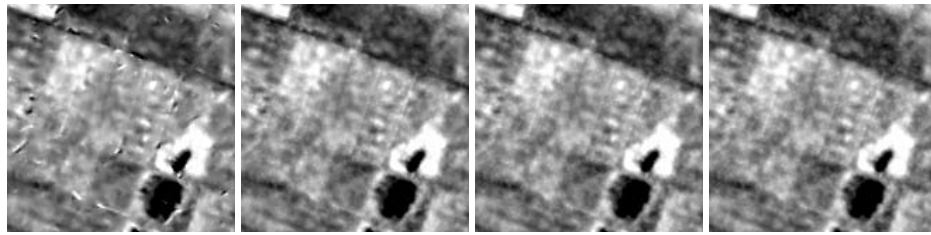
- **MAP-GMRF results**

The model order 2 of the GMRF gives good quality of despeckled images from window size of 8×8 pixels. The model order 4 of the GMRF, shown in Figure 5.11(b) requires larger window sizes to allow the convergence of the evidence maximization algorithm to the right texture parameters of the scene. It is clear to see that some distortions were introduced with estimation windows 2×2 and 8×8 . The visual impressions are acceptable when using estimation windows larger than 16×16 pixels size.

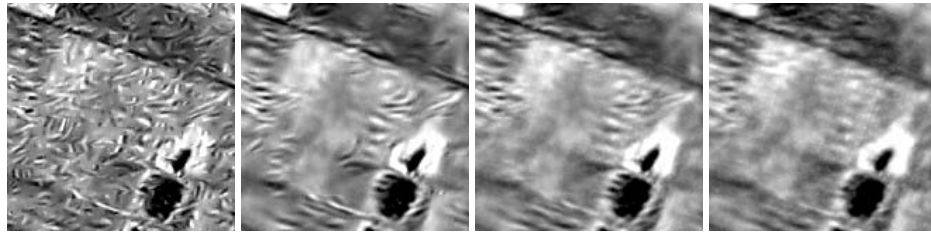
- **MAP-ABM results**

The speckle remains in the reconstructed images, when the model order 2 of the MAP-ABM method is used. The evidence maximization algorithm models the speckle instead of the scene. By increasing the model order larger structures are modelled and speckle is removed from the image.

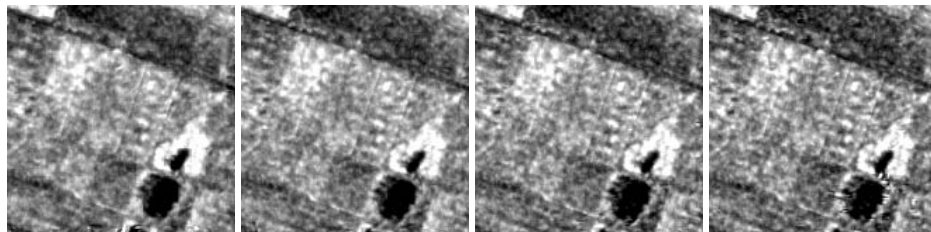
The model order 4 with window size of 32×32 pixels is a good compromise between the quality of despeckling and the complexity of the methods. MAP-ABM properly works with model order three or higher.



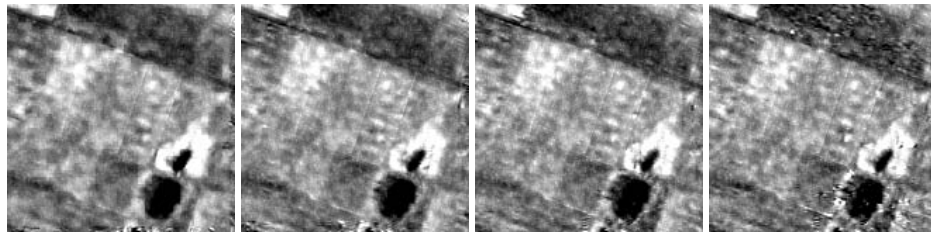
(a) Despeckled images using MAP-GMRF with Model Order 2.



(b) Despeckled images using MAP-GMRF with Model Order 4.



(c) Despeckled images using MAP-ABM with Model Order 2.



(d) Despeckled images using MAP-ABM with Model Order 4.

Figure 5.11: Despeckling results obtaining using (a-b) MAP-GMRF and (c-d) MAP-ABM method with estimation window of (from left to right) 2×2 , 8×8 , 16×16 and 32×32 pixels.

2. Quantitative measurements versus computation time

Table 5.7 shows the objective results for the despeckling, where the estimation window and the model order of MAP-GMRF and MAP-ABM models were changed. The experimental results were performed using two synthetic textures from Brodatz mosaic (B7 and B4 in Figure 5.4(b)) and two real SAR images from TerraSAR-X mosaic (T1 and T7 in Figure 5.5).

- **MAP-GMRF results**

The experimental results obtained with the MAP-GMRF method reported in Table 5.7(a) show that by increasing window size the fidelity decreased and the best results on average are obtained with a window size of 32×32 pixels. The fidelity is also

decreasing by increasing the model order. The best results were obtained with a model order 4. The reconstructed images are becoming smoother by the increasing the window size and model order, which indicates that the speckle is well removed. Speckle noise is well estimated, because reconstructed images have almost no bias and the mean of ratio is very close to 1. The SSIM value is very close to 0.9, which indicates that the reconstructed images are similar to the original images.

- **MAP-ABM results**

The MAP-ABM method behaves similarly to the MAP-GMRF method, when the model order and the window sizes are changed. The experimental results in Table 5.7(b) show, that the fidelity is decreasing by increasing the window size and increasing the model order of the ABM. The speckle was well estimated, the images were not biased, and the mean of estimated speckle was close to 1. The SSIM increased when the model order gets higher and images were not over smoothed by changing the window size, which means that some textures are well preserved. **The MAP-ABM method gives better objective results, but it is 3–5 times more computationally demanding than the MAP-GMRF method.** The reason is the estimation of the textures and the numerical computation of the MAP.

3. Level of preserving details

In order to complete the evaluation, in Figure 5.12, we display the despeckled image from small parts of the Hamburg sub-scene. Here, we can notice that the level of despeckling and detail preservation. Despeckled images with higher model order and larger window sizes produce more detailed despeckled images, meanwhile model order 2 and small window size produce blurred images, as shown in Figures 5.12(a) and 5.12(b), using MAP-GMRF method.

The despeckling results obtained with the MAP-ABM, presented in Figures 5.12(c) and 5.12(d) show that the quality of despeckling is improving when using larger window sizes. Details and strong scatterers are well preserved using model 4 and window size of 32×32 pixels.

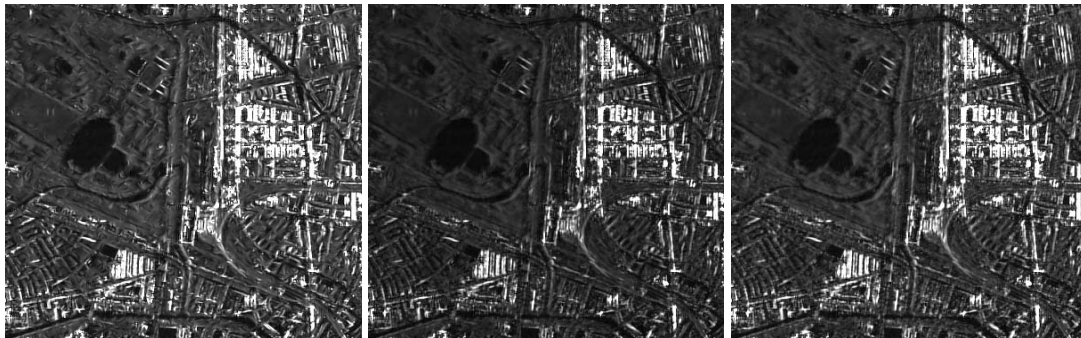
5.3.3 Conclusions

The goal of this section is to conclude which method (MAP-ABM or MAP-GMRF) gives the best despeckling results. The MAP-GMRF and MAP-ABM methods were compared using synthetically generated data, speckled Brodatz textures and real SAR data. The efficiency of MAP methods using texture models depends on how well the texture model adapts to the original scene. The texture estimation algorithm for MAP-GMRF and MAP-ABM methods is based on second level Bayesian inference. The despeckling performance of MAP-GMRF and MAP-ABM methods depends on the image textures, selected window size, and the model order of the Markov Random Field. In general, we should estimate the complexity of the scene to select the most convenient model order and window size.

- In this thesis we used fixed model order and window size, which were defined experimentally using trade-off between the complexity and efficiency of the methods.
- The **MAP-GMRF method** well estimates textures, the mean value of which does not change dynamically, meanwhile the MAP-ABM method better performs on the

Table 5.7: Quantitative despeckling criteria versus run-time processing, estimation window and model order in the case of MAP-GMRF and MAP-ABM despeckling methods.

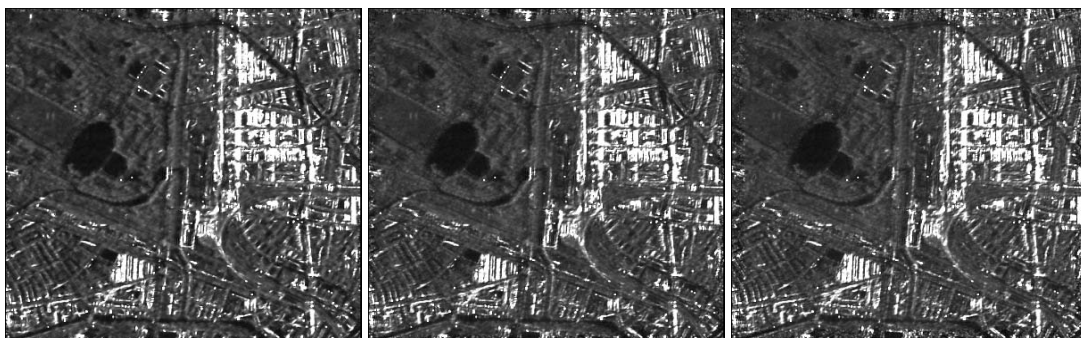
(a) MAP-GMRF							(b) MAP-ABM						
Fig. 5.4(b)-B7							Fig. 5.4(b)-B7						
Model Order	Criterion	Estimation window					Model Order	Criterion	Estimation window				
		2 × 2	4 × 4	8 × 8	16 × 16	32 × 32			2 × 2	4 × 4	8 × 8	16 × 16	32 × 32
2	Time(s)	11.1	13.7	23.2	49.9	140.2	2	Time(s)	112.32	219.76	555.12	1892.38	5072.89
	Bias	0.6	0.6	0.59	0.59	0.58		Bias	0.34	0.31	0.46	0.59	0.51
	Ratio	1.00	1.01	1.01	1.01	1.00		Ratio	1.11	1.11	1.10	1.09	1.09
	Smoothness	4.82	5.51	5.08	4.57	4.44		Smoothness	12.86	9.60	6.91	5.05	4.90
	Fidelity	3993.75	3973.52	3960.65	3912.97	3894.14		Fidelity	3214.98	3163.44	3096.65	3027.77	2994.98
	SSIM	0.44	0.45	0.46	0.46	0.46		SSIM	0.41	0.42	0.44	0.46	0.49
3	Time(s)	12.9	15.5	27.6	64.7	185.3	3	Time(s)	163.12	307.52	800.33	2790.64	7583.60
	Bias	0.64	0.51	0.46	0.59	0.51		Bias	0.34	0.31	0.46	0.59	0.51
	Ratio	1.68	1.69	1.71	1.75	1.77		Ratio	1.09	1.09	1.08	1.07	1.07
	Smoothness	2.17	2.24	2.04	2.04	1.92		Smoothness	12.99	10.27	7.06	4.90	4.74
	Fidelity	3974.10	3917.56	3905.76	3985.11	3901.07		Fidelity	2963.64	2958.28	2946.84	2942.56	2822.17
	SSIM	0.44	0.44	0.45	0.45	0.45		SSIM	0.41	0.43	0.45	0.48	0.50
4	Time(s)	24.2	31.5	58.0	131.4	340.4	4	Time(s)	253.99	474.58	1281.53	4318.00	11886.13
	Bias	0.53	0.53	0.46	0.59	0.51		Bias	0.60	0.59	0.56	0.59	0.57
	Ratio	1.67	1.68	1.70	1.75	1.77		Ratio	1.07	1.07	1.06	1.05	1.05
	Smoothness	2.16	2.21	2.03	2.03	1.91		Smoothness	15.32	11.31	7.23	4.96	4.66
	Fidelity	3876.67	3868.50	3825.87	3813.75	3794.31		Fidelity	2933.74	2909.37	2915.09	2861.90	2794.35
	SSIM	0.44	0.44	0.44	0.44	0.44		SSIM	0.41	0.43	0.45	0.60	0.63
Fig. 5.4(b)-B4							Fig. 5.4(b)-B4						
Model Order	Criterion	Estimation window					Model Order	Criterion	Estimation window				
		2 × 2	4 × 4	8 × 8	16 × 16	32 × 32			2 × 2	4 × 4	8 × 8	16 × 16	32 × 32
2	Time(s)	10.8	13.4	23.5	53.2	142.7	2	Time(s)	80.69	157.31	395.79	1199.30	3539.06
	Bias	0.2	0.2	0.21	0.21	0.21		Bias	0.52	0.52	0.52	0.52	0.52
	Ratio	1.01	1.01	1.01	1.01	1.01		Ratio	1.14	1.15	1.17	1.18	1.18
	Smoothness	4.91	5.58	5.17	4.66	4.88		Smoothness	243.41	677.73	149.11	149.03	85.18
	Fidelity	1483.61	1413.26	1352.29	1267.20	1174.42		Fidelity	700.42	719.20	766.86	835.41	854.01
	SSIM	0.92	0.92	0.91	0.89	0.88		SSIM	0.67	0.66	0.66	0.65	0.66
3	Time(s)	12.7	16.3	31.0	71.5	192.4	3	Time(s)	112.98	221.58	579.71	1881.03	6194.95
	Bias	0.19	0.18	0.17	0.14	0.13		Bias	0.52	0.52	0.51	0.51	0.52
	Ratio	1.01	1.01	1.01	1.01	1.01		Ratio	1.14	1.15	1.17	1.19	1.19
	Smoothness	4.84	5.53	5.06	4.57	4.59		Smoothness	271.04	754.26	142.16	91.08	40.67
	Fidelity	1346.12	1248.28	1176.93	1092.10	1003.08		Fidelity	699.73	718.04	770.50	850.84	886.94
	SSIM	0.91	0.91	0.91	0.89	0.88		SSIM	0.67	0.66	0.66	0.66	0.67
4	Time(s)	20.4	27.9	56.4	133.5	354.7	4	Time(s)	186.317	367.54	966.08	3143.31	10913.78
	Bias	0.02	0.02	0.00	0.14	0.21		Bias	0.56	0.52	0.51	0.51	0.51
	Ratio	1.00	1.01	1.01	1.01	1.00		Ratio	1.14	1.15	1.17	1.19	1.19
	Smoothness	4.82	5.51	5.08	4.57	4.44		Smoothness	11.80	11.99	11.50	11.40	11.61
	Fidelity	1233.75	1190.52	1173.65	1012.97	994.14		Fidelity	930.16	915.61	866.27	845.92	595.63
	SSIM	0.90	0.91	0.91	0.89	0.88		SSIM	0.59	0.60	0.67	0.66	0.66
Fig. 5.5-T1							Fig. 5.5-T1						
Model Order	Criterion	Estimation window					Model Order	Criterion	Estimation window				
		2 × 2	4 × 4	8 × 8	16 × 16	32 × 32			2 × 2	4 × 4	8 × 8	16 × 16	32 × 32
2	Time(s)	6.0	6.6	10.9	25.5	74.3	2	Time(s)	26.03	47.07	111.06	295.76	875.61
	Bias	0.19	0.20	0.25	0.22	0.15		Bias	8.27	7.70	7.11	6.67	5.87
	Ratio	0.99	0.99	0.99	0.99	0.99		Ratio	0.93	0.93	0.94	0.94	0.94
	Smoothness	316.34	412.51	601.24	341.39	337.79		Smoothness	271.88	305.74	431.18	316.59	317.83
3	Time(s)	6.4	7.7	13.9	33.2	97.8	3	Time(s)	39.92	73.21	177.61	478.98	1551.342
	Bias	0.17	0.19	0.24	0.26	0.18		Bias	8.77	8.04	7.38	6.82	6.01
	Ratio	0.99	0.99	0.99	0.99	0.99		Ratio	0.92	0.93	0.93	0.94	0.94
	Smoothness	234.44	268.78	814.61	426.36	445.15		Smoothness	338.24	386.75	553.78	374.74	401.87
4	Time(s)	8.4	12.6	23.6	65.7	171.9	4	Time(s)	70.36	120.72	314.37	842.02	2785.59
	Bias	0.24	0.26	0.24	0.29	0.19		Bias	9.57	8.67	7.76	7.08	6.17
	Ratio	1.00	0.99	0.99	0.99	0.99		Ratio	0.92	0.93	0.93	0.94	0.94
	Smoothness	122.40	175.35	1012.00	494.19	496.07		Smoothness	402.80	485.10	697.62	456.27	488.99
Fig. 5.5-T7							Fig. 5.5-T7						
Model Order	Criterion	Estimation window					Model Order	Criterion	Estimation window				
		2 × 2	4 × 4	8 × 8	16 × 16	32 × 32			2 × 2	4 × 4	8 × 8	16 × 16	32 × 32
2	Time(s)	6.8	8.9	14.7	33.0	87.5	2	Time(s)	73.78	153.49	399.44	1160.64	3410.69
	Bias	0.57	0.24	0.07	0.91	0.44		Bias	27.82	28.93	30.97	32.46	34.59
	Ratio	1.00	1.01	1.01	1.01	1.02		Ratio	1.12	1.12	1.14	1.15	1.17
	Smoothness	8.13	6.07	5.69	5.27	5.91		Smoothness	20.42	17.18	14.68	12.85	12.16
3	Time(s)	9.0	10.4	17.5	38.6	107.7	3	Time(s)	106.40	212.15	536.42	1395.08	3996.9
	Bias	0.29	1.19	0.03	0.20	0.65		Bias	28.61	30.01	32.34	33.42	35.27
	Ratio	1.00	1.00	1.00	1.00	1.00		Ratio	1.12	1.13	1.15	1.16	1.17
	Smoothness	5.88	4.12	4.06	3.21	3.53		Smoothness	18.27	15.39	12.29	11.70	11.85
4	Time(s)	10.8	17.5	31.0	68.8	171.7	4	Time(s)	168.12	343.83	962.23	2437.73	6831.13
	Bias	0.94	0.50	0.56	0.75	0.23		Bias	28.21	29.61	32.09	33.43	35.29
	Ratio	1.00	1.00	0.99	1.00	1.00		Ratio	1.12	1.13	1.15	1.16	1.17
	Smoothness	6.47	6.80	6.79	6.03	6.17		Smoothness	20.37	16.95	13.40	11.98	12.42



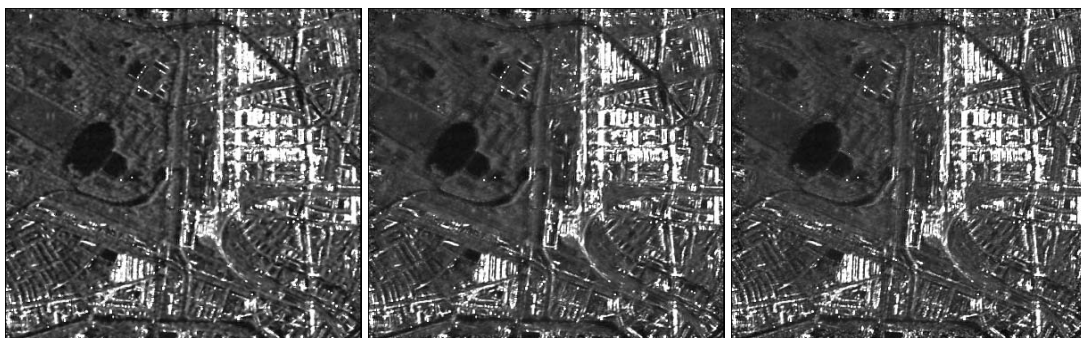
(a) Despeckled images using MAP-GMRF with Model Order 2.



(b) Despeckled images using MAP-GMRF with Model Order 4.



(c) Despeckled images using MAP-ABM with Model Order 2.



(d) Despeckled images using MAP-ABM with Model Order 4.

Figure 5.12: Despeckled images provided by (a-b) MAP-GMRF and (c-d) MAP-ABM method with model order 2 and 4, and estimation window of (from left to right) 4×4 , 16×16 , and 32×32 pixels, respectively. A good quality of the despeckled images is obtained with estimation windows bigger than 4×4 pixels. A good compromise is model order 4 and estimation window 32×32 .

blob-like textures, which appear often in real SAR images. The synthetic case shows that the MAP-GMRF can introduce bias, if the analyzing window does not correspond to the structures in the scene.

- The **MAP-ABM method** is able to model local statistics and better preserves point-like characteristics of the original image. The MAP-ABM model better adapts to the local changes, because it can model larger variety of textures. The MAP-ABM better preserves textures in the urban scenes and is more suitable for real SAR scenes such as the cities, scenes with real structures, etc.
- The impact of the window size was addressed to the synthetic and real SAR images. The images are better modelled by increasing the model order and the window size. **A good compromise is model order 4 and estimation window 32×32 pixel size.**

5.4 Quantitative evaluation of SAR information extraction

In the previous chapter the importance of texture analysis was presented. The texture analysis concerns mainly feature extraction and image coding. The texture analysis methods have been categorized in geometrical, statistical, signal processing and model-based methods (Tuceryan and Jain, 1998). However, the main goal of all these methods is to provide representative features for classifying the image content.

In this thesis, we focus on the model-based methods. In our work, the texture modelling is based on Gibbs-Markov Random Fields, where the texture is characterized by the model vector parameter (θ) (Descombes and Zhizhina, 2008). With this approach, the information extraction is defined as a problem of stochastic parameter estimation, or in other words as, the estimation of stochastic model parameters from the (image) data. Thus, the content of the image is not only described by spectral attributes but also by estimated parameters based on texture model (Schröder et al., 1998).

As previously presented in section 5.1, the evaluated methods (MAP-GMRF and MAP-ABM) provide a set of estimated parameters ($\hat{\theta}$), which corresponds to the estimation of texture, at the second level of Bayesian inference. After having the estimated model vector parameter, the evaluation focuses on the quality of the information extraction, which relies on the accuracy of the estimated parameters and classifications derived from the extracted texture features.

In this section, the quantitative evaluation of SAR information extraction is presented and it entails 1) the evaluation of the accuracy of the estimated model vector parameter ($\hat{\theta}$) using synthetic data, 2) determining the robustness of the texture extraction in terms of classifications and confusion matrices. 3) Moreover, since the results depends on the tuning of the input parameters of the evaluated methods, we address the impact of the estimation window and the model order on feature extraction and classifications. In the following, we start with defining the quantitative metrics used during the evaluation. The robustness of feature extraction evaluation in terms of classifications is presented in section 5.4.2 and the impact the model order and estimation windows is discussed in 5.4.3.

It is worth mentioning that previous works have been presented in (Espinoza-Molina et al., 2011), (Espinoza-Molina et al., 2010a), (Espinoza-Molina et al., 2009).

5.4.1 Quantitative Metrics

The classifications are assessed by means of confusion matrices.

5.4.1.1 Classification and Confusion Matrix

The interpretation of image is only possible if the classifiers can effectively label previously unseen objects. However, the recognition ability of classifiers depends on the quality of the selected features. Therefore, the goodness (robustness) of texture extraction is evaluated by using the estimated model parameters as primitive features to classify the image content. Later, the classification results can be quantified using confusion matrices.

A confusion matrix provides a complete way to describe the performance (accuracy) about the classification result. In addition, the confusion matrices offer an idea about the classes that a classifier has more difficulties to distinguish. Table 5.8 shows an examples of confusion matrix. Here, for a data set with k classes, the matrix has $k \times k$ dimension and each element (w_{ij}) represents the number of examples from class i that were misclassified as belonging to class j (Mitchell, 1997). The row indexes (j) of a confusion matrix correspond to true classes and the column indexes (i) correspond to predicted classes produced by applying the model. For any pair of true/predicted indexes, the value indicates the number of records classified in that pairing.

Table 5.8: Example of a confusion matrix.

Class _{i,j}	$i = 1$	2	3	...	k
$j = 1$	w_{11}			...	w_{k1}
2		w_{22}		...	w_{k2}
3			w_{33}	...	\vdots
\vdots	\vdots	\vdots	\vdots	\ddots	\vdots
k	w_{k1}	w_{kk}

The classifications can be performed by unsupervised and supervised classifiers. In the unsupervised classification case, a conventional pattern recognition algorithm, K-means (MacQueen, 1967), with a pre-defined number of clusters (k) can be used as classifier. In the case of supervised classification the maximum likelihood method will be used. Here, we can give regions of interest corresponding to the evaluated class as input.

5.4.2 Robustness of feature extraction evaluation in terms of classifications

The evaluation of the robustness of feature extraction parameter is performed in terms of unsupervised and supervised classifications, since the accuracy in the classification is strongly depended on the used primitive features. Both methods provide texture as primitive feature. The Brodatz texture mosaic shown in Figure 5.4(b) and the TerraSAR-X mosaic image, shown in Figure 5.5, were used for estimating texture parameters using both methods.

In the case of MAP-GMRF, Figures 5.13(a) and 5.13(b) show how appear the estimated texture parameter θ_1 obtained using Brodatz and TerraSAR-X mosaic, respectively. In the case of MAP-ABM, Figures 5.13(c) and 5.13(d) display how appear the estimated texture parameter θ_1 obtained for the Brodatz and TerraSAR-X mosaic, respectively.

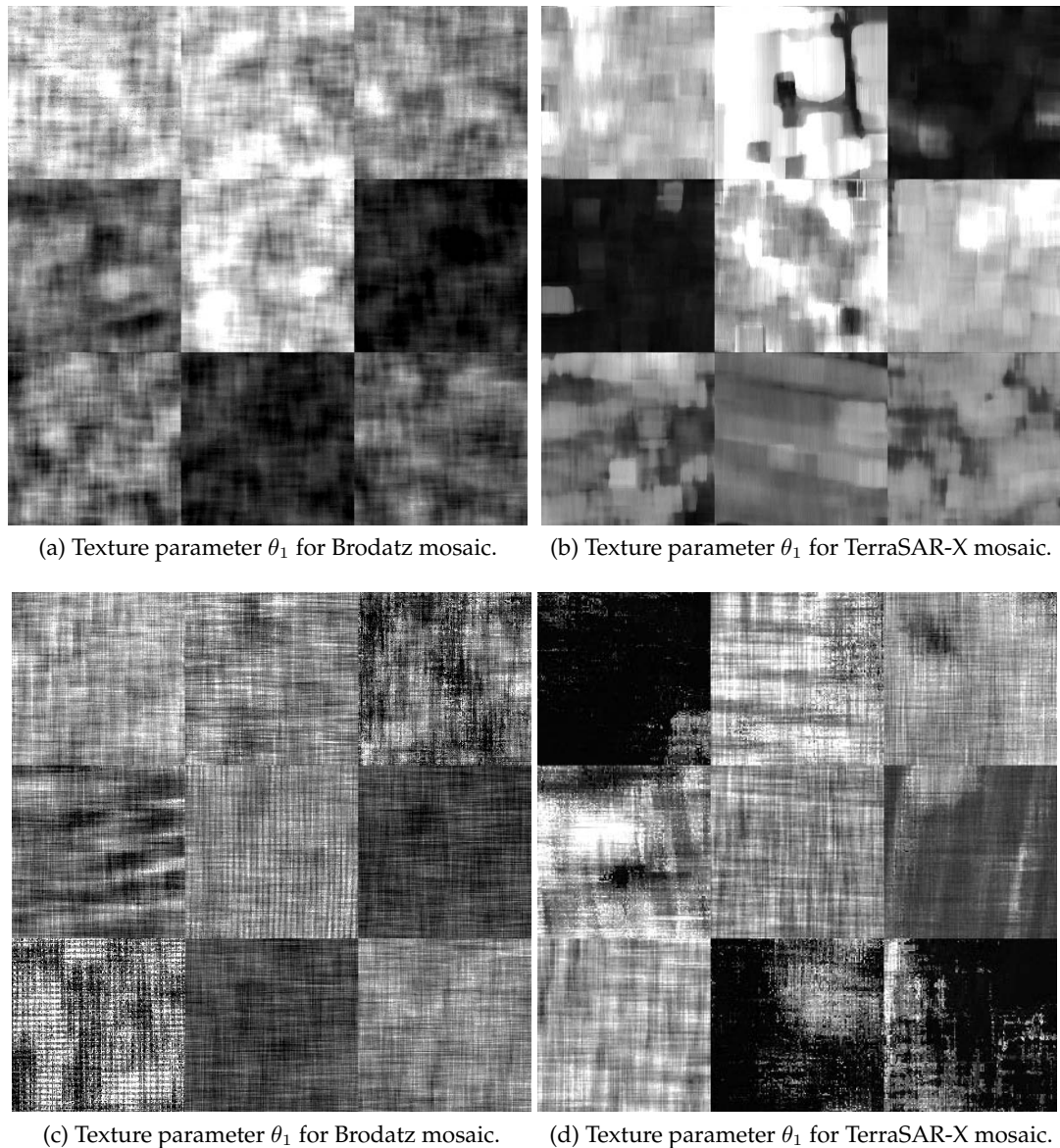


Figure 5.13: Texture parameter θ_1 for Brodatz mosaic (see Figure 5.4(b)) and TerraSAR-X mosaic (see Figure 5.5) obtained using (a-b) MAP-GMRF and (c-d) MAP-ABM.

The estimated texture parameters provided by both methods were classified using supervised and unsupervised classifiers. A K-means (MacQueen, 1967) algorithm with 9 classes was used in the case of unsupervised classification. The supervised classification was performed using a Maximum likelihood method, using 9 classes. Here, half of the data were used for training (ground-truth).

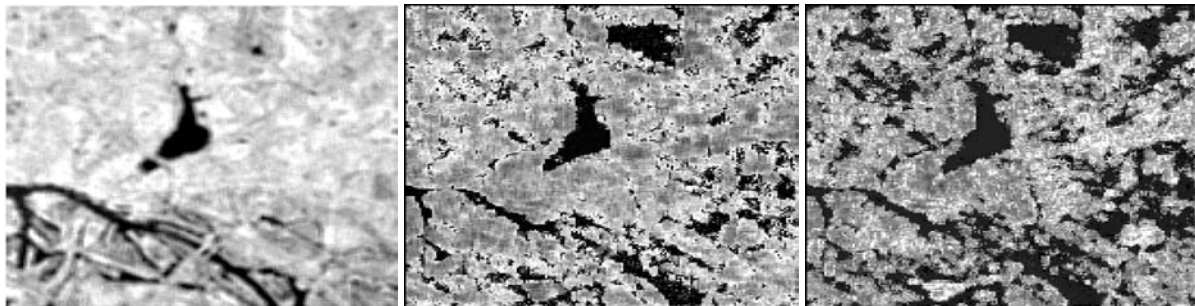
The TerraSAR-X sub-scene over Hamburg depicted in Figure 5.6 was also used for estimating texture parameters. Figure 5.14 shows the variance and the first two estimated texture parameters (θ_1 and θ_2) obtained using both methods.

