Information Extraction and Modeling from Remote Sensing Images

Application to the Enhancement of Digital Elevation Models

Zur Erlangung des akademischen Grades eines

DOKTOR-INGENIEURS

von der Fakultät für

Bauingenieur-, Geo- und Umweltwissenschaften

der Universität Fridericiana zu Karlsruhe (TH)

genehmigte

DISSERTATION

von

Dipl.-Ing. Cyrille MAIRE

aus

DIJON (Frankreich)

Tag der mündlichen Prüfung:
9. Februar 2010

Hauptreferent:	Prof. Dr.-Ing Stefan Hinz
Korreferent:	Prof. Dr.-Ing Pierre Grussenmeyer
Korreferent:	Prof. Dr.-Ing Peter Reinartz

Karlsruhe (2010)

Abstract

The analysis of remote sensing images aggregates numerous disciplines which were, before the advent of Earth observation (EO) techniques, distinct fields. The synergies between Cartography, Photogrammetry, Image processing and analysis, Computer vision, etc. allow to observe, monitor and study our environment.

In this dissertation, several methods to analyse, interpret and enhance satellite images are considered. Those different axes are motivated by a crucial need of efficient and novel methods for the extraction and the modeling of relevant information from standard remote sensing images in order to fully exploit those data presenting metric resolution over large areas.

During this research work, an innovative, fast and robust image processing system has been fully developed, tested and validated. It consists in, (1) a robust region-based segmentation approach, using the Mumford and Shah formalism to generate meaningful partitions from large remote sensing images, (2) a novel dynamic algorithm to extract and express the regions with the complex cellular model. A tree structure is introduced to store the regions while reflecting their topological (adjacency, inclusion) and geometrical properties. (3) information learning methods to analyse and interpret the extracted and modeled primitives.

The interest of such a processing line is emphasized on various EO data ranking from decametric to metric resolution (e.g. multispectral SPOT 5, Ikonos or QuickBird, panchromatic Cartosat) presenting various coverage types (e.g. agricultural fields, communication networks, man-made structures, urban areas).

The proposed system can be used for different applications and its potential is demonstrated for the improvement of world-wide digital elevation models (DEM)s, corrupted by noise, artifacts and inconsistencies. Image denoising techniques are investigated and a protocol to evaluate DEM's filtering performance is sketched. However, while resolution increases filtering limitations arise; statistical methods generally fail to deal with non-stationary data and the elevation data suffer from inconsistencies which have to be semantically corrected by means of semi-automatic methods.

A DEM enhancement system is presented; The object-based description derived from a remote sensing image is linked to 3D semantic models by means of information learning approaches. Having at disposal the tree structure augmented by semantic information, the region's 3D structure is estimated and integrated in the DEM by using multi-level B-spline refinement techniques. The presented system enables to improve accuracy and consistency of digital elevation models.

Zusammenfassung

Die Analyse der Fernerkundungsbilder betrifft viele Fächer, die vor Beginn der Erdbeobachtungstechniken unterschiedliche Felder waren. Die Synergien zwischen Kartographie, Photogrammetrie, Bildverarbeitung, Bildanalyse, Computer Vision usw. ermöglichen die Beobachtung, Überwachung und Untersuchung unserer Umwelt.

In dieser Dissertation werden mehrere Methoden zur Analyse, Interpretation und Verbesserung von Satellitendaten betrachtet. Die vollständige Modellierung und Extraktion relevanter Information aus den von modernen Fernerkundungssatelliten erzeugten riesigen Datenmengen erfordert effiziente und neuartige Methoden.

In der vorliegenden Arbeit wird ein neues, schnelles und robustes Bildverarbeitungssystem entwickelt, getestet und validiert. Dieses besteht aus:

(1) einem effizienten region-basierten Segmentierungsverfahren, welches den Mumford & Shah Formalismus verwendet, um relevante Regionen aus große Fernerkundungsbildern zu generieren,

(2) einem neuartigen dynamischen Algorithmus um die Regionen zu extrahieren und diese mit dem zellulären Komplexmodell zu beschreiben. Die Regionen werden unter Berücksichtigung ihrer topologischen (Angrenzen, Inklusionen) und geometrische Eigenschaften in einer Baumstruktur organisiert,

(3) Klassifikation mittels Support Vector Machines (SVM)-Methoden, zur Analyse und Interpretation der extrahierten und modellierten Regionen.

Das entwickelte System liefert auf verschiedenen Datensätzen (z.B. multispektrale SPOT 5, Ikonos oder QuickBird Szenen, panchromatisches Cartosat Szenen) mit unterschiedlichen Landbedeckungen (z.B. landwirtschaftliche Felder, Straßennetzwerke, künstliche Strukturen, Stadtgebiete) sehr gute Ergebnisse.

Dieses System ist die Grundlage für weitere Anwendungen. In dieser Arbeit wird es zur Verbesserung von durch Artefakte und Inkonsistenzen gestörten weltweiten digitalen Höhenmodellen (DHM) verwendet. Klassische Filter und Rauschunterdrückungsverfahren werden untersucht, und eine Methodik zur Bewertung der Ergebnisse gezeigt. Dabei treten Probleme mit höher aufgelösten Daten (Auflösungsbereich 1-10 m) auf, weil in der Regel konventionelle statistische Verfahren für ortsveränderliche Daten nicht optimal sind, und Inkonsistenzen im Höhenmodell mit Hilfe von semi-automatischen Methoden verbessert werden müssen.

Ein System zur Verbesserung von Höhenmodellen unter Verwendung von komplementären, aus optischen Bilddaten modellierten Regionen wird entwickelt. Die dreidimensionale Struktur der Regionen wird unter Verwendung der Topologie und zusätzlicher semantischer Information geschätzt und mit einem B-spline basierten Verfahren in das Höhenmodell integriert. Dadurch wird die Genauigkeit und systematische Konsistenz der Höhenmodelle verbessert.

Acknowledgments

Contents

Introduction

With the development of Earth Observation (EO) techniques, new possibilities are offered to observe, monitor and study Earth surface, atmosphere and environment. Information is captured using a sensor operating in a specific range of the electromagnetic spectrum, on board of airborne or spaceborne platforms. Compared to cartographic surveying, the interests of EO technologies have already been proved to be multiple: short observation cycles on each part of the globe even under critical weather conditions or relief, wide range of spectral analyses, large remotely sensed areas, etc.

However, the data are free from a-priori knowledge; it induces important processing chains to extract meaningful information. Indeed these processes are different from traditional mapping which generates relevant information and deduces the area, while remote sensing (RS) survey acquires the area and then extracts significant information within boundaries. To optimally exploit RS data, this remark emphasizes the importance of the pre-processing, extraction and interpretation phases. The availability of more complex and various EO data (see the optical and SAR metric images presented in Fig. 1.1) induces some migration of classical RS image analysis methods from large scale modeling based on global statistical models to computer vision techniques with non-linear objects or features, especially in man-made areas. In this purpose, the RS community is faced with challenging issues when developing methods to

- handle an increasing amount of data archives due to data higher complexity and content,
- filter and enhance the acquired data,
- extract, model relevant information and exploit synergies between data.

1.1 Elevation Data Enhancement

The scope of remote sensing is not restricted to the acquisition of raw data such as optical, radar or thermal images, but also to the generation of complementary information by exploiting synergies between multiple acquisitions and physical characteristics of the electromagnetic spectrum employed. In particular, by means of Interferometric Synthetic Aperture Radar (InSAR) or optical stereoscopy, large and homogeneous digital elevation models (DEM) are now available. SRTM or Cartosat are respectively typical examples of Earth- and continent-scale elevation data-sets. With the upcoming TanDEM-X mission, a new level is expected to deliver a homogeneous sub-decametric digital elevation model of the Earth.

(a) SAR Image (TerraSAR-X)   (b) Optical Image (Ikonos)
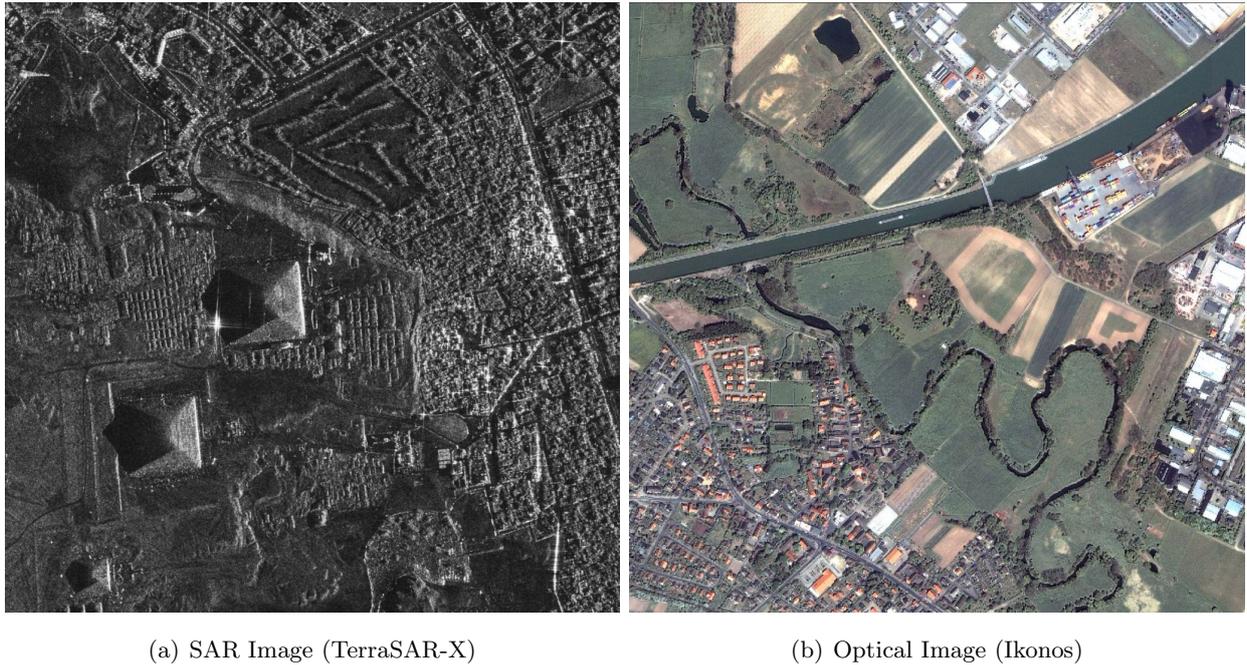
Figure 1.1: Metric EO Images (size 1536×1536). (a) SAR image of the pyramids of Gizeh in Egypt. The image was acquired by the TerraSAR X-band sensor with the "High Resolution Spotlight" mode (resolution of 1 meter, HH polarization). An analysis of the scattering mechanisms involved is presented in [9]. (b) Optical image acquired by the Ikonos satellite over Germany, panchromatic and 3.2 meters multispectral data are fused to deliver a 1 meter multispectral image.

If both methods have been reaching an operational stage, artifacts, inconsistencies and noise remain. Surprisingly, in the literature, few attention has been paid to the enhancement of elevation data which constitutes, in geo-spatial analysis, one of the most essential and stable information. Image denoising can significantly contribute to reduce the impact of such disturbing effects. Limitations arise when image resolution increases. Statistical methods mainly fail to properly handle objects and/or structures. In this context, the variety of approaches to handle sub-metric elevation data (mainly LIDAR) points out the topic complexity.

The representation of the third dimension provides important information about the relations between land coverages: shapes, structures and slopes appear to describe scene objects such as waterways, surface material or vegetative growth. Indeed, our perception automatically associates an object with its corresponding shape. For instance, a lake should be a flat surface with the lowest elevation according to its direct neighbors. Implicitly, the access to semantic information can be determinant to improve DEM quality. To enhance elevation data, it has been demonstrated [114] that having at disposal an existing Geographical Information System (GIS) can lead to significant improvements of the elevation data in terms of accuracy, consistency and realism. However, even when such databases exist for a studied area, it is time-consuming to access and collect suitable information.

The state of the art of EO images (quality and resolution) should be an alternative to the use of GIS databases. Since land-cover or man-made structures are easily recognized in an optical image, we propose to achieve the DEM regularization by using only an available optical image. It consists in extracting relevant information from the optical image and aggregating them in the filtered DEM. In this context, innovative methods based on the analysis of EO images have to be proposed and developed to:

- extract land-cover or man-made structures from large optical images,
- model and encode the resulting extracted features,
- develop semi-automatic learning methods,
- link semantic content to geometrical regularization constraints,
- integrate the regularization constraints on elevation data.

1.2 Advances in Image Information Retrieval and Modeling

The initial objective of building a computer vision system emerged in the 1960's and is still an unresolved issue. Indeed, to retrieve and model the original 3D scene, numerous problems arise given a set of two-dimensional discrete observations corrupted with physical inaccuracies, noise and quantizations. The statement of image analysis as an ill-posed inverse problem, combined with the non-stationarity properties of images, may explain the interest of researchers for the Bayesian analysis [73].

If the earliest methods proposed were mainly characterized by highly heuristic approaches, the contribution of Marr [146] develops a more consistent and rigorous formalism, based on results collected from psychology and neurophysiology, to address the problem of vision. The information processing system proposed distinguishes three distinct and complementary stages with increasingly symbolic scene representation. This organization into different levels (bottom-up approach) is still up-to-date, even if it suffers from a lack of validation in the reverse direction (top-down).

The first stage of a computer vision system consists in the extraction of low-level primitives by means of feature extraction (e.g. textures, [82]) and segmentation techniques. Segmentation processes aim to subdivide an image into disjoint and homogeneous entities, called regions, by defining sharp boundaries between regions using some decision criterion. This duality leads to represent image segmentation as a compromise between those two objectives. If edge-based approaches [24], [48] provided successful applications (e.g. machine vision), further post-processing stages are required to retrieve the corresponding regions. Region-based approaches [166] exploit directly spatial information in order to group homogeneous pixels into closed and connected regions.

Performances have been significantly increased by the formulation of the image segmentation as an estimation problem, introduced in [154]. However, there still does not exist a solution to the Mumford and Shah functional; only approximations are derived by studying its bounding cases ("snakes", [106], [26]), ("cartoon limit", [115]) or by using partial differential equations to

iteratively converge to a solution ("level sets", [27]). Other approaches have been proposed, based on structural analysis [151], entropy [126], hybrid classification [37] and multiscale Bayesian [19] techniques, or unified Bayesian/MDL formalisms [218].

The partition of the image into disjoint and connected regions, resulting from the segmentation stage, requires to be modeled into higher levels of information, expressing shape, statistics and topological properties of the regions. Among spatial discrete models proposed in the literature [116], [117], the complex cellular approach [117] is particularly well-suited to achieve geometrical and topological image description since it eludes the connectivity paradox and topological problems with $\mathbb{Z}^2$ (i.e. with discrete grid). The plane ($\mathbb{R}^2$) is decomposed in regular complex cells, which replace the pixel model ($\mathbb{Z}^2$). A discrete model is generated that preserves the topology of $\mathbb{R}^2$ in a similar way as manipulating vector models.

The analysis of data is motivated by two main purposes: (1) exploration, to understand and evaluate information content of raw images without any pre-specified models, (2) inference, to validate a model derived from some low-level operations such as region-based segmentation or texture patterns. In this scope, classification techniques play a major role. There are two basic approaches to classification, "supervised" (e.g. Support Vector Machine) and "non-supervised" (e.g. K-Means), depending on whether a set of prototypes is available or not. The former being particularly interesting since it enables to ingest *a priori* knowledge in the system.

1.3   Outline and Main Contributions of this Thesis

To enhance DEMs, the possibility to exploit synergies between optical images and elevation data implies the elaboration of a robust workflow to extract and model information from RS images. For this reason, most of the efforts have been devoted to (1) segment efficiently large images, (2) model the segmented image, (3) analyse the data and the results obtained.

A segmentation algorithm is presented, based on the Mumford and Shah functional. It enables the generation of meaningful partitions, even on large images, which become a standard in remote sensing with the resolution increase and technological improvements[1].

In order to deal with regions expressed as objects, their geometrical and topological characteristics need to be extracted. If such properties can be retrieved during the segmentation process, a sequential approach presents the advantage to be independent of the segmentation method used. In this scope, a method to generate an object description of a region-based segmentation is presented. Based on the complex cellular model, we propose to complete the object representation by investigating the inclusion relationships between regions. In this purpose, a tree structure is introduced to store the regions while reflecting their topological properties [140], [141].

---

[1]The original Ikonos image (Fig. 1.1) is 12000×14000 pixels large with 11 bits range for each of its 4 channels.

The analysis of optical images and the extraction of features/regions presenting similar properties is investigated by means of information learning approaches. The self-similarity properties of the relief is investigated by means of the fractal analysis to derive quality estimators able to deal with non-stationary data such as DEMs.

The enhancement of elevation data is addressed in two stages: (1) filtering, promising techniques such as wavelet shrinkage [51] or Bayesian filtering [208] need to be evaluated and compared [142] to classical filtering approaches. (2) integration of complementary information [140].

A DEM regularization is achieved by using only an optical image. It consists in extracting relevant information from the optical image and aggregate them in the filtered DEM. Using the object-based description, structures to be integrated into the DEM are interactively selected and classified among a set of user-thematics. Each thematic is associated with a corresponding elevation modeling and enables to estimate the region's 3D structure. Contrary to the approach proposed in [114] based on a triangulated DEMs, the presented work is realized on a classical regular grid representation. In this scope, interpolation methods are studied to integrate 3D structures in digital elevation models.

1.4   Organization of the Document

Some background notions, shared all along the dissertation, are presented in Part I. Chapter 2 mainly deals with basics on image representation and analysis while chapter 3 is devoted to EO data. Introductions in Bayesian image analysis and SAR Imagery are provided respectively in annex B.1 and annex A.1.

Part II is related to the analysis of images. An overview on image segmentation techniques is presented in chapter 4. An algorithm is proposed in chapter 5 for the segmentation of large images and illustrated with some examples using *Ikonos* and *SPOT 5* data-sets. Image modeling is addressed in chapter 6 to elaborate a complete object representation of segmented images, by encapsulating both a geometrical and a topological description of regions. In addition, classification techniques applied mostly on multispectral images are investigated.

The enhancement of DEM is presented in Part III. DEM post-processing operations are investigated in chapter 7 and evaluated based on X-SRTM and Cartosat data. An important issue is to enhance DEM information, the data itself being not sufficient to correct the data. A DEM regularization is sketched in chapter 8 by using only RS images.

Part I

Preliminaries

Image and Computer Vision

Abstract

An overview on digital images is presented, beginning with image formation processes. Some specific image representations (an image is itself a representation...), characteristics and properties are explored and emphasized by typical image analysis methods.

Definitions and concepts used throughout the dissertation are discussed. Such background knowledge is related to linear algebra, image processing/analysis and computer vision.

We present possible hierarchical approaches adopted in computer vision and address the examples of local analysis by means of contour and texture extraction techniques. A sketch of the image analysis flowchart discussed in this dissertation is introduced.

Introduction

The term "image" is employed to qualify a representation or imitation of a person or thing. Conjointly with technological development, its support evolved and may have a physical description (mirror reflection, painting, picture, statue, code of bits) or mental representation (concepts, ideas, impressions, symbols, etc.). Let's denote a function expressed by $f : \mathcal{D} \Rightarrow \mathcal{V}$ where $\mathcal{D}$ and $\mathcal{V}$ are at least topological space domains. In the case of the image formation process, the **forward** model f leads to a set of observations y,

$$y = f(x) + n$$

x being some physical parameters of the scene sensed and n noise or disturbing effects. Similarly, applying an image operator f to a given image x produces an output y in form of an image or some features, patterns, partial computations, etc. Those two examples (image formation and operation) emphasize typical tasks in image processing or computer vision; given an observed image y:

- *Scene Understanding* aims at retrieving the scene properties (e.g. object geometry),

- *Image Understanding* focuses on the extraction of patterns from the uncorrupted image x.

Both cases appear as ill-posed problems with critical constraints for the estimation of the **backward** model f^{-1}; their direct inversion often results in non-uniqueness and/or unstable solutions. Indeed, f is rarely a bijective process since (1) data are generally pervaded with noise, (2) one or several quantizations on finite support are involved (the codomain $\mathcal{V}$ generally presents a lower space dimension than $\mathcal{D}$); both cases inducing a loss of information.

2.1 Image Representation

Fig. 2.1 depicts the axes of analysis and images properties discussed in the present section. We provide for each of them a typical image processing method.
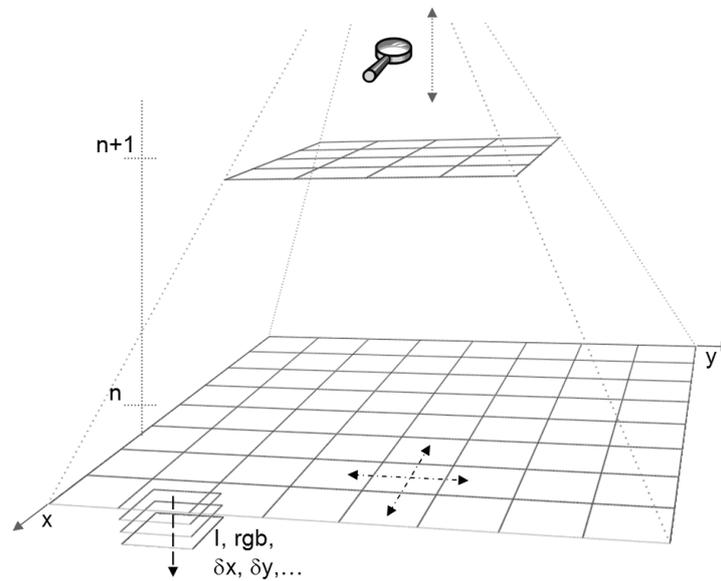
Figure 2.1: Representation axes of an image, a discrete planar xy function: scale-resolution analysis (dot arrows) at levels n, $n+1$, etc. spectral (dash arrows) and structural (dash-dot arrows) analyses.

2.1.1 Image Formation

The information reflected or emitted from an observed scene is collected by an instrument. Active sensors are acquisition instruments which emit themselves the measured signal, passive sensors receive a reflected or emitted radiation signal from natural sources.

Electromagnetic Spectrum
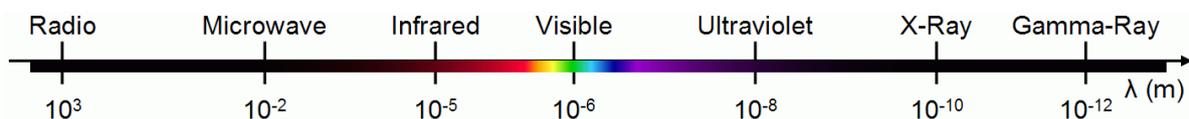
Figure 2.2: Electromagnetic spectrum. Domains sensed according to their wavelength λ.

The information collected from the scene is generally stored as images and constitutes physical measures on a specific domain of the electromagnetic spectrum (Fig. 2.2). Humans are able to perceive two of them (sounds and lights). The observation of those domains with instruments leads to several $\{1...n\}$-dimensional signal techniques applied for sky studies, medical diagnoses (X-Ray, PET, ultra-sound, MRI), night vision in Infrared, optical photography, etc. Each of

them, depending on its own properties and interactions with materials, enables the acquisition of complementary information and leads to dedicated applications or synergies.

Although most of the following methods have been successfully applied on different spectrum domains or even other types of information, we are focusing our attention in this dissertation mainly on optical and microwave images, commonly used in the remote sensing field. In this scope, microwave characteristics and SAR processing are presented in annex A.1.

Image Sampling and Quantization

The image definition space, $\mathcal{V}$, is typically expressed on $\mathbb{Z}^n$, i.e. on discrete support. A classical processing chain is depicted in Fig. 2.3 to convert a continuous analog signal into a digital one. We chose this example - perhaps far away from CCD arrays embedded in digital camera or satellite - to at least mention the sequence and the process ordering involved.

Figure 2.3: Conversion chain from Analog to Digital signals.

To properly reconstruct the original signal, the choice of a suitable sampling rate is crucial. In this scope, a sufficient bounding condition - the Nyquist frequency f_s (half the sampling rate) - is given by the Nyquist-Shannon sampling theorem. It states that frequencies above f_s can be sampled without loss of information, whereas below f_s, it is getting ambiguous to distinguish the signal from the other components. This issue, called aliasing, breaks the sampled structural information. A low-pass filtering applied in the earlier stage prevents from aliasing effects [98]. As pointed out by Morel in [49], if the Shannon theory made explicit the conditions to reconstruct a continuous image from a finite number of values (discrete image representation), the amount of information in a digital image is finite and its resolution is bounded.

Quantization denotes the approximation of a continuous range of values to a discrete one [98]. According to the range and the sensitivity of the sensor, a certain bit depth is adopted, generally in the form of integer values encoded in base 2. This operation is not always a loss of information, some accurate sensors are already able to count the number of photos collected. The amount of information is linearly dependent on the depth and the sampling rate adopted.

For the generation of 2-dimensional discrete signals like images, many configurations are available: sampling the signal on regular square or hexagonal grid - see Fig. 2.6 -, moving or rotating a 1-d array in its normal direction or interleaved sampling (also referred as quincunx sampling). A specific mode of the SPOT 5 satellite for instance directly exploits this strategy in acquiring from the same location two signals shifted of half sampling rate. By interpolating interleaved missing information, the generated image is delivered with a twice higher resolution.

Fourier Analysis and Interpolation

Although the Fourier analysis was firstly devoted to continuous signal, its discrete and optimized version, the Fast Fourier Transform (FFT), constitutes a powerful and widespread tool for the analysis of n-dimensional signals. The Fourier transform is linear and invertible, decomposing a signal into a sum of components of different frequencies. The representation of a signal in the frequency domain offers numerous interesting properties; the convolution of two infinite sequences is obtained as the inverse transform of the product from individual transforms (convolution theorem).

The Fourier transform is used in almost all signal/image processing fields such as spectral analysis, noise reduction, compression, partial differential equations, interpolation, etc. However, for the past two decades, this technique has been greatly challenged by an ever more promising one, the wavelet analysis. Fig. 2.4 depicts the Lena's image decomposition in the Fourier domain, before and after the use of a low-pass filter. The representation of digital images on a continuous support is often required in various image transformations (e.g. rotation, re-sampling, rectification, etc.). Interpolation operators are defined on local neighborhood; nearest-neighbor, bilinear, bicubic are classical methods. Robust estimations are computed by polynomial, rational or trigonometric functions. In chapter 8, B-Spline techniques are introduced for the interpolation of scattered data.

2.1.2   Image Scale Representation

Because scale and resolution notions are quite often mixed up, those two concepts are presented. A short overview of *fractals* and *wavelets* is given. These two strong mathematical tools rely on scale and resolution notions and present many possibilities: image re-sampling, compression and filtering, multi-scale analysis, landscape or image synthesis (see Fig. 2.8(a)), etc.

Concepts of Scale and Resolution

Resolution is a pure signal processing notion related to the effect of a low-pass filter on a signal. The Dirac impulsion, filtered with an ideal low-pass filter, results in a sinus cardinal signal. Since the Dirac distribution is of infinite amplitude with a null bandwidth, the corresponding image will contain very fine details on a continuous support. In opposition, the ideal low-pass filter operation (elimination of high frequency in Fourier domain) reduces the possibility to distinguish and separate fine details. This limitation is linked to the first lobe width of the sinus cardinal function. As a consequence, the image appears with a blurring effect, depending on the acquisition system and conditioning its resolution. This traduces a loss of information.

The **scale** is a mathematical concept based on the notion of contraction/dilation of the signal. All information is preserved by applying a scale transformation to a continuous signal. From the mathematical point of view, the scale transform cannot be directly applied to 2D discrete signals such as images and requires interpolations (i.e. low-pass filter). Since sub-sampling is a

combination of scale and resolution transforms, the scale transform for discrete signal does not preserve the whole information. In addition, a particular system geometry can lead to various scales, even for a coherent and homogeneous acquired data-set. Remote sensing images, defined on regular lattices, are often registered and projected into a geographic ellipsoidal coordinates system. In this case, the resulting sampled information covers different ground widths.

Fractals are mathematical objects characterized by self-similarity properties. A fractal is a geometric shape that has symmetry of scale; its examination at all possible scales will result in the same structure (and statistics) [145]. This qualitative definition allows to emphasize the fractals phenomenon and the basis of fractal analysis algorithms: to observe the fractal behavior at several scales. The relief's self-similarity was already figured out before the fractal concepts had been defined. In this scope, if the fractal analysis is interesting to generate (landscape synthesis) or analyze elevation data, it needs to be carefully applied since pure self-similarity is not a property of real landscapes [75]. A fractal object is characterized by its degree of irregularity or fragmentation, also called fractal dimension. Dimension objects given by integers are non-fractal (1 for curves, 2 for surfaces, etc.), whereas fractals have fractional dimensions since they represent transition structures between curves and surfaces, surfaces and solid bodies, etc.

Multi-scale Representation

Like fractals, **wavelets** rely on the analysis of the data at several scales [144]. The wavelet transform is similar to the Fourier's one. The latter decomposes the signal into a sum of sinusoidal waves; functions are localized in the Fourier space (frequency) which is enough for stationary signals. The main interest of the wavelet transform is its ability to deal with non-stationary signals since it uses functions that are localized in both real (time) and Fourier space.

Given a signal $x(t)$, a specific sequence $A_m x(t)$ is constructed to represent the signal $x(t)$ at a given resolution. The signal space $V(x(t))$ is decomposed into a nested sequence of approximation spaces $V_m \subseteq V_{m+1}$, $m \in \mathbb{Z}$, obtaining an expression of the analysed signal as a linear combination of basis functions of V_m where $A_m x(t)$, the projection of $x(t)$ in the subspace V_m, constitutes the best approximation of $x(t)$ at the resolution 2^m. Since the approximation at the scale m is recovered at the scale $m + 1$, the wavelet transform gives a description of the "closest" signal approximation at each scale without loss of information. In defining at the resolution 2^m the detail signal D_m by $D_m x(t) = A_{m+1} x(t) - A_m x(t)$ we obtain

$$D_m x(t) = \sum_n x_n^m \ \psi_n^m(t) \qquad x_n^m = \int_{-\infty}^{\infty} x(t) \ \psi_n^m(t) dt \qquad (2.1)$$

where $\psi_n^m(t)$ denotes a set of basis functions built for each level m. An interesting property is that all of them are constructed from a single function called "mother wavelet" ψ, by dilating and translating ψ by different amounts (a_0, b_0): $\psi_n^m(x) = a_0^{m/2} \ \psi(a_0^m \ x - n \ b_0)$. Similarly, the basis functions for V_m must be derivable from a so-called *scaling function* ϕ (father wavelet). The inverse

(a) Fourier's domain   (b) Fourier's domain filtering   (c) Wavelet decomposition, $m = 3$
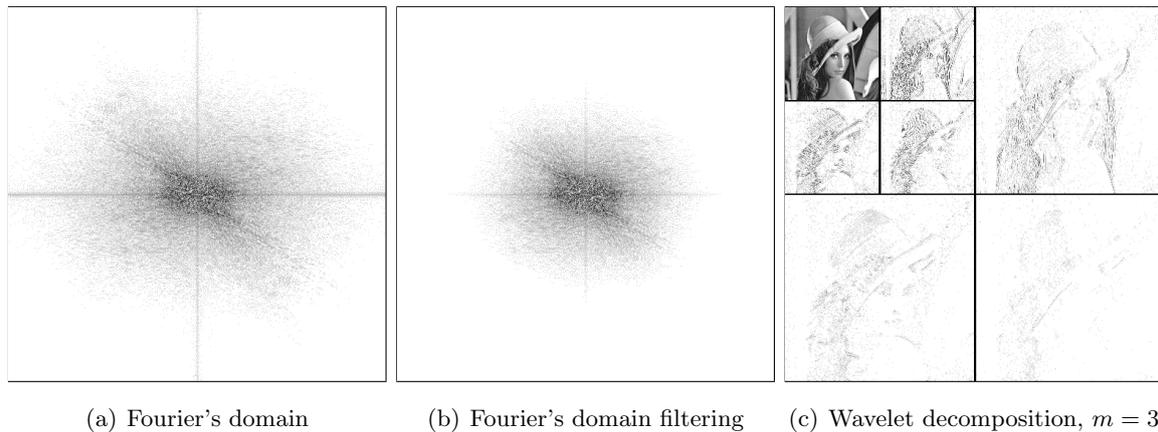
Figure 2.4: Image representation in Fourier and Wavelet domain, applied on Lena image. All images have been inverted (dark areas represent large magnitude coefficients).

discrete wavelet transform is given by $x(t) = \sum_m \sum_n x_n^m \ \psi_n^m(t)$

By analogy with the Fourier transform, the concept can be extended to the 2D discrete case (see Fig. 2.4). However, it exists an infinite number of functions satisfying the properties required by the wavelet theory (orthogonal wavelet with compact support in the discrete case). Gabor and Haar functions were the first ones used. More recently, the family introduced by Daubechies [47] ($a_0 = 2, b_0 = 1$) has proved efficiency in various applications, one of them is part of the lossless image format JPEG 2000. The wavelet decomposition of an image can be interpreted as a signal decomposition in a set of independent, spatially oriented frequency channels. The decomposition is performed on $2^i \times 2^j$ basis with no volume increase introduced. The reader is referred to [144], [87], [47] for further explanations. We mainly investigate such mathematical tools to filter and analyse digital elevation data.

2.1.3 Image Structure Analysis

Image sampling and quantization do not ensure a statistical independence between picture elements. Numerous image techniques such as image filtering, segmentation, restoration or compression rely on the neighboring pixel analysis and the extraction of local properties. Before illustrating this analysis axis with morphological operations, some topological concepts are presented to properly deal with various elements such as pixels, chain of pixels, points or objects (see Fig. 2.5).

A **Graph** is an ordered pair $G := (N, E)$ where N denotes a set of "nodes" or vertices and E a set of "edges" with $E \subseteq N \times N$. When E is a set of ordered pairs of vertices, the graph is said to be "directed": the directed edge or arc $e \in E$ with $e = (x, y)$ is oriented from x to y ("path"). In such cases, the vertices x and y are connected[1].

---

[1]Primitives data structures [66], [67] as well as topological modeling [117], [116], are presented more in detail in chapter 6, pages 87-94.

(a) Regular Grid   (b) TIN   (c) Binary Tree   (d) Tree of Regions
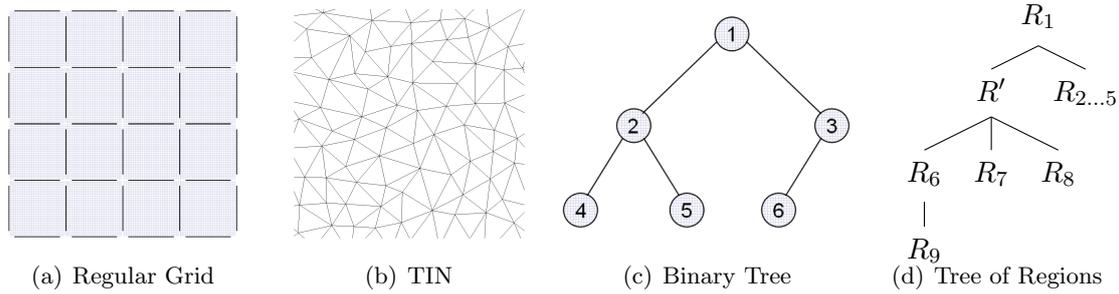
Figure 2.5: Some Data structures: (a) regular grid, (b) Triangular Irregular Network (TIN), (c) example of tree structure, here a binary tree. The root node (1) (at the top of the tree) is the parent of (2) and (3). (2) and (3), the children of (1) are siblings. Height of the tree is 2. Depth of (2) and (3) is 1. (d) Tree of regions, symbolic representation, see example page 97.

Neighborhood Systems and Regular Grid

A "lattice" is a graph whose vertices are the points in the plane (resp. 3D space) with positive integer coordinates. Such a definition appears extremely well suited to model digital images. Let S be an image defined on a regular lattice $\mathcal{L}$ whose sites are pixels. A neighborhood system in the topological space $\mathcal{L}$ is defined by

$$\mathcal{N} = \{\mathcal{N}_r | \forall r \in \mathcal{L}\} \tag{2.2}$$

where $\mathcal{N}_r$ is defined as the set of neighboring sites of r. $\mathcal{N}_r$ is built with the properties:

- A site does not belong to its associated neighboring $i \notin \mathcal{N}_i$
- It exists a mutual relationship between sets of neighboring sites $s \in \mathcal{N}_r \Leftrightarrow r \in \mathcal{N}_s$

(a) 4 connectivity   (b) 8 connectivity   (c) hexagonal connectivity
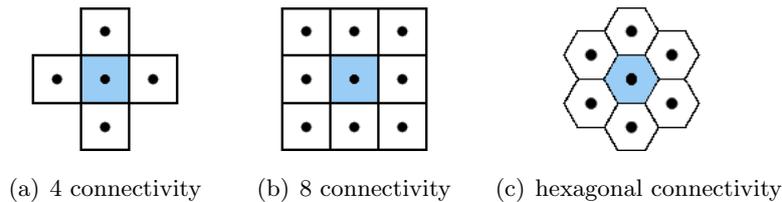
Figure 2.6: Classical configurations for neighboring system representation.

The **regular grid** is the most usual way for describing and manipulating images, composed of cells called pixels regularly sorted. In the plane (resp. 3D space), each pixel (voxel) is associated with a lattice point. Such a discrete representation of the signal is depicted in Fig. 2.5(a). Two distinct lattice points in the plane [178] are said:

- 8-*adjacent* if their corresponding coordinates differ by at most one unity,
- 4-*adjacent* if they are 8-adjacent and differ in at most one of their coordinates.

Fig. 2.6 depicts classical configurations, called respectively *4-, 8 connectivity* systems, employed for the analysis of local pixels interactions [176]. The cell element has a predefined size and shares boundaries with 4 neighbors (for 2-dimensional representation), each of its 4 nodes are shared

with three other cells. By boundary, we mean the inter-pixel connection between two adjacent cells. 4-connectivity neighborhood systems are implicitly preferred since diagonal cell's intersection (occurring with 8-connectivity system) is reduced to a node[1]. Let's denote R_i the n pixels composing an image defined on a regular lattice Ω and Γ the set of boundaries, we have

$$\bigcup_{i=1}^{n} R_i = \Omega \setminus \Gamma$$

Sequence: upper chain	lower chain
022006646000242000600	000000000
$0_1 2_2 0_2 6_2 4_1 6_1 0_3 2_1 4_1 2_1 0_3 6_1 0_2$	0_9

4 Connectivity   8 Connectivity
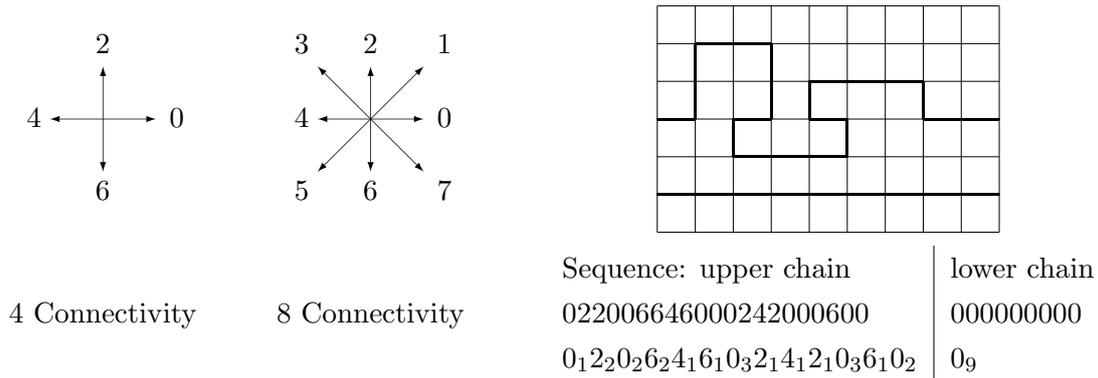
Figure 2.7: Freeman Code on 4- (resp. 8-) connectivity systems. Two examples of chain encoding (direction from left to right) are given using only Freeman's orientation (left) or by combining it with direction length (right), the index of the orientation being its number of consecutive occurrences.

A classical method to represent boundaries on a lattice is the **chain code of Freeman** [67]. It consists in describing a sequence of consecutive connected points by their relative orientation direction, as shown in Fig. 2.7. Let's consider a sequence of m connected points $\{p_1, ... p_m\}$ which describe a curve under the connectivity system $\mathcal{N}$. p_{i-1} and p_{i+1} are the neighbors of p_i $\forall i < m$. In the case of a closed curve, the neighbors of p_m are obviously p_{m-1} and p_0. Each jump from one point to the next one is represented using a vector $\overrightarrow{p_i p_{i+1}}$ and encoded using an integer value indicating the relative direction from p_i to p_{i+1}. It results that the overall curve is represented by a sequence of $m - 1$ vectors or m if the last one is made explicit. Numerous relations are found in the literature (e.g. [98]) to derive geometrical features from the Freeman code.

Set of Points

A graph which can be drawn in a plane without crossing edges is called a "planar graph". A "dual graph" H of a given planar graph G is a graph which has a vertex for each plane region of G, and an edge for each edge in G joining two neighboring regions. A celebrated planar graph, the foundation of Computer graphics, is the **Delaunay triangulation**. For a set of points P, it consists in a subdivision of the plane into triangles so that no point of P is inside the circumcircle of any triangle. Indeed, this constraint penalizes thin angles. The dual graph of a Delaunay triangulation is a **Voronoï diagram**. Given a set of points P - called Voronoï sites s - it relies on finding over the plane all the locations equidistant to the two closest sites. The resulting segments partition the

plane into regions $V(s)$ which represent the closest area to each site s. Such structures play a major role for solving various and different geometric problems. The representation in the RGB-space of a clustering into k classes - see Fig. 2.9(d) - can be interpreted as a Voronoï diagram.

Data Handling

A "cycle" is a path without self intersections. A connected graph without cycle is a tree. **Trees** are hierarchical and recursive structures presenting nodes (leaves), relationships (branches) and properties between the connected nodes. The root is the starting node (at the top of the tree, see Fig. 2.5(c)). It can be connected to other nodes, each of them being considered as a root of a sub-tree. Connections between nodes are modeled by disjoint "branches". Each node stems only from another node (only one parent (the "ancestor"), vertical relation, bottom-up direction). A node can have several children (at most two "descendants" in the case of binary trees, vertical relation, top-down direction), called "siblings" of each other. Connections in adjacent levels (horizontal relations) are allowed.

"Terminal nodes" are the nodes without children. The "path" is a sequence of nodes $n_1...n_k$ so that n_i is the parent of n_{i+1}, $\forall i \in \{1...k-1\}$. Its length is 1 less than the number of nodes on the path. The "height" of a tree is the longest path length between the root and its terminating nodes. The "depth" of a node denotes the length of the unique path from the root to the node. The node's ordering of a tree is unique (transversal property to visit each node only once).

A **heap** is a specialized tree-based data structure satisfying the heap property; if B is a child-node of A, a *max-heap* follows $A \geq B$, ($A \leq B$ for *min-heap*). Such a structure promotes the largest (resp. smallest) value to the root.

Those classical structures are used throughout the dissertation, applied on pixels or on higher level elements/primitives such as boundaries.

Morphological operations

Mathematical morphology is a widely used technique, involved for instance in feature detection, image segmentation or noise removal processes. It refers to certain operations applied on an object within an image with a structuring element. With the analysis of the reduced object, some shape or size properties of the studied object are then derived. The "structuring element" $\mathcal{N}$ is a subset of $\mathbb{R}^2$, generally defined as a neighborhood system (*4-8 connectivity*) or a shape. Most morphological operations can be defined using two elementary operations called *dilation* and *erosion* [184]. Let $u(x)$ be an image, the erosion of u by $\mathcal{N}$ is the image t defined by

$$t(x) = \inf_{\delta \in \mathcal{N}} u(x + \delta)$$

Similarly, its "reverse" operation, the dilation, is

$$t(x) = \sup_{\delta \in \mathcal{N}} u(x + \delta)$$

(a) Raw Image   (b) Erosion   (c) Dilation   (d) Opening   (e) Closing   (f) Skeleton

(g) Raw Image   (h) Erosion   (i) Dilation   (j) Opening   (k) Closing
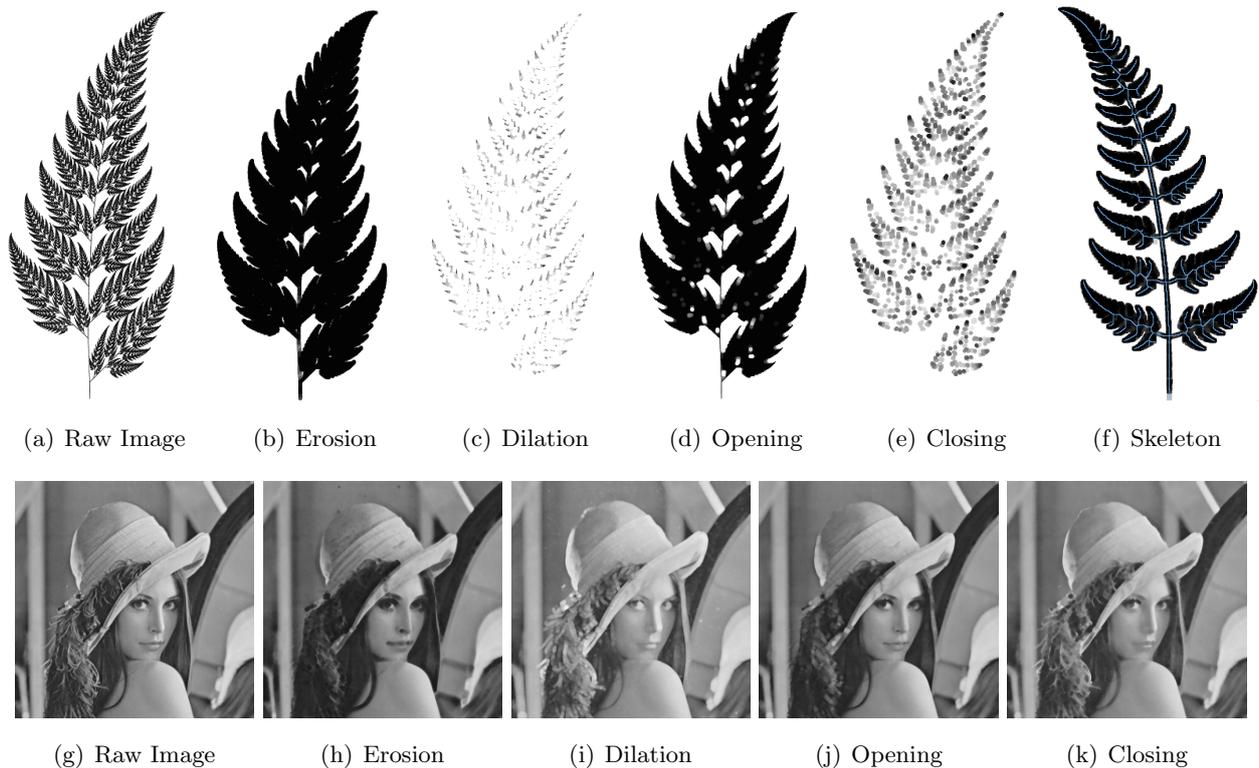
Figure 2.8: Some morphological operators applied on Barnsley's fractal synthesis fern images (size 256×512) with $n = 2$ and Lena's image $n = 3$ (size 512×512), using a 4 connectivity system as structuring element.

Erosion can be seen as a shrinkage operation whereas dilation is an expansion one. If we consider a particular object in a image, an erosion of the object is equivalent to the dilation of the background (*Duality* property). It should be mentioned that dilation and erosion are not inverse to each other. Those operations have interesting properties, such as *invariance in translation, distributivity, iteration, increasing*. Applying a morphological operation f with a size n is equivalent to iterate n times f. An *opening* operation is obtained by applying sequentially n erosions and n dilations on the processed image. The *closing* operation denotes its dual, both having increasing $(A \in B \rightarrow f(A) \in B)$, idempotent $(f(f(A_i)) = f(A_i))$ and (anti-) extensive (opening: $f(A) \in A$, closing: $A \in f(A)$) properties. Using range images, some examples are given in Fig. 2.8. *Skeletonizing, pruning, thinning or thickening* are complementary morphological operations useful in image processing; interesting presentations and illustrations can be found in [84], [98]. For a more advanced coverage of mathematical morphology, the reader is referred to [184].

2.1.4 Image Information Content

A large type of information can be stored in digital images. It can be physical measurements sensed during a short time interval in one or several spectral bands in the form of real or complex

values. In many applications, these raw data are directly exploited; e.g. intensity or amplitude of the received signal, travel time from an emitted wave, phase analysis (Doppler variations), thermal information. Some techniques based on many acquisitions require intensive computations to retrieve a higher scene information level; for instance, image pair stereoscopy to get 3-dimensional surface description or computer tomography to reconstruct volumetric scene representation.

For the purpose of image analysis, information in each picture element (pixels) often represents local features, statistical descriptors (e.g. gradients, variance, etc.), thematic or semantic content [2]. The discrete image representation is explicitly used for positioning semantic contents and if possible expresses their topological relationships. At an even more abstracted level, logical operations between images are extremely convenient and useful, a simple example being the definition of masks with Boolean values. If images are mostly represented in the Euclidean space, efficient image techniques operate on other spaces, such as the Fourier domain or Hough parameter space.

Digital images are numerical information handled and processed using a computer. Each physical measurement, pattern or semantic information is interpreted as codes of bits, i.e quantified. According to the content to be stored as well as accuracy and size constraints, pixels are encoded with integers, floating or complex values. Numerous compression schemes exploit the code redundancy in the numerical support to reduce the amount of space required for its full description.

Feature Domain and Color Space

The most common color representation is the RGB space where colors are represented by their red, green and blue components in an orthogonal Cartesian space. It corresponds to a three-component additive color model, where any color is defined by its coordinates, also known as the tristimulus theory [192]. The human visual system operates in the same way, i.e. photoreceptors called cone cells acquire color information by means of three band-pass filters centered on each primary spectral wavelength. If the RGB space matches the acquisition of the primary components, human color interpretation (i.e. perception) is better represented in the components of another space called HSI expressed by hue, saturation and intensity [98][3].

Measure of Similarity

The similarity or proximity between two data groups in multidimensional space can be explicitly quantified by a distance measure between their corresponding center of mass. Based on a data partition in several sub-groups, the pairwise distances can be further compared and ranked to define highest similarities. Most of the image classification or segmentation methods are based on such similarity evaluations. The Euclidean distance is generally used as criterion function. In the

---

[2]Often coupled with some object lists, (e.g. color table, clusters, set of points, lines, edges, regions, etc.)

[3]In the literature, several other color spaces were proposed (HSV, $CMYK$, etc.) as well as the transformations between the different color spaces [192]. In [191], an interesting review is presented with dedicated image analysis processes.

| Distance | Properties | |
	Equation	Complexity
Manhattan	Distance traveled along every axis in the feature space $$|m_1 - m_2|$$	+
Euclidean	Equally variable weighting $$(m_1 - m_2)^T \ (m_1 - m_2)$$	++
Mahalanobis [138]	Correlation between pt_1 and pt_2 integrated: scale invariant $$(m_1 - m_2)^T \ \Gamma_2^{-1} \ (m_1 - m_2)$$	+++
Bhattacharyya [16]	Equivalent to the distance between two Gaussian distributions $$\tfrac{1}{4}(m_1 - m_2)^T \ (\Gamma_1 + \Gamma_2)^{-1}(m_1 - m_2) + \tfrac{1}{2}log\frac{\tfrac{1}{2} \ |\Gamma_1 + \Gamma_2|}{\sqrt{|\Gamma_1| \ |\Gamma_2|}}$$	++++

Table 2.1: Distance metrics in multidimensional space.

literature, several other distances are given (Manhattan, Mahanalobis, sup, etc.) We reported in Table 2.1 a list of metrics employed with the following notations: In a p-dimensional space, for a given set of points $< pt_i >$, m_i denotes its vectorial mean and Γ_i its covariance matrix.

The celebrated *K-Means* method [137], applied with $k = 15$ clusters on a color image is presented in Fig. 2.9. Aim of clustering methods is to subdivide the features space in a limited set of clusters through iterative processes without spatial dependencies assumptions. The resulting classification encapsulates thematic content; it can be seen as a data reduction or a compression process. The corresponding spectral partitioning in the RGB feature color space is shown in Fig. 2.9(d). Clustering methods are described more in detail in section 4.1.1, page 55.

Image Visualization and Compression

Visualization and Encoding are probably the best examples of applications of indexed data with images. Image grey values or color are indexed with a particular *keyword* referring to a Look Up Table (LUT). Hence, providing at most one code per pixel, computational complexity or/and space representation can be drastically reduced. According to the nature of the data to be visualized, LUT can be used for scaling efficiently the data to the range of a monitor (usually 8 bits per color).

In compression, *Huffman codes* are based on the idea of entropy (i.e. quantifying the amount of information, see section 4.3.1) to elaborate a variable-length code table for the encoding of symbols (prefix code) in a binary tree. Symbols frequencies are sorted out and used to set up the tree's bottom nodes. Other nodes are filled out by a recursive bottom-up approach until the root node is reached. Pairs of nodes with the smallest probabilities are replaced by a parent which inherits the sum of their probabilities. The binary tree hierarchy provides a uniqueness code for each symbol. Such a code is optimal for the purpose of symbol-to-symbol encoding. The use of dictionary to archive most occurred sequences or groups of symbols generally offers a better compression rate than Huffman's code, the dictionary can be encapsulated in the encoded data-set (*LZW* algorithms).

Sequence occurrences can also be linked to spatial relationships. The *Run Length Encoding* method is generally applied line by line; instead of archiving data for each pixel, only new data content is saved with a length indicating its occurence. It should be noticed that the compression rate without loss of information is directly dependent on the data content and on its ordering. In some cases, (e.g. highly textured image) the compression is larger than the original data represented on a regular structure (rectangular grid). The data operations (e.g. insertion, fusion, deletion) generally require dedicated processes with higher computational cost and the disposal of both indexed image and dictionary.

(a) Original RGB image (size 979×454)

(b) *K-Means* clustering with 15 clusters.
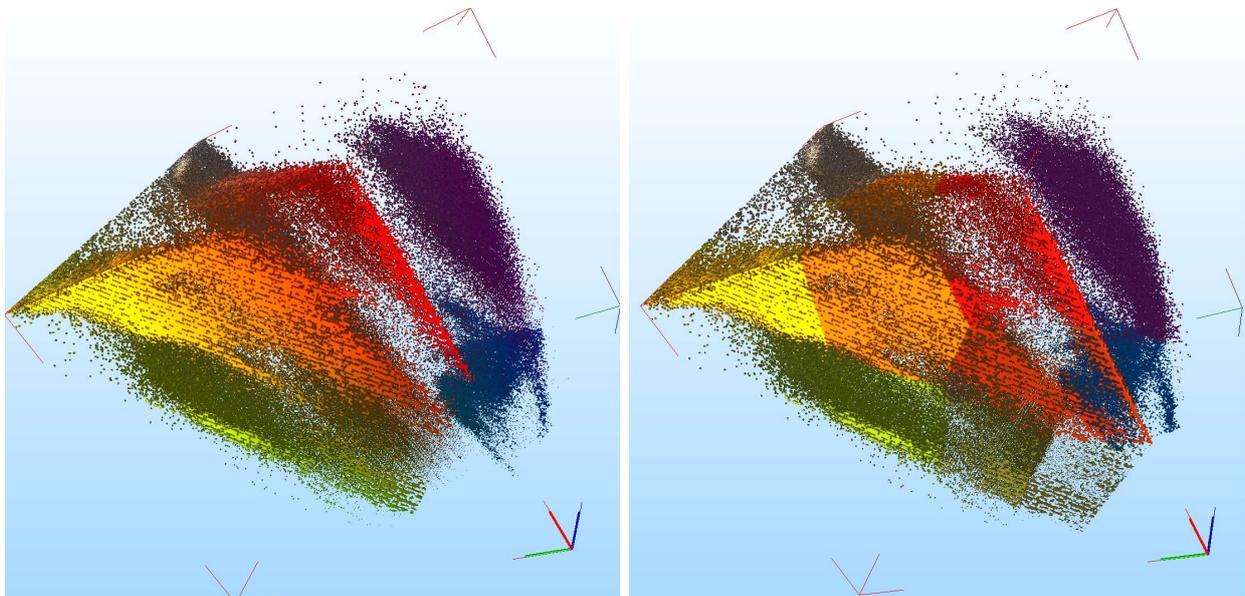
(c) Fig. 2.9(a) in the RGB feature space

(d) Fig. 2.9(b) in the RGB feature space

Figure 2.9: Image classification (b) of a color image (a) and its 3-dimensional representation in the RGB feature space (c), (d). The referential origin is on the bottom-right side of each image where every axis is associated to a color (pixel intensities are mapped as point coordinates in the RGB space). In (d), points which belong to the same cluster are mapped with its cluster centroid "color", reflecting a Voronoi tesselation.