In the following, we present the results of classifications using the data set described in section 5.2, starting with the case of simulated SAR data and later real SAR data. All experiments were performed using an estimation window of 32×32 pixels size, and a

model order of 4, which were chosen experimentally. According to the evaluation provided in last section, an estimation window of 32×32 pixels size is a good trade-off between accuracy of the estimation and results.



(a) Texture parameters obtained with MAP-GMRF.



(b) Texture parameters obtained with MAP-ABM.

Figure 5.14: From left to right (a) variance σ , θ_1 and θ_2 texture parameters provided by MAP-GMRF. (b) Parameter a , θ_1 and θ_2 texture parameters obtained using MAP-ABM.

5.4.2.1 Classification results based on Simulated SAR data

The Brodatz mosaic shown in Figure 5.4(b) was despeckled and the estimated texture parameters classified using unsupervised and supervised classifiers. The classification results will be discussed in terms of 1) confusion matrices and 2) the stability of the estimated parameters. This will be analyzed by computing the means and variances.

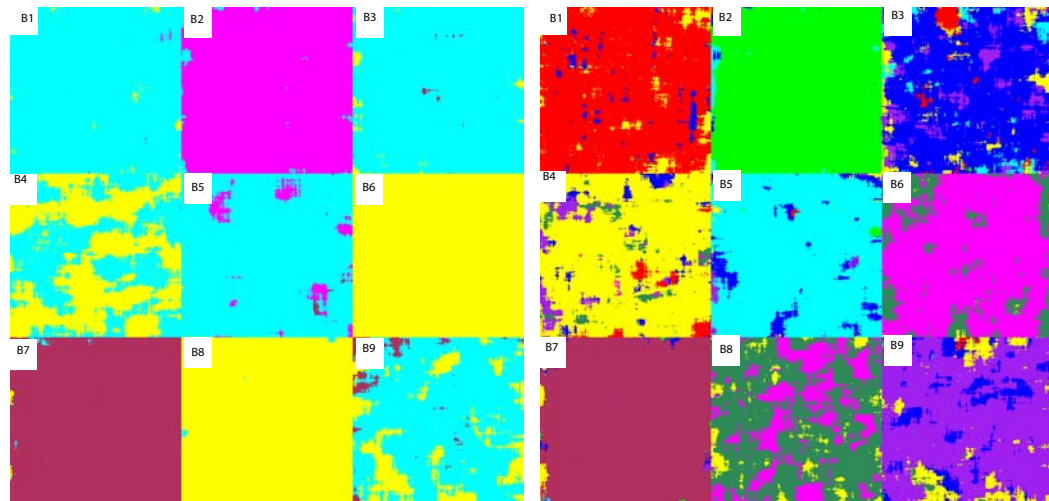
1. Confusion matrices

In the following the unsupervised and supervised classifications are discussed in terms of confusion matrices. It is worth noting that the confusion matrix presented in the case of unsupervised classifications shows only the recognized classes.

- **MAP-GMRF results**

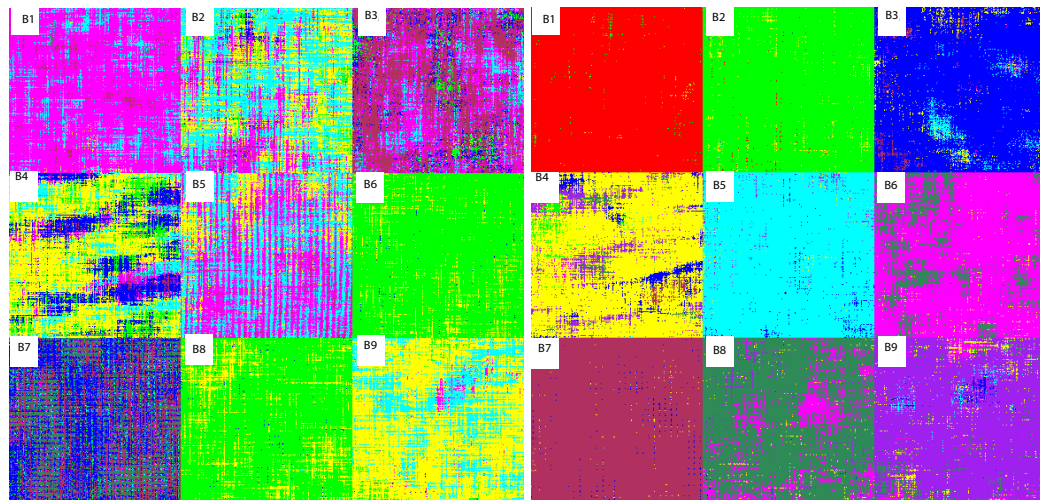
The results obtained with the MAP-GMRF method are shown in Figures 5.15(a)-(b) for unsupervised and supervised classifications, respectively. The classified texture parameters shown in Figure 5.15(a) recognized four different textures (B1,B2,B6,B7) well. All textures were separated into distinct percentages using supervised classification, as shown in Figure 5.15(b). Table 5.9 shows the confusion matrices for both classifications. It is worth noting that, during the unsupervised classification,

scene B6 was 100 percent recognized. In the supervised classification, the best separation of the textures corresponds to textures B2 and B7, whilst B3 and B8 were insufficiently recognized.



(a) Unsupervised classification of Brodatz mosaic. Average accuracy **99.10**

(b) Supervised classification of Brodatz mosaic. Average accuracy **83.87**



(c) Unsupervised classification of Brodatz mosaic. Average accuracy **60.69**

(d) Supervised classification of Brodatz mosaic. Average accuracy **89.51**

Figure 5.15: Brodatz mosaic classification based on the estimated texture parameters provided by (a,b) MAP-GMRF and (c,d) MAP-ABM method. MAP-GMRF is superior in the case of unsupervised classification, while MAP-ABM outperforms to MAP-GMRF in the case of supervised classification.

- **MAP-ABM results**

The MAP-ABM method was able to separate six textures (B1,B2,B3,B6,B7,B9) using unsupervised classification, as shown in Figure 5.15(c) and the confusion matrix summarized in Table 5.10, but, some classes were poorly defined. However, the supervised classification depicted in Figure 5.15(d) shows the superiority of the

MAP-ABM method, having recognized all the textures with better quality than the MAP-GMRF. The confusion matrices for both classifications are presented in Table 5.10, which shows that in the unsupervised case the better recognized class was B6, whilst scenes B2, B3, and B7 were only percent 40 recognized. In the supervised classification, all classes were separated at more than 80%, with the exception of scene B4 (78 percent).

Table 5.9: **Confusion Matrices for Brodatz textures** obtained in the case of unsupervised and supervised classifications using estimated texture parameters provided by MAP-GMRF.

(a) Unsupervised classification.					(b) Supervised classification.										
Class:	B1	B2	B6	B7	Class:	B1	B2	B3	B4	B5	B6	B7	B8	B9	
Cyan	98.99	0.12	0.89	0.00	B1	89.96	0.02	3.36	6.59	0.04	0.00	0.00	0.00	0.03	
Magenta	1.93	98.07	0.00	0.00	B2	0.12	97.48	0.69	0.03	1.64	0.00	0.00	0.00	0.04	
Blue	98.25	0.01	1.45	0.29	B3	2.91	0.23	74.74	6.54	4.97	0.00	0.00	0.01	10.59	
Purple	45.41	0.00	54.58	0.00	B4	3.08	0.00	4.00	82.39	0.02	0.38	0.00	4.58	5.55	
Red	94.65	4.90	0.03	0.43	B5	0.19	0.42	7.44	0.57	90.07	0.00	0.01	0.00	1.30	
Yellow	0.00	0.00	100.00	0.00	B6	0.00	0.00	0.00	0.30	0.00	80.64	0.00	19.06	0.00	
Maroon	0.40	0.00	0.24	99.35	B7	0.12	0.18	0.66	1.81	0.07	0.00	96.80	0.00	0.35	
Sea Green	0.28	0.00	99.72	0.00	B8	0.00	0.00	0.00	6.27	0.00	26.72	0.00	66.65	0.36	
Green	77.39	0.03	18.50	4.09	B9	0.42	0.05	13.04	9.36	0.10	0.00	0.09	0.78	76.16	
Average accuracy				99.10	Average accuracy										83.87

Table 5.10: **Confusion Matrices for Brodatz textures** obtained in the case of unsupervised and supervised classifications using estimated texture parameters provided by MAP-ABM.

(a) Unsupervised classification.							(b) Supervised classification.										
Class:	B1	B2	B3	B6	B7	B9	Class:	B1	B2	B3	B4	B5	B6	B7	B8	B9	
Magenta	82.10	13.98	2.63	0.00	0.00	0.79	B1	99.19	0.62	0.00	0.19	0.00	0.00	0.00	0.00	0.00	
Cyan	20.45	41.95	0.97	3.88	1.94	30.60	B2	0.38	97.43	0.01	2.05	0.05	0.04	0.00	0.00	0.05	
Maroon	33.44	9.52	42.56	3.88	8.63	0.95	B3	0.00	0.00	89.85	3.18	3.57	0.18	2.59	0.06	0.57	
Purple	5.63	13.65	1.27	12.32	20.50	46.24	B4	0.01	2.27	5.86	77.86	0.28	1.47	1.09	2.15	9.01	
Red	45.50	37.14	2.95	0.01	0.07	14.26	B5	0.00	0.02	1.59	0.81	95.97	0.00	0.00	0.01	1.60	
Green	0.00	0.04	0.00	91.82	0.27	7.87	B6	0.00	0.03	0.01	1.14	0.00	80.65	0.00	17.21	0.96	
Blue	3.54	5.08	28.02	2.39	37.34	0.20	B7	0.00	0.00	0.95	0.47	0.00	0.01	98.58	0.00	0.00	
Sea Green	0.00	0.12	0.00	83.51	0.38	15.99	B8	0.00	0.00	0.10	1.24	0.00	16.28	0.00	79.72	2.65	
Yellow	1.94	29.64	0.08	2.24	0.07	66.00	B9	0.00	0.21	0.93	3.72	1.25	2.83	0.00	4.72	86.34	
Average accuracy						60.29	Average accuracy										89.51

Comparing both methods, the average accuracy in the unsupervised classification shows that MAP-GMRF (99.10%) is superior than MAP-ABM (60.29%), but in the case of supervised classification MAP-ABM (89.51%) outperforms to MAP-GMRF (83.87%).

2. Stability of the estimated parameters

The mean and variances of the different texture parameters obtained using the MAP-GMRF and MAP-ABM methods are shown in Tables 5.11(a) and 5.11(b), respectively. The variance of texture parameters obtained with the GMRF method was much lower than obtained with the MAP-ABM, therefore, the texture parameters were only slightly deviated around the mean values, as shown in Figure 5.13(a)-(b). The texture parameters obtained with the MAP-ABM method were much more deviated around the mean

5.4.2.2 Classification results based on Real SAR data

In the following, the classification results using the mosaic of TerraSAR-X images and TerraSAR-X sub-scene over Hamburg are presented. The classification results will be discussed in terms of 1) confusion matrices and 2) the stability of the estimated parameters will be analyzed by computing the means and variances. It is worth noting that the confusion matrix presented in the case of unsupervised classifications shows only the recognized classes.

1. Mosaic of TerraSAR-X images

The mosaic of TerraSAR-X images depicted in Figure 5.5 shows different texture scenes. The texture parameters of this mosaic were extracted and later, classified. In the following the results are presented.

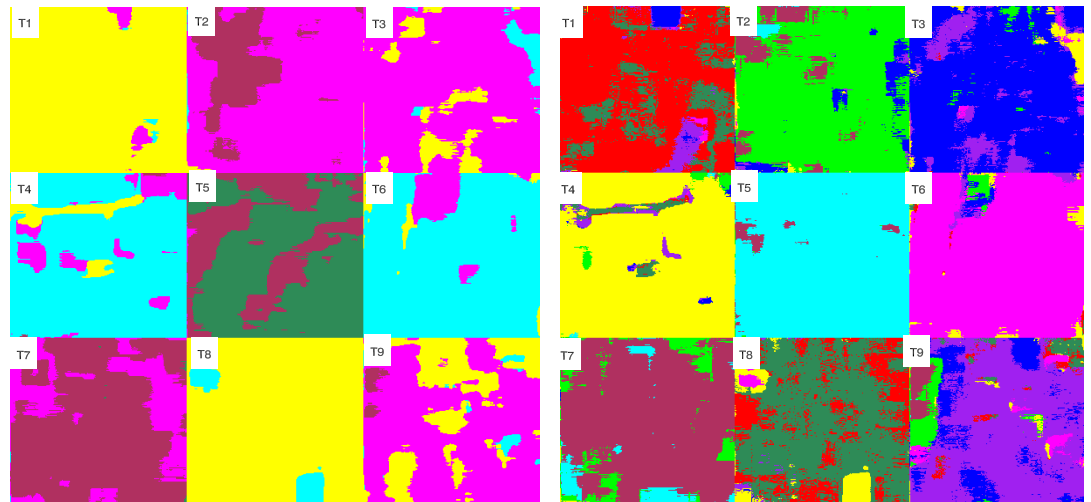
- **MAP-GMRF results**

The unsupervised and supervised classifications results using the estimated texture parameters provided by MAP-GMRF, are presented in Figures 5.16(a)-(b), respectively. According to the confusion matrix of the unsupervised classification summarized in Table 5.12, this method was able to well-recognize four classes (T1, T4, T5, and T7). However, in scene T2, the small houses were well-recognized and some buildings were separated in the upper part. These kinds of buildings can be found in scene T7. T7 was in fact recognized by the great majority. Scene T3 was confused with scene T9, which represented different kinds of vegetation. Scenes T4 and T6 had two different types of agricultural fields, but, the method placed all fields into one class. The confusion matrix presented in Table 5.12 explains the goodness of the extracted features. Scene T5 is separated, with 80 percent as a new class. Scenes T3, T6, T8 were unrecognized. The results for supervised classification show a lower quality compared with unsupervised classification, for example, image T1 was only recognized in 73 percent and some confusion with other classes was introduced. However, scenes T5 and T9 were better recognized, scene T5 was classified in 97 percent, and scene T9 in 66 percent. In addition scenes T3, T6, and T8 were also recognized.

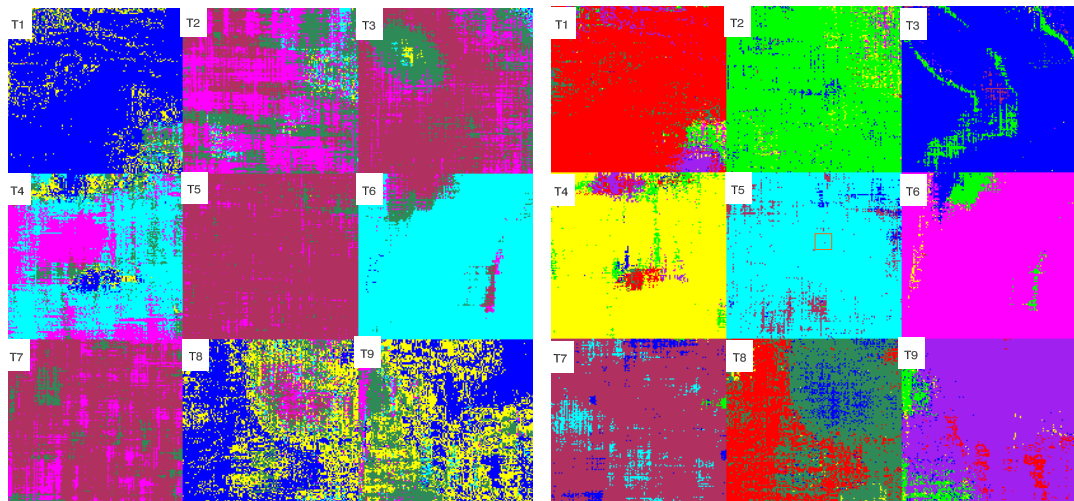
- **MAP-ABM results**

Figures 5.16(c)-(d) show the unsupervised and supervised classifications using the estimated parameters provided by the MAP-ABM. The unsupervised classification shows lower quality compared to the unsupervised classification obtained using the estimated parameters of the MAP-GMRF method. The MAP-ABM method separated scene T1, scene T5, and scene T6 well, but it also recognized scene T2, T3, and T9 with more than 30 percent. However, the supervised classification obtained using the MAP-ABM was superior. The confusion matrix, presented in Table 5.13 shows the results for classification. Eight textures were separated with more than 85 percent accuracy, and only scene T8 was classified with only 55 percent.

Comparing both methods, the results are similar that ones obtained with Brodatz mosaic. The average accuracy in unsupervised classification shows the superiority of MAP-GMRF (76.96% vs 61.17%), while in the supervised case MAP-ABM (85.20%) outperforms MAP-GMRF (81.30%).



(a) Unsupervised classification of TerraSAR-X mosaic. Average accuracy **76.96**
 (b) Supervised classification of TerraSAR-X mosaic. Average accuracy **81.30**



(c) Unsupervised classification of TerraSAR-X mosaic. Average accuracy **61.17**.
 (d) Supervised classification of TerraSAR-X mosaic. Average accuracy **85.20**.

Figure 5.16: TerraSAR-X mosaic classification based on the estimated texture parameters provided by (a,b) MAP-GMRF method and (c,d) MAP-ABM method. MAP-GMRF is superior in the case of unsupervised classification, while MAP-ABM provides better results in the case of supervised classification.

- **Stability of the estimated parameters**

The result of the TerraSAR-X mosaic classification can be explained using the mean and variances of the texture parameters as shown in Tables 5.14(a) and 5.14(b), obtained using the MAP-GMRF and MAP-ABM methods, respectively. In these tables, the texture parameters are represented by their mean value. Figures 5.13(c)-(d) depict examples of the first texture parameter. The texture parameters of the MAP-ABM had much higher variance and were deviated around the mean value. Therefore, K-means classifier has more difficulties separating the classes.

Table 5.12: **Confusion Matrices for TerraSAR-X textures** obtained in the case of unsupervised and supervised classifications using estimated texture parameters provided by MAP-GMRF.

(a) Unsupervised classification.						(b) Supervised classification.										
Class:	T1	T2	T4	T5	T7	Class:	T1	T2	T3	T4	T5	T6	T7	T8	T9	
Yellow	95.57	4.12	0.31	0.00	0.00	T1	73.19	0.00	2.17	0.43	0.00	0.57	0.00	19.16	4.48	
Magenta	0.35	29.12	0.43	0.00	70.10	T2	0.00	81.14	5.76	0.84	1.39	0.04	9.94	0.01	0.88	
Blue	4.22	93.02	2.77	0.00	0.00	T3	0.07	1.58	80.36	3.46	0.00	0.92	0.00	0.00	13.60	
Cyan	6.86	7.84	83.37	0.00	1.92	T4	0.44	1.78	0.70	90.64	0.22	0.27	0.09	2.92	2.94	
Sea Green	0.00	0.11	0.00	79.85	20.04	T5	0.00	0.19	0.00	0.07	97.30	0.00	2.44	0.00	0.00	
Red	2.97	10.16	86.86	0.00	0.00	T6	0.39	1.56	1.67	3.59	0.00	90.68	0.00	0.53	1.58	
Maroon	0.00	2.13	0.06	0.90	96.91	T7	0.00	9.65	0.08	0.40	6.19	0.00	83.67	0.00	0.01	
Green	95.21	0.00	4.78	0.00	0.00	T8	23.28	0.00	0.09	6.29	0.00	0.87	0.00	68.66	0.81	
Purple	17.20	76.08	1.43	0.00	5.28	T9	2.09	5.81	15.22	1.48	0.02	2.98	3.34	2.96	66.11	
Average accuracy					76.96	Average accuracy										81.30

Table 5.13: **Confusion Matrices for TerraSAR-X textures** obtained in the case of unsupervised and supervised classifications using estimated texture parameters provided by MAP-ABM.

(a) Unsupervised classification.							(b) Supervised classification.										
Class:	T1	T2	T3	T5	T6	T9	Class:	T1	T2	T3	T4	T5	T6	T7	T8	T9	
Blue	88.63	0.49	3.58	0.26	1.18	5.87	T1	85.80	2.64	0.20	0.93	0.00	0.53	0.01	3.15	6.75	
Magenta	0.14	36.50	38.10	18.87	3.46	2.94	T2	0.16	91.64	3.12	3.04	0.00	0.52	0.09	0.00	1.42	
Sea Green	0.12	8.74	34.60	54.90	0.63	1.01	T3	0.00	6.98	91.30	0.18	0.01	0.01	1.02	0.20	0.28	
Red	5.81	30.74	7.84	0.75	48.33	6.53	T4	2.94	5.35	0.65	87.00	0.00	0.17	0.00	0.96	2.94	
Maroon	0.00	2.76	4.40	92.80	0.00	0.04	T5	0.00	0.03	1.87	0.00	89.81	0.00	8.30	0.00	0.00	
Cyan	0.11	1.35	6.25	3.09	88.62	0.58	T6	0.03	4.13	2.33	1.04	0.00	92.46	0.00	0.01	0.00	
Green	0.00	12.05	17.88	70.07	0.00	0.00	T7	0.00	0.39	3.74	0.03	7.42	0.00	88.42	0.00	0.00	
Purple	53.85	4.25	10.82	3.46	2.17	25.45	T8	37.85	0.07	6.31	1.12	0.00	0.02	0.00	54.61	0.02	
Yellow	43.34	3.37	18.12	0.51	2.78	31.88	T9	3.83	6.39	2.72	1.08	0.00	0.01	0.15	0.03	85.80	
Average accuracy						61.17	Average accuracy										85.20

2. TerraSAR-X sub-scene over Hamburg

This study case refers to a sub-scene over Hamburg in Northern Germany with 2000×2000 pixels size (see Figure 5.6). Both methods for despeckling and information extraction were applied over this image. The goal is to recognize urban areas from the sub-scene and obtain a land use classification.

In order to evaluate the efficiency of the extracted features, an unsupervised classification was performed, using the K-means algorithm with 5 classes. The supervised classification was performed using the maximum likelihood method by selecting some regions of interest. The proposed classes were: water (blue), high buildings (cyan), small building (yellow), vegetation (green).

Figures 5.17 and 5.18 depict the classification results provided by both methods. Here, it should be noted that, according to the unsupervised classification results, both methods well-separated the high buildings class. The MAP-GMRF recognizes small-buildings and forest better than the MAP-ABM. The MAP-ABM well recognizes the water class. The supervised classification shows the superiority of the MAP-ABM because it recognizes well all the classes, for example, in the right upper part of the image there is a small lake, which was correctly classified as water between the vegetation class, which was also well-defined. However, in the MAP-GMRF supervised classification, the vegetation

Table 5.14: Mean and variance values of texture parameters for each tile (T1-T9) in the case of TerraSAR-X mosaic using MAP-GMRF and MAP-ABM methods.

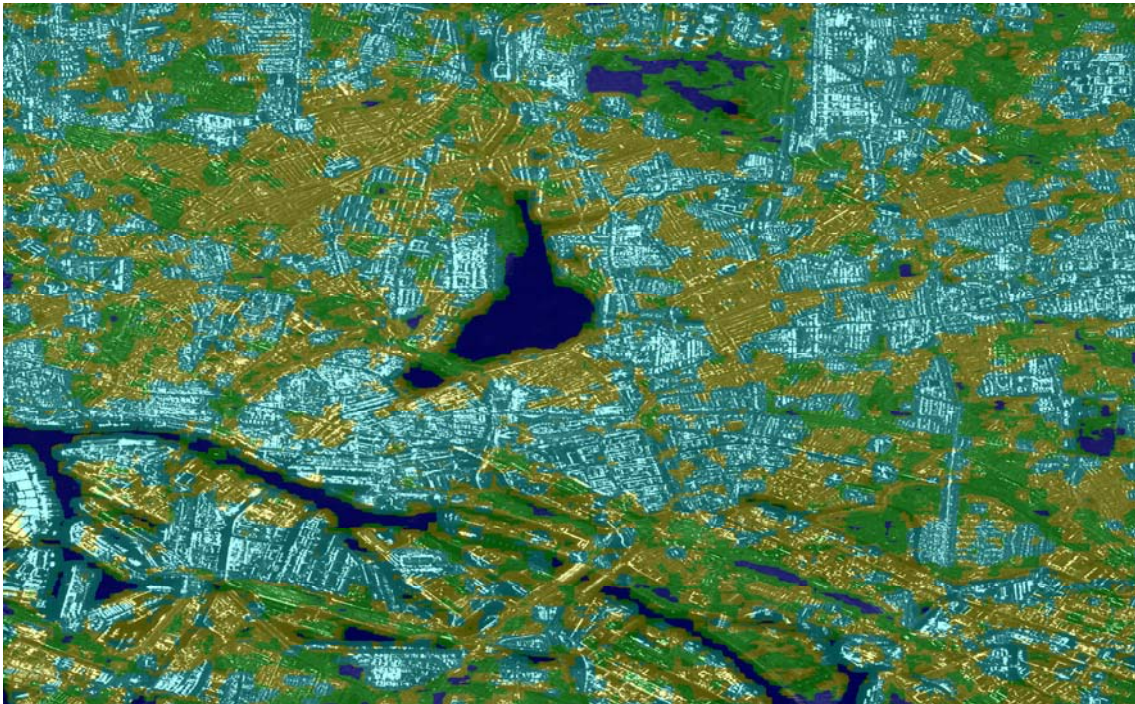
(a) MAP-GMRF

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
Mean										
T1	0.12	0.09	0.07	0.07	0.07	0.00	0.03	-0.00	-0.00	0.04
T2	0.27	0.28	0.04	0.15	-0.05	-0.03	-0.04	-0.06	-0.03	-0.02
T3	0.23	0.20	0.09	0.13	-0.00	-0.02	-0.01	-0.04	-0.03	-0.04
T4	0.11	0.31	0.05	0.03	0.01	0.10	0.03	-0.08	-0.00	-0.05
T5	0.44	0.28	-0.14	0.10	-0.04	-0.03	-0.09	-0.04	0.01	0.00
T6	0.11	0.19	0.04	0.12	-0.02	0.08	-0.02	0.06	-0.02	-0.03
T7	0.33	0.28	0.01	0.14	-0.06	-0.05	-0.06	-0.04	-0.02	-0.00
T8	0.11	0.09	0.06	0.08	0.06	0.01	0.05	0.01	-0.00	0.02
T9	0.22	0.19	0.08	0.13	0.01	-0.02	-0.01	-0.04	-0.03	-0.04
Variance										
T1	0.00129	0.00048	0.00024	0.00031	0.00057	0.00070	0.00049	0.00082	0.00053	0.00075
T2	0.00068	0.00185	0.00202	0.00083	0.00064	0.00043	0.00031	0.00048	0.00052	0.00069
T3	0.00093	0.00106	0.00033	0.00031	0.00090	0.00050	0.00068	0.00045	0.00025	0.00033
T4	0.00454	0.00746	0.00073	0.00241	0.00060	0.00291	0.00071	0.00268	0.00142	0.00091
T5	0.00633	0.00426	0.00261	0.00127	0.00088	0.00052	0.00096	0.00102	0.00095	0.00056
T6	0.00112	0.00072	0.00057	0.00030	0.00011	0.00090	0.00036	0.00089	0.00057	0.00030
T7	0.00145	0.00148	0.00132	0.00090	0.00061	0.00056	0.00034	0.00055	0.00053	0.00055
T8	0.00092	0.00045	0.00014	0.00021	0.00096	0.00104	0.00067	0.00062	0.00054	0.00045
T9	0.00166	0.00196	0.00043	0.00037	0.00100	0.00043	0.00103	0.00060	0.00022	0.00052

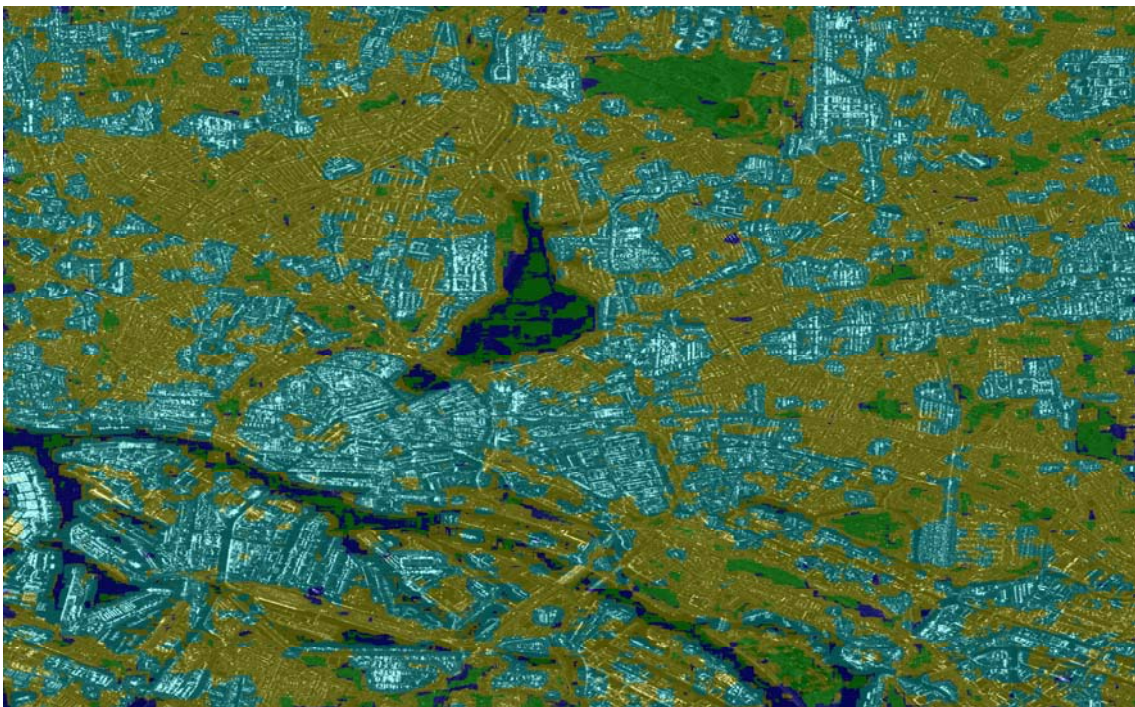
(b) MAP-ABM

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}
Mean										
T1	0.47	0.27	0.14	0.17	0.12	0.13	0.16	0.17	0.17	0.22
T2	1.83	1.40	-0.65	0.15	-0.55	-0.24	-0.12	-0.10	0.09	0.23
T3	1.90	1.23	-0.47	0.28	-0.57	-0.36	-0.15	-0.11	0.13	0.15
T4	1.16	1.22	-0.45	-0.08	-0.27	0.27	0.01	-0.05	0.21	0.05
T5	1.85	1.15	-0.42	0.53	-0.27	-0.28	-0.32	-0.15	0.06	0.04
T6	1.05	0.75	-0.34	0.25	-0.31	0.40	-0.08	0.31	0.08	-0.05
T7	1.90	1.31	-0.50	0.41	-0.50	-0.30	-0.26	-0.12	0.02	0.20
T8	1.00	0.49	-0.01	0.10	-0.07	0.01	0.07	0.10	0.12	0.15
T9	1.20	0.50	-0.03	0.08	-0.13	-0.03	0.07	0.10	0.15	0.20
Variance										
T1	0.23	0.04	0.03	0.01	0.03	0.02	0.00	0.00	0.00	0.01
T2	0.11	0.14	0.06	0.05	0.04	0.03	0.01	0.03	0.01	0.01
T3	0.05	0.06	0.04	0.03	0.03	0.02	0.02	0.01	0.01	0.01
T4	0.20	0.25	0.08	0.02	0.05	0.05	0.02	0.05	0.02	0.01
T5	0.04	0.05	0.03	0.02	0.02	0.02	0.01	0.01	0.01	0.01
T6	0.13	0.04	0.05	0.01	0.02	0.03	0.00	0.02	0.00	0.02
T7	0.05	0.06	0.04	0.01	0.02	0.01	0.01	0.01	0.01	0.01
T8	0.45	0.17	0.07	0.02	0.07	0.04	0.02	0.01	0.01	0.01
T9	0.61	0.18	0.12	0.04	0.11	0.05	0.02	0.01	0.01	0.01

and water are mixed.

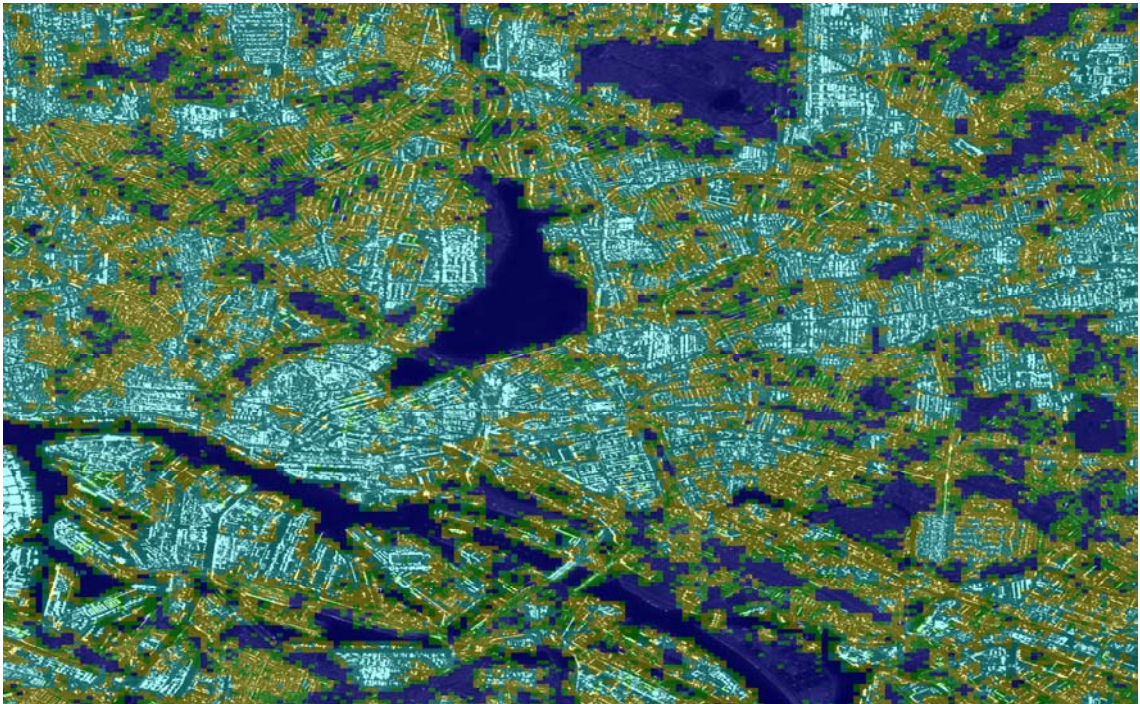


(a) Unsupervised classification of TerraSAR-X Hamburg scene using GMRF estimated parameters.

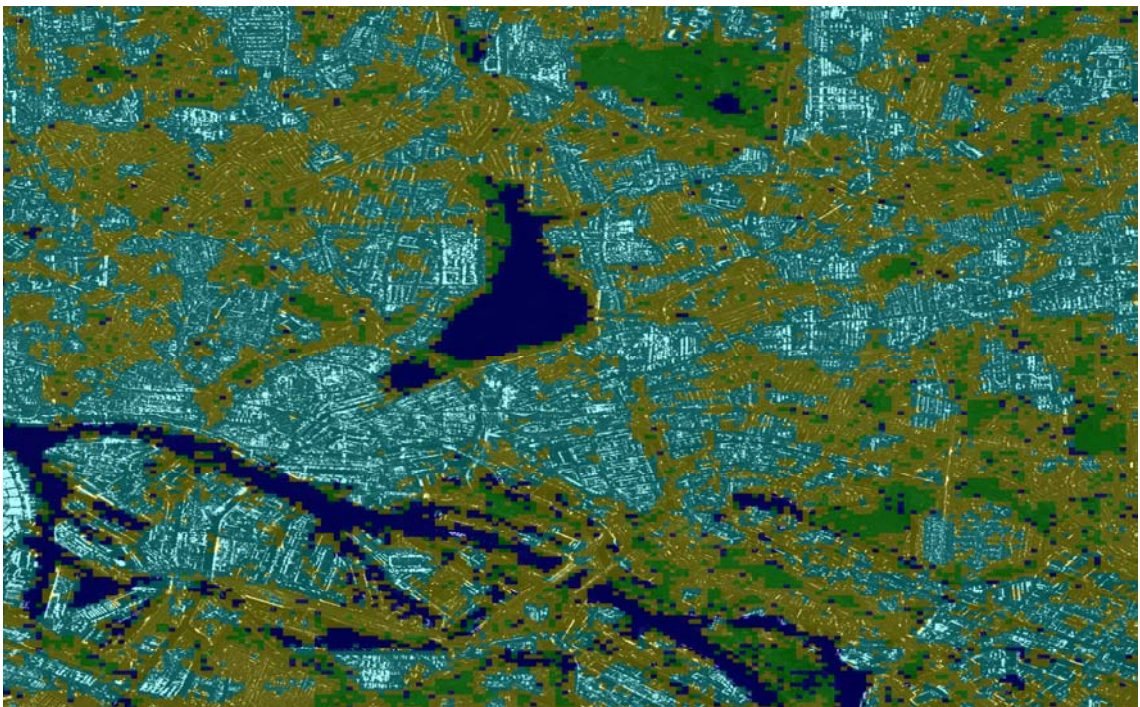


(b) Supervised classification of TerraSAR-X Hamburg scene using GMRF estimated parameters.

Figure 5.17: TerraSAR-X Hamburg classifications using the estimated parameter provided by MAP-GMRF. (a) Unsupervised classification. (b) Supervised classification. The classes and their associated colors are: water (blue), high buildings (cyan), small buildings (yellow), vegetation (green). Urban area, small and high buildings, are well recognized. There is confusion between river and forest.



(a) Unsupervised classification of TerraSAR-X Hamburg scene using ABM estimated parameters.



(b) Supervised classification of TerraSAR-X Hamburg scene using ABM estimated parameters.

Figure 5.18: TerraSAR-X Hamburg classifications using the estimated parameter provided by MAP-ABM. (a) Unsupervised classification. (b) Supervised classification. The classes and their associated colors are: water (blue), high buildings (cyan), small buildings (yellow), vegetation (green). Urban area, small and high buildings, are well recognized as well as river and vegetation.

5.4.3 The impact of window size and model order on feature extraction

The selection of the appropriate model order and estimation window size is determined by the complexity of the scene to be classified. We can guess that a larger extent of the analyzing window will lead to a higher accuracy of the estimator, that being a consequence of using more data for the computation. This reasoning is valid in the assumption of homogeneous images. However, the small details and structures present in high resolution SAR images need to be preserved. The use of higher model order affects the capability to capture high structural complexity in the images, i.e., high-complexity models are defined on large neighborhoods, which implies the use of large analyzing windows. Then, a good compromise between analyzing window and model order can improve the classification results. However, they depend on what the user expects to have either the greater number of classes or the higher accuracy in the classification.

In this context, the impact of the estimation window and the model order on feature extraction is discussed using model order from two to six and an estimation window from (4×4) to (64×64) pixel size. We start by presenting the results of Figure 5.19, in here, 1) the visual inspection is used to evaluate the feature extraction and gives an idea about how the estimation window size and the model order affect the classification results. Later, 2) the performance of the classifications is discussed in terms of the accuracy and the number of recognized classes. These results are presented in Figure 5.20 and Fig. 5.21 in the case of MAP-GMRF and MAP-ABM, respectively.

1. Visual inspection

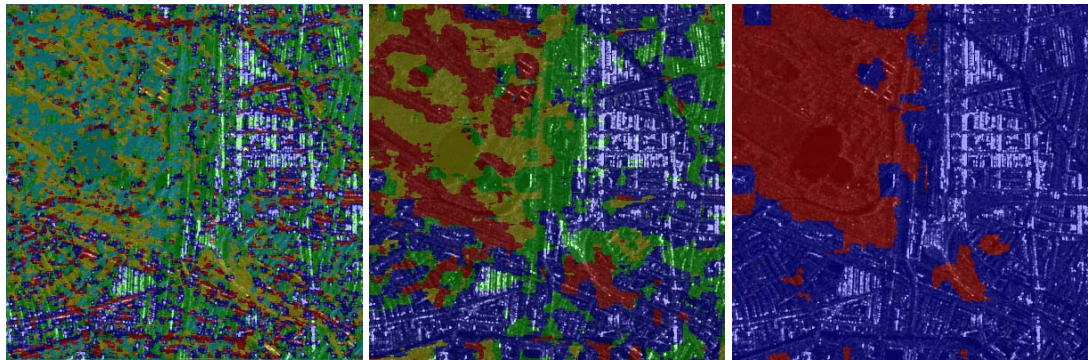
We chose a sub-image of the TerraSAR-X Hamburg scene to extract the texture primitive features and to demonstrate how the different window sizes and model orders have an affect on the unsupervised classification results. The results are summarized in Figure 5.19. Here, the classification results were overlapped with the original TerraSAR-X scene in order to ease the visual inspection.

- **MAP-GMRF results**

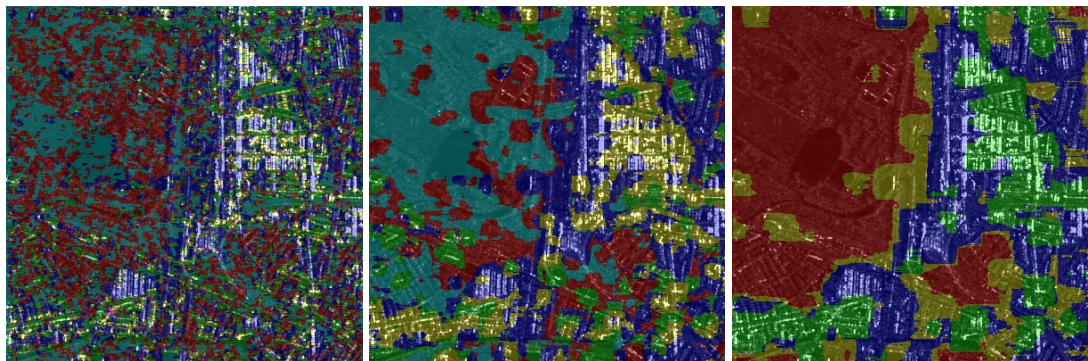
With small window sizes 4×4 pixels, MAP-GMRF models only the local statistics of the SAR image, therefore, the textures are not well separated. The MAP-GMRF method with window size 16×16 separates textures such as city and forest for both model orders 2 and 4. The MAP-GMRF with model order 2 and window size 32×32 distinguishes two classes: homogeneous in terms of structures (e.g vegetation) and heterogeneous (e.g buildings), as shown in Figure 5.19(a). MAP-GMRF with model order 4 and window size 32×32 pixels separated forest area from the city, and further found three classes within heterogeneous areas, as shown in Figure 5.19(a).

- **MAP-ABM results**

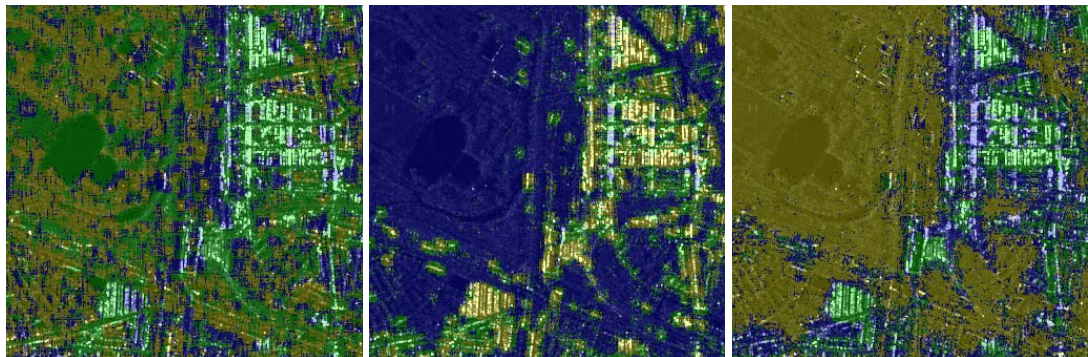
MAP-ABM using small window sizes 4×4 pixels models only the local statistics of the SAR image, therefore, the textures were badly separated. The MAP-ABM classification well-separated textures from homogeneous areas (e.g vegetation and water), as shown in Figures 5.19(c) and 5.19(d). The higher the window size gets more the textures that can be separated, however, with a window size too big the small details can be blurred. The homogeneous in terms of structures and heterogeneous areas are well discriminated for model order 4 and 2 with a window size of 32×32 pixels.



(a) Unsupervised classification based on MAP-GMRF texture parameters using Model Order 2.



(b) Unsupervised classification based on MAP-GMRF texture parameters using Model Order 4.



(c) Unsupervised classification based on MAP-ABM texture parameters using Model Order 2.



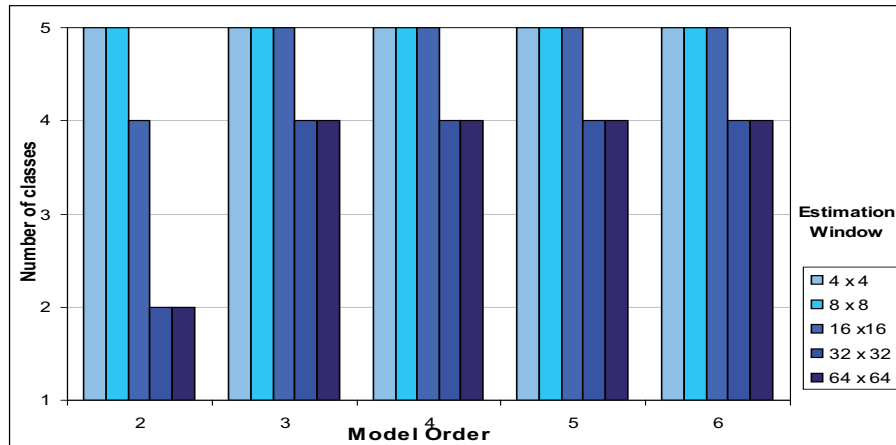
(d) Unsupervised classification based on MAP-ABM texture parameters using Model Order 4.

Figure 5.19: Unsupervised classification using the estimated texture parameters provided by (a-b) MAP-GMRF and (c-d) MAP-ABM method with model order 2 and 4, and estimation window of (from left to right) 4×4 , 16×16 , and 32×32 pixels, respectively.

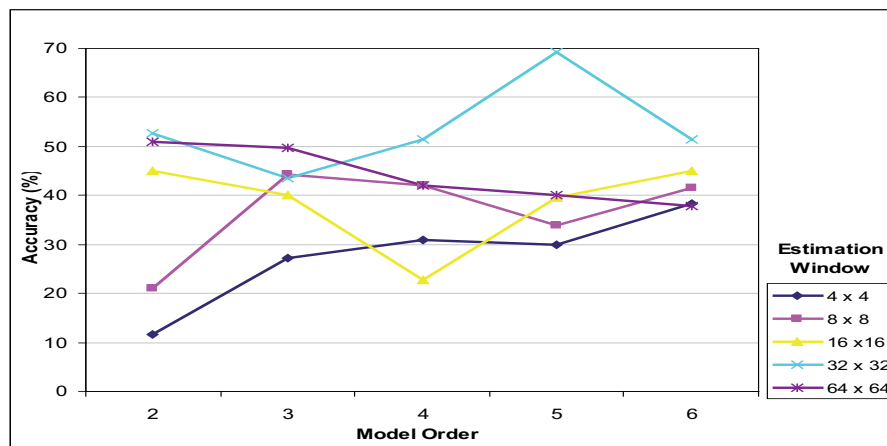
2. Performance of the classification

A resume of the classification performances versus the model order and the estimation window size is presented in Figure 5.20 in the case of MAP-GMRF and Figure 5.21 in the case of MAP-ABM. Both figures show the number of recognized classes and their accuracy. The accuracy was computed from the confusion matrix obtained by comparing the ground-truth against the distinguished class.

• MAP-GMRF results



(a) Number of recognized classes versus the model order and the estimation window using MAP-GMRF.



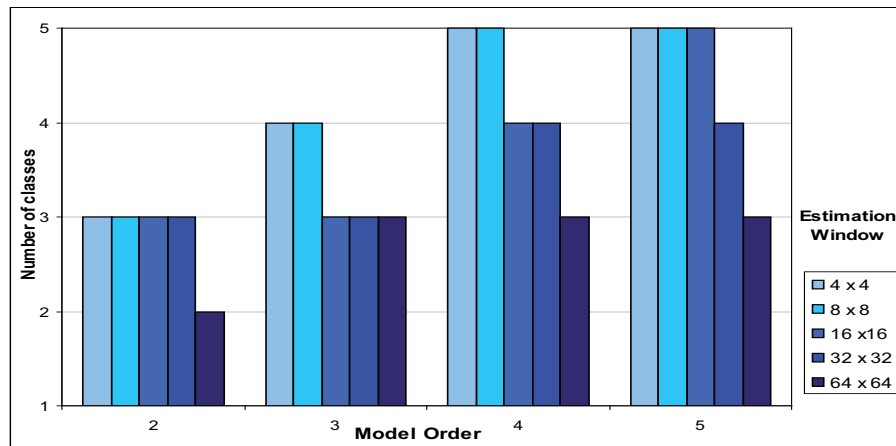
(b) Accuracy of the recognized classes versus the model order and the estimation window using MAP-GMRF.

Figure 5.20: Summary of the classification using the estimated texture parameters provided by MAP-GMRF with model order between 2 and 6 and the size of the estimation window ranging from 4×4 , 8×8 , 16×16 , 32×32 and 64×64 pixels. (a) Number of recognized classes as function of model order and estimation window. (b) Overall accuracy of the classification versus the model order and estimation window.

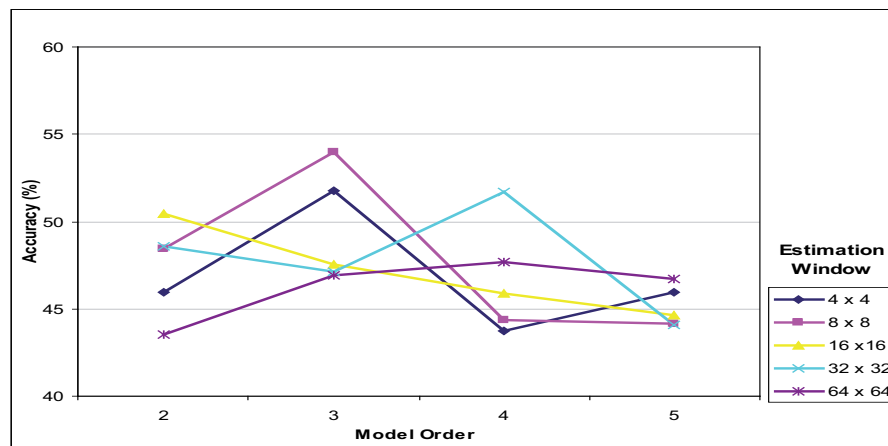
In the case of MAP-GMRF, in Figure 5.20(a), we can observe that using all model orders it is possible to recognized all the classes (5) with different degrees of accuracy. In the case of model order 2, the number of recognized classes decreases when

the size of the estimation window increases, but the accuracy increases, therefore, the two distinguished classes are better defined (homogeneous and heterogeneous). In the case of model order three and higher, the classifications show that the minimum of recognized classes is four, but their accuracy differs significantly. In Figure 5.20(b), the overall classification accuracy is depicted. In here, the highest accurate data belongs to model order 5 with an estimation window of 32×32 pixels. The lower accuracy is obtained using model order 2. With model order higher than 3 it is recommend to use an estimation window higher than 16×16 pixels size. It is worth noting that an estimation window of 64×64 decreases the accuracy in the classification, this is because the small details could not be well estimated.

• MAP-ABM results



(a) Number of recognized classes versus the model order and the estimation window using MAP-ABM.



(b) Accuracy of the classification

Figure 5.21: Summary of the classification using the estimated texture parameters provided by MAP-ABM with model order between 2 and 5 and the size of the estimation window ranging from 4×4 , 8×8 , 16×16 , 32×32 and 64×64 pixels. (a) Number of recognized classes as function of model order and estimation window. (b) Overall accuracy of the classification versus the model order and estimation window.

In the case of MAP-ABM, Figure 5.21(a) shows that using model order two it is possible to recognize three classes. Better results are achieved using model order higher than 2. The five classes are recognized using model order 4 and 5 and estimation windows size between 4×4 and 32×32 . In Figure 5.21(b), the overall classification accuracy is depicted. In here, we can note that an estimation window of 64×64 decreases the accuracy in the classification, because of small details could not be well estimated. The best compromise between number of recognized classes and accuracy is obtained using model order 4 and estimation windows 32×32 pixels.

5.4.4 Conclusions

From the information extraction section, we could make the following conclusions

- The MAP-ABM method well separates homogeneous and heterogeneous areas, and well distinguishes different textures. It well preserves strong scatterers, therefore, the MAP-ABM also models textures within the homogeneous areas.
- The MAP-GMRF well models homogeneous and heterogeneous areas, but is unable to model the point-like characteristics of images compared to MAP-ABM. The MAP-GMRF groups textures together, because of the larger window sizes capture those strong scatterers that do not belong to the homogeneous areas but are included because of the window size.
- The feature extraction from real SAR images using texture models is presented using the GMRF and ABM models. Practical use of the models shows that the GMRF model groups areas with strong scatterers, and is able to distinguish homogeneous from heterogeneous areas. The ABM model is also able to separate homogeneous, and heterogeneous area but is also able to model different textures with strong scatterers. The MAP-ABM is more appropriate for use with real SAR images.

5.5 Conclusions

This chapter presents evaluations of Bayesian despeckling and texture extraction methods based on Gauss Markov and Auto-binomial Gibbs Random Fields, using the application to TerraSAR-X data. The performances of the MAP-GMRF and MAP-ABM methods were assessed in terms of the quality of the despeckled image and the quality of the feature texture parameter' extractions provided by both methods. In the case of evaluating despeckling quality, we compared the degree of smoothness and structure preservation in the images and used some quantitative measurements, such as mean-squared error as a fidelity measurement, the bias and the ratio, the smoothness, the sharpness criterion, structural similarity index, etc. The estimated texture parameters of the GMRF and ABM models were analyzed using simulated and real SAR images. The accuracies of the model parameters were analyzed together with supervised and unsupervised classifications, where visual impressions and the confusion matrix have allowed us to formulate our comments about the results.

1. MAP-GMRF outperforms MAP-ABM

- The MAP-GMRF method estimates textures well, the mean value of which does not change dynamically. The synthetic case shows that the MAP-GMRF can introduce bias, if the analyzing window does not correspond to the structures in the scene.
- The GMRF model is more appropriate for modeling natural scenes (i.e. vegetation). Moreover, the experimental results showed that the GMRF texture parameters are more suitable for unsupervised classification case, because they are more stable than ABM ones, which deviate more around their mean values. This could be explained by the fact that the image modelled with ABM has a maximum value G two times higher than the mean value of the image, which is not the case of SAR images. Therefore the texture parameters deviate much more around the mean value.
- The GMRF model is also less complex, because MAP estimates can be evaluated analytically, whilst ABM requires numerical computation. The MAP-GMRF method requires 3-5 times less computationally demanding than MAP-ABM.

2. MAP-ABM outperforms MAP-GMRF

- The despeckling results of real SAR images obtained with ABM are much better than the ones obtained with GMRF. The experimental results showed that the ABM outperforms the GMRF model when computing objective measures as MSE (fidelity), smoothness, etc on synthetic data sets. The GMRF model estimates well textures, when the size of analyzing window is very large, whilst a good performance of the ABM model requires smaller window sizes.
- The ABM estimates better blob-like textures, whilst the GMRF model is more suitable for natural textures. The ABM model is also able to estimate larger sets of textures than the GMRF model. The supervised classification results showed the superiority of the ABM model for extracting textures from both synthetic and real SAR images.

3. Impact of model order and estimation windows on the performance of the methods

An experiment was carried out using different window sizes and model orders for the same image. The despeckling and classification was observed and we found out that by increasing the window size, the information in the image is better describe, but it depends on complexity of the scene. By increasing the model order the reconstructed images are becoming sharper and information extraction can better separate textures and heterogeneous areas from homogeneous areas. The model order 4 with window size of 32×32 pixels is a good compromise between the quality of despeckling, texture extractions, computation time and the complexity of the methods.

4. Overall recommendation

The MAP-GMRF and MAP-ABM methods do not require human interactions for estimating texture parameters and despeckling. The MAP-ABM outperforms the MAP-GMRF in despeckling and feature extraction for real SAR images, therefore, we recommend to use the MAP-ABM method for despeckling and feature extraction.

Chapter 6

User-driven evaluation of Image Information Mining systems

Image Information Mining (IIM) and Content-Based Image Retrieval (CBIR) systems were presented in chapter 3. Most of these systems have functions to extract relevant information from the data, to generate a content-index and to perform the search process in an interactive way based on a graphical human-computer interface. This leads us to group their functions in two parts, an objective part since it involves feature extraction and classification methods performed automatically by handling the data, and a subjective part since it involves the human-machine communication. It was also pointed out in chapter 3 that the overall evaluation of IIM systems could be divided in the evaluation of their components (Smeulders et al., 2000), which does concern the data-driven evaluation (assessment of feature extraction methods), and the user-driven evaluation (assessment of the probabilistic search) involving the user and his conjectures in the evaluation process. The **data-driven evaluation** of suitable feature extraction methods for SAR images was performed in chapter 5.

The present chapter will conclude the overall evaluation by presenting the **user-driven evaluation**, which involves the creation of validation scenarios by the user as study cases. It is worth mentioning that the data-driven evaluation of information extraction methods provides us with important recommendations such as appropriated parameters according to the scene, optimal analyzing window, and model order; which are taken into account during the creation of the scenarios.

In the following, we start by presenting the Knowledge-based Image Information Mining system, called **KIM** in this thesis, as an IIM tool, which serves us as a prototype to perform this evaluation. The main theoretical concepts, its architecture, as well as an example of operation are presented in section 6.1. The **effectiveness** of KIM retrieving images based on their content is matter of this evaluation, therefore we present a review of the metrics used in evaluating retrieval systems in section 6.2. The **subjective** nature of this evaluation is justified in section 6.3. It involves the interaction of real users and provides a feedback about their satisfaction degree with the retrieved results. Therefore, KIM requires a methodology for evaluating its effectiveness. Moreover, since the user is the main actor in using KIM system, we need to define the user roles either as interpreter or as evaluator, as well as a standard scheme for annotating images, and to create the ground-truth. The proposed methodology is presented in section 6.3. Finally, with the collaboration of the user, two examples of scenarios in a disaster monitoring framework,

the oil spill in Gulf of Mexico and the Nepal flooding, using KIM and TerraSAR-X images are presented in section 6.4. Here, each scenario is configured taking into account the recommendation provided in chapter 5. Section 6.5 provides some conclusions. It is worth mentioning that more validation scenarios were created during the development of this work, which are summarized in appendix C.

6.1 Knowledge-based Image Information Mining system as an Image Information Mining tool

The Knowledge-based Image Information Mining (KIM) theoretical concept and first proto-type system have been developed by the Swiss Federal Institute of Technology ETH Zurich and the German Aerospace Center (DLR) from 1996 to 2000. The results have been published in (Datcu et al., 2003).

KIM allows us to manage and explore large volumes of remote sensing images with high diversity of the content that are in general stored into a huge database. Therefore, KIM requires of robust information extraction and mining methods. It is designed to operate on large archives of remote sensing data in a non-application specific concept. This means that it could be adapted for several requirements depending on the user interests. Nowadays, KIM is integrated with the TerraSAR-X ground segment.

In the following, we start by presenting the theoretical concepts behind KIM in section 6.1.1, as well as its architecture from the technical point of view in section 6.1.2. Then an example of configuration and working with KIM is presented in section 6.1.3.

6.1.1 KIM Concept

The concept applied in KIM aims at building a system free from application specificity, in order to enable its open use in almost any scenario using remote sensing images, and also to accommodate new scenarios required by the development of new sensor technology or growing user expertise. These goals are reached by defining a **hierarchy of information representation**, as presented in Figure 6.1, enabling the communication between the image archive and the users. The **concept of information representation on hierarchical levels** of different semantic abstraction is based on the four-level Bayesian learning model (Schröder et al., 2000), where the levels are the following:

- **Level 0: Data level**

This level corresponds to the image data (D). Here several remote sensing images from different satellites (e.g. Landsat, Ikonos, TerraSAR-X, etc) are stored in order to be processed in the next levels. This thesis focuses on the use of SAR images, specifically TerraSAR-X (see section 2.4), therefore the KIM system models and reconstructs the backscattered image that is speckle-free using stochastic models, which are implemented in the next level.

- **Level 1: Feature space**

The feature space extracts the primitive image features from the image data using several stochastic signal models (M), these models are given as parametric data models $p(D|\theta, M)$ and assign the probability to a given realization of the image data D for a particular value of the parameter vector θ .

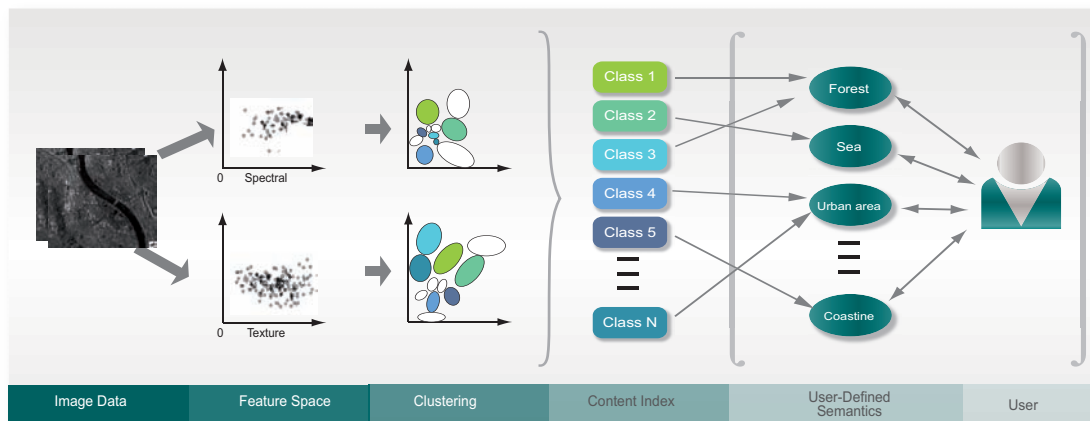


Figure 6.1: Image Information Mining represented as hierarchical modeling of the data content. In the feature space, the primitive features are extracted from the data. Later, this primitive features are grouped in different clusters, which form the content index. The user defines the semantic labels linking between the user interest (semantic labels) and the generic classes.

The feature extraction inside the KIM follows the Bayesian approach by performing the information extraction from remote sensing images through the selection of the prior model that best explains the image content (structures, features, etc.) and the estimation of the Maximum *a posteriori*, resulting in **image primitive features that characterize the image content**.

The implemented feature extraction methods for SAR images were already explained and evaluated in chapter 5.