2.2   Image Analysis and Computer Vision

As suggested in the example of a computer vision system depicted in Fig. 2.10, definitions and concepts used throughout the thesis are made explicit.
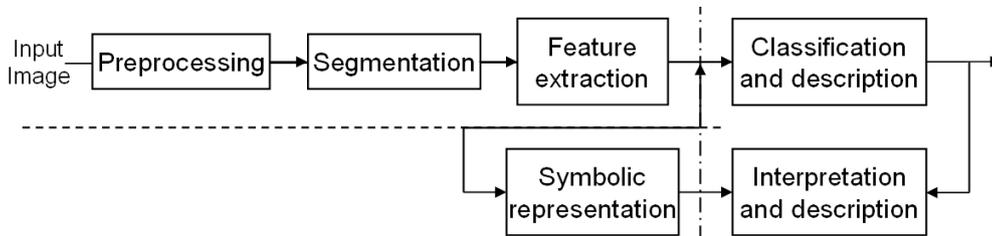
Figure 2.10: An example of flowchart for a computer vision system. The dash line denotes the separation between image analysis (top) and image understanding (bottom) while the dash-dot line distinguishes data analysis (left) and interpretation (right).

Any image operator system can be designed in the form of an input/output system as presented in Fig. 2.11, f being a forward model. Filtering/restoration, enhancement/transformations or coding are typical ***image processing*** operations. Such methods mainly act as pre-processing stages and consist in "image to image transformation" where both input and output are images.
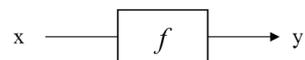
Figure 2.11: Abstract formulation of any image processor, modeled by an input/output system.

In ***image analysis***, by means of segmentation, feature extraction and classification techniques (see Fig. 2.10), results are produced in terms of sets of features, patterns, regions. Those outputs are further investigated and analyzed, with the assumption that this higher level of information expressed on a much compact form represents meaningful content. The reconstruction, interpretation and understanding of a 3D scene from its corresponding 2D image(s) is the objective of a ***computer vision*** system. It attempts to elaborate an explicit and meaningful description of physical objects using some quantitative measurements of the structures in the 3D scene.

Despite intensive research, image analysis and computer vision are still unresolved problems: difficulties lie in (1) their formulation as ill-posed inverse problems as well as in (2) noise and uncertainties which pervade images. Another important limitation for the analysis of "natural" images is their non-stationarity properties, which require local analyses and adaptive noise models. The image analysis formulation as an inverse problem, combined with the non-stationary image characteristics, explains the growing interest for Bayesian methods in image processing. An introduction to Bayesian analysis is provided in Annex B.1, its application for noise removal is studied in chapter 7.

2.2.1   Linear Problems Estimation

In this section, we give a short summary of some linear algebra concepts which are used as base foundations in numerous computer science techniques. Let's denote the following linear system given the vectors x (n variables), b (m variables):

$$A\ x = b \tag{2.3}$$

A indicates a $m \times n$ matrix defining a linear mapping from the vector space x to the vector space b. The subspace of b which can be "reached" is called the "range" space $\mathcal{R}(A)$, its dimension is named the "rank". A non-singular matrix ensures a rank equal to n while a singular one requires the definition of a "null" space $\mathcal{N}(A)$ to complete the missing range components. Given the matrix A, a more formal definition of both subspaces is

$$
\begin{aligned}
\mathcal{R}(A) \quad &= \{x \in \mathbb{R}^m | A\ x = b \quad \text{ for some } x \in \mathbb{R}^n\} \\
\mathcal{N}(A) \quad &= \{x \in \mathbb{R}^n | A\ x = 0\}
\end{aligned}
$$

It comes out that $\mathcal{N}(A)$ is the set of homogeneous solutions to Eq. 2.3. As a consequence, the set of solution to Eq. 2.3 is a particular solution increased by $\mathcal{N}(A)$.

Least square (LS) problems are formulated in finding an approximation of x which minimizes the quadratic loss[4]

$$\min_{x \in \mathbb{R}^n} \ || \ b - A\ x ||^2 \tag{2.4}$$

Various techniques have been suggested, we focus our attention on a particular one, known for its robustness and efficiency.

Singular Value Decomposition

Several methods have been formulated to decompose sets of equations or matrices into more useful elements. Compared to *Gauss Elimination*, *LU* or *QR* decomposition, the so-called singular value decomposition (SVD) method [194] appears to be extremely robust to provide a diagnostic and a numerical result for any studied system.

SVD consists in a factorization of a given $m \times n$ matrix A into the product $U \ \Sigma \ V^T$ of a $m \times n$ column orthogonal matrix U, a diagonal $n \times n$ matrix Σ and a $n \times n$ unitary matrix V^T. The existence of this decomposition is proved for any matrix A, with $m \geq n$, the diagonal matrix Σ being populated of positive or zero diagonal elements w_j, $j \in \{1...n\}$ ($w_j > 0$ being the *singular values*). An interesting property of SVD lies in the estimation of the eigenvalues; if A has eigenvalues λ_i, its singular values are $|\lambda_i|$. Indeed, SVD builds orthonormal bases for both range- and null-space respectively characterized by $w_j \neq 0$ and $w_j = 0$.

---

[4]Statistical estimators can help to quantify the estimate accuracy. In section 7.2.1 some of them are introduced, such as *Mean Square Error* (MSE) or the *signal to noise ratio* (SNR).

This factorization in unitary matrices[5] leads to numerous relevant properties and applications: stability to approximate A, minimal low rank estimation of A, null-space determination, suitable manipulations and analyses, pseudo-inverse or least square fitting computation, etc.

When A is a square matrix (i.e. $m = n$), the inverse of A is easily computed since inversion of the square matrices U, V and Σ is trivial

$$A^{-1} = V \left[diag \left(\frac{1}{w_j} \right) \right] U^T$$

A solution to the system (Eq. 2.3) is then computed by

$$x = A^{-1} b$$

where elements of $diag(1/w_j)$ with $w_j = 0$ are simply replaced by zeros. Special attention has to be paid to possible round-off approximations. In the case of an overdetermined set of linear equations ($m > n$, i.e. having at disposal more observations b than unknowns x), the same rules are applied to compute a least square solution, $\{b, x\}$ being vectors with $\{m, n\}$, and U^T a $n \times m$ matrix. The SVD method always provides a solution; the one among all possible solutions which minimizes $\| x \|$ (in the minimal mean square error (MMSE) sense).

Principal Component Analysis (PCA)

PCA, also known as the discrete Karhunen-Loève or Hotelling transform, is a wide-spread method for the analysis of multidimensional data-sets [54]. Very often, due to possible correlations, the relevant data almost lies within a much smaller-dimensional space. Such a projection or at least a decorrelation of the data presents several interests in image processing: better separability for feature or pattern extraction, image compression, data reduction, etc.

A least square formulation of this problem is the following; *find a subspace that captures most of the summed squared vector lengths of the data.* In least square terms, the PCA is the optimal linear orthogonal transform: a new subspace is built by finding the directions which maximize the variance of the data projected onto that direction. Axes are sorted according to the magnitude of their computed variance. As shown in Fig. 2.12, the one which presents the largest variance is called first principal component. In this scope, for n-dimensional analysis, the new subspace is found in computing the eigenvalues of the covariance matrix C_x. C_x is a $n \times n$ symmetric matrix, where the diagonal terms are the variance of x_i

$$c_x(i,i) = E \left\{ (x_i - m_i)^2 \right\}$$

,the other terms measure the degree of the linear relationship between x_i and x_j (covariance)

$$c_x(i,j) = E \left\{ (x_i - m_i) \ (x_j - m_j) \right\}$$

---

[5]Normal $n \times n$ matrix M which satisfies $M^{-1} = M^T$ and $M^{-1}M = M \ M^{-1} = I_n$

(a) SPOT5 Image   (b) New Spectral Space   (c) 1st component

Figure 2.12: Principal component analysis applied on a multispectral optical Image (SPOT 5-Fusion) (a) (size 512×512, resolution of 2.5 meters). In (b), the principal component is the red channel, the second one is the blue channel. (c) is a grayscale image of the principal component.

Interestingly, if we use the SVD, we notice that the product of x^T by x leads to

$$x^T \ x = V \ \Sigma^2 \ V^T \tag{2.5}$$

which is the expression of the covariance matrix when x is centered (zero mean).

The SVD is a suitable and efficient method for solving LS problems, eliminating correlations or reducing data volume to its main relevant content. Under the assumption that the variance, which represents the data dynamic, is somehow a way to "quantify" information content, PCA appears to be an efficient method. However, PCA will fail when largest variances do not correspond to the meaningful axes. An alternative is the use of "Independent Component Analysis" (ICA) [39].

2.2.2 Low vs. high Level of Analysis

Marr's pioneer work on computational vision, summarized in his book [146], dramatically influenced vision research by providing founding bases for the description and the design of a vision system. According to him, visual perception is decomposed in a sequence of increasingly symbolic scene representations, ranking from a **primal sketch** of the retinal image, through a $2.5D$ **sketch**, to simplified object's $3D$ **models**. The first level consists in extracting primitives and features through the location of intensity changes. To tackle this level, he suggests in [147] the use of the Laplacian of a Gaussian. Indeed the "zero-crossings" are carefully interpreted as contours or boundaries, mainly because of the noise sensitivity and the meaningless relationship between local intensity changes and edges.

Results obtained are used as input in the second stage and run over processes such as object recognition. At this level, local depth and surface orientation are made explicit. The 3D information detected is combined to perceive the localised depth of objects. Finally, by combining all visual

information and scene conceptual hypotheses, the scene interpretation is achieved by the generation of $3D$ object-oriented models describing the scene.

If most image analysis methods introduced by Marr for each detailed module have been outperformed or encountered severe limitations, his strategy of dividing the problem in several levels of analysis is still up-to-date. It enables more robust and consistent framework than heuristic or/and empirical, "Ad hoc" methodologies previously employed.

It should be noticed that the presented modeling applied as a sequential process, and particularly the lower levels employed, are generally considered as **bottom-up** approaches. Unfortunately, **top-down** approaches both within the same level or between the levels are mostly neglected even if synergies and/or feedbacks are highly expected.

Recently, other theories already formulated before the advent of computer vision have been addressed for solving the visual perception paradigm. We refer to Morel's work [153], [49], [25] for readers further interested in Helmoltz's principle, the economy principle of Ernst Mach (XIXth century) and Gestalt theory or Shannon's information theory (XXth century).

2.2.3 Local Analysis and Edge Detection

As noticed in Fig. 2.10, feature extraction (e.g. textures, edges) and segmentation[6] processes are applied on the earlier stage of a computer vision system. Objects embedded in images are characterized by several features describing radiometric and/or textured properties as well as shape information (geometry and relationships).

Edge Detection

Low-level image analysis aims to identify image discontinuities which correspond to physical boundaries between regions or objects. Under the formulation of an image as a continuous function, the computation of the image gradient is used as the foundation of edge detection algorithms. A further stage, called *non-maximum suppression* (NMS), identifies edges based on the assumption that contours correspond to local maxima (gradient's norm) in the gradient direction.

The Sobel edge detector [99] computes the gradient by means of 3×3 convolution kernels in respectively horizontal and vertical direction. Despite its high computational efficiency, this edge operator is extremely sensitive to noise and requires first an image convolution with a low-pass filter. In the 1980s, more analytical techniques were suggested based on Gaussian derivatives. In [147], Marr and Hildreth discussed the zero-crossing computation of the image's Laplacian. Canny formulated in [24] an optimal edge detector in terms of localization, uniqueness and robustness to

---

[6]Segmentation processes aim to subdivide an image into disjoint and homogeneous entities (regions) presenting simultaneously sharp boundaries' discontinuity and high "internal" similarity. Region based techniques are explored more in detail in chapters 4- 5.

noise with the incorporation of noise and contour models. In [48], Deriche implemented the optimal Canny's criteria in a recursive filter.

Laplacian's method presents the main advantage of defining closed contours (zero-crossing is defined as a connected isophote). However, the performance of this operator is strongly limited by its too high sensitivity to noise and local variation (textures for instance). In addition, its isotropic property[7] prevents the extraction of T junctions. If the Canny-Deriche filter appears to be more robust to noise (controlled by the σ parameter), it requires further processing stages to extract edges (NMS) and close them with heuristic techniques (e.g. *hysteris thresholding*) which rapidly become extremely complex[8].

To extract some edge junctions (corners) or invariant points, another celebrated method related to image gradient computation is the Harris detector [98]. Using the image gradient, the eigenvalues of its covariance matrix are estimated. When both values are high, the local auto-correlation function is sharply peaked, indicating that the signal changes simultaneously in both directions. To reduce the computational complexity, a good approximate of the eigenvalues product is derived based on the determinant and the trace of the covariance matrix.

As noticed by Marr himself, the detection of edges for a fixed scale is not sufficient and more generally, edge detection algorithms often give "meaningless" contours. Another drawback of the presented methods concerns the blurring effect resulting from the Gaussian filtering.

Texture Analysis

In human visual perception, textures act as a determinant feature. The term of texture denotes a visual phenomenon caused by the repetition of a structural element according to a certain rule. Indeed, most natural surfaces show regular structural patterns (e.g. grass, forest, wall of brick, marble, etc). However, classical methods introduced in the previous sections (K-Means classifier, Edge extraction) are ineffective for the extraction of such areas. In this scope, analyses of textures have been extensively made in computer vision (see [82], [81], [99]) and modeled in images as a spatial distribution of gray-level variations[9].

Several properties are usually employed and combined for the description of textures. Among them, one can distinguish for instance roughness, uniformity, density, regularity, directionality, frequency, etc. Unfortunately, even by using data analysis operators (PCA, ICA) to reduce the dimensionality to the most meaningful features, the exploitation of those characteristics is strongly limited by their interdependency and representation at various scales.

In the literature, texture discrimination is mainly subdivided in three directions: geometrical, model-based and statistical approaches. Textures are modeled in geometrical approaches by a composition of primitives or structuring elements. Firstly primitives are extracted by means of

---

[7]No directional information taken into account, rotation invariant.

[8]See for instance the extraction of road or network from aerial images in [89].

[9]Such a model can be expressed as well in form of stochastic process, introduced in annex B.1.

spatial-, Fourier-, wavelet- methods (frequency analysis, e.g. edges density, gradient amplitude, Laplacian of Gaussian at various scales, etc.). Secondly, their relationships are analyzed to retrieve the placement rules which characterize each of the studied areas. Markov random fields as well as fractal dimension analysis denote model-based approaches. Both of them are discussed in the dissertation (see respectively sections B.1 and 2.1.2).

Statistical methods provide two powerful tools for the analysis of textures; two-dimensional histogram and co-occurrence matrix [81], both of them being explicitly related to the gray-level's spatial distributions within images. The derivation of **first-order statistics** is achieved by histogram techniques; mean, standard-deviation, high order moments, energy, entropy, etc. are classical features which belong to this category, all of them being locally computed without considering any interaction between neighboring pixels. **Second-order statistics** characterize pairs of gray-values within a defined image area. The occurence of the pairs of intensity values is stored as coordinates in a symmetric matrix. The co-occurrence matrix, bounded to $[0,1]$ after normalization, is further used to compute contrast, homogeneity, correlation features, etc. The computational complexity is generally reduced by previously decreasing the image intensity range (image quantization). We refer to [81], [99] for finding explicit equations of those features.

2.2.4 Classification Techniques

Image classification is a classical process in digital image analysis widely studied and reported in numerous publications. The aim of image classification is to categorize all pixels in a digital image map which describes one or several land cover classes. The resulting thematic image shows the pixels that belong to similar spectral patterns or features. Jain in [99] and [100] addressed two excellent state of the art in image classification techniques. Duda's book [54] constitutes a solid reference for further theoretical points.

Image classification is often realized on multi-spectral images; a feature space can be sketched in constructing an image description corresponding to a spectral or color band for each axis. Mapping all image pixels in this defined feature space will result in the emergence of "packets" or groups of pixels close to each other called clusters. A *cluster* is a collection of objects which are all "similar" (between them) and are "dissimilar" to the objects belonging to other clusters. The objective of image classification is to distinguish and extract those clusters.

These patterns do not necessarily correspond to direct meaningful characteristics of the scene, such as contiguous and easily recognizable areas of a particular soil type or land use. They are simply clusters of pixels with similar spectral characteristics. In some cases, it is more efficient to identify groups of pixels with similar radiometric characteristics than to sort the pixels into recognizable categories. Classification methods are generic, they can be applied on spectral contents, sets of statistics, texture parameters or shape descriptors. It remains only to define a suitable measure of similarity and if necessary apply normalization constraints between the various features to obtain homogeneous entities.

The described algorithms and concepts are restricted to exclusive classification where each object of the data-set to be classified belongs exactly to one subset of clusters. Non-exclusive classification, such as fuzzy clustering, is not dealt with here, the reader is referred to Shepard and Arabie in [186] for an interesting review of non-exclusive or overlapping classification.

The two main pixel-based classification approaches are non-supervised and supervised methods. Both are presented respectively in section 4.1.1, page 55 and section 6.3, page 108.

2.2.5 Summary and Outline in Image Analysis

For our purposes, what can be summarized from Marr's vision theory and the image properties, and from analyses presented in this chapter? Vision problems are subdivided into respectively low-, intermediate- and high- levels of image description. At the low-level, the description is sought in terms of edges, corners, ridges, elementary (part of) primitives such as lines or circles. The features are extracted by means of edge detection, texture analysis, segmentation (chapters 4- 5) techniques. Object recognition and full three-dimensional scene description (chapter 8) is expected at the high-level of analysis by means of, for instance, geometrical and topological image representation or image classification techniques (chapter 6).
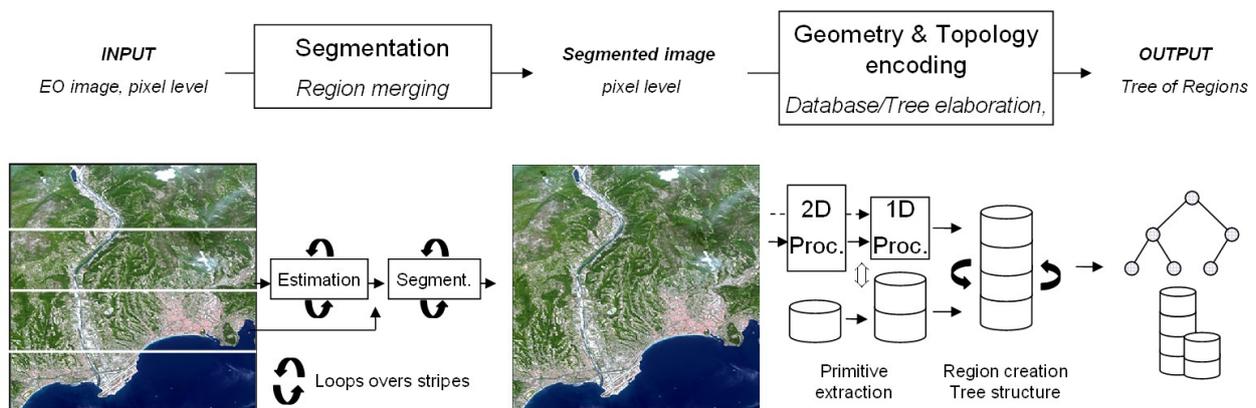
Figure 2.13: Flowchart of the computer vision system proposed. Starting from a large EO image, two independent processes, representing respectively low-, intermediate- level of analysis, are designed to achieve an object-oriented description of the image.

Fig. 2.13 depicts the flowchart of the computer vision system proposed in this dissertation, mostly applied for optical color or range images. Based on the results obtained, image analysis methods such as clustering or supervised classification are investigated to reach a higher description and understanding level of the studied data-set. Finally, one keeps in mind that images are pervaded with noise and artifacts which need to be tackled by an early image filtering step (see chapter 7).

Earth Observation

Abstract

An overview of Earth Observation techniques is presented. Compared to classical topography campaigns, remote sensing offers numerous advantages such as repetitiveness, all time-weather capabilities, wide range of spectral analyses, automation, flexibility of acquisition as well as important orbital accuracies. The potential carried by automatic acquisition to deliver material for the elaboration of cartographic products has to be reinforced by improved techniques and processes to generate reliable and enhanced data-sets. In particular, the interests and the availability of large and homogeneous digital elevation models are emphasized.

Methods to generate elevation information from EO images are investigated, their characteristics and properties are described. According to the resolution and the type of acquisition, the information content embedded within elevation data is summarized. Indeed, the latter is mainly composed of relevant information, typical errors as well as artifacts and noise resulting from several listed conditions. Improvement of digital elevation models is one of the main objectives of the present work.

Introduction

According to the available techniques, several methods have been developed to describe the relief in detail and quantitatively. The first representations were drawn on maps, called topographic maps, based on military topographic surveys, since relief information was of vital importance to assist in planning for battle and identifying strategic emplacements. Over the years, modern nations established national institutions to extend and improve territory descriptions. Topographic maps became a basic resource in planning infrastructure and resource exploitation.

In parallel with relief surveying, Earth's surface cartography integrates always more and more accurate and sophisticated natural, man-made, administrative features such as forests, rivers, roads, buildings, cadastral or political boundaries. Some thematic and semantic descriptions were added (e.g. nature of the vegetation, name or complementary information about the various mapped objects, etc.). Conjointly, a large amount of data and statistics were collected and archived to describe physical phenomena such as climate or human characteristics (population density, health issues). Their integration on geographical systems enables to analyze their respective spatial variations and interrelationships while providing a "visually" meaningful and convenient way to handle databases or information systems.

To define accurately the location of each topographic feature in 3D coordinates, geographic coordinate systems have been set up using local or global ellipsoid approximations of the Earth and datums. Territories are subdivided in portions of local ellipsoids which match a surface defined according to an arbitrary elevation (zero level's sea) and a planimetric set of reference points. Interlinked networks of physical reference points settled by triangulation and conversion tables enable to associate surveyor's mapping or photogrammetric campaigns to homogeneous and unified coordinate systems, which relaxes transition constraints between local and nearly global references.

An alternative to the physical network of triangulated reference points is possible with the advent of the Global Positioning System (GPS) and its associated ellipsoid (WGS84). Based on an ascending system composed of an antenna at the studied location and a constellation of satellites, it permits accurate tridimensional location at every clear position. Centimetric accuracies in both planimetry and altimetry are reached with differential GPS techniques.

Airborne first, then spaceborne platforms carry accurate sensors to monitor and cover landscape, forest, network or city areas. Some of the resulting images of the remotely sensed areas provide sub-metric details for several parts of the electromagnetic spectrum (e.g. optical, radar, infra-red waves). Techniques as SAR Interferometry (InSAR) or optical stereoscopy have reached an operational level to deliver "acceptable" elevation data.

Compared to cartographic surveying, interests of Earth Observation technologies have already been proven to be multiple: short observation cycles on each part of the globe even under critical weather conditions or relief, wide range of spectral analyses, large and homogeneous coverage, etc. However, the data are free from a-priori knowledge; it induces important processing chains to extract the required information. Indeed these processes are different from traditional mapping which generates relevant information and deduces the area whereas remote sensing (RS) survey acquires the area and then extracts significant information within boundaries. This remark emphasizes the importance of pre-processing, extraction and interpretation phases to optimally exploit RS data.

The availability of more complex and various EO data induces some migration of classical RS image analysis methods from large scale modeling based on global statistical models to computer vision techniques with non-linear objects or features, especially in man-made areas. Optical and SAR metric images presented in Fig. 1.1 emphasize those remarks. In this scope, the RS community is faced with challenging issues when developing methods to

- handle an increasing amount of data archives due to data's higher complexity and content,
- filter and enhance the acquired data,
- extract and model relevant information and exploit synergies between data,

Objective of this chapter is mainly to address a foundation of EO data and more especially on elevation data, derived mostly from InSAR or optical stereoscopy techniques. An overview related to SAR processing is given in annex A.1, (pages 168- 174). Information content as well as possible sources of errors are discussed.

3.1   Elevation Information

In geo-spatial analysis, terrain elevation is one of the most essential and stable piece of infor-
mation. Discrete two-dimensional functions enable to create a digital representation of the terrain
surface. Its numerical representation of heights is based on a 3-dimension plot and further additional
geometric information to model the surface.

3.1.1   Taxonomy

In the 1950s, the earliest definition of a digital terrain model (DTM) was proposed [152] as
"*a statistical representation of the continuous surface of the ground by a large number of selected
points with known xyz coordinates in an arbitrary coordinate field*". Later, other concepts such
as digital elevation model (DEM) or digital surface model (DSM) have been introduced to define
in a more suitable way the discrepancies between various information content or the different
meanings resulting from context-dependent analysis and applications. To clarify these concepts,
we summarize some of the definitions found in the literature:

In [108], [132], a **Digital Terrain Model** is an umbrella concept covering models of elevations
and other geographical elements as well as natural features, such as rivers and other break lines.
It can also include derived data about the terrain, such as slope, aspect, curvature, visibility, etc.

In [149], a **Digital Elevation Model** is a generic term covering digital topographic (and
bathymetric) data in all its various forms as well as the method(s) for interpreting implicitly the
elevations between observations. It usually implies elevations of bare earth without vegetation and
buildings, but can include other man-made features, such as road embankments. Elevations of
hydrological features (e.g. lakes and rivers) normally concern a free water surface.

In [132], a DEM is the most fundamental component of DTM. In [108], it corresponds to a
numerical representation of a topographic surface arranged as a set of regularly spaced points -
usually in a square grid or hexagonal pattern - expressed as three-dimensional coordinates. The
latest definition introduced, **Digital Surface Model** is a model depicting elevation of the top of
reflective surfaces, such as buildings and vegetation [149]. Such a definition is mainly used in works
related to airborne laser scanning.

All of them are considered as "2.5"-dimensional representation, i.e in the nature of 2-dimensional
data but contain 3-dimensional (3D) information.

3.1.2   Data Representation

Digital elevation data can be represented and encoded in different data structures (see Fig. 3.1)
such as contour lines, triangular meshes, regular grids or scattered data points.

Contour lines

Originating in analogue topographic maps, the traditional representation of the relief was achieved using elevation isolines stored in contour lines. All points located at a same altitude above or below a zero level are linked together giving a contour line. Using contour lines which "follow" and "overlap" the terrain, structural information is implicitly included and constitutes the most "comprehensive" terrain description. Slopes and local relief's variations are qualitatively and quantitatively described (steepness, orientation). It is not a real 3D product but gives directly relevant information about the area. Some interpolation problems can arise between contour lines.

Triangular meshes

Also known as Triangular Irregular Network (TIN), the structure is based on a connected set of elementary planes (triangular facets) with vertices at the sample points known in xyz coordinates. Such planes are described when joining the three adjacent points from a set of elevation points, and are constructed using Delaunay triangulation. Neighboring and topological information is implicitly represented while preserving local variations in points density or accuracy. The data-structure is therefore able to match locally terrain roughness, to include strong discontinuities and to set defined regularization constraints. A set of algorithms for two-dimensional Delaunay refinement is presented in [187]. In [188], combinatorial foundations of constrained Delaunay triangulations are laid out. TINs enable to improve the level of realism and model the terrain more properly, but the computational abilities are limited and require more elaborated definitions and sophisticated structures to manipulate correctly and efficiently the data-set. For those reasons, computer graphics (e.g. game, 3D rendering), GIS or hydrological applications use triangular meshes most of the time. However, despite its accuracy, this representation fails in describing morphological structure.

Regular grid

The data are given on a square or rectangular grid with constant sampled grid width. It constitutes the most classical type of representation, useful in terms of operational management since it constitutes simple elevation matrices which implicitly encode topological relations between data points locations. From a computational point of view, efficiency and optimization of processes are easier formulated and achieved. But operating on a fixed grid presents strong limitations in terms of reliability since structural information is lost or ambiguous (slope computation for instance) due to the fixed rectangular structure. It does not reflect the variation of accuracy over the grid; the sampling rate of the regular grid has to be adapted to the highest terrain frequency to preserve information and lead to a high volume of data. However, it still represents the standard structure to handle elevation data to be shared between all the various relief data representations.

3.1.3 Applications

Remote sensing provides a wide range of information relating to surface topography. Since elevation models describe the surface topography (geometry), most studies related to remote sensing

(a) Regular grid   (b) Contour lines   (c) Triangular meshes
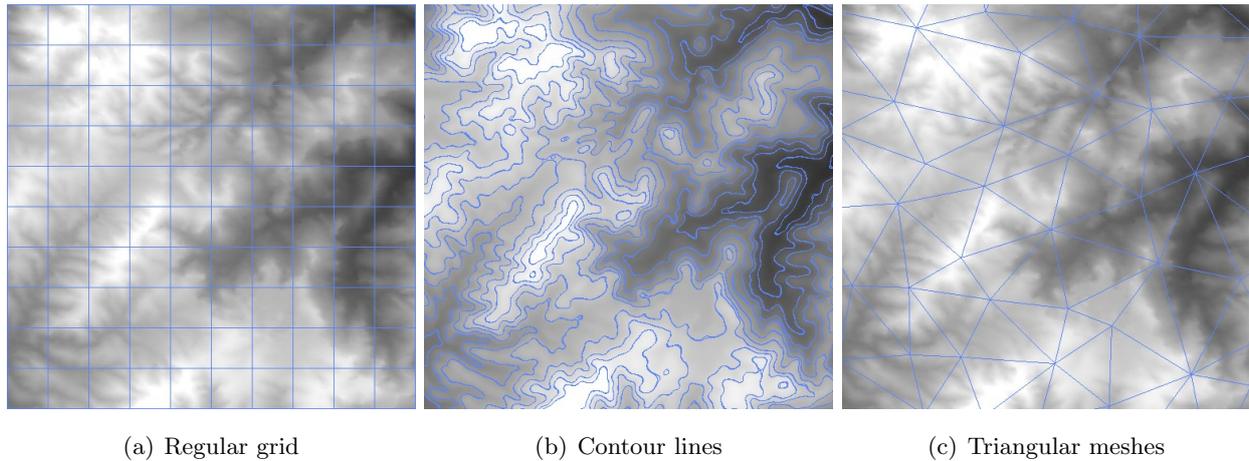
Figure 3.1: Models of elevation data representation. A DTM is inserted in background (size 1024×1024). Bright areas denote high elevations while dark areas represent lower ones. The contour lines extracted (b) are 40 meters distant.

data and applications can be improved or even require elevation information. Elevation data are intensively used in all the stages of EO data analyses or applications;

Image Reconstruction and Analysis: Interferometry processes improvement (SAR processing), orthorectification, image matching in photogrammetry. Elevation data can contribute to the scope of information extraction in bringing complementary information useful for the recognition of features, landscape thematics, or man-made structures such as buildings.

3D Rendering: Three dimensional visualizations are simulated using digital elevation model data overlaid by a corresponding optical or radar image. Several applications such as flight simulations, planning or construction in Civil Engineering have become popular. Intensive research [32], [33], [74] is being carried out for real time 3D rendering of large volume of data.

Integration and Synergies in **Geographic Information Systems** (GIS) between various EO data and databases. It enables the production of cartography in the case of risk monitoring [169], [214], disasters management [143], [139], etc. In addition, the possibility to derive references from elevation data for gravitation surface models is essential for other complementary domains such as Hydrology, Geology, Geophysics, etc. More exhaustive and detailed reviews of DEM applications are sketched in [95], [132].

3.2 Principle and Analysis of DEM Generation from EO Data

Before the advent of spaceborne systems, analytical photogrammetry was the alternative to classical topographic methods for producing accurate elevation information. Both techniques were limited to small areas and numerous manual processing tasks, which induced expensive costs. The possibility to generate large and homogeneous DEM reinforces the relevance of satellite images.

RS optical and radar methods for elevation extraction are based on the exploitation of at least two suitable images. An exception is the method called "clinometry" which might only require one of these sets of data. As the "shape from the shading" method used in computer vision, it is based on the determination of the ground local orientation taken from radiometric information gathered by optical or radar images. A reflexion model is required and some constraints have to be fixed (homogeneity, slope behavior) in order to invert the model. Experiments to derive elevation data from SAR polarimetry ([216], [201]) are presented in [197].

Another method, based only on one acquisition, is the use of LIDAR (Light Detection and Ranging) techniques. Such an active system operates in infra-red and enables to compute DSM with high accuracy. However its use is mainly restricted to localized urban area studies since acquisition costs are important (aerial campaign). Further details are given in section 8.1.1, page 142. Radargrammetry (radar images) and stereoscopy are based on the relation between ground elevation and parallax in image pairs. These two methods mainly differ in their geometric equations and image specific characteristics. By means of phase difference analysis between two coherent radar signals, Interferometric SAR (InSAR) processes ensure a good extraction of the relief.

We restrict our investigations to optical stereoscopy and InSAR methods, both being well established and operational. For complementary studies, we refer to the reviews [197] and [195] about the generation of elevation information from optical and radar images. Both methods are affected by geometrical distortions and the resulting DEM is usually defined in a geographic coordinates system to be integrated in databases.

Topographic DTM content

The formerly most used method to produce elevation data was the interpolation of topographic digitised level curves. If level curves are useful for terrain visualisation and interpretation, their representation as numerical surfaces is difficult. As a consequence, interpolation techniques based on splines functions with constraints (various levels of geomorphological and topological modeling) or gridding algorithms are designed to sample the data on a regularly spaced grid. Such processes generally induce some undesirable effects (e.g. artifacts, smooth behaviors), observed in the Fourier domain, which do not accurately represent the fractal nature of the terrain. Accuracy of level curves relies on the digitization of topographic products whose quality is strongly correlated with the landscape use. For instance, compared to landscape areas, a better accuracy is expected in urban areas since human interests are of higher importance. A compromise has to be reached to set up a fixed resolution according to the heterogeneity and the interests of the DTM coverages.

Interpolation processes result in important data simplifications since high frequencies, objects, break-lines, ridges, or drainage lines are not well integrated, though bring relevant information. "Terraces" in the terrain may occur due to the isoline tracing and its resulting interpolations. In addition, the computed information is different from EO DEM; it refers only to the terrain elevation information, while forest or urban elevation are not included.

Orthorectification

With the improvement of EO technologies, the remote sensing community is faced with the same challenges as photogrammetrists dealing with aerial images. Modern sensors on board of platforms are able to produce high quality and fine detailed images. However, those data-sets are corrupted with geometrical distortions resulting from image and sensor geometry, especially when the satellite operates with important incidence angles of acquisition. Computer analysis on raw EO data is limited and prevents synergies between the extracted information and other sources or supports (images, vectors, maps, etc.). In this scope, geometrical distortions as well as the integration in geo-referential systems are required, conditioning the quality of the application investigated.

The term "ortho-image" denotes a geocoded image[1]. Geometrical transformations are based on sensor modeling as well as accurate orbit positions and orientations. External data, generally reference points, are used to perform the conversion of raw images to georeferenced and corrected images. A detailed review on geometrical processing from RS images is given in [196].

The rectification under the assumption of a mean height plan is not valid in mountainous regions: some local relief geometric distortions must be taken into account (e.g. an existing DEM) to obtain suitable and consistent results. The magnitude of the planimetric error Δx resulting from an altimetric error Δh of the DEM may be sizeable: in a first approximation of a flat terrain and for a constant incidence angle, we obtain the relation $\Delta x = \Delta h / tan(\theta)$ [2].

3.2.1 Optical Stereoscopy

This section gives a short introduction to the stereoscopy principle for the generation of digital elevation models from a pair of EO images. For a deeper and more consistent presentation as well as the latest works on this topic, the reader is referred to [44], [121], [130].

Introduction and Principle

Stereo-vision, the foundation of photogrammetry, directly exploits the same principle as the human vision to retrieve 3-dimensional information. Our "binocular" vision operates as observing a scene from two distinct positions S_1 and $S_2 = S_1 + dX$ as described in Fig. 3.2. The shift dX causes a human's physiological parallax, also called stereoscopic or X-parallax, X being the shift direction. A further process, called "convergence", is required and denotes our ability to focus the pairwise information located in each image to a single object. The amount of parallax computed is exploited to derive elevation or depth information on the area sensed by the two acquisitions.

---

[1] An image where all distortions including relief are corrected, i.e. can be overlaid by topographical maps

[2] Under these assumptions, with the acquisition characteristics of the European Space Agency SAR satellite ERS, an error of 100 meters in elevation results in an error of 235 meters in planimetry...
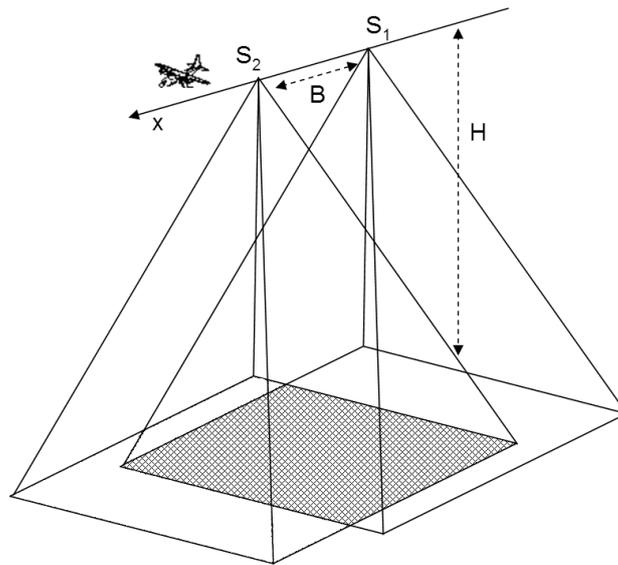
Figure 3.2: Image Stereoscopy principle based on at least two images acquired from slightly different positions (S_1, S_2), B denotes the baseline.

An accurate image acquisition modeling is necessary and decomposed in two parts called respectively internal and external calibrations [119]. The first one is related to the sensor properties themselves, whereas the second one denotes its positions and orientations. Using ground control points as references, parametric models or polynomial/rational functions are generally employed [173] to estimate and calibrate internal and external orientations. Before the advent of even more accurate EO data covering always larger territories, the stereoscopy principle was applied on aerial photographs with analytical stereo-plotter. In this way, the two photographs are fitted to the human's binocular vision and it remains to manually plot the corresponding points in both images.

Faced with a growing amount of data, this task, which requires experienced operators, is quickly becoming time-consuming, even with the "help" of epipolar constraints. Indeed, the localization of a point in picture 2, given its corresponding point P_1^i in image 1, is reduced to an epipolar line; the intersection of the picture with the 3D plane formed by (P_1^i, C_1, C_2), C_j being the focal point of the picture's j sensor system. Applied on RS images, photogrammetric principles are kept. Solving the correspondence problem, a field also called "image matching" has been intensively studied to set up automatic processes as an alternative to computer-assisted methods. It mainly consists in correlation methods based either on image intensity or features.

Image Acquisition Properties

The combination of parallax and convergence principles leads to compute the intersection of two incidence optical paths of the same sensed object. By triangulation, the estimation of this angle enables to deduce the elevation since the positions S_1 and S_2 are known. The distance dX, also called "baseline" is thereafter noted B.

Depending on B and the sensor elevation from the object H, the resulting angle will vary. The ratio B/H is a determinant criterion for the accuracy of the produced DEM. If too small, it results in a poor elevation accuracy; if too large, a better accuracy is expected but mage matching becomes less reliable and some occlusions appears as a consequence of too different viewing points or sharp transitions in relief. Indeed, relief itself plays a major role; as already mentioned, elevation cannot be computed for regions presenting occlusions in at least one of the pair images (steep slopes, buildings, etc.). Furthermore, a larger parallax is obtained for higher elevations compared to lower ones.

Stereo-pairs can be significantly altered when one or both images are tilted with respect to an exterior coordinate system. This effect is often generated by platform motion or by image pairs presenting slightly different scales. Some studies reported in [195] have been carried out by using sensors operating at different "ground" resolutions or spectral domains.

Image Matching

The rectification process enables to simplify the correspondence between stereo pair images. It denotes a transformation of each image plane so that pairs of conjugate epipolar lines become parallel and collinear to one of the image axes. The epipolar constraint is relaxed in a 1D search problem along the horizontal raster lines or column of the rectified images. In addition, the image pair is generally pre-processed to reduce radiometric differences. Among the methods addressed to solve the correspondence problem between image stereo pairs (see [131]), two main families emerge: correlation- and feature-based methods.

The first group generally follows a maximization of the intensity cross-correlation between two stereo images along common epipolar lines. Local correspondence, computed with the use of a defined windows size, leads to a disparity map out of which depth information can be retrieved. Recently, methods which treat the disparity estimation map as a single energy minimisation problem avoid the planar surface assumption inside the correlation window [180], [90]. This is especially important for metric datasets. Temporal correlation, stereo base size, as well as acquisition conditions are critical factors which condition the matching performances.

Feature based methods directly exploit computer-vision methods. The extraction of primitives enables a significant data reduction. The basic assumption is to focus the matching process on relevant features and on their possible dependencies. Edges-, lines-, corners-extraction algorithms are first applied to both images processed separately. Then, instead of computing the disparities directly on the image information itself, the matching process is achieved on the collected image primitives. Indeed, several features have almost invariant properties and the analysis of their relationships enables the derivation of complementary information or logical rules.

Other hypotheses are investigated to improve the algorithm robustness and avoid the so-called ghosts, aliasing effects, etc. caused by erroneous matching. Such assumptions, formulated as con-
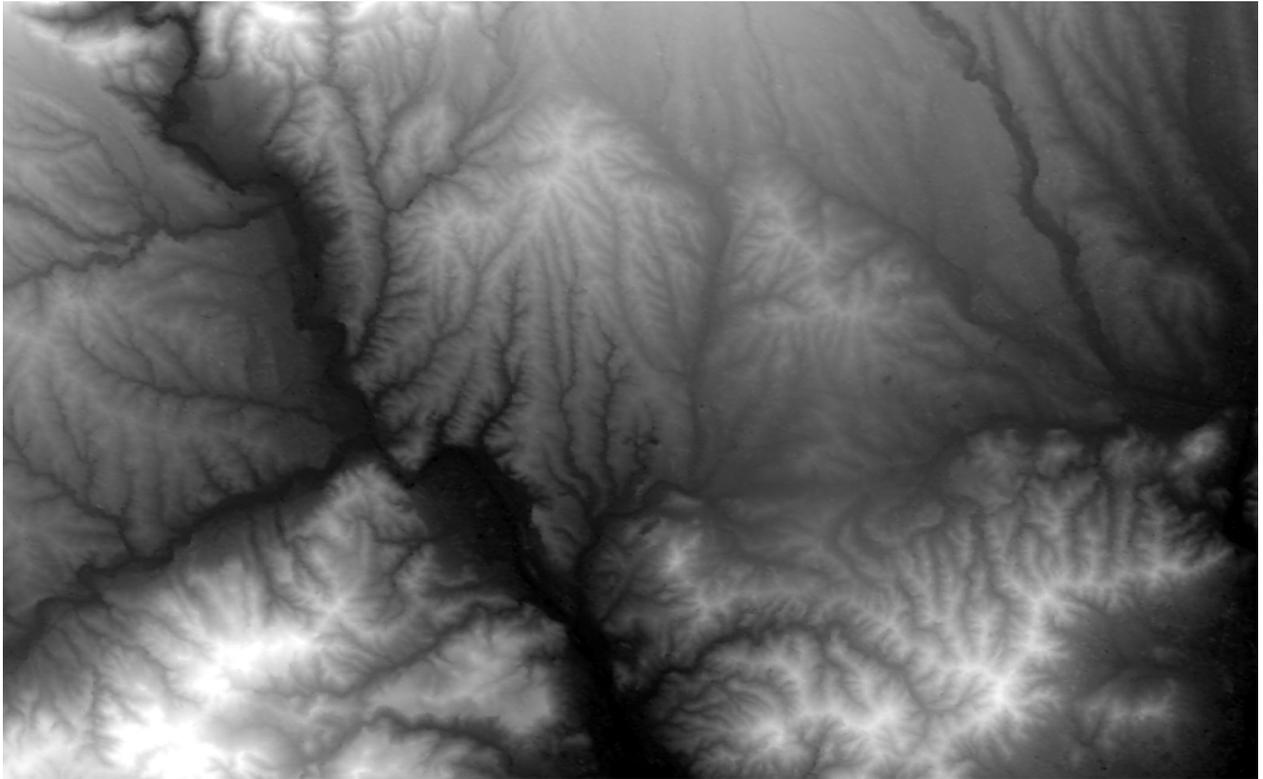
Figure 3.3: Example of Optical Stereoscopy DEM based on a pair of SPOT 5 panchromatic images (nadir looking, 5 meter resolution) [173] sampled at 15 meter resolution. The depicted area is located near Barcelona (SPAIN) and represents a 30 kilometer-large coverage with altitude ranking from about 30 to 550 meters (a Gamma correction was applied to improve DEM's visualization).

straints, are classified into two sequential levels to reject or validate a correspondence between features and solve ambiguities. Firstly, similarity measures (e.g. intensity correlations, etc.) are computed to find possible correspondences. Secondly, to validate or not the pair studied, uniqueness, ordering and completeness constraints are tested, as well as boundary conditions of, respectively, the epipolar range and the resulting disparity map. To maximize the image matching, subsequent heuristic strategies have been put forward, combining several primitive types extracted at various scales and applied until satisfactory results are obtained. The level of knowledge to be ingested depends on the accuracy required and on the type of applications or data.

Summary in optical stereoscopy

Interests of optical stereoscopy for the generation of elevation/depth information have been established in photogrammetry, remote sensing as well as in computer vision. Intensive research is carried on the improvement of image matching methods and the development of automatic systems for the production of accurate high resolution DEM. A DEM generated [173] from a pair of SPOT 5 panchromatic images is presented in Fig. 3.3. An automatic processor is currently

in development for the Indian IRS-P5 satellite with the Cartosat instrument on board. This along-track stereoscopic push-broom scanner (2.5 meter resolution) appears to be well-suited to derive homogeneous elevation data models on continent scale. In [44], preliminary results are promising with location accuracy (planimetry and altimetry) within an order of 3-5 meters outside mountainous areas without ground control points. The alternative to ground control points derived from topographic survey is the use of lower resolution data available on almost each part of the globe such as SRTM DEM or Landsat ETM+ data. SPOT 5 and Cartosat DEMs, available on the same area, are used in the thesis (chapters 7- 8).

3.2.2   SAR Interferometry

A radar sensor is an active and coherent system operating in the microwave frequencies with all- weather and time capabilities. The principle of imaging radars is to generate a reflectivity map (of an illuminated area) by means of transmission and reception of electromagnetic energy. The development of the radar system was limited by its very low resolution. These constraints were by-passed using two specific techniques: Synthetic Aperture Radar (SAR) [211] and pulse compression [136]. For a comprehensive understanding of this field, which radically differs from our natural perception/vision, some background related to radar imagery is discussed in annex A.1.

Information content of SAR images is not only "embedded" in the amplitude of the back-scattered emitted microwaves. Two important applications such as topographic mapping and change detection exploit phase information in studying interference patterns created by the combination of two sets of radar signals (phase of a single radar signal is not exploitable).

Interferometric SAR (InSAR) has been suggested and successfully demonstrated as a topographic mapping technique by Graham [77], Zebker [215] and Bamler [11]. For a couple of years, SAR and InSAR processing techniques, generating back-scatter intensity, DEM and interferometric coherence samples, have been reaching an operational stage.

Introduction and Principle

A radar interferometer is formed by relating the signals from two spatially separated antennas observing the same area. In an interferometric radar system, this spatial extent B - called *baseline* - is one of the major limiting factors in terms of performance. If B is too short, the sensitivity to signal phase differences is undetectable; if too long, additional noise due to spatial decorrelation corrupts the signal. Two distinct approaches have been discussed for topographic radar (across-track interferometers) and differ in how the baseline is elaborated (see Fig. 3.4):

1 - "Single-pass" - The baseline is formed by two physical antennas which illuminate simultaneously a given area on the ground. This approach is usual for aircraft implementations where physical mounting structures enable a sufficient baseline [215]. It has been the solution retained for the SRTM Mission on board of the Nasa Space Shuttle with a 60-meter extended mast. However, it

Antenna	Single-Pass		Repeat-Pass	
A_1	$E_m - R_e$	t_0	$E_m - R_e$	t_0
A_2	R_e	t_0	$E_m - R_e$	t_{0+dt}

S_i	Radar Echoes
R_i	Ground Ranges
E_m	Signal Emission
R_e	Signal Reception
B	Baseline Length
α	Baseline Angle
θ	Look Angle
H	Sensor Altitude
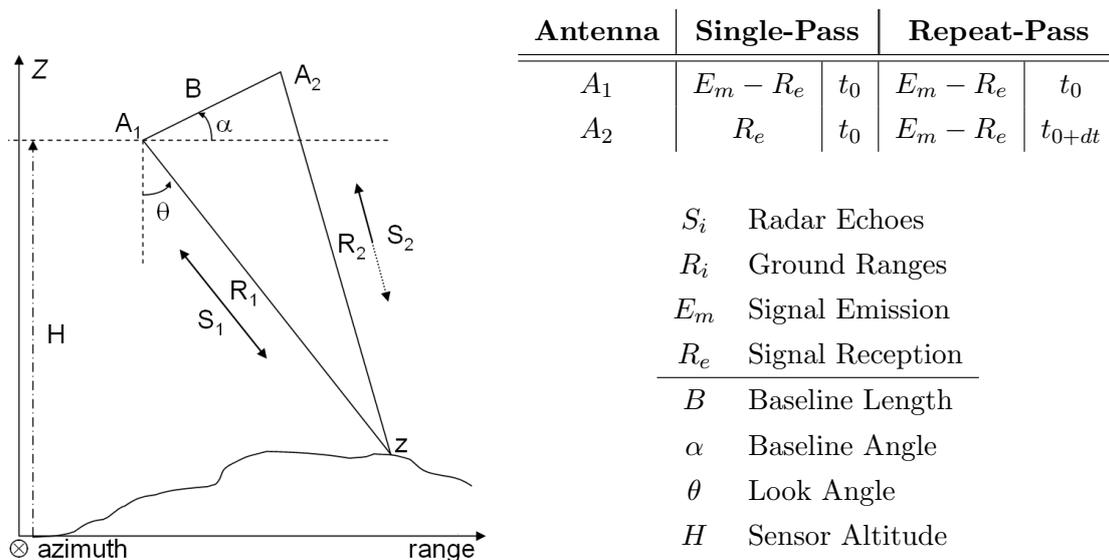
Figure 3.4: Interferometric SAR principle. Antenna's configurations (single-, repeat-pass) are given by E_m and R_e, being respectively the emission and reception of signals for a given antenna A_i.

is technically impossible to use this method for a space-borne operating in a sun-synchronous orbit (Height between 600 and 900 km). Alternative is to use a pair of identical satellites in constellation, like the upcoming TanDEM-X mission [122].

2 - "Repeat-pass" - A single satellite antenna is employed in a similar orbit, forming an interferometer baseline by relating radar signals on repeated passes over the same site. Loss of coherence induced by the time lap between the two acquisitions can strongly degrade the accuracy of reconstruction. Much research has dealt with interferometry using different sensors [197].

By measuring the exact phase difference $\Delta\Phi$ between two back-scattered signals observed from the same location in nearly similar spatial and temporal conditions (fairly modest spatial baseline i.e. $\leq$ 1km), the path length difference is accurately estimated (sub-wavelength scale), allowing to retrieve the elevation information.

Interferometric Process

The conjugate multiplication of two registered SAR complex images, called "interferogram", is composed of its normalised amplitude (used to derive the "coherence") and of its angle - the phase difference - which corresponds to elevation differences. The transformation from relative elevation reference to terrain elevation is critical and subject to numerous artifacts. Prior to the interferogram generation, the co-registration of the two complex SAR images is expected with sub-pixel accuracy and strongly determines the quality of the resulting InSAR process. Many approaches have been put forward to carefully re-sample the slave image to the master one.

Since the real phase values are "wrapped" around a 2π interval, the number of cycles N has

(a) Un-filtered DEM   (b) Modulation of (a)   (c) Modulation of (d)   (d) filtered DEM
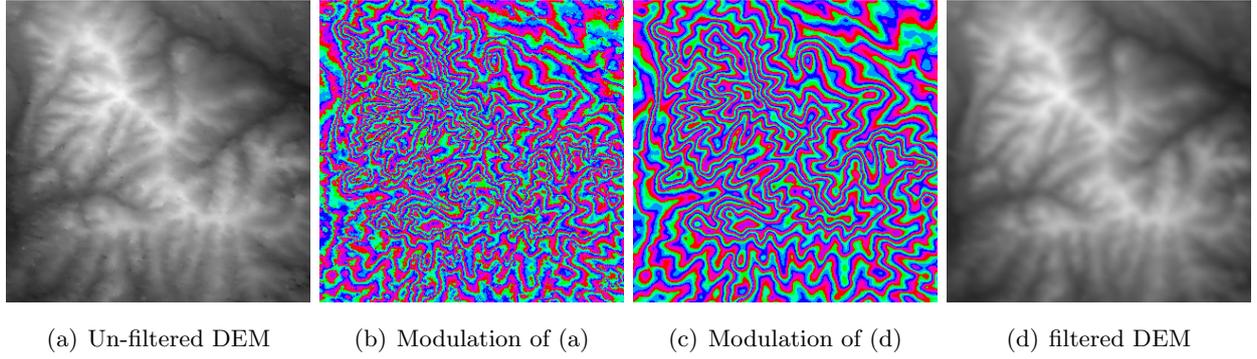
Figure 3.5: Phase Unwrapping: (a) raw DEM, (d) filtered DEM (size 320×320). Valleys are clearly recognized by their "fractal-like" behavior. (b) and (c) represent the "reversal" DEM modulation to simulate an interferogram and to emphasize the complexity of the PU process.

to be added to the principal values to calculate the height of each point of the image. This stage, called "phase unwrapping", is the most delicate step in the generation of elevation maps using SAR interferometry, especially with low coherence images (e.g. images corrupted by noise, sharp terrain transitions). The coherence, which is a measure of correlation (term defined between 0 and 1), is a reliable indicator of the interferogram quality. It provides an approximation of the signal-to-noise ratio (SNR) relative to the interferometric process and a control of the phase unwrapping step [3].

To retrieve absolute terrain elevation from the phase difference image and obtain satisfying results, specific operations applied to the interferogram are required prior to the "phase unwrapping" stage. They consist of interferogram band-filtering in range and azimuth directions [170], phase pre-filtering, multilooking and "flat earth correction" methods to decrease noise influence and remove acquisition bias. Once the interferogram is unwrapped (absolute phase difference $\hat{\Delta\Phi}$), the proportion (Eq. 3.1) between $\Delta\Phi$ and the travel path difference $2\Delta R$ is exploited for the elevation computation (Eq. 3.2) in taking into account only geometric differences in path lengths

$$\Delta\Phi = \frac{2\,\pi\,k}{\lambda}\,\Delta R \quad , \text{ with } k = 1 \text{ (single-pass)}, k = 2 \text{ (repeat-pass)} \qquad (3.1)$$

$$z = H - \frac{\left[\left(\frac{\lambda\Delta\Phi}{4\,\pi}\right)^2 - B^2\right]\,cos(\theta)}{2\,B\,sin(\alpha - \theta) - 2\,\left(\frac{\lambda\Delta\Phi}{4\,\pi}\right)} \qquad (3.2)$$

where θ denotes the look angle, α the baseline angle and H the SAR altitude.

The transformation of $\Delta\Phi$ into re-sampled elevation data referenced in a global or local system of coordinates (Geodesy) is achieved by geocoding techniques [181][4].

[3]Unfortunately, this measure depends on the window size used in its calculation [198], its computation should therefore be carefully achieved.