- **Level 2: Clustering**

From level 1 the content of remote sensing images can be extracted by parametric models, which produce a large amount of data (primitive features) that can not be handled in practice, therefore, the estimated images parameters must be reduced. This task is performed through clustering methods in level 2. These methods group the different primitive features by their common characteristics/signatures giving as result a **set of unsupervised classes or clusters, which have no direct semantic meaning**. In this work, the group of classes is referred as the **“generic content index”**.

- **Level 3: User-defined semantics**

Finally, the user specifies his interests by creating semantic labels (L) and assigning them to the unsupervised classes resulted from level 2. Thus, **the link between the generic content index and the user interests is created**.

From an implementation point of view, the KIM system relies on the **processing chain**, which starts with the extraction of primitive features (Level 1) using the methods described in section 5.1 in the case of SAR images. The result consists of a best estimated despeckled image together with its texture parameters. Then, the resulting features are grouped by similarity using a K-means clustering algorithm (Level 2). The

clusters have no direct meaning, since they group points in an n-dimensional space of non-commensurable variables. However, they represent characteristics of the image seen as a multi-dimensional signal. It is possible to associate meaning with these clusters through training. A user can tell the system that a specific and weighted combination of clusters represents a feature derived from the data. By making this association, it is possible to select all images in the database that have that specific combination and may therefore contain the feature that the user is searching for.

Levels 0 to 2 constitute a complete unsupervised characterization of the image and they are performed off-line, while **level 3 consists of a supervised learning performed by the user**. In this supervised learning, the KIM system works with an active learning principle based on positive and negative training examples given by the user and indications of learning. During the semantic definition, the user creates a link between the searched image content with generic classes, assigning meaning to the information and storing it as knowledge. This link is created using Bayesian networks (Schröder et al., 2000).

Later, the user can query the database for relevant images and obtain a **probabilistic classification** of the entire image archive as an intuitive information representation based on the stochastic link. Moreover, the system provides a **posterior map classification** based on the combination of different primitive features and their associated semantic labels. It gives the user a feedback of how strong and accurate the labels are defined. The user interaction with the system provides a subjective factor to take into account in the system performance.

In the following section we describe the KIM system with its main components from a technical point of view.

6.1.2 KIM Architecture

The KIM system integrates concepts of Content-Based Image Retrieval (CBIR) (Smeulders et al., 2000), Knowledge Database Discovery (KDD), and scene understanding into a complete and automatic data processing chain provided with a Graphical User Interface (GUI) allowing the users to search for images based on their content (Datcu et al., 2003), where a successful image retrieval means likely a genuine image understanding. The KIM system was developed on a web platform in order to support the human-machine interaction via internet.

A scheme of the system architecture is presented in Figure 6.2. Here, the KIM system is composed of the server-side and the client-side. **In the server-side, the system relies on a processing chain** providing access to relevant information extracted from remote sensing imagery. This processing flow is named the **ingestion chain**, which entails 1) cutting and storing of the images into the database, 2) **extraction of primitive features** from the images, 3) clustering of the extracted primitive features, and 4) storing of the class catalogue (generic content-index) into the database. The ingestion chain is performed off-line.

KIM has several methods implemented to perform the primitive feature extraction, which are selected by the user according to the image type. For example, in the case of SAR images, Gauss Markov Random Field is used as primitive features extraction method. This method was previously presented in section 5.1.2. This method extracts **texture parameters**, which characterize the image content, and its performance depends on the set of input parameters such as model order and estimation window size. It is

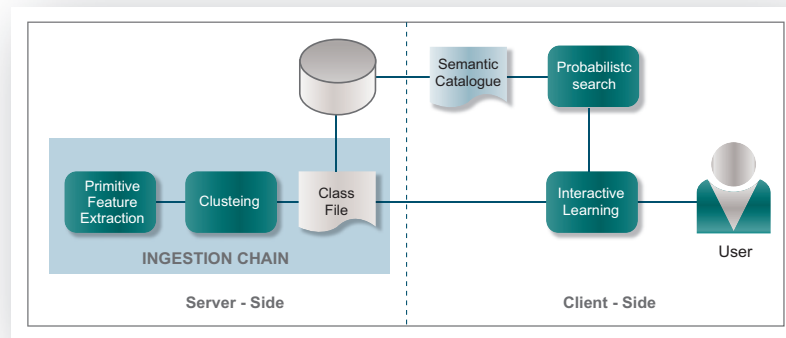


Figure 6.2: KIM client-server system architecture.

worth noting that the performance of the primitive feature extraction is highly dependent on the selected input parameters, thus, properly set parameters ensure the accuracy of the results. In this context, the proper selection of the parameters is a matter of evaluation, therefore, the evaluation of the off-line components was presented in chapter 5.

The **client-side** allows users to interact with the system and access the data collections through the **GUI**. Here, the common performed tasks are 1) the **interactive learning** of the KIM by training examples provided by the user, 2) the **probabilistic search** based on the required content according to user specifications, and 3) the **creation of semantic catalogue**, which is stored into the database.

6.1.2.1 GUI Components

In this thesis, we only mention the main GUI components related with scene retrieval and classification, since they are used by the end user during the creation of validation scenarios and the evaluation.

- **Collections Management (Ingestion Chain)**

This component (see Figure 6.3) allows the user to start a new ingestion in the server-side. The term **ingestion** refers to the extraction of primitive features from Earth-Observation (EO) products. Through this component it is possible to add a new collection, to select the primitive feature extraction method, and to configure their relevant parameters. The term **collection** refers to a set of one or more images grouped by a common purpose and assigned to a specific project.

- **Image Browser**

The image browser displays all available image collections, allowing the user to navigate through the image collections stored inside the database.

- **Feature Label Designer**

The feature label designer allows the users to define the semantic labels by training the **probabilistic information mining system** with positive and negative examples.

The user is able to train the system in order to get similar scenes or to make classifications through this tool. A **feature label**, named in this work **semantic label**, is a weighted combination of contributing primitive features specified by a semantic term and a description.

- **Probabilistic Search**

After saving the semantic labels, two kinds of searches in the collection can be performed: 1) search based on similarity labels, and 2) search based on extracted primitive features. The system retrieves the images based on the feature labels, that were defined by the user. **The probabilistic search returns a ranked list** ordered by probability, separability or coverage as result. The evaluation will focus on the effectiveness of the retrieved results.

- **Product Explorer**

The product explorer opens a window used in order to explore the original image and a mosaic of the selected semantic label, thematic map or primitive feature. This tool **allows to export the maximum a posteriori map with the classification of the image content** according to defined semantic labels.

In the following, in order to clarify the KIM concept and its architecture, an example of configuration and working with the KIM system is presented.

6.1.3 Operation of KIM

At the beginning, the end-user of KIM needs to be authenticated by the system itself in order to grant him the access to various tools, and provide him with information tailored to his profile. After the registration by filling the information such as his username and password, the user can work with KIM.

6.1.3.1 The Server-side

On the **server-side** the process starts when **the user performs an ingestion of the EO product through the collection manager**. During the ingestion, the image will be cut in several tiles, and from each tile the primitive feature will be extracted. An example is shown in Figure 6.3. Here, the user will ingest a TerraSAR-X image over Berlin as an example. In this case, the parameters selected are GMRF as the primitive feature extraction method, with model order 4 and estimation windows size of 40×40 pixels. The size of the tiles is also set at 500×500 pixels. The duration of the ingestions depends on the type and size of the images. When the image is ready, **a new collection is available in the Image Browser**, enabling thus the work on the **client-side** and the dialog between user and KIM can be performed.

6.1.3.2 The Client-side

On the **client-side**, the **browsing and query engine in combination with the interactive learning system** module provide the following user functions: 1) define a semantic label and classify the image content, 2) search the archive for relevant images according to the defined label, 3) query the archive for certain objects and structures, and 4) analyze the returned images.

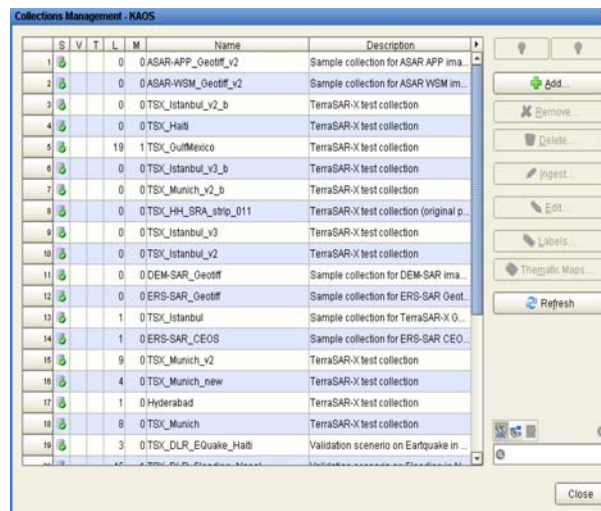


Figure 6.3: KIM components: The collection management. The ingestion chain is performed through this tool, which starts the process on the server-side.

Then, the workflow starts when the user chooses an image collection by using the **Image Browser** (See Figure 6.4) and selects a specific tile to search for image content according to his interest, to train the system and to define semantic labels. In

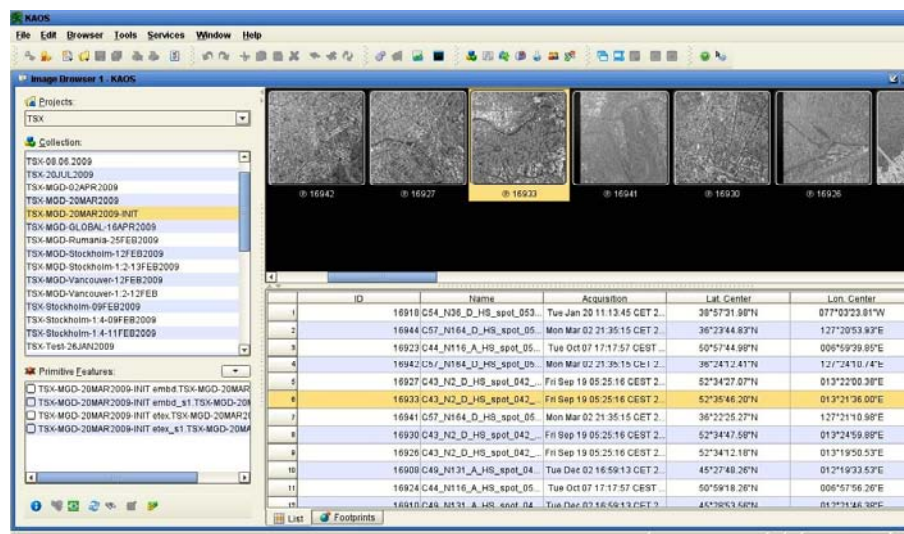


Figure 6.4: KIM components: The image browser shows the whole image collection as tiles with their metadata.

this example, the user selects an image tile highlighted in Figure 6.4, which contains **high buildings** and he will look for similar content in the whole image collection. At the beginning, it is needed to define a semantic label and to associate it with a generic class in order to search for a specific content and to retrieve the results. Then, the user defines a new semantic label by selecting different primitive features, which are described in Table 6.1 in the case of SAR images, and performs the training. An example of training is shown in Figure 6.5.

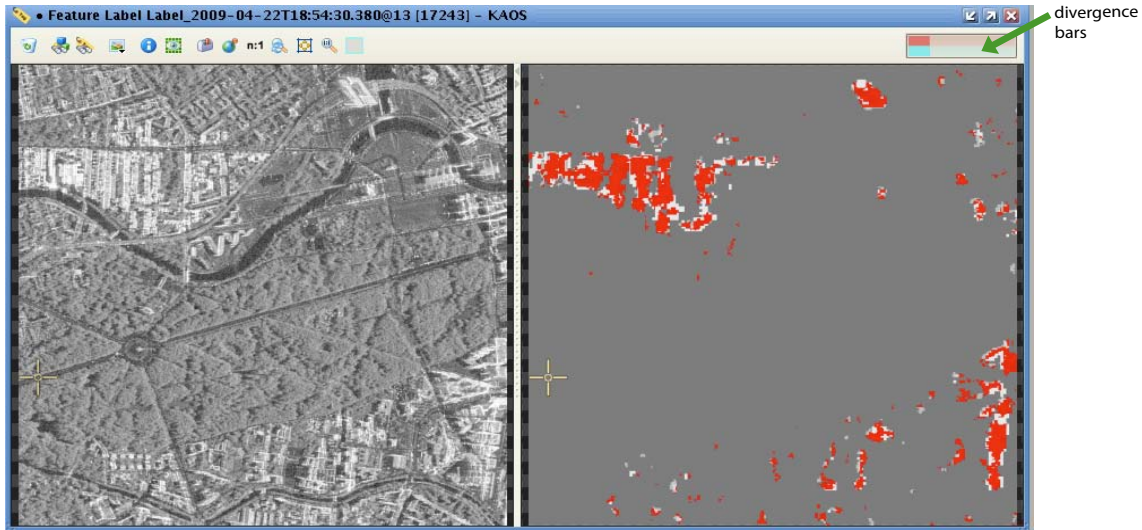


Figure 6.5: KIM components: Training of **high buildings** semantic label. The user gives positive and negative examples by clicking on the image content, while the system learns and shows a feedback in the divergence bars in the right-upper part.

The left side of Figure 6.5 shows the selected tile and the identified features are displayed in red (posterior map) on the right. The training is performed by giving positive and negative training samples using the left and right mouse button. After each click, the posterior map is updated with different colors indicating the probability of each point (red means positive example, black means negative example, grey means not assigned). Additionally, the divergence bars change according to the user-performed actions. Both the posterior map and divergence bars give the user a feedback about the quality of the performed training and the state of the semantic label.

Table 6.1: Primitive features provided by KIM in the case of SAR images.

Primitive Feature	Abbreviation	Meaning
Intensity 0	<i>embd0</i>	The despeckled image with original resolution.
Intensity 1	<i>embd1</i>	The despeckled image with two time reduced resolution.
Texture 0	<i>tex0</i>	The texture parameters provided by GMRF prior model at original scale.
Texture 1	<i>tex1</i>	The texture parameters provided by GMRF prior model at two time reduced resolution.

Later, when the training is finished and the user is satisfied with the defined semantic label, the selected primitive features are used in order to search for similar image content in the whole collection. In our example, the results are presented to the user as a gallery of top ranked images, containing "high buildings" ordered according to their probability as shown in Figure 6.6. The user can refine the semantic label cover-type and the probabilistic retrieval list by selecting another tile and repeating the interactive learning training.

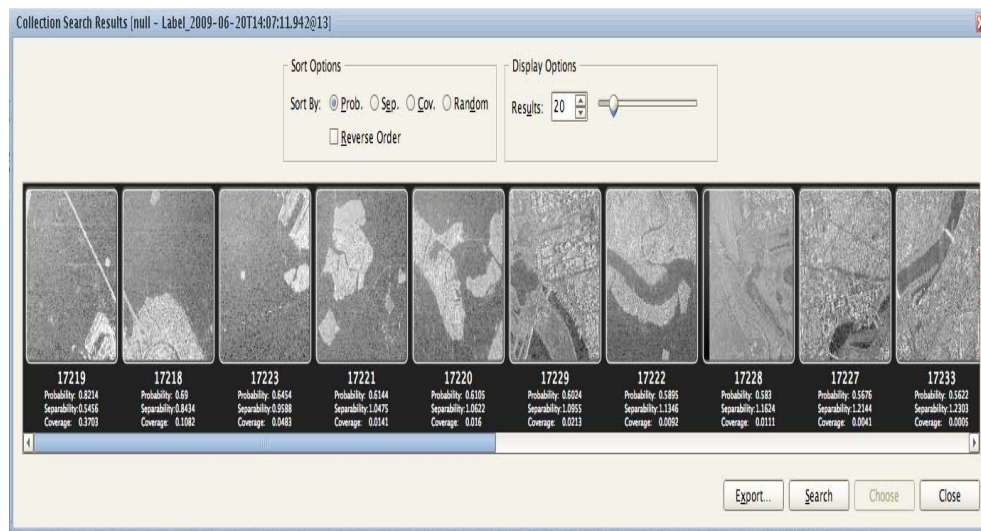


Figure 6.6: KIM components: The collection search results. This component shows an ordered list with the top relevant images.

Through the training and definition of semantic labels, the user makes a supervised classification of the image content. This supervised classification can be analyzed by using the product explorer as shown in Figure 6.7. In here, the complete image is classified according to the semantic labels and later the probabilistic supervised classification map can be exported for further analysis. Moreover, using the same tool it is also possible to combine the semantic labels in a thematic map and export it.

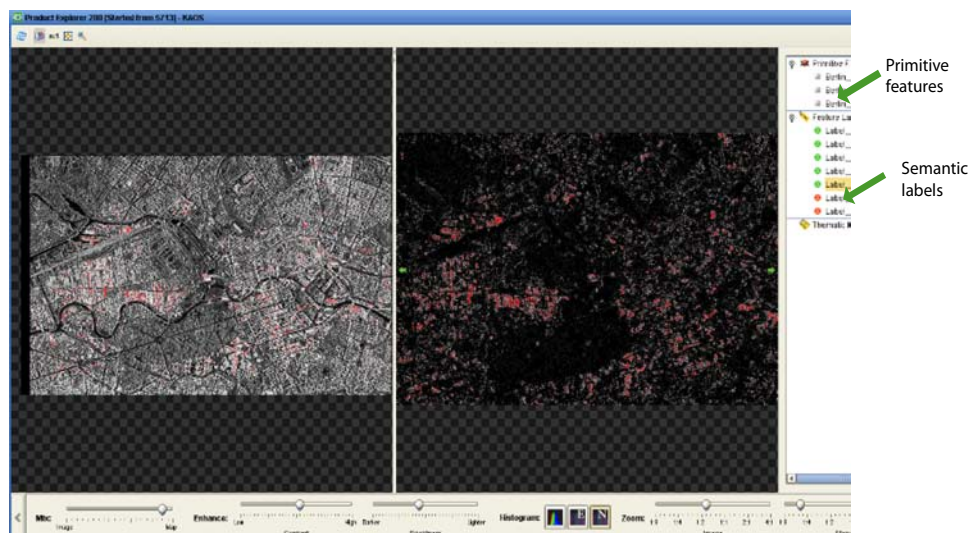


Figure 6.7: KIM components: The product explorer shows the complete image, the primitive features, and the defined semantic labels. This content is possible to export for further analysis. (Left) TerraSAR-X image over Berlin. (Center) Classification map according to the semantic label. (Right) Primitive features extracted from the image and semantic labels.

6.2 Effectiveness evaluation metrics

In the context of information retrieval, one of the primary distinctions to be made in the evaluation of search engines is between **effectiveness** and **efficiency**. While **effectiveness** measures the ability of the search engine to find the right information, efficiency measures how quickly this is done. For a given query, and a specific definition of relevance, we can more precisely **define effectiveness as a measure of how well the ranking produced by the search engine corresponds to a ranking based on user relevance judgments** (Croft et al., 2010). Efficiency is defined in terms of time and space requirements for the algorithm that produces the ranking. We are interested in evaluating the effectiveness of KIM.

In the following, the precision and recall measures as effectiveness evaluation measure and metrics derived from them are explained.

6.2.1 Precision and Recall

The two most common effectiveness measures, recall and precision, were introduced in the Cranfield studies (Cleverdon et al., 1966) to summarize and compare search results. Intuitively, recall measures how well the search engine is doing at finding all the relevant documents or images for a query, and precision measures how well it is doing at rejecting non-relevant documents. The definition of these measures assumes that, in the entire collection, for a given query, there is a set of documents or images that is retrieved and a set that is not retrieved (the rest of the documents), where part of them are **relevant**. This definition is explained in Figure 6.8 and Table 6.2, where the possible results of the query are summarized.

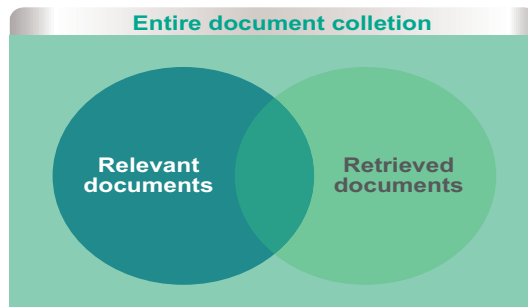


Figure 6.8: Precision and recall graphical explanation. The query retrieves a set of documents, where some of them are relevant.

	Relevant	non-Relevant
Retrieved	true positive (tp)	false positive (fp)
Not retrieved	false negatives (fn)	true negatives (tn)

Table 6.2: Definition of contingency tables

In the entire document or image collection (see Figure 6.8), a set of documents is judged as **relevant** according to the user interests, and another set is **retrieved** by the performed query. The intersection of both of them derives some quantitative measures

and the possible combinations of results are expressed in the contingency tables (see Table 6.2). The precision and recall measures are defined in the following as:

- **Precision (P)**

This is the fraction of retrieved images that are relevant, it is expressed as

$$\text{Precision} = \frac{\#(\text{relevant items retrieved})}{\#(\text{retrieved items})} \quad (6.1)$$

or

$$\text{Precision} = \frac{(\text{tp})}{(\text{tp} + \text{fp})}. \quad (6.2)$$

- **Recall (R)**

This is the fraction of relevant images that are retrieved, it is given by

$$\text{Recall} = \frac{\#(\text{relevant items retrieved})}{\#(\text{relevant items})} \quad (6.3)$$

or

$$\text{Recall} = \frac{(\text{tp})}{(\text{tp} + \text{fn})}. \quad (6.4)$$

6.2.2 Accuracy and F-measure

An alternative metric for evaluating the retrieval system is its **accuracy**. In terms of the contingency table (6.2), the accuracy is giving by

$$\text{Accuracy} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{fp} + \text{fn} + \text{tn}}. \quad (6.5)$$

The measures of precision and recall concentrate the evaluation on the return of true positives, asking what percentage of relevant documents have been found and how many false positive have also been returned. The two quantities show a clear trade-off against one another; you can always get a recall of 1, but a very low precision, by retrieving all documents for all queries.

A single measure that does a trade-off between precision and recall is **F-measure**. This later is the weighted harmonic mean of precision and recall and is given by

$$F = \frac{1}{\alpha \frac{1}{P} + (1 - \alpha) \frac{1}{R}} = \frac{(\beta^2 + 1)PR}{\beta^2 P + R}, \quad (6.6)$$

where

$$\beta^2 = \frac{1 - \alpha}{\alpha} \quad (6.7)$$

and $\alpha \in [0, 1]$ and thus $\beta^2 \in [0, \infty]$. The default balanced F-measure equally weights precision and recall, which means making $\alpha = 1/2$ or $\beta = 1$. It is commonly written as F_1 , which is sort for $F_{\beta=1}$, even though the formula in terms of α more transparently

exhibits the F measure as a weighted harmonic mean. When using $\beta = 1$, the formula on the right side can be simplified to:

$$F_{\beta=1} = \frac{2PR}{P+R}. \quad (6.8)$$

Values of $\beta < 1$ emphasize precision, while values of $\beta > 1$ emphasize recall.

6.2.3 Probability to over-retrieve and to forget an image

Other interesting measures derived from precision and recall are the **probability to over-retrieve** an image P_o and the **probability to forget** an image P_f . Whereas *precision* and *recall* deliver the retrieval performance using the queried results from the database, P_o and P_f give measurements of lost and confused images in the database (Daschiel and Datcu, 2005), respectively. They are given by

$$P_o = \frac{T \setminus R}{T} \quad (6.9)$$

and

$$P_f = \frac{R \setminus T}{R} \quad (6.10)$$

where T is the set of returned images and R is the set of relevant images to query (See Figure 6.8).

6.2.4 Target and misclassified images

A queried image is said to be a **target** if it contains the trained label from an operator point of view, while a **misclassified** image was retrieved but does not contain the trained label. The relevant image can be determined by visual inspection as well as by ground-truth information.

Recall, precision and F-measure are inherently measures between 0 and 1, but they are commonly written as percentages, on scale between 0 and 100 and plotted in graphics.

6.3 Evaluation Methodology

The evaluation methodology involves 1) working with the user as an interpreter of the ground-truth or as an evaluator of the KIM system, 2) creating the ground-truth, and 3) performing the system evaluation methods.

In the following, section 6.3.1 deals with the role of the user. In section 6.3.2 the different steps to create the ground-truth for evaluating KIM are presented. Section 6.3.3 describes what components of KIM will be evaluated.

6.3.1 Role of the user

The user is the main actor in the system operation. During the user-driven evaluation, the user has two roles: 1) as interpreter (expert): he creates the reference database by annotating manually the image tiles. This database will act as a ground - truth during the evaluation, 2) as evaluator: he creates the validation scenarios by interacting with KIM, and later judges the results.

6.3.1.1 The user as interpreter of the ground-truth

During the evaluation, the creation of a ground-truth is a critical point. We should count with a reliable ground-truth, which ensures the validity of the results. A way to create a ground-truth could be by annotating manually the image content. Therefore, we need one expert or more that understand SAR images and help us by annotating the image content.

The user as interpreter provides a definition of the image content according to his experience and knowledge of SAR images. Moreover, since the visibility in SAR images is difficult, the annotating task could be supported by optical images.

6.3.1.2 The user as evaluator of the system

The end-user or evaluator will define the semantic labels by training the KIM with positive and negative examples. He will query the image catalogue looking for relevant data according to his interest and needs. He will also provide his degree of satisfaction using the KIM, which is an indicator of the system performance (Griffiths et al., 2007). The user could have different levels of expertise in understanding remote SAR images and in working with CBIR systems.

After the training, the user will be asked to rank the retrieved results with the **qualitative scale** ranging from low value to high values as follows: Unsatisfied (U), Acceptable (A), Good (G), Very Good (VG) according to his satisfaction degree. The user will also try to use VG, only when he is completely satisfied with the results.

6.3.2 Creation of the ground-truth

The creation of the ground-truth involves the following tasks: the first task consists in the **definition of the relevance criterion**, when is an image considered relevant in a search? or which consideration needs to be taken into account in order to judge an image as relevant?. Later, when the relevance is decided, the second task which involves the **definition of standards for annotating** the image tiles takes place. Finally, a **structure to store all information** about the image content needs to be created.

In the following, we start with the discussion of the factors to be taken into account to define an image as relevant in section 6.3.2.1. Later, a scheme for annotating the image tiles is proposed in section 6.3.2.2. Finally, the storing of all this information in a reference database is described in section 6.3.2.3.

6.3.2.1 Definition of the relevance criterion

The standard approach for the evaluation of information retrieval systems deals with the notation of **relevant** and **irrelevant** images. For example, the evaluation metrics previously presented in section 6.2.1 need the definition of a **relevance** criterion for their computation.

The relevance is assessed relative to the information need, not the query. An image is relevant if it addresses the stated information need. The user can judge the evaluated image collections on basis of their relevance to it. In this context, the **relevance assessment** is a crucial part of the user-driven evaluation because it provides the link between an image collection and the representative tasks (validation scenarios) according to the user needs and allows to compute the effectiveness measures during the evaluation.

In the relevance assessment the images are judged whether they are relevant or not, creating the so-called *ground-truth* for each validation scenario. There are two approaches for defining the relevance judgments (Grubinger, 2007):

- **Multi-grade judgements**

The majority of test collection have marked with relevance judgements as binary ("relevant" or "not relevant"), whereby images are judged as relevant if any part of it is relevant, regardless of how small the proportion of that part is in relation to the entire image. However, it is possible to assign a different degree of relevance depending on the content, for example: relevant, partially relevant, not relevant (Voorhees and Harman, 1998).

- **Assessor disagreement**

Relevance may seem to be a very subjective concept, and it is known that relevance judgements can vary depending on the person making the judgements or even for the same person at different times. Therefore, it is recommended to adopt multiple judgement to reduce the influence of subjectiveness (Shatford, 1986).

The use of multi-grade judgements and several assessors inherently leads to the creation of several sets of relevant images per query topic. In this thesis, we define **two levels of relevance**, relevant and irrelevant, where the relevance is assessed according to the validation scenario and the user requirements. The user as expert supervises the images and decides which classes are contained in the image.

6.3.2.2 Definition of the scheme for annotating images

A hierarchical scheme is proposed for annotating image tiles. It consists of a two-level hierarchy using two digit numbers as a code for identifying a semantic class inside the image. The **hierarchical classification scheme is presented in Table 6.3**. This is an adapted version of the Anderson scheme (Anderson et al., 1972), which has a tree-level hierarchy of a classification scheme for urban development.

In this work, the selected classes are only these ones involved in the validation scenarios. In addition to the hierarchical classification scheme, an **annotation convention** is defined, which will help us to annotate the image content in a standard manner. Each tile will be annotated with a maximum of three defined classes according to their percentage of coverage.

Figure 6.9 shows an example of an annotated tile. This tile is composed of three main classes. Here, the classes are 1) the primary class, which covers the image is agricultural field (not specified further) and its code corresponds to 30 according to Table 6.3. 2) The secondary class corresponds to forest with code 40, and 3) the minor class which covers the image is flooding, with code 64. The number codes are given in Table 6.3.

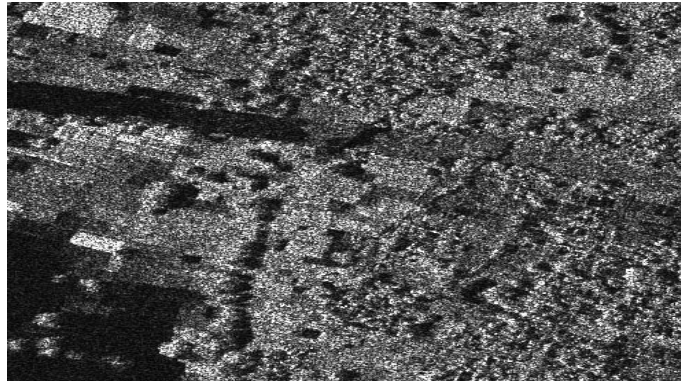
6.3.2.3 Creation of the reference database

In order to evaluate the retrieved results provided by KIM, we created a parallel database to store the information of the annotated images and to provide a link with the information stored in KIM.

The reference database requires to correctly identify the image content and properly design a scheme. Therefore, the proposed structure is shown in Figure 6.10, where the

Level I	Level II	
1 Urban built-up	10	Not specified further
	11	High-density residential area
	12	Medium-density residential area (row buildings)
	13	Low-density residential area (single family, treeded)
	14	Informal settlements - solid (favelas, kibera) and transitional.
	15	Refugee camps (traditional housing, structured)
	16	Commercial area
2 Transportation	17	Industrial area
	20	Not specified further
	21	Roads
	22	Railroads
	23	Bridges
3 Agriculture	24	Ports
	25	Airports
	30	Not specified further
	31	Cropland (all types)
4 Forest	32	Pasture
	33	Stubble / bare agriculture land
	40	Not specified further
	41	Coniferous forest
	42	Broadleaf forest
5 Bare ground	43	Mixed forest
	44	Regrowth
	45	Clearcut
	50	Not specified further
6 Water	51	Brush / Rangeland (dense, scattered)
	52	Barren, soil or sand
	53	Barren, rock / bare agriculture land
	54	Snow / ice covered
	60	Not specified further
7 Oil	61	River / channel
	62	Lake (natural or artificial)
	63	Ocean
	64	Flooding areas
	70	Not specified further

Table 6.3: Two-level hierarchical classification scheme based on Anderson scheme (Anderson et al., 1972). The cited classes are used in the reference database and validation scenarios.

(a) TerraSAR-X tile 500×500 pixels size.

Primary class	Secondary class	Tertiary class
30	40	64
Agriculture fields not specified	Forest	Flooding Areas

(b) Hierarchical annotation of the tile.