[4]Geocoding generally involves spatial topographic systems (GPS, Doris) to compute orbit positions and to refine the baseline estimation, meta-data (image geometry) and reliable ground control points if available and captured in images. Interpolation methods are required for conversion of the slant range to the ground range information.
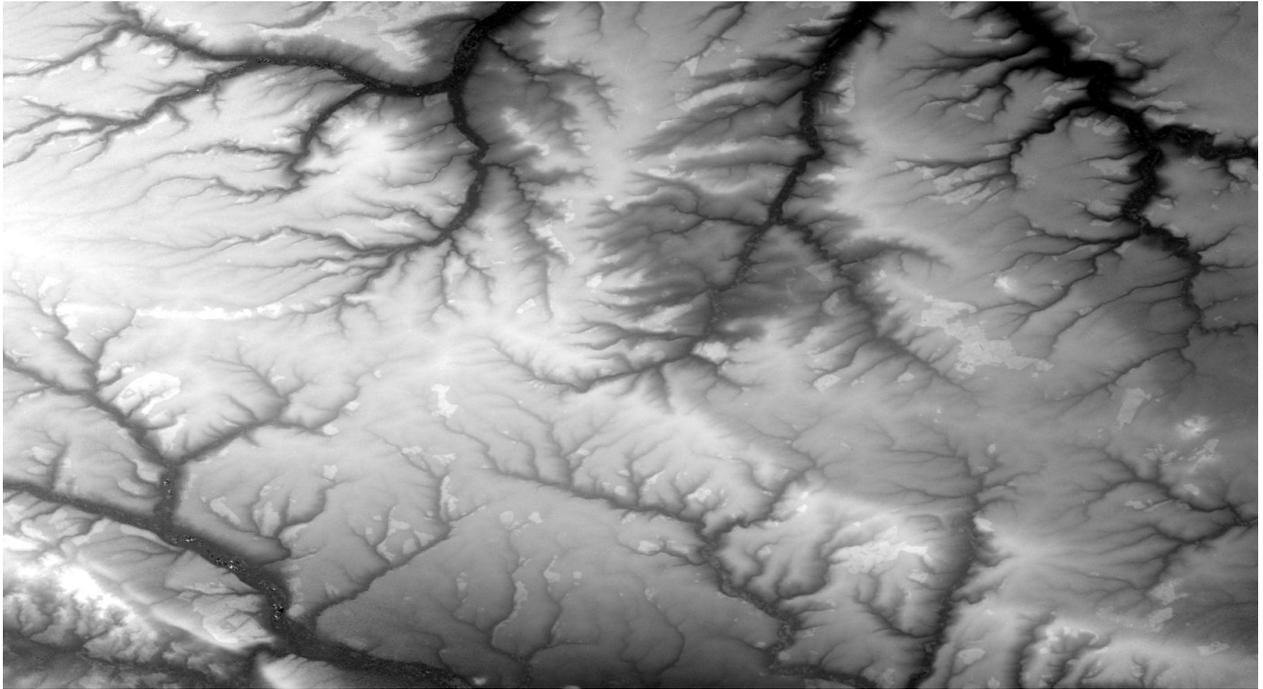
Figure 3.6: Example of InSAR SRTM DEM (X-band, single-pass configuration) [10]. The studied area is a smooth region near Amiens (France) about 50 kilometers-large with altitude variation ranking from ca. 80 to 245 meters. Data are geocoded on Ellipsoid/Datum WGS 84, with a 1/3600 degree resolution (ca. 20 meters at the Amiens's latitude).

Influence of Baseline Estimation and Decorrelation

Interferometric fringes are generated only from a pair of scattered coherent echoes reflected by the target sensed. According to Huygens, to be kept coherent, those echoes have to stay within a ray bounded by the critical baseline (ca. 1km for ERS). A compromise has to be reached for the baseline choice between coherent signals with less height sensitivity and noise level with better altimetric resolution. The baseline is used to interpret phase as elevation and requires to be known within the wavelength order, involving robust methods to estimate it accurately. The knowledge about the orbit position enables the baseline estimation. Any errors in the orbit estimation result in additional errors in each stage of the InSAR process.

The amount of decorrelation observed in repeat-pass interferometers is important for two reasons: (1) it constitutes an important source of errors during the InSAR process with a coherence loss. (2) This reduction is not always perceived as a drawback, since it allows applications like surface change monitoring on the wavelength scale (vegetative growth, earthquake, glacier motion, etc.). A coherence lower than 0.4 or 0.5 leads to problems with phase unwrapping algorithms. Under a critical level of the SNR ratio, phase measurements are unreachable unless using multi-look SAR techniques. The sensitivity of SAR and InSAR data (phase) to atmospheric conditions is reported in [197]. The co-registration stage is also determinant for the final quality of the product.

Exploitation of InSAR Elevation Data

The SRTM program was a successful mission where the NASA, the JPL and the DLR were involved. It consisted in pay-loading the space shuttle equipped with two SAR single-pass acquisition systems (American, C band and X band for the German). This ambitious technological challenge (mast of 60m, longest structure spread in the space) had allowed the generation of two world DEM coverages by interferometry (complete for the American DEM and 80% for the German). Due to the technological limitations, large and homogeneous extents are generated only by means of single-pass interferometric systems. For the first and only time, the SRTM mission had allowed to merge advantages of single pass measurements (coherence increased) with a global world and coherent coverage acquisition. Despite these improvements, non-stationary noises and artifacts remain and have to be filtered.

The dissertation mainly focuses on the German SRTM data; an example is provided in Fig. 3.6. Due to the smaller acquisition wavelength, this data-set represents an incomplete world DEM with a higher relative accuracy in comparison with the other generated SRTM-C DEM. The data-set consists of elevation and error maps (see Fig. 3.9, page 49). The latter is based on the accuracy of the coherence image, the baseline (distance between the two acquisitions) and the unwrapped phase. In June 2007, The German Space Agency (DLR) launched the satellite TerraSAR-X[5], a metric SAR sensor which operates in X-Band with full polarimetry capabilities [9] (an example is given in Fig. 1.1(a), page 2). With the launch of a second synchronised satellite, the upcoming TanDEM-X mission appears to be extremely promising and ambitious for the generation of a homogeneous digital surface model covering the globe with metric resolution.

3.3 Characterization of EO Data and Information Content

Modern systems are generally based on one or several push-broom camera(s). The imaging system is composed of *Charge Coupled Device* CCD arrays operating on a particular domain of the electromagnetic spectrum (see Fig. 2.2). On board of the SPOT 5 satellite, 3 arrays capture information in *red, green* and *near infra-red* channels. A further one, twice as accurate, operates in "panchromatic" mode (wavelength between $0.51 - 0.73 \ \mu m$) [6].

In addition, by means of two "High Resolution Stereo" panchromatic instruments ($\pm 20\,^{\circ}$ in forward and aft viewing angle of the satellite over a 120 km swath, 10 meter resolution). The resulting stereo-pair, acquired nearly simultaneously along- or across- track enables the generation of elevation data, see Fig. 3.7(d). Similarly, the "Ikonos" satellite operates in *red, green, near infra-red* and *blue* channels sampled to 3.2 meters. A panchromatic mode (spectral bandwidth $0.45 - 0.9 \ \mu m$) delivers 0.82 meter images.

---

[5]On 15th June 2007 , the German radar satellite TerraSAR-X was launched on a sun synchronous orbit at an altitude $H = 514km$ with an incidence angle varying from 15 to 60 degrees depending on the acquisition mode. The X-band sensor's center frequency is set to 9.65 GHz ($\lambda = 3.11cm$) with a bandwidth of 150MHz.

Image fusion is a field which deals with the generation of high resolution images in combining panchromatic and multispectral information. Geometry (high resolution, sharp images) is provided by the panchromatic image while features and information content come from multispectral channels acquired at a lower resolution. In the literature, many image fusion techniques have been investigated based on different image domains (wavelet, Fourier (high-pass filtering), multi-scale (Laplacian pyramid), IHS, PCA) or image analysis techniques (morphology, Markov) each of them having its own advantages and drawbacks. We refer to [50] for further explanations and performance analyses/comparisons. Purposes or interests are multiple; in this dissertation we have at disposal SPOT 5 "super-mode" data (2.5 meters) as well as pan-sharpened Ikonos images (1 meter), see example in Fig. 1.1(b).

3.3.1   Analysis by Resolution and Scale, from statistical to Object Modeling

The ability to provide large dynamics from the radiometric quantization is decisive for numerous feature extraction/discrimination processes (e.g. image classification). Apart from specific physical acquisition characteristics which strongly condition information content, the planimetric resolution acts as a determinant criterion. Several examples are given, corresponding to various resolution data-sets (Fig.1.1, Fig.3.7 to Fig.3.11). Some dependencies between resolution and the information content encapsulated in elevation data are summarized in Table 3.1.

The spatial planimetric resolution gives somehow an order related to the nature and the representation of the captured information. At a larger scale, both SAR and optical images are highly textured; such areas (e.g. field, forest with pure or mixed stands, etc.) are identified by classical statistical modeling which delivers reliable feature/pattern descriptions. Their discrimination is investigated in numerous publications using classification or texture schemes.

Local Adaptive Contrast Adjustment

By means of a locally adaptive contrast enhancement, information content, noise or errors encapsulated in elevation data are visually underlined. Locally (neighborhood system $\mathcal{N}$), the Wallis filter adjusts "brightness" values to fit local mean and standard deviation $(m_{\mathcal{N}}, \sigma_{\mathcal{N}})$ to user-specified target values (m, σ) based on a weighted average (α, σ_{max}) of the original image I.

$$I_W = \alpha\, m + (1 - \alpha)\, m_{\mathcal{N}} + \frac{(I - m_{\mathcal{N}})\, \sigma}{\sigma/\sigma_{max} + \sigma_{\mathcal{N}}} \qquad (3.3)$$

This filter performs a good enhancement throughout the image and reduces the overall contrast between bright and dark areas. Applied to elevation data, local features, alluvial drainage patterns as well as unwrapping and artifacts errors in shadowed areas are emphasized, Fig. 3.7(b)-(e).

---

[6]As already mentioned section 2.1.1 (page 9), SPOT 5 disposes of two "High Resolution Geometry" instruments with 5 meter resolution in panchromatic mode (multispectral information is sampled to 10 meters). Interleaved sampling combining both panchromatic acquisitions enables a resolution increase from 5 meters to 2.5 meters.

Data Resolution	Designation / Comments	Elevation Content	Applications	Optical / SAR Image Comments
> 25 m	*DTM* — Methods: Topographic InSAR/Stereo	Terrain model, drainage	Hydrology, primary reference for specific processes (geocoding, data reconstruction)	2D land cover, strong textures
~ 10 − 25 m	*DEM* — Except topographic data, hard to describe it	Large buildings, road networks, forests, bridges	Landscape analysis, flight simulations, GIS	Large structures are 3D, industrial areas, few textures
~ 1 − 10 m	Stereo: surface model InSAR: research	Objects, textures	Urban, landscape	Optical: 3D Objects SAR: research
< 1 m	*DSM* — Nice but occlusions: expensive campaigns, artifacts LIDAR data, airborne photogrammetry	Objects: cars, tiled roofs, etc.	Building extraction	Nice but complicated: 3D Objects, ridges-, edges-, lines- primitives, strong targets/textures, corners, scatterers

statistical description ← → computer vision
natural scenes — *man−made scenes*

Table 3.1: Information content in DEM data according to planimetric resolution (first column). The next columns report respectively the designation, the type of reconstruction, some features/content encapsulated in EO images, and some typical uses of the derived data. The content noticed for elevation data is integrated in order of "apparition".

Medium Resolution

Large scale elevation data originating from either topographic campaign, optical stereoscopy or SAR interferometry, represent mostly the terrain elevation. Vegetation or man-made structures, being too small or inaccurately measured, are almost ignored in the resulting data content. A typical example is the world-wide C-band SRTM DTM sampled to around 90 meters.

Medium resolution data (in remote sensing about 15-30 meters) appears less textured. Elevation data generated by means of EO images partially incorporate an umbrella to the terrain surface. Large buildings, road networks, bridges, dams are typical examples captured in DEM (see Fig. 3.7(a)- 3.7(d)). In SAR processing, according to the wavelength used, the information collected is partially composed of the vegetation cover (canopy) or in the opposite case, soil properties due the SAR penetration capabilities.

The corresponding DEM associated "reflects" such information with a geometrical representation. Fig. 3.7(a) exhibits the additional information carried by SAR images (SRTM-X DEM, $\sim$ 25m resolution). Road networks and bridges are encapsulated in the data. In addition, vegetated areas are very clearly emphasized in the data (as shown in Fig. 3.7(b) - 3.7(e), forestry areas and trees along the river borders are identified). Analyses of SRTM-X DEM accuracy reported in [86] a shift of 3 meters compared to an elevation reference with a standard deviation below 4 meters over non-vegetated or urban areas.

In Fig. 3.7(d), the factory (automotive industry) is encapsulated in the stereoscopic DEM derived from SPOT 5 data. The authors reported in [173] an absolute accuracy of 5 to 9 meters with standard deviations of 4 to 7 meters in comparison with Laser scanner DEM (reference) using only ancillary data. They investigated the fusion of SRTM and SPOT 5 elevation data to combine InSAR's absolute elevation accuracy (sea level calibration) with stereo's lower standard deviation, especially in mountainous areas. Similarly to [92], they noticed an improvement of the resulting InSAR-stereo fused DEM compared to the individual source.

High Resolution

The analysis of high resolution data, characterized by spatial resolution ranking from 1 to 10 meters, - see Table 3.1 -, is limited by the difficulties to properly describe and model the information content. The latter is partially composed of features and patterns typically related to decametric data-sets as well as few objects and/or primitives. These weaknesses are well illustrated by naming confusions between techniques applied on different resolution data-sets.

A 10 meters stereo DSM, computed from Cartosat data [44], [130] is shown in Fig. 3.7(f) over the same area as the SPOT 5 stereo DEM (Fig. 3.7(d)). Indeed, the chosen test site is particularly interesting since it presents a huge variety of terrain configurations (urban, landscape, smooth/hilly/mountainous areas). In addition, both data-sets were generated with the same image matching algorithm described in [131].
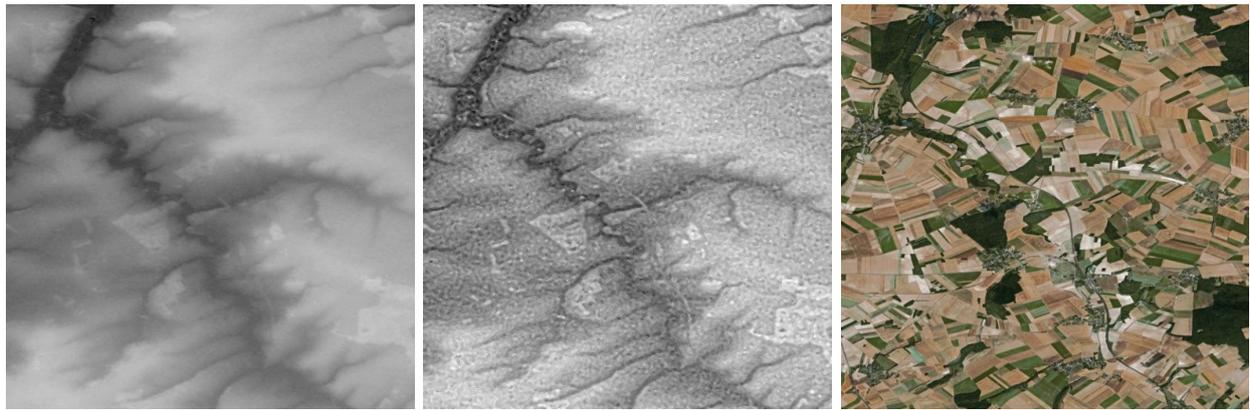
(a) InSAR SRTM DEM   (b) Local contrast applied on (a)   (c) Ortho-image
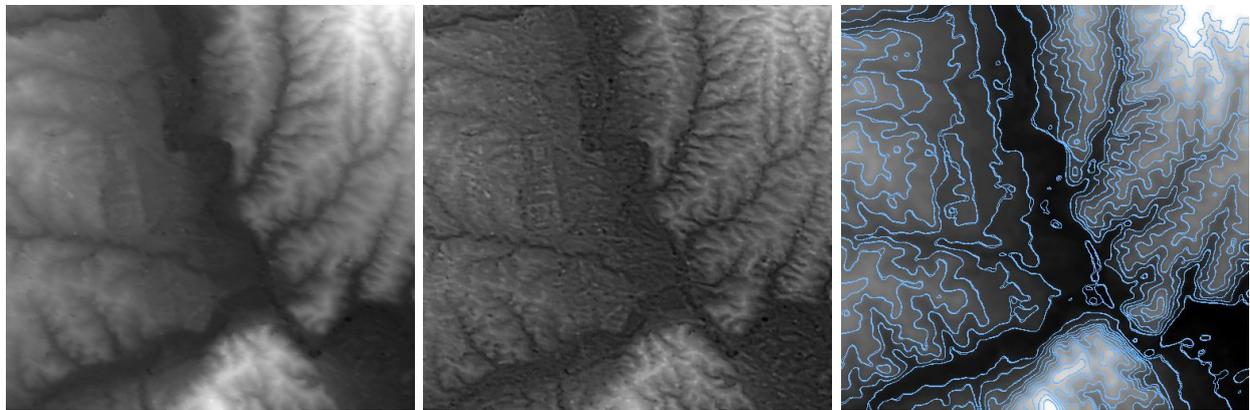
(d) Optical stereo DEM   (e) Local contrast applied on (b)   (f) Isolines
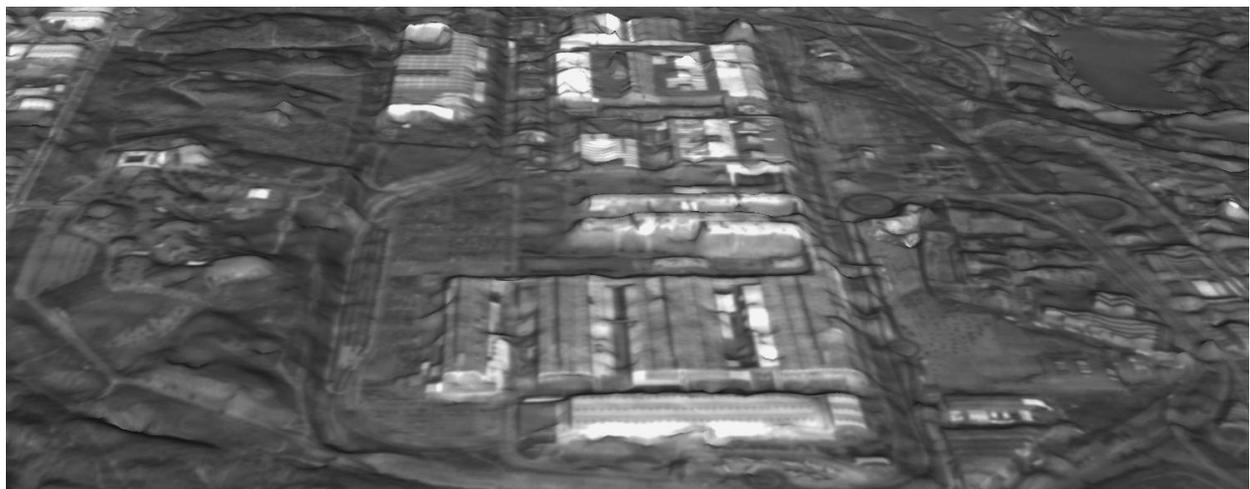
(g) 3D perspective using high resolution stereo DSM (10 meters) computed from Cartosat data, same area as (d)

Figure 3.7: Information content on decametric InSAR (a) and stereo (d) (g) DEM. (a) SRTM-X DEM (384×384 tile extracted from Fig. 3.6). (d) SPOT 5 DEM (from Fig. 3.3, page 37, size 512×512). (b), (d) Local contrast adjustment (7×7 windows size). (f) contour lines (20 meters distant). Valleys, main road network/bridges, forest areas are identified in SRTM data while large buildings are encapsulated in SPOT 5 DEM. Compared with (d), the factory (g) is more properly captured. Road networks are emphasized, smaller buildings still appear as blobs.

Significant improvements are noticed for the identification of man-made features (factory, building, road network, etc.), terrain variations or vegetated areas. SPOT 5 DEM is smoother but as expected from its 15 meter resolution, results in urban areas are outperformed by the Cartosat DSM. However, the latter suffers from a partial description of the buildings in terms of accuracy and precision, appearing as blobs which mostly indicate their presence and an "idea" about their shape. The 3D rendering scene - Fig. 3.7(f) -, composed with Cartosat data (DSM (geometry) overlaid by a panchromatic image (texture)), emphasizes how much information has been captured and reconstructed. Further examples are given with this data-set in chapter 7.8, page 133.

Very High Resolution (VHR)

The generation of VHR (<1 meter) elevation EO data is made possible by means of airborne campaigns. Photogrammetry is a well-established field that reached an operational stage using InSAR, stereoscopy or LIDAR techniques. Due to the spatial resolution increase of the computed DSM, nature and representation of the information content drastically differ from traditional mapping based on lower resolution. As a consequence, this remark concerns also the analysis techniques involved to enhance the surface model or to extract relevant information.

Fig. 3.8 depicts airborne optical data acquired and processed by the French National Geographic Institute (IGN) over the city of Amiens in the north of France. This data-set is made up of one optical image and two DSM computed by using different techniques (LIDAR sampled to 50cm and 20cm for the stereoscopic DSM). The interest for very high resolution (VHR) data is not only restricted to cartography, urban planning or military applications. In [143], InSAR data (25cm) are investigated for the purpose of forestry monitoring and disaster management. However, expensive costs related to the airborne acquisition mostly focus the applications/interests on man-made areas such as cities where the lack or the update of cartographic databases (GIS) is required. In the literature, buildings or road networks extraction are perhaps the most studied cases. For those reasons, a state of the art related to VHR urban data-set is presented in section 8.1.

Indeed, even considering that VHR SAR/Optical images are highly textured - see Fig. 1.1(a) and 3.11(b) -, classical statistical models fail to represent the elements/objects which populate the data, and particularly elevation data. In a similar way as several computer vision systems, the information is mainly encapsulated in shape, sharp transitions, primitives, in form of lines, corners, ridges, edges elements which compose objects or elements.

Suitable techniques and methodologies based on computer vision are required to properly extract and model non-linear objects or structures. Another remark concerns 3D rendering applications with VHR data-sets (remotely sensed nearly in nadir acquisitions). Shadows are not neglected anymore and increase difficulties for extracting information (see Fig. 3.11(b)). For a higher level of realism, the geometry provided by VHR LIDAR data is not sufficent. 3D models are required with facets of the building facades, the use of simulated textures being somehow limited to compensate architectural descriptions.

(a) Optical image   (b) LIDAR DSM   (c) STEREO DSM

(d) Optical image   (e) LIDAR DSM   (f) STEREO DSM
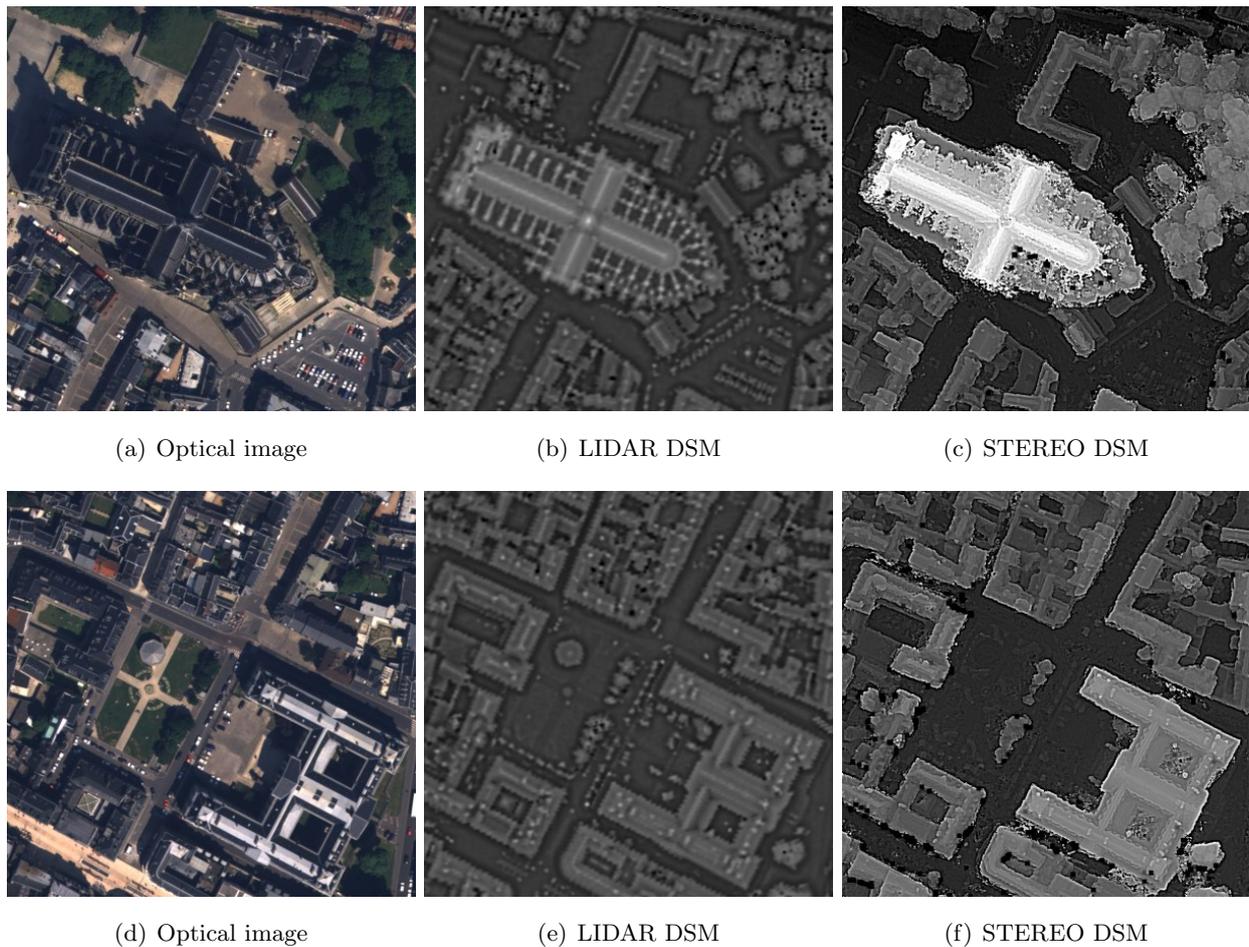
Figure 3.8: Information content on VHR LIDAR and stereo DSM over a $(200 \times 200 m^2$ area). (a)(d) Optical image (10cm resolution). (b)(e), (c)(f) are local contrast adjustments applied respectively on LIDAR (50cm resolution) and stereo (20cm resolution) DSM.

3.3.2 Analysis by Noise and Artifacts, physical Principles

3.3.2.1 InSAR DEM

SAR content exploitation is complicated due to the speckle phenomenon which can be seen as a multiplicative noise affecting all coherent imaging systems (laser, acoustics, radar images). This is the consequence of random interferences between electromagnetic waves from different element targets. In SAR images, this effect is characterized by a granular noise, only visible in the detected amplitude or intensity signal (see Fig. 1.1). Based on these remarks further developed in annex A.1, several problems are expected and corrupt DEM generated by SAR Interferometry.

Thermal Noise, Signal Properties: Numerous problems encountered during the phase unwrapping stage are due to acquisition characteristics of the pair coherent signals (suitable baseline, uncertainties in its estimation, etc.) as well as microwave signal properties (L-, X-, C-bands). In SAR imagery, the SNR ratio is affected by a thermal noise added by the receiver and ground ar-

(a) InSAR SRTM-X DEM   (b) High Error Map   (c) Ortho-image
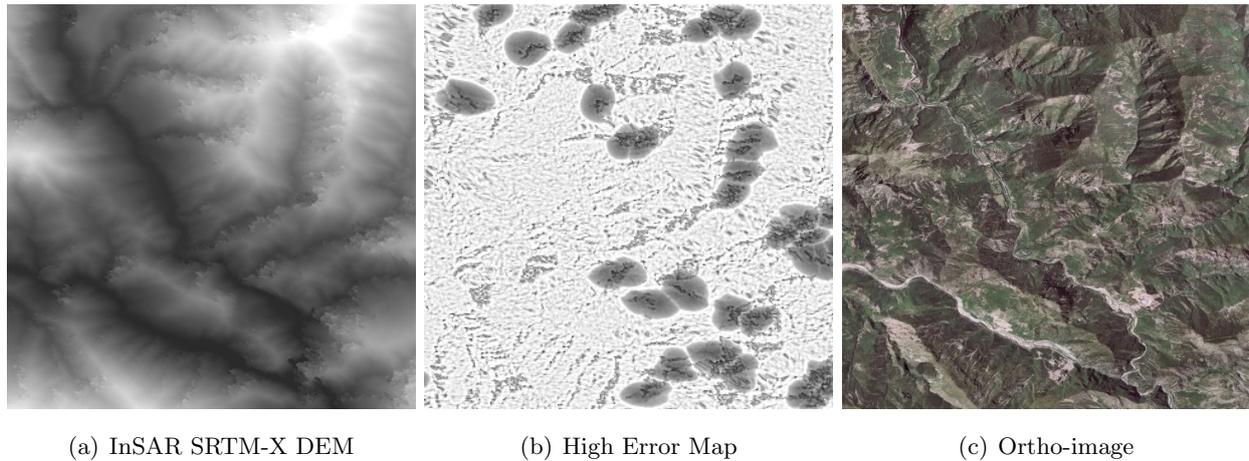
Figure 3.9: SRTM DEM over mountainous areas (occlusions). (a) SRTM-X DEM (size 320×320, altitude range about 240 to 1530 meters). (b) Error Map Quality, dark areas present low coherence while bright ones depict reliable elevation. (c) aerial optical image (5 meter resolution, IGN).

eas with low reflectance. Such a noise, which impacts both SAR and InSAR data, is observed in Fig. 3.7(a) by its granular visual effect.

Image Geometry and Relief Influence: Geometrical effects presented in Fig. A.3 such as layover lead to discontinuities in the phase variations. Indeed, relief pervades all of SAR information (geometric, amplitude and phase). As a consequence, due to the incidence angles, occlusions lead to several characteristical errors appearing in mountainous areas. As shown in Fig. 3.9, such areas present a lack of information (i.e. low SNR) affecting the elevation determination. Inversions and displacements can occur depending on the acquisition conditions and relief (layover, elongation, foreshortening). In those cases, elevation determination is intricate and the corresponding area is often corrupted by strong artifacts, characterized by an excessive texture's roughness (too high fractal dimension and local variance).

In Fig. 3.9(a), an example of InSAR DEM in mountainous areas is given. The French Riviera is located along the Mediterranean sea. The Alps just North of Nice are 4000 meters high above sea level. This test site was chosen because it represents almost the worst case encountered during the elaboration of the SRTM-X DEM data-set (high relief and geode variations, important incidence angle, along a stripe border, only one single-pass acquisition, etc). In [57], Eineder addresses more in detail the issue of generating InSAR DEM over mountainous areas. An interferogram covering a flat area presents a linear trend in the phases in the range direction, and the fringe rate depends only on the baseline separation between the satellite positions. To get rid of this effect, a reference interferogram is computed to eliminate the earth curvature influence. Any inaccuracy in this step would lead to additional residual slope in the final DEM.

Temporal Decorrelation and Surface Characteristics: The nature itself of the target affects the coherence analysis and the sensitivity to temporal decorrelation with variable time scales.

<div align="center">

(a) InSAR SRTM-X DEM   (b) Optical Image   (c) 3D Perspective

</div>
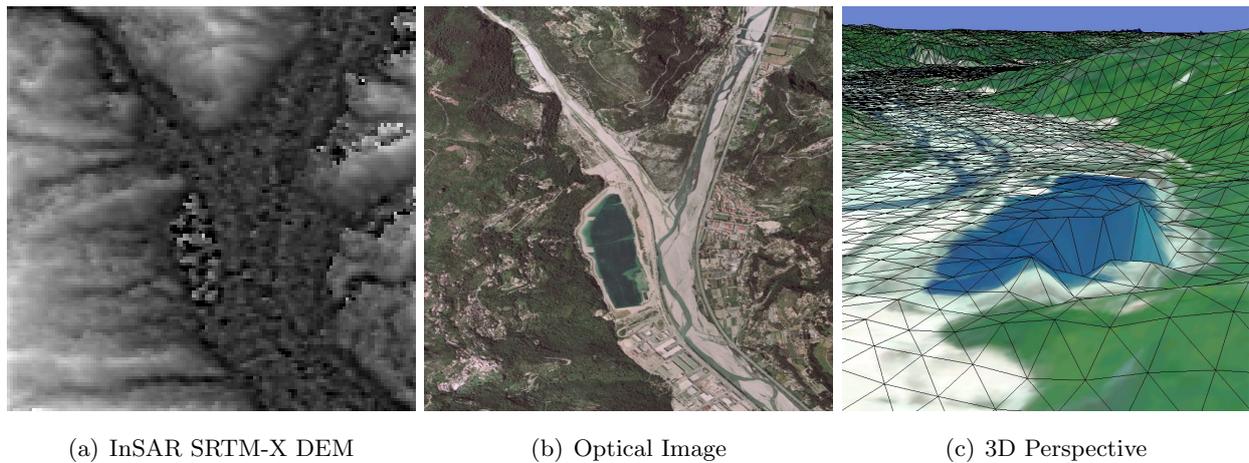
Figure 3.10: Specular reflection on flat surfaces. (a) InSAR SRTM-X DEM (size 128 × 128), with Wallis filter. (b) optical ortho-image, (c) 3D perspective using a SPOT 5 image.

The temporal decorrelation in unstable areas -expressed by a low coherence- directly influences the resulting interferogram. Unreliable results are expected from active agricultural regions or huge vegetated areas (see Fig. A.2). Penetration capabilities in soil or canopy of the radar signal (according to its wavelength) affect the computed height by integrating an average measure of the density of penetration. Multiple bouncing (see Fig. A.2), often occurring in urban areas, leads to erroneous elevation estimation, characterized by peaks and artifacts. In the same way as in the shadowed areas, specular reflection on flat surfaces (low SNR) results in severe artifacts in the unwrapping phase (example of lake, rivers in Fig. 3.10).

3.3.2.2   Optical Stereoscopy

Thermal noise as well as alteration or erroneous calibration of optical sensors can corrupt the collected information. Arrays are placed perpendicular to the satellite track. It results in acquisition un-synchronization which is neglected in most of the cases, except when sensed objects are moving as shown in Fig. 3.11(a) with the airplane example. Such a behavior is not always a drawback, target monitoring as well as velocity estimation applications are investigated (mainly traffic monitoring, see [88]). Further examples of Cartosat DSM data are given in section 7.2.2, pages 127- 134. In particular, Fig. 7.7 summarizes information content encapsulated in the DSM.

Image Matching: Quality and accuracy of elevation data computed from stereoscopy is directly dependent on the success of the image matching stage. The use of large incidence angles, which leads to higher parallax accuracy, is limited for several reasons; occlusions in urban or mountainous areas (see in Fig. 3.8(c), 3.8(f)), matching performances (in terms of detection and computational time, data content modeling). Indeed, reliable and robust algorithms optimized for "landscape" imaging [131] with statistical modeling partially fail in capturing man-made features or patterns, characterized by sets of primitives/objects having sharp and non-linear transitions.

(a) Ikonos Image (Germany)   (b) Aerial Image (Spain)

Figure 3.11: (a) Moving target in optical images acquired with push-broom techniques. (b) Textured/marked roofs on VHR optical image, important shadows of factory, trees or even telegraph pole. Trees, cars, railway network are clearly identified.

Image matching in pairwise-images is limited for flat or/and homogeneous areas such as meadows, fields. The absence of singular features or patterns prevents solving the correspondence problem and results in a lack of information, appearing as "holes" within the generated DEM/DSM. To fill missing information, weighted-distance interpolation techniques based on complementary information (e.g. world wide SRTM-C DEM) are generally employed. In chapter 8, a method based on multi-scale B-Splines is proposed as an alternative and tested over HR and VHR DSM.

Category of Errors: Blunders occurred during elevation computation are mostly tracked and removed during or right after the image matching stage. Based on the best extracted tie points, approximation of the affine transformations between the image pair are computed and further used to detect, by comparison with the overall matching, points and blunders which exceed some arbitrary threshold variance. Such an approach may not be robust enough with VHR data presenting important objects densities. Artifacts or numerical errors generated during the interpolation stage are locally characterized patterns. In opposition, systematic errors are spread over much larger areas and usually follow linear or constant trends. Random errors and thermal noise pervade data and make the extraction of the elevation more complicated.

Summary: Elevation Data Enhancement

All these aspects (model formation, acquisition, thematic and relief) encountered during the phase unwrapping/image matching stages severely affect the quality and accuracy of the resulting DEM. Chapters 7 and 8 are mainly devoted to the enhancement of DEM data, with respectively filtering techniques and integration of complementary information.

Part II

Image Analysis

Image Segmentation

Abstract

Image segmentation - one of the earliest stage of a computer vision system - is a determinant process conditioning the performance of the overall processing chain. This chapter addresses a partial state of the art in image segmentation techniques and requires several concepts already presented in chapter 2 as well as annex B. Objective is to partition an image into disjoint and homogeneous regions, presenting sharp and accurate boundaries.

This field is still an open issue and perhaps the one which gathers the largest amount of publications in image analysis and computer vision. Among this important literature, we mainly focus on clustering, region-based and variational approaches. More particularly, the formulation of image segmentation as an optimization problem based on the *Mumford-Shah* functional appears well adapted for our purpose. It should be mentioned that a comparison between them is made difficult since they are related to different "applications" and the notion itself of segmentation is still not clearly defined.
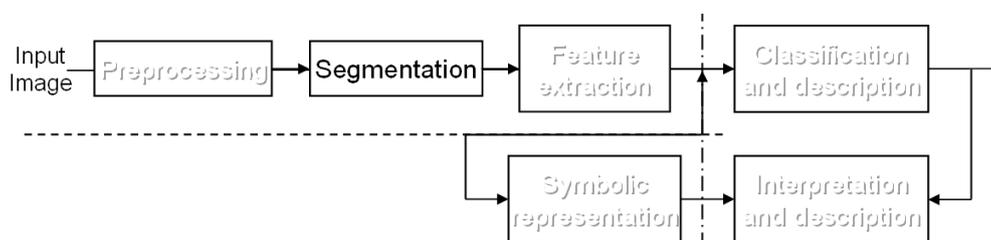
Introduction

In several image analysis and computer vision applications, image segmentation is a critical operation since it represents the first step of low-level processing. Segmentation processes aim to subdivide an image I into disjoint and homogeneous entities, called Regions R_i, by defining boundaries[1] between features and objects using some Boolean predicates $\mathcal{P}()$. $\mathcal{P}()$ is related to a particular homogeneity measure which can be computed based on image intensity, color, texture, local statistical features, etc.

---

[1]See section 6.1, page 87.

A Variety of Image Segmentation Techniques

It has been proved that extracting boundaries between regions is an equivalent process. This duality leads to represent image segmentation as a compromise between those two objectives. In this scope, Haralick [83] pointed out that instead of a general theory on image segmentation, it exists several *ad hoc* techniques which are suited for dedicated and specific applications or analyses.

To segment images, Meyer and Beucher presented in [151], [150] the so-called *watershed* algorithm, based on morphological operations (thinning). Some methods [37], [72] are for instance directly related to a given image formation or representation model. Since knowledge about the image considered is determinant for image segmentation problems, Pavlidis [166] excluded the perspective of a purely analytical solution. For those reasons, this fundamental computer vision problem has generated much research on a wide range of domains based on various image types and purposes.

The aim of the present chapter is not to give an exhaustive review of image segmentation techniques but to present some of them in order to focus our interest on particular methods. Many papers [219], [69], [83], [159], [191] are addressed to collect, classify and synthesize investigated research in image segmentation. According to [83], [159], five main approaches generally emerge.

Organization

The chapter is organized as follows: clustering is introduced and used as foundation to describe **hybrid** and **spatial clustering** methods encountered in the purpose of image segmentation. Then, classical **region-based** approaches are presented as the complement to **boundary-based** methods (already discussed in section 2.2.3, page 24).

Working with regions has advantages: image partitioning into sub-regions leads to closed boundaries, while edge detectors based on threshold or local filtering such as the Sobel or Canny operators [24], [48] often result in discontinuous boundaries.

Finally, some of the most popular **global optimization** methods such as the Minimum Description Length (MDL) criteria and variational approaches are introduced. A particular attention is paid to the Mumford-Shah functional; its application to active contours and region-based approaches is studied.

4.1 Segmentation Based on Clustering

Histogram and thresholding techniques [98], closely related to image clustering are not presented. An exhaustive review for image color segmentation is given in [191]. Examples of the clustering approaches described below are provided page 115.

4.1.1 Image Clustering

Clustering is a special kind of classification [107] (see page 26). It aims at dividing a large set of data into smaller non-overlapping groups of data (clusters) characterized by similar properties. The division of the data-set into clusters is called a partitioning. A cluster is modeled by its center of mass (centroid) and an attribute to define the membership relation of the data elements.

The similarity or proximity between a data element p and a cluster can be explicitly quantified by a distance measure between p and the cluster centroid (some of them are reported in Table 2.1, page 18). Regarding the set of clusters, p is assigned to the cluster which provides the smallest distance. Given a priori a fixed number of clusters (let's assume k clusters), several methods have been introduced for their estimation. As mentioned by Jain in [99], clustering algorithms are subdivided in two main families: hierarchical and partitional clustering.

Hierarchical Clustering

The simplest automated method for partitioning the data-set in k clusters is to incrementally agglomerate data elements or clusters by pairs of clusters until k clusters remain. The main drawback of Ward's method [210] is its slowness, since all combinations have to be computed before any merging. Various approaches have been reported to improve result consistency and computational time. Among them, the single link algorithm is one of the most popular methods [94], [99].

The alternative to agglomerative approaches is divisive methods. It consists in starting with a single cluster that contains all data elements. New clusters are created by dividing the existing ones until having the desired number of clusters. Two crucial questions arise when using this top-down approach: which cluster to be divided and how to perform the division? The *Median Cut* algorithm proposed by Heckbert in [85] achieves the division along a selected dimension of the feature space. In [213], the principal component analysis is applied to select the division.

Inspired by the works of Wong [212] in clustering by melting, Ciucu proposed in [34] an accurate and efficient algorithm by considering the feature space as a thermodynamical ensemble. Having the temperature as scale parameter, data are grouped by minimizing the free energy. To improve data compactness, a tree structure is adopted to model the data and the clustering evolution by encapsulating each temperature to a corresponding level of the tree. The melting procedure is initialized using a partitional algorithm to reduce the computational complexity.

Partitional Clustering

Partitional strategies are mainly based on square-error criteria; i.e. given a defined number of clusters, find the partition which minimizes the square-error or equivalently maximizes the variations between clusters. Even with small data-sets to be partitioned in few clusters, the number of

possible combinations prevents feasible combinatorial implementations. The limitation can be get round by means of iterative methods. Most of them are based on the flowchart suggested by Forgy in [63] to classify a data-set into a certain number k of clusters:

1. **Initialization:** Define k clusters using a random selection based either on:
   - randomly chosen patterns or points inside the bounded data-set,
   - histogram analysis.
2. **Assignation:** Assign each pattern to the nearest cluster centroid
3. **Update:** Recompute the cluster centers using the current cluster membership
4. **Convergence:** Repeat steps 2-3 until the convergence criterion is reached

Algorithm 1: Flowchart of a K-Means algorithm.

The convergence criterion may be related to the minimal decrease of the squared error, or as most frequently reported in the literature, to the reassignment's evolution of patterns to new clusters. The widely used *K-means* algorithm introduced by Mac Queen in [137] is a variant of the Forgy's method. Some other implementations recompute the centroid after each new assignment, which is extremely convenient when data are continuously ingested [148]. Although there is no guarantee that such iterative algorithms reach a global minimum, the presented methods always converge [183] and therefore provide robust computations. Regarding the clustering performance, K-means algorithms are highly conditioned by two key points: the fixed number of clusters and the *Initialization* step.

The two limitations are partially circumvented by the *Isodata* algorithm introduced by Hall and Ball [8]. It follows the same procedure while providing more flexibility since it enables to create new clusters or to merge existing ones. To control both splitting and merging operations during the *Update* step, the user assigns conditions by means of a set of thresholds. The defined rules can be: (1) a cluster is split if it has too many patterns or an excessive variance along at least one feature space axis, (2) two clusters are merged if their centroids are too close from each other.

Regarding data consistency and compactness, several authors have paid attention to outliers which may result from erroneous data acquisition (e.g. artifact or noise phenomenon). To avoid possible clustering distortion or the generation of small meaningless clusters, outliers must be identified and removed [103], [168]. Considerable research has been investigated to improve the *Initialization* stage [7], [20], [104], [110], [172] and to propose efficient implementation schemes.

4.1.2 Clustering and Bayesian Segmentation Approaches

In [83], Haralick pointed out similarities between image segmentation and classification techniques. While the grouping is done in the feature space for clustering, image segmentation operates in the spatial domain. The likeness between techniques stimulated the emergence of many hybrid methods to combine the strength of one given method with the requirements of the other.

Extended versions of the *K-Means* algorithm are addressed by Pappas in [162], [161] and Marroquin in [148]. Both include spatial constraints to take into account local intensity variations and remove small outliers. Spatial interactions are encapsulated by means of Gibbs random field (GRF) models. Pappas used an 8-neighbor GRF and proposed an iterative procedure based on a hierarchical implementation for preserving most significant features while removing irrelevant details. Marroquin states the problem in the Bayesian formalism using as prior discrete Markov random fields (MRF) expressed on larger neighborhood systems than Pappas's algorithm.

To improve the algorithm's efficiency and quality, he also suggested the idea of multi-scale approaches based on pyramidal processes. Later on, Bouman and Shapiro introduced in [19] a consistent Bayesian multi-scale formalism by replacing respectively the MRF model and the Maximum *a posteriori* (MAP) estimator with a multi-scale random field and sequential MAP estimator. The method is illustrated with results derived from both synthetic and multi-spectral EO images and further detailed page 121. Lavalle applied in [125] a Bayesian segmentation methodology for parametric image models such as Gauss Markov random fields (GMRF) and polynomial surfaces.

4.1.3 Mean Shift

To segment color images, Comaniciu further exploited in [37], [38] the *Mean Shift* procedure initially formulated in [70]. Contrary to *K-Means* clustering, the *Mean Shift* is a non-parametric clustering mode; i.e. neither an assumption on the shape distribution nor the number of clusters is required. The image to be segmented is considered as an empirical probability density function (pdf) in the feature space, where data are seen as a set of d-dimensional points.

Intuitively, local maxima or modes of the underlying pdf distribution indicate dense regions in the feature space. For each data point, gradient ascent procedures on the local estimated density are iterated until convergence. Stationary points emphasize distribution modes and, by extension, clusters populated by the data points that converged to the stationary points.

Given n data points x_i in d-dimensional points, using radial symmetric kernels $K(x) = c_k k(||x||^2)$, the multivariate kernel density is estimated

$$\hat{f}_K = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{4.1}$$

where h denotes the kernel radius called *bandwidth* and c_k acts as normalization term. Deriving the gradient of the density estimator given in Eq. 4.1 leads to

$$\bigtriangledown \hat{f}(x) = \left[\frac{2c_{k,d}}{nh^{d+2}} \cdot \sum_{i=1}^{n} g\left(\left\|\frac{x - x_i}{h}\right\|^2\right)\right] \left[\frac{\sum_{i=1}^{n} x_i g\left(\left\|\frac{x - x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{x - x_i}{h}\right\|^2\right)} - x\right] \tag{4.2}$$

where $g(x) = -k'(x)$ is the selected kernel profile derivative. In Eq. 4.2, the left term reflects the density estimate centered around x. The right term is called the *mean shift* vector m. Hence, for a

given point x, m (computed using Eq. 4.2) is oriented to the maximal increase in density direction and with an amplitude proportional to the density gradient estimate using the kernel $K(x)$. For a given point x_i, an iterative procedure $(t \rightarrow t+1)$ is designed;

1. **Estimate:**   - Compute the mean shift vector $m(x_i^t)$

2. **Update:**   - Translate the density estimation window $x_i^{t+1} = x_i^t + m(x_i^t)$

3. **Convergence:**   - Repeat steps 1-2 until convergence criterion $\bigtriangledown \hat{f}(x_i) = 0$ is reached

Algorithm 2: Flowchart of the Mean Shift algorithm.

A proof of this convergence is given in [38]. Identifying neighbor points at a current position in the feature space with bandwidth h appears to be the most time consuming operation. In addition, the selection of a suitable bandwidth is rather difficult; Comaniciu addressed a possible locally adaptive bandwidth estimation in [36]. The method also covers adaptive smoothing or tracking applications.

4.2 Region-based Approaches

4.2.1 Definitions

Although there is no clear definition, Haralick and Shapiro pointed out in [83] what is expected from a good segmentation: *"Regions of an image segmentation should be uniform and homogeneous with respect to some characteristics such as gray tone or texture. Region interiors should be simple and without many small holes. Adjacent regions of a segmentation should have significantly different values with respect to the characteristic on which they are uniform. Boundaries of each segment should be simple, not ragged, and must be spatially accurate."*

A more formal definition is given by Zucker in [219]: If I is the set of all pixels and $\mathcal{P}()$ a uniformity predicate defined on groups of connected pixels, a segmentation is a partitioning of I into a set of connected subsets or regions $(R_1, R_2, , R_n)$ abided by the relations

$$
\begin{aligned}
S &= \{R_1, ... R_n\} && \text{with } R_i \neq \{0\} \ \ \forall i \in \{1...n\} \\
\bigcup_{i=1}^{n} R_i &= I && \text{with } R_i \cap R_j = \emptyset \ \ i \neq j \\
\text{and} \\
\mathcal{P}(R_i) &= true && \forall i \in \{1...n\} \\
\mathcal{P}(R_i \cup R_j) &= false && \text{if } R_i \text{ and } R_j \text{ are adjacent}
\end{aligned}
\tag{4.3}
$$

In relation 4.3, the two first propositions are related to the image partition into non overlapped connected regions. The two last propositions describe the predicate which controls conjointly region homogeneity (third proposition) and dissimilarity with neighbor regions.

4.2.2 Measure of Homogeneity

The Boolean predicate is generally formulated by the imposition of a threshold t to be compared to a particular homogeneity measure h

$$\forall i \in \{1...n\}, \ \ t \in \mathbb{R}, \ \ \ \mathcal{P}_h^t(R_i) \ \text{is true} \ \ \ \Leftrightarrow \ \ \ h(R_i) \le t$$

Using the notations previously introduced, some of the predicates found in the literature are summarized in Table 4.1.

Type of Predicate	Definition $h(R_i) =$
Global Contrast	$max_{R_i}(I) - min_{R_i}(I)$
Local Contrast	$\|I(i) - I(j)\|, \ \ \forall(i,j)_{R_i} \ \text{neighbors}$
Standard Deviation	$\sqrt{\dfrac{1}{N}\sum_{R_i}(I-m)^2}, \ \text{with} \ N = Card(R_i) \ \text{and} \ m = \dfrac{1}{N}\sum_{R_i} I$
Entropy	$-\sum_{R_i} p(f)\log p(f)$
Histogram	Inter-quartile range between first and third quartile

Table 4.1: Some classical predicates.

The mathematical definition 4.3 guarantees neither a unique solution nor a unique method. Hence, for a given image, several partitions are found using the same predicate. Combinatorial computations prevent from generating all the partitions to validate the predicate adopted. Based on these remarks, the quality of a segmentation is extremely difficult to establish; among all the possible partitions, which one does provide the best results? The primary answer was the use of heuristic approaches, focusing on different optimization schemes according to the application, such as minimizing the number of regions, the region size or heterogeneity between neighbor regions.

4.2.3 Region-Based Segmentation Evolution

During the segmentation process, a region R_i may evolve. In this case, it follows one of the two elementary operations:

Operation	Definition	Level of Details
region merging	union of R_i with one of its adjacent regions	-
region splitting	R_i is divided in two or more regions	+

As mentioned by Pavlidis in [166], the strategy to build a partition is determinant and leads to many approaches, based on one operator only or combinations of operators with different region initialization schemes (partial or on the whole image).

Trivial Partition Initialization

The **Split** method is a *top-down* approach. It begins with its corresponding trivial partition, i.e. the entire image is one region. Iteratively, the regions which do not meet the homogeneity criterion are split. Unfortunately, segmentations based exclusively on region splitting generally provide poor results, mainly due to the decomposition into regular quadrants. **Merge** (*bottom-up*) approaches are initialized by considering that every pixel is a region. Adjacent regions with similar properties are merged together. The similarity between the regions can be seen as a merging cost; it remains for the algorithm to sort properly the pair of connected regions to be merged.

Split and Merge techniques use a sequential combination of both methods. Such a strategy is generally applied on a pyramidal structure called *quad-tree* [93]. Many authors [31], [91], [193] further improved such segmentation schemes to combine statistical and spatial information. Similarly to the regular quadrants, methods based on *quad-tree* suffer from the too rigid data structure. In [19], Bouman and Shapiro used a multiresolution stochastic image model in the development of a Bayesian segmentation algorithm. More generally, Split and Merge techniques are heuristic methods; their efficiency largely depends on the combination of the operators.

Seeds Initialization

Contrary to the techniques previously mentioned where the initial state matches the image segmentation definition expressed as an image partition into disjoint regions, **region growing** methods follow a different work-flow; seeds are chosen along the image and expressed as regions. Based on a predicate $\mathcal{P}()$, the *top-down* process evolves inductively from the seeds to merge unlabeled neighboring pixels. The progressive pixel aggregation enables to fit properly edges and ridge between regions.

Results are highly conditioned by the choice of the seeds. As reported in the literature, they usually correspond to similar groups of neighboring pixels. Neighbor regions are often handled with stronger constraints on the predicate or even not merged, preferring over-segmented results to heterogeneous regions. Depending on the strength of the predicate associated, some methods [1], [14] do not "produce" enough seeds to label all the pixels and result is an incomplete image partition. This case is justified for several applications such as object tracking or face recognition involving to ignore irrelevant background areas.

Even if the principle is simple and intuitive, the almost infinite options involved in the process ([219], [167]) as well as the various applications and their computation costs make the implementation of an efficient algorithm extremely difficult. Some of the numerous approaches published which emerge are referred in [83].

4.3 Variational Methods

As presented in the previous sections, image segmentation is an "ill-posed" problem which can explain the variety of methods envisaged or their high heuristic level by combining different regions and edge-based approaches [167]. This lack of formalism has been compensated by variational methods which emerged once the classical approaches limitations had been reached. In the continuation of the celebrated work of Geman and Geman [73], which operates on lattice domain, variational methods formulate image segmentation as an estimation problem defined on the plane domain using functions to represent images.

The integration of a rigorous mathematical formalism using continuous and well studied models constitutes a significant improvement. The use of continuous models in image processing (e.g. wavelets, B-Splines, etc.) is more and more important and enables in a better and more convenient way the use of global models without being dependent on the discrete parametrization.

If the problem is now clearly defined and stated, its resolution is mostly not directly reachable. Instead of the hypotheses of particular boundary settings which simplify the problem, mathematical theories closely related to physical concepts (conservation laws, front propagation) are used, mainly based on partial differential equations solved by gradient descend methods.

Several image functionals were formulated in [17], [154]. After presenting the Minimum Description Length principle, the dissertation is mainly focused on one of the most studied models, the functional suggested by Mumford and Shah in [154].

4.3.1 Minimum Description Length

Rissanen addressed in [174], [175] and [13] the *Minimum Description Length* (MDL) theory. This technique is closely related to the overall purpose of image reconstruction, which is not only restricted to noise filtering but also to the reduction of information to its essential meaningful content. The MDL principle is based on the simple idea that the best way to capture regular features in data is to construct a model in a certain class which permits the shortest description of the data and the model itself. Let's take a function $\mathcal{L}(\mathcal{X})$ called *Language* computing the information quantity to encode X. Quantification generally expresses the average number of bits of description per bit of input. Having some data D and models M_i, the MDL strategy is to choose the models M_i that minimize

$$\mathcal{L}\left(D|M_i\right) + \mathcal{L}\left(M_i\right) \qquad (4.4)$$

Leclerc in [126] established the equivalence between the posterior maximization within the Bayesian formalism and the code length minimization expressed in relation 4.4. He applied in this paper the MDL theory to image segmentation problems by constructing the shortest complete and stable image description. More recently, Refregier and Galland obtained promising results on SAR images and demonstrated the strength of the MDL theory in [72].