Figure 6.9: Example of annotated TerraSAR-X tile. (a) This tile contains agriculture fields, forest and flooding areas. (b) The information is annotated according to the hierarchical scheme presented in Table 6.3.

main objects are represented by tables (composed of fields). They are described in the following:

- **IMAGE**

In this table the information about the original image used for annotation is stored. This image will be cut in several tiles. The stored data are: its name, its path, the coordinate systems expressed in latitude/longitude and some comments or descriptions about the image.

- **LABEL**

This table contains the **definition of semantic class** the user is interested in. This object allows the user to associate a name with a meaning for a searched class. The stored data are an identification, a semantic name, and a description.

- **SUBSCENE**

The original image is divided in several tiles, which are stored in this table. The stored data are rows and columns, which give us its localization in the main image, the size of the tile, some comments, the path, and identification of the main image to which it belongs.

- **SUB-LAB**

This table contains the relation between *labels* and *sub-scene* tables. It stores how many semantic labels are associated with a sub-scene.

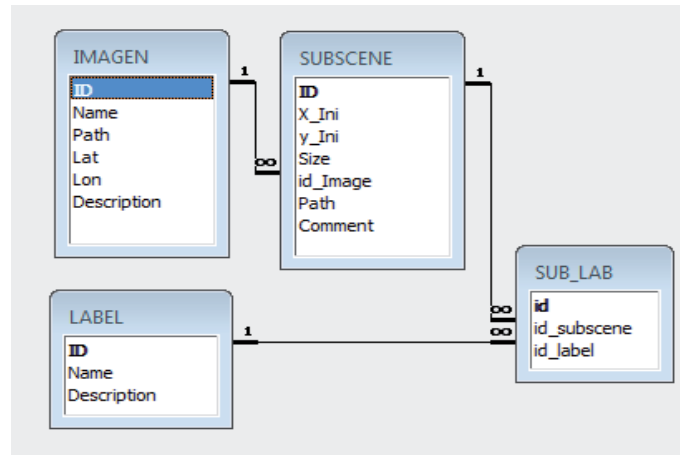


Figure 6.10: Database structure for the annotated image tiles.

6.3.3 System evaluation methods

In the KIM client-server architecture, shown in Figure 6.2, it can be observed that the user is involved in the interactive learning, defining the semantic labels and performing queries (probabilistic search) based on the image content. Therefore, the user-driven evaluation of KIM will focus on the **evaluation of the probabilistic search and the interactive learning**, which is equivalent to evaluate the quality of the user defined labels.

In order to evaluate KIM, we create different scenarios, which show the user-KIM interaction through the interactive learning and the probabilistic search. The proposed methodology for creating the scenarios entails three steps: 1) configuration of the scenario, 2) semantic labels definition through the training performed by the user, and 3) assessment of the effectiveness of the retrieved results and classifications. These steps are described in the following:

6.3.3.1 Configuration of the scenario

In this step, the user defines the goal of the scenario and his interests. Later, according to this information, the user selects the image, generates the ground-truth, and ingests the data into KIM.

6.3.3.2 Training performed by the user and definition of semantic labels

The end-user trains the KIM system during the interactive learning. Here, the user defines the semantic labels according to his understanding about the image content and interests, creating a link between the classes in the generic catalogue and the defined label (see Figure 6.1). Therefore, **each semantic label can be seen as a supervised classification of the entire image database**, where its accuracy indicates the objective quality of the system to detect the cover-type (different classes) in the whole image.

The semantic label is interactively defined by the user. The user starts selecting a certain combination of primitive features, which were provided by the features extraction methods. Later, the user will repeatedly give positive and negative examples by different clicks on the image content stopping when he is satisfied. Finally, the user searches for similar image content in the collection and qualitatively ranks the results.

When the user searches for relevant images, a probabilistic search is performed. It gives as result a top ranked list. The user judges the relevance of an image in the ranked list. If the user considers that an image is irrelevant in the ranked list, he can improve the results by choosing a new image and performing a new interactive training. This concept is similar to a relevance feedback framework (Salton and Buckley, 1990).

6.3.3.3 Assessment of the probabilistic search and interactive learning

After training the system through the interactive learning, the semantic cover type labels are defined. Later, based on the stochastic link between semantic label and generic content index, the whole data repository can be searched in a probabilistic way for relevant images. The end-user (system operator) searches for an image based on its content. Later, the query results provided in the probabilistic search will tell us how well the semantic labels defined by the user are and how effective the retrieval process is.

In this thesis, the effectiveness of the probabilistic search (retrieved results) are evaluated by computing the effectiveness measures proposed in section 6.2. The quality of interactive learning, which is equivalent to the quality of supervised classification of the whole image, is evaluated by the user through visual impressions and confusion matrices.

The user satisfaction degree is obtained by questioning the evaluators to rank qualitatively the results of the probabilistic search and supervised classification using a **qualitative scale**: Unsatisfied (U), Acceptable (A), Good (G), and Very Good (VG).

6.4 Examples of validation scenarios: Disaster monitoring

The German Remote Sensing Data Center (DFD) of the German Aerospace Center (DLR) has set up a dedicated crosscutting service, the so-called **Center for satellite-based Crisis Information (ZKI)**, to facilitate the use of its EO capacities in the service of major disaster situations, humanitarian relief efforts, and civil security issues (Voigt et al., 2007). ZKI operates in the context of disaster-response. For example more than 11 scenarios which sum up a total of 104 alerts were processed by ZKI till 2010. The scenario categories are: Flood, Earthquake, Wild Fire and Burn Scars, Landslide, Storm and Hurricane, Tsunami, Volcanic Eruption, Technical Accident, Humanitarian Crisis, Exercise, and Others.

In the context of disaster monitoring, we selected two examples based on the developed work of ZKI. The first scenario corresponds to oil spill detection in the Gulf of Mexico, and the second one is Nepal-Embankment breach, flooding monitoring.

In the framework of Oil spill detection, several publications have been presented using SAR data, as for example the work of (Mercier and Girard-Arduin, 2006), a semi-supervised oil-slick detection technique has been proposed using single class support vector machines into a wavelet decomposition of a SAR image. A specific kernel is developed to perform accurate segmentation of the local sea-surface wave spectrum by using both radiometric and texture information. The algorithm has been applied on ENVISAT Advanced SAR images. It yields accurate segmentation results even for small slicks, with a very limited number of false alarms. In (Solberg et al., 2007), the authors presented algorithms for automatic detection of oil spills in SAR images, specifically using Radarsat and Envisat. The algorithms consist of three main parts: 1) detection of dark spots, 2) feature extraction from the dark spot candidates, and 3) classification of dark spots as oil spills or look-alikes. In the first step, the authors used a multiscale approach for dark

spot segmentation based on adaptive thresholding algorithm to segment each level in the scale pyramid. The features extraction was based on features that describe slick shape, contrast, homogeneity, and surroundings. Finally, the classification approach is a mix between a statistical classifier with subclasses based on wind and shape and a rule-based approach. The average classification accuracy is 78 for oil spills.

In the framework of Flooding monitoring using TerraSAR-X many publications have been presented as for example the use of TerraSAR-X images for water bodies detection is presented in (Hahmann et al., 2008). Here, the authors showed that the TerraSAR-X SAR sensor has enormous potential for flood detection in disaster mapping activities due to its high geometric resolution and repetition rate. A review of pixel and segmentation-based classification, as well as thresholding methods for flood extraction and their adaptability for high resolution SAR data are presented. The conclusions show that in the case of smooth water surfaces the thresholding approach works satisfactorily. However, the development of new segmentation-based and texture-based methods becomes necessary due to the large amount of visible image details in high-resolution (Hahmann et al., 2008).

Mason et al. (2010) proposed an algorithm to detect urban floodwater in a TerraSAR-X image and to estimate the accuracy of this detection and the degree to which this accuracy was tempered by the presence of shadow and layover. The authors mention that TerraSAR-X is good at identifying those flooded areas that are visible to the SAR and reasonably good at identifying all the flooded urban areas.

A procedure to obtain flooding mapping using multi-temporal TerraSAR-X images was presented in (Lu et al., 2009). The method extracts the water bodies from the pre and post-disaster images. The changes of the water body areas correspond to the flooded areas. The main process is to extract accurately the water bodies from the images using mainly mean shift and k-means algorithms to segment the image. A similar approach is presented in (Herrera-Cruz and Koudogbo, 2009). Here, the authors proposed an object oriented methodology, which consists of segmentation, water mask extraction and flood extraction using pre-event and post-event data. As a conclusion, we remark that the high resolution of TerraSAR-X offers enormous potential in the domain of flood mapping. However, the improved spatial resolution of the SAR data results in a large variety of very small-scaled image objects, which makes image processing and analysis even more challenging. The ongoing research activity is the development of information extraction methods, which allow SAR data processing and reliable water bodies detection and flooding mapping.

In the following validation scenarios, which were motivated from ZKI work, are presented. The oil spill detection in Gulf of Mexico is presented in section 6.4.1, and the flooding in Nepal due to embankment breach is presented in section 6.4.2.

6.4.1 Study case 1: Oil Spill in Gulf of Mexico

An explosion on April 20, 2010, aboard the Deepwater Horizon, a drilling rig working on a well for the oil company BP one mile below the surface of the Gulf of Mexico, led to the largest accidental oil spill in history. An estimated amount of up to 800.000 liters of oil were leaking from three places of the well and threatened the marine ecosystem of the Gulf of Mexico, the ecologically sensitive Mississippi delta as well as the coast of Louisiana, USA. Figure 6.11 depicts the magnitude of the tragedy.



(a) Explosion aboard the Deepwater Horizon.

(b) Oil films over the Atlantic Ocean.

Figure 6.11: Oil Spill in Gulf of Mexico, on April 20, 2010. From left to right, Explosion on Deepwater Horizon and oil spilled in the Atlantic Ocean. (images taken from time.com)

6.4.1.1 Description of data set, parameters and ground-truth

During several overpasses over the Gulf of Mexico the German Satellite TerraSAR-X acquired SAR images showing the extensive oil spill on the sea surface caused by the oil rig disaster. In this scenario, **the user is interested in identifying the oil surface from the water**. Therefore only two classes need to be detected: water bodies and spilled oil.

- **Data set**

The analyzed TerraSAR-X scene used in this study case is shown in Figure 6.12. This image corresponds to a **GEC_RE_SC_S** product, which is a radiometrically enhanced single polarization **ScanSAR mode in Geocoded Ellipsoid Corrected TerraSAR-X product** (DLR, 2009). It is a post-disaster image and has been acquired on May 11, 2010.

- **Selected parameters**

The whole scene was ingested in KIM system and during the **ingestion** it was divided in 50 tiles with 1000×1000 pixels size each one. **MAP-GMRF** (See section 5.1.2) was used as the primitive feature extraction method, whose input parameters were **model order 5 and estimation window** size of 31×31 pixels.

- **Ground-truth**

During the **ground-truth definition**, the TerraSAR-X image was divided in 50 tiles in order to have the same tiles as the ones ingested in KIM system. Each tile was annotated by the interpreter-user. In this study case, the tile's content is either Ocean water (code=63) or oil (code=70). The codes are given in Table 6.3. It was found that 43 tiles contain ocean water and 31 tiles contain oil. This is summarized in Table 6.4.

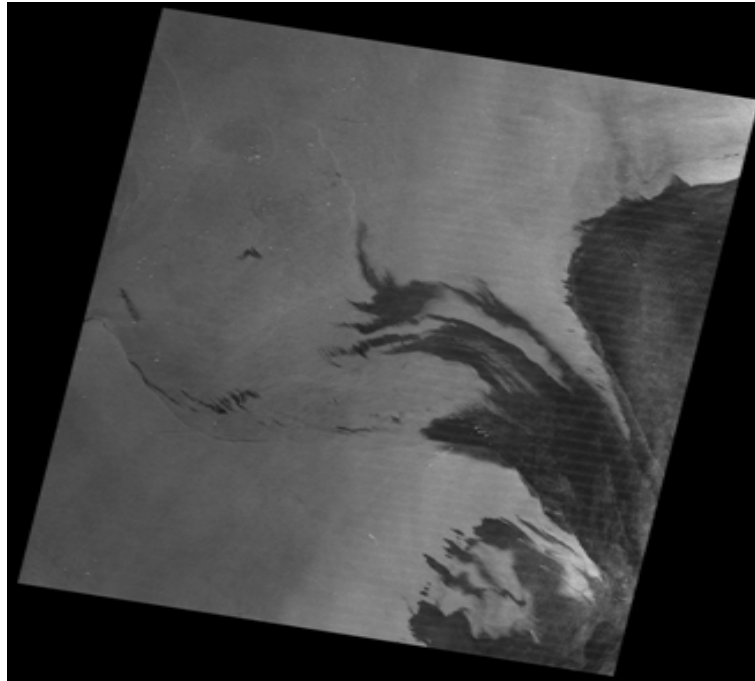


Figure 6.12: TerraSAR-X image over Gulf of Mexico taken on May 11, 2010. The image shows an extensive oil film east of the Mississippi delta.

Table 6.4: Resume of Mexico Gulf collection ground-truth

Level	Code	Semantic class	Total of relevant images
II	63	Ocean water	43
I	70	Oil	31

6.4.1.2 Training performed by the end-user

The creation of the validation scenario begins with the interactive training performed by the end-user or evaluator. The end-user must train the system in order to query the database and classify the image content. In this scenario, it is required to have two classes: oil and ocean water.

Table 6.5: Semantic class labels and primitive features used in the training performed by the user in the case of oil spill scenario.

Semantic labels	Primitive features
Oil & Ocean water	embd0 tex0 embd1 tex1 embd0 + tex0 embd1 + tex1 tex0 + tex1

During the training and definition of the semantic labels (Ocean water and Oil), several combinations of primitive features were used, as shown in Table 6.5. In the case of Oil class, the user performed one iteration with the probabilistic search (kind of relevance feedback), 3 clicks on average taking an approximate time of 6 minutes for each primitive feature in the training. An example of an user click sequence (progressive clicks) is displayed in Figure 6.13. Here, the desired class is **oil**. The red zone is increased according to the image understanding by the system. Here, it can be noted how KIM adapts according to the user clicks.

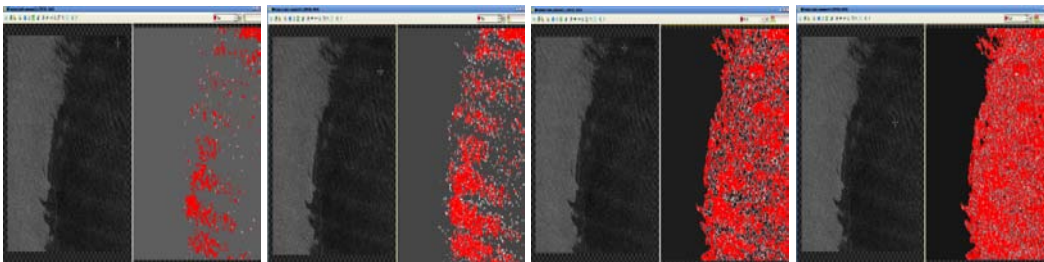


Figure 6.13: Example of training, the user performs several clicks on the selected image tile looking for a specific class (oil in this case). From left to right, the identified features are trained according to the first, second, fifth and sixth click.

6.4.1.3 Evaluation results of the probabilistic search

The interactive training (See Figure 6.13) stops when the user is satisfied with the identified class. The next step is to look for similar image content in the whole image collection and evaluate the effectiveness of the retrieval system in terms of the metrics previously presented in section 6.3.2.2.

In the following, the effectiveness results for both classes using all primitive features defined in the training is summarized in Table 6.6, here also the user criteria were in-

cluded¹. Later, we analyze the performance of user-defined classes in terms of Precision and Recall (PR) curves. We start with presenting the precision of all the primitive features used in the training sessions. Then, in order to summarize the results, we detail the PR curves for the best primitive feature according to Table 6.6 versus the number of retrieved images, and PR versus probability.

Table 6.6: **Oil spill in Gulf of Mexico: Effectiveness evaluation of the probabilistic search**, considering the top ranked list for oil and ocean class versus the primitive features. The user satisfaction (column U-satis) is expressed as Unsatisfied (U), Acceptable (A), Good (G), and Very Good (VG).

Semantic label:		Oil					
Primitive feature*	Targets	Misclas.**	Po	Pf	Accuracy	F-measure	U-satis.
embd0	95,00	5,00	0,05	0,05	0,73	0,75	VG
tex0	90,00	10,00	0,06	0,05	0,69	0,71	G
embd1	60,00	40,00	0,08	0,05	0,44	0,47	A
tex1	80,00	20,00	0,06	0,05	0,60	0,63	G
embd0 + tex0	75,00	25,00	0,07	0,05	0,56	0,59	A
embd1 + tex1	75,00	25,00	0,07	0,05	0,56	0,59	A
tex0 + tex1	90,00	10,00	0,06	0,05	0,69	0,71	VG
Semantic label:		Ocean					
Primitive feature*	Targets	Misclas.	Po	Pf	Accuracy	F-measure	U-satis.
embd0	100,00	0,00	0,05	0,05	0,52	0,63	VG
tex0	90,00	10,00	0,06	0,05	0,44	0,57	G
embd1	80,00	20,00	0,06	0,05	0,35	0,51	G
tex1	90,00	10,00	0,06	0,05	0,44	0,57	G
embd0 + tex0	90,00	10,00	0,06	0,05	0,44	0,57	G
embd1 + tex1	85,00	15,00	0,06	0,05	0,40	0,54	G
tex0 + tex1	90,00	10,00	0,06	0,05	0,44	0,57	G

* The primitive features are presented in Table 6.1. **Misclas refers to misclassified metric.

- **Oil class results**

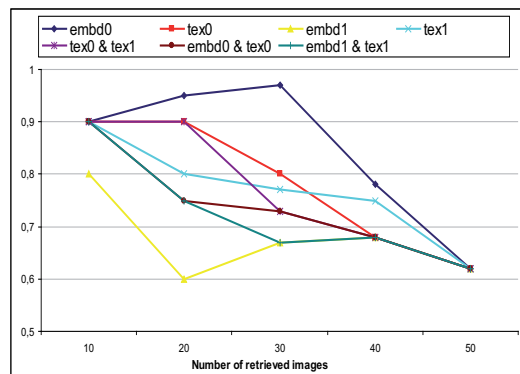
Table 6.6 shows that the best results are obtained with **embd0** as primitive feature. This does mean that the intensity image is enough for recognizing this class. Moreover, the effectiveness of KIM retrieving oil is good since all the metrics have high values. Indeed, with the exception of *embd1*, all the primitive features succeed in recognizing more than 75 percent of the targets. These results are confirmed in Figure 6.14(a), where the precision of all training sessions are plotted. Here, it can be observed that the highest precision was obtained using **embd0** as primitive feature. Moreover, it is worth to observe that, all the primitive features have a precision range between 0.7 and 0.90.

The precision and recall values for the best primitive feature (**embd0**) are displayed in Figure 6.14(b)-(c). In Figure 6.14(b), we can notice that the 30 top ranked images are retrieved with a precision higher than 90 percent. Moreover, in Figure 6.14(c) it

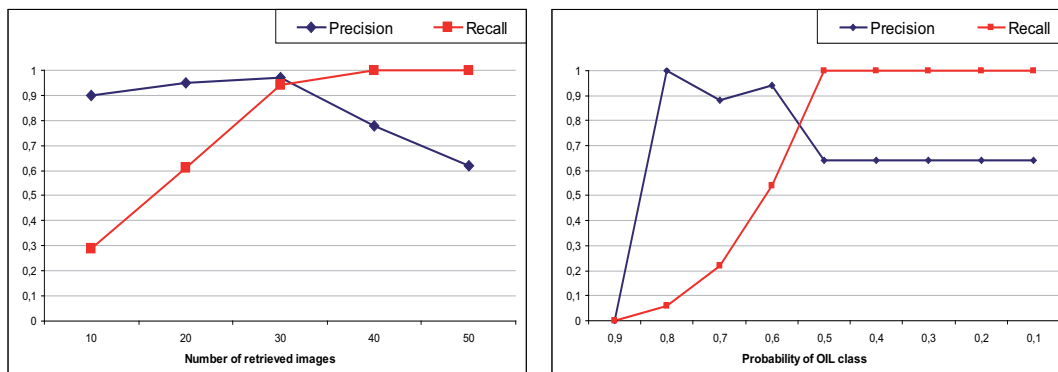
¹Observe that the values of U-satis are an average of responses from evaluators

can be observed that when the probability is 80 percent, the precision reaches the highest value (1). Then it decreases, getting stable with 50 percent.

According to the user opinion, which is also presented in Table 6.6, the *embd0*-based retrieved results are very good (VG). Therefore, there is coincidence between the user criteria and the effectiveness results. Also, a VG user satisfaction has been obtained with the combination of textures at different scales (*tex0* + *tex1*).



(a) PR versus number of retrieved images using all primitive features in the case of Oil class definition.



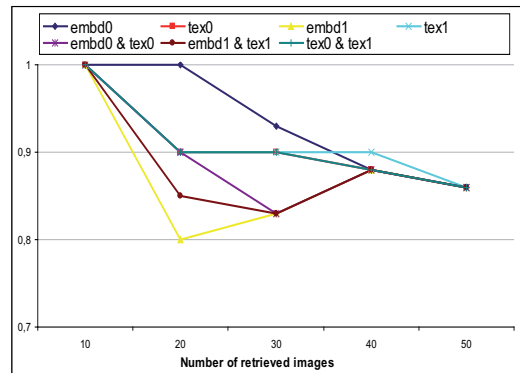
(b) PR versus number of retrieved images using embd0 as primitive feature. (c) PR versus probability of Oil class using embd0 as primitive feature.

Figure 6.14: Effectiveness retrieving Oil class: Precision and Recall curves versus (a) Number of retrieved images from the collection using all primitive features, (b) Number of retrieved images from the collection using intensity (*embd0*) as primitive feature, and (c) probability of oil using intensity (*embd0*) as primitive feature.

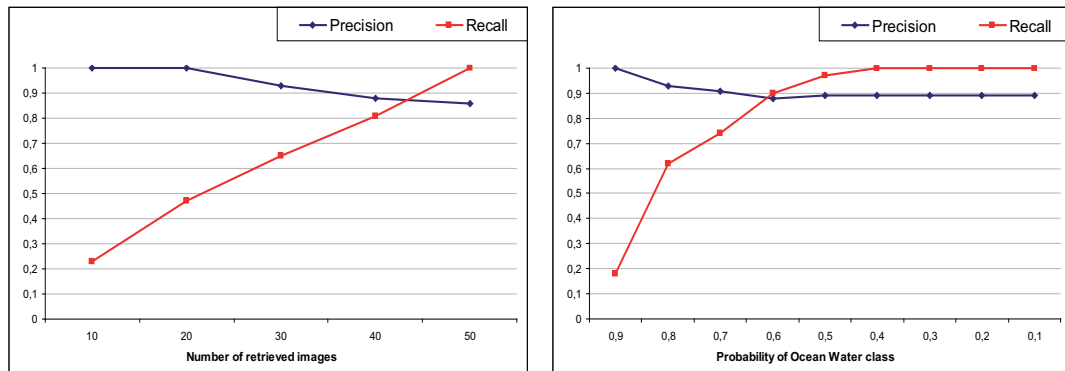
• Ocean water class results

In the case of the ocean water class, in Table 6.6, we notice that the results are similar to the ones obtained with the oil class. In fact, the best primitive feature is intensity at original scale, *embd0*, giving a 100 percent of recognized targets. Moreover, all the primitive features give high effectiveness (more than 80 percent). Figure 6.15(a) shows also that the highest precision corresponds to intensity at original scale *embd0*. This feature was selected to plot the precision and recall values, which are presented in Figure 6.15(b)-(c). Here, the curves show that the system retrieves all the images containing ocean water with a precision superior to 80 percent and

shows high values in their probability (more than 0.8) retrieving this class. The user satisfaction degree is expressed in Table 6.6. The results show that the user is quite satisfied with the retrieved results and there is coincidence between the user criteria and the effectiveness results.



(a) PR versus number of retrieved images using all primitive features in the case of Ocean class definition.



(b) PR versus number of retrieved images using embd0 as primitive feature.

(c) PR versus probability of Ocean class using embd0 as primitive feature.

Figure 6.15: Effectiveness retrieving Ocean class: Precision and Recall curves versus (a) Number of retrieved images from the collection using all primitive features, (b) Number of retrieved images from the collection using intensity as primitive feature, and (c) probability of Ocean class using intensity as primitive feature.

In the case of both classes according to the effectiveness results, the performance of KIM using two classes of level I (see Table 6.3) is quite good (more than 90%).

6.4.1.4 Classification results

The interactive training can be seen as a supervised classification of the whole image content. In the following, we present the classifications of oil and water, using the best primitive feature according to the previous metrics (intensity *embd0* in both cases), and in order to consolidate the previous results, the classification accuracy is evaluated in terms of confusion matrices and user satisfaction degree. Figures 6.16 and 6.17 show the results in the case of Oil and Ocean, respectively.

- Oil class results

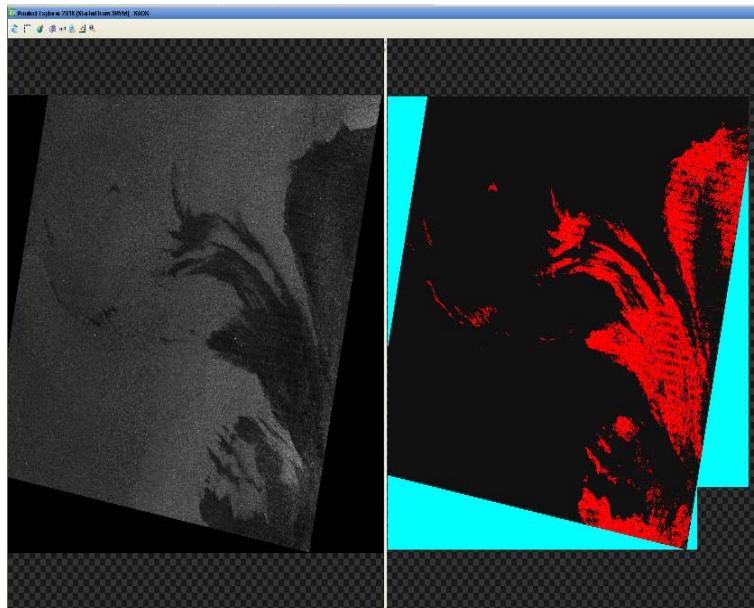


Figure 6.16: Oil classification presented in the Product Explorer. Left part shows the whole TerraSAR-X scene and right part presents the classification results (in red) after the training performed by the user, using intensity (*embd0*) as primitive feature.

In the left part of Figure 6.16, the area recognized as oil is presented in red. Here we can notice that the oil class is well recognized in its majority showing that the extracted primitive features are strong enough for recognizing this class. It is worth mentioning that, this result was obtained with *embd0* as primitive feature, 3 clicks on average, and one iteration with the relevance feedback, taking an approximate time of 6 minutes for the training.

- Ocean water class results

In the case of the ocean water class, three iterations in the probabilistic search were needed in order to obtain an acceptable classification result. The different iterations are presented in Figure 6.17(a)-(c). From the last classification (see Figure 6.17(c)), we can observe that the ocean water class is not well separated. There is a confusion (see left-bottom part) caused by some shadows in the image. This result was obtained with *embd0* as primitive feature, three iterations with the probabilistic search, and 9 clicks on average for each iteration, taking an approximate time of 20 minutes for the training.

Both supervised classifications of the image content were combined into a thematic map of the oil spill disaster as shown in Figure 6.18. The confusion matrix was computed and is presented in Table 6.7. Here, it reports an **overall accuracy of 88.7928** and **Kappa coefficient of 0.6809**.

6.4.1.5 User satisfaction degree

In order to consolidate the user-driven evaluation, a group of 11 evaluators were asked to rank **qualitatively the degree of satisfaction regarding the results of the probabilistic**

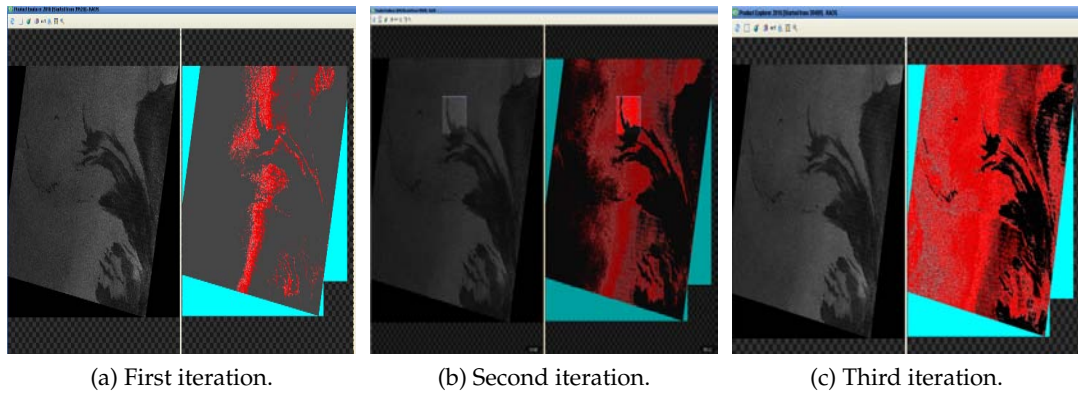


Figure 6.17: Ocean water classification obtained in (a) first (b) second and (c) third iteration of the probabilistic search. The accuracy of the classification is improved with the iterations stopping when the user is satisfied. Intensity image is used as primitive feature. There is a confusion (see left-bottom part) caused by some shadows in the image

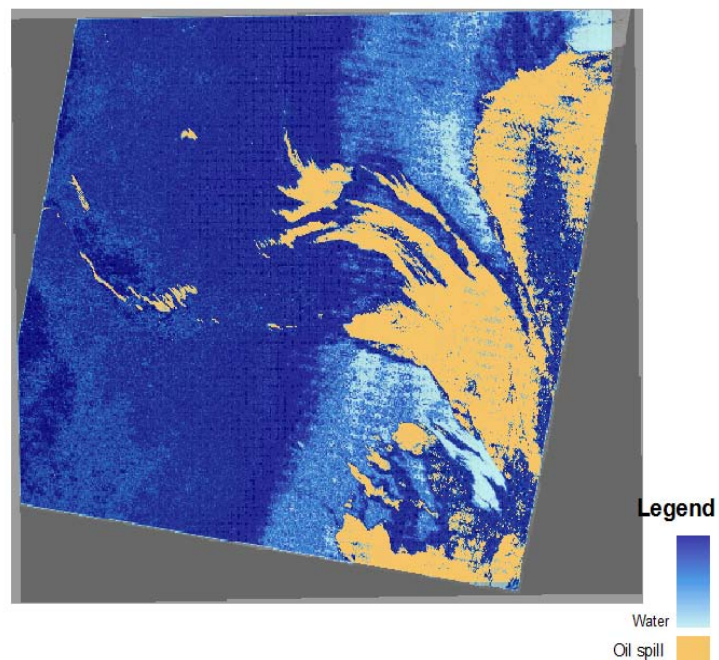


Figure 6.18: Oil spill detection using the KIM System and TerraSAR-X images. The oil is presented in brown and the water in degraded blue.

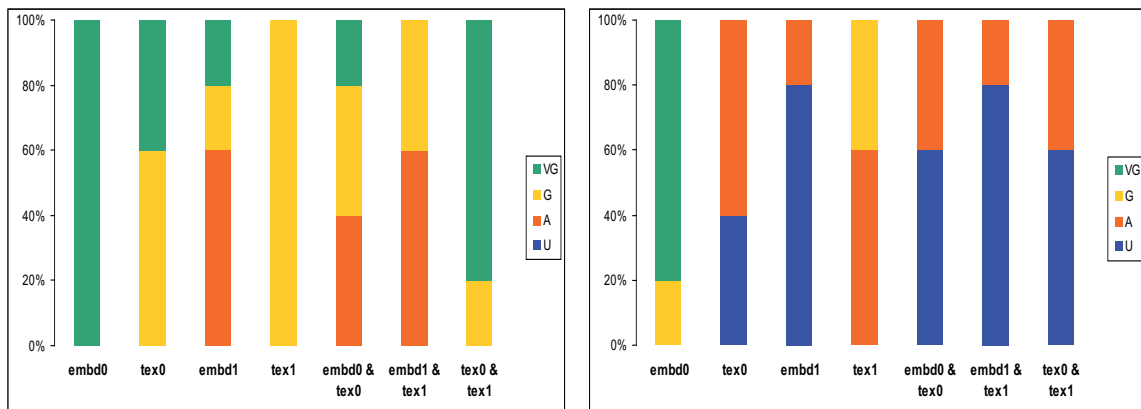
Table 6.7: Confusion matrix of OIL spill detection.

	Oil	Water
Oil	60.90	1.71
Water	39.10	98.29
Total	100	100

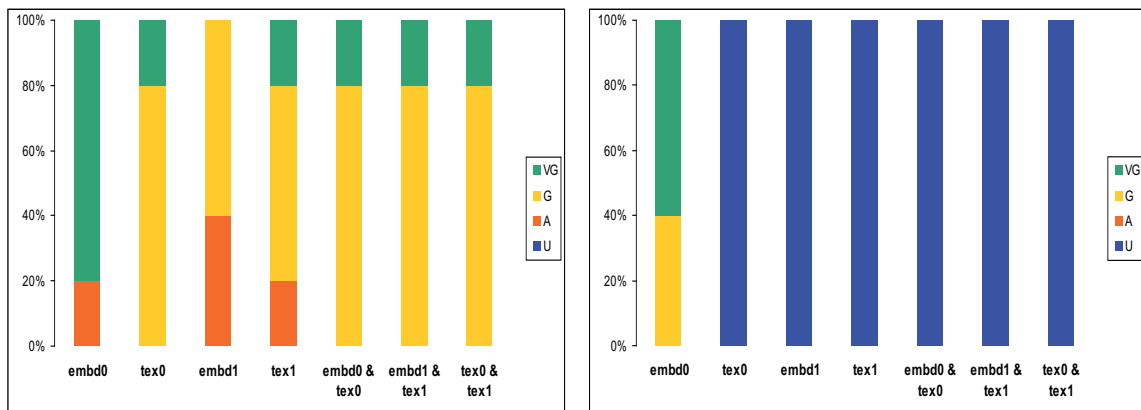
search and the classification, in the case of oil and ocean water classes, using the qualitative scale ranging from low value to high values as follows: Unsatisfied (**U**), Acceptable (**A**), Good (**G**), Very Good (**VG**) according to their satisfaction degree.

In the following, we start by presenting two cases: 1) the user satisfaction degree in the case of both semantic classes versus the different primitive features used during the training, and 2) the user satisfaction degree versus the effectiveness metrics previously computed.

1. User satisfaction degree versus primitive features



(a) Oil class: User satisfaction versus probabilistic search results. (b) Oil class: User satisfaction versus classification results.



(c) Ocean class: User satisfaction versus probabilistic search. (d) Ocean class: User satisfaction versus classification results.

Figure 6.19: User satisfaction degree regarding to (a,c) probabilistic search, and (b,d) classification results in the case of (a,b) OIL semantic class, and (c-d) OCEAN semantic class versus different primitive features. The values of the user satisfaction degree correspond to Unsatisfied (**U**), Acceptable (**A**), Good (**G**), Very Good (**VG**).

Figure 6.19(a) shows that the degree of user satisfaction in retrieving oil images ranges between good and very good, with the exception in the case of the combination of features (intensity and texture). In the case of retrieving ocean water, the user satisfaction is good in most of the cases and using embd0 as primitive feature is very good as is shown

in Figure 6.19(c). However, Figures 6.19(b) and (d) point out that in the case of classification, when only using `embd0` as primitive feature, the satisfaction of the user is very good. The others results are unsatisfying for the user in the case of both classes.

2. Overall user satisfaction degree versus KIM effectiveness

In order to compare the user satisfaction degree (qualitative index) with the effectiveness results of KIM (quantitative metrics), the effectiveness results were converted to a qualitative scale as follows: Unsatisfied (**U**) values between (0 – 25%), Acceptable (**A**) values between (26 – 50%), Good (**G**) values between (51 – 75%), Very Good (**VG**) values between (76 – 100%). The results are presented in Figure 6.20. In the case of Oil class, according to the effectiveness measurements the results are good and very good, however, the user criteria express that most of the results are good and very good but also are acceptable. In the case of Ocean class, the user satisfaction degree ranges between good and very good and low percentage is acceptable, while the effectiveness measurements of all the results are very good.

The global user satisfaction degree compared with the effectiveness metrics presents a moderate coincidence between both results, which ranges between good and very good.

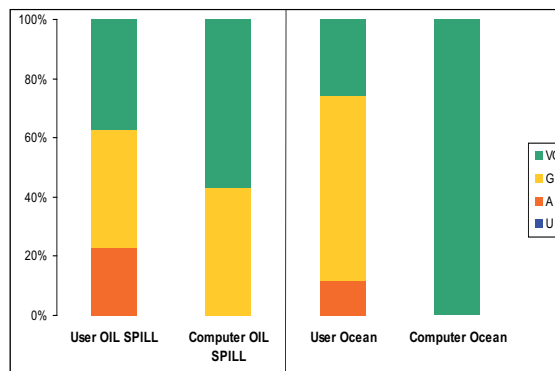


Figure 6.20: User satisfaction degree versus effectiveness of the probabilistic search in the case of both classes: OIL and OCEAN classes. The values of the user satisfaction degree correspond to Unsatisfied (**U**), Acceptable (**A**), Good (**G**), Very Good (**VG**).

6.4.2 Study case 2: Nepal-Embankment breach

On August 18, 2008, the eastern embankment of Kosi river was breached near Kusaha village, Nepal, about 12 km upstream of the barrage. The water inundated the adjoining villages of Kusha and Haripur (Nepal) immediately and reached Bhimnagar, Birpur (India). In Birpur, flood water rose alarmingly to a height of 5 – 6 ft within a short time after the breach. Subsequently, the flood water spread over a great extent covering several districts, causing severe damages to life and property. In fact, after the August 18 breach, the Kosi river has altogether taken a new course (D. V. Reddy and Mandal, 2008).

6.4.2.1 Description of the study area, data set and ground truth

The study area is located in the north of India at the border with Nepal. In this scenario the user is interested in recognizing the water bodies (river and the flooded areas) and verifying if the flooded area can be distinguished from the river water.

- **Data set**

The analysis is based on TerraSAR-X data acquired on August 30, 2008. The image corresponds to **spatially enhanced** single polarization **High Resolution Spotlight products with a Multi-look Ground range Detected**, whose pixel spacing = 0.75 m, range resolution = 1.58 m, azimuth resolution = 1.68 m, and ENL = 2.65 (DLR, 2009). The TerraSAR-X image is shown in Figure 6.21. This scene corresponds to a post-disaster image.

- **Selected parameters**

This TerraSAR-X scene was ingested in KIM system and during the **ingestion** it was divided into 130 tiles with a size of 500×500 pixels each. From each tile the primitive feature extraction was performed by the MAP-GMRF method (see section 5.1.2), whose input parameters are **model order 5 and estimation window size of 30×30 pixels**. We remark that in the data-driven evaluation, previously presented (see section 5), one of the conclusions proved that GMRF has a good performance in detecting natural scenes and it has been recommended to use a model order higher than 3.

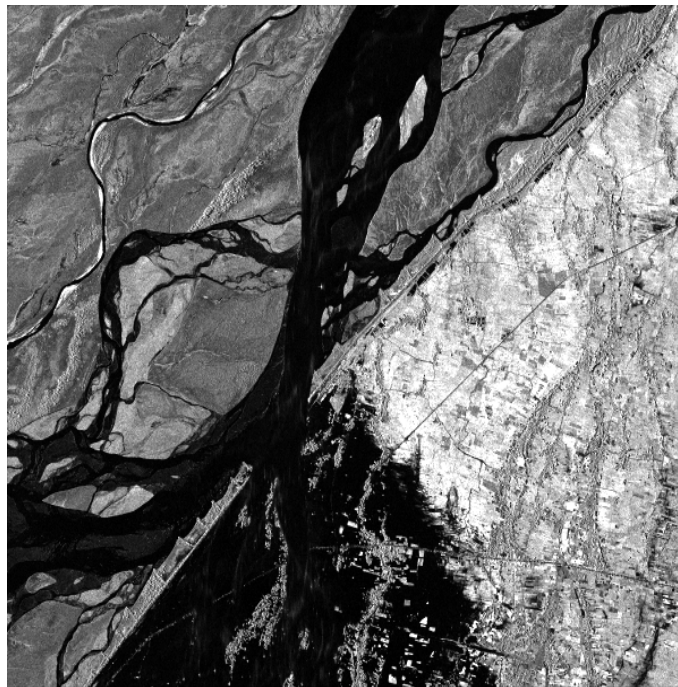


Figure 6.21: TerraSAR-X image over Nepal Embankment breach as seen on August 30, 2008.

- **Ground-truth**

During the **ground-truth definition**, the TerraSAR-X image was divided into 130 tiles in order to have the same tiles as the ones ingested in the KIM system. Each tile was annotated by the user, who judged and decided which classes were contained in the tile according to the specifications provided in section 6.3.2. After the annotation, the information concerning the image tiles, the semantic labels and the relation between them were stored in the reference data base. Table 6.8 summarizes

the ground-truth for this scenario. In this scenario it is worth mentioning that we are interested in determining the precision of classes with level II (see Table 6.3), which is the case of flooded areas and river.

Table 6.8: Resume of Nepal collection ground-truth

Level	Code	Semantic class	Total of relevant images
I	60	Water bodies	
II	61	River	55
II	64	Flooded Area	39
I	30	Agriculture fields	49
I	40	Forest	61

6.4.2.2 Training performed by the end-user

During the interactive training, the end-user can interactively define a semantic label of his interest based on a certain combination of pre-selected **primitive features** such as intensity and texture (see Table 6.1). The **interactive learning** is obtained by giving positive and negative examples by clicking on the image tiles. Figure 6.22 shows examples of the training performed by the user. The red part in the images represents how well the selected primitive features are adapting to the semantic labels. During each training 8 clicks and 20 minutes were needed on average. The different primitive features used in the training sessions are presented in Table 6.9 for each semantic label.

Table 6.9: Semantic class labels and primitive features used in the training performed by the user in the case of flooding scenario.

Semantic labels	Primitive features
River & Flooded areas	embd0
Agriculture & Forest	tex0
	embd1
	tex1
	embd0 + tex0
	embd1 + tex1

6.4.2.3 Evaluation results of the probabilistic search

After defining the semantic labels, the user can search for relevant images based on their content, and KIM performs a probabilistic search giving as result a ranking list ordered according to probability, separability and coverage.

In order to obtain a measurement for the quality of the probabilistic search, we first analyze the queried images according to **relevant** or **irrelevant**. Relevant images are said to be targets. We assume that each retrieved image is either a target or a misclassified one, independently on how strong they are covered by the semantic label. In this thesis, we evaluate the effectiveness of the query results in terms of the measurements presented in section 6.2.1.

In the following, we derived the **effectiveness metrics by comparing the retrieved ranking image list with the ground-truth**, this is presented in Table 6.10. In addition

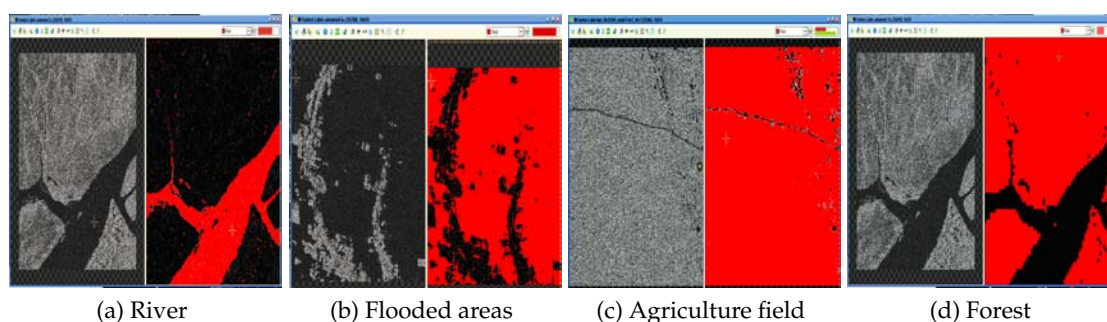


Figure 6.22: Example of user training. When the user is defining (a) a river semantic label, (b) flooded areas, (c) agriculture fields, and (d) forest. The red part in the images represents how well the selected primitive features are adapting to the semantic label.

to these metrics the **user satisfaction degree**² is also presented in Table 6.10. Later, we compute the precision versus the number of retrieved images for all the training sessions performed by the user. Finally, in order to summarize, we select the best results from Table 6.10 and plot the precision and recall values versus the probability and the number of retrieved images. These curves detail the effectiveness of the KIM system retrieving the query images. We start by presenting the results of retrieving water bodies, which include river and flooded areas, later we present the agriculture and forest results.

- **River class results**

In the case of river class, from Table 6.10 it can be noted that the highest recognition rate is 35% and it is obtained when using only intensity (*embd0*). Moreover, the same results can be achieved using texture (*tex1*) and a combination of texture and intensity (*tex1 and embd1*). However, the effectiveness of KIM retrieving images containing only river is very low (under 40 percent).

The precision versus the number of retrieved images for all the training performed by the user is presented in Figure 6.23(a). Here, we point out that the combination of features *tex1 and embd1* gives the highest precision along the retrieved process, which is ranging from 0.35 at 0.56.

According to the user, most of the training sessions are unsatisfied (U), which is confirmed by the low values of the effectiveness metrics, nevertheless the most satisfactory results (acceptable) are obtained using the combination of texture and intensity at two scales (*embd0+tex0, embd1+tex1*) as primitive features. The precision and recall values using intensity and texture at second scale (*embd1+tex1*) are presented in Figures 6.23(b) and 6.23(c). Here, they show a lower precision retrieving this class, only the 56 is reached when 80 images are retrieved.

- **Flooded areas results**

In the case of flooded areas using *embd1*, the intensity image is enough to retrieve 80 percent of the targets, with an accuracy of 79 percent as shown in Table 6.10.

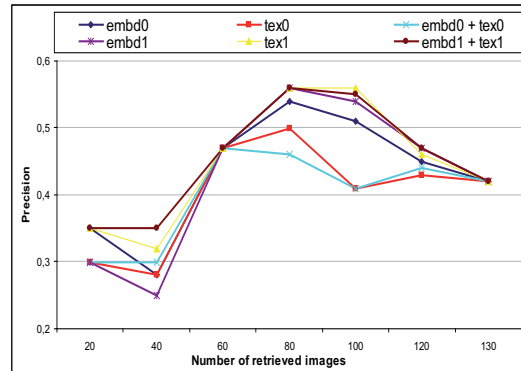
Precision ranging from 60 to 80 percent can be reached when the 40 first images are retrieved using all the primitive features as shown in Figure 6.24(a). In here, it can also be observed that the intensity (*embd0*) feature provides the highest precision.

²Observe that the values of U-satis are an average of responses from evaluators.

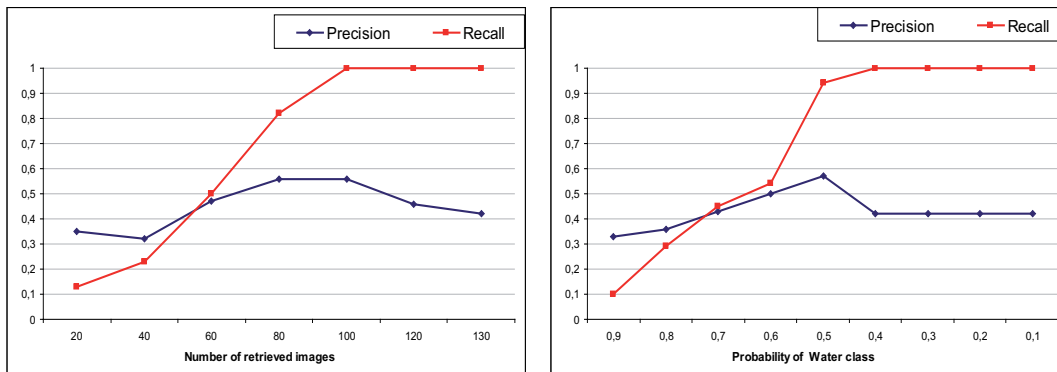
Table 6.10: **Nepal-Embankment breach scenario: Effectiveness evaluation of the probabilistic search** considering the top ranked list. The user satisfaction (U-satis) is expressed as Unsatisfied (U), Acceptable (A), Good (G), and Very Good (VG).

Semantic label: River							
Primitive feature*	Targets	Misclas.	Po	Pf	Accuracy	F-measure	U-satis.
embd0	35,00	65,00	0,14	0,05	0,53	0,19	U
tex0	30,00	70,00	0,17	0,05	0,52	0,16	U
embd0 + tex0	30,00	70,00	0,17	0,05	0,52	0,16	A
embd1	30,00	70,00	0,17	0,05	0,52	0,16	U
tex1	35,00	65,00	0,14	0,05	0,53	0,19	U
embd1 + tex1	35,00	65,00	0,14	0,05	0,53	0,19	A
Semantic label: Flooded areas							
Primitive feature	Targets	Misclas.	Po	Pf	Accuracy	F-measure	U-satis.
embd0	75,00	25,00	0,07	0,05	0,78	0,51	G
tex0	70,00	30,00	0,07	0,05	0,76	0,47	VG
embd0 + tex0	60,00	40,00	0,08	0,05	0,73	0,41	G
embd1	80,00	20,00	0,06	0,05	0,79	0,54	VG
tex1	60,00	40,00	0,08	0,05	0,73	0,41	G
embd1 + tex1	65,00	35,00	0,08	0,05	0,75	0,44	G
Semantic label: Agriculture							
Primitive feature	Targets	Misclas.	Po	Pf	Accuracy	F-measure	U-satis.
embd0	95,00	5,00	0,05	0,05	0,76	0,55	VG
tex0	95,00	5,00	0,05	0,05	0,76	0,55	VG
embd0 + tex0	75,00	25,00	0,07	0,05	0,70	0,43	VG
embd1	80,00	20,00	0,06	0,05	0,72	0,46	VG
tex1	70,00	30,00	0,07	0,05	0,68	0,41	VG
embd1 + tex1	95,00	5,00	0,05	0,05	0,76	0,55	VG
Semantic label: Forest							
Primitive feature	Targets	Misclas.	Po	Pf	Accuracy	F-measure	U-satis.
embd0	55,00	45,00	0,09	0,05	0,55	0,27	A
tex0	55,00	45,00	0,09	0,05	0,55	0,27	VG
embd0 + tex0	60,00	40,00	0,08	0,05	0,56	0,30	G
embd1	60,00	40,00	0,08	0,05	0,56	0,30	G
tex1	45,00	55,00	0,11	0,05	0,52	0,22	G
embd1 + tex1	65,00	35,00	0,08	0,05	0,58	0,32	G

* The primitive features are presented in Table 6.1.**Misclas refers to misclassified metric.



(a) PR versus number of retrieved images using all primitive features in the case of river class definition.



(b) PR versus number of retrieved images using in-

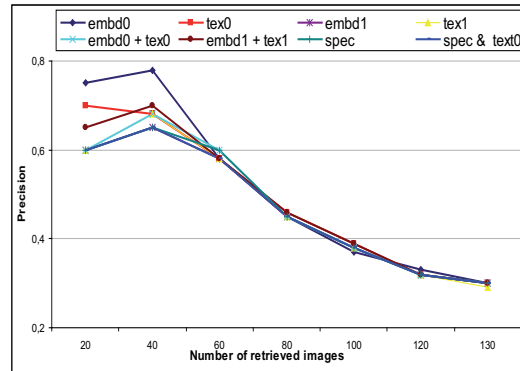
(c) PR versus probability of river class using intensity and texture as primitive features.

Figure 6.23: Effectiveness retrieving river class. (a) Precision of all training session defining River class. (b) Precision and Recall curves versus the number of retrieved images in the collection (c) probability of **River class**, using intensity and texture ($embd1 + tex1$) as primitive features.

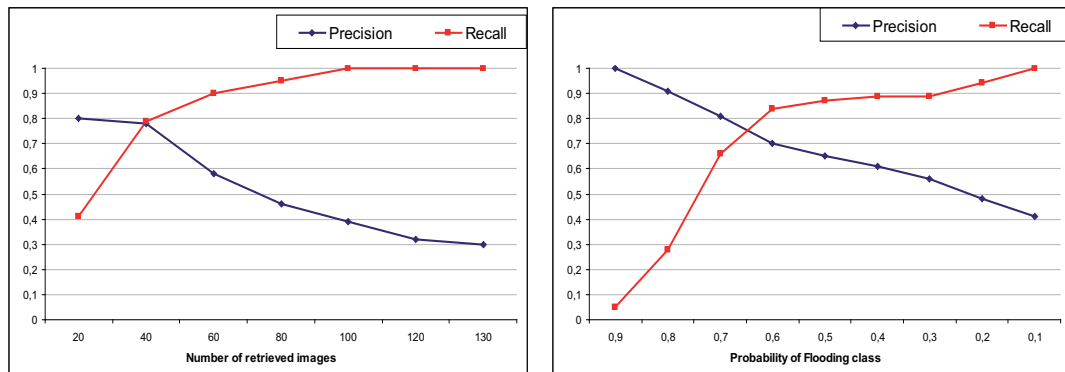
Table 6.10, shows that there is coincidence between the criteria of the user and the effectiveness metrics. Indeed, the best primitive feature is **embd1** in both cases. This result is confirmed in Figure 6.24(b), and Figure 6.24(c), which show that the top ranked retrieved images as flooded area are correct since the precision reaches the highest value (1) with 0.9 of probability, using the intensity image ($embd1$) as primitive feature.

- **Water bodies results (River water and Flooded areas)**

The previous results (river and flooded areas) correspond to second level classes in the hierarchical scheme presented in Table 6.3. Now, we are interested in calculating the effectiveness of KIM retrieving water bodies without distinction of flooding or river. An experiment was carried out in retrieving water bodies (river and flooded areas) using all the primitive features. In here, we computed and plotted the precision for all the primitive features, these results are shown in Figure 6.25(a). From the curves it can be noted that the first 80 images were retrieved with the same precision using all the primitive features, however, the precision of $embd1$, $tex1$ and



(a) PR versus number of retrieved images using all primitive features in the case of flooding class definition.



(b) PR versus number of retrieved images using in- (c) PR versus probability of flooding class using in-
tensity as primitive feature. intensity as primitive feature.

Figure 6.24: Effectiveness retrieving flooded areas. (a) Precision of all training session defining Flooded area class. (b) Precision and Recall curves versus the number of retrieved images in the collection (c) probability of **Flooding** class, using intensity (*embd1*) as primitive feature.

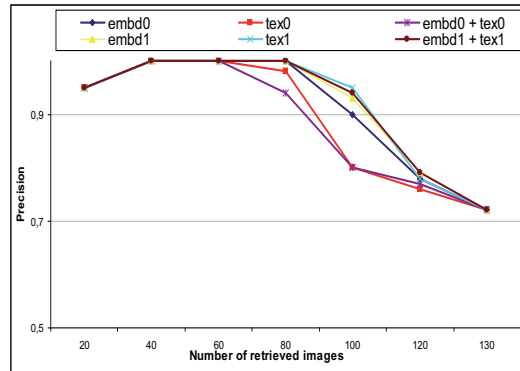
embd1 + tex1 are higher. Even though the highest precision is obtained with **tex1** as primitive feature.

Figures 6.25(b)-(c) plot the precision and recall values using **tex1** as primitive feature. Here it can be observed that precision retrieving water bodies is higher than 0.9; showing a very high performance of KIM retrieving only water bodies, without distinction between river and flooded areas.

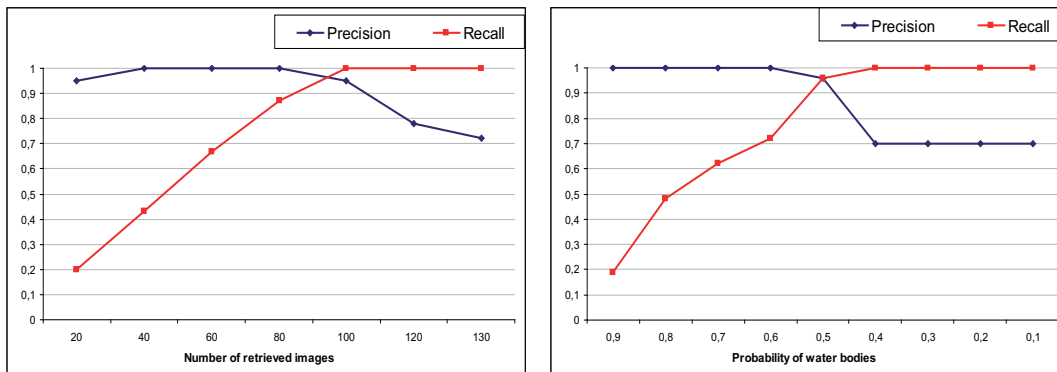
• Agriculture Fields results

From Table 6.10 it can be noted that the best results are obtained using texture (*tex0*) and intensity (*embd0*) separately, and the combination of both at second scale (*embd1 + tex1*). Indeed, it is possible to recognize more than 90 percent of the targets with an accuracy of 76 percent.

The precision of all the primitive features is plotted in Figure 6.26(a). Here it is important to observe that texture (*tex0*) and the combination of texture and intensity (*embd1 + tex1*) provides the highest precision (0.95) along with the increasing num-



(a) PR versus number of retrieved images using all primitive features in the case of water bodies class definition.



(b) PR versus number of retrieved images using texture as primitive feature. (c) PR versus probability of water bodies class using texture as primitive feature.

Figure 6.25: Effectiveness retrieving water bodies. (a) Precision of all training session defining water bodies class. (b) Precision and Recall curves versus the number of retrieved images in the collection (c) probability of **water bodies** class, using texture (*tex1*) as primitive feature.

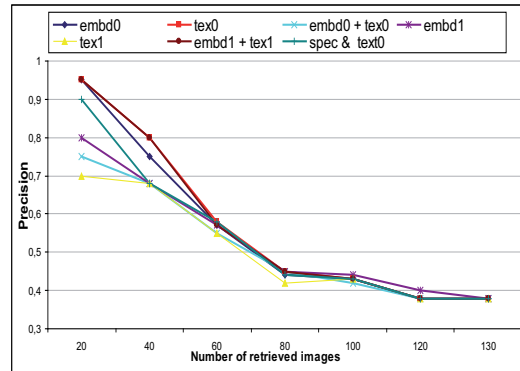
ber of retrieved images. Therefore, we can assume that the contribution of texture is strong enough for recognizing this class.

According to Table 6.10, the user is very satisfied with the results. All the training sessions are marked as very good (VG). The PR curves using intensity and texture (*embd1 + tex1*) as primitive features are presented in Figures 6.26(b) and 6.26(c). Here, it can be noted that, the 40 first images are retrieved with more than 0.8 of probability and when retrieving the 20 first images, the precision is 1, which justifies the satisfaction of the user, since all the retrieved images are correct at first look.

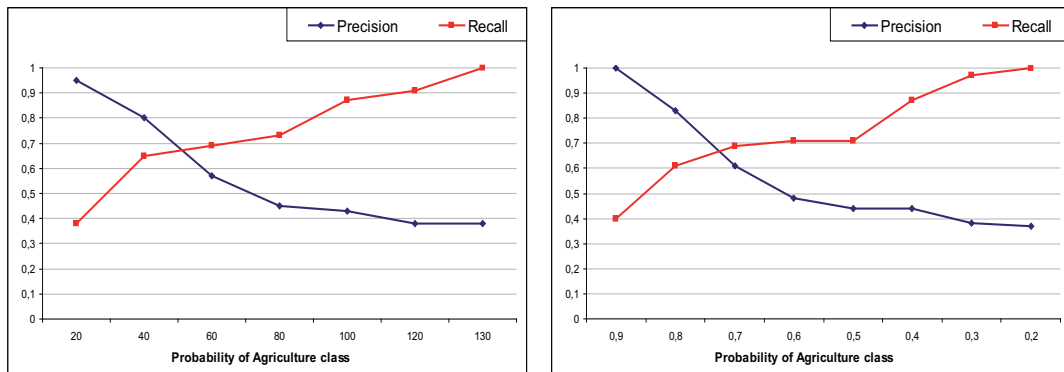
• Forest results

By combining intensity and texture at second scale (*embd1+tex1*), the number of target images containing forest class reach the highest rate (65%). A good effectiveness of KIM can also be concluded from Table 6.10. The other primitive features show lower effectiveness (less than 60%) in the retrieved results.

Figure 6.27(a) depicts the precision for all the training sessions using different prim-



(a) PR versus number of retrieved images using all primitive features in the case of agriculture class definition.



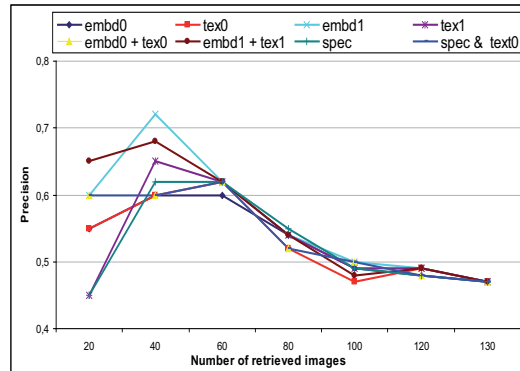
(b) PR versus number of retrieved images using in- (c) PR versus probability of agriculture class using intensity and texture as primitive feature.

Figure 6.26: Effectiveness retrieving agriculture fields. (a) Precision of all training session defining Agriculture fields. (b) Precision and Recall curves versus the number of retrieved images in the collection (c) probability of **Agriculture fields**, using intensity and texture at second scale (*embd1 + tex1*) as primitive feature.

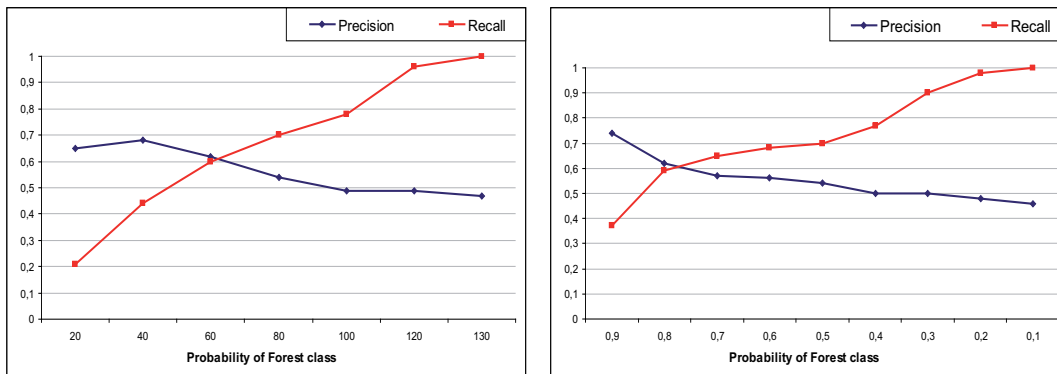
itive features. Here, the highest precision is reached using intensity at second scale (*embd1*) as primitive feature, reaching a precision of 0.7 when retrieving 40 images.