A variational and statistical model based on snake models and region growing algorithms is studied by Zhu and Yuille in [218]. They attempt to unify several formalisms by using a generalized Bayesian/MDL criterion to derive their evolution equations by the Euler-Lagrange equation technique. The interest was to integrate region-based criteria into the snake model. However, their active contour formulation suffers from topology changes. In [164], [163], Paragios and Deriche combined this formalism with geodesic contours and level-sets methods. The next sections present those techniques and their relation with the functional presented by Mumford and Shah in [154].

4.3.2 Mumford Shah Functional

In [154], Mumford and Shah formulated image segmentation as an optimization problem and defined it simultaneously as a smoothing and an edge detection problem. The image I to be segmented is approximated by piecewise smooth functions $\{u_i\}$, describing disjoint regions R_i bounded by a corresponding discontinuity set Γ, (the regions boundaries, see Fig. 2.5(a), page 13). The idea of Mumford and Shah [154] is expressed as minimizing the functional

$$E(u, \Gamma) = \mu^2 \int_\Omega (u - I)^2 \; dx + \int_{\Omega - \Gamma} |\bigtriangledown u|^2 \; dx + \nu \; |\Gamma| \tag{4.5}$$

where μ, ν are strictly positive parameters and $|\Gamma|$ denotes the Hausdorff measure (i.e. length in 2D, surface in 3D) of the boundary set Γ. The two first terms control the regions fidelity and their regularity outside the edges. The last term carries some constraint smoothness regarding the discontinuity representation Γ by minimizing its overall length, somehow in a similar way as Leclerc's idea [126] in the MDL approach. Good results are expected and obtained under the functional hypotheses; contrasted images which are populated with piecewise smooth surface objects.

Existence and Convergence

Unfortunately, mainly due to the nature of the unknowns (a function and a curve), the analysis and numerical implementation of the Mumford-Shah problem are very challenging. Indeed, as mentioned in [153], [4], [204], [27], this minimization problem is not convex and therefore has no exact solutions. To prove the existence of minimizers, conjectured in [154] and still open, Morel and Solimini establish in [153] several inequalities to prove the existence of a "less weak" minimum of E. In [2], Ambrosio reformulates the Mumford-Shah functional by introducing special bounded variation functions and proves the equivalence between the minimizers of the two formulations. This work leads to prove in [3] that the Mumford-Shah functional can be approximated, in the sense of a Γ convergence, by regular functionals defined on *Sobolev* spaces. It constitutes a major improvement regarding the numerical computation of minimizers of E. Progresses have been made, as shown for instance by the work of Bonnet [18]. For a more detailed and accurate presentation and analysis on the Mumford-Shah conjecture, the reader is referred to [154], [2], [3], [153], [18], [4].

Some Properties and Functional Limits

In the same references, several interesting properties are reported and related to the boundary shapes, the nodes characteristics of the boundary intersections. Using the notations given by Mumford and Shah in [154], two reduced cases have been studied and described in numerous papers:

(1) the functional E_0, restriction of $\mu^{-2}E$ to piecewise constant functions $u_i = a_i$, is deduced from Eq. 4.5 and presented separately in chapter 5 (pages 66-87).

(2) the functional E_∞, which only relies on Γ is given by

$$E_\infty(\Gamma) = \lim_{\mu \to \infty} E(u, \Gamma) = \int_\Gamma \left[\nu_\infty - \frac{\delta I}{\delta n}^2 \right] ds \tag{4.6}$$

where ds refers to an arc-length parametrized curve and the second term denotes a gradient computed in the normal direction of Γ. We notice that this functional is only dependent on Γ. The minimization of the presented functional is seen as a generalized geodesic problem; it leads to minimize Γ_i boundaries through the constant term ν_∞ while finding the largest gradients. Under certain restrictions on Γ, this optimization is a well-posed problem which is solved by using quadratic functions admitting a unique minimum with boundary conditions.

Some particular applications of the Mumford and Shah functional are presented in the next section with the active contours and level-sets methods.

4.3.3 Snakes and geodesic Contours

Starting from an initial position around an object, *active contour* (or *snake*) models aim to detect the closest contours by deforming a curve γ to match locally sharp image intensity variations. Under the convergence of an evolution process, γ follows the edges of the extracted object. The snake model was introduced by Kass in [106] using either closed or open curves defined by $\gamma(p) = \{x(p), y(p)\}$ where $p \in [0, 1]$ is an arbitrary parametrization. The curve evolution is derived by minimizing an energy functional defined as the sum of two terms: external and internal energy to control respectively contrast $g(s) = 1/(1 + s^2)$ and smoothness requirements along the curve.

A drawback of Kass's model is its formulation dependence with the curve parametrization. To circumvent this limitation, Caselles introduced *geodesic active contour* models in [26]. The geometric form of the geodesic snake is expressed as an arc-length parametrized curve, ds replacing the arbitrary term p. A compromise between length and contrast is obtained by minimizing an energy functional

$$E(\gamma) = \int_0^{L(\gamma)} g\left(|Du(\gamma(s))|\right) ds \tag{4.7}$$

where g is a positive contrast function, $L(\gamma)$ the length of the curve γ and $|Du(\gamma(s))|$ the magnitude of the gradient computed on the image u. A disadvantage of such a model is that minimizing this functional forces to shrink the snake and decrease its length.

In the model of Kimmel and in [112], the isotropic term $|Du(\gamma(s))|$ is replaced by the term $u_n(s) = Du(\gamma(s)).\vec{n}(s)$ where $\vec{n}(s)$ is the curve's normal vector. Using contrast functions in form $g(t) = t^\alpha$ with $\alpha > 0$, the *average* contrast is maximized by the functional

$$E(\gamma) = \frac{1}{L(\gamma)} \int_0^{L(\gamma)} g\left(u_n(s)\right) \ ds \tag{4.8}$$

We notice that the case $\alpha = 2$ refers to the Mumford-Shah functional. Higher power leads to faster convergence but increases sensitivity to noise. Weak accuracy is expected as well as possible curve duplication on areas which present the strongest contrast variations along the curve.

To maintain the object consistency, curve re-parametrization is essential during the numerical scheme based on a gradient descent. The parametric curve description prevents or strongly limits topological changes (curve splitting or merging for instance). In [35], Cohen introduces a constant force called *balloon force* which pushes the contour out (or in) along its normal $\vec{n}(s)$. An energy is derived from this force and added in the energy functional to increase the convergence speed by relaxing initialization incidence and to detect non-convex objects. An example of application of geodesic contour is given page 144.

4.3.4 Bayesian Equivalence

An attempt for unifying active contours and region growing approaches has been formulated by Zhu and Yuille in [218]. In this publication, the authors proposed a *region competition* algorithm based on Bayes / MDL criterion using the variational principle. They emphasized a link between Bayesian approaches for image segmentation and the cartoon limit case of the Mumford Shah functional. Their work was completed in [23], by a demonstration of the equivalence between the Mumford Shah functional and the first-order approximation of a Bayesian model with probability densities estimated in local windows. The resulting generalization enables to approximate the image with piecewise smooth Gaussian distribution $\{m, \sigma\}_i$.

4.3.5 Level-Sets Methods

In the same way as *SVM* methods in image classification, levels-sets procedures have been extremely popular over the last years. The level-set method was originally proposed by Osher and Sethian in the celebrated paper [158]. Contrary to snakes or geodesic contour techniques previously presented, the main interest of this approach relies on its ability to handle topological changes while computing front propagation. The closed curves or surfaces evolution is achieved on a Cartesian grid, avoiding any re-parametrization steps of the driven object or curve duplications. Shape splitting, holes creations or their reverse operations are therefore monitored in a convenient way. In this formalism, a closed curve Γ is modeled as the zero level set of a Lipschitz continuous function ϕ, called *level-set function* $\Gamma = \{x \in \Omega \mid \phi(x) = 0\} = \phi^{-1}(0)$

Γ can consist of many disjoint and connected components or loops which divides the finite lattice Ω in two complementary parts $\Omega^{\pm} = \{x \in \Omega \mid \pm\phi(x) > 0\} \quad \Rightarrow \quad \Omega = \Omega^+ \cup \Gamma \cup \Omega^-$

Let $H(y) = 1_{y>0}(y)$ be the Heaviside function for $y \in \mathbb{R}$, applying the property

$$H(y) + H(-y) = 1 \quad \forall \, y \neq 0$$

to the level-set function we obtain

$$H(-\phi(x)) = 1 - H(\phi(x)) \quad \forall \, x \in \Omega \setminus \Gamma$$

Applying such a formalism to the Mumford Shah model defined in Eq. 4.5 leads to

$$
E(u^+, u^-, \phi | u_0) = \int_\Omega (u_0 - u^+)^2 \, H(\phi) \, dx + \int_\Omega (u_0 - u^-)^2 \, (1 - H(\phi)) \, dx +
$$
$$
\mu \left[\int_\Omega |\bigtriangledown u^+|^2 \, H(\phi) \, dx + \int_\Omega |\bigtriangledown u^-|^2 \, (1 - H(\phi)) \, dx \right] + \nu \int_\Omega |\bigtriangledown H(\phi)|
$$
(4.9)

A solution to Eq. 4.9 is obtained using *partial diffusion equations* (PDE)s. A two-level scheme is iteratively solved until convergence $(\phi, u^{\pm})_t$:

1. For a given evolution state (i.e. ϕ fixed), the piecewise smooth approximations $u^{\pm}$ are separately solved by Euler Lagrange equations with Neumann boundary conditions

$$u^{\pm} - u_0 = \mu \triangle u^{\pm} \quad \text{on} \quad \pm\phi > 0 \,, \quad \frac{\delta u^{\pm}}{\delta \overrightarrow{n}} = 0 \quad \text{on} \quad \phi = 0$$

2. Once $u^{\pm}$ are estimated, the motion of the level set is computed

$$\frac{\delta\phi}{\delta t} = \delta\phi \left[\mu \bigtriangledown (\frac{\bigtriangledown\phi}{|\bigtriangledown\phi|}) - ((u^+ - u_0)^2 - (u^- - u_0)^2 + \mu(|\bigtriangledown u^+|^2 - |\bigtriangledown u^-|^2) \right]$$

In [179], Sansom et al. suggested a supervised image classification method based on a variational approach using level set functions. Many authors extended this formalism as an image partitioning problem by means of multiple level-set functions. Chan and Vese introduced in [28] level-methods to compute Mumford-Shah segmentation. They provided several issues to enhance the use of level-sets methods compared to snake models. The presented approach, based on one level-set function, is mainly used to extract singular objects from the background image.

Using the Mumford and Shah functional, Vese and Chan extend in [204] the 2-phase segmentation model to a higher number of regions to be segmented. In opposition to [179], they define a multiphase level-set framework which avoids vacuum and overlaps problems by construction and requires, to represent n phases, $log(n)$ level set functions in the piecewise constant case and only two in the piecewise smooth case using the *Four-Color Theorem*.

At the present time, this field is generating considerable research, not only based on optical data. To segment SAR images, a significant contribution is given in [5] and [6]. All authors mentioned that the initialization stage is critical, which in any case scales the convergence speed and may condition the results obtained.

An Image Segmentation Framework

Abstract

A boundary case of the Mumford and Shah functional is described and used as foundation for an image segmentation algorithm. This chapter requires some concepts presented in sections 2.2, 4.2 and 4.3.

The proposed method is mainly addressed for the segmentation of multispectral remote sensing images, but it can also be useful for more general purposes using grayscale images (panchromatic, X-ray), SAR polarimetry images or volumetric data-sets generated by computer tomography. In particular, we sketch a protocol for the processing of large images resulting from metric satellite sensors (i.e. couple of gigabytes in the case of Ikonos) and give some examples.

This segmentation stage acts as a low-level operation with a little set of parameters, followed by image analysis processes detailed in the next chapter.

5.1 The piecewise constant Case

As already mentioned in section 4.3.2, page 62, there are boundary conditions where the Mumford-Shah functional becomes a well-posed problem.

5.1.1 Hypothesis Relaxation and Problem Statement

E_0, the restriction of E to piecewise constant functions u, is the natural limit functional of E when the set of boundaries Γ, is fixed and $\mu \to 0$ (see Eq. 4.5). Applied to the image I, defined as a scalar function on the image domain Ω, this case is named the cartoon limit [154] and the functional is rewritten

$$E_0(u, \Gamma) = \sum_i \int\!\!\int_{R_i} (u_i - I)^2 \, dx \, dy + \lambda \, |\Gamma| \tag{5.1}$$

where $\lambda = \nu/\mu^2$ acts as normalized scale parameter and R_i are the regions of the partition, with $\bigcup_{i=1}^{n} R_i = \Omega \setminus \Gamma$. The piecewise constant values u_i taken by the solution are the mean values defined on R_i and $|R_i|$ is the corresponding surface area

$$u_i = \frac{1}{|R_i|} \int\!\!\int_{R_i} I \, dx \, dy = \text{mean}_{R_i}(I) \tag{5.2}$$

In the case where u is restricted to the values ± 1, a solution to Eq. 5.1 is known by using simulated annealing methods (Ising's problem). Since the values u_i are simply the means of the image restricted to R_i, the boundary set Γ is sufficient to describe the segmentation, we can write $E(u, \Gamma) = E(\Gamma)$ and Γ is used to denote the segmentation as well. Γ denotes

- a *normal* segmentation when every sub-segmentation Γ' of Γ satisfies

$$E(\Gamma) \leq E(\Gamma') \tag{5.3}$$

- a *2-normal* segmentation if for every pair of connected regions, Γ', the segmentation resulting from their merging, verifies Eq. 5.3.

A consequence is that any merging operation results in an energy increase. Below, we address the problem of solving *2-normal* segmentations.

5.1.2 Multi-scale Approach

In Eq. 5.1, $E(\Gamma) = 0$ implies simultaneously $\lambda = 0$ and the set of piecewise constant functions exactly fitting the data. Any increase of $E(\Gamma)$ results in coarser segmentations with larger approximation errors and fewer boundaries.

λ, called the "scale" parameter by Koepfler et al. [115], indicates the coarseness of a segmentation. This term represents a compromise between boundary regularization and internal homogeneity, it partially addresses the segmentation issues (see Haralick's definition, page 58). Keeping λ small enables weak constraints on the amount of boundaries and results in segmentations close to the original data. With the increase of λ, "fine" details are progressively smoothed, coarser representations (like cartoons) are derived.

To partition an image I into disjoint regions, most of the segmentation algorithms are based on energy minimization: among several operations performed on the image partition, the one which gives the minimal energy increase (e.g. smallest MSE in the least square sense) is applied.

Let us recall some basic operations applied on image partitions expressed with piecewise constant functions $u_i = \text{mean}_{R_i}(I)$. Following notations used in Eq. 5.1, the mean of two piecewise constant functions u_a, u_b is

$$u_{ab} = \left(\frac{|R_a|u_a + |R_b|u_b}{|R_a| + |R_b|} \right) \tag{5.4}$$

Integration relations over a defined surface obviously follow

$$\int_{R_i} u^2 \, dx = |R_i| u_i^2 \tag{5.5}$$

Given two arbitrary connected regions $\{a, b\}$, the energy increase resulting from their merging

is given by

$$
\begin{aligned}
\Delta E = \ & E\left(\Gamma \setminus \delta_{ab}\right) - E\left(\Gamma\right) \\
= \ & \int_{R_{ab}} (u_{ab} - I)^2 \ dx - \sum_{i=\{a \ b\}} \int_{R_i} (u_i - I)^2 \ dx - \lambda \ \delta_{ab} \\
= \ & \int_{R_{ab}} u_{ab}^2 \ dx - 2\int_{R_{ab}} u_{ab} I \ dx + \overbrace{\int_{R_{ab}} I^2 \ dx - \int_{R_a} I^2 \ dx - \int_{R_b} I^2 \ dx}^{=0} \\
& - \int_{R_a} u^2 \ dx - \int_{R_b} u^2 \ dx + 2\int_{R_a} uI \ dx + 2\int_{R_b} uI \ dx - \lambda \ \delta_{ab}
\end{aligned}
$$

using Eq. 5.2, Eq. 5.5 and simplifying gives: $\qquad$ (5.6)

$$
= \ -\left(|R_a| + |R_b|\right) u_{ab}^2 + |R_a| u_a^2 + |R_b| u_b^2 - \lambda \ \delta_{ab}
$$

replacing the weighted mean computed in Eq. 5.4, we obtain:

$$
= \left(\frac{|R_a| \cdot |R_b|}{|R_a| + |R_b|} \cdot \left(u_a^2 + u_b^2 - 2u_a u_b\right) \right) - \lambda \ \delta_{ab}
$$

$$
\Delta E = \left(\frac{|R_a| \cdot |R_b|}{|R_a| + |R_b|} \right) (u_a - u_b)^2 - \lambda \ \delta_{ab}
$$

with δ_{ab}, the boundary shared by R_a, R_b[1], removed from Γ during the fusion of R_a and R_b.

The energy variation resulting from the merging of two regions does not involve terms related to the original image; it only relies on the piecewise constant functions and the boundary set Γ. As suggested in [115], [153], minimizing the merging cost for two connected regions leads to

$$
E\left(\Gamma \setminus \delta_{ab}\right) - E(\Gamma) = 0 \quad \Leftrightarrow \quad \lambda = \frac{|R_a| \times |R_b|}{|R_a| + |R_b|} \frac{||u_a - u_b||^2}{\delta_{ab}} \tag{5.7}
$$

Hence for every pair of connected regions, it exists a value λ_{ab} which corresponds to no energy increase. A fine to coarse strategy can be driven by the scale parameter used as decision criterion for merging operations.

The term itself of "scale" for the regularization parameter λ should be interpreted in this context: a value which controls and enables to "navigate" into a *bottom-up* approach starting from fine details and evolving to rougher and rougher image descriptions (fewer boundaries and larger region areas).

(a) $\lambda = 36.07$, 16384 regions (4×4)   (b) $\lambda = 171.4$, 4096 regions (8×8)   (c) $\lambda = 539.3$, 1024 regions (16×16)

Figure 5.1: Cartoon limit applied on a Lena image (size 512×512) with increasing λ values, corresponding to segmentations with n regions. The size $m \times m$ refers to the average area per region. (a) Under-segmentation with $n = 16384$ (i.e. $m = 4$), (b) Intermediate segmentation, (c) Over-segmentation with $n = 1024$ (i.e. $m = 16$).

5.1.3 Merging Algorithm combined with multi-scale

Starting from an initial configuration, (e.g. each pixel is a region), the proposed segmentation algorithm iteratively merges pairs of connected regions $\{a, b\}$ by ordering all the corresponding λ_{ab} values in a list $\{\lambda_{ij}\}$ and by selecting the pair which provides the lowest energy increase (smallest λ value) at each fusion stage. The pair candidate is merged, the resulted region and all of its adjacent regions are updated (boundary description, re-computation of their merging criterion) and the list $\{\lambda_{ij}\}$ is ordered again to take into account deleted or modified boundaries. The procedure is recursively applied until it reaches a user-defined λ value.

There is no warranty that the obtained solution is the optimal one, however this work-flow is reasonable and prevents excessive combinatorial computations. The irreversibility induced by the merging operations over the λ schedule is the main drawback of the presented method. Information fused cannot be retrieved, making somehow challenging the setting of the stopping criterion. However, such a scheme presents several interesting properties; a full image partition is guaranteed within relatively short computational time; a unified and single process is involved which encapsulates through Eq. 5.7 both boundary regularization and region homogeneity.

5.2 An Algorithm

We propose an algorithm based on the cartoon limit. Objective is to seek the segmentation process at a low level of analysis and use the resulted segmentation as input for further image

---

[1]Thereafter, the term "boundary" is used to represent inter-pixel connections between two adjacent regions (see section 6.1, page 87).

analysis processes, presented in the next chapter. The uncoupling between image segmentation and image analysis is motivated by several reasons: to include neither heuristics on the earliest stage of a computer vision system, nor a priori information. It is an asset in terms of robustness and universality, preventing irreparable issues for the overall flowchart when dealing with images which slightly differ from what was expected or with various information contents.

Differences with the algorithm detailed in [115] are mainly related to implementation issues and data structures involved. Koepfler et al. exploited the consistency and the sufficiency of the boundary set to describe the segmentation. Our approach goes in the opposite direction; instead of modeling Γ, we focus on the region definition. Despite the duality between both formulations, each of them has pro and cons. If the boundary description is definitely more "elegant" and offers further possibilities (e.g. topological analysis, boundary regularization processes using B-Splines or Fourier descriptors), such a modeling is rapidly defeated with an important amount of data.

Indeed, a global functional potentially requires to have the full image at disposal; such cases are related to the scale of the data content and occur in presence of sparse information, spread all over the image. Some typical examples in remote sensing images are forestry areas, motorways or rivers. When the image to be segmented becomes too large, the initialization stage on a fixed grid constitutes the bottleneck of boundary description implementations.

A regular grid where nodes, single and connected boundaries are encoded enables to preserve topological consistency while performing time efficient operations[2]. However, physical memory on computer is bounded; a compromise has to be reached between respectively high/low complexity data structure and addressed memory/time efficiency.

For these reasons, we deliberately choose to keep the data representation as simple and minimal as possible, to reduce the amount of data (at least in a first stage). In Alg. 3, a flowchart for an image segmentation algorithm is proposed. Let's describe each step more in detail.

1. **Initialization:**
   - label regions,
   - extract connected boundaries, fill list $\{\lambda_{ij}\}$

2. **Sorting:**   update/sort $\{\lambda_{ij}\}$, take smallest λ, $\lambda = \lambda_{ur}$

3. **Merging:**   merge R_u and R_r into R_u
   - merge labeled regions, remove δ_{ur}
   - update boundaries of the updated region R_u
   - recompute λ for all neighbored regions of R_u

4. **Convergence:**   repeat steps 2-3 until convergence criterion λ_c is reached

Algorithm 3: Flowchart of the proposed segmentation algorithm.

---

[2]We retrieve here some of the distinctions between snakes and level sets approaches, presented page 63-65.

5.2.1 Data Modeling

An essential issue in image analysis and computer vision is the identification of "meaningful" regions, features or primitives. Image labeling algorithms index image's area or shape according to their data content with a particular *code* referring to look up tables (LUT) or indexed arrays. Interests are multiple: region or clusters identification (from image partitioning and/or classification), boundary and chain code extraction, data manipulation and analysis, etc.[3].

A more compact data description results from this low level process with the possibility to link image location to elaborated object's modeling. It constitutes a pre-ordering of the data content with a higher level of information representation.

Indexing of regions

The labeling of images consists in assigning a specific identifier or keyword i to each pixel composing a group of adjacent pixels with homogeneous data content. The process is driven using a structuring element $\mathcal{N}$ [4] and the predicate

$$\mathcal{P}(R_a, R_b) = \text{true} \quad \Longleftrightarrow \quad \text{Distance } \mathrm{d}(R_a, R_b) = 0$$

The most intuitive method for region labeling is the use of a region growing algorithm (see section 4.2.3). The image is scanned in one pass: each time a position p is not labeled, a region is created with an incremented label i. A seed at the position p is generated and propagated to each adjacent pixel $p_j = p_{\mathcal{N}_j}$ that matches the segmentation criterion.

When the predicate is verified, the label i is assigned to the position p_j and the procedure is repeated at the given position. The Seed's recursive propagation is stopped when no unlabeled pixel is found (i.e. all the pixels composing the studied regions being labeled). The recursive behavior is implemented in several ways by using for instance:

(1)   a recursive function to propagate the region at $p = p_{\mathcal{N}_j}$ when $\mathcal{P}(p, p_{\mathcal{N}_j})$ is verified,

(2)   an incremental approach[5] based on temporary index tables.

In the case of image labeling, the partition resulting from the region growing methods is complete and enables to properly handle empty areas (e.g. border of registered or rotated images) in defining a

---

[3]In section 2.1.4, some applications are presented (encoding, visualization) using indexed table, dictionaries, LUT, etc. for the modeling of images.

[4]$\mathcal{N}$ describes neighborhood relationships for the studied position p, $p_{\mathcal{N}_j}$ denotes one of its adjacent positions. For region-based labeling, $\mathcal{N}$ is generally a $4-$connectivity system, $j \in \{1...4\}$.

[5]A pseudo-code of the incremental method is given in Alg. 6, page 181. Indeed, the presented algorithm is not only restricted to image labeling purposes. A region growing algorithm based on iterative scans can be obviously derived using for each pass a predicate with decreasing strength applied on the previous segmentation result. Unfortunately, such a method would have at least three major drawbacks; (1) seeds and, (2) predicate sequence initialization and (3) the sequential region merging is driven by the image scanning convention (i.e. no specific ordering according to the pair of regions to be merged in priority during a given pass).

second predicate to check data validity. The use of recursive functions is generally not recommended in terms of efficiency, and the risk of excessive interlinked functions calls can run the system out of stack space pretty easily. The incremental method presents a high computational efficiency but requires additional memory resources.

An intermediate implementation is chosen which avoids both incremental and recursive issues. The related pseudo code is given in Alg. 7, page 182; it follows

1. **Initialization:**   scan line by line $\mathcal{N}$ and label the index map

2. **List of Regions:**   create regions related to each generated label
3. **Merging:**   scan line by line $\mathcal{N}$, if predicate $\mathcal{P}()$ between adjacent regions is verified, merge the regions

Algorithm 4: Flowchart of the proposed image labeling algorithm.

In Alg. 4, two passes are required; the first one is applied to the index map prior to the region initialization, the second one aims at merging regions which have been labeled using several seeds, as for instance *Y-shaped* regions. Both Alg. 6 and Alg. 7 remain unchanged when $\mathcal{N}$ is a 8-connectivity system[6] and extensible to high-dimensional data (e.g. volumetric data-sets described with voxels) by defining N-dimensional connectivity systems. In comparison with recursive or incremental implementations, the proposed algorithm is proved to be more effective in terms of space and computational time.

Region and Boundary Description

The definition of an image on a regular grid simplifies the representation of regions and boundaries as depicted in Fig.2.5(a), page 13. Each region is identified by a label, its location is indicated in the index map. The region definition encapsulates its related piecewise constant function (scalar in the case of multispectral image) and a list of all its neighbored regions defined by their boundaries shared with the studied region. In addition, a list of positions (pixels/voxels belonging to the region) is created[7]. Another formulation is to model each region by its external closed contour, defined by a set of oriented and connected boundaries.

An important issue is the pyramidal evolution of the method; starting from an initial set of boundaries Γ, the boundary δ corresponding to the smallest λ is removed during the merging procedure and results in $\Gamma \setminus \delta$. Since no new boundary is created, a graph description of Γ is sufficient. It consists of the pair of regions sharing the boundary, its length and the computed scale parameter λ.

---

[6]The description of higher level of information is discussed in chapter 6.

[7]Indeed such information is redundant with the use of an index map. However, to improve computational efficiency with faster data accesses, both data location representations are maintained.

5.2.2   Multi-scale Evolution

Numerous algorithms exploit the image's regular structure in applying to every pixel p a function f corresponding to a kind of convolution. Methods whose f_p are independent of each other enable massive parallelism implementations. On the contrary, dependencies result in sequential approaches which require to order $\{f_p\}$ according to some rules. For a region merging algorithm, the function is obviously "*merge two adjacent regions*"; it remains to define the rule to iteratively apply the sequence of f_i.

As suggested by Koepfler et al. in [115], the segmentation criterion is the smallest λ defined for each boundary between pairs of regions. The multi-scale evolution can be seen as a sorting problem based on the set of boundaries. The amount of data to be sorted is consequent. Given an image of size $w \times h$, the list contains in the worst case $n = 2wh - w - h$ elements (i.e. with an image of size 2048×2048 about 8 millions of boundaries).

Information Sorting

In computer science, information ordering is an essential operation. To handle it, a large scope of methods has been investigated and can be subdivided into two families, according to their complexity[8]. The analysis of a sorting algorithm emphasizes several properties: *stability* (equal items are not reordered), *in-place* (no extra-space required), *adaptive* (algorithmic complexity almost in $O(n)$ when the data are nearly sorted or few different items), *robustness* according to initial conditions (i.e. data distribution impacts on average and worst complexity). Unfortunately, there is no algorithm that has all of these properties, the choice of a suitable sorting algorithm is application dependent.

The first class, in $O(n^2)$ includes *bubble*, *insertion*, *selection*, and *shell* sorts. *Heap*, *merge*, and *quick* sorts belong to the second class, operating in $O(n \log(n))$. The oldest, simplest (but slowest) sort in use is the **bubble** sort. It consists in comparing each item in the list with the item next to it and swapping them if required. The process is repeated until a pass all the way through the list without swapping is obtained.

Merge and **quick** sorts are based on divide-and-conquer strategy using recursive sort. Both methods split the list to be sorted into two sub-lists which are recursively sorted. The *merge* sort needs a further stage to merge back each equal half list. In addition, it requires additional arrays: one to store each half of the data-set and another one to store the sorted list (*merge* is not an in-place sort). *Quick* sort is the fastest of the common sorting algorithms; it is an in-place, massive recursive sort using an element in the list called "pivot" to define the two sub-lists (smaller/larger elements than the pivot). The choice of the pivot has a considerable impact on the efficiency of the

---

[8]Algorithmic complexity is generally written in a form known as $O()$ notation with n the size of the set the algorithm is run against. An algorithm that runs in time $O(n)$ has a linear complexity; it takes twice as much time to operate on a set of 50 items than it does on a set of 25 items. An algorithm with $O(n^2)$ would take 4 times longer.

method which falls, in the worst case, to $O(n^2)$ (e.g. operating on lists that are mostly sorted in either forward or reverse order). Like the *merge sort*, *quick* sort is recursive and therefore unsuitable for applications that run on machines with limited memory.

Although the **Heap** sort is the slowest of the $O(n \log(n))$ sorting algorithms, this method is very attractive for very large data-sets since it does not require massive recursion or multiple arrays, and its complexity both in average and worst cases remains $O(n \log(n))$. The method presents two stages based on a binary heap structure (see section 2.1.3).

The *heap* sort consists in (1) building a binary heap which is filled with the data-set. Similarly to the Peter's principle, small *(min-heap)* / large *(max-heap)* are promoted to the top of the structure, the root being the smallest/largest, (2) storing in the sorted list the root node extracted from the heap and then update the heap. The procedure is repeated until no items remain in the heap (i.e. the sorted array is full). The *heap* sort is not stable; in almost sorted cases, the original order is destroyed during the initialization stage (1).

The *merge* sort is slightly faster than the *heap* sort for larger sets, but it requires twice as much memory because of the second array. Most of the time, the *quick* sort is the preferred option, the *heap* sort being a better solution for very large sets.

Heap of Boundaries

We implement a *heap* sort algorithm (min-heap) with a binary heap whose nodes are boundaries with the scale parameters as keys. In addition to the remarks presented previously, such a structure particularly matches our requirements: extract the boundary presenting the smallest scale parameter among a large amount of boundaries, and manipulate a list of boundaries whose scale parameters (the values to be sorted) are modified during the multi-scale evolution. Indeed, by construction, even if the heap is partially sorted, the root is undoubtedly the smallest feature. A total ordering prior to every merging procedure is not required, the sorting being achieved during the segmentation evolution.

Construction: The construction of the heap of boundary is realized by promoting every inserted boundary from the first empty terminating node to the root. The bottom-up approach stops when the inserted boundary presents a λ value larger than the one of its parent in the heap.

The in-place property is satisfied using the same array to store both boundary set and heap structure. A heap position is inserted to the boundary definition. By manipulating the boundary set/heap, namely with update and node removal operations, it only remains to "bubble-up" the smallest keys.

Update: When a boundary is updated (λ recomputed) or moved to an empty node of the heap, its relation within the hierarchical structure has to be checked. According to its sorting value, it will either be promoted to the top of the heap or fall to the bottom (large value) while taking into account both left and right node children (siblings of the binary heap are sorted as well).

Deletion: When a boundary is deleted, its node in the heap becomes empty. To maintain complete the heap structure, its element is taken to fill in the empty node before removing the last terminating node. The inserted element is then updated according to the **Update** step previously defined. If the removed element was the root node, the inserted element can either stay at the root node or sink in the heap.

Since children are sorted, the inserted element is swapped with the smallest value of the two children until both children have larger keys. Indeed, bottom's heap elements are expected to have important values; in most of the cases, they go back to the bottom of the heap, however this long "way" enables to partially sort the heap structure and promote the smallest candidate.

Merging procedure

The proposed algorithm is split into several structures: (1) region definition by means of an indexed map, (2) a list of regions and boundaries to carry the relationships between the regions and (3) a binary heap structure to store boundaries and to control the iterative region merging procedure. All of them are linked together by a set of unique region identifiers. Given two connected and disjoint regions R_u and R_r, and δ_{ur} the shared boundary corresponding to the smallest scale parameter, their fusion into the updated region R_u - R_r being removed - follows:

1.	**Region description**	using the list of positions $\{p_r\}$, update the index map $i_r \leftarrow i_u$
		append $\{p_r\}$ to the list of positions $\{p_u\}$
		compute the resulting piecewise constant function u_u using Eq. 5.4
2.	**List of boundaries per region**	fuse the two lists of boundaries $\{b_r\}$ and $\{b_u\}$
		• remove δ_{ur} from both lists
		• append $\{b_r\}$ to $\{b_u\}$
		• update in $\{b_u\}$ multiple boundaries δ_{ui}^u, δ_{ui}^r with R_i being adjacent to R_u and R_r: $|\delta_{ui}^u| \leftarrow |\delta_{ui}^u| + \delta_{ui}^r$
3.	**Heap update**	recompute scale parameters for the list $\{b_u\}$, update the related nodes
		remove the root δ_{ur}, update the heap
		remove the nodes for the list δ_{ui}^r and update the heap

Algorithm 5: Flowchart of the proposed merging procedure.

Multi-scale Evolution

The algorithm is based on a merging scheme to build image partitions by removing at each aggregation the boundary with the smallest cost. The procedure is repeated until a defined energy threshold is reached or, as a limit case, until it remains only one region in the image.

(a) Original Spot 5 image

(b) Segmented image, $\lambda = 453.05$, 24576 regions (4×4)

(c) Segmented image, $\lambda = 1019.33$, 10922 regions (6×6)

Figure 5.2: Cartoon limit applied on a SPOT 5 image with increasing λ values, corresponding to segmentations where n regions remain. The size $m \times m$ refers to the average area per region. (a) is the original image (Region of Nice (France), 2.5 meter resolution, size 1024×384), (b) and (c) are segmented images with respectively $\lambda = 453.05$ and $\lambda = 1019.33$.

The resulting pyramidal sequence of image segmentations presents coarser and coarser partitions with the increase of the scale parameter's values. Some examples are given in Fig. 5.2 to 5.4, all of them being computed from the original SPOT 5 multispectral image given in Fig. 5.2(a).

Two image partitions are presented in Fig. 5.2 corresponding respectively to $\lambda = 453.05$ and $\lambda = 1019.33$. In both images, the relative good behavior of the segmentation on landscape areas (forests with various tree types, fields, rivers) or on man-made features (road networks, tennis area or houses) is emphasized. In the coarser segmentation (Fig. 5.2(c)), the textures (plantation fields) on the bottom-right image's part are lost and some of the small ways are broken. However, it should be noticed that the investigated image is sampled to 2.5 meters (fusion of multispectral-panchromatic SPOT 5 data). According to the resolution, some thin lines or curves cannot be described using a 4-connectivity system analysis and remain as groups of disconnected regions (see Fig. 5.4).

Fig. 5.2(b) is ①, Fig. 5.2(c) is ②, ③ ④ ⑤ are Fig. 5.4:   ③   ①   ⑤ ② ④

Figure 5.3: Scale parameter evolution according to the number of regions in the segmented images (gray dot-line). ○ denotes the segmentations presented in Fig. 5.2 and Fig. 5.4 in the hierarchy of segmentations. The black solid-line gives a smooth trend of the curve.

The coarseness of the segmentation is emphasized in Fig. 5.3. It illustrates the scale parameter evolution during the sequence of segmentations according to the number of regions that remain in the partition. The curve was built by taking one sample at every ten iterations. It is not a strictly increasing function since disconnected regions with similar content may become adjacent by consecutive merging. In this case, the updated configuration results in lower λ values than the previous state. This behavior can be seen as the removal of a weak membrane that separates some similar liquid in two parts. Three examples of important λ steep decrease observed in Fig. 5.3 are presented in Fig. 5.4. All of them refer to thin objects, hardly described and extracted on a regular grid structure with 4-connectivity system analysis.

Merging sequence (case ③) with $n = 46974$ and $\lambda = 206.08$ for (b)-(d), $\lambda = 13.63$ (e)

(case ④) $n = 5800$, $\lambda = 1751.10$ and $\lambda = 905.37$

(case ⑤) $n = 16034$, $\lambda = 619.06$ and $\lambda = 48.01$

(a) Original image   (b) Segmented image   (c) Studied regions   (d) Merging   (e) Pair of regions

Figure 5.4: Analysis of some local decreases of the scale parameter curve (cases ◯ in Fig. 5.3) during the merging process of a SPOT 5 image (a). (b) segmented image prior to curve's decrease (boundaries between regions are drawn in black). (c) configuration of the three related regions to be consecutively merged. (d) merging step which generates a curve decrease on the next iteration (e).

The curve is smooth with weak values when the segmentations still present fine details (i.e. any small merging cost) and the fusion does not "destroy" any data structure. As the number of regions in the partition decreases, the curve slope increases and merging costs are more and more important. When few regions remain (in this example less than 20% of the initial amount), the curve presents an asymptotic trend which indicates important changes with the removal of objects and structures. Fig. 5.3 also depicts the limitations of the method which is based on "local" energy minimization[9] instead of searching for a global minimization with combinatorial approaches on the Mumford and Shah functional [79].

5.3 Analysis and Evaluation

Several modifications of the functional are presented to improve segmentation capabilities and solve some particular cases. In addition, the problem of segmenting large images is investigated.

---

[9]Energy minimization along the merging sequence, the current state depends only on the previous one.

5.3.1 Texture Discrimination and Surface Modeling

The Euclidean distance, computed in Eq. 5.7, can be replaced by the Mahalanobis distance (see p. 18), the main interest being to integrate second order statistics in the criterion and distinguish for instance two regions presenting the same mean values but different variances (or noise level). The segmentation behavior should be improved in textured areas and for images having different illuminations (see parrots p. 19). Unfortunately, if those assessments were mostly matched, we report a higher sensitivity between regions, resulting into inaccurate and rough boundaries.

Small regions

The analysis of the merging cost λ between two adjacent regions $\{R_a, R_b\}$, computed according to Eq. 5.7, presents three components. The first one is related to the data content with a term representing a spectral distance measure. The second one $R_w = \dfrac{|R_A| \times |R_b|}{|R_a| + |R_b|}$ is related to the weight of each region. When both regions are comparable in terms of area, we get $R_w \approx |R_a|/2$; when one of the region is much larger than the other $|R_a| >> |R_b|$, R_w becomes equivalent to $|R_b|$. As a consequence, a region will more likely merge with one of its adjacent regions presenting a smaller size than with one of similar or larger size. The last contribution concerns the relative disposition of the pair of regions with the length δ_{ab} and favors clamped regions; hence, given two adjacent regions, a $||$ shape's configuration is preferred to the $\perp$ one's.

Some tests have been conducted to promote the merging of small regions, especially when the segmentation becomes coarser in switching the heap sorting after a given state. If interesting results are obtained, such a regularization approach presents the major drawback to be highly heuristic when defining the sequence. Alternative is to first merge regions which are smaller than a threshold area. This stage is directly applied after the heap initialization and presents the advantage to avoid down-sampling operations as pre-processing step. Starting from the root, the heap is traversed; when a boundary having a small region is found, we remove it and update the heap. The procedure is repeated until no small region exists. Based on this result, the overall segmentation is proceeded until the selected λ is reached. These investigations were carried out to use an other region modeling which requires several samples (e.g. at least 3 points are required to define a plane).

Surface Description

In many cases, regions descriptions based on piecewise constant functions leads to satisfying results. However, the region homogeneity may present small biases or trends, caused by lighting reflexion, information content, etc. The use of c_1 piecewise functions partially circumvents such conditions. In addition, they can be evaluated within the cartoon limit case of the Mumford and Shah functional like c_0 functions. The equation of a plane in a 3-dimensional space is

$$\mathcal{P}(X, Y) = a_0 + a_1\, X + a_2\, Y$$

(a) Original image   (b) with $\mathcal{N}(0,5)$ (c) c_0 (d) c_1
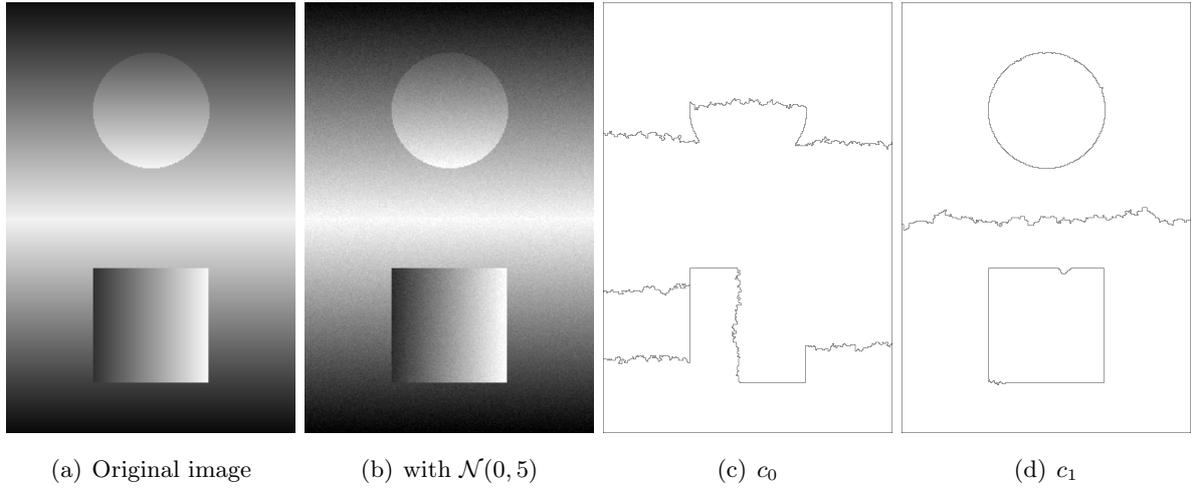
Figure 5.5: Comparison of piecewise c_0 and c_1 functions on a synthetic image with a Gaussian white noise $\sigma = 5$.

where $\{X, Y\}$ denotes the planimetric position; the data content acts as third coordinate.

In the case of n multispectral data, regions are described by means of n planes. Instead of using a direct relation as Eq. 5.7, the energy corresponding to each region is computed by (1) fitting the best plane using a Singular Value Decomposition (see page 21) and (2) storing the fitting error

$$error_a = \int \int_{R_a} (I - \mathcal{P})^2 \; dx \; dy$$

To evaluate the merging of two regions, the best surface for the union $\{a, b\}$ is computed (in the least square sense). The scale parameter corresponding to the union of regions is estimated by

$$\lambda = \frac{error_{a \cup b} - error_a - error_b}{\delta_{ab}} \tag{5.8}$$

An example is given in Fig. ?? where the modeling of regions with piecewise constant functions fails. The use of planes to approximate the regions delivers much better results. It should be noticed that computational time efficiency drastically drops, mainly due to the SVD required for estimating every surface. Another limitation concerns the initialization: with few samples, the least square surface fitting algorithm is sensitive to noise and therefore inaccurate. In [15], the initialization stage is carried out by means of mean and Gaussian curvature analysis prior to a region growing scheme. Higher orders of bi-variate polynomial surface models are employed as well, using a more general formulation

$$\mathcal{P}(X, Y) = a_0 + a_1 X + a_2 Y + a_3 XY + a_4 X^2 + a_5 Y^2 + a_6 X^2 Y + a_7 XY^2 + a_8 X^3 + ...$$

Such an analysis is generally limited to the fourth-order[10] for computational tractability and to avoid side effects (oscillation artifacts, numerical limitations).

---

[10]Bi-quartic functions, the 15 first parameters a_i considered.

5.3.2   Application to large Images

Since a scale parameter is defined for every boundary, sorting $\{\lambda_{ij}\}$ for an image of size $w \times h$ leads in the worst case to order $n = 2wh - w - h$ elements. Instead of starting the region merging procedure like Koepfler et al. [115] with the definition of one region per pixel (i.e. n boundaries), the region labeling stage enables to reduce the amount of boundaries by finding the exact set of regions. If this distinction presents minor differences when applied on the original image, it may become relevant when the image is already a partition. Using increasing λ values, such an iterative approach consumes less resources since a more compact data description is generated at each stage. It can be seen as a filtering process to reduce the amount of information without down-sampling the data and therefore reducing the segmentation accuracy.

Figure 5.6: Scalable image segmentation based on stripe decomposition. Each image shows an input of an iteration. The presented image is (6144×6144) big. Using a decomposition in three parts, the process requires for this data-set up to 1.6 Gigabyte of computer memory.

The subdivision of the image into multiple tiles or stripes to be separately processed and then recombined is a possibility to segment large images. Hence, having at disposal a bounded quantity of resources (memory), iterations are realized on decreasing images cut (see Fig. 5.6): starting from an initial configuration with n stripes, the overall process is achieved in n iterations, the iteration i corresponds to a segmentation based on the previous partition with $n - i + 1$ stripes, the last iteration being applied to the whole image coverage. Such an approach is validated by the slow evolution of the scale parameter when the region's density is important. In addition, the largest sequence of merging concerns small regions which are nearly similar (see Fig. 5.3 and Fig. 5.7).

To maintain the level of segmentation homogeneous in the whole image, each iteration except the last one starts with an estimation of the region's density per stripe. Fig. 5.6 presents the segmentation work-flow applied to a SPOT 5 image initialized with 3 stripes. The upper part of the image is mainly composed of landscape areas, the middle part of the city of Nice (France), and the last one of the Mediterranean sea, the coast and the airport. Fig. 5.7 illustrates the corresponding curve evolution for each part; as expected, the urban area presents the steepest
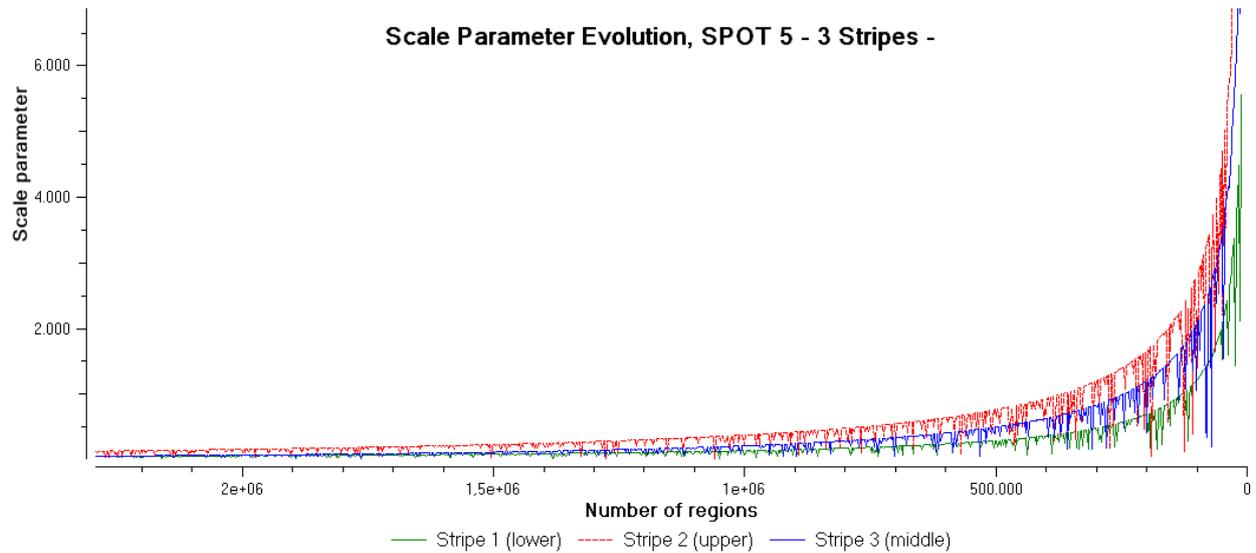
Figure 5.7: Scale parameter evolution according to the number of regions in the segmented image stripes (extracted from the original image). The stripes are shown in Fig.5.6 and correspond to the upper part of the image (stripe 1), middle part (stripe 2) and lower part (stripe 3). Each of them have 12582912 pixels (6144 × 2048), the depicted merging sequence in the curves starts with a region's average size of 5 pixels.

evolution curve with man-made features and strong targets. Conversely, large and homogeneous objects or patterns like sea or forest areas keep the curves on lower values[11]. It results in higher region densities for the parts where λ is larger (Table 5.1).

Since the parameter λ_s is highly dependent on the data content, setting such a threshold is delicate to stop the merging process. Its dual, the number of regions in the resulting partition is used to control the process since an average memory space per region is estimated. Expecting as result a segmentation with r_s regions, an estimation of the region's repartition is achieved based on a linear regression on the set of n evolution curves, with $\sum_i^n r_i = r_s$ and $\lambda_i = \lambda_s$, $\forall\{i\} \in \{1...n\}$. Then the region merging algorithm is applied to each stripe separately until λ_s (resp. r_i) is reached.

The corresponding configuration is given in Table 5.1. At each stage, the resulting amount of regions is computed to maintain constant the resources required for each stripe process (region and boundary descriptions and index map) and to reduce the volume of data. This progressive approach is followed in order to finalize the segmentation on the whole image when the merging costs become important. In Table 5.1, the last iteration is applied on a partition with a region density of about 20%. At this density, the behavior of the scale parameter is still smooth as depicted in Fig. 5.6 (where the last 20 % are represented for each stripe). The process could be achieved in one iteration, however a further regularization stage would be necessary for the regions which overlap multiple stripes. Applying several iterations enable to nearly cancel the influence of the stripe's processing.

---

[11] λ can also be used to locally quantify the amount of information within an image.

Stripe	Iteration 1 size 6144 × 2048				Iteration 2 ('), Iteration 3(") size 6144 × 3072, 6144 × 6144			
	regions	boundaries	r (%)	b (%)	regions	boundaries	r (%)	b (%)
1	12177883	24633040	96.78	97.88				
1'					9918667	22738695	52.55	60.23
2 1"	12390643	24927117	96.47	99.05	7546314	18013857	20.00	23.86
2'					8956550	19815356	47.45	52.49
3	11333666	23507711	90.07	93.41				

Table 5.1: Region and Boundary's densities per stripe for the SPOT 5 image prior to each iteration. For the first iteration, the estimated criterion $\lambda = 7.952$ corresponds to half the amount of regions (in the worst case). For the second iteration, $\lambda = 67.00$ corresponds to a reduction by five.

5.3.3 Some Examples

Experiments were done with larger data-sets, such as Cartosat panchromatic data (8 bits, size ca. 15000×15000, 2.5 m) and Ikonos multispectral images with 4 channels (12 bits, 1 m) of comparable image size. Using a computer with 8 Gigabytes memory, each data-set required to be initially subdivided in 5 stripes. According to the raw data-set processed and its content, we reported a range of memory allocation from 115 to 150 bytes per pixel. The last iteration was performed on the whole image using the previous segmentation, containing about 20 millions of regions. In Fig. 5.8 - 5.10 some results are presented, extracted from three levels of segmentations noticed (b), (c) and (d) in Table 5.2.

Iteration	Segmentations	regions	boundaries	r (%)	b (%)	$\overline{|R_i|}$	λ
0	Image grid	168849928	337699856	100	100	1	-
	Original data	164127734	328343760	97.203	97.229	1.03	-
... 4	(a)	21106025	52847701	12.5	15.649	8	2628.36
5	(b)	750444	1939045	0.44	0.574	225	47524.79
5	(c)	422124	1063756	0.25	0.315	400	73590.01
5	(d)	270160	659648	0.16	0.195	625	102352.99

Table 5.2: Settings and statistics to segment an Ikonos image (size 14072×11999).

Other tests have been carried out on polarimetric SAR images from the ENVISAT satellite. Compared to single band SAR information, the content captured in multiple polarization significantly improves the use of such a method with SAR data. Results obtained are promising, even if the SAR artifacts affect the quality of the segmented image.