The user marked most of the results as good (G), showing a high satisfaction degree, as is presented in Table 6.10. Here, it can also be observed that all the experiments using texture give to the user a major satisfaction degree; specifically, when using texture (*tex0*) the retrieved images were marked as very good (VG).

The precision of retrieved results based on intensity and texture (*embd1 + tex1*) ranges between 0.6 and 0.7 as is plotted in Figure 6.27(b). Here, it shows 0.7 as the highest precision during the whole experiment. The probability versus the precision of intensity and texture 1 is plotted in Figure 6.27(c), where the top ranked list, with 0.9 probability, gives only a precision of 0.7. Moreover, the precision is decreasing showing a moderate effectiveness of KIM retrieving this class.



(a) PR versus number of retrieved images using all primitive features in the case of forest class definition



(b) PR versus number of retrieved images using in- (c) PR versus probability of forest class using intensity and texture as primitive features.

Figure 6.27: Effectiveness retrieving forest class. (a) Precision of all training session defining Forest class. (b) Precision and Recall curves versus the number of retrieved images in the collection (c) probability of **Forest** class, using intensity and texture at second scale ($embd1 + tex1$) as primitive features.

6.4.2.4 Classification results

After the user had performed the interactive training and is satisfied with the retrieved ranked image list, the results can be seen as a supervised classification of the whole scene. In the following, instead of presenting the classification results for each class, we group them into classification of water bodies (river and flooded areas) and vegetation (agriculture and forest).

The primitive features selected for creating the classification maps were the ones chosen by the user with the highest satisfaction degree shown in Table 6.10. These correspond to texture ($embd1 + tex1$) in the case of River class, intensity ($embd1$) in the case of flooded areas, the combination of intensity and texture ($embd1 + tex1$) in the case of agriculture, and texture ($tex0$) in the case of Forest. The classification results for water bodies and vegetation are shown in Figure 6.28.

Unfortunately, in this scenario it was not possible to derive a confusion matrix due to the unavailability of a ground-truth. Nevertheless, from the visual impressions we can observe that both classes (water bodies and vegetation) are well separated.

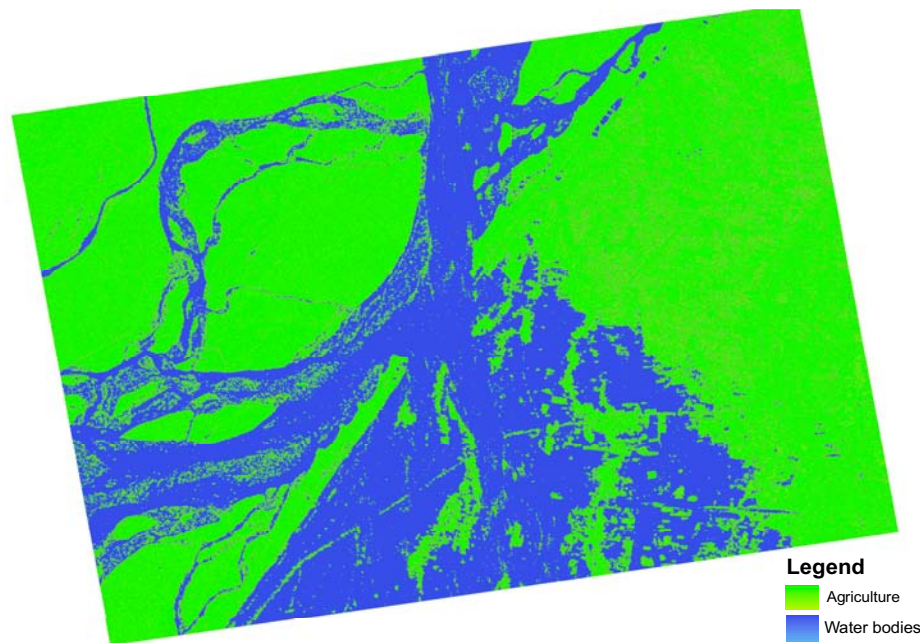


Figure 6.28: Classification results in the case of flooding in Nepal scenario. (blue) Water bodies and (green) vegetation.

6.4.2.5 User satisfaction degree

Since the user opinion is very important in the evaluation, the user was asked to rank the retrieved image list according to his satisfaction degree. Thus, the user gives us the feedback about the interactive training and the performance of the KIM. Furthermore, the group of evaluators judged qualitatively the results of the probabilistic search and the classifications in the case of all defined semantic classes.

In the following, we start by presenting two cases: 1) the user satisfaction degree in the case of all semantic classes, and 2) the user satisfaction degree versus the effectiveness metrics previously computed.

1. User satisfaction degree versus primitive features

In the case of probabilistic search, the degree of user satisfaction regarding the **river class** is unsatisfied as is presented in Figure 6.29(a). The reason could be that the users expected to have scenes only containing river, but KIM retrieved scenes containing water bodies in general. This exposes the necessity of more context and information for retrieving this class. In the case of river classification, the results are depicted in Figure 6.29(b). Here, the user satisfaction presents good results using `embd0`, `tex0` and `tex1` as primitive features.

In the case of the **flooding class**, the probabilistic search results are good and in two cases very good according to the user criteria, this is shown in Figure 6.29(c). On the contrary, the degree of user satisfaction classifying flooded areas ranges between unsatisfied and acceptable as shown in Figure 6.29(d).

The most satisfactory results according to the user criteria are achieved in the case of retrieving **agriculture fields**. This is presented in Figure 6.29(e). Here, all the results are very good showing a high performance of KIM with this class. The user satisfaction degree with agriculture classification is plotted in Figure 6.29(f). In here, we can observe that the user is unsatisfied using intensity and texture (embd0 + tex0) as primitive features. The better classification is obtained either with embd0 or tex0 separately, according to the user criteria.

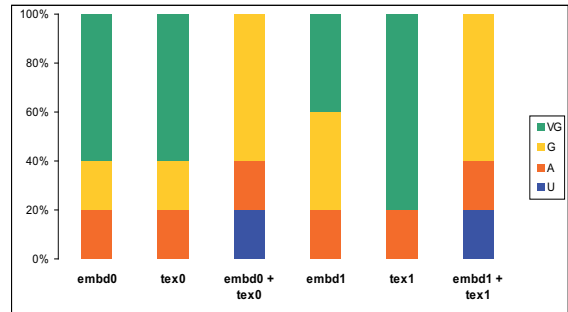
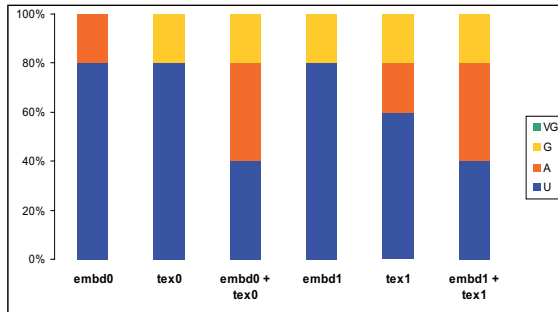
The user satisfaction when retrieving **forest** ranges between acceptable and very good as shown in Figure 6.29(g). Here, we can observe that most of the results are good with exception of when using embd0 as primitive feature where the user satisfaction is acceptable. The highest user satisfaction degree is obtained with texture. On the contrary, most of the results for forest classification are unsatisfactory according to the user. This is depicted in Figure 6.29(h).

2. Overall user satisfaction degree versus KIM effectiveness

In order to compare the user satisfaction degree (qualitative index) with the effectiveness results of KIM (quantitative metrics), the effectiveness results were converted to a qualitative scale as follows: Unsatisfied (**U**) values between (0 – 25%), Acceptable (**A**) values between (26 – 50%), Good (**G**) values between (51 – 75%), Very Good (**VG**) values between (76 – 100%). The results are presented in Figure 6.30(a)-(b). Here, we analyzed the user criteria against the effectiveness metrics 1) in the case of river and flooding classes, and 2) in the case of agriculture and forest classes.

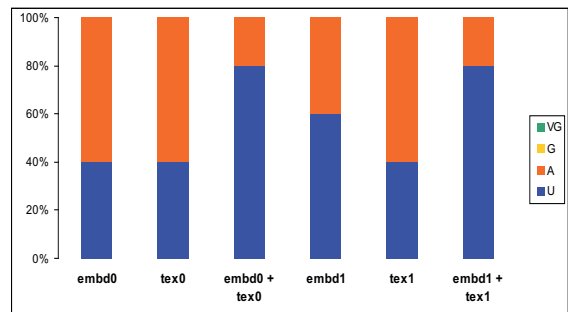
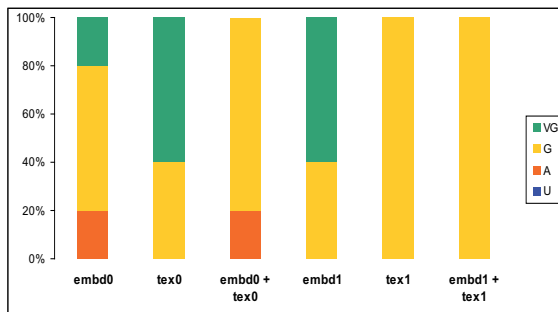
From Figure 6.30(a) it can be noted that in the case of river class, the user is unsatisfied with the retrieved results and that the effectiveness metrics are unsatisfying as well demonstrating that both results are consistent. Moreover, a coincidence between both metrics is also observed in the case of the flooding class, where both results are good.

In the case of agriculture and forest classes, the results are detailed in Figure 6.30(b). Here, it can be observed that the user criteria coincide with the effectiveness metrics. In the case of agriculture both are very good and in the case of forest both are good.



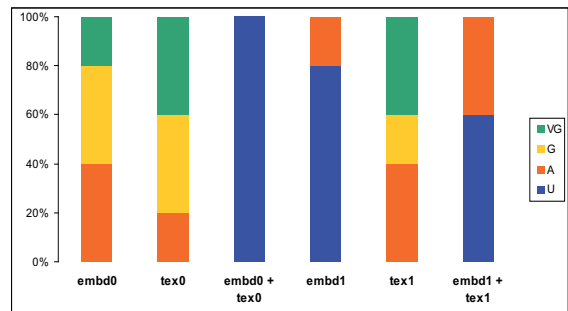
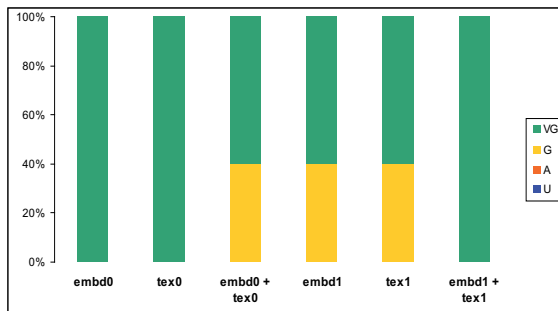
(a) River class: User satisfaction vs. probabilistic search.

(b) River class: User satisfaction vs. classification.



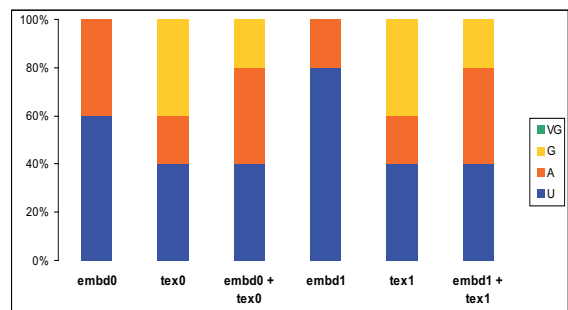
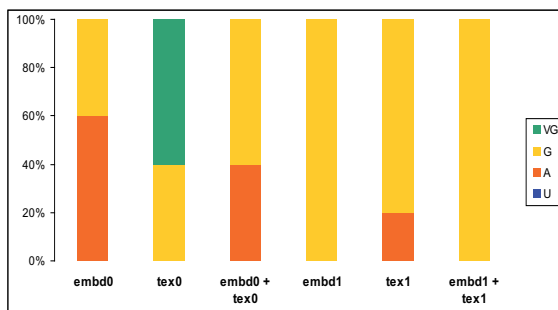
(c) Flooded areas: User satisfaction vs. probabilistic search.

(d) Flooded areas: User satisfaction versus classification.



(e) Agriculture class: User satisfaction vs. probabilistic search.

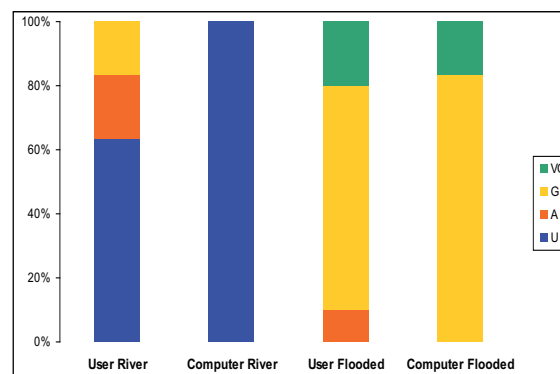
(f) Agriculture class: User satisfaction versus classification.



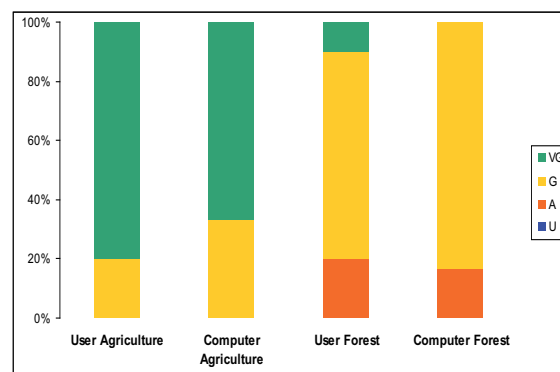
(g) Forest class: User satisfaction vs. probabilistic search.

(h) Forest class: User satisfaction versus classification.

Figure 6.29: User satisfaction degree versus (First column) probabilistic search, and (second column) Classification results in the cases of the (a-b) river, (c-d) flooding, (e-f) agriculture fields, and (g-h) forest semantic classes. The values of the user satisfaction degree correspond to Unsatisfied (U), Acceptable (A), Good (G), Very Good (VG).



(a) River and flooding classes: User satisfaction versus effectiveness of KIM.



(b) Vegetation and forest classes: User satisfaction versus effectiveness of KIM.

Figure 6.30: User satisfaction degree versus effectiveness of KIM in the case of (a) [River and flooding classes, and (b) vegetation and forest classes. The values of the user satisfaction degree correspond to Unsatisfied (U), Acceptable (A), Good (G), Very Good (VG).

6.5 Conclusions

This chapter presents the user-driven evaluation of KIM by creating validation scenarios using TerraSAR-X data, and involving the user in the evaluation process. In the system evaluation methods we distinguish the evaluation of the probabilistic search and the evaluation of the supervised classification of the image content, where both are results from the defined labels by the user. The evaluation was performed in terms of effectiveness of KIM in finding similar image content and retrieving the result. The opinion of the user was qualitatively ranked. In the validation scenarios, the work performed by ZKI motivated us to perform similar products in order to have a reference for comparing the results. Here, two study cases of disaster monitoring were created using KIM. We deduce from the above work the following points:

- In this thesis, we presented the evaluation of the effectiveness of the probabilistic search by means of the precision and recall, targets and misclassified images, accuracy and f-measure, etc. The classification results were evaluated by visual inspection. The evaluator qualitatively judged the classifications. The user satisfaction degree was incorporated as a performance metric.
 - The first case corresponds to the oil spill in the Gulf of Mexico. In here, the user was satisfied with the results, since the satisfaction degree is reported as good (G). The effectiveness metrics show that KIM has an effectiveness of more than 80 percent retrieving classes of Level I.
 - The second study case corresponds to the Nepal-Embankment breach (flooding in Nepal). In here, the user was looking for flooded areas, since KIM retrieved with a high precision the images containing flooded areas, the user was very satisfied ranking the results as good. However, the user was not satisfied with the classification results, since the system recognize well only two classes (water and vegetation), both of level I. In this scenario, it is important to note that the precision of KIM retrieving classes of Level II ranges from (0.2) to (0.5) , but the precision retrieving classes of Level I ranges from (0.7) to 1, which is also confirmed in the first scenario.
 - According to the user criteria, in both scenarios there is no strong coincidence between the satisfaction degree in the probabilistic search and the classification accuracy. This shows that the query and mining capabilities of KIM for various primitive feature combinations are stronger than its capabilities in grouping the classes. Nevertheless, the classification results could be improved by several interactions in the probabilistic search (relevance feedback).
 - Comparing the user criteria with the effectiveness metrics, we can conclude that the user criteria regarding the probabilistic search results reflect the performance of KIM, which in both cases is good.
-

Chapter 7

Conclusions

7.1 Summary of the Thesis

This thesis provides an evaluation of an Image Information Mining (IIM) system using high resolution Synthetic Aperture Radar (SAR) data. The evaluation is mainly based on two approaches. First, the data-driven evaluation, which involves the assessment and validation of despeckling and information extraction methods applied to high resolution SAR images. Second, the user-driven evaluation, which involves the generation of study cases including the end-user in the elaboration and assessment of the results.

1. Data-driven evaluation approach

The very high resolution of SAR imagery is exploited at best with automatic information extraction methods that understand the image content and deal with specific characteristic of this kind of images, i.e speckle, geometry, etc. In this context, the Gibbs Random Field (GRF) family of stochastic models allow to model the image content and to extract information through the parameter estimation. These models are based on parameters that are able to characterize the image content. The suitability of two GRF models Gauss Markov Random Field (GMRF) and Auto-binomial (ABM) for automatic information extraction of high resolution SAR data has been analyzed and evaluated in this work.

Both models are implemented in despeckling and feature extraction methods following the two levels of Bayesian inference. This is a model-based approach, which relies on a prior model of the scene. The first method uses a Gauss Markov Random Field (GMRF) as prior, here it is named MAP-GMRF, and the second is based on an Auto-binomial model (ABM), referred in this thesis as MAP-ABM. Both methods perform the despeckling of an image whilst modelling the image content preserving its texture and structures. They calculate a Maximum *a posteriori* estimation and determine the best model using an evidence maximization algorithm.

The goal of these methods with texture models (GMRF or ABM) is to estimate texture parameters and produce despeckled products at the same time. The estimated parameters can be used for classifications and indexing of the image content. Therefore, during the **data-driven evaluation, the performances of the MAP-GMRF and MAP-ABM methods are assessed in terms of despeckling and parameter extraction qualities.**

The data-driven evaluation has been carried out using the following types of data: 1) synthetically generated SAR data, by means of texture images formed by a known prior model (GMRF or ABM) with its respective parameters, and the likelihood of modelling

the speckle (Gamma distribution). Here, the data contain large-scale structures and small-scale structures. 2) Simulated speckled textures taken from the Brodatz album, which contain different kinds of texture. 3) Real SAR data (TerraSAR-X) follow a known speckle distribution, thus allowing us to prove how well the prior model adapts to the data. These images mainly contain natural-scenes (i.e. vegetation) and man-made structures (i.e. urban area). 4) An application-oriented TerraSAR-X scene containing urban area.

In order to have quantitative results about **the quality of the despeckled images**, several quantities are proposed as quality metrics. These metrics determine 1) the preservation level regarding the details and structures (sharpness), and the mean value (bias) within an image; 2) the level of similarity between two images (structural similarity index); 3) the level of speckle removal without affecting homogeneous areas (smoothness) and inserting distortions (ratio). Moreover, in the case of synthetic SAR data also the mean-squared error is used as a fidelity measurement, and the signal to noise ratio and peak signal to noise ratio are used as quality metrics.

MAP-ABM gives better results in terms of smoothness and sharpness criteria, showing a compromise between despeckling homogeneous areas and detail preservation; whereas MAP-GMRF provides a better result in terms of bias and ratio criteria, showing the mean value of the original images is preserved. Although the results of both methods enhance the image during the despeckling process.

Using stochastic models **the information extraction** is expressed as a parameter estimation problem, where the evaluation focuses on determining the robustness of texture estimated parameters using unsupervised and supervised classifications derived from the extracted texture features, and confusions matrices. This gives us the clues to decide which model to be used according to the kind of structure in the SAR image to be extracted.

The classification results show that both methods are suitable for high resolution SAR information extraction, since the accuracy of the classifications is superior than 60 percent in the case of unsupervised classification and 80 percent in the case of supervised classification for both methods. However, the despeckling and feature extraction performances of MAP-GMRF and MAP-ABM methods depend on the image textures (kind of structure contained in the SAR image), and input model parameters such as selected window size, and the model order of the Markov Random Field. In general, we should estimate the complexity of the scene to select the most convenient model order and window size. In this thesis we used a fixed model order (4) and a window size of (32×32) pixels, which are defined experimentally using trade-off between the complexity and efficiency of the methods. The analysis of the impact of model order and estimation window on despeckling and feature extraction is also evaluated.

- **In the case of MAP-GMRF** it can be concluded that this model adapts the synthetic Brodatz speckled textures well, most of which are homogeneous and with small-scale structures. The synthetically generated texture case shows that the MAP-GMRF can introduce bias, when estimating large-scale structures.

Using real SAR data, the MAP-GMRF method estimates textures well, when its mean value does not change dynamically (homogeneous areas), which often appears in natural scenes, showing that MAP-GMRF is more suitable for SAR scenes containing i.e. vegetation, forest, etc. From the classification experimental results,

we can conclude that GMRF texture parameters are more appropriate for the unsupervised classification case, because they are more stable than the ABM ones, which deviated more around their mean value.

The GMRF model is also less complex, because MAP estimates can be evaluated analytically, whilst ABM requires numerical computation.

- **The MAP-ABM method** estimates well the synthetically-generated textures with ABM model as well as the ones with GMRF model, degraded by the speckle, demonstrating that the ABM model is also able to estimate larger sets of textures than the GMRF model.

In the case of real SAR data, the MAP-ABM method is able to model local statistics and better preserves point-like and blob-like characteristics of the original image, which appear often in high resolution SAR images. Moreover, this method better preserves textures in the man-made scenes and is more suitable for real SAR scenes containing urban area. The supervised classification results show the superiority of the ABM model for extracting textures from both synthetic and real SAR images.

The MAP-ABM method gives better objective results, but it is 3 – 5 times more computationally demanding as the MAP-GMRF method.

The impact of the model order and estimation window size is addressed using different window sizes and model orders for the same image. Despeckling and classification are performed and it has been discovered that by increasing the window size, the information in the image is described in a better way, depending on complexity of the scene. By increasing the model order the reconstructed images become sharper and information extraction can better separate textures and heterogeneous (in terms of mean value dynamic) areas from homogeneous areas. In the case of both methods, a model order 4 with window size of 32×32 pixels is a good compromise between the quality of despeckling, the complexity of the methods and computation time.

2. User-driven evaluation

The user-driven evaluation addresses the problem when an IIM system deals with large databases of very high resolution SAR imagery and the user interacts by searching for the image content, which reflects his interests.

The user-driven evaluation involves the end-user by creating two validation scenarios as study cases in the framework of disaster monitoring. The first validation scenario corresponds to the oil spill in the Gulf of Mexico, and the second one to the flooding in Nepal. In both scenarios, ScanSAR and High Resolution Spotlight TerraSAR-X post-disaster images are used, respectively. Since both scenarios contain natural scenes (vegetation and water bodies), MAP-GMRF is used as feature extraction method during the scenario configuration.

The user-driven evaluation focuses on the effectiveness of the KIM retrieving the scene according to queried content and the user satisfaction degree regarding the probabilistic search and classification. In order to evaluate the effectiveness of KIM the information retrieval metrics such as precision and recall and other ones derived from them are computed. In the case of the user satisfaction degree, the user has been asked to rank the results with a qualitative scale ranging from low value to high values as follows: Unsatisfied (U), Acceptable (A), Good (G), Very Good (VG).

The experimental results shows that the effectiveness of KIM in the case of the oil spill scenario is about 95% in the retrieval results and the degree of user satisfaction range between good and very good. However, the classification results are unsatisfying for the user. Similar results are observed in the case of the flooding detection scenario. Here, the effectiveness of KIM retrieving water bodies without distinction of flooding or river is more than 90%. However, when considering flood and river as two different classes, the effectiveness decreases to 50% – 70%. These results are confirmed with the user satisfaction degree, which is very good in the case of water bodies and between unsatisfied to acceptable in the case of retrieving only the river class. In the case of vegetation (agriculture and forest) the satisfaction degree is between good to very good.

The user criteria point out that, in both scenarios there is no strong coincidence between the satisfaction degree in the probabilistic search and the accuracy of the classification. This can be explained by the fact that the probabilistic search is based on the content of the images without considering the coverage level of the searched content, while the classification is based on pixel content.

7.2 Future work

The data-driven and user-driven evaluation approaches proposed in the frame of this thesis have provided recommendations about the use of different models according to SAR structure as well as the strengths and weaknesses of the evaluated system. In this context, we could propose as a perspective, to include the MAP-ABM method as a feature extraction method in the processing chain of the KIM system since this method shows promising results modeling SAR scenes containing urban areas. However, the current MAP-ABM implementation needs to be improved by developing a method which will find the optimum for estimating the maximum value (G) of the ABM model. The current implementation assumes that the G value is 8 times the mean of the SAR images, which does not correspond in all the cases. Moreover, the quantitative quality despeckling evaluation measurements as for example ratio, bias, smoothness as well as the effectiveness metrics as for example precision and recall could be incorporated in the KIM system as assessment tools.

The user-driven evaluation and the user satisfaction degree pointed out the necessity of improving the pixel based classification results. This introduces the next perspective in our work, which consists in developing proper clustering methods for very high resolution SAR data.

Appendix A

TerraSAR-X image content

Tables A.1-A.2 show typical examples of TerraSAR-X image content. Thus, Table A.1 presents scenes with low backscatterers, which appear in natural areas such as rivers, lakes, and medium brightness patterns which appear in vegetation areas. Locally extended brightness patterns that are typical for urban structures are depicted in Table A.2. Here, some examples of infrastructures are also presented. Most of the sub-scenes were taken from a Munich TerraSAR-X scene, radiometrically enhanced Stripmap product in Multi-look Ground range Detected with single polarization.

Table A.1: Different types of **natural scenes** in TerraSAR-X images (side by side with optical Google-Earth reference images).

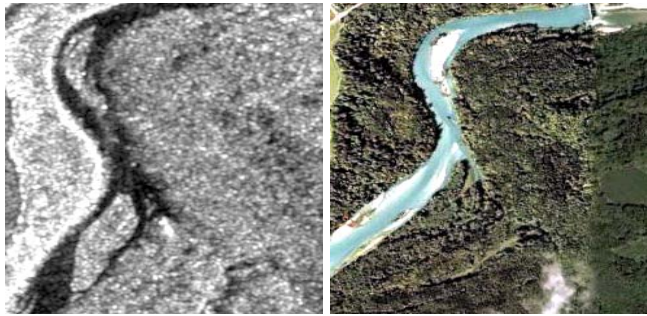
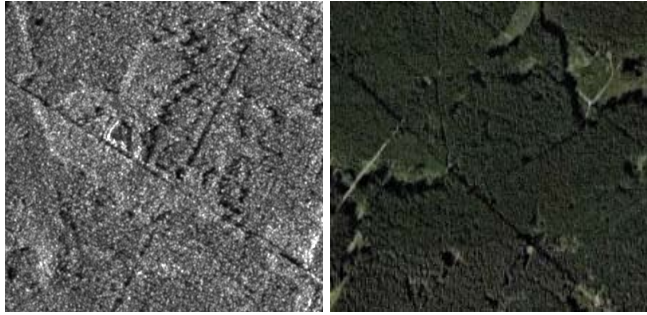

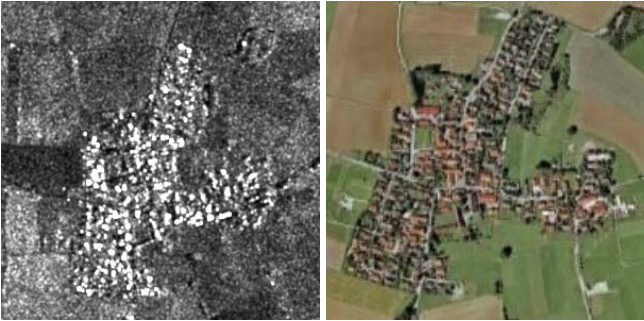
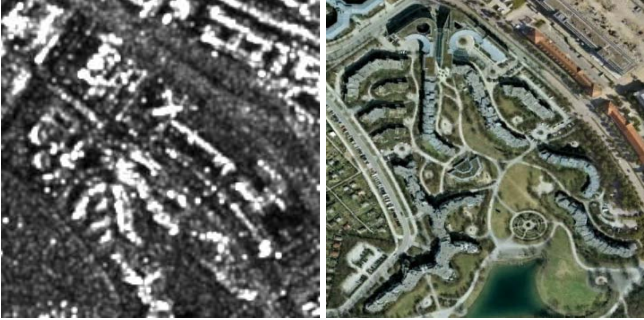

TerraSAR-X sub-scene	Description
<p>(a) TerraSAR-X image (left) and corresponding Google Earth image (right) of a river.</p> 	<p>The image corresponds to the Isar River in Munich. Here, the presence of speckle is moderate and the dark regions are the water surfaces represented in a SAR image. The bright regions correspond to sand and dry areas around the river.</p>
<p>(b) TerraSAR-X image (left) and corresponding Google Earth image (right) of woods and hedgerows.</p> 	<p>The south-west of Munich has a big extension of woods and hedgerows, which are represented in a SAR image with medium brightness and some kinds of texture.</p>

Table A.2: Different types of **man-made scenes** in TerraSAR-X images (side by side with optical Google-Earth reference images).

TerraSAR-X sub-scene	Description
<p>(a) TerraSAR-X image (left) and corresponding Google Earth image (right) of small houses with gardens.</p> 	<p>In this sub-scene there are small houses with gardens. Here the bright scatterers appear like distributed points.</p>
<p>(b) TerraSAR-X image (left) and corresponding Google Earth image (right) of small houses.</p> 	<p>The sub-scene corresponds to a small town in the southern part of Munich. In the center, the urban part appears as bright in the SAR image, since the backscatterers of the houses are strong. Surrounding the city, there are agricultural fields represented as medium bright.</p>
<p>(c) TerraSAR-X image (left) and corresponding Google Earth image (right) of high buildings.</p> 	<p>This image corresponds to high buildings arranged in a form of curve.</p>
<p>(d) TerraSAR-X image (left) and corresponding Google Earth image (right) of a stadium</p> 	<p>This image corresponds to the Allianz Arena stadium in Munich. One of the special features of the TerraSAR-X satellite is the Spotlight Mode, providing images of the Earth's surface with a high resolution of up to one meter, which causes an effect like 3D in the objects in the image.</p>

Appendix B

Evidence maximization approach

The Evidence Maximization (EVM) algorithm is used for the estimation of texture parameters. Each parameter θ_i is changed by a value of $delta$, which is usually set to 0.01. The variable $delta$ has a dimension of the number of model parameters. It also determines the step size of the EVM algorithm. The evidence is computed using the MAP estimate obtained with the initial parameters. Each parameter θ_i is increased or decreased by the value $delta$. If the evidence is increased, the direction of the changing parameter θ_i is kept, otherwise the direction of the changing parameter is changed. The algorithm repeats until the evidence increases.