The presented results emphasize the interest of the proposed method for a low-level processing of large metric data-sets. The generated partition can be further used as input for image analysis methods which are described in the next chapter.
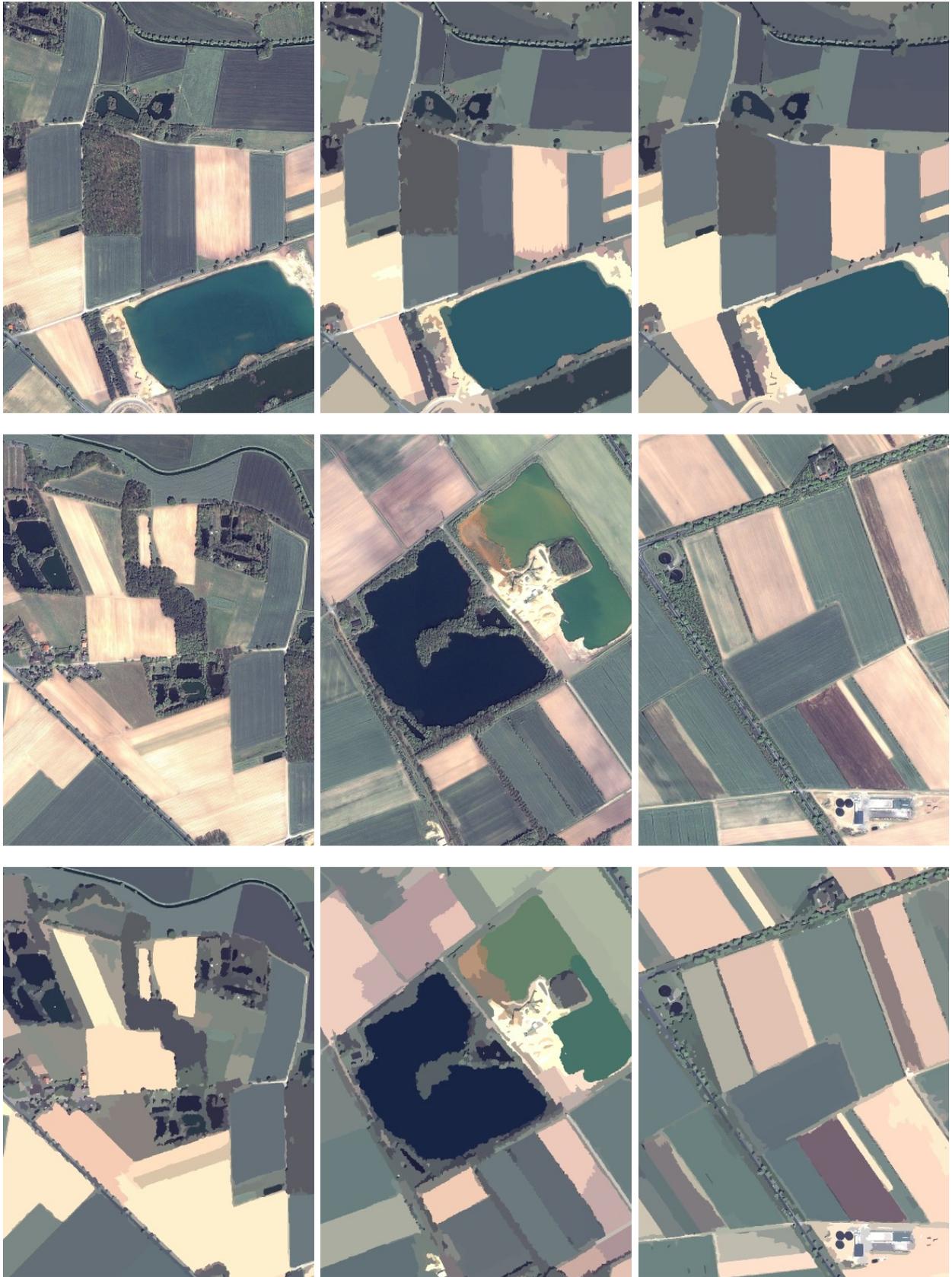
Figure 5.8: Ikonos segmentations over landscape area (size 768×1024). Top: sequence $\overline{|R_i|} = 225$ - $\overline{|R_i|} = 625$. Bottom: $\overline{|R_i|} = 400$.
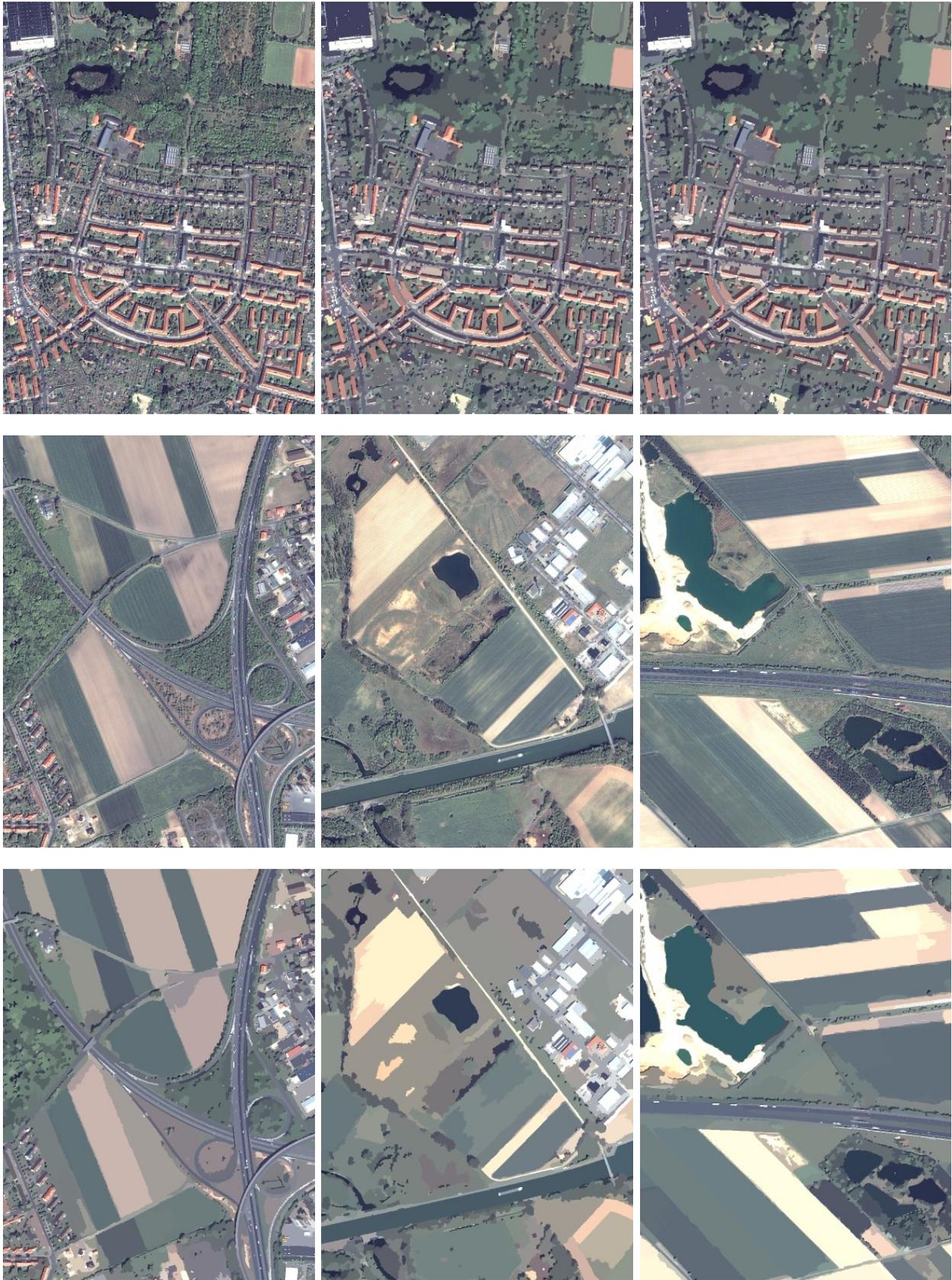
Figure 5.9: Ikonos segmentations over urban and man-made area (size 768×1024). Top: sequence $\overline{|R_i|} = 225$ - $\overline{|R_i|} = 625$. Bottom: $\overline{|R_i|} = 225$.
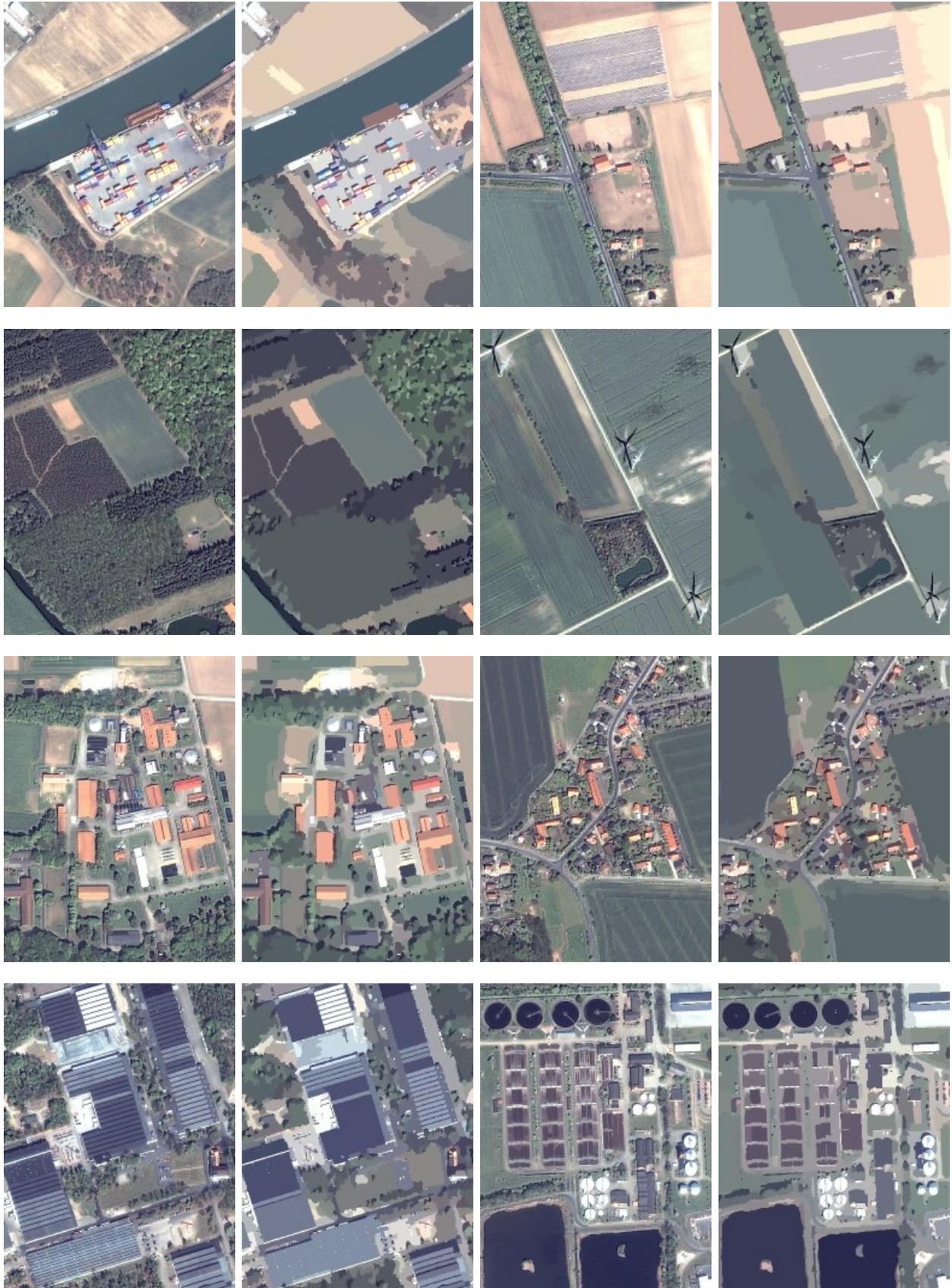
Figure 5.10: Ikonos segmentations (size 384×512), some localized details extracted from segmentation (b) $\overline{|R_i|} = 225$

Image Topology and Image Analysis

Abstract

According to the principle stated by Marr [146] to subdivide a computer vision system into operations with an increasing level of descriptions, a stage is investigated to model the regions resulting from a low-level image partition. The pixel-wise region definition is replaced by a set of primitives, relations and objects, allowing to express additional properties and characteristics. Such complementary information - hardly retrievable based on pixel representation - is generated using geometrical modeling and topological analyses.

A dynamic algorithm is proposed to retrieve, extract and encode the segmented regions in a tree structure which describes their topological relations (adjacency, inclusion). The overall collected information constitutes a consistent and independent hierarchical database which enables to link the extracted objects to semantic information.

In addition, the interest of the proposed data structure is demonstrated by applying efficiently further image analysis processes. In particular, (un-)supervised classifications are presented to derive spectral similarities. In this chapter, we refer to several concepts presented in sections 2.1.3, 2.1.4 and 4.1.1.
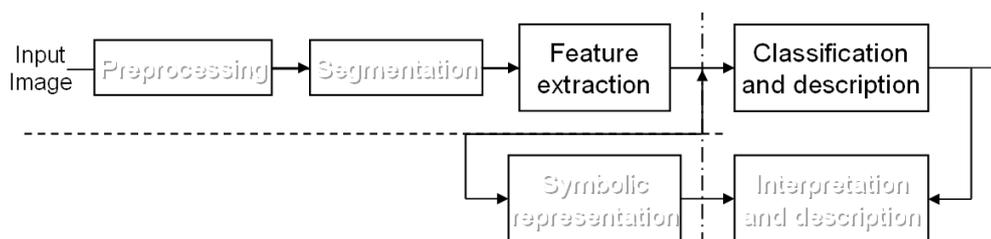
6.1 Information Representation

Dimension reduction methods are an essential issue to cope with an important amount of information like images. Indeed, the principle itself of capturing a scene with a sensor and then try to get back to the imaged "reality" hardly avoids such a stage. The segmentation algorithm presented in the previous chapter[1] leads to efficient dimensionality reduction while preserving most of the meaningful information.

---

[1] A large scope of methods are useful to reduce the dimensionality of an image by investigating one or several "axes" ( ) of data content. Such techniques are not restricted to image segmentation (locally). Interesting results are derived from PCA (largest variance) or classification (spectral) methods.

The pixel-wise labeling of the region partition allows a more compact image description[2]. Instead of dealing with an important amount of data - regularly spaced on a grid - more abstract elements, namely primitives and relations, can be used to more properly express shape, features and regions. The jump from pixel to object representation presents numerous advantages, such as more compact description, generation of additional information, ability to describe elaborated data structures (as it is the case with content embedded in high resolution EO images).

In addition, neighboring relationships are not restricted to a local set of pixels anymore, but to objects themselves, providing a better understanding and modeling of the captured scene. A hierarchical organization of the information improves data access and increases the computational efficiency of machine learning processes.

The duality between regions and boundaries makes relevant the description of a segmented image by its set of boundaries. Two main discrete approaches have been reported in the literature; (1) marking the boundary on the pixels [66], (2) expressing the boundary between the pixels [22]. The former generally follows edge-based segmentation processes[3] while the latter, based on "inter-pixel" boundary modeling, is adopted in most of the region-based segmentation approaches for its non ambiguous discrete representation.

6.1.1   Graph-based Approach and Connectivity Paradox

Topology was introduced in the late 1960's for the analysis of digital pictures [176] and emerged as an established field called "digital topology" [177]. Objective is mainly the determination of topological properties - such as "connectedness" - of objects (and their complements) within an image. A state of the art in digital topology is given in [116].

Rosenfeld in [176], [178] suggests to consider a digital image as a graph, called the Region Adjacency Graph, whose nodes are discrete points (pixels in 2D, voxels in 3D) and arcs being the neighboring relationships between pixels. On a regular 2D lattice, 4-, 8- connectivity systems are generally employed, it results in the so-called 4- and 8- neighborhood graphs (see Fig.6.1). Such formulations are, among all the approaches to model boundaries, the most widely used.

By applying such a neighborhood analysis on an edge map using some contour following techniques [166] (presented page 101), chains of connected pixels are generated, describing edges. The extracted chain is oriented and can be used to represent a path between the two extremities of the boundary. As shown in Fig.6.2(a), a similar protocol is applied as well on image partitions to describe, for a given region, its inside 8-boundary (frontier) which is a closed curve.

---

[2]A segmented Ikonos image with an average region size of 225 pixels (raw image almost 1.4 Gigabytes large) can be saved on a file about 20 Megabytes big using a LZW lossless compression algorithm.

[3]Edge detector filters and techniques such as Canny-Deriche, Sobel, zero-crossings, etc. are discussed page 24, snakes are presented page 63. They can be employed as well to model watershed algorithm results (morphological thinning operation).

(a) Continuous $\mathbb{R}^2$ (b) $\mathbb{Z}^2$, 4- neighborhood graph   (c) $\mathbb{Z}^2$, 8- neighborhood graph
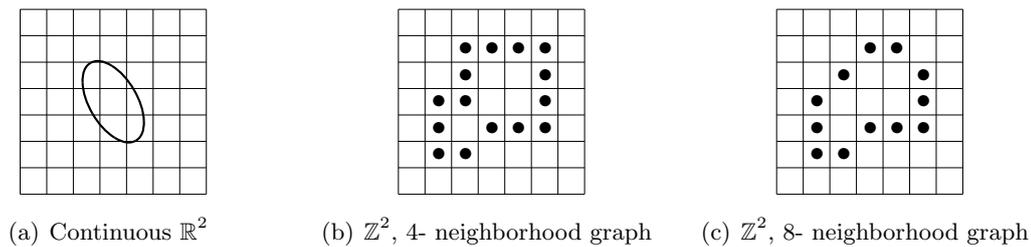
Figure 6.1: Jordan Theorem in the Euclidean plane, continuous and discrete cases.

An essential topological property of closed curves is given by the Jordan theorem. It states that a simple closed curve in an Euclidean plane separates the space into two connected components, the inside and the outside of the curve (see Fig. 6.1(a)). The Jordan theorem can be extended to higher dimensions. However, using graph representation, limitations arise when transposing such a property from n-dimensional space $\mathbb{R}^n$ to $\mathbb{Z}^n$. In addition, if the Jordan theorem is verified using 6-neighborhood graphs (hexagonal sampling), Fig. 6.1 illustrates the resulting connectivity paradox with both 4 and 8 connectivity systems.

The discrete 4-connected closed curve in Fig. 6.1(b) separates the Euclidean 2-dimensional space $\mathbb{Z}^2$ into one outside and two inside components. In addition, by removing a corner of the closed curve, inside and outside components remain disjoint. On the contrary, with 8-neighborhood graph, inside and outside components are not always distinguished (Fig. 6.1(c)). Some further contradictions appear to model boundaries with graphs (Fig. 6.2). Indeed, for a given closed discrete curve, one can define either the inside or outside boundaries, using 4- or 8- neighborhood systems. It results into four combinations, but none of them respects the Jordan theorem.

(a) 8 inside   (b) 4 outside   (c) extended
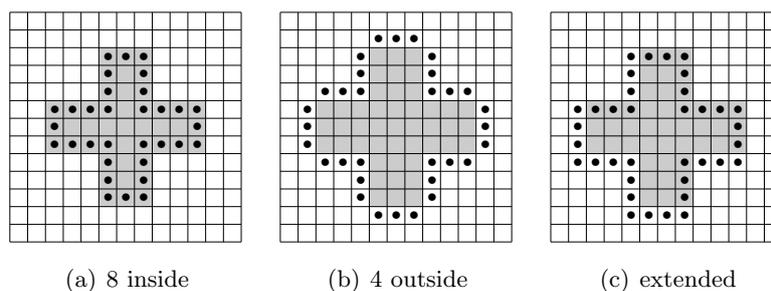
Figure 6.2: Some boundary models.

A solution to overcome such problems is given in [178]. It consists in choosing a different adjacency for the foreground (shape) and the background.

- A simple 4-curve (with more than 4 pixels) separates $\mathbb{Z}^2$ in two 8-components,
- A simple 8-curve separates $\mathbb{Z}^2$ in two 4-components.

An alternative to the multiple boundary definitions (4-/8-, inside/outside) is proposed in [166] and illustrated in Fig. 6.2(c). We remark that the "extended" boundary is equivalent to inter-pixel boundaries in the graph representation shifted from half a pixel in both directions.

6.1.2 Complex-Cell Model and combinatorial Maps

In digital topology, a rather different approach to the Rosenfeld's graph concept is the use of inter-pixels [22]. Such a method, more theoretically formulated by Kovalevsky in [117], is based on the cellular decomposition of the Euclidean n-dimensional space $\mathbb{R}^n$ into a regular grid, the cellular complex, denoted $\mathbb{C}^n$. Some of the essential foundations are presented in this section, the reader is referred to [117], [118] for further explanations.

Definition 1. An abstract cellular complex $C = (E, B, dim)$ is a set E of abstract elements provided with an antisymmetric, irreflexive, and transitive binary relation B, $(B \subset E \times E)$ called the bounding relation (or the face relation) and with a dimension function dim: $E \to I$ from E into the set I of the non-negative integers so that $dim(e') < dim(e'')$ for all pairs $(e', e'') \in B$.
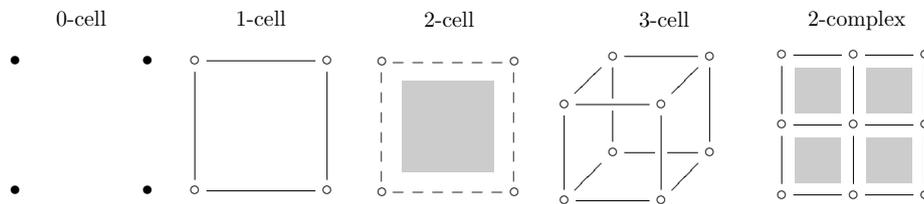
Figure 6.3: Complex cell modeling.

Some examples of elements e' are given in Fig 6.3, characterized by their dimension $dim(e')$=d, (e' is then called a d-dimensional element or a d-cell). The relation B defines a partial order of the elements (e', e'') in E, denoted $e' < e''$. It indicates that e' is a part of the boundary of e''. In an image, the 2-dimensional elements are the area elements, they are related to pixels while the area of 1-dimensional elements is zero. A complex whose all its elements have a dimension inferior or equal to k is called a k-complex. A 2-complex is presented in the right illustration of Fig.6.3.

Definition 2. A sub-complex $C' = (E', B', dim')$ of $C = (E, B, dim)$ is a complex whose set E' is a subset of E, the bounding relation B' is the intersection of B with $B \subset E' \times E'$ and $dim'(e)=dim(e)$, $\forall e \in E'$.

As a consequence, defining a sub-complex C' of C implies only to define the corresponding subset E'. In other words, all sub-complexes of C (E') can be considered as subset of E.

A "topological space" is a pair (E, SY) consisting of a set E of abstract elements and a system $SY = \{S_1, S_2, ...\}$ of "open" subsets S_i of E and satisfies the following axioms:

- The empty set $\emptyset$ and E belong to SY (i.e. are open)

- For every family F of subsets S_i belonging to SY, the union of all subsets which are elements of F must also belong to SY.

- If some subsets S_1 and S_2 belong to SY, the intersection $S_1 \cap S_2$ must also belong to SY (only with a finite number of subsets).

A topological space has the "separation" property when for any two elements $\{e, e'\} \in E^2$, there exists in SY an open subset S' so that exactly one of the elements is in S'. In the cellular complex theory, the notions of "open" and "open star" employed in topology are expressed as follows:

Definition 3. A sub-complex S of C is called "open" in C if for every element e' of S all elements of C which are bounded by e' are also contained in S.

Definition 4. The "open star" of e' in C is the open subset consisting of e' and all elements of C which are bounded by e'.

In a complex, the simplest open subsets are open stars; it results that all the other open subsets are unions of some open stars. Using definition 4, a unique formulation of a boundary is:

Definition 5. The "boundary" of a subset $S \subset C$ relative to C is the subset $Fr(S,C)$ of all elements e' of C so that any open star of e' contains elements both of S and of its complement $C \setminus S$.

Such a definition is quite convenient for modeling regions within the lattice. Its symmetry property (one can easily swap S and its complement $C \setminus S$) leads to a singular boundary. 4-, 8-neighborhood, inside and outside boundaries - which appear as dilemma in Rosenfeld's approach - are neither considered nor required at all.

A further useful topological property is directly reachable by means of the cell's theory,

Definition 6. Two sub-complexes S_1 and S_2 of a complex C are called "incident" to each other if they do not intersect and there are two elements $e_1 \in S_1$ and $e_2 \in S_2$ so that one of them bounds the other.

It allows to reformulate the classical adjacency definition [178], [166] by

Definition 7. Two i-cells are "adjacent" if they are both incident to a common j-cell with $\dim(j) < \dim(i)$

In the purpose of structure encoding, let's recall the notions of path and connected elements:

Definition 8. A sequence of elements of a subset S of a complex C beginning with e' and finishing at e'' is called a path in S from e' to e'' if, for every two elements which are adjacent in the sequence, one of them is bounding the other.

Definition 9. A subset S is called connected if for any two elements e' and e'' of S there exists a path in S from e' to e''.

In neighborhood graphs, the connectivity paradox results from the fact that 0-elements are not taken into account. Instead of using the geometrical definition of adjacency, no paradox arises under the cellular complex modeling since paths are linked by 0-elements.

As a consequence, within the Kovalevsky's theory, all meaningful topological notions for finite sets can be transferred to complexes. Kovalevsky demonstrates that any cellular complex is a topological space and the only finite topological structure. This consistent and compact description of topological spaces,, enables for a given complex, to define properties and relationships of sub-complexes (such as *intersection, union, complements*).

In the literature, other approaches have been proposed, such as the Khalimsky's topology [109] or combinatorial maps [61], [21]. A Khalimsky grid, defined on $\mathbb{Z}^2$, associates a face with every pixel - see Fig. 6.4 -. Given a lattice of size (w, h), the corresponding grid $(2w + 1, 2h + 1)$ is composed of faces (two even coordinates), nodes (two odd coordinates) and edges (remaining coordinates). The latter bounds two neighboring faces while nodes bound their 8 neighbors (4 faces and 4 edges). The equivalence between the cellular complex topology and the Khalimsky's topology is demonstrated in [109]. However, we do not retain this approach since it becomes inefficient when images get larger.

(a) 2D grid   (b) Kovalevsky   (c) Khalimsky
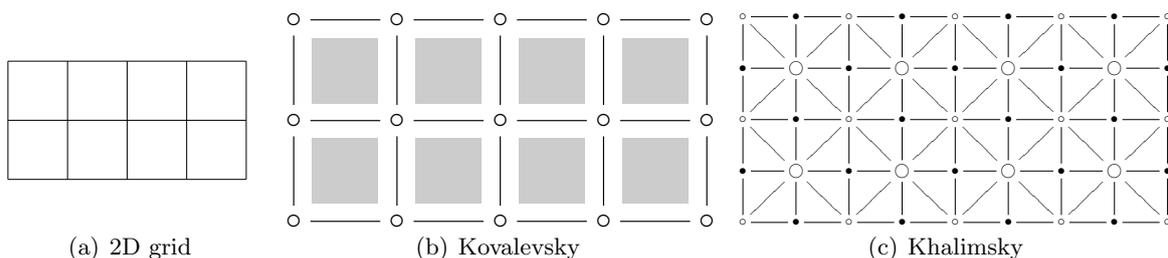
Figure 6.4: Khalimsky's grid (c) and complex cell modeling (b) corresponding to the 2D lattice (a). In (c), ○ denotes the nodes, ● the edges and ◯ the faces. The neighborhood relationships between the heterogeneous elements of the grid are expressed by the arcs — .

Combinatorial maps [55], [43] make explicit some data structures to represent labeled images on $\mathbb{C}^n$. A decisive advantage is that such a map is defined using only one element called "darts" and some basics operators denoted β_i, the relation between two i-dimensional elements (oriented or not).

Definition 10. A nD combinatorial map is a (n+1)-tuple $M = (D, \beta_1, \beta_2, ...\beta_n)$ where

    (1) D is a finite set of darts

    (2) β_1 is a permutation on D

    (3) β_i is an involution on D, $\forall i \in \{2, .., n\}$

    (4) $\beta_i \circ \beta_j$ is an involution, noted β_{ij}, $\forall i \in \{1, .., n\}$, $\forall j \in \{2, .., n\}$

(a) Simplified map   (b) with β_1, β_2 relations   (c) β_1 and β_2 relations
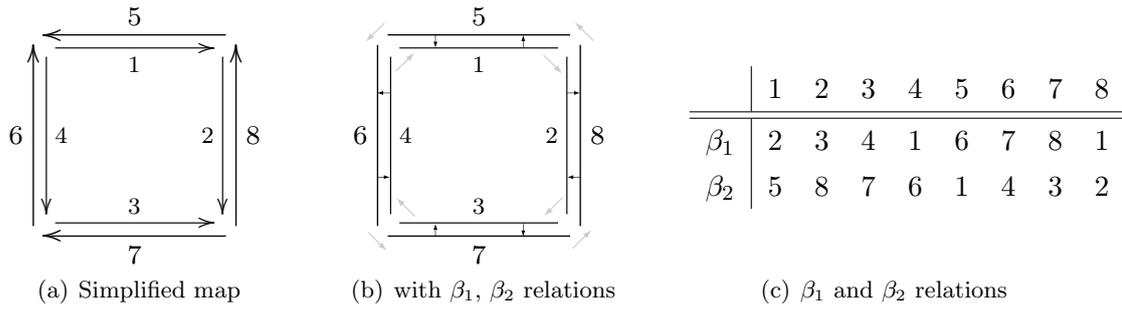
Figure 6.5: Example of Combinatorial Maps [43]: (a) implicit representation, the numbered oriented segments denote the darts, (b) full representation. Black arrows depict β_2 relations between 2-sewn darts while gray arrows express the β_1 relations between 1-sewn darts. (c) makes explicit the corresponding β_1 and β_2 relations.

In definition 10, a permutation on a set S is a one-to-one mapping from S to S. An involution f is a permutation so that $f = f^{-1}$. When two darts are linked by β_i, they are said to be β_i-*sewed*. An edge is connected to its following edge by the relation β_1 (in the same face). Two faces incident to the same edge are connected by the relation β_2. It results that every space cell is represented by a set of darts using the notion of "orbits".

Definition 11. Let $\Phi = \{f_1, ..., f_k\}$ be some permutations on D and $< \Phi >$ the permutation group generated by Φ. $< \Phi >$ is the set of permutations obtained by any composition and inversion of permutations contained in Φ. The orbit of a dart d relatively to Φ is defined by

$$< \Phi > (d) \; = \; \{ \; \phi(d) \mid \phi \in \Phi \}$$

Given a 2-map and a dart d, using particular orbits, all the cells incident to d are retrieved; the vertex incident to d is defined by $< \beta_{21} > (d)$, the edge by $< \beta_2 > (d)$ and the face by $< \beta_1 > (d)$.

In opposition to a classical grid of pixels, whose topology is only implicit and carried by a pure geometrical definition of adjacency, such structures make explicit the topology. Combinatorial maps describe only object topology. Geometry can be inserted to the model by providing some geometrical description of the orbits.

The separation of geometry and topology offers numerous advantages, as for instance the possibility to distinguish topological operations from geometrical ones. A merging operation applied on two connected regions is a topological operation, the boundary geometry is not affected. An example is the segmentation algorithm presented in chapter 5 based only on region modeling and the sorting of all the pairs of adjacent regions.

6.2 Information Modeling

Region-based segmentation methods generate a partition $\Omega_{w,h}$ of the image into disjoint regions expressed on a pixel level (w, h denote image size). If the region description as a set of features and labeled pixels is efficient for low-level processing, such a basic modeling is rapidly defeated with higher-level operations/requests occurring in image interpretation processes. A more consistent region modeling can be achieved by means of geometrical and topological description. Indeed, such a type of information is already accessible in numerous segmentation schemes. However, to keep the segmentation as a low-level process and be independent on the method and its implementation, an algorithm to generate an object description of a region-based segmentation is presented and applied sequentially once the segmentation process is finished.

Among the various discrete spatial models [116], [117] presented in the last section, the complex cellular approach [117] is particularly well suited since it turns out the connectivity paradox and topological problems with $\mathbb{Z}^2$ (i.e. with discrete grid). The plane ($\mathbb{R}^2$) is decomposed in regular complex cells which replace the pixel model ($\mathbb{Z}^2$). A discrete model is generated that preserves the topology of $\mathbb{R}^2$ in a similar way as manipulating vector models.

| (a) Primitives | (b) Frontier of R_i | (c) Summary |
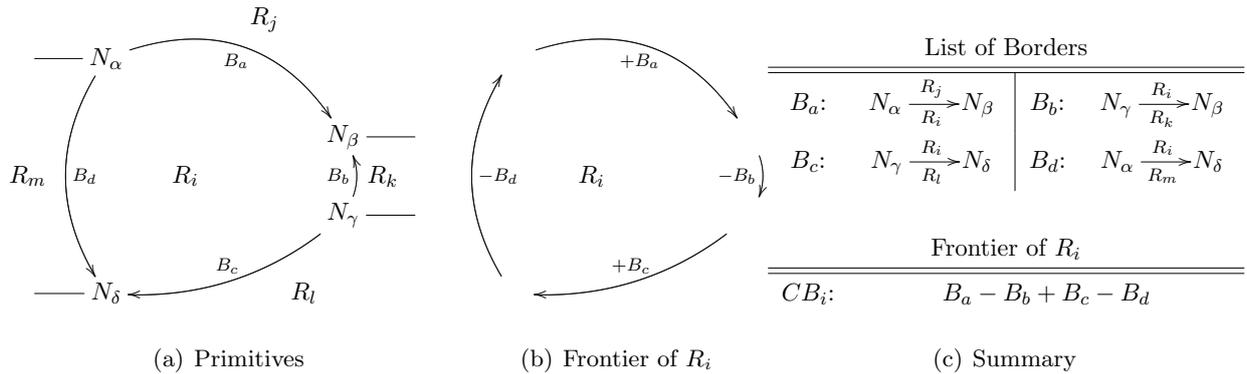
Figure 6.6: Topological Region Modeling.

Based on this modeling, geometry and adjacency relations of the regions can be expressed. Fig. 6.6 sketches a rough illustration of the primitives involved to model regions. In opposition to the segmentation algorithm described in chapter 5 where the modeling is based on explicit description of 2-cells (pixels representation through the image labeling), the proposed method makes implicit the 2-cells, the resulting regions being modeled only by set of 0- and 1- cells elements.

We propose to complete the object representation by investigating the inclusion relationships between regions. In this purpose, a tree structure is introduced to store the regions while reflecting their topological properties.

6.2.1 Topological Description

Using the complex cellular modeling [117], $\Omega_{w,h}$ is expressed by an inter-pixel representation $\Omega'_{w+1,h+1}$ composed of Nodes N ($0-cells$), Boundaries B (set of connected $1-cells$) and Regions R ($2-cells$). A cell representation is provided in Fig. 6.3. Using the extracted set of nodes and boundaries, borders CB are constructed (closed chain of connected boundaries) to model the regions - see Fig.6.6-. The proposed system is built according to relations 6.1- 6.10 with $\{i, j, k, l\} \in \mathbb{Z}^4$.

Node (0-cell)

From the regular set of 0-cells ($N \in \mathbb{Z}^2$) which covers the lattice $\Omega'_{w+1,h+1}$, only the ones incident to at least two regions are investigated. The corresponding subset can be as well decomposed into two parts, according to their relationships with the 1-cells. The first part concerns 0-cells used to connect 1-cells which are incident to the same pair of regions and form a path between two regions (the boundary). This set predominates when the partition is not excessively dense. 0-cells belonging to the second part are called "Node".

Under the inter-pixel representation, a node N (see Eq. 6.1) is the extremity of at least one boundary (closed boundary) or maximum four of them

$$B_i \bigcap B_j = \left\{ \begin{array}{c} \emptyset \\ \{N_k\} || \{N_k, \ N_l\}, k \neq l \end{array} \right. \quad , \text{ with } \quad i \neq j \tag{6.1}$$

Such a definition of a node can be interpreted as a 0-cell which is incident to at least three regions (rank ≥ 3). An exception is presented in Fig. 6.7(a) where the region R_2 is included in R_1; its frontier, composed of 1-cells connected by 0-cell, should not contain any node. To maintain the formalism valid, a single node is added to properly bound its unique boundary.

Boundary (1-cells)

B delimits two adjacent regions R_{left}, R_{right} and has two nodes N_{begin}, N_{end}. A boundary is composed of a sequence of 1-cells connected by 0-cells, forming a path from N_{begin} to N_{end}. The orientation of the boundary $N_{begin} \xrightarrow[R_{right}]{R_{left}} N_{end}$ is necessary to ensure the uniqueness of the boundary.

Indeed, it exists several paths incident to two regions R_i and R_j between two nodes N_α and N_β. An example of multiple paths is given in Fig.6.7(a) with the region R_4 that shares nodes with regions R_3 and R_5. The two boundaries (b_4, b_5) of R_4 present the same extremities (N_3, N_4) and are incident to the same regions (R_1 and obviously R_4). The maximal number of multiple paths is two, since the intersection between two boundaries is restricted to $\emptyset$ or their extremity (node(s)), as stated in relation 6.1.

Border or Frontier (1-cells)

A border CB_i is a closed chain of connected boundaries. The chain is clockwise oriented: for each boundary B_j which belongs to CB_i a direction δ_j is associated (Eq. 6.2) using Eq. 6.7.

$$CB_i = \sum_j \delta_j \; B_{N_k \xrightarrow{R_j/R_i} N_l} \qquad \delta_i = \begin{cases} 1 & \text{if} \quad R_i = R_{right} \quad \Leftrightarrow \quad N_l^j = N_k^{j+1} \\ -1 & \text{if} \quad R_i = R_{left} \quad \Leftrightarrow \quad N_l^j = N_l^{j+1} \end{cases} \qquad (6.2)$$

The closed chain CB_i (Eq. 6.2, Eq. 6.3) delimits the region R_i from its external neighboring regions as shown in Fig.6.7(a). The entire area within CB_i (in direction of R_{right}) is called the interior region R'_i.

$$CB_i \bigcap CB_j = \begin{cases} \{B\} & \Leftrightarrow R_i \text{ and } R_j \text{ are adjacent} \\ N || \emptyset \end{cases} \qquad (6.3)$$

The border CB_i can be interpreted as the external frontier of the corresponding region R_i, separating R_i from the other regions (Jordan curve). Since the border is composed of boundaries, the intersection of borders is also restricted to nodes or $\emptyset$. Due to the inter-pixel representatin, the 4-connectivity neighborhood system is fixed. It maintains disjoint regions only connected by 0-cells, independently of the spectral signature of those regions. An example is given in Fig.6.7 with the disjoint regions R_3, R_4 and R_5.

Region

A region R_i is defined by one external border CB_i and a collection of internal disjointed borders $\{CBj\}$ (Eq. 6.4): CB_i encloses R_i while $\{CB_j\}$ excludes "holes" inside R'_i.

$$R_j \subset R_i \,, \; R_i = CB_i \bigoplus \sum_j CB_j \quad \Leftrightarrow \quad R_i = R'_i \setminus \cup_j R'_j \qquad (6.4)$$

Such a modeling enables to investigate region inclusion relationships and to define a region by means of its frontier CB_i and the ones of its included regions $\{R_j\}$. To circumvent the case where a group of adjacent regions $\{R_a\}$ constitutes a compact hole, a new frontier is created using the set of boundaries in order to represent the external frontier of $\{R_a\}$. Its corresponding region is a dummy region only composed of the holes $\{R_a\}$. It enables to maintain the area size consistency, its area contribution being neglected. An example of creation of such a region and its frontier is given in Fig.6.7(b) (R', $CB_{R'}$) which encapsulates regions R_6, R_7 and R_8.

The inclusion relationship between regions can be retrieved as well by means of the intersection of their interior regions

$$R'_j \bigcap R'_i = R'_j \quad \Leftrightarrow \quad R_j \subset R_i \qquad (6.5)$$
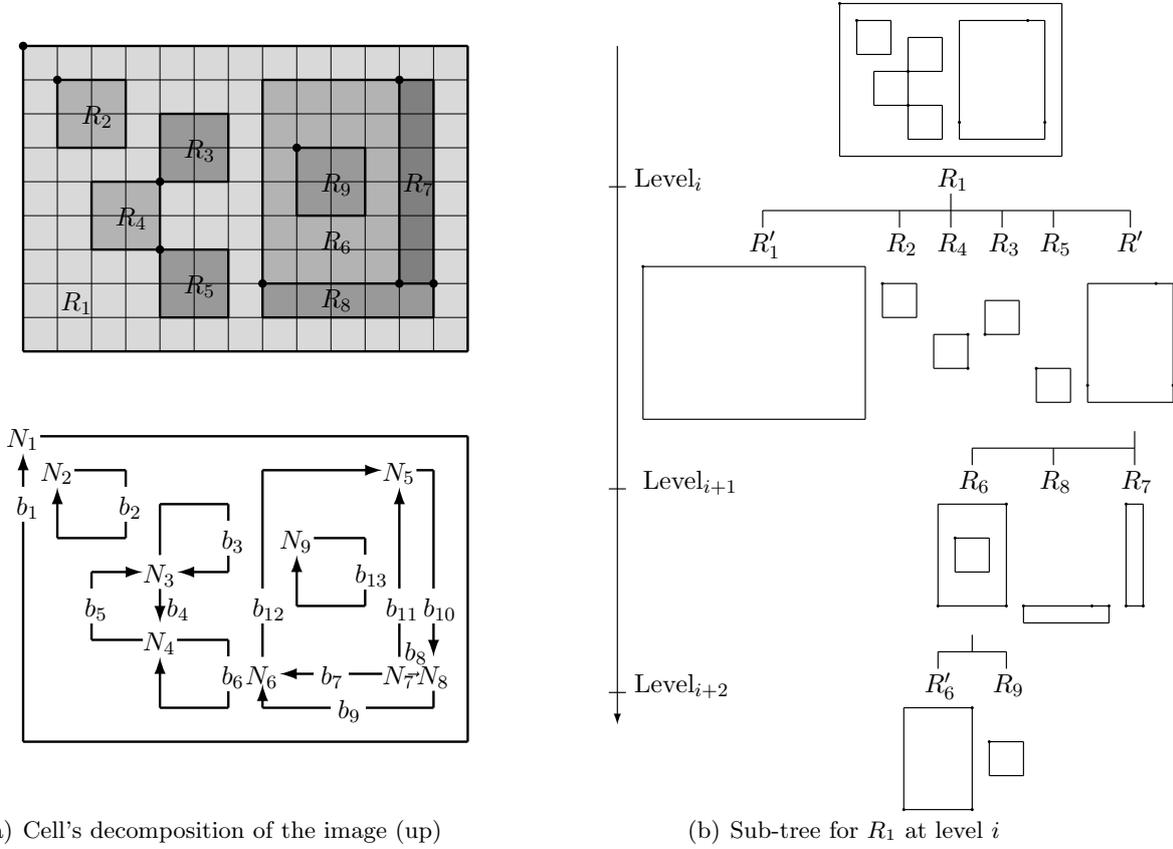
(a) Cell's decomposition of the image (up)

(b) Sub-tree for R_1 at level i

(c) List of Boundaries

(d) List of Borders (External Frontier of R_i)

Figure 6.7: Example of Topological analysis on a segmented image (a) using cellular complex modeling [117]. Representation of the region decomposition into 0-, 1- cells. (b) depicts the sub-tree with the root R_1. In the tree, the interior region R_1' of R_1 is placed as first child of R_1. The region R' denotes the union of regions R_6, R_7 and R_8. The external frontier of R' is composed of the boundaries of $R_{6...8}$ having R_1 as incident region. The symbol $\bullet$ expresses the nodes N_j between the connected boundaries. (c)(d) dress a summary of the extracted frontiers and boundaries .

6.2.2   Geometrical Description

The model proposed is sufficient for describing in a consistent framework regions resulting from an image partitioning. Since some metrics can be incorporated and the regions are spatially localized, it is useful to integrate geometrical information to the model. Such complementary features are linked to the model by means of the 0- and 1- cells. Nodes encapsulate some absolute location ((x, y) coordinates in the inter-pixel map) while boundary shapes are described in storing their relative travel path with a Freeman code (see page 14) by means of a kernel function depicted in Fig.6.8.

(a) Freeman code [67]   (b) Trace R_i, $i \neq l$, $i \neq j$ (c) Trace R_i, $i \neq l$, $i \neq k$ (d) Trace R_i, $i \neq l$
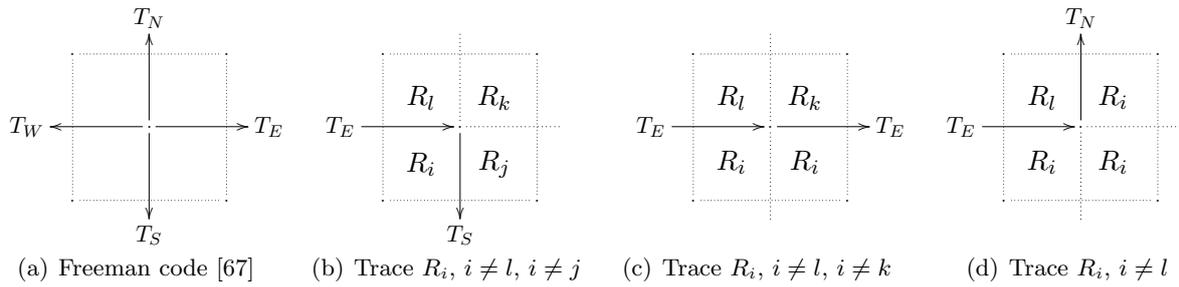
Figure 6.8: Elementary trace function *findPath()* on inter-pixel representation (4-connectivity). Given an initial direction T_E, three directions are possible to encode the frontier of R_i. According to the cellular complex modeling, the initial direction T_E leads to $i \neq l$. In (b), the curve "frontier of R_i" separates the plane into two connected components R_i and $\Omega \setminus R_i$(Jordan theorem); in the case $k = i$, it results $j \neq l$ (the two 2-cells of R_i are not connected).

A Freeman code (Eq. 6.6) describes the boundary shape from N_{begin} to N_{end} using l pairs $\{L, T\}$, where L denotes the number of 1-cells in a given tracking direction T.

$$
N_{\text{end}} = N_{\text{begin}} + \sum_{k=1}^{l} L_k \ T_k
$$

$$
\text{, with} \quad T \begin{pmatrix} x \\ y \end{pmatrix} = \left\{ \begin{array}{cccc} 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 \end{array} \right\} \quad (\{T_E = 0; T_N = 2; T_W = 4; T_S = 6; \})
$$

$$(6.6)$$

Several geometrical features are derived from the chain code; The area $|B|$ below the boundary is computed as the sum of the consecutive trapeziums between the boundary and the abscissa axis x in $\Omega'_{w+1,h+1}$,

$$
|B| = \frac{1}{2} \sum_{k=1}^{l} (y_k + y_{k+1}) \ L_k \ T_{k_x}
$$

$$
\text{, with} \quad \begin{cases} y_1 & = & y_{\text{begin}} \\ y_{k+1} & = & y_k + L_k \ T_{k_y} \\ y_{l+1} & = & y_{\text{end}} \end{cases}
$$

$$(6.7)$$

Indeed, the notion of area is only related to i-cells with $i > 1$. Introducing an area underneath a boundary presents numerous advantages to improve computational efficiency. A simple way to

control the boundary uniqueness is ensured by the area consistency; if it exists several boundaries between two regions and two nodes, they are characterized by different areas[4]. Since the boundary is oriented, its area obviously follows

$$|B| + |B'| = 0 \quad , \text{ with B defined as } \quad N_k \xrightarrow[R_i]{R_j} N_l \quad \text{ and B' as } \quad N_k \xleftarrow[R_i]{R_j} N_l \qquad (6.8)$$

Since the chain is oriented in the clockwise direction, the area $|CB_i|$ below the border is strictly positive and computed according to

$$|CB_i| = \sum_j \delta_j \left| B_{N_k \xrightarrow{R_j/R_i} N_l} \right| \qquad (6.9)$$

It characterizes the area of R'_i. In comparing areas of R'_i with the one of R_i, one easily retrieves an indication of the presence of possible regions included in R_i.

The resulting algorithm dynamically extracts and stores the various objects ($0 - 1cells$) in a database $\mathcal{D}$ containing both geometry and adjacency relations. Region representation is achieved by collecting elements from $\mathcal{D}$ and completed by investigating object inclusion relationships. With this in mind, a tree structure is developed to store the regions and expresses their topological properties.

6.2.3 Tree of Regions

A root tree structure is designed to encode in a simple and "natural" way the regions according to their topology, but also to speed up data access. Nodes of the tree are regions and "branches" reflect inclusion and adjacency relationships between the regions.

Properties

The root denotes the border of the processed image. It can be seen as an empty region that covers the whole image where all the regions are included. In the tree construction, the "directed" notion is added. In a rooted directed tree, from a given region stored in the tree, there exists only one path to the root. Given this directed notion to a rooted tree, a rooted sub-tree can be defined for each region of the tree. As a consequence, for each region which contains included regions, a sub-tree is recursively generated as depicted in Fig. 6.7(b).

$$\left. \begin{array}{rcl} R_j & \subset & R_i \\ \text{depth}_j & = & \text{depth}_i + 1 \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} R_j \text{ is a child of } R_i \\ R_i \text{ is the parent of } R_j \end{array} \right. \qquad (6.10)$$

In the tree structure, the level order indicates the level of appearance (depth), as shown in Fig. 6.11. It corresponds to the vertical distance from the root (depth=0). For a given sub-tree of

---

[4]An alternative to the area test for controlling the boundary uniqueness is to check either the first or last orientation direction of the boundary with the one of the compared boundary. If they are identical, the two boundaries represent the same object.

the region R_i (depth$= k$), the direct children (depth$= k+1$) are siblings and have a unique parent R_i (see Eq. 6.10). In this case, the interior region R'_i is placed as first child of R_i as shown in Fig. 6.7(b) (regions R'_1, R'_6).

Indeed, the interior region can be implicitly retrieved, such a region being only defined by means of its external frontier. In [140], [141], we decide to make it explicit since it gathers some complementary information. Another motivation concerning the interior region is related to the image formation model of Caselles [25] within the Gestalt theory; "occlusion" and "transparency" constitute basic elementary operations (e.g. inclusion can be interpreted as an occlusion while shadows express some transparency phenomenon).

Data access and modifications

The tree directly reflects region relationships. Given a region, adjacent regions (horizontal navigation), the parent region, and possible included regions (i.e. vertical navigation in the tree) can be directly accessed. Once the tree structure is completed, further complementary user-editing and modifications are possible. Based on the tree structure and the database $\mathcal{D}$, a region merging process can be achieved to complete or refine the segmentation process.

The pyramidal decomposition of the image partition into the tree significantly reduces the number of combinations to be solved for updating the tree. Indeed, depending on their relative position within the tree, adjacent regions are subdivided into two types of relations; (1) inclusion relation, (2) neighbored regions with the same parent. As a consequence, the merging operation slightly differs and follows;

- **ancestor-descendant regions**: $\updownarrow$ tree's relation

Merging R_i, R_j with $R_j \subset R_i$ ($CB_i \cap CB_j \neq \emptyset$): If the sub-tree of R_j exists, its level order (depth) is decremented. When R_j has adjacent regions, enclosed by the frontier CB', three cases are considered. (1) The adjacent regions of R_j are becoming disconnected, so there is no need to build interior frontiers and CB' is removed. (2) Regions enclosed by CB' still form one group of regions, then only an update of CB' is required. (3) The group of regions is split; then CB' is removed and new interior frontiers can be created to update the tree structure.

- **sibling regions**: $\leftrightarrow$ tree's relation

Merging neighbored regions R_i, R_j ($R'_i \cap R'_j = \{B\}$) into a new region R_{ij}: since no additional information is needed, CB_{ij} is easily computed and used to build the new interior region R'_{ij}. The spatial separation induced by the tree structure implies two behaviors: (1) Except interior regions R'_i and R'_j which are removed, the respective sub-trees are kept. (2) Adjacent regions to R_i or R_j can move vertically in the tree (to become included in R_{ij}). In dropping, such a group of regions may carry with them regions disconnected from R_i and R_j. A trivial example is the case of a peninsula formed by some districts. If the peninsula is becoming an island, all the districts are included in the sea, even those which are not on the coast.

6.2.4   Flowchart of the Method

There is a clear discrepancy between the purpose of contour following based on Canny's edges and the extraction of inter-pixel boundaries of an image partition into regions. Differences are multiple; boundary based segmentation approaches result in some disconnected edges marked on the pixel using a gradient computation followed by a non maxima suppression along the steepest local slope. In general, further post-processing operations are required (e.g. edge thinning, handling of T junctions, etc. [166]) prior to the generation of a neighborhood graph.

From graph to complex cell's modeling

Critical issues appear during the contour following stage for the neighborhood graph elaboration. Indeed, such an operation rapidly becomes complex and heuristic, even when applied on a partition of the image into disjoint regions. Contour following approaches are generally based on a Freeman code. The boundary is extracted iteratively in checking, according to the current direction, the next possible direction in (anti-)clockwise.

Such a local scanning presents different behaviors according to the orientation adopted. The exterior frontier of a region is extracted in the clockwise direction while the interiors are encoded using the reverse direction. In presence of intersections or T junctions, contour following approaches turn into combinatorial problems since ambiguities arise and some pixel location requires to be visited more than once.

Within the cellular complex modeling, which implicitly induces a 4-connectivity neighborhood analysis, a proper topological framework for the description of spatially discrete data is constructed. The proposed algorithm is devoted to the description of region-based segmented images and eludes all the critical issues encountered with neighborhood graphs.

The dynamical process [140], [141] is decomposed into three sequential steps: (1) pre-processing to generate temporary image maps used for the (2) extraction and the encoding into a database $\mathcal{D}$ of the 0- 1-cells elements forming closed curves. Finally, using the primitives achieved in $\mathcal{D}$, the regions are edited and recursively stored in the topological tree (3).

Step 1: pre-processing, generation of labeled and inter-pixel images

A pre-processing stage is first applied, it results in the generation of two temporary images and a description from the spectral content for every region. Such results encapsulate sufficient information to simulate again the input image. The image $I_{w,h}$ is labeled using the algorithm 4 proposed in the last chapter as first step for an image segmentation process (description page 72 and pseudo code in Annex, page 7).

$$Node_{Init} \longrightarrow Node_{Begin} \longrightarrow \boxed{\begin{array}{ll} \text{Geometry:} & N_{begin} \longrightarrow N_{end} \\ \text{Topology:} & (R_i, R_j) \end{array}} \longrightarrow Node_{End} \longrightarrow$$

$$\text{Boundary: } (R_i, R_j, N_{begin}, N_{end}), i \neq j$$

$$Node_{Init} \neq Node_{End}$$

$$CB_i = \sum_j \delta_j \; B_{N_k \xrightarrow{R_j/R_i} N_l}$$
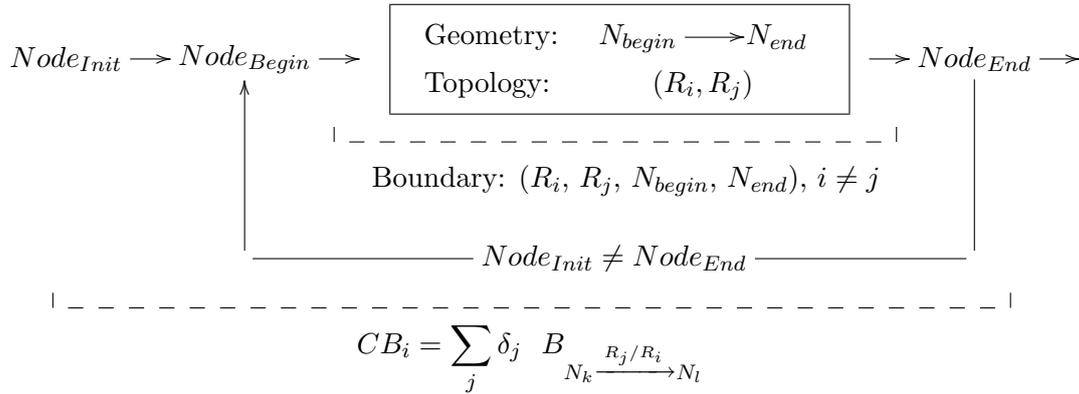
Figure 6.9: Trace function *buildBorder()* for the generation of the external frontier CB_i associated to the region R_i. The internal block *buildBoundary()* denotes a recursive call of the contour following function *findPath()* (depicted in Fig. 6.8) along the frontier of R_i until the tracked pair of incident regions changes. The generated path from N_{Begin} to N_{end} describes a boundary incident to R_i and R_j. Having at disposal a starting node N_{Init}, the boundaries which compose CB_i are extracted. The process is terminated when N_{Init} is reached again.

Inter-pixel Map

Based on the labeled image $\Omega_{w,h}$, an inter-pixel image $\Omega'_{w+1,h+1}$ whose pixels denote 0-cells is created. It consists in a skeleton which contains information related to the possible presence of 0- and 1-cells. The lattice Ω' is filled with pre-codes describing the comparison of the content for the four 2-cells incident to each 0-cell.

To improve computational efficiency of the primitives extraction (boundaries and frontiers), the number n of incident regions ($3 \leq n \leq 4$) to each node is encoded as well in Ω'. The resulted inter-pixel map encapsulates a description and a location of the paths and their junctions. Some examples of labeled images and inter-pixel maps are depicted in Fig. 6.7(a) and Fig. 6.10.

Processed image type

At this stage, except the spectral information of the segmented image, all the computed data are at the pixel level. In the present work-flow, if we use as example a region described into piecewise constant functions, classified images are processed similarly. The thematic map acts as input image while the clusters signatures are further used to assign the corresponding spectral information to every group of connected pixels presenting identical clusters (regions). In the case of segmentation with non-linear or local region modeling, a labeled image is provided to explicitly express the region's coverage, and the region content description is maintained.

Such a method can also be used on segmented images with partially complete partitions. It remains only to define a label for areas ignored, the corresponding regions placed in the tree acting as an "empty" region. When such empty areas are located on the image's border (e.g. georeferenced or rotated image) and split on the image corners, a specific process is designed to obtain the convex

hull of the image content.

Step 2: Boundaries extraction and creation of the frontiers

The objective of this stage is to collect the primitives required for an inter-pixel boundary description of regions composing an image partition (see Fig.6.6) by means of several encapsulated operations applied on Ω and Ω'. Such operations, satisfying relations 6.1 - 6.9, enable to retrieve conjointly geometrical features (using a Freeman code) and topological relations of the extracted primitives, composed by sets of nodes $< N >$, boundaries $< B >$ and closed borders $< CB >$ stored in the database D.

Knowing the amount of regions (n) inside the partition $I = \bigcup_{i=1}^{n} R_i$, a list of external frontiers $< CB > = \{CB_1, ..., CB_n\}$ is initialized and dynamically generated by means of a trace function *buildBorder()* applied for every labeled region R_i as illustrated in Fig.6.9. Since a closed border is formed by linked boundaries, representing a path between two incident regions R_i and R_j, *buildBorder()* consists of successive calls to the function *buildBoundary()*. The latter is itself a sequence controlling elementary operations *findPath()* recursively applied on Ω'.

Path extraction

The purpose of *findPath()*, depicted in Fig. 6.8, is to connect the consecutive elements of the tracked path. Indeed, given an initial direction along the frontier of a region, three cases are possible to link a 0-cell to the next one, the reverse direction being forbidden. The direction between two consecutive elements is deduced from the pre-codes of Ω'. When the frontier is tracked clockwise (the studied region being always on the right-hand side of the tracing direction), the possible Freeman code directions are scanned counterclockwise.

Starting from an initial node $N_{init} = N_{begin}$, the kernel function is recursively applied along the frontier of R_i until the tracked pair of incident regions changes (i.e. a node N_{end} is reached). To preserve node uniqueness, N_{end} is appended to the list $< N >$ only when none of the stored nodes presents the same coordinates.

Boundary generation

Similarly, *buildBoundary()* checks in the list $< B >$ whether the generated boundary B_a[5] is already inserted. If a boundary B_b of $< B >$ presents the same pair of regions and nodes, a uniqueness test is made, based on the comparison of the area underneath the boundaries. When relation 6.8 is satisfied, B_a and B_b represent the same boundary, B_b being traced in the reverse direction during the generation of CB_j.

This duality indicates that R_i and R_j are adjacent but not included. In this case, the function *buildBorder()* appends B_b to the list of boundaries composing CB_i with an orientation flag $\delta = -1$.

---

[5]The path from N_{begin} to N_{end} is built with *findPath()*, R_i being the studied region. Based on the path and its incident regions R_i and R_j (deduced from Ω), a boundary is created.

If the path is tracked for the first time, B_a is inserted to the lists $< B >$ and CB_i (with $\delta = 1$).

Then, *buildBoundary()* is applied again with N_{end} as new starting point. The overall procedure is terminated when the last extracted node N_{end} matches N_{init} (i.e. when the curve CB_i is closed). An example of the sequence is given Fig. 6.7, page 97, based on a simplified image partition (a) which produces the lists $< N >$, $< B >$ (c) and $< CB$ (d).

Sequence initialization

For each closed border CB_i to be tracked, *buildBorder()* is initialized using a first call to *buildBoundary()* with a starting node having the coordinates of the first 2-cell of R_i when the index image is scanned line by line from the labeled image origin. Using this orientation convention, a 0-cell -N_{init} - of the frontier presents the same coordinates in Ω' as the first 2-cell in Ω. The corresponding node N_{init} presents a rank $n \geq 2$.

Another important initialization issue is the direction of the border tracking. We predicate that all borders are described clockwise. Such a direction is obtained without ambiguity from the starting point using 0 (T_E) as first Freeman code direction.

If the rank of N_{init} is $n = 2$ (no intersection between boundaries), the function *buildBoundary()* finds, if it exists, the next node with a rank $n \geq 3$. In case the function reaches N_{init} without finding any node, (1) N_{init} is stored in $< N >$, (2) the generated boundary B_c is closed (it presents the same node as extremities), (3) the resulting closed border CB_i is composed of only one boundary inserted into CB_i ($\delta = 1$) and $< B >$. In such cases, R_i is included in its adjacent region.

It should be noticed that the geometry is carried only by the set of nodes and the Freeman code, while the topology is encapsulated within the boundaries. Indeed, both information are extracted by the kernel function *findPath()*. The two other detailed functions mainly control the dynamical extraction process, the construction of the borders, and ensure the database consistency in ingesting or not the proposed extracted objects (Eq. 6.8 - 6.9). At this stage, the set of nodes, boundaries and borders are stored in the database D.

Step 3: Inclusion relationships and tree of regions

If for various purposes, the knowledge of region adjacency is sufficient, it can be relevant to distinguish the ones which are included. However, the inclusion relationship cannot be derived only by checking, among the adjacency relations, the ones which do not belong to the exterior frontier of the studied region (example of the island composed of several districts, page 100). Indeed, the inclusion relationship does not imply an adjacency property. The possible holes within a region must be modeled and result in a "spatial" sorting of the data represented by means of a tree structure.

Interior frontier construction

The extraction of the interior frontier can be achieved exactly with the same protocol as the one previously described for the exterior frontier. The only difference would be that paths of the interior frontier (clockwise orientation) of R_i should be traced by scanning the Freeman direction in the clock-wise turn, with R_i on the "left side" of the tracked path. However, such boundaries are already extracted since they constitute the external frontier of the included regions. Indeed, no new primitives have to be derived from Ω and Ω': the interior frontiers are directly generated from the sets $< N >$ and $< B >$.

Instead of tracing again the interior frontier of the regions whose areas differ from the interior region areas (relation 6.9), the set of boundaries $< B >_i^{Int}$ incident to the studied region R_i which does not belong to CB_i is first extracted from $< B >$. Among $< B >_i^{Int}$, some of the inclusion relationships are already identified in the previous step, it concerns the regions composed only of one exterior boundary.

The remaining boundaries compose the external frontiers $< CB' >_i$ of some groups of included regions, called "virtual" regions. Their generation is stated as a combinatorial problem, which is significantly simplified by the fact that all borders are oriented clockwise. Since all the boundaries of $< B >_i^{Int}$ are already oriented in the same direction, they can be directly linked by their extremities with $N_{end}^n = N_{begin}^{n+1}$, n representing the consecutive elements of the created "virtual" borders.

Inclusion description

The generated interior frontiers bound R_i from its included regions and complete the region model. The set of included regions is computed by propagation of the borders adjacency relations having one or several elements in $< B >_i^{Int}$; this function is called *sameLevel()*. Therefore, for each region, its direct children are retrieved. A tree is built to "unwrap" those inclusion relationships and expresses in a single structure all the regions, ordered according to their relationships.

The tree is generated level by level. The image can be seen as a group of regions included in its complement. A "virtual" frontier $< CB'_0 >$ along the image is created, its corresponding dummy region $< R'_0 >$ is placed on the tree root node. Using *sameLevel()*, the regions extracted are placed into the first level of the tree, having $< R'_0 >$ as parent. The next levels are generated by recursively scanning the list of interior frontiers for each region inserted in the previous level.

The procedure is repeated until no "holes" are found. Since the tree implicitly induces a pyramidal decomposition, the "spatial" sorting of the data allows to build in parallel every sub-tree. As described in Fig.6.7(b), in case of derivation of a new sub-tree (inclusion relationship) for region R_i, the interior region is placed first in the list of the sibling regions, having R_i as unique parent. Such a construction remains unchanged when R_i is a dummy region.

6.2.5 Summary

Experiments have been achieved using first synthetic images. An intuitive "representation" of the hierarchical topological image description is presented in Fig. 6.11, resulting from the analysis of the synthetic image (Fig. 6.10). The third "dimension" describes the inclusion relationship between different images areas, or objects. Such a formalism is very interesting for the analysis of segmented images or to provide an image segmentation framework.

The proposed algorithm efficiently builds tree structures from segmented images without any parameters in 4-neighborhood connectivity. The tree architecture enables the analyzing of tiled images or multiple no-overlapped images.

The topological processing does not require any assumptions of the segmentation algorithm and leads to relevant object and spatial information. The segmentation algorithm can be refined or based on other region-based approaches without changing the data structure obtained by the topological algorithm.

The data representation combined to the tree structure enables many applications such as queries by content, direct object clustering or even, data merging of different data types / resolutions as presented in the next section. Tests have been successfully done on large remote sensing images.

(a) Original image   (b) Inter-pixel image   (c) Image labeling
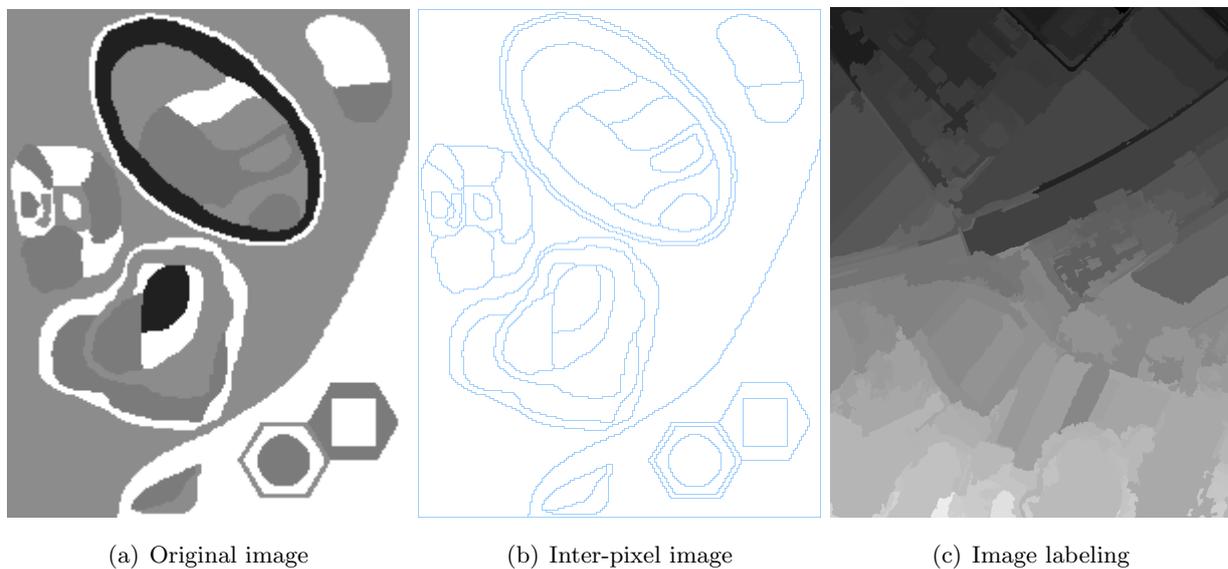
Figure 6.10: A synthetic image is used to generate deep inclusions and various neighboring relationships (size 201×256). "Natural" EO images, (e.g. landscape $SPOT\ 5$) do not present such a "complexity". It is not the case for urban VHR data. (b) is the inter-pixel map (size 202×257). (c) is a labeled image from a segmented satellite image. Bright areas depict large labels, the image labeling starts at the top position from left to right: assigned labels increase in the scan direction (about 580 regions).
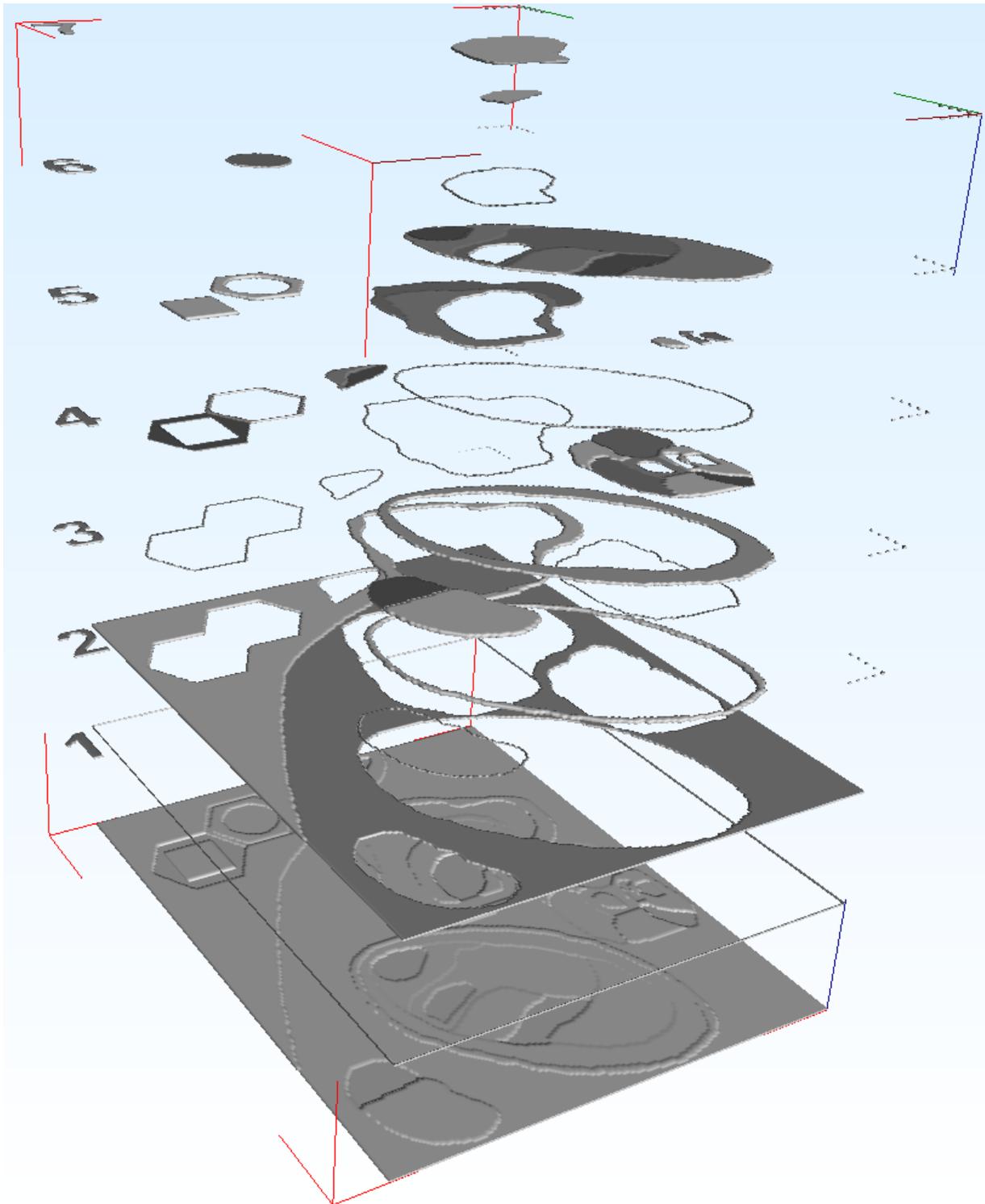
Figure 6.11: Tree representation: The figure depicts the several inclusion levels generated from the topological analysis of an image I_0, presented in Fig. 6.10. Each number denotes a specific order of inclusion, the level $i = 0$ -at the bottom of the figure- is the original image. Surfaces represent regions while contour lines indicate the creation of a "virtual" region through their exterior closed contour representation. Such regions, stored at level $i+1$, are generated to collect several adjacent regions which constitute an included object in the region stored in level i.

6.3   Information Learning

The two main pixel-based classification approaches are supervised and unsupervised methods. The differences are mainly related to the hypotheses and knowledge about the classes to be extracted. Supervised processes learn to assign instances from predefined classes, trained by an operator and iteratively refined while clustering methods do not require any class description, the task being to learn a classification from the data.

The assessment and the evaluation of accurate remotely sensed image classification is presented in numerous papers [199], [135]. They investigate and report some factors which affect classification performance such as: (1) methods and protocols adopted, (2) methodology for the training area definitions, (3) diverse landscape complexities to be classified, (4) selected data used.

Due to their generic properties, classification techniques can be applied directly on raw images or on results obtained from low-level operations (e.g. texture or edge features). The latter is often preferred since applying learning processes to a large amount of data is time-consuming. In this scope the interest for applying classifications to segmented images is determinant. Another reason is related to the segmentation process itself. Results obtained are generally promising; however, a further stage is generally required to properly exploit and link the regions to semantic contents.

6.3.1   Supervised Classification

In the case of supervised classification, samples called training sites of the studied land cover type are identified in the image. The definition of the training areas, characterized by small local extensions considering the full image to be classified, is determinant to ensure satisfying results. For each cover type, a signature based on the respective statistical characterization is derived. The training signature assigned to each class should fully describe its variability within the image: differences in vegetation type, health, soil type or moisture may lead to inaccurate thematic maps. This description enables to decide which cover type presents the most important similarities to the pixel to be sorted.

The decision is generally based on parallelepiped or maximum likelihood criteria. The concept of the first one is easily interpreted by the definition of a bounded box in the feature space centered on the signature class. If the pixel to be classified falls inside the bounded box, it is assigned to the class. Despite its velocity, overlapping between signature classes and pixels in the gaps between parallelepipeds limits such a method that generally leads to poor accuracy results.

Improved results are expected from Maximum Likelihood pixel-based classifiers. Aim is to classify unknown pixel-based classes of interest on multivariate probability density functions (pdf). The class's pdf is estimated using statistical properties derived from the training areas. Significant improvements are achieved by ingesting prior probabilities (simple case is equally Gaussian distribution) and assumptions between the different land covers.

Unknown pixels are assigned to the class which presents the highest probability; overlapping signature problems can be partially avoided. Results accuracy is strongly dependent on the signature class modeling and for both methods on the training set defined. The reader is referred to [182] for an interesting presentation of *Naive Bayes* techniques in the scope of probabilistic information content retrieval in large RS image archives. For *Neural Networks* methods, we refer to [54].

6.3.2 Support Vector Machines

In the literature, many authors reported that classification based on Support Vector Machines (SVM) outperforms classical supervised methods in numerous fields such as biomedical imagery, pattern recognition or remote sensing. However, like all other methods, the quality of the results is highly conditioned by the learning phase.

Since the method is fundamentally a binary classification, let's consider a set of data $< x >$ in $\mathbb{R}^m$ to be labeled into two classes represented by $y_+ = 1$ and $y_- = -1$. The notation x_i denotes the ith vector in the training set composed of n samples $\{(x_i, y_i)\}_{i \in 1...n}$, where y_i is the label associated with x_i. Based on the training set, the goal of SVM is to determine a discriminant function $f()$ which separates $< x >$ into two classes $y_i = \pm 1$. Instead of using a criterion based on the error rate minimization (empirical risk minimization, see page 176) of the training set, SVM are built on the structural risk minimization [203]. Such a formalism prevents the case of over-fitting[6]. A hyperplane is built which maximizes the distance *(margin)* of the training set from the hyperplane and therefore penalizes the generalization error.

Linear case

If it exists at least a $m - 1$-dimensional hyperplane which separates the m-dimensional space into two classes, the training set is said to be linear. Such a linear process is called a maximum margin classifier, based on a linear discriminant function written $f(x) = w^T x + b$, where w is a weight vector and b the bias. A $m - 1$-d hyperplane $\{x : f(x) = 0\}$, normal to w, divides the space into two parts. The sign of $f(x)$ denotes on which side of the hyperplane x is.

Among the multiple hyperplanes, SVM finds the one which maximizes the closest distance from the hyperplane on each side. Such a hyperplane, called "Optimal Separating Hyperplane" (OSH), is determined as the solution of a convex Quadratic Programming Problem (QPP) with inequality constraints based on the maximization of the geometric margin $2/||w||$

$$\min_{w,b} \frac{||w||^2}{2} \quad \text{subject to: } y_i(w^T x_i + b) \geq 1, \forall i \in \{1...n\} \tag{6.11}$$

where the geometric margin represents the distance between $f(x) = -1$ and $f(x) = 1$. Such an

---

[6]Over-fitting occurs when f reflects too closely the structure of the training data (too many details or noise); it implies a loss of generalization power. To reduce it, post-processing based on pruning operations is generally adopted.

optimization problem can be solved using Lagrange multipliers λ_i, we obtain to the "dual problem"

$$\text{maximize} \quad \sum_{i=1}^{n} \lambda_i - \frac{1}{2} \sum_{i,j=1}^{n} \lambda_i (y_i y_j \; x_i^T x_j) \; \lambda_j \; , \text{ subject to } \lambda_i \geq 0 \;\; \text{and} \;\; \sum_{i=1}^{n} \lambda_i y_i = 0 \qquad (6.12)$$

The hyperplane is determined by a relatively small number of training points since everything inside the convex hull of each set of samples is irrelevant. Those points are characterized by $\lambda_i = 0$, the others ($\lambda_i > 0$) denote the points x_i that are on the margin (the nearest points to the separating function), the so-called *Support Vectors*. w is derived from $w = \sum_{i=1}^{n} \lambda_i y_i x_i$. For any training point x_i with $\lambda_i > 0$, we have $y_i(w^T x_i + b) = 1$, which delivers the value of b. The resulting decision function is given by

$$f(x) = \text{sign}(w^T x + b) = \text{sign}(\sum_{i=1}^{n} \lambda_i y_i x_i^T x + b) \qquad (6.13)$$

Non-linear case

In general, the linear hypothesis for the training areas is too restrictive since the latter can be corrupted by noise, outliers or measurement errors. To reduce their impact, *slack variables* $\{\xi\}_{i \in 1...n}$ are introduced to penalize some samples when $\xi_i \neq 0$. The *soft margin* method [203] reformulates Eq. 6.11

$$\min_{w,b,\xi} \frac{||w||^2}{2} + C \sum_{i=1}^{n} \xi_i \quad \text{subject to: } y_i(w^T x_i + b) \geq 1 - \xi \; , \;\; \xi \geq 0 \; , \;\; \forall i \in \{1...n\} \qquad (6.14)$$

where the regularization parameter C, $C \geq \lambda_i$ acts as compromise between margin maximization and margin errors (weaker tolerance for small errors such as outliers).

To circumvent the case where the training data-set needs to be separated with more elaborated (non-linear) shapes, a technique called *kernel trick* is applied to the maximum-margin hyperplane by replacing the *dot product* ($x^T x = x.x$) by a non-linear kernel function. Using a function Φ, the input space $\mathbb{R}^m$ is mapped in a higher dimensional space $\mathbb{F}$ where the data become linearly separable.

$$\Phi \; : \;\; \mathbb{R}^m \;\; \longrightarrow \;\;\; \mathbb{F}$$
$$x \;\;\; \longmapsto \;\; \Phi(x)$$

If the new training set $\{(\Phi(x_i), y_i)\}_{i \in 1...n}$ is linearly separable, the derived OSH parameters are

$$w \;\; = \sum_{i=1}^{n} \lambda_i \; y_i \Phi(x_i)$$
$$f(x) = \text{sign}(\sum_{i=1}^{n} \lambda_i \; y_i \Phi(x_i).\Phi(x) + b) \qquad (6.15)$$

Indeed, the explicit knowledge of Φ is not required, Φ contributes only to the dot product $K(x, x') = \Phi(x).\Phi(x')$, called kernel. Such a process belongs to the category of *kernel* methods since SVM depends on the data only through dot products ("." in Eq. 6.13 and Eq. 6.15). $K(x, x')$ is a symmetric definite positive function (according to Mercer's theorem); some examples of kernels are presented in Table 6.1.

Extension to several classes

A solution to overcome the use of SVM for the classification into multiple classes $< c_1 ... c_l >$, $l > 2$ is to reformulate the problem as several binary problems. Two cases are possible, either considering the sample of the studied class c_i against all the other samples $< c_j >_{j \neq i}$, or comparing every pair of class $(c_i, c_j)_{j \neq i}$, the assigned class being the one which won the largest number of "duels".

6.3.3 Analysis and Evaluation

Some investigations on supervised classification and clustering methods are presented using Ikonos multispectral data. Due to the restriction imposed by the implementation of the Mean Shift algorithm (computation in a non-linear feature space) and the Median Cut approach, the data are rescaled to 256 values per channel and the infra-red component is ignored (see Fig. 6.12).

Some results obtained with SVM

If the linear kernel is useful to evaluate the data complexity, it can be used as well as starting point for an analysis based on a non-linear kernel. Indeed, there is no principled method to select a particular model; several kernels are tested, the one which provides the best error rate -measured using cross-validation- and provides the best qualitative results is retained. Nevertheless, among the investigated kernels (table 6.1) run on different images, promising results were obtained using a Gaussian kernel.

Linear	Radial Basis Function	Gaussian RBF	Polynomial	Tan. Hyperbolic
$x \cdot x'$	$\exp\left(-\gamma \lVert x - x' \rVert^2\right),\ \gamma > 0$	$\exp\left(-\dfrac{\lVert x - x' \rVert^2}{2\sigma^2}\right)$	$(x \cdot x' + 1)^\gamma$	$\tanh(\alpha x \cdot x' + \beta),$ $\alpha > 0,\ \beta < 0$

Table 6.1: Some kernel functions $K(x, x')$ with x the data to be classified, x' the learned samples.

The latter is governed by the parameter $\gamma = 1/2\sigma^2$, it represents the ability of the decision boundary to fit the data (a small value provides nearly linear results while a too large one leads to over-fitting). The setting of this parameter is slightly difficult and dependent on the analysed data. In addition, it has to be combined with another parameter, the soft-margin constant C. Small values for C allow to increase the margin (i.e. points too close from the decision boundary are ignored). However, considering the pair $\{C, \gamma\}$, similar results are obtained from different combinations. As a consequence, several tries are required to select the kernel model, its parameter and the soft margin constant.

Another critical issue is related to the data consistency. If the dot product is very attractive to handle heterogeneous data-sets, the relative contribution of each feature has to be integrated and

taken into account. A typical approach is to apply a standardization (substract to each feature its mean and divide it by its standard deviation) but it is not well-suited in case of heterogeneous data since it destroys the data sparsity.

The normalization of the data is always a sensitive stage to prevent the possible degradation of the SVM accuracy. Within the SVM, it can be done separately for each feature either on the data or the kernel itself by reducing vectors to unit length. Indeed, the inner product in 2D between two vectors of unit length returns the cosine of the angle between them. In Eq. 6.16 we notice that for linear kernels, the normalization in the input- or feature-space are equivalent. In addition, the Gaussian kernel does not require strictly a normalization since it already satisfies $K(x,x) = 1$; however, even with such kernels, a normalization can increase the classification accuracy.

$$\hat{x} = \frac{x}{||x||} \quad \text{, with } ||x||^2 = \sum_{i=1}^{n} x_i^2 \quad \hat{K}(x,y) = \frac{K(x,y)}{\sqrt{K(x,x)K(y,y)}} \qquad \frac{C_+}{C_-} = \frac{n_-}{n_+} \tag{6.16}$$

<div align="center">Input Space (data)   Feature Space (kernel)   Un-balanced data</div>

The case of un-balanced data (e.g. important amount of positive samples, compared to negative ones), can be compensated by weighting the soft margin constant (see in Eq. 6.16 where n_+ denotes the number of positive samples (resp. n_- for negative samples) and C_+ its associated soft margin constant). To solve Eq. 6.14 and therefore compute the weighted coefficients corresponding to a binary classification, we use the algorithm described in [105]. Once the coefficients are derived[7], given a kernel model and $\{C, \gamma\}$, the data are processed using a classifier we implement.

The two general rules to extend binary SVM to multiple classes have been studied. Only few variations between the "one versus all" and "one versus one" results were reported. The interest of the "one versus all" model is motivated by its computational efficiency. Indeed given n learned classes, only n kernels are computed, compared to the $n(n-1)/2$ required for the "one versus one" model. However, the increase of learning samples impacts the solver performances to determine Eq. 6.15.

We present some tests based on a metric resolution Ikonos image (Fig. 6.12). 6 classes are trained with a relatively few amount of samples (3 to 6 per class) and computed with both multiclass schemes. Table 6.2 emphasizes the similarity of results using a Gaussian kernel with $\gamma = 0.003$. To reduce the sensitivity of the learned samples, especially when few of them are used to characterize a class, we try to set the samples on a segmented image through a Graphical User Interface. Indeed, working on regions presents several advantages: it depicts some local behaviors, the region's spread makes easier the sample selection, and the learned samples get the regions radiometric properties, which are -in the case of the proposed algorithm (chapter 5)- the weighted average of the pixels belonging to the selected region. Using regions instead of single points adds another criterion to the SVM: the segmentation granularity controled by its scale parameter. In Table 6.2, we report about 5% difference using learned samples extracted from a segmentation image with $\overline{|R_i|} = 64$ (1792 regions), $\lambda = 1595$.

---

[7]i.e. among the learning samples, only the support vectors are kept $w_i \neq 0$.
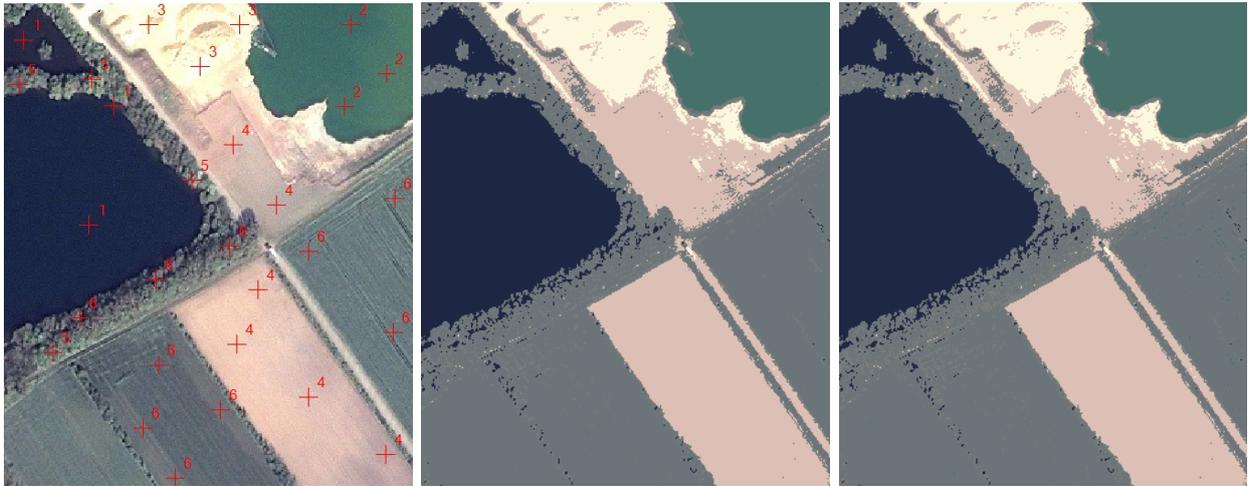
(a) Ikonos image, 6 trained classes   (b) SVM, "one versus all"   (c) SVM, "one versus one"

Figure 6.12: SVM classification from an Ikonos image over landscape area (size 310×370).

classes	1	2	3	4	5	6
1	23000	0	0	0	0	20
2	0	10347	0	0	10	221
3	0	0	7072	29	0	0
4	0	0	53	26208	0	0
5	0	0	0	48	6156	1945
6	0	0	0	0	2	39589

one (column) versus all (line).

-class repartition, from 114700 pixels-

classes	1	2	3	4	5	6
1	23000	0	0	0	0	263
2	0	10239	0	0	17	180
3	0	0	7120	578	0	0
4	0	0	5	25558	448	0
5	0	0	0	149	6156	5485
6	0	108	0	0	877	35847

one (column) versus all (line)

-based on samples taken from a segmentation-

Table 6.2: Comparison of SVM results (assignation comparison (in pixel)) using an Ikonos image (see Fig. 6.12) based on 6 training samples, respectively around 2% (left) and 7% (right) differences.

An implementation of the K-Means algorithm

A version of the K-Means algorithm is investigated for the two main purposes of data analysis: (1) exploration, to understand and evaluate information content of raw images without any pre-specified models, (2) inference, to validate a model derived from some low-level operations such as region-based segmentation or texture patterns. Instead of using a random clustering selection for the initialization step, the cluster centroids are set by computing the first and second order statistics (m_l, σ_l^2) for each channel l

$$c_l^i = m_l + \sigma_l \, \frac{2\,i-k}{k} \; , \quad \forall i \in \{1...k\} \tag{6.17}$$

where c_l^i denotes the $l-$coordinate of the ith cluster's centroid in the feature space. Such an initialization enables to reduce the risk of empty clusters. Indeed, by positioning the set of seeds around the gravity center according to the standard deviation of each axis, clusters are willing to expand.

The update of clusters is performed only once all the data are assigned. Two further processes are integrated within the update step, acting mostly for the earliest iterations. A stage controls for each pair of clusters their relative proximity compared to their distribution. It mainly corresponds to a compactness test. According to a given user-fixed threshold, the pairs of clusters too close from each other are merged. The remaining empty cluster is re-used to split the one presenting the largest deviation into two clusters along its principal component (see PCA analysis, page 23). Outliers, characterized by small objects, are removed since they can affect the classification results.

Unfortunately, if the presented approach significantly reduces the number of iterations and provides satisfactory results, it does not prevent the convergence of the criterion function to local minima. The analysis of the classification quality is not trivial since classical estimators, such as the mean squared error (MSE), are related to the number k of clusters.

The proposed algorithm is compared to the implementation of the Median-Cut algorithm available within the *TIFF* library. If for some images (e.g. parrots in Fig. 2.9) the latter outperforms the K-Means approach when the number of clusters is excessively low, the K-Means delivers more consistent results in the case of RS images (see Fig. 6.13). Indeed, those different behaviors are directly related to the data content itself and this comparison mainly reflects the ability of the methods to capture clusters distant from the principal component axis.

The potential gain from the Median-Cut algorithm on some images and its computational efficiency are however balanced by the introduction of a bias in the reconstructed data. It results in brighter and more unrealistic colors with a larger mean square error. The difference in the missing clusters (when k is too low) is mainly related to the initialization stage. Using the spectral partition of the median-cut algorithm as cluster seeds for the K-Means algorithm enables to reach similar results without any bias and to present a smaller MSE.

As the number of clusters increases and reflects much better the information encapsulated in the image, the overall quality of the K-means algorithm becomes superior to the divisive approach for a wide spread of images. The proposed initialization stage delivers slightly better results and presents a smaller MSE. Indeed, those investigations highlight the lack of such an approach: in case of few classes, initial clusters fail to properly migrate in the periphery. This effect is circumvented when the number of clusters is more in emphasis with the partitioned data. Compared to random processes, an advantage of the initialization along the main statistical axes is a more confident analysis of the amount of information in the data while setting k. A plot of the MSE versus the number of iterations emphasizes the clustering behavior depending on k and can help to set the number of clusters.

The K-means approach can be used as well as post-processing stage to perform segmentation regularizations. Indeed, prior to design an iterative work-flow for the segmentation of large images, a classifier was used to fuse regions being on the border of each tile: such corresponding and adjacent regions, belonging to the same cluster, were merged.

(a) Original Image   (b) K-Means, 14 classes   (c) K-Means, 6 classes

(d) Median Cut, 14 colors   (e) Mean Shift, 14 colors   (f) Mean Shift, 6 colors

(g) Median Cut, 22 classes   (h) Segmentation, $\overline{|R_i|} = 144$ (i) Segmentation, $\overline{|R_i|} = 625$
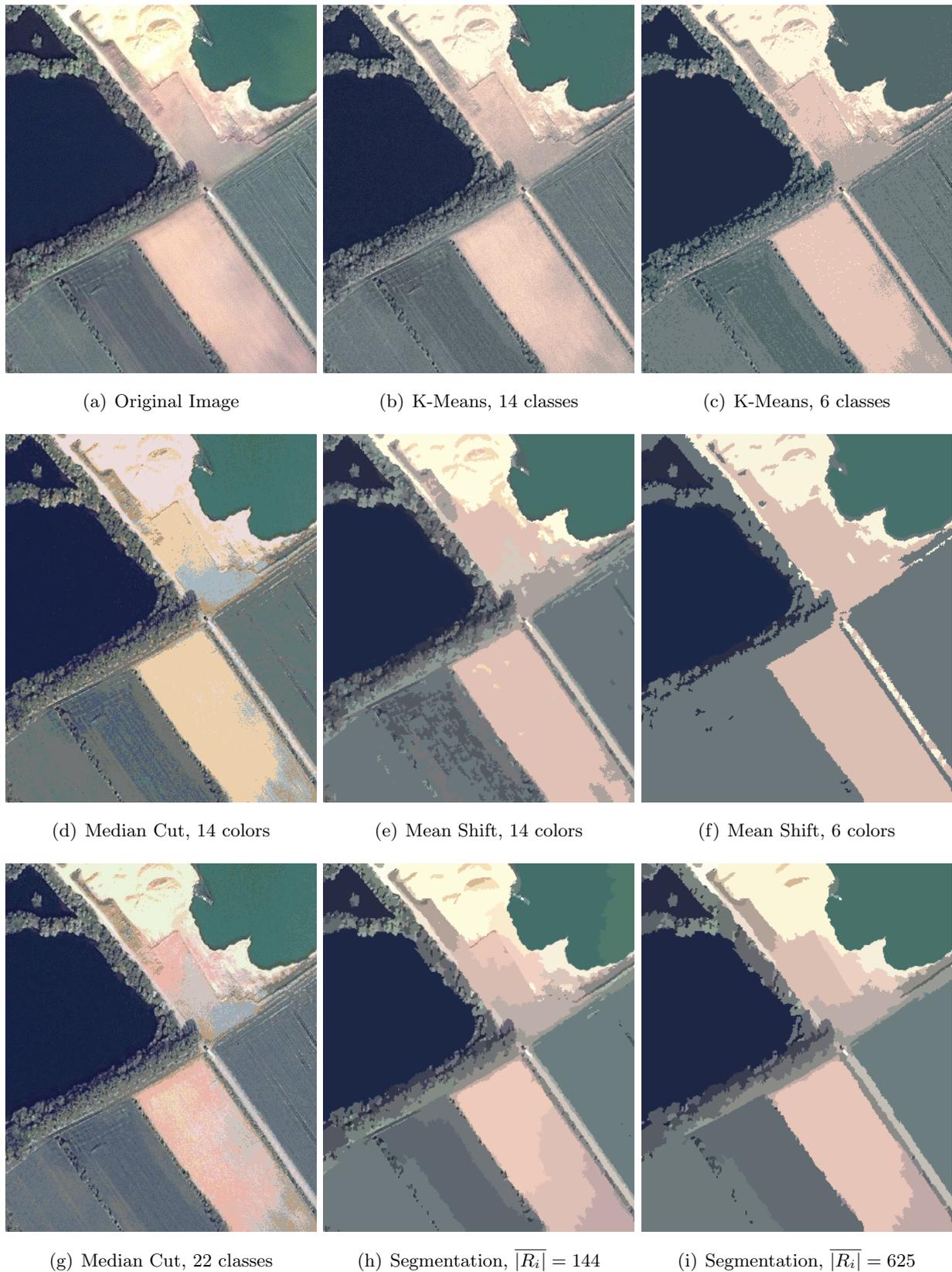
Figure 6.13: Clustering of an Ikonos image using K-Means and Median Cut algorithms. Some results derived from our segmentation (see chapter 5) and the Mean Shift approach. (e) and (f) refer respectively to "Over-segmentation" and "Under-segmentation" cases mentioned in [37].

Comparison with segmentation methods

Based on the same data-set, some results derived from the Mean Shift algorithm [37] and our region-based approach are presented in Fig. 6.13. The region merging algorithm was set to reach the same amount of regions as each of the two computed Mean Shift modes[8]. Indeed, both cases correspond to over-segmentation; i.e. too many details are removed, the results presenting only the dominant components of the image. In [37], a further mode called "quantization" is distinguished and provides a segmentation into 40 colors (about 20247 regions) for the example shown in Fig. 6.13.

method	classes	mean	σ	regions	method	colors	mean	σ	regions
K-Means	14	-0.070	6.432	27553	Mean Shift	14	-0.261	12.516	717
Median Cut	14	-3.820	13.008	19959	Segmentation	$\overline{|R_i|} = 144$	-0.006	10.691	796
K-Means	6	-0.033	10.503	7785	Mean Shift	6	0.580	16.924	136
Median Cut	22	-3.166	10.136	30671	Segmentation	$\overline{|R_i|} = 625$	-0.038	13.151	183

Table 6.3: Some statistics related to the results presented in Fig. 6.13. Mean and standard deviation are derived by the average of each channel difference with the raw data.

Summary in Image Analysis

The results presented in Table 6.3 cannot be strictly compared since objective of segmentation and classification differ. We notice a better behavior of the K-means approach, compared to the Median Cut, which is still attractive due to its velocity. The hybrid method Mean Shift delivers interesting results, circumventing classical drawbacks of the pixel-wise classification approaches. However, numerous details are removed and the results appear "smooth" with inaccurate boundaries. Such an approach can be relevant for the purpose of data simplification or reduction.

Very promising results were obtained using the region merging algorithm, applied on a large broad of data-sets with different contents (optical or polarimetric SAR images, range data such as X-Ray images, volumetric data-sets generated by computer tomography, etc.).

The motivation and the interest of a separated and universal workflow for the modeling of region based segmentations is emphasized by the possibility to apply the topological and geometrical analysis to every region-based approaches listed in chapter 4.

SVM techniques present a high interest to develop semi-automatic methods. In particular, learning methods can be run on image partitions, and discriminate regions of objects presenting specific patterns or similarities.

---

[8]Examples with K-Means and Median Cut approaches provided in Fig. 6.13 are computed with the same amount of classes (or colors).

Part III

DEM Analysis

Filtering of Digital Elevation Models

Abstract

Techniques for SAR or optical image denoising are investigated and classified according to their ability to properly model and suppress noise. In particular, we focus on multi-scale and Bayesian approaches for their ability to deal with non-stationary data.

Performances of the described filtering methods are evaluated and compared using another type of image information, digital elevation models. Indeed, there is much interest for the improvement of such EO data-sets which constitute a stable time reference and may cover almost completely and homogeneously the Earth (e.g. SRTM, upcoming TanDEM-X).

To cope with most of the inconsistent artifacts (e.g. InSAR unwrapping errors, image matching) and noise (thermal and low coherence), a filtering process is required. Based on statistical, fractal and qualitative analyses, a methodology to qualify, quantify and validate DEM filtering accuracy is given using optical stereoscopic and InSAR DEM.
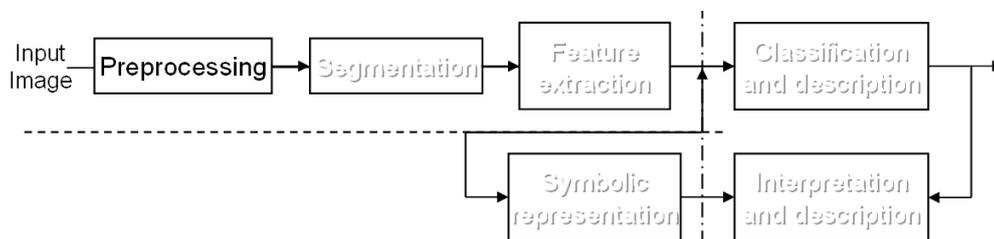
Introduction

Like every system collecting physical measures, data acquired by imaging sensors are pervaded with noise and errors. In addition, physical and structural properties, relief of the remotely sensed scene affects the accuracy and the reliability of the corresponding measure (see section 3.3).

The ultimate objective of image denoising is to produce an estimate $\hat{f}$ of the unknown noise-free image f. Based on some function of a local neighborhood, information content is modified, involving a loss of information as well as an expected enhancement. Filtering and restoration processes are distinct, the latter is devoted to the emphasis of particular image features. Filtering is an intensive field of research in image processing and still an open issue to improve high complexity data. Curiously, less attention has been paid to the filtering of digital elevation models, whose noise modeling is unknown.

7.1 Image Filtering

Linear filtering can be considered as a controlled scaling of the signal components in the frequency domain (see section 2.1.1). Filtering in the Fourier domain presents several advantages since the Fourier coefficients can be changed apart from each other. If wider filter supports are applied without drastically increasing computational time, in most of the applications reported in the literature, the convolution is performed in the spatial domain with an approximated convolution kernel truncated by an appropriate windowing function.

7.1.1 Stationary Assumption

The simplest filter class concerns filters that do not exploit any noise modeling. The same operation is applied using scanning windows of a predefined size centered on the processed pixel. A stationary hypothesis is considered for the data. Two linear and non-linear filters are presented.

Mean Filter and Gaussian Smoothing

The "Mean filter" replaces the pixel considered by the arithmetic average of all values within a sliding window. Although the average process reduces extreme artifacts, it also decreases the spatial resolution. Due to the linear approach of this filter, the window size determination represents a trade-off between noise and resolution reduction. As a consequence, even if the local variance is decreased by the square of the window size, poor results are obtained. They can be improved by computing locally the coefficient of variation to adapt the window size.

Smoothing signals with a Gaussian distribution have been intensively studied in image processing and computer vision. A smoothing process applied on a signal consists in suppressing its high frequencies and enhancing the low frequencies. The smoothing function is sometimes referred to as a low-pass filter applied either on the frequency domain or on the spatial domain with a truncated approximated Gaussian kernel (see Eq. B.10, page 180). The degree of smoothness is carried out by σ, its determination is highly dependent on the information content.

Median, Geometrical Filters

The non-linear "Median" filter replaces the processed pixel by the median value of the estimation window. No new value is computed, and in comparison with the mean filter, edges and shapes are much better preserved. However, isolated point scatterers and local artifacts are handled in a similar way. Its non-linearity makes the analysis of the median smoothing rather difficult, some theoretical analyses are discussed in [155], [71].

Geometrical filters such as the Crimmins filter [42] exploit non-linear combinations using morphological operations (see section 2.1.3 page 15) of the pixel centered on the windows estimation.

An iterative process is performed to reduce the level of noise in valleys or hills/mountains. It preserves the structural information in a better way than the above filters without any noise modeling considerations. These corresponding algorithms enable extremely fast computations but lead to poor filtering performances. Intuitively, the incorporation of knowledge in the filtering process should lead to improved results. In many cases, the assumption of an additive Gaussian noise model is reasonable. In particular, we made this assumption to filter elevation data (see section 7.2).

Concerning remote sensing, since most of the research in image filtering focuses on SAR data which are corrupted by multiplicative speckle noise, the following overview is not restricted to additive noise models. Indeed, homomorphic filtering is used in the log-spectral domain to separate signal from the noise; by applying a logarithmic transformation, a multiplicative noise model is turned into an additive one.

7.1.2 Adaptive Filters

Limitations of the presented filters are explained by the fact that the stationary assumption is not valid Adaptive filters are more suited for non-stationary signals which are considered as locally stationary. Statistical estimators (mean and variance) are computed in a fixed windows size and further used to control locally the filtering. Since structure and texture (e.g. shapes, terrain's roughness) contribute to local statistics, additional structural information is often required to preserve edges and features and filter their surroundings. Statistical filters partially exploit some noise modeling using the assumption of a Gaussian distribution expressed in Eq. (B.10).

For SAR denoising, Kuan et al. [124] adopted a method based on a Minimum Mean Square Error (MMSE) criterion. A MMSE estimate is first developed for an additive noise model $y = x + n$. In order to adapt the MMSE method to the multiplicative noise model, it is further considered under the form $y = x + (n - 1)x$ from which the corresponding linear filter is deduced. The Kuan filter is optimal when both the scene x and the detected intensities y follow Gaussian distribution. In this case, the estimate $\hat{x}$ is

$$\hat{x} = k \, y + (1 - k) \, \mu_y \quad , \text{ with } \quad k = \frac{\sigma_x^2}{\sigma_x^2 + (\mu_y^2 + \sigma_x^2) \ / \ L} \tag{7.1}$$

The variance of the signal is derived by $\sigma_x^2 = (L \, \sigma_y^2 - \mu_y^2) \ / \ (L+1)$, where L denotes the equivalent number of looks of the speckle noise. Among the proposed variants of the Kuan approach, the Lee filter [127], [128] differs only in the weighting factor k, a linear approximation of the multiplicative noise model leads to

$$k = \frac{\sigma_x^2}{\sigma_x^2 + \mu_y^2/L} \tag{7.2}$$

The comparison of the two filters does not lead to significant differences. It explains the interest and the popularity of the simplified form compared to the Kuan filter. Compared to the previous methods, these adaptive filters constitute a strong improvement. Wiener filters are optimal linear

filters in the sense of MMSE for stationary signals corrupted by additive noise. The noise modeling is of higher level since the knowledge of its power spectra and the noise-free scene are required. Correlated additive noise should theoretically be handled by a full Wiener filter.

In practice, optimal results are obtained with Gaussian distribution for both noise model and noise-free scene. To deal with non-stationary images, efficient methods are proposed to design optimum filters for each region. The well-known Frost filter [68] is an adaptive Wiener filter derived for SAR speckle reduction. Pixel values within a fixed size window are convolved with an exponential impulse response $h(r)$ given by

$$h(r) = N \exp\left(-k \; \frac{\sigma_I^2}{\mu_I^2} \; |r| \; \right) \tag{7.3}$$

where k is the filter parameter (damping factor), N acts as normalization constant and $|r|$ is the radial distance from the center pixel to be filtered. This response comes from an auto-regressive exponential model assumed for the scene reflectivity x. Compared to a full Wiener filter, this assumption limits the performance of the Frost filter.

7.1.3 Multi-scale Approaches

Multi-scale approaches are based on the analysis of the image at different resolutions. In this scope, inter-scale dependencies are better expressed using a pyramidal decomposition to represent the image. To merge and exploit the relevant information extracted from each level of the pyramidal scale, numerous approaches are investigated [19].

Most filters based on *multi-resolution decomposition* follow a "coarse to fine" approach: Since it is difficult to directly capture relevant image content, information extracted at lower scale is exploited in higher scale, allowing to generate further knowledge and a priori information. The basic statistical estimators (mean and variance) are more relevant and provide the main behavior of the image thanks to varying neighborhood sizes. As for the information extraction, the filtering is performed at every scale.

At lower scale, the filtering is simplified due to the noise variance decrease (proportional to the square of the resolution reduction). Progressively, the analysis is performed at a higher and higher scale. Filtered images at each scale are recombined to form the final filtered result. Compared to the corresponding single-scale analysis methods, multi-scale techniques can both improve performance and reduce computation. Using multi-scale models [19], the incorporation of inter-scale dependencies delivers promising results.

Wavelet filters are based on a pyramidal discrete wavelet transform [144]. The use of wavelets to filter images with additive noise is well documented in the literature. It should be noticed that generally wavelet coefficients are large in irregular regions and small in uniform regions. If the signal is pervaded with noise, this noise dominates the small coefficients at finer scale and only large coefficients are linked to the strong similarity of the underlying function.

In [51], Donoho introduced a simple non-linear algorithm for image denoising, based on a wavelet decomposition under the assumption of additive Gaussian noise. The wavelet coefficients are shrunk in order to reduce the noise energy in different sub-bands using a *soft-thresholding* technique:

$$\eta_s(w, t) = \begin{cases} w - t & w \geq t \\ 0 & , \quad |w| < t \\ w + t & w \leq -t \end{cases} , \text{ with } \quad t = \gamma \sigma \sqrt{2 \log(n)} \tag{7.4}$$

where n denotes the number of signal samples, γ is a constant related to the wavelet basis (set to 1 in [53]) and σ^2 is the noise variance that can be estimated from the wavelet's finest level coefficients. An alternative to Eq. 7.4 consists in *hard-thresholding*

$$\eta_h(w, t) = \begin{cases} w & |w| \geq t \\ 0 & |w| < t \end{cases} \tag{7.5}$$

Both methods attempt to remove noise while preserving the large coefficients. The difference relies on a continuous function which shrinks all observations (soft-thresholding) while hard-thresholding removes only the small ones. The process described in [51] is achieved in three steps: (1) to decompose the signal into a pyramidal wavelet basis, (2) to threshold the coefficients (Eq. 7.4) with a particular parameter t and finally (3) to recover the signal by applying the inverse wavelet transform. As for any problem of this type, to set properly the threshold parameter is a major issue. For instance, an adaptive threshold that could be adapted to each sub-band of the wavelet decomposition should bring more efficient results. However, many limitations have to be considered such as the introduction of wavelet artifacts and the large choice of wavelet decompositions as well as thresholding techniques (type and settings).

7.1.4 Bayesian Approaches

Gamma MAP Filter

For SAR denoising, Lopes et al. proposed in [134], [133] a Maximum A Posteriori (MAP) filter based on a Bayesian analysis of image statistics. The Gamma MAP filter assumes that both radar reflectivity (via the prior, Eq. 7.6) and speckle noise (via the likelihood function, Eq. (B.11)) follow a Gamma distribution. The expression of the prior as Gamma distribution is

$$p(X_i = x_i \mid \mu_x) = \frac{\nu^\nu \, x_i^{\nu-1}}{\mu_x^\nu \, \Gamma(\nu)} \exp\left(-\frac{\nu \, x}{\mu_x}\right) \tag{7.6}$$

where μ_x is the mean value of the cross section and $\nu = (\mu_x/\sigma_x)^2$ reflects its roughness. The "superposition" of these distributions results in a K-distribution. This choice is reinforced by the fact that for the uniform cross section, the speckle pdf corresponds to a gamma distribution. From a practical point of view, the K-distribution matches a large variety of radar back-scattered of land and ocean targets. The noise-free data cross section $\hat{x}$ (estimate of x), is given by

$$\hat{x}_{i_{MAP}} = \frac{\alpha - L - 1}{2 \, \alpha} \mu_y + \sqrt{\frac{(\alpha - L - 1)^2}{4 \, \alpha^2} \mu_y^2 + 2 \, L \, y_i \mu_y} \tag{7.7}$$

with L being the number of looks of the data, $\alpha = (L + 1)/(L \; \sigma_y^2 \; / \; \mu_y^2 - 1)$. The mean μ_y and the variance σ_y^2 of each pixel are estimated by means of sliding windows which extract the neighborhood behaviors. The Gamma-Gamma Map filter is a well-established speckle filter based on two parameters. However, in spite of a more accurate noise modeling, comparable results are obtained with the traditional Kuan or Lee filters.

Bayesian Filter with GMRF Prior

Walessa adopted in [46], [209] and [207] a Bayesian approach to filter SAR images using as prior Gauss Markov random fields (GMRF)[1] and a gamma likelihood model. Assuming an uncorrelated noise model (via the likelihood pdf), the filtering process is considered as an ill-posed inverse problem, and formulated in the general frame of Bayesian inference at two levels:

1. **Model fitting**: Maximum A Posteriori method with a Gaussian distribution for the likelihood and GMRF models are employed as prior.

2. **Model selection**: Evidence Maximization computed for a library of models centered on a guess of the parameter to be estimated.

The two-level scheme is explained as follows: To filter the image y, using the Bayes equation (B.3), a MAP estimation of the data (first level, maximization of the numerator in Eq. (B.3)) is used choosing (second level) different prior parametric models. It enables retrieving the noise-free image x. Indeed, there is no direct analytical solution to the Bayesian equation. The MAP estimation (first level) generally considers the evidence term as normalization. Assumed a given model, the computed guess fits the data. However, its quality is highly dependent on the model used. A model selection (second level of inference) is performed in order to get the model that best describes the data through the calculation of the evidence. The two stages are iteratively computed until the evidence reaches a maximum.
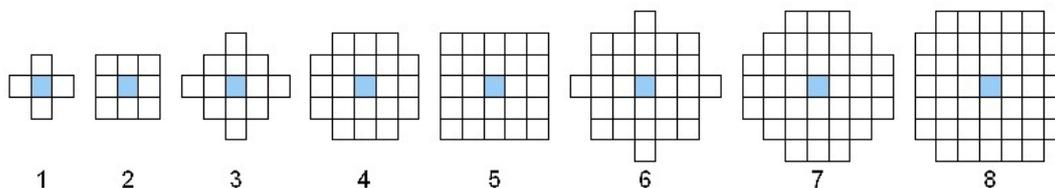
Figure 7.1: GMRF model order and the corresponding vicinity studied $\mathcal{N}_n$. A GMRF is a symmetric model composed of single cliques (see a definition page 178). Applied on a regular lattice at position (x, y), its parameters θ_n are spread on $\mathcal{N}_n$ to each pair of pixel $(x + i, y + j), (x - i, y - j)$ with $i, j \geq 0$. A GRMF model is described by $\{\theta_1, \theta_2\}$ with an order $n = 1$, $\{\theta_1...\theta_4\}$ with $n = 2$, etc.

The structural complexity of the model is defined by its order (i.e. neighborhood size, see Fig. 7.1 and section B.1.3, pages 177-180). For analytical tractability, Walessa uses a library of

---

[1]we refer to Annex B.1.3 for presenting more details about Bayesian analysis.

small models. This choice is motivated for several reasons: the theoretical concepts (Occam's razor, curse of dimensionality) prevent the use of GRMF within a large neighborhood size (i.e. high order). Indeed, expanding the neighborhood is not a practical solution because parameter estimation and subsequent interpretation are more difficult. Another consequence of using a misfit model (due to a large window of analysis) is the incapacity to extract high structural complexity within the image.

More efficient results are reached by the extraction of sharp edges and strong scatterers. Such non-linear features have to be preserved and do not have to influence speckle filtering since they limit the ability of GMRF model to fit the SAR image. Hence, a region growing algorithm is performed to extract the edges, whereas strong scatterers detected by the prior model during the filtering are excluded. When the filtering is achieved, the characteristic features extracted in the previous steps are reinserted. In Fig. 7.2, a flowchart of the described algorithm is given.
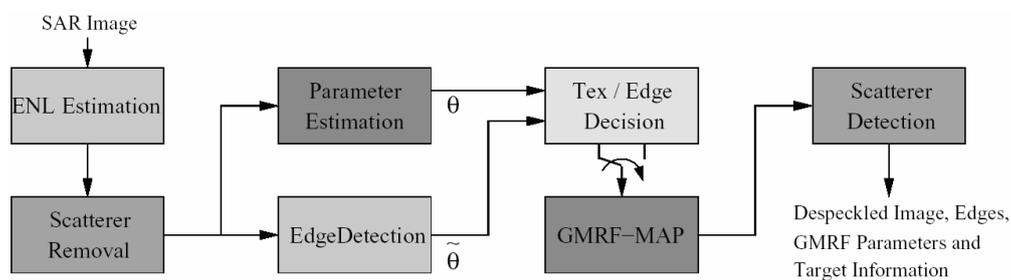
Figure 7.2: Model Based De-speckling filter: Flowchart of the SAR filter with the various image processes (reproduced from [207]).