Algorithm 1 Evidence Maximization Algorithm

```

delta[*] = 0.01
Compute MAP estimate
initialevidence =  $-10^{10}$ ; maxiter = 0;  $\theta_{old} = \theta$ 
repeat
  i=0; oldevidence = initialevidence
  repeat
     $\theta_i = \theta_i + delta[i]$ 
    Estimate evidence, GMRF (5.11), ABM (5.19)
    if evidence > initialevidence then
      initialevidence = evidence
    else
       $\theta_i = \theta_i - delta[i]$ 
       $delta[i] = -delta[i]$ 
    end if
    i=i+1
  until i < numberpara
  maxiter=maxiter+1
  if initialevidence > oldevidence then
     $\theta_{old} = \theta$ 
  else
     $\theta = \theta_{old}$ 
  end if
until (initialevidence > oldevidence) and (maxiter < 2000)

```

Appendix C

Validation scenarios examples

Several scenarios were created in the framework of the Partner Information Mining Services System (PIMS) project. One of the goals of the project was to validate PIMS with TerraSAR-X data. This entails the use of images with worldwide coverage showing the typical TerraSAR-X diversity of information content. Then our work contributed to methods for accessing the TerraSAR-X data archive, to assess the quality and information content of TerraSAR-X products, and to perform classifications, target detection and structure discovery in the observed scenes.

During the validation and evaluation more than 50 TerraSAR-X scenes, in Spotlight and Stripmap mode using Geo-coded Ellipsoid Corrected (GEC) and Multi-look Ground range Detected (MGD) products were used in order to create the following study cases: flooding and urban areas detection, environment and security applications, detection of artifacts in Digital Elevation Model (DEM).

In the study case of urban area detection, the evaluation was performed on more than 40 scenes, of 34 cities. The data ingested correspond to the following cities: Anshan, Bangkok, Basel, Berlin, Bogota, Bremen, Cologne, Czestochowa Daejeon, Dalian, Dubai, Havanna, Jinan, Kuala Lumpur, Las Vegas, Lodz, Lyon,, Moscow, Munich, Oldenburg, Oran, Perm, Pune, Reno, Rostov, Shenyang, Tashkent, Tiajin, Timisoara, Tokyo, Tula, Venice, Washington. In this scenario with various combinations of primitive features it is possible to define more than 20 categories of structures or objects. Figure C.1 shows an example of recognized man-made structures.

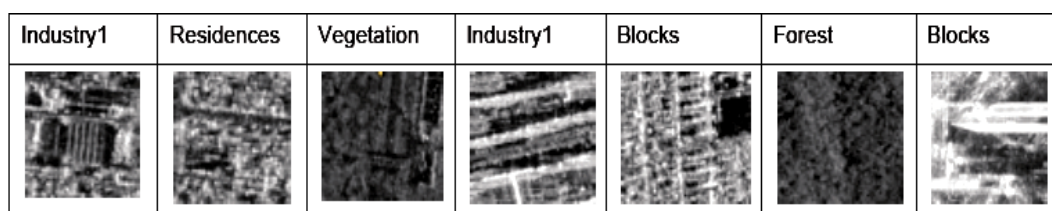


Figure C.1: Examples of man-made structures identified in TerraSAR-X sub-scenes.

In the study case of detection of artifacts in DEM, the evaluation aimed at the assessment of KIM possibilities to describe DEM geomorphology and detect artifacts. KIM was validated and evaluated with more than 30 SRTM DEMs, selected on areas with broad geomorphological diversity. A complete description of several validation scenarios related to flooding, DEM ingestion and examples using synthetic data refers to the ESA project (ESA, 2009).

Acronyms

ABM	Auto-binomial model
CBIR	Content-Based Image Retrieval
CV	Coefficient of Variation
DEM	Digital Elevation Model
DLR	German Aerospace Center
EO	Earth-Observation
ENL	Equivalent Number of Looks
EVM	Evidence Maximization approach
GEC	Geo-coded ellipsoid corrected
GLCM	Gray Level Co-occurrence Matrices
GRF	Gibbs Random Field
GMRF	Gauss Markov Random Field
GUI	Graphical User Interface
IIM	Image Information Mining
KIM	Knowledge-based Image Information Mining
MAP	Maximum a <i>Posteriori</i>
MAP-GMRF	Maximum a Posteriori despeckling using Gauss Markov Random Field and evidence maximization
MAP-ABM	Maximum a Posteriori despeckling using Auto-binomial model and evidence maximization
MGD	Multi-look Ground range Detected
MLE	Maximum Likelihood
MMSE	Minimum Mean Squared Error
MRFs	Markov Random Fields
PIMS	Partner Information Mining Services System
SAR	Synthetic Aperture Radar
UTM	Universal Transversal Mercator
ZKI	Center for satellite-based Crisis Information

Bibliography

- Achim, A., Kuruoglu, E., and Zerubia, J. (2006). SAR image filtering based on the heavy-tailed Rayleigh model. *IEEE Transactions on Image Processing*, 15(9):2686–2693.
- Achim, A., Tsakalides, P., and Bezerianos, A. (2003). SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 41(8):1773 – 1784.
- Agouris, P., Carswell, J., and Stefanidis, A. (1999). An environment for content-based image retrieval from large spatial databases. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(4):263 – 272.
- Aksoy, S. (2001). *A Probabilistic Similarity Framework for Content-Based Image Retrieval*. PhD thesis, University of Washington, Seattle, WA.
- Anderson, J. R., Hardy, E. E., and Roach, J. T. (1972). *A Land Use And Land Cover Classification System For Use With Remote Sensor Data*. U.S. Geological Survey Professional paper 671.
- Argenti, F. and Alparone, L. (2002). Speckle removal from SAR images in the undecimated wavelet domain. *IEEE Transactions on Geoscience and Remote Sensing*, 40(11):2363 – 2374.
- Argenti, F., Bianchi, T., and Alparone, L. (2006). Multiresolution MAP Despeckling of SAR Images Based on Locally Adaptive Generalized Gaussian pdf Modeling. *IEEE Transactions on Image Processing*, 15(11):3385 –3399.
- Bamler, R. and Schättler, B. (1993). *SAR Geocoding: Data and Systems*, chapter SAR Data Acquisition and Image Formation, pages 53–102. Wichmann-Verlag, Karlsruhe-Germany.
- Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society*, 36(2):192–236.
- Besag, J. (1986). On the Statistical Analysis of Dirty Pictures. *Journal of the Royal Statistical Society*, 48(3).
- Besag, J. (1989). Towards Bayesian image analysis. *Journal of Applied Statistics*, 16(3):395–407.
- Breit, H., Eineder, M., Fritz, T., Schattler, B., Huber, M., and Mittermayer, J. (2006). TerraSAR-X Products and Product Performance Update. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1921 –1921.
-

- Brent, R. P. (1973). *Algorithms for Minimization without Derivatives*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Brodatz, P. (1966). *Textures: A Photographic Album for Artist and Designers*. Dover Publications, New York.
- Burl, M. C., Fowlkes, C., and Roden, J. (1999). Mining for Image Content. In *Systemics, Cybernetics, and Informatics / Information Systems: Analysis and Synthesis*, Orlando, FL, USA.
- Casella, G. and George, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistician*, 46(3):167–174.
- Chellappa, R. and Chatterjee, S. (1985). Classification of textures using Gaussian Markov random fields. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 33(4):959 – 963.
- Chellappa, R. and Jain, A. (1993). *Markov Random Fields: Theory and Application*. Academic Press.
- Clausi, D. and Jernigan, M. (1998). A fast method to determine co-occurrence texture features. *IEEE Transactions on Geoscience and Remote Sensing*, 36(1):298 –300.
- Clausi, D. and Yue, B. (2004). Comparing cooccurrence probabilities and Markov Random Fields for texture analysis of SAR sea ice imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 42(1):215 – 228.
- Cleverdon, C. W. (1991). The significance of the Cranfield tests on index languages. In *Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '91*, pages 3–12, New York, NY, USA. ACM.
- Cleverdon, C. W., Mills, J., and Keen, E. M. (1966). Factors determining the performance of indexing systems. Technical report, University of Cranfield, Aslib Cranfield Research Project.
- Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603 –619.
- Cox, I. J., Miller, M. L., Minka, T. P., Papathomas, T. V., and Yianilos, P. N. (2000). The Bayesian Image Retrieval System, PicHunter: Theory, Implementation, and Psychophysical Experiments. *IEEE Transactions on Image Processing*, 9(1):20–37.
- Croft, B., Metzler, D., and Strohman, T. (2010). *Search Engines: Information Retrieval in Practice*. Addison-Wesley.
- Cross, G. R. and Jain, A. K. (1983). Markov Random Field Texture Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(1):25 –39.
- Cui, J. and Wong, W. (2006). The Adaptive Chirplet Transform and Visual Evoked Potentials. *IEEE Transactions on Biomedical Engineering*, 53(7):2818–2835.
- D. V. Reddy, D. Kumar, D. S. and Mandal, M. K. (2008). The 18 August 2008 Kosi river breach: an evaluation. *Current Science*, 95 (12):1168–1669.
-

- da Cunha, A., Zhou, J., and Do, M. (2006). The Nonsampled Contourlet Transform: Theory, Design, and Applications. *IEEE Transactions on Image Processing*, 15(10):3089–3101.
- Dai, M., Peng, C., Chan, A., and Loguinov, D. (2004). Bayesian wavelet shrinkage with edge detection for SAR image despeckling. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1642–1648.
- Daschiel, H. and Datcu, M. (2005). Information Mining in Remote sensing image archives: system evaluation. *IEEE Transactions on Geoscience and Remote Sensing*, 43(1):188–199.
- Datcu, M., Daschiel, H., Pelizzari, A., Quartulli, M., Galoppo, A., Colapicchioni, A., Pastori, M., Seidel, K., Marchetti, P., and D’Elia, S. (2003). Information mining in remote sensing image archives: system concepts. *IEEE Transactions on Geoscience and Remote Sensing*, 41(12):2923–2936.
- Datcu, M., D’Elia, S., King, R., and Bruzzone, L. (2007). Introduction to the special section on image information mining for earth observation data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(4):795–798.
- Datcu, M., Seidel, K., and Walessa, M. (1998). Spatial Information Retrieval from Remote-sensing Images Part I: Information Theoretical Perspective. *IEEE Transactions on Geoscience and Remote Sensing*, 36(5):1431–1445.
- Datta, R., Joshi, D., Li, J., James, and Wang, Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40(2):1–60.
- De Vries, F. P. (1998). Speckle reduction in SAR Imagery by various multi-look techniques. Technical report, FEL-96-A01, TNO Physics and Electronics Laboratory.
- Derin, H. and Elliott, H. (1987). Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(1):39–55.
- Descombes, X. and Zhizhina, E. (2008). The Gibbs fields approach and related dynamics in image processing. *Condensed Matter Physics*, 11(2(54)):293–312.
- Deselaers, T., Keysers, D., and Ney, H. (2004). Classification error rate for quantitative evaluation of content-based image retrieval systems. In *Proceedings of the 17th International Conference on Pattern Recognition*. ICPR, volume 2, pages 505–508.
- DLR (2009). *TerraSAR-X Description our Products*. "<http://www.dlr.de/TerraSAR-X/>".
- Dowe, J. (1993). Content-based retrieval in multimedia imaging. In *Proceedings of SPIE Storage and Retrieval for Image and Video Databases II*, volume 2, pages 164–167.
- Dyson, M. C. and Box, H. (1997). Retrieving symbols from a database by their graphic characteristics: Are users consistent? *Journal of Visual Languages and Computing*, 8(1):85–107.
-

- Eakins, J. and Graham, M. (1999). Content-based image retrieval. Technical Report 39, University of Northumbria at Newcastle, JISC Technology Applications Programme.
- ESA (2009). *Partner Information Mining Services System*. <http://earth.eo.esa.int/rtd/Projects/PIMS-DLR/index.html>.
- Espinoza-Molina, D. and Datcu, M. (2010). Impact of Model Order and Estimation Windows for indexing TerraSAR-X images. In *Proceedings of SPIE Image and Signal Processing for Remote Sensing XVI*, volume 7829, pages 7–15.
- Espinoza-Molina, D., Datcu, M., and Schwarz, G. (2009). Experience gained with texture modeling and classification of 1 meter resolution SAR images. In *Proceedings of SPIE Image and Signal Processing for Remote Sensing XV*, volume 7477.
- Espinoza-Molina, D., Gleich, D., and Datcu, M. (2010a). Assessment of two Gibbs random field based feature extraction methods for SAR images using a Cramer-Rao bound. In *8th European Conference on Synthetic Aperture Radar EUSAR 2010*.
- Espinoza-Molina, D., Gleich, D., and Datcu, M. (2010b). Gibbs Random Field Models for Model-Based Despeckling of SAR Images. *IEEE Letters on Geoscience and Remote Sensing*, 7(1):73–77.
- Espinoza-Molina, D., Gleich, D., and Datcu, M. (2011). Evaluation of Bayesian despeckling and texture extraction methods based on Gauss Markov and Auto-binomial Gibbs random fields: Application to TerraSAR-X data. *IEEE Transaction on Geoscience and Remote Sensing*, to be published.
- Fjortoft, R., Lopes, A., and Adragna, F. (2000). Radiometric and spatial aspects of speckle filtering. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 4, pages 1660 – 1662.
- Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., and Yanker, P. (1995). Query by Image and Video Content: The QBIC System. *IEEE Computer Special Issue on Content-Based Retrieval*, 28(9):23–32.
- Foucher, S., Benie, G., and Boucher, J.-M. (2001). Multiscale MAP filtering of SAR images. *IEEE Transactions on Image Processing*, 10(1):49 –60.
- Frost, V., Stiles, J., Shanmugan, K., and Holtzman, J. (1982). A Model for Radar Images and its Application to Adaptive Digital Filtering of Multiplicative Noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-4(2):157–166.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741.
- Gevers, T. and Smeulders, A. W. M. (2000). Pictoseek: Combining color and shape invariant features for image retrieval. *IEEE Transactions on Image Processing*, 9(1):102–119.
- Gleich, D. and Datcu, M. (2007). Wavelet-Based Despeckling of SAR Images Using Gauss-Markov Random Fields. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):4127 –4143.
-

- Gleich, D. and Datcu, M. (2009). Wavelet-Based SAR Image Despeckling and Information Extraction, Using Particle Filter. *IEEE Transactions on Image Processing*, 18(10):2167 – 2184.
- Gleich, D., Kseneman, M., and Datcu, M. (2010). Despeckling of TerraSAR-X Data Using Second-Generation Wavelets. *IEEE Letters on Geoscience and Remote Sensing*, 7(1):68 –72.
- Goodman, J. W. (1975). Statistical properties of laser speckle patterns. In *Laser Speckle and Related Phenomena*, volume 9 of *Topics in Applied Physics*, pages 9–75, Berlin. Springer-Verlag.
- Griffiths, J. R., Johnson, F., and Hartley, R. J. (2007). User satisfaction as a measure of system performance. *Journal of Librarianship and Information Science*, 39(3):142–152.
- Grubinger, M. (2007). *Analysis and Evaluation of Visual Information Systems Performance*. PhD thesis, School of Computer Science and Mathematics Faculty of Health, Engineering and Science, Victoria University.
- Gupta, A. and Jain, R. (1997). Visual Information Retrieval. *Communication of the ACM*, 40(5):70–79.
- Hahmann, T., Roth, A., Martinis, S., Twele, A., and Gruber, A. (2008). Automatic Extraction of Water Bodies from TerraSAR-X Data. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, volume 3, pages 103 –106.
- Hammersley, J. M. and Clifford, P. (1971). Markov Field on Finite Graphs and Lattices.
- Haralick, R. (1979). Statistical and Structural Approaches to Texture. In *Proceedings of the IEEE*, volume 67, pages 786–804.
- Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6):610–621.
- Harman, D. (1993). Overview of the first TREC conference. In *Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 36–47, New York, NY, USA. ACM.
- Hassner, M. and Sklansky, J. (1980). The use of Markov Random Fields as models of texture. *Computer Graphics and Image Processing*, 12(4):357 – 370.
- Hebar, M., Gleich, D., and Cucej, Z. (2009). Autobinomial Model for SAR Image Despeckling and Information Extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 47(2):2818–2835.
- Herrera-Cruz, V. and Koudogbo, F. (2009). TerraSAR-X rapid mapping for Flooding events. In *ISPRS Hannover Workshop 2009: High-Resolution Earth Imaging for Geospatial Information*, volume XXXVIII-1-4-7/W5.
- Hsu, W., Lee, M. L., and Zhang, J. (2002). Image Mining: Trends and Developments. *Journal of Intelligent Information Systems*, 19(1):7–23.
-

- Huang, X., Zhang, L., and Wang, L. (2009). Evaluation of Morphological Texture Features for Mangrove Forest Mapping and Species Discrimination Using Multispectral IKONOS Imagery. *IEEE Letters on Geoscience and Remote Sensing*, 6(3):393–397.
- Huijismans, D. P. and Sebe, N. (2005). How to Complete Performance Graphs in Content-Based Image Retrieval: Add Generality and Normalize Scope. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2):245–251.
- Jain, A. K. and Vailaya, A. (1996). Image Retrieval using Color and Shape. *Pattern Recognition*, 29(8):1233–1244.
- Jeng, F.-C. and Woods, J. (1991). Compound Gauss-Markov random fields for image estimation. *IEEE Transactions on Signal Processing*, 39(3):683–697.
- Jermyn, I. H., Shaffrey, C. W., and Kingsbury, N. G. (2002). Evaluation Methodologies For Image Retrieval Systems. In *Proceedings of Advanced Concepts for Intelligent Vision Systems*, pages 9–11.
- Joemon, J. M., Furner, J., and Harper, D. (1998). Spatial querying for image retrieval: a user-oriented evaluation. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '98, pages 232–240.
- Kachouri, R., Djemal, K., Maaref, H., Masmoudi, D., and Derbel, N. (2008). Feature extraction and relevance evaluation for heterogeneous image database recognition. In *First Workshops on Image Processing Theory, Tools and Applications. IPTA 2008.*, pages 1–6.
- Kandaswamy, U., Adjeroh, D., and Lee, M. (2005). Efficient Texture Analysis of SAR Imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 43(9):2075–2083.
- Karkanis, S., Magoulas, G., Iakovidis, D., Karras, D., and Maroulis, D. (2001). Evaluation of textural feature extraction schemes for neural network-based interpretation of regions in medical images. In *Proceedings of International Conference on Image Processing*, volume 1, pages 281–284.
- Kato, T. (1991). Cognitive view mechanism for multimedia database system. In *Proceedings of First International Workshop on Interoperability in Multidatabase Systems*, pages 179–186.
- Kato, T. (1992). Database architecture for content-based image retrieval. In *Proceedings of SPIE Conference on Image Storage Retrieval Systems*, volume 1662, pages 112–123.
- Kowalski, G. and Maybury, M. T. (2000). *Information Retrieval Systems: Theory and Implementation*. Springer, second edition.
- Kuan, D., Sawchuk, A., Strand, T., and Chavel, P. (1985). Adaptive Noise Smoothing Filter for Images with Signal-Dependent Noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):165–177.
- Kuan, D., Sawchuk, A., Strand, T., and Chavel, P. (1987). Adaptive Restoration of Images with Speckle. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35(3):373–383.
-

- Lee, J. (1980). Digital Image Enhancement and Noise Filtering by Use of Local Statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(2):165–168.
- Lee, J. (1981). Refined Filtering of Image Noise Using Local Statistics. *Computer Graphics and Image Processing*, 15(4):380–389.
- Lew, M. S., Sebe, N., Djeraba, C., and Jain, R. (2006). Content-based multimedia information retrieval: State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2(1):1–19.
- Li, J. and Narayanan, R. (2004). Integrated spectral and spatial information mining in remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 42(3):673–685.
- Li, S. and Shawe-Taylor, J. (2005). Comparison and Fusion of Multiresolution Features for Texture Classification. *Pattern Recognition Letters*, 26(5):633–638.
- Li, S. Z. (1995). *Markov Random Field Modeling in Computer Vision*. Springer-Verlag, London, UK.
- Li, Y. and Bretschneider, T. (2007). Semantic-sensitive satellite image retrieval. *IEEE Transactions on Geoscience and Remote Sensing*, 45(4):853–860.
- Lopes, A., Nezry, E., Touzi, R., and Laur, H. (1993). Structure detection and statistical adaptive speckle filtering in SAR images. *International Journal of Remote Sensing*, 14(9):1735–1758.
- Lopes, A., Touzi, R., and Nezry, E. (1990). Adaptive Speckle Filters and Scene Heterogeneity. *IEEE Transactions on Geoscience and Remote Sensing*, 28(6):992–1000.
- Lu, X., Wang, J., Wang, Z., and Sun, H. (2009). Flooded area detection using multi-temporal TerraSAR-X data. In *2nd Asian Pacific Conference on Synthetic Aperture Radar 2009*, pages 155–159.
- Ma, L. and Shan, Y. (2008). Integration of Fractal and Grey-Level Features for Texture Segmentation. In *Congress on Image and Signal Processing CISP 08*, volume 3, pages 687–691.
- MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297. Berkeley, University of California Press.
- Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693.
- Martin, F. J. and Turner, R. W. (1993). SAR speckle reduction by weighted filtering. *International Journal of Remote Sensing*, 14(9):1759–1774.
- Mason, D., Speck, R., Devereux, B., Schumann, G.-P., Neal, J., and Bates, P. (2010). Flood Detection in Urban Areas Using TerraSAR-X. *IEEE Transactions on Geoscience and Remote Sensing*, 48(2):882–894.
-

- Medeiros, F. N. S., Mascarenhas, N. D. A., and Costac, L. F. (2003). Evaluation of speckle noise MAP filtering algorithms applied to SAR images. *International Journal of Remote Sensing*, 24(24):5197–5218.
- Mercier, G. and Girard-Ardhuin, F. (2006). Partially Supervised Oil-Slick Detection by SAR Imagery Using Kernel Expansion. *IEEE Transactions on Geoscience and Remote Sensing*, 44(10):746–755.
- Minka, T. P. and Picard, R. W. (1996). Interactive Learning with a "Society of Models". In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 447–452, Los Alamitos, CA, USA. IEEE Computer Society.
- MIRA (1999). Evaluation frameworks for interactive multimedia information retrieval applications.
- MIT (2001). *Fort Hood Datasets*. "<http://www.mbvlab.wpafb.af.mil/public/sdms/datasets/fthood/index.htm>".
- Mitchell, T. (1997). *Machine Learning*. McGraw Hill Science, first edition.
- Müller, H., Müller, W., Squire, D. M., Marchand-Maillet, S., and Pun, T. (2001). Performance evaluation in content-based image retrieval: overview and proposals. *Pattern Recognition Letters*, 22(5):593–601.
- Ogle, V. E. and Stonebraker, M. (1995). Chabot: Retrieval from a Relational Database of Images. *Computer*, 28(9):40–48.
- Oliver, C. and Quegan, S. (1998). *Understanding Synthetic Aperture Radar Images*. Artech House, Norwood, MA.
- Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*. New York: McGraw Hill, third edition.
- Pennec, E. L. and Mallat, S. (2005). Sparse Geometric Image Representations with Bandelets. *IEEE Transactions on Image Processing*, 14(4):423–438.
- Pentland, A., Picard, R. W., and Sclaroff, S. (1996). Photobook: content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3):233–254.
- Porcello, L. J., Massey, N. G., Innes, R. B., and Marks, J. M. (1976). Speckle reduction in synthetic-aperture radars. *Journal of the Optical Society of America*, 66(11):1305–1311.
- Raney, R. K. and Wessels, G. J. (1988). Spatial Considerations in SAR Speckle Simulation. *IEEE Transactions on Geoscience and Remote Sensing*, 26 (5):666–672.
- Razniewski, S. and Strzelecki, M. (2005). Evaluation of texture features based on mutual information. In *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis. ISPA*, pages 233 – 238.
- Rignot, E. and Kwok, R. (1990). Extraction of Textural Features in SAR Images: Statistical Model and Sensitivity. In *Proceedings of International Geoscience and Remote Sensing Symposium*, pages 1979–1982, Washington, DC.
-

- Salton, G. and Buckley, C. (1990). Improving retrieval performance by relevance feedback. *Journal of the American Society for Information Science*, 41(4):288–297.
- Saracevic, T. (1995). Evaluation of evaluation in information retrieval. In *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 138–146, New York, NY, USA. ACM.
- Schröder, M. (2000). *Stochastic Modeling of Image Content in Remote Sensing Image Archives*. PhD thesis, University of Ulm.
- Schröder, M., Rehrauer, H., Seidel, K., and Datcu, M. (1998). Spatial Information Retrieval from Remote-Sensing Images Part II: Gibbs Markov Random Fields. *IEEE Transactions on Geoscience and Remote Sensing*, 36(5):1446–1455.
- Schröder, M., Rehrauer, H., Seidel, K., and Datcu, M. (2000). Interactive Learning and Probabilistic Retrieval in Remote Sensing Image Archives. *IEEE Transaction on Geoscience and Remote Sensing*, 38(5):2288–2298.
- Schreier, G. (1993). Geometrical Properties of SAR Images. In G.Schreier (Ed.) *SAR Geocoding: Data and Systems*, pages 103–134. Wichmann-Karlsruhe.
- Sharma, M. and Singh, S. (2001). Evaluation of texture methods for image analysis. In *The Seventh Australian and New Zealand Conference on Intelligent Information Systems*, pages 117 – 121.
- Shatford, S. (1986). Analyzing the subject of a picture: a theoretical approach. *Cataloging and Classification Quarterly*, 6(3):39–61.
- Shirahatti, N. V. and Barnard, K. (2005). Evaluating image retrieval. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 955–961, Washington, DC, USA. IEEE Computer Society.
- Shyu, C.-R., Klaric, M., Scott, G. J., Barb, A. S., Davis, C. H., and Palaniappan, K. (2007). GeoIRIS: Geospatial Information Retrieval and Indexing System-Content Mining, Semantics Modeling, and Complex Queries. *IEEE Transactions on Geoscience and Remote Sensing*, 45(4):839–852.
- Silva, D. S. (1996). *A Bayesian Tutorial*. Oxford: Clarendon Press.
- Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain, R. (2000). Content-Based Image Retrieval at the End of the Early Years. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380.
- Smith, J. R. and Chang, S.-F. (1996). VisualSEEK: A Fully Automated Content-Based Image Query System. In *Proceedings of the fourth ACM international conference on Multimedia*, pages 87–98.
- Smith, J. R. and Chang, S.-F. (1997). Image and video search engine for the world wide web. In *Storage and Retrieval for Image and Video Databases (SPIE)*, pages 84–95.
- Smith, J. R. and Li, C.-S. (1998). Image Retrieval Evaluation. In *Proceedings of the IEEE Workshop on Content - Based Access of Image and Video Libraries*, pages 112 –113.
-

- Solberg, A. H. S., Brekke, C., and Husoy, P. O. (2007). Oil Spill detection in Radarsat and Envisat SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, 45(3):746–755.
- Solberg, S., Anne, H., and Jain, A. (1997). Texture fusion and feature selection applied to SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 35(2):475–479.
- Tague-Sutcliffe, J. M. (1996). Some perspectives on the evaluation of information retrieval systems. *Journal of the American Society for Information Science*, 47(1):1–3.
- Touzi, R. (2002). A Review of Speckle Filtering in the Context of Estimation Theory. *IEEE Transactions on Geoscience and Remote Sensing*, 40(11):2392–2404.
- TREC (1999). Text retrieval conference.
- Tsai, C.-F., McGarry, K., and Tait, J. (2006a). Claire: A modular support vector image indexing and classification system. *ACM Transactions on Information Systems*, 24(3):353–379.
- Tsai, C.-F., McGarry, K., and Tait, J. (2006b). Qualitative evaluation of automatic assignment of keywords to images. *Information Processing and Management*, 42(1):136–154.
- Tuceryan, M. and Jain, A. (1990). Texture segmentation using Voronoi polygons. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(2):211–216.
- Tuceryan, M. and Jain, A. K. (1998). Texture analysis. In Chen, C. H., Pau, L. F., and Wang, P. S. P., editors, *Handbook of pattern recognition & computer vision*, number 2, pages 207–248, River Edge, NJ, USA. World Scientific Publishing Co., Inc.
- Ulaby, F., Moore, R., and Fung, A. K. (1981). *Microwave Remote Sensing: Active and Passive, Vol. I Microwave Remote Sensing Fundamentals and Radiometry*. Addison-Wesley Publishing Company.
- VanTrees, H. L. (1968). *Detection, Estimation and Modulation Theory Part I*. John Wiley and Sons.
- Vasconcelos, N. (2007). From Pixels to Semantic Spaces: Advances in Content-Based Image Retrieval. *Computer*, 40(7):20–26.
- Voigt, S., Kemper, T., Riedlinger, T., Kiefl, R., Scholte, K., and Mehl, H. (2007). Satellite Image Analysis for Disaster and Crisis-Management Support. *IEEE Transactions on Geoscience and Remote Sensing*, 45(6):1520–1528.
- Voorhees, E. M. and Harman, D. (1998). Overview of the Seventh Text REtrieval Conference (TREC-7). In *Proceedings of the Seventh Text REtrieval Conference*, pages 1–24.
- Walessa, M. (2001). *Bayesian Information Extraction from SAR Images*. PhD thesis, Universitat Siegen.
- Walessa, M. and Datcu, M. (2000). Model-based despeckling and Information Extraction from SAR Images. *IEEE Transactions on Geoscience and Remote Sensing*, 38(9):2258–2269.
-

-
- Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.
- Winkler, G. (1995). *Image Analysis, Random Fields and Dynamic Monte Carlo Methods: A Mathematical Introduction*. Springer.
- Xiao, J., Li, J., and Moody, A. (2003). A detail-preserving and flexible adaptive filter for speckle suppression in SAR imagery. *International Journal of Remote Sensing*, 24(12):2451–2465.
- Xie, H., Pierce, L., and Ulaby, F. (2002). SAR speckle reduction using wavelet denoising and Markov random field modeling. *IEEE Transactions on Geoscience and Remote Sensing*, 40(10):2196 – 2212.
-

Publications

1. Journal Papers

- Espinoza-Molina, D., Gleich, D., and Datcu, M. (2011). Evaluation of Bayesian despeckling and texture extraction methods based on Gauss Markov and Auto-binomial Gibbs random fields: Application to TerraSAR-X data. *IEEE Transaction on Geoscience and Remote Sensing*, to be published.
- Espinoza-Molina, D., Gleich, D., and Datcu, M. (2010). Gibbs Random Field Models for Model-Based Despeckling of SAR Images. *IEEE Letters on Geoscience and Remote Sensing*, 7(1):73-77.

2. Conferences

- Espinoza-Molina, D. and Datcu, M. (2010). Impact of Model Order and Estimation Windows for indexing TerraSAR-X images. In *Proceedings of SPIE Image and Signal Processing for Remote Sensing XVI*, volume 7829, pages 7-15.
 - Datcu, M., Espinoza-Molina, D., de Miguel, A., and Schwarz, G. (2010). Texture estimation in sar images: The impact of scale and model order. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 2844-2847.
 - Espinoza-Molina, D., Gleich, D., and Datcu, M. (2010). Assessment of two Gibbs random field based feature extraction methods for SAR images using a Cramer-Rao bound. In *8th European Conference on Synthetic Aperture Radar EUSAR 2010*.
 - Espinoza-Molina, D., Datcu, M., and Schwarz, G. (2009). Experience gained with texture modeling and classification of 1 meter resolution SAR images. In *Proceedings of SPIE Image and Signal Processing for Remote Sensing XV*, volume 7477.
 - Espinoza-Molina, D., Datcu, M., and Gleich, D. (2009). Cramer-Rao Bound-Based Evaluation of Texture Extraction from SAR Images. In *16th International Conference on Systems, Signals and Image Processing IWSSIP*, pages 1-4.
-