This algorithm allows to filter SAR images by removing the speckle/noise and preserves relevant features such as the strong scatterers and the edges. Additional information is also provided, consisting of texture parameters (GRMF model parameters), scatterers and edge maps. Unfortunately, the filter fails in the case of correlated noise since this hypothesis is not taken into account in the likelihood function.

7.2 DEM Filtering

DEMs are pervaded with errors and artifacts mainly due to the acquisition/generation techniques and lead to inaccurate and/or unrealistic data-sets. To cope with most of noise (thermal and low coherence) and inconsistent artifacts (e.g. InSAR unwrapping errors), a filtering process is required. To filter non-stationary data such as DEMs, an evaluation and a comparison of the methods previously introduced is determinant.

In this purpose, a methodology to qualify, quantify and validate DEM filtering accuracy has to be set up. Multi-resolution techniques like fractal analysis enable, through the Hausdorff's measure and dimension, to get an estimate of the terrain roughness. An evaluation of the filters is presented in [142] which illustrates the efficiency of Walessa's algorithm [209], [208]. The chosen decametric
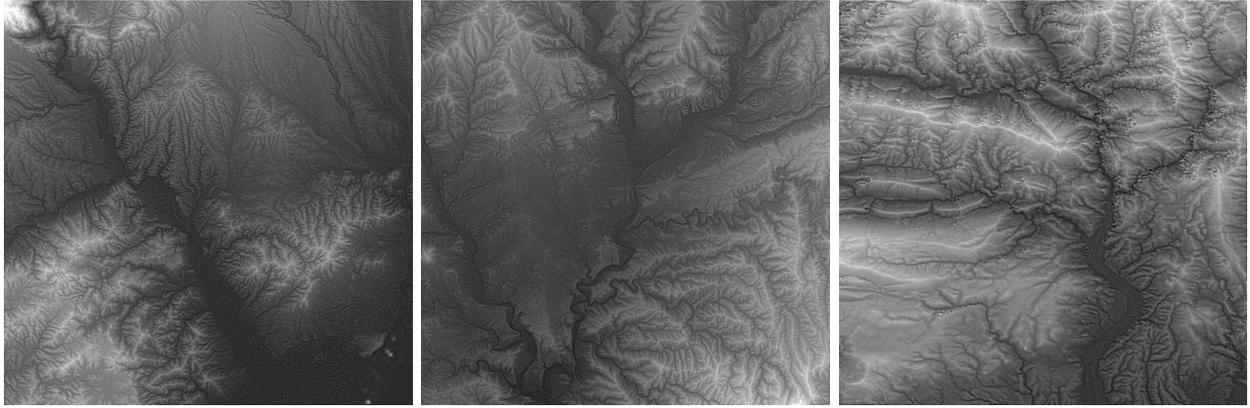
(a) SPOT5 (15m), section 7.2.1   (b) Cartosat (10m), section 7.2.2   (c) SRTM-X (30m), section 7.2.3

Figure 7.3: Decametric data-sets investigated. (Local contrast adjustment) (a), (b) are stereoscopic DEM generated from optical sensors [131], [173], [44], [130] on Barcelona area (Spain), (size 2048×2048). (c) is a SRTM-X DEM [10] over Nice (France). It mainly depicts mountainous areas (size 1024×1024).

data-sets (Fig. 7.3) are interesting since each of them presents specific behavior and data content with various level of noise/artifacts, generated from InSAR or optical stereoscopy processes.

7.2.1   Internal Validation

Most of the criteria (statistical, texture, visualisation) employed for the filtering evaluation are presented. The use of wavelet and Bayesian methods for DEM filtering is described.

Statistical Estimators

Retrieving the unknown noise-free image x, the overall purpose of image denoising can be stated as an estimation problem by minimizing the error between $\hat{x}$ and the uncorrupted data. In this scope, the minimization of the *mean squared error* (MSE) (see section 7.1.2, page 120) leads to

$$MSE = \frac{||x - \hat{x}||^2}{N} = \frac{1}{N} \sum_{u=1}^{N} (x_i - \hat{x}_i)^2 \tag{7.8}$$

where N is the number of elements or pixels studied. An estimate of the guess's "quality" is the *signal to noise ratio* (SNR), expressed in dB

$$SNR = 10log_{10} \frac{\hat{x}^2}{||x - \hat{x}||^2} = 10log_{10} \frac{\hat{x}^2/N}{MSE} \tag{7.9}$$

The *peak signal to noise ratio* (PSNR) corresponds to Eq. 7.9 where $\hat{x}^2$ is replaced by $max(x)^2$ ($max(x) = 2^8 - 1$ is taken for 8 bits images). In the literature, those estimates are mainly employed to describe signal/image compression performances.
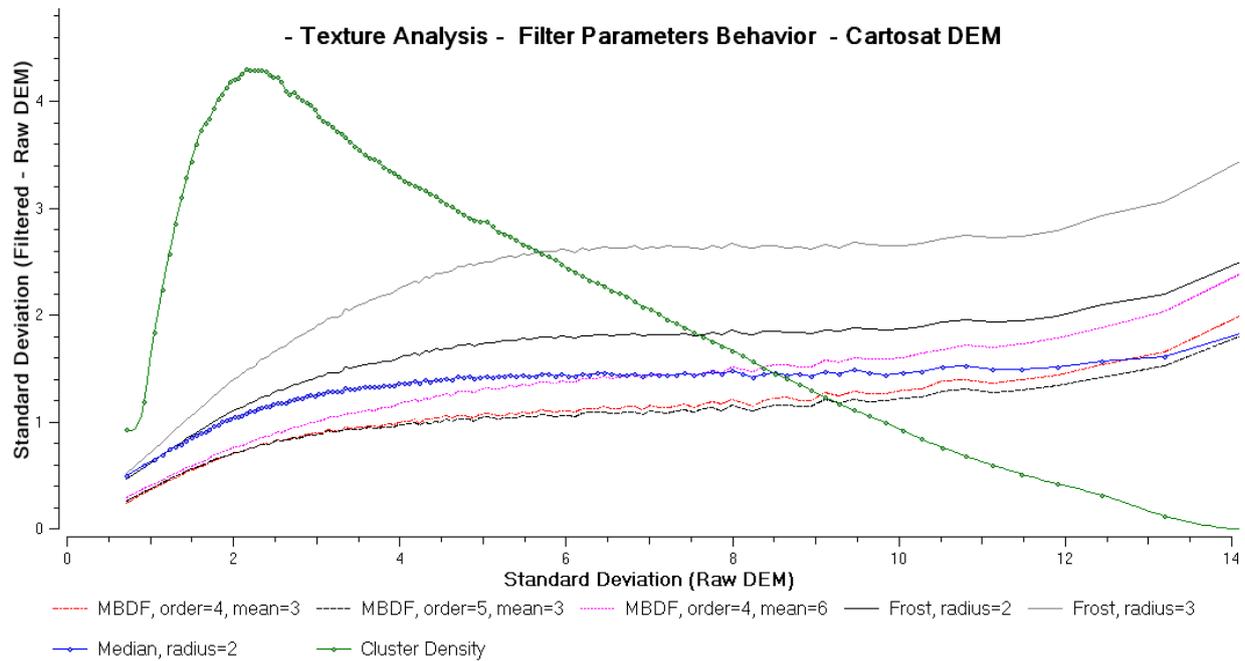
Figure 7.4: Filtering analysis using texture features on Cartosat DSM computed using 7×7 windows size. Filtering amplitude (variance *filtered-original* DSM) is expressed according to the local variance of the original DEM. Parameter's influence for Bayesian filtering (model order and windows radius for the local mean computation used in Eq. 7.11), Median and Frost filtering using radius $= \{2, 3\}$. The curve with dot symbols depicts the information density.

Texture analysis

The local standard deviation of the topographic surface provides a "rough" quantification of the DEM roughness. Indeed, this measure incorporates structures, ridges, patterns (signals) as well as noise (small variance) and artifacts (high variance). Instead of deriving a list of statistics, filtering performances with non-stationary data such as elevation data can be better understood in analysing the filtering rejection according to the local DEM texture. In this scope, the local standard deviation is computed using narrow window sizes on the original data (7×7). The resulting texture image is classified into equally populated clusters σ_i. Finally, for each cluster i, a standard deviation of the DEM difference *filtered-original*, σ_i^Δ is computed.

By means of a plotted curve f, a "signature" is obtained for each filtering process as shown in Fig. 7.4, page 126 with the curve $f(\sigma_i) = \sigma_i^\Delta$. It depicts the ability of each studied filter to deal with artifacts or/and noise. Such curves are obviously strongly correlated with the nature itself of the data (level of noise, errors, resolution, etc.). In varying the windows size for the texture parameter computation, the overall curve shape is kept with an abscissa dilation in taking larger windows size (i.e. standard deviation increase). All diagrams were generated using 128 clusters. The curve with small dots in Fig. 7.4 depicts the cluster centroids. Their repartition emphasizes the density of information and the level of roughness of the studied data.

Visual Analysis

The listed criteria (variance, MSE, SNR, PSNR, etc.) enable to derive some statistical information about the studied image. Such approaches, purely quantitative, drastically differ from the evaluation carried out through human perception, with terms and/or meaning like "sharpness", "smoothness", etc as well as a priori considerations or hypotheses concerning the analysed data-set. Indeed, those distortion measures do not always perfectly correlate the perceived quality.

Applied to elevation data - see Fig. 7.3, 7.5(a) -, the use of local contrast adjustment is particularly well-suited to emphasize local elevation variations. Simple visual analyses of DEM differences give a determinant indication on the filter behavior with non-stationary data. In Fig. 7.5, the Bayesian filter (named "MBDF" below) adapts its filtering relatively well to the relief. In opposition to multi-scale or median filters, valleys and mountain chains do not clearly appear (i.e. filtering rejection being less sensitive to the relief). It should be noticed that the decametric DEM employed was a smooth DEM with few image matching artifacts.

7.2.2 DEM Filtering with multi-scale and Bayesian Techniques

The adaptation of the presented methods for the filtering of EO elevation data is presented and illustrated with results derived from Cartosat DSM.

Parameter Choice and Behavior

All filters are controlled by a set of parameters, mainly related to the sampled window size for the local estimation of the noise to be removed and some weighting coefficients. The increase of data content, mainly in the Cartosat DSM, prevents the use of large windows of analysis to preserve structural information and features. In this scope, Frost, median and Gaussian filters were applied (convolution) with 5×5 and 7×7 windows size [2].

Since the mountainous InSAR DEM presents important shadowed areas and noise distribution of elevation data are unknown, the Frost filter (usually adressed for SAR denoising) is taken into account. The damping constant k in Eq. 7.3 is set to 1 and local statistics are estimated with windows size w larger than the one used for the filtering convolution. Influence of both factors k, w (under small variations) is weakened since only minor differences are reported.

Wavelets

Soft- and Hard-Thresholding techniques (resp. Eq. 7.4, Eq. 7.5) depend on a parameter whose amplitude expresses the filtering strength. In the case where the value of the input noise

---

[2]In the section, size of the convolution windows is given by its radius length i and indicated as follows: $filter_{r=i}$. In the case of the Bayesian filter, the model order o is written $filter_{m=o}$ (see Fig. 7.1, page 123).

(a) Raw DEM

(b) Median - Raw

(c) Wavelet - Raw

(d) MBDF - Raw
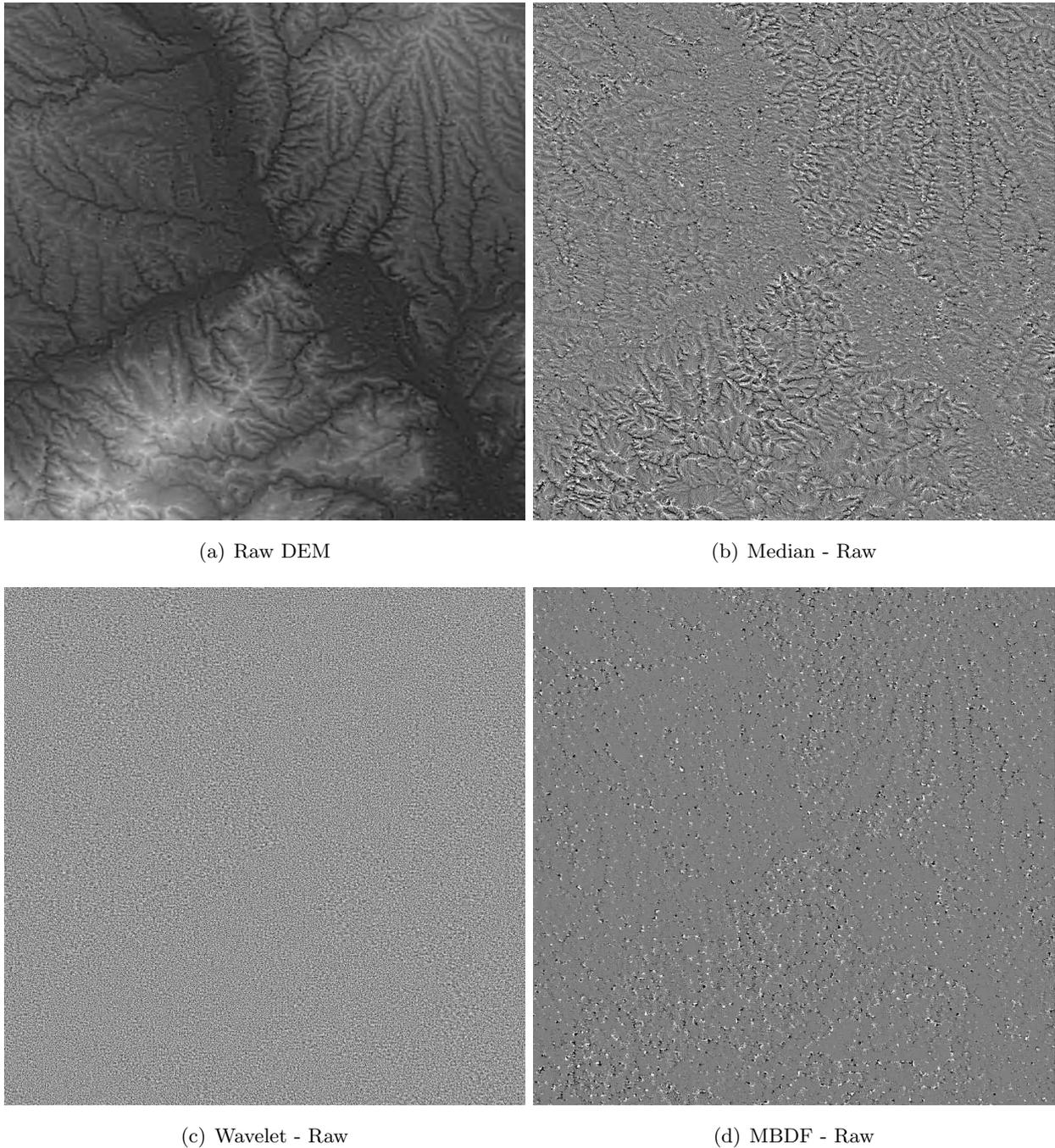
Figure 7.5: Filtering evaluation by DEM difference: $filtered - original$ using a SPOT5 stereoscopic DEM sampled to 15 meters (1024×1024). (a) depicts a local contrast adjustment applied on raw DEM. (b) denotes a median filtering with a neighborhood size defined by a circle of radius 3. (c) Wavelet hard-thresholding. (d) Bayesian filter.

variance σ^2 is unknown, a robust estimator is a median measurement since it copes with isolated outliers (defined by locally high variation amplitude). The highest frequency sub-band noted HH_1 (lower-right one in Fig. 2.4, page 12) of a wavelet decomposition is a proper location for the noise estimation, since its coefficients w mainly characterize noise. In [52], an estimate of σ is

$$\hat{\sigma} = \frac{MAD(w_{HH_1})}{0.6745} \tag{7.10}$$

where the term MAD denotes the "Median Absolute Deviation" and 0.6745 being the 0.75 quantile of the standard normal distribution. Later, the same authors introduced in [51], [53] other threshold criteria, mainly based on Eq. 7.4. For every level of a 1-D wavelet transform, two complete transforms are possible resulting either from the original bank or from its shifted version (1 pixel). Indeed, a simple signal translation results in important changes in the wavelet coefficients. To overcome aliasing, over-complete wavelet transforms are used, ensuring, at each level of decomposition, a sufficient sampling using the 2^L possibilities of a L-level decomposition. For two dimensional signals (4 complete transforms per level) 4^L combinations are obtained.

To denoise an image, since a complete transform is invertible, each of the 4^L transforms is evaluated, its coefficients are shrunk and the results of each inverse transform are averaged. Experiments have been mainly carried out using symmetric bi-orthogonal Daubechies 9/7 wavelets with 4 levels [47]. Other wavelet bases have been investigated (Haar, Antonini), but only results obtained with the Daubechies wavelets are presented.

Bayesian Approach

In [208], Walessa derived this approach to deal with non-stationary data such as InSAR DEM. Edge and strong scatterers detectors used with SAR image are removed. Locally adaptive noise estimation appears to be very well-suited to remove noise from the InSAR or Stereoscopic DEM data while preserving structural information. If an additive white Gaussian noise model with variance σ_n^2 is assumed for the data, a Gaussian likelihood is used to describe the relation between noise-less pixel values and their noise-corrupted instances. Using a GMRF model as prior information with the model parameters σ and $\bar{x}$, the value that maximizes the local posterior is a weighted mean

$$\hat{x}_{i_{MAP}} = \frac{y_i \ \sigma^2 + \bar{x}_i \ \sigma_n^2}{\sigma^2 + \sigma_n^2} \tag{7.11}$$

The noise variance is locally estimated by comparing the DEM image with a rough guess of the noise free image resulting from a mean filter. Due to the inaccuracies carried by the rough guess, the resulting filtering is conditioned by the quality of this local estimation. Through the evidence maximization (marginalization of the prior model), an iterative process is designed to choose the most suited model among all the estimated ones. The selected model is used to filter the data. The convergence can be relatively slow, especially when DEM's texture varies abruptly. The model order defines the vicinity of analysis (see Fig. 7.1, page 123). Numerical limitations prevent the use

of large models (dimensionality reduction); for these reasons, only models order ranking from 3 to 5 are evaluated (respectively 6, 10 and 12 cliques). Other filter settings rely on the size of local analysis for the computation of first and second order of noise statistics ($\{\bar{x}_i, \sigma_n\}$ in Eq. 7.11).

Filter comparison and Parameter behavior

Table 7.1 presents some statistical features computed on the Cartosat DSM using different filter parameters settings. The features are subdivided into two categories according to the data analysed; either on the filtered DSM, or the difference between filtered and noisy elevation data.

If the evaluated filters depict various behaviors, larger vicinity analysis results for all of them in smoother data-sets (decrease of the absolute σ and $PSNR$, larger relative σ and MSE). Except for the wavelet soft-thresholding ($\Delta\ mean = -0.1639$), no bias is generated by the filtering process. This trend results from the modification of all wavelet coefficients. The amplitude of the filtering varies as well, ranking from small values for the hard-threshold filter (only small wavelet coefficients are removed), to important variations for the Median and Frost filters.

	Absolute (m)			**Relative**: $filter_i - RawData$ (m / dB)				
Filter	min	max	σ	$mean$	$\pm\delta$	σ	MSE	$PSNR$
Raw Data	186	1066	117.299	434.7615				
Median$_{r=2}$	208	1061	117.271	-0.0221	5.0	1.3266	1.7602	-24.6810
Median$_{r=3}$	209	1058	117.248	-0.0311	6.0	1.7527	3.0730	-30.2529
Gauss$_{r=2}$	207.7	1061	117.244	0.0003	4.1	1.2120	1.4689	-22.8465
Wavelet$_H$	186.4	1064.8	117.294	-0.0003	**2.0**	0.6702	0.4491	-10.9971
Wavelet$_S$	193	1060.9	117.269	**-0.1639**	3.6	1.1887	1.4399	-22.6724
Frost$_{r=2}$	208.8	1061	117.215	0.0004	5.3	1.6106	2.5939	-28.5327
Frost$_{r=3}$	209	1061	117.149	0.0005	**7.1**	2.2561	5.0901	-35.2741
MBDF$_{m=3,r=3}$	206.4	1061	117.280	0.0016	4.3	1.1517	1.3263	-21.8254
MBDF$_{m=4,r=3}$	206.9	1061	117.299	0.0058	3.8	1.0337	1.0687	-19.6651
MBDF$_{m=4,r=6}$	207.6	1061	117.279	-0.0051	4.6	1.2641	1.5980	-23.6889

Table 7.1: Filtering comparison using the Cartosat DSM. The column δ gives the bounded domain covering 99% of the values (the lower and upper 0.5% are not taken into account).

Such an analysis, summarized in Table 7.1, enables to emphasize the boundary limits of each method, however, it is not sufficient to discriminate and analyse more in detail the behavior of the Bayesian approach compared to the smoothing with a Gaussian. Using the same data-set and filters, a more accurate evaluation is achieved, based on the analysis of the filtering rejection according to the data roughness, as illustrated in Fig. 7.4. As pointed out in Table 7.1, results obtained with the Bayesian filter with fourth and fifth model order are nearly identical. Discrepancies are noticed only for important DEM coarseness. Using a larger window of analysis for the noise estimation gives larger rejections when the data variance grows -see Eq. 7.11- but the curves are similar for
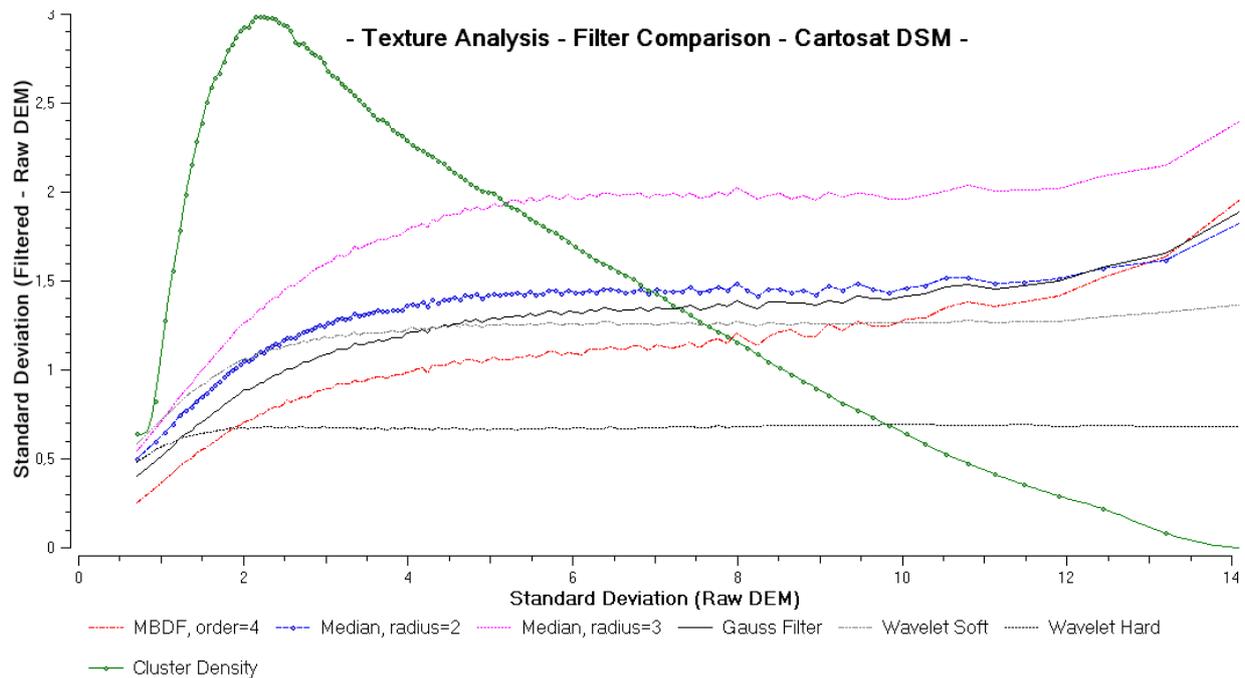
Figure 7.6: Comparison of Filters with Cartosat DSM data-set. (same protocol as Fig. 7.4)

the larger part of the data.

Compared to the median filter, its adaptive behavior is highlighted: textures derived from the median filter are larger with small variances but less important over higher ones while its overall amplitude is larger than the MBDF's one (Table 7.1: $\pm\delta_{MBDF_{m=4,r=6}} = 4.6$, $\pm\delta_{Median_{r=2}} = 5$). Results given by the Frost filter clearly state that such a SAR filter is not applicable for the current data-set.

The same protocol is followed (see Fig.7.6) by introducing three other methods and using a radius of 3 for the median filter. Independently from the DEM roughness, the latter performs a stronger filtering than the one with a radius of 2 and some "terraces" appear in extending the windows of analysis. For small values, Gaussian smoothing acts as a compromise between MBDF and median$_{r=2}$, however their behaviors are similar when dealing with important artifacts.

As expected, a correlation is observed in both diagrams between DEM roughness and filtering rejection. Results obtained with the wavelet methods significantly differ from all other approaches. Except for low values, the wavelet denoising delivers a homogeneous filtering. However, both multi-scale methods fail with large artifact removals and differ in terms of filtering amplitude.

Some examples

Since traditional signal quality measurements are not always adequate to evaluate the filtering operation, a subjective method, namely visual analysis, is necessary as "ultimate truth" and states a perceived quality.

(a) Cartosat Image   (b) Original DSM   (c) Median$_{r=2}$

(d) Gaussian Filter   (e) Bayesian Filter$_{m=4,r=3}$ (f) Wavelet (soft)

(g) Difference (d)-(b)   (h) Difference (e)-(b)   (i) Difference (f)-(b)
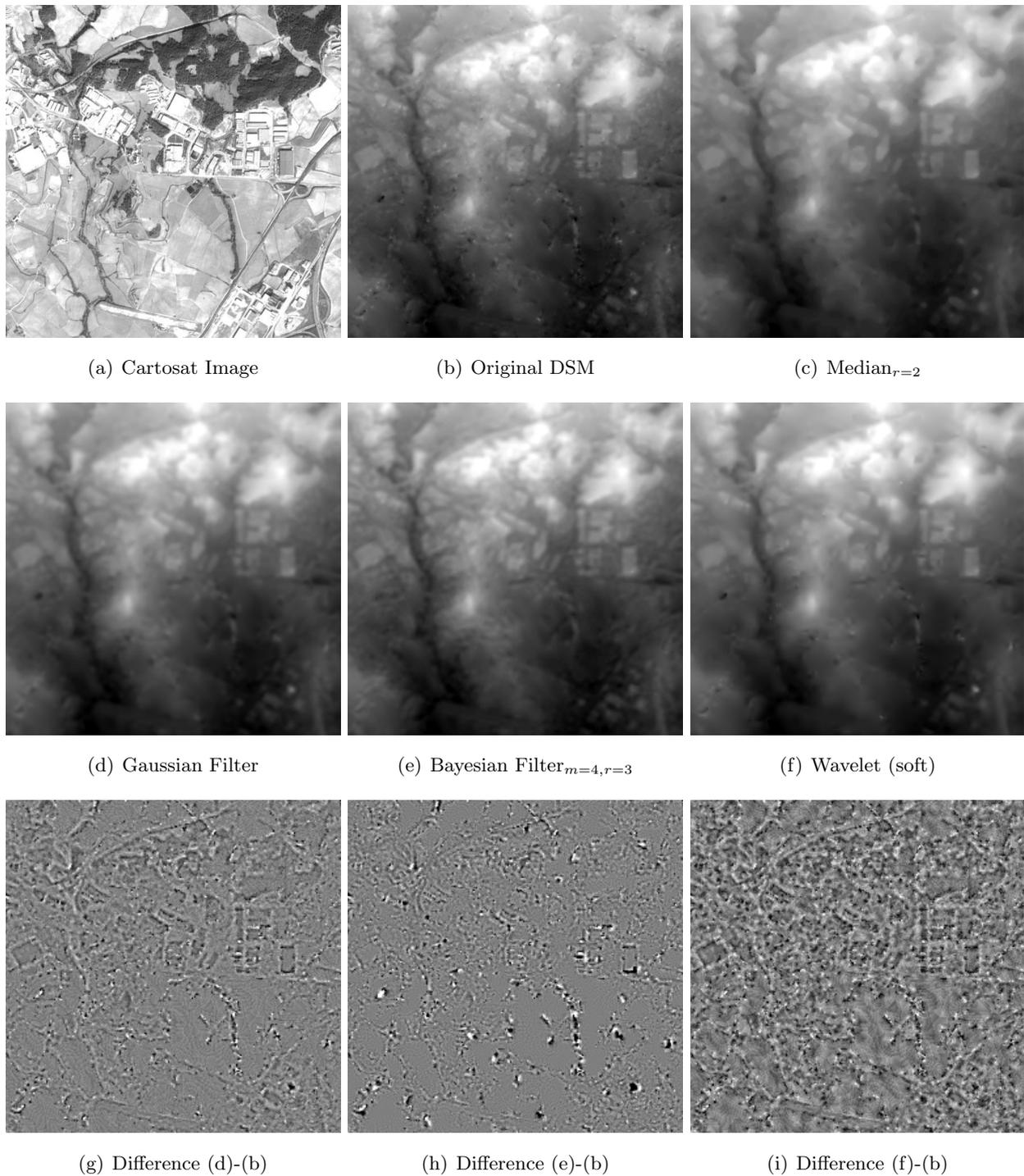
Figure 7.7: Filtering comparison applied on the original Cartosat DSM using Gaussian smoothing, MBDF (model order 4, local mean=3) and wavelet soft-thresholding filters. The area (size 256×256) presents a smooth relief (elevation range $314 - 398$ meters, $\sigma = 18.62$) and numerous interesting features (forest, road network, large buildings, etc.).

Optical stereoscopic Cartosat DSM

InSAR SRTM-X DEM

(a) Raw Data   (b) Wavelet Soft   (c) Median$_{r=3}$ (d) MBDF$_{m=3,r=3}$
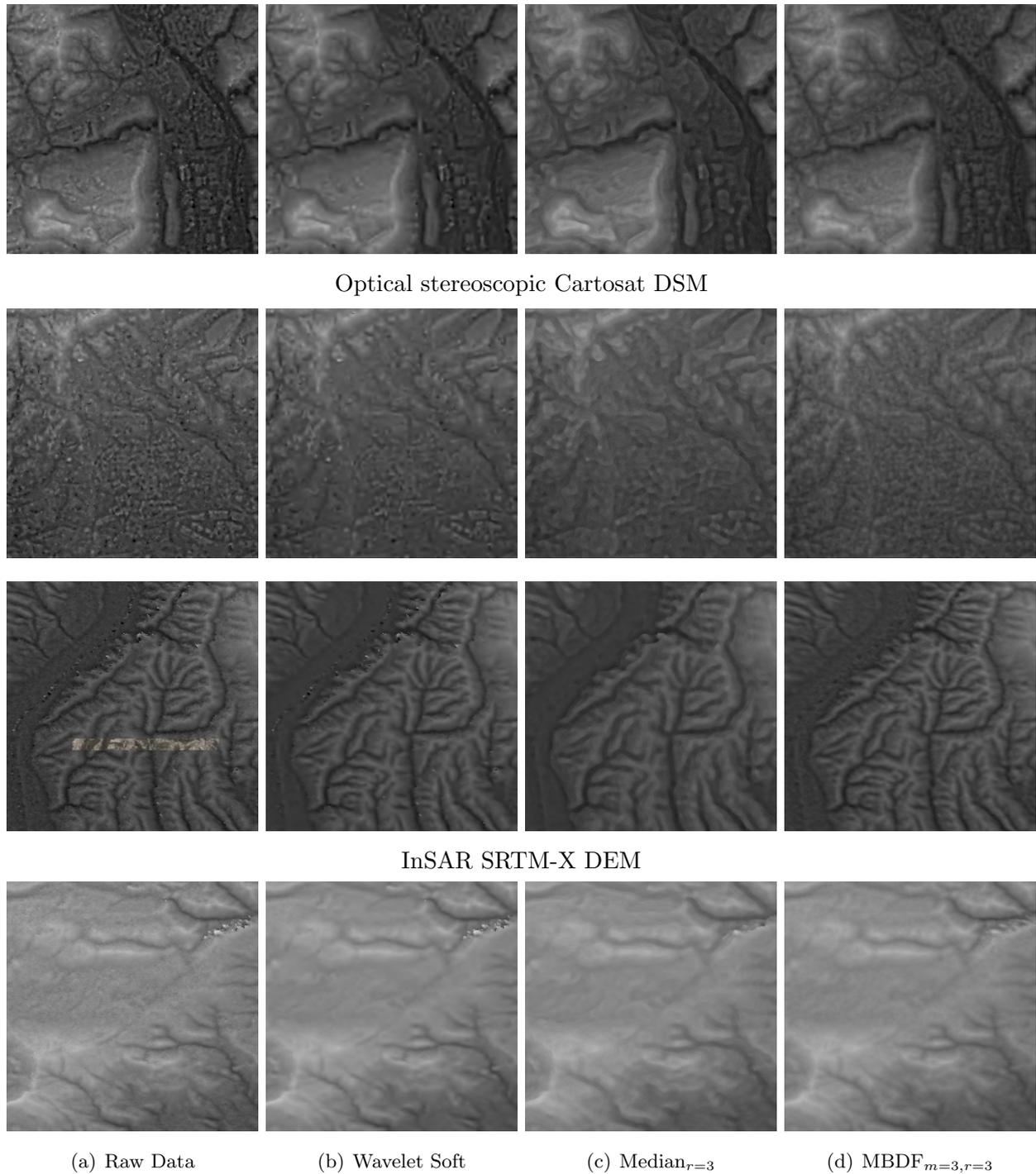
Figure 7.8: Local contrast applied on Cartosat DSM (Up) and raw SRTM-X DEM (Bottom) on their corresponding filtering processes with Wavelet Soft Thresholding (b), Median (c) and MBDF (d) filters (size 256×256). The small optical image stripe in the upper raw SRTM-X DEM indicates the location of the profile given in Fig. 7.10, page 137.

In Fig. 7.7, an area extracted from the Cartosat DSM is studied. It covers interesting features such as forests, road networks, large buildings. The evaluation of the filters in presence of such an information content is determinant. If the Gaussian smoothing delivers promising statistical results and achieves acceptable denoising, its behavior with object and structures is more delicate. Building structures, road networks are clearly recognizable in the DEM difference images. Those effects are even more pronounced on the wavelet filter. The latter generates some artifacts which, combined with the removal of high frequencies, give the impression of dealing with synthetic data.

We notice the presence of some typical smoothed area in the image difference for all filters and particularly for the wavelet filter. It mainly corresponds to fields with smooth relief where the image matching process fails. The missing information is re-interpolated from an available DEM (here the world coverage's SRTM-C DEM) to the resolution of the Cartosat data. When removing high frequencies, the weighted interpolation within the wavelet DSM is emphasized. In Fig. 7.8, further filtering examples based on the Cartosat DSM and InSAR DEM are presented.

7.2.3 External Validation

This section is related to the analysis of SRTM-X data since it was the only decametric data-set where another source of elevation data (topographic DTM) was at disposal.

| | **Absolute** (m) | | | **Relative**: $filter_i - RawData$, (m / dB) | | | | |
Filter	min	max	σ	$mean$	δ	σ	MSE	$PSNR$
Raw Data	22	1854	349.765	730.0794		-	-	-
Median$_{r=3}$	62	1841	349.217	-0.1433	-59.0:53.0	14.4641	155.3742	-41.8730
Gauss	64	1846.1	349.249	-0.0015	-49.0:47.0	10.0314	100.6283	-37.5163
Wavelet$_H$	73.4	1844.1	349.415	0.0194	-47.0:45.0	12.4621	155.3042	-41.8431
Wavelet$_{H(E)}$	22.9	1853	349.755	0.0001	-3.90:3.90	1.4612	2.1352	1.0380
Wavelet$_S$	72.3	1831.5	349.159	-2.8008	-47.7:39.6	12.1420	155.2734	41.8920
Wavelet$_{S(E)}$	37.2	1848.8	349.635	-0.3518	-9.69:8.86	3.2243	10.5197	14.9218
Frost$_{r=2}$	69.3	1842.5	349.037	0.0087	-58.2:54.7	12.0577	145.3882	-41.1960
GMap	65.5	1828.2	345.487	-8.1479	-49.4:32.6	9.3682	154.1520	-41.9857
MBDF$_{m=4,r=6}$	68.2	1848.1	349.288	0.0027	-52.7:51.7	11.0243	121.5345	-39.4039
DTM	67.7	1849.2	350.235	-4.9196	-141.:48.0	21.2454	475.5697	-53.0471

Table 7.2: Filtering comparison using SRTM-X DEM and topographic DTM (size 1024×1024).

Statistics with InSAR DEM and topographic DTM

Some statistical measures about the filtering of SRTM-X DEM are reported in Table 7.2. The studied area - extracted from Fig. 3.9, page 49 - represents almost the worst case for SRTM data (mountainous area, far range). Compared to the Cartosat data-set, the level of noise and artifacts

is higher, however this magnitude is not only related to the accuracy of InSAR process; SRTM data exhibits a 30 meters planimetric resolution over the alpine region, which is geologically a young mountain range characterized by sharp gullies, steeply and narrow ridges (in Table 7.1: {range 186:1066 m, $\sigma = 117.3$}$_{\text{Cartosat}}$, {22:1854 m, $\sigma = 349.7$}$_{\text{SRTM}}$.

As a consequence, most filtering operations result in stronger noise/artifacts rejections with amplitude variations almost 10 times larger $\delta_{MBDF} = \{-52.7 : 51.7\}$. For those reasons, investigations are conducted with the Gamma-map filter. Its parameter is set in order to obtain a MSE equivalent to the median filter. This method is also used to estimate the wavelet thresholds for the soft and hard methods. The classical "MAD" operator defined in Eq.7.10 appears to drastically underestimate the level of filtering ({$\Delta\sigma = 3.22$, $MSE = 10.52$}$_{Wavelet-H}$ compared to {$\Delta\sigma = 10.03$, $MSE = 100.63$}$_{Gauss}$).

Fig.7.9(a) underlines each filter operation for a given DEM roughness. If the noise rejection is similar for both filters, the wavelet filter cannot cope with a high local variance of the DEM, which characterizes the SAR artifacts and high relief transition. On the contrary, the Frost filter highly depends on the relief and leads to a rough response for high DEM variances. It constitutes a relevant example where the minimization of the mean square error is not the solution to every problem. Even if better results are obtained with the soft-thresholding technique, the use of wavelet thresholds according to a MSE equivalent to the median filter does not provide satisfactory results. The strength of the multi-scale filtering is too important for small variance but not enough to deal with InSAR artifacts.

The comparison between InSAR DEM and topographic DTM is emphasized in Fig.7.9(b), the latter being used as reference. Filtering rejections are plotted according to the DTM variance, which is, due to the interpolation, smooth. As expected, the raw SRTM DEM curve presents the steepest and roughest shape. Under weak DTM variances, the trend of the wavelet curve is above all the others, characterizing excessive filtering and the generation of artifacts.

Profile Analysis

The analysis of profiles, combined with an external reference, gives a sketch of the filtering performances as well as a better understanding of the data content (Fig. 7.10). Influence of relief generates geometrical displacements and topographic errors within the InSAR DEM. Shadows areas are clearly recognized in presenting simultaneously planimetric and altimetric errors confronted to the DTM.

Compared to the other filters, the adaptive behavior of the Bayesian filter is well illustrated in the profiles of Fig. 7.10. Gullies V-Valleys and sharp ridges are preserved while wide U-valleys are properly smoothed. It reinforces assumptions gathered on the superiority and interests of the MBDF filter when dealing with digital elevation models.
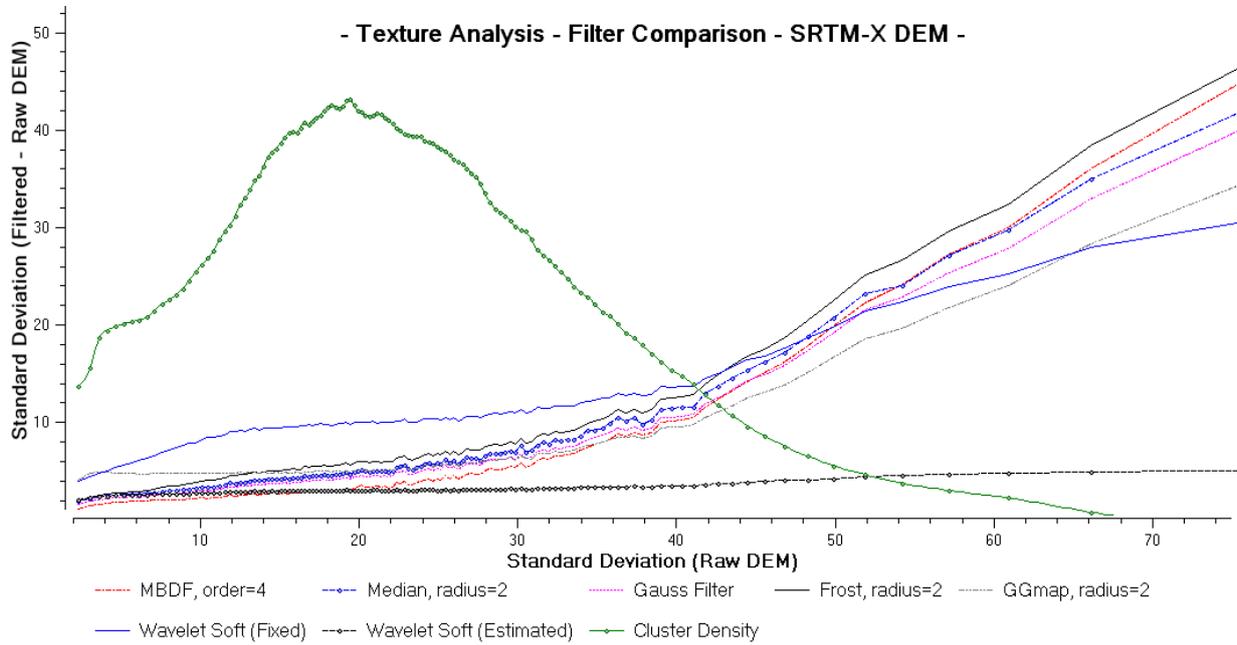
(a) $\sigma = 70$, curves ordering: $\sigma_{Frost} > \sigma_{MBDF} > \sigma_{Median} > \sigma_{Gauss} > \sigma_{GGmap} > \sigma_{Wave(F)} > \sigma_{Wave(E)}$
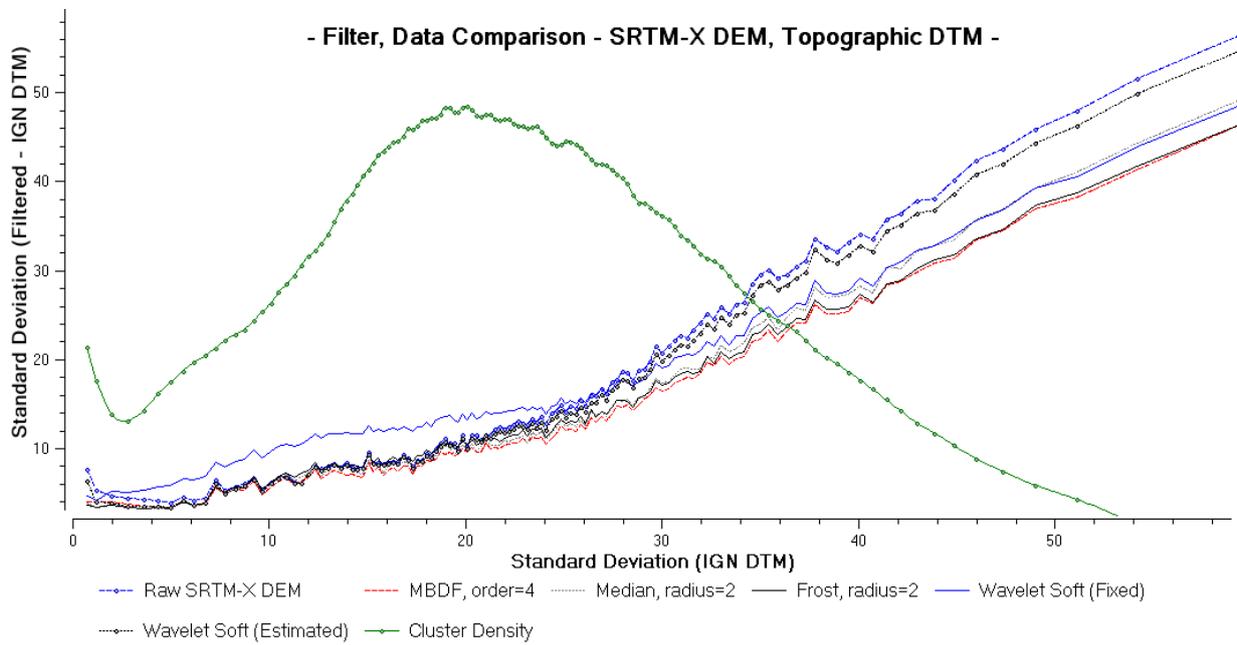
(b) $\sigma = 57.5$, curves ordering: $\sigma_{RawSRTM} > \sigma_{Wave(E)} > \sigma_{Median} > \sigma_{Wave(F)} > \sigma_{Frost} > \sigma_{MBDF}$

Figure 7.9: (a) Comparison of filters with SRTM-X DEM data-set. (b) Comparison of the InSAR DEM filtering according to the terrain "roughness", using a topographic DTM provided by IGN.
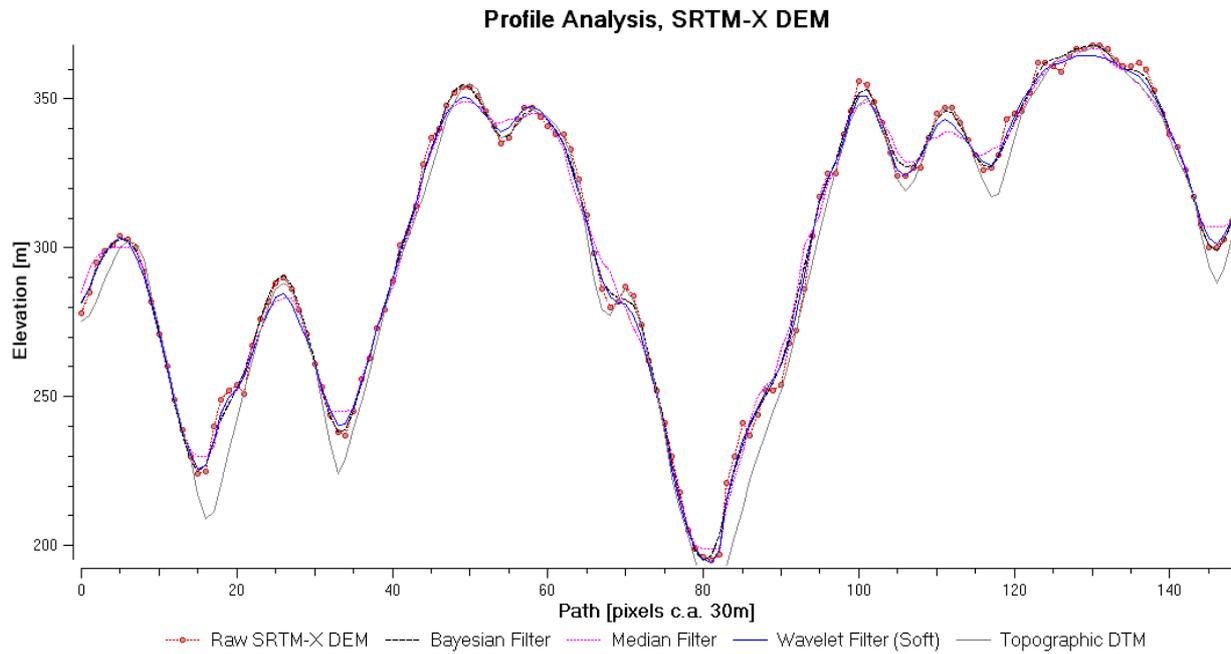
(a) Optical Image

(b) Topographic DTM
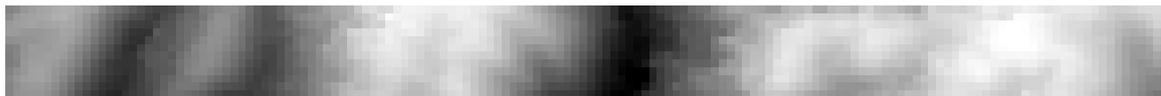
(c) Raw SRTM-X DEM
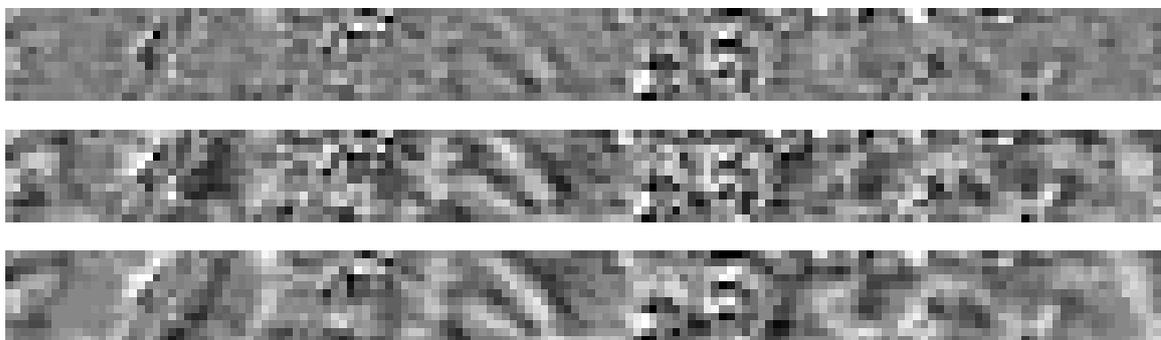
(d) Bayesian (up), wavelet (center), median (bottom) filters

Figure 7.10: Profile Analysis. The stripe (b) presents a topographic DTM, generated from level curves. Filtering comparison applied on a SRTM DEM (X-band) (c): (d) depicts median$_{r=3}$, Bayesian$_{m=4,r=3}$ filters and Wavelet soft thresholding ($\sigma = 5.71$).

7.2.4 Fractal Analysis

Several investigations have been conducted to describe DEMs by means of the fractal theory. In a limited resolution range, experiences show that specific fractal models fit the statistics of roughness measurements [45]. In this context, the simplest possible fractal curve - the fractional Brownian model (fBm) - is adopted (see fBm simulations in Fig. 7.11).

fBm

| (a) H = 0.1 | (b) H = 0.3 | (c) H = 0.6 | (d) H = 0.9 |
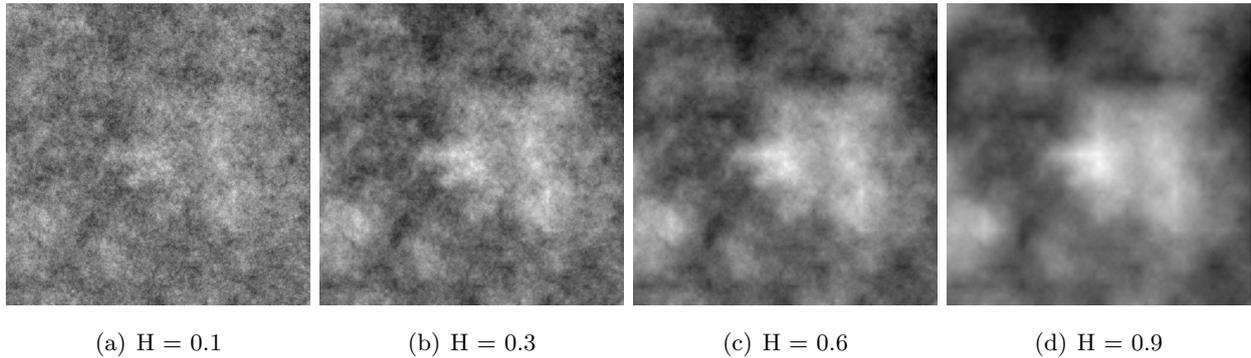
Figure 7.11: Simulation of a fractional Brownian model according to the Hurst exponent (H).

Due to the non-stationarity property of the fBm, its analysis is difficult. Indeed, the result of this model is related to the time-dependent analysis, whereas the self-similarity is related to the scale-dependent analysis. fBm describes a signal $B_H(t)$ characterized by the fact that the variance of the increment between t_1 and t_2 is proportional to the power of the time lag $|t_2 - t_1|$ [205],

$$E\left\{|B_H(t_2) - B_H(t_1)|^2\right\} \propto |t_2 - t_1|^{2H}, \qquad D = n + 1 - H \qquad H \in \ [\,0, 1\,[\qquad (7.12)$$

where n is the topological dimension of the space representation, H is called "Hurst Exponent" and linked to the fractal dimension D. H measures the agitation of a signal with values between 0 and 1. A low H indicates so much that the object analysed almost fills the next-higher dimension. As a consequence, the fractal dimension is a typical estimation of the signal's roughness; the deeper the value of D, the smoother the signal. In two dimensions, it reflects the roughness of a landscape. Most fractal analysis methods are focused on accurate determination of the fractal dimension.

To estimate the Hurst Exponent, the "Rescaled Range" analysis provides interesting results: however, such an approach is mainly devoted to the analysis of one-dimensional signals. The extension of such methods to deal with two-dimensional signals was not really convincing even when using different data organization schemes[3] to come back to 1D signal and therefore avoid discontinuities. If numerous other methods have been proposed, based on temporal, frequency or geometrical approaches, we mainly focused our investigations on multi-scale analysis.

---

[3]Image rows (resp. columns) are alternatively scanned back and forth to form a single one-dimensional representation of the sampling windows. Other attempts are considering quincunx neighborhood whose nodes express the distance from its origin to compute local Hurst coefficient approximations.

Relation to Wavelets

A method for the estimation of D is based on the wavelet decomposition of a fBm. The wavelet decomposition of an image can be interpreted as a signal decomposition in a set of independent, spatially oriented frequency channels (see section 2.1.2, pages 10- 12). Using a wavelet decomposition on a non-stationary process such as fBm whitens the signal: the transformed samples at a given resolution 2^m become weakly correlated and stationary. Furthermore, Mallat demonstrates in [144] that for a fBm, the variance of detail (1-D) signals follows

$$\text{Var}(D_j) = 2^{2H+1}\text{Var}(D_{j+1}) \tag{7.13}$$

with D_j being the detail signal at resolution j as shown in Fig. 2.4, page 12. The direct estimation of the fractal dimension is then based on the fact that the variance of the wavelet coefficients (factorization of Eq. 7.13) at level 2^m is given by

$$\text{Var } x_n^m = \sigma^2 2^{-\gamma m}, \quad \sigma^2 > 0 \tag{7.14}$$

with σ^2 a constant depending on the fractal decomposition used and $\gamma = 2H + n$. γ can be obtained using a maximum likelihood estimation applied on the variance computed at different levels 2^m. An alternative relies on the fact that for real fractals, a log-log plot of $\text{Var } x_n^m$ against the resolution level m gives a straight line with a slope proportional to $-\gamma$.

Experiments

To estimate the fractal dimension, the DEM is first decomposed into a 4-level wavelet basis using the symmetric bi-orthogonal Daubechies 9/7 wavelets. For each complete transform, an estimation of the variance is computed at every detail level. Results derived from the wavelet transforms are either averaged together and then plotted on a log-log diagram or drawn individually. Linear regressions on the curves generated by both methods provide similar results (difference of the order of 10^{-4}).

Some simulations of a fractional Brownian model according to the Hausdorff measure are presented in Fig. 7.11. Such synthetic data are used to test the quality of the estimated fractal dimension (see Table 7.3). The computed fractal dimensions of simulated fBm images are slightly overestimated. A bias is present but a small standard deviation is obtained which is, in the purpose of data comparison and evaluation, determinant.

In the literature, the typical range for elevation data observed lies between 2.2 and 2.4 (we obtained $D = 2.18$ for the Cartosat DSM) and rougher data-sets reach 2.6 as in the case of the mountainous InSAR DEM (2.52). For the topographic DTM, a measure of 1.86 is reported, below 2. It can be argued by the fact that smoother interpolations of level curves neither match exactly a real landscape nor the properties of fractals over the different scales of analysis (few high frequencies).

Among denoising methods, for both test sites, the wavelets approaches present the higher fractal dimensions with results far away from the others. It traduces a weak filtering and still the presence

simulated fBm		real data			
Accuracy Control		**SRTM-X DEM**		**Cartosat DSM**	
Theoretical	Estimated	size 1024×1024		size 2048×2048	
2.05	2.09739	Original Data	2.521390	Original Data	2.181578
2.10	2.15476	IGN DTM	1.868301	$Median_3$	2.019069
2.15	2.18838	$Median_2$	2.056595	$Median_2$	2.024320
2.20	2.25153	Gauss	1.759579	Gauss	1.932136
2.25	2.29939	$Frost_2$	1.747191	$Frost_{r=2}$	1.919108
2.30	2.33335	Wavelet $Soft_F$	2.112751	Wavelet $Hard_E$	2.046473
2.35	2.39824	Wavelet $Soft_E$	2.446424	Wavelet $Soft_E$	1.963991
2.40	2.47690	$Mbdf_{m=4,r=3}$	1.855768	$Mbdf_{m=4,r=3}$	1.939676

Table 7.3: Fractal analysis using simulated fBm, SRTM-X DEM and Cartosat DSM.

of artifacts that keep the signal extremely disturbed. An adaptive determination of the threshold should lead to more efficient results.

Although the estimation of the filtered data is smaller than expected, their relative behavior is interesting to validate hypotheses about filter comparison, evaluation and more generally about the data itself. Using this signal's statistical feature, filtering methods or data-sets presenting un-typical characteristics are distinguished. Even when taking into account these information discrepancies between both products, the fact that the roughness of the filtered DEM is close to IGN DTM's reinforces the assumptions about the quality of the MBD filter.

Summary

Several classical filters have been presented in the purpose of enhancement of digital elevation models using quantitative and qualitative analyses. Such an evaluation was also suitable to describe more in detail the EO data tested and their corresponding elevation generation processes.

The adaptive method based on Bayesian analysis appears to be the most appropriate approach to filter DEM. The noise restoration is very efficient due to the complete and local modeling of the noise. The global aspect of the DEM is more realistic. Nevertheless, the filter process limits seem to be reached with correlated noise or in the presence of important artifacts carried by the phase unwrapping stage. The only solution for such areas is to down-sample the data, which results in a resolution decrease, but leads to a better data homogeneity.

Despite the performance of the filtering process, limitations arise with very high resolution data that contain structures. Indeed, texture or radiometry description based on statistical modeling fails to fully capture and represent non-linear features. To significantly enhance the DEM, complementary information must be gathered from other sources.

Image Information and DEM Aggregation

Abstract

To remove artifacts and noise from an InSAR DEM, the non-stationary Bayesian filter presented in the last chapter is efficient to cope with most of the artifacts and noise in the DEM data, while preserving structural information. However, a denoising process is not sufficient to properly enhance EO elevation data.

Three main limitations emerge; (1) The reconstruction methods (InSAR, stereoscopy) fail to extract elevation for specific areas. (2) High resolution DEMs contain non linear shapes and severely limit performances of processes based on statistical modeling. We address a short summary about techniques investigated to deal with very high resolution DSM, mainly based on LIDAR data. (3) Land-cover or man-made structures are easily recognized in an optical image. Their geometry encapsulated in the DEM differs from our implicit perception and generally leads to inconsistencies and unrealistic 3D rendering.

To improve this, a DEM regularization is achieved using only a remote sensing image. It consists in a sequence of operations to: (a) extract and model meaningful information from a complementary image (see chapters 4), (b) classify the collected structures among a set of user-thematics, (c) define elevation modeling for the thematic studied, (d) estimate the region's 3D structure and (e) integrate them in the digital elevation model using a B-spline interpolation method.

We emphasize the interest of the proposed method to improve digital elevation models, using a SPOT 5 multispectral image and a SRTM-X DEM.
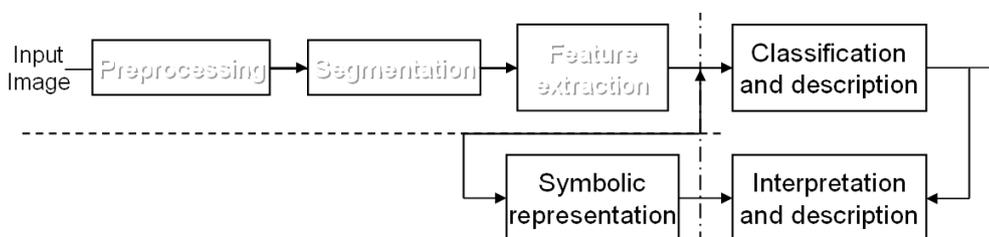
8.1 Very High Resolution DEM

In the case of very high resolution data which exhibit information content such as structures, edges or objects, stationarity and even the relaxed assumption of local stationary process are not fulfilled anymore. It can result in erroneous noise estimation, inadequate prior models, and lead to unrealistic filtering behavior.

8.1.1 LIDAR Data

LIDAR (Light Detection and Ranging) system is becoming an alternative to classical aerial photogrammetry techniques for acquiring terrain elevation data. Like SAR sensor, it's an active system that emits a signal with generally a wavelength in the infra-red domain and receives the reflected signal intensity. The amplitude of the collected returned signal depends on the incidence angle and the surface characteristics.

For a given emitted pulse signal, multiple returns can be collected. Discrepancies between the returned pulses emphasize surface characteristics or particular thematics such as vegetation where the foliage partially reflects the signal. Coupling a LIDAR sensor on board of an airborne with a differential GPS and an inertial navigation sensor, a data-set of irregularly-spaced 2.5D points can be generated. This set describes the top reflective surface which includes buildings, man-made objects, roads, bridges and partially the vegetation. Since the surface differs by its information content from a bare terrain modeling (DEM), a specific formulation was established to denote such surface elevation, called DSM.

An essential issue is to classify the acquired point clouds that can reflect bare-terrain, vegetation or man-made surfaces. This classification process is called filtering. It differs from the filtering methods previously presented where the aim was to remove noise and local artifacts that pervade the data. In the case of VHR LIDAR data, the filtering notation is not restricted the removal of noise and outliers, but encapsulates the extraction of some objects superposed to the bare terrain. The "filtering" process is motivated by two purposes: either retrieve the bare earth in removing the overlaid objects, or extract the corresponding objects (e.g. urban reconstruction).

8.1.2 DSM Information Extraction

Various methods are found in the literature; they mainly differ in the modeling of the transitions between bare terrain and object surface. Two categories are distinguished; the points clouds are either individually processed comparing to their closest neighbors by means of slope or height differences, or the homogeneousness of a group of points is considered as a criterion to merge points that belong to the same surface characteristics. The last category denotes segmentation methods and generally follows a region growing scheme based on seed initialization.

(a) Optical image

(b) LIDAR DSM pseudo-color

(c) LIDAR DSM
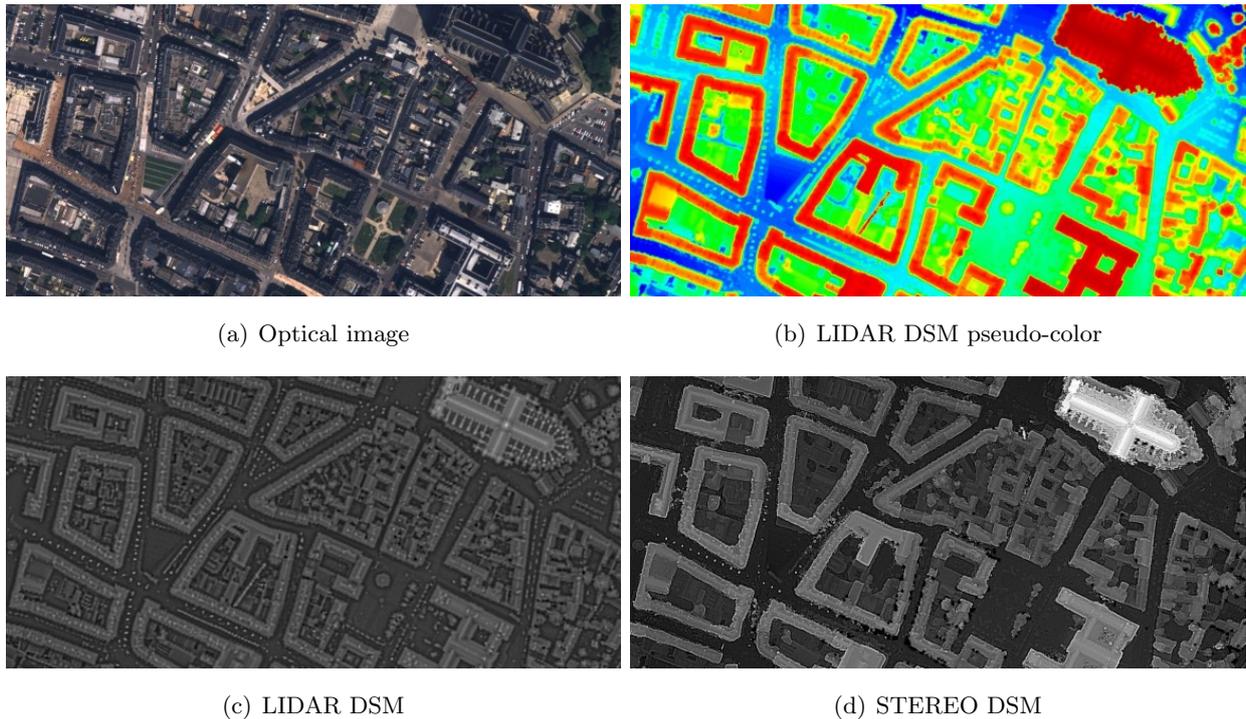
(d) STEREO DSM

Figure 8.1: Information content on sub-metric LIDAR (b)-(c) and stereo (d) DSM. (a) Optical image (5782×2774, $\sim 10\ cm$ resolution). (b) LIDAR DSM in pseudo color, warm colors depict high elevation while cold ones represent low elevations. (c) and (d) are respectively local contrast adjustments applied on LIDAR (50 cm resolution) and stereo DEM (20 cm resolution).

Extraction of the terrain points from LIDAR elevation data

In [111], [206], morphological approaches are investigated for the classification of laser data. Lowest points are considered as bare-terrain under the assumption of a locally flat terrain. By means of the mathematical morphological operators "erosion" and "dilation", the points clouds which do not belong to the terrain are identified. The ability of morphological filters to provide a good estimate of the elevation terrain depends on the windows size and the distribution of the objects in the data (building or vegetation areas).

A linear least-square interpolation scheme is used in [120] to extract the points clouds that belong to the bare earth in forested areas by iteratively fitting surfaces and weighting the points clouds according to their inverse distance to the fitted surface. A similar iterative approach, improved by an elimination of the trend, is proposed in [165]. If suitable results are obtained for objects of relatively small extension in the case of smooth terrain, these algorithms which work on small vicinity fail in separating the terrain from large objects.

A progressive method is suggested in [217] to estimate locally terrain elevation threshold in varying iteratively the size of the morphological filter. Active contour models were applied in [60] to estimate the bare terrain elevation. In [190], a segment-based classification approach which

exploits the neighborhood height differences is proposed to better preserve the bare Earth surface discontinuities. According to the authors, this method provides satisfactory results in urban areas. The same authors dressed in [189] a comparison of the different LIDAR filtering techniques.

The presented filtering techniques aim to retrieve the ground terrain elevation from LIDAR data. Such a post-processing step is determinant since DEMs are often the basis of man-made object or vegetation extraction. Those applications consist in relative elevation analyses where the ground surface acts as reference or normalization surface[1].

Object Extraction from LIDAR elevation data

Over the last decade, the availability of very high resolution LIDAR data over urban areas has stimulated much research to extract and reconstruct man-made structures like buildings, roof or road networks. In the literature, many investigations are described, based only on LIDAR data.

In [64], the RANSAC algorithm [62] is used to detect roofs modeled as planar segments. In [157], Ortner provided a stochastic method to extract rectangular buildings. Based on a spatial point process framework, urban areas are modeled as configurations of an unknown number of rectangles. An energy is defined, taking into account elevation data and some geometrical constraints about the building dispositions. The algorithm converges to a configuration of rectangles (i.e. "buildings") by minimizing the formulated energy using simulated annealing.

Application with Geodesic Contours

The case of the extraction of buildings was chosen to investigate the potential of snake methods. We implement the geodesic contour approach proposed in [112] (described page 63) which aims at maximizing the average contrast (Eq. 4.8) along an arclength parametrized curve γ.

To compute the first and second derivatives, the recursive Canny filter suggested by Deriche in [48] is implemented. It consists in two sequential recursive 1D functions for each direction $\{\Delta X, \Delta Y\}$: edge and smoothing operators, the former being applied perpendicular to the computed direction. The parameter α - the width of the impulse response - is set to 1, it represents a good compromise between detection and localization of edges.

The curve evolution, shown in Fig. 8.2, is based on a gradient descent. Since the number of points in the curve is maintained constant during the evolution process, an appropriate length is required. The contrast function is expressed in the form $g(t) = |t|^\alpha$ with $\alpha > 0$. The selection of a suitable α value is difficult; when α is low (e.g. $\alpha < 1$), reliable contours are extracted but the convergence is slow and sensitive to noise.

When α gets larger (e.g $\alpha > 1.25$), the sensitivity to low contrasted areas is lost and the re-

---

[1]In [143], by means of subtraction between DSM (airborne InSAR DSM sampled to 25cm resolution) and DTM data-sets, we proposed to produce a cartography of damages occured during the storms of 1999 in forestry areas.

(a) Raw LIDAR DSM   (b) Initialization, $e = 79.84$ (c) $i = 12$, $e = 294.24$ (d) $i = 39$, $e = 432.59$

(e) $i = 63$, $e = 524.37$ (f) $i = 111$, $e = 671.84$ (g) $i = 300$, $e = 764.53$ (h) $i = 675$, $e = 828.76$
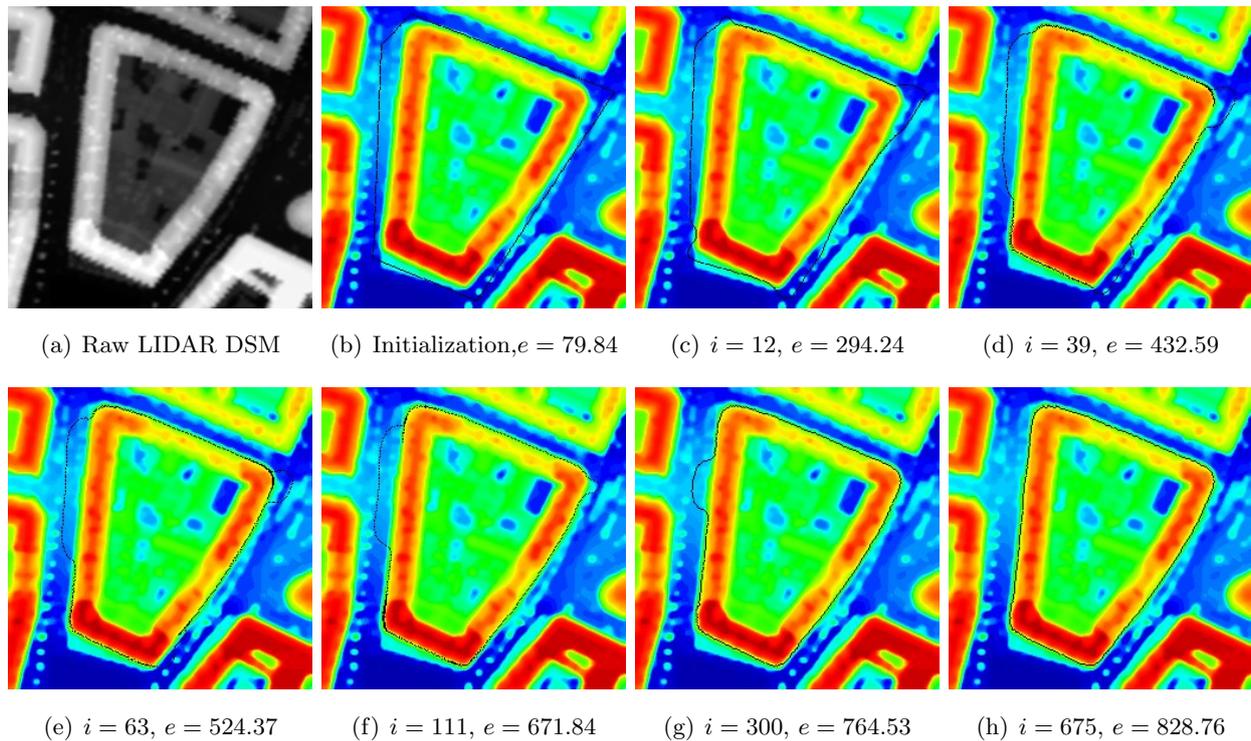
Figure 8.2: Geodesic contour evolution on a LIDAR DSM (range: 52.3-58.8 meters, image size 256×256 cropped from Fig.8.1(b)). The black curve represents the extracted contour at iteration i with a contrast power of 1.5, e denotes the computed energy (Eq. 4.8).

parametrization of γ becomes difficult since no topological changes are handled. As a consequence, some areas present thick curves (Fig. 8.2(f) with $\alpha = 1.5$); those areas with sharp boundaries agglomerate an important density of points from the curve with multiple overlapping and intersections. In the same way, the curve can follow the same contour in both directions (top-right image part of Fig. 8.2(e)), which results into the delineation of region parts with a zero thickness.

To reduce those effects as much as possible while preserving the curve evolution, some heuristics methods are introduced. The curve is sampled on a fixed grid in order to take into account such topological properties when the curve is re-parametrized. It mainly includes contours following and morphological operators, applied every $n \approx 50$ iterations.

If geodesic contours are very appealing methods to optimize the accuracy of boundaries, for the extraction and the detection of contours, such approaches are still extremely difficult (initialization stage, sensitivity to local maxima, no topological change handling) for the extraction and the detection of objects.

Fusion of Complementary Information

To improve extraction and/or reconstruction processes from elevation data, numerous papers - mainly concentrated on urban areas- pointed out interests and synergies gathered by the availability

of complementary information. In [80] the introduction of cadastral maps give an accurate building localization to model buildings detected from LIDAR data. The use of cadastral maps information to estimate roof from LIDAR data using the RANSAC algorithm was further investigated in [56]. In a similar way, some authors focused on different topographic objects. Reconstruction of road networks is investigated in [59] using topographic maps combined with LIDAR data.

8.1.3 DEM Integration

A limitation to the filtering processes presented in chapter 7 is the recognition of 3D structures in the images. Its corresponding geometry, encapsulated in the DEM, differs from our implicit perception and generally leads to unrealistic 3D rendering even if globally the accuracy is very high. Additional knowledge extracted from optical or SAR images can be relevant to define regularization constraints to correct DEMs in the same way as a priori knowledge is ingested in topographic DTM.

Semantic DEM Correction

Koch proposed in [113], [114] to investigate the semantic content of the objects depicted in the elevation data and to verify their correctness and consistent representation. The importance of the semantic description is emphasized by the example of a lake. The shape structure is not only driven by internal constraints (the lake surface should be flat) but also by relation dependencies with the neighborhood objects. Depending on the semantics, geometrical equality and inequality constraints are derived to model the object represented by means of local slopes and neighborhood relations.

The extension of the object and its given semantic is given by a GIS vector database and described by a closed polygon. Elevation data are stored in form of a triangular irregular network. The homogenization of the two data-sets is driven by the defined constraints and formulated as an inequality constrained least square adjustment for all the objects points and the corresponding neighbored DTM TIN points.

The object integration in the TIN data is achieved using a constrained Delaunay triangulation. New points are inserted in the elevation model and correspond to the intersection between respectively the TIN DTM and the object borders. Along the object border, TIN points are re-triangulated to preserve the representation and locally make the meshes denser.

8.2 Image Interpolation

Although interpolation is a classical task in image processing, it represents a challenging problem to circumvent the compromise between shape smoothing and approximation accuracy. Among the various techniques investigated for the estimation of a continuous two-dimensional function from its sample values, B-splines techniques appear extremely promising.

8.2.1 B-spline Interpolation

A severe limitation of polynomial interpolation (even with high orders) relies on the non-continuity of the interpolated curve in all its derivatives. Splines circumvent such drawbacks by the integration of additional derivative constraints at the grid points. Since B-splines are separable, we present such a parametric curve $S(t)$ in the case of one-dimensional signal. A B-spline is a linear combination of piecewise polynomial functions (basis B-spline $b_{i,n}$ of degree n) constructed along a sequence (t) of m knots $< P_i >_{i=0...m-1}$, called control points ($i \leq m - n + 1$)

$$S(t) = \sum_i^{m-n-1} P_i b_{i,n}(t) \quad , \quad \forall t \in [t_i, t_{i+n+1}[\quad b_{i,n}(t) > 0 \quad , \quad 0 \quad \text{otherwise} \tag{8.1}$$

Its shape is determined by the position of the knots which bound the curve (contained in the convex hull of $< P_i >$). We clearly see in Eq. 8.1 that any control point P_i modification affects $S(t)$ only locally since $b_{i,n}(t)$ is zero outside its compact support). The elements $b_{i,n}$ can be seen as weight contribution for each control point, based on the distance to the location of interpolation. A classical uniform "cubic B-spline" basis function, derived from the Bernstein polynomial is

$$
\begin{array}{c|c|c|c|c}
n = 3 & b_0(t) & b_1(t) & b_2(t) & b_3(t) \\
\hline
t \in [0, 1[& \frac{1}{6}(1-t)^3 & \frac{1}{6}(3t^3 - 6t^2 + 4) & \frac{1}{6}(-3t^3 + 3t^2 + 3t + 1) & \frac{1}{6}t^3
\end{array}
\tag{8.2}
$$

If the implementation of fast (matrix factorization) and efficient B-splines techniques to interpolate images is achieved with a minuscule amount of mathematical machinery, the task is harder in case of irregularly spaced data points. An algorithm, based on multilevel B-splines, is presented in [129] for the interpolation of scattered data. A coarse-to-fine hierarchy of control lattices is used to generate a sequence of bi-cubic B-spline functions whose sum approaches the desired interpolation function. A B-spline refinement is used to reduce the sum of these functions into one equivalent B-spline function. Let us give more details about the algorithm [129] we implemented.

B-spline approximation

Considering a set of scattered points $< P = \{x, y, z\} >$, its approximation is formulated using a uniform bi-cubic B-spline function f defined by a control lattice Φ [2] overlaid on domain Ω

$$f(x, y) = \sum_{k=0}^{3} \sum_{l=0}^{3} b_k(x - \hat{x}) b_l(y - \hat{y}) \ \Phi(\hat{x} + k - 1, \ \hat{y} + l - 1) \tag{8.3}$$

where $\hat{x}$ denotes the largest integer satisfying $\hat{x} \leq x$ and b_i are the basis functions presented in Eq. 8.2.

To obtain the interpolation function, the values of the control lattice $\Phi(k, l)$ have to be estimated. $\Phi(k, l)$ is solved by considering only the points in its 4×4 neighborhood $< P(k, l) >_{4 \times 4}$.

---

[2]Ω denotes the rectangular bounding domain of $< P >$ (size $\{m+1, \ n+1\}$). The control lattice Φ is $\{m+3, \ n+3\}$ big. The node $\Phi(0, 0)$ coincides with $\Omega(1, 1)$.

Given a point $P_c = P(x_c, y_c, z_c)$ of $< P(k, l) >$, the values $\Phi(k, l)$ in Eq. 8.3 must satisfy $f(P_c) = z_c$, an approximation is computed using the pseudo-inverse

$$\Phi(c) = \frac{w_c \, z_c}{\sum_{a=0}^{3} \sum_{b=0}^{3} w_{ab}^2} \quad , \text{ with } \quad w_c = w_{kl} = b_k(x_c - \hat{x}_c) \, b_l(y_c - \hat{y}_c) \tag{8.4}$$

Given a node $\Phi(i, j)$, based on Eq. 8.4, an approximation $\Phi(c)$ is computed for each point of $< P(i, j) >_{4 \times 4}$. A least square solution to $\Phi(i, j)$ is then derived by minimizing the local approximation error $\sum_c (w_c \Phi(i, j) - w_c \Phi(c))^2$ and leads to

$$\Phi(i, \, j) = \frac{\sum_c w_c^2 \, \Phi(c)}{\sum_c w_c^2} \tag{8.5}$$

Multilevel B-spline approximation

The direct interpolation of scattered data using a lattice of control points is conditioned by the size of the lattice. Too small, it results in important smoothing; too large, a good accuracy is obtained, but only on the scattered points. To propagate the scattered data across the image, multilevel B-splines are used. A single control point lattice Φ is replaced by a hierarchy of control lattices $\{\Phi_{0...p}\}$ [3] (defined from level 0 to p and overlaid on Ω). The approach starts by computing the coarsest lattice Φ_0 and continues incrementally to Φ_p using for each level k, a function f_k (derived from Eq. 8.5) to approximate the data-set $< \Delta^k P = P(x_c, y_c, \Delta^k z_c) >$ where

$$\Delta^k z_c = z_c - \sum_{i=0}^{k-1} f_i(x_c, y_c) = \Delta^{k-1} z_c - f_{k-1}(x_c, y_c) \text{ , with } \Delta^0 z_c = z_c \tag{8.6}$$

The resulting interpolation $f = \sum_{i=0}^{p} f_i$ represents the sum of finer and finer approximations.

Instead of a sequence of nested B-splines, a single B-spline function is preferred for saving computational time and space. A B-spline refinement is used to reduce the sum of these functions into one equivalent B-spline function. A lattice Ψ_h [3] is refined into the next lattice in the hierarchy Ψ'_h (identical size as Ψ_{h+1}, both Ψ and Ψ' representing the same B-spline) using

$$
\begin{bmatrix} \Psi'_{2i,2j} \\ \Psi'_{2i+1,2j} \\ \Psi'_{2i,2j+1} \\ \Psi'_{2i+1,2j+1} \end{bmatrix} = \frac{\Psi_{i,j}}{16} \begin{bmatrix} 9 \\ 6 \\ 6 \\ 4 \end{bmatrix} + \frac{1}{32} \begin{bmatrix} 3 & 3 & 3 & 3 \\ 2 & 0 & 12 & 2 \\ 0 & 2 & 2 & 12 \\ 0 & 0 & 8 & 8 \end{bmatrix} \begin{bmatrix} \Psi_{i,j-1} \\ \Psi_{i-1,j} \\ \Psi_{i+1,j} \\ \Psi_{i,j+1} \end{bmatrix} + \frac{1}{64} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 4 & 0 & 4 \\ 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 16 \end{bmatrix} \begin{bmatrix} \Psi_{i-1,j-1} \\ \Psi_{i+1,j-1} \\ \Psi_{i-1,j+1} \\ \Psi_{i+1,j+1} \end{bmatrix} \tag{8.7}
$$

The B-spline refinement avoids the definition of functions f_k at every level which have to be summed to obtain the approximation function f. The summation is done at every level k, the lattice Φ_k is computed with Eq. 8.5 on the data set $< \Delta^k P >$ (Eq. 8.8) and the control lattice Ψ_{k-1} is refined to Ψ'_{k-1} using Eq. 8.7. Both lattices, of the same size, are summed to update the lattice Ψ_k to be used at the next level

$$\Psi_k = \Phi_k + \Psi'_{k-1} \text{ , with } \Psi'_0 = 0 \tag{8.8}$$

---

[3] Let Φ_h be a lattice of size $\{m+3, n+3\}$, Φ_{h+1} is $\{2m+3, 2n+3\}$ big. $\Phi_h(i, j)$ coincides with $\Phi_{h+1}(2i, 2j)$.

(a) Original data   (b) Sampling   (c) Interpolation   (d) Difference (c)-(a)
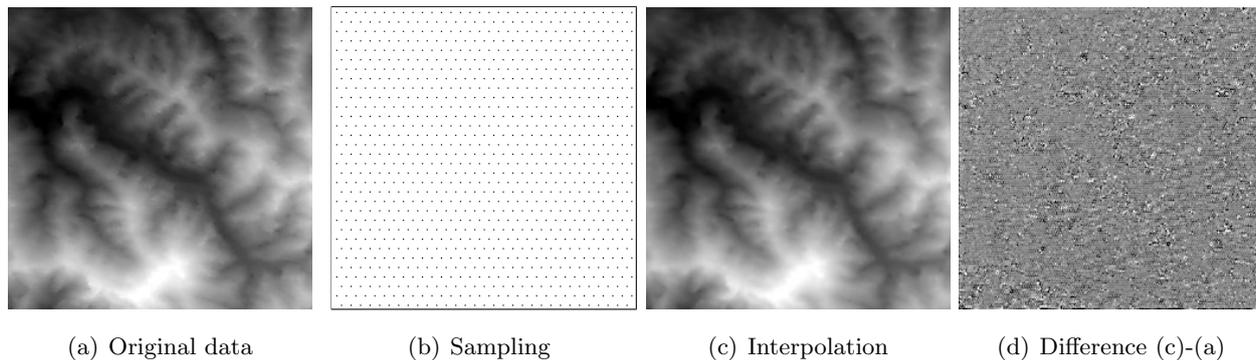
Figure 8.3: 5-level B-spline refinement (c) applied on an interleaved sampling (s=2) of a regular grid of 3D points ((a) 256×256, elevation range: $153 - 465$ meters, $\sigma = 68.9$) using an initial lattice size of $\{m_0 = 16, n_0 = 16\}$. (d) depicts the comparison with the original data-set.

The resulting control point lattice Ψ_o expresses the coefficients of the interpolation function f.

8.2.2 Application to Elevation Data

Though this method was initially presented for the purpose of image warping, it can also be adapted to handle elevation data. We tested the B-spline approach to smooth DEMs and to fill missing information in stereoscopic DEMs.

Surface Approximation and Incidence of the Control Points Lattice Size

An example of interpolation from a set of regular 3D points (here a SPOT 5 DEM) is given in Fig. 8.3. Based on an interleaved sampling of the 3D data-set, a 5-level B-Spline refinement using an initial lattice size of $\{m_0 = 16, n_0 = 16\}$ is adopted and compared to the original data-set. Difference of elevations are reported in a range of -7.3 to 5.4 meters. If a small bias is observed ($mean_\Delta = -0.63$ meters), the uniformity of the approximation is important with a standard deviation of $\sigma_\Delta = 1.78$. Such an approach can be used as smoothing operator.

The incidence of the initial lattice Ψ_0 size is presented in Fig. 8.4. Using a 3-level refinement and Ψ_0 with different sizes, several control point lattices Ψ_3 are generated, characterized by decreasing level of smoothness (according to the density of control points)[4].

This example emphasizes the interpolation behavior of the presented method. The level of smoothness is controled by the resulting lattice size Ψ_o obtained in varying either the number of refinement levels o or the size $\{m_0, n_0\}$ of Ψ_0. Based on our experiment, the option of relatively small size for Ψ_0 (e.g. ≤ 16) combined with high values of o is preferred to a large lattice Ψ_0 with few refinement levels. This observation is particularly justified in the case of scattered points with

---

[4]Since both control point lattice and the resulting surface present the same size, a representation of the 3D points set is given in Fig. 8.4(d).

(a) $\{m_0 = n_0 = 8\}$, $o = 3 \Rightarrow \Psi_3(m_3 = n_3 = 32)$

(b) $\{m_0 = n_0 = 16\}$, $o = 3 \Rightarrow \Psi_3(m_3 = n_3 = 64)$

(c) $\{m_0 = n_0 = 32\}$, $o = 3 \Rightarrow \Psi_3(m_3 = n_3 = 128)$

(d) $\{m_0 = n_0 = 64\}$, $o = 3 \Rightarrow \Psi_3(m_3 = n_3 = 256)$
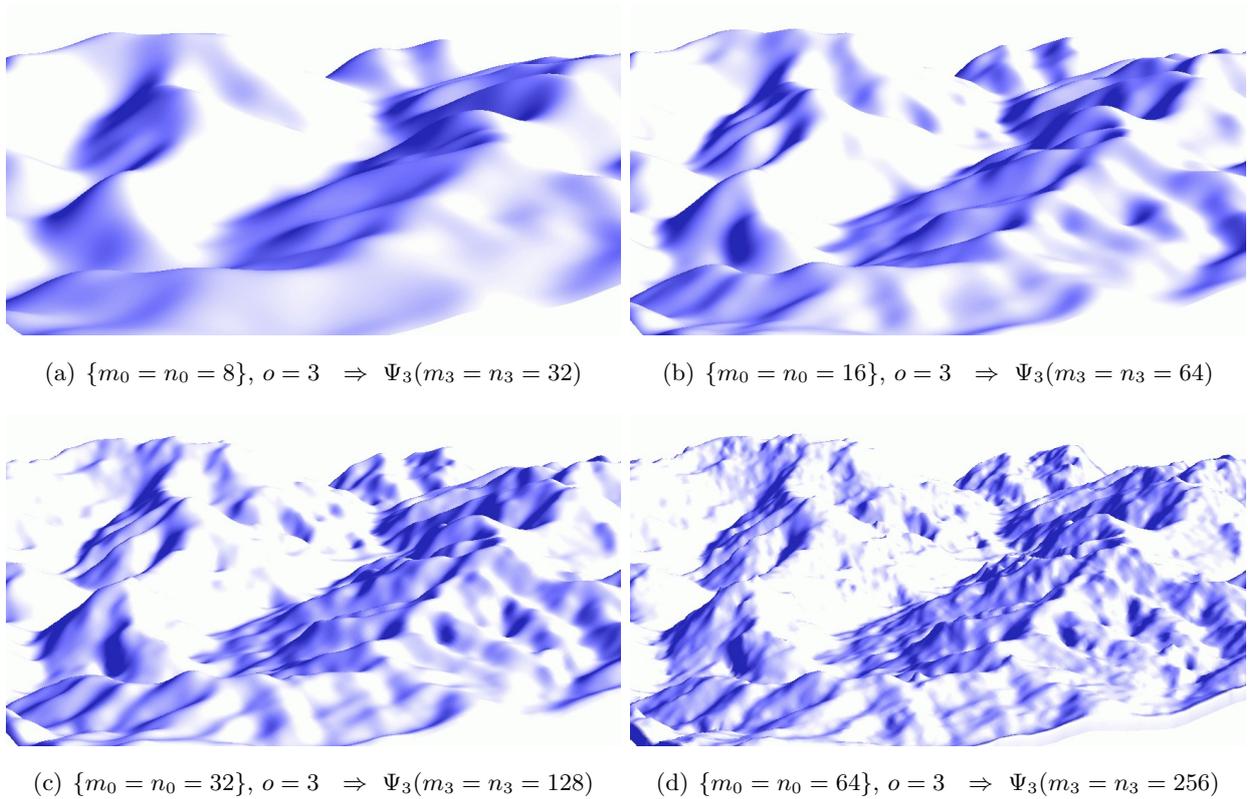
Figure 8.4: Incidence of the initial lattice Ψ_0 size $\{m_0, n_0\}$ using a 3-level B-spline refinement scheme on the data-set presented in Fig. 8.3. The control point lattices Ψ_3 are then used to sample a surface to the same size as the input.

wide areas without any information. With small $\{m_0, n_0\}$ values, the risk of empty areas during the construction of the initial control point lattice is reduced. In that case, the B-spline approximation leads to better results.

Filling Missing Information

As already mentioned in chapter 3, the elevation cannot be derived by stereoscopy for regions whose correspondence problem (image matching) failed. The missing areas need to be interpolated from the available elevation areas or integrated from other existing sources of elevation.

The preliminary Cartosat data-set at disposal is provided with a mask indicating the areas whose elevation information is missing (Fig. 8.5(a)). Such locations were filled with the SRTM-C DTM (90 meter resolution) using an inverse distance weighted interpolation algorithm (IWD, see Eq. 8.11, page 158). As shown in Fig. 8.5(c), some interpolation artifacts[5] are reported, resulting from the important sampling difference between the two sources of elevations.

---

[5]Presence of vertical and horizontal lines which provide an unrealistic rendering. The observed patterns can result from the resampling of the SRTM DEM with a bilinear interpolation to the Cartosat resolution during its reprojection to another geographical coordinate system.

(a) Generated Cartosat DSM   (b) Integration of SRTM DEM (B-spline)

(c) Local contrast (IWD)   (d) Local contrast (B-spline)
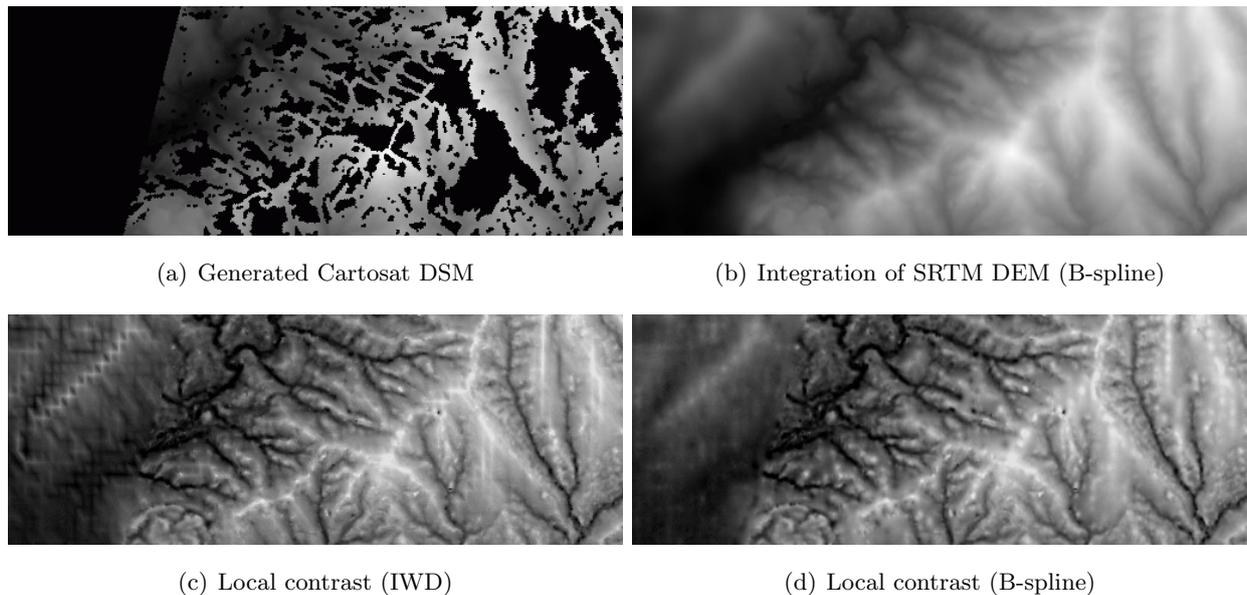
Figure 8.5: Cartosat DSM (size 512×192, elevation range: 329-650 meters, $\sigma = 70.6$). Missing information (black areas in (a)) filled with SRTM-C DTM using a B-spline method (b). Local contrast adjustement to compare the preliminary DSM (c) with our method (d).

Based on the results presented in Fig. 8.3 and Fig. 8.4, we propose, instead of an inverse distance weighting [185] the use of the multilevel B-spline refinement. To fill the missing information, two scenarios are investigated depending or not of the availibility of another elevation data-set (e.g. SRTM-C DTM). The interest of the method is that both cases are handled identically since it allows to simultaneously injest multiple sources of elevation presenting different resolutions.

A list of 3D points $< P >$ is generated by collecting all the Cartosat DSM information. To ensure a suitable smoothing of the empty areas, we choose a small initial control point lattice size Ψ_0 (e.g. 2×2). Based on $< P >$, Ψ_0 is refined until a control point lattice Ψ_o with a size at least as large as the DSM is obtained. The latter is used to compute the filled DSM (Fig. 8.6(e)).

To integrate elevation information from another source (here SRTM-C DTM), the same protocol is followed. The list $< P >$ is augmented by 3D points from the areas where no information is available in the Cartosat DSM (see Fig. 8.6(a)). Such an integration assumes that no registration errors (in planimetry and altimetry) occur between the data-sets[6].

To improve the results, the set of SRTM data is densified (see Fig. 8.6(b)). This stage is justified by the important resolution difference between both elevation sources. Since some local height difference between the data-sets may arise, a minimal distance from the Cartosat elevation points is used when collecting the 3D points from the SRTM DTM. To generate the densified points, the SRTM is first reinterpolated using the B-spline technique to the Cartosat DSM resolution. Since

---

[6]In our example, elevations from the SRTM-C DTM are measured from the ellispoid WGS 84 while elevations of the Cartosat DSM are already mapped from the local approximation of the geoid. To compensate for the difference of elevation between the datums, the assumption of a constant shift was done (the test is achieved on a small area).

(a) Cartosat and SRTM samples   (b) Cartosat and SRTM densified   (c) IWD

(d) In Painting only Cartosat   (e) B-spline only Cartosat   (f) B-spline resulting from (b)
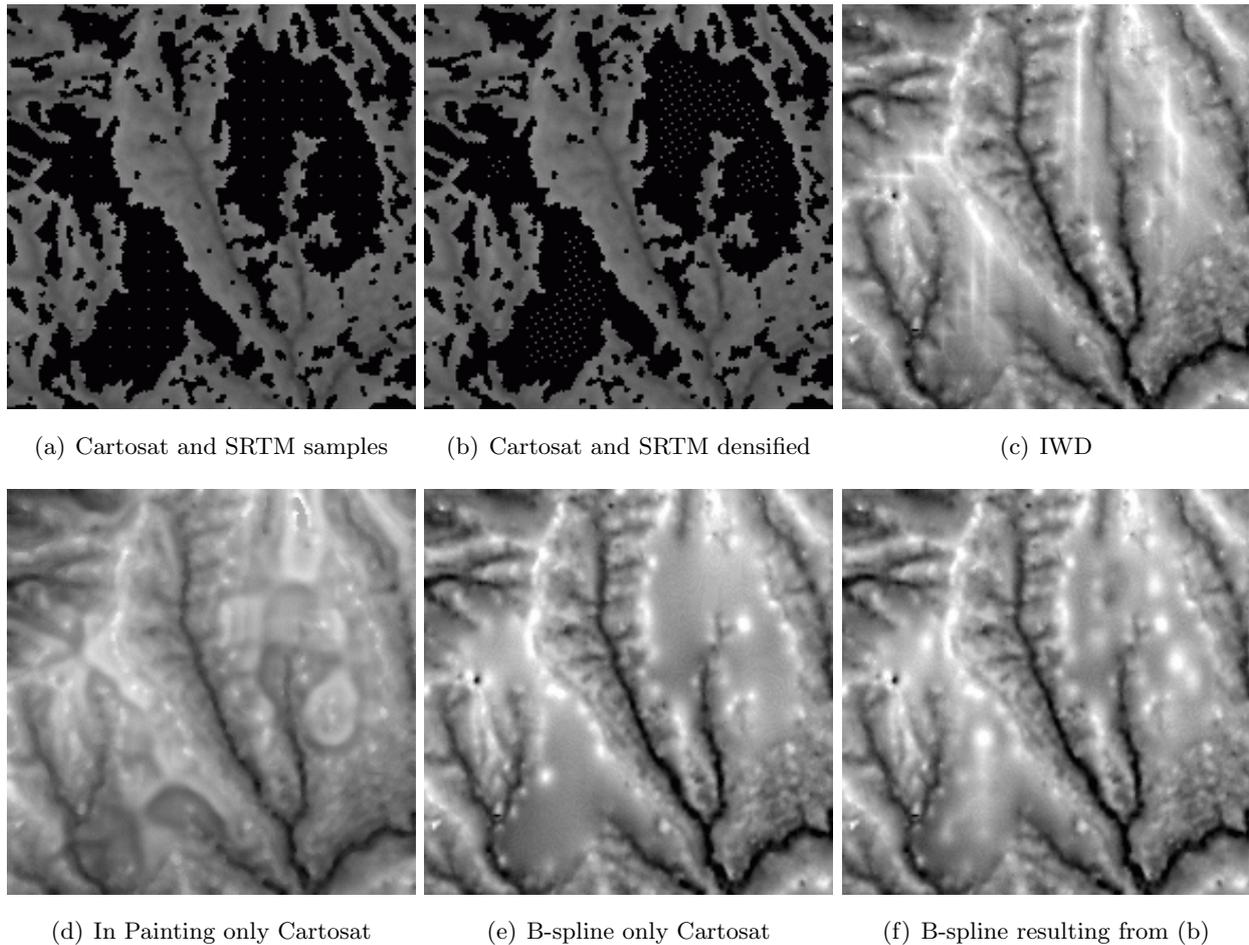
Figure 8.6: Missing information filled based on Cartosat data augmented with repectively SRTM-C data (a) and SRTM-C data densified (b) (the regularly spaced points). A local contrast adjustment is applied on the results (c)-(f) to emphasize the interpolation behavior.

by construction the height difference between the interpolated source and the control points is minimized, the denser information is directly taken from the reinterpolated data-set.

In Fig. 8.6, some results are given from a cropped area of size 192×192 with an elevation range of 440-647 meters and $\sigma = 38.4$. The local contrast adjustement shows important changes according to the methods. Indeed, the range of height difference between the IWD and the B-spline methods are ± 10 meters with $\sigma_\Delta = 2.6$. An In-painting algorithm [27] was implemented to analyse the potential of such an approach to fill empty areas in DEMs. Unfortunately, as shown in Fig. 8.6(d), difficulties arise when (1) the density of the related coverages is important with narrow information in between, (2) the areas are large and thick.

The presented method constitutes an alternative to the classical IWD method. The quality of the results is highly dependent on the 3D registration accuracy. Its advantage is that both re-interpolation and integration are performed in one stage using sources of elevation with different resolutions.

8.3 Integration of Semantic Information in DEM

The representation of the third dimension provides important information about the relations between land coverages: shapes, structures and slopes appear to describe scene objects such as waterways, surface material or vegetative growth. Indeed, our perception automatically associates an object with its corresponding shape. For instance, a lake should be a flat surface with the lowest elevation according to its direct neighbors. In RS images, we obtain such inconsistencies (see Fig. 8.7).

| (a) InSAR SRTM-X DEM | (b) Optical Image | (c) 3D Perspective |
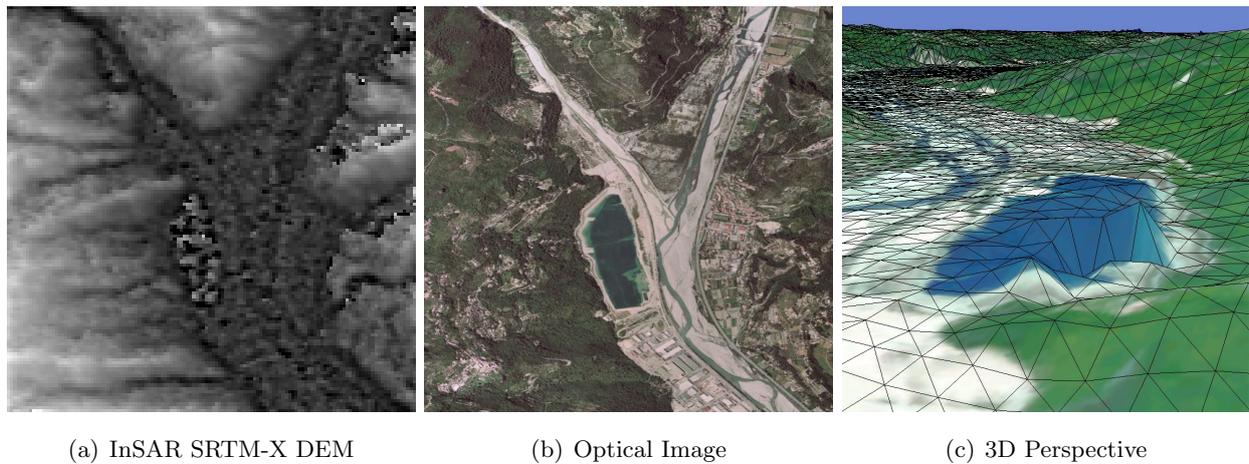
Figure 8.7: Lake example (a Wallis filtering is applied to the DEM, duplicated from Fig. 3.10).

8.3.1 Flowchart of a DEM Enhancement System
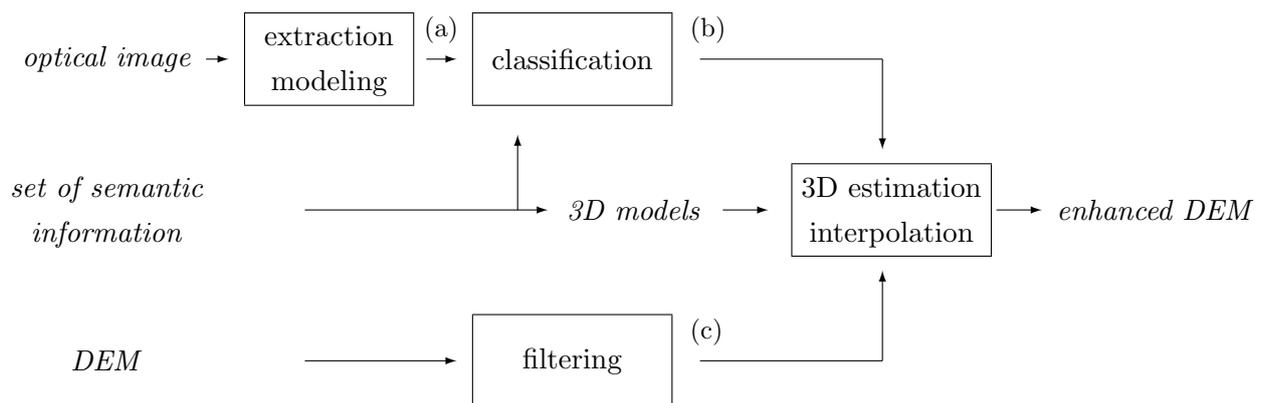
Figure 8.8: Flowchart of the DEM enhancement system. (a) denotes the object modelization of the segmented image (see Fig. 2.13 page 27 and chapters 4 to 6). (b) consists in (a) augmented by semantic information (SVM, section 6.3). (c) represents the filtered DEM (chapter 7). Based on 3D semantic models, (b) and (c), the DEM is enhanced in two sequential stages; 3D surface estimation and B-spline interpolation.

Elevation constraints have to be integrated in the DEM according to objects thematic. Based on an existing vector database, a semantic DEM regularization is introduced in [113], [114]. Our proposed method for DEM regularisation is only based on an optical image. In Fig. 8.8, a DEM enhancement system is presented, based on the independent processes previously proposed.

The results achieved by the processing line allow to delineate the 2D coverage (in the complementary image) of the objects to be regularized in the elevation data. We assume proper registration between DEM and image. In case of large misregistrations, the quality of the enhancement is affected by weaker surface estimations.

The estimation of the object elevation surface must be carried out and depends on: (1) DEM information, (2) region shape and its associated 3D semantic model, (3) relations to the neighboring regions. The first point is modeled by absolute elevation constraints, while the others provide observations in the form of relative elevations.

8.3.2   3D semantic Modeling and Object Description Linkage

Table 8.1 summarizes some of the 3D structures with semantic constraints observed in EO images (medium to high resolution, see page. 44). Two types of constraints are modeled (internal / external). The former are inferred by the object itself (its geometry), whereas the latter are related to its topological properties

Object,	Constraints	
border elevation H_{obj}	**Internal**	**External (neighbors)**
Sea, lake, canal		$H_{obj} < H_{neigh}$
Large building roof		$H_{obj} > H_{neigh}$
Airport, parking, football ground...	Horizontal surface	-
Bridge, dam		$H_{obj} < H_{neigh}$
Street, motorway	Tiled surface,	$H_{obj} < H_{neigh}$
River	(small slope)	$H_{obj} < H_{neigh}$
Building, house	2 adjacent tiled surface (sym. slope)	$H_{obj} > H_{neigh}$
Forest	-	$H_{obj} > H_{neigh}$

Table 8.1: Some man-made objects and natural features with elevation constraints.

The image analysis processing line proposed (segmentation and modeling) can be used as basis for the generation of semantic information. To apply DEM regularization constraints, objects which

present a semantic having a possible impact on the elevation data (see Table 8.1) need first to be labeled (e.g. user interests, such as "forest", "bridge", "road", "rivers", etc.). This semantical information is then added to the geometric and topological tree structured model.

User-defined semantic contents are trained on the extracted region using a supervised classifier (SVM). A multiclass "One versus All" scheme is designed using a Gaussian kernel with $\gamma = 0.001$ (see section 6.3, page. 112). The resulting classification can be further improved through a graphical user interface. An example of the resulting thematic map is given in Fig. 8.10.

The geometrical and topological modeling of regions, handled by the tree structure, is augmented by an additional set of information, which completes the description of regions by linking them to their respective user-defined cover types. Using the learning process, the user's "knowledge"and his specific interests are ingested in the system.

If the classification results are satisfying, a further stage is often required to validate and correct the different learned thematics if necessary. Indeed, several semantics, such as "bridge" or "road" are first trained using a more global description such as "road network". Later, the classification is refined in order to discriminate such semantics using the geometrical/topological properties of regions and their linked thematic.

Another case concerns areas populated from different individual semantics which represent a thematic. An example of such a thematic is given in Fig. 8.10(b) with the "airport", composed of bare-terrain elements (runways, runwaymarkings, apron, fields, parkings, etc.) and some buildings, planes, etc. The tree structure is used to retrieve the former elements using a polygonal delineation: all included regions which are not assigned to the semantic "Building" are collected.

Once regions with similar spectral signatures are distinguished and the thematic assigned, the neighbored regions with identical semantics are merged. Such a process can be interactively refined until satisfying cover type assignations are obtained.

8.4 3D Estimation and Regularization

Based on the modeling presented in Table 8.1 and the generated thematic map, the consistency of elevation data can be improved. In this scope, we propose a DEM regularization process achieved in two stages: (i) surface elevation estimation, (ii) transition control, according to the 3D geometry of the inserted objects.

8.4.1 Problem Statement

The 3D semantic models represent either horizontal or slanted (Eq. 8.9) surfaces according to the defined theme (Table 8.1). Both can be described by means of a plane surface

$$X_i = (x_i, y_i, H_i) \ , \quad a \ x_i + b \ y_i + c \ H_i + d = 0 \tag{8.9}$$

Similarly to 3D computer graphic models, many 3D structures can be decomposed with a relatively compact set of facets. In remote sensing applications with high resolution data (≥ 1 meter resolution), such a modeling is sufficient and justifies the restriction to planar surfaces.

Eq. 8.9 is used to estimate the elevation information: the planimetric information provided by the objects extracted from the complementary image is neither corrected nor regularized. This assumption prevents the integration of possible registration errors or inaccuracies resulting from the object's delineation performed by the region-based segmentation approach. If this relaxation is reasonable using a 2.5 meter SPOT 5 image, when dealing with higher resolution data, this hypothesis should be reviewed. Another issue, which is dependent on the DEM content, is related to the height estimation according to the DEM planimetric resolution. It seems useless to extract precisely small houses from a metric image while expecting an accurate roof estimation in 30 meter resolution DEM.

There exist several approaches for the estimation of plane surfaces. The 3D Hough transform [97] or the random sample consensus (RANSAC) [62] are extremely appealing for their robustness to outliers. However, both are based on a discrete method to estimate the continous surface function formulated in Eq. 8.9: the dilemma between accuracy and robustness is also valid here.

In [40], a Bayesian approach is proposed to estimate multi-plane models, using a Gaussian likelihood, a non-informative prior and a spatially independent data model. Such a process is applied to estimate the roof parameters of buildings. Based on our implementation of this method, we notice several observations:

- stable convergence,
- the parameter initialization affects only the convergence velocity,
- accurate results for small volume of 3D points

We did not pursue this direction since the method makes only implicit the use of spatial dependencies and the regions are processed sequentially.

8.4.2   Elevation Data with semantic Constraints

A least square problem (LSP) is formulated to estimate the elevation surface of each region selected. Since elevation and image data can have different resolutions, the least square adjustment is applied to the DEM grid. Because the elevation samples are taken from the filtered DEM[7], the sensitivity of LSP to outliers is relaxed. Indeed, only large artifacts have to be handled. The latter depict areas without reliable information and are therefore not considered during the height estimation. Their extraction is performed by estimating a plane surface for each region without assumptions regarding its slopes. The fitted surface is then compared to our a priori knowledge about the region (semantic), if the slope comparison exceeds a given threshold, those areas are removed.

---

[7]The DEM is filtered using the non-stationary Bayesian approach (see section 7.2).

The plane surface estimation depends on the assigned theme, the region neighboring and the corresponding DEM observations. The latter are collected and classified in two categories (inside / outside). Unfortunately, based on such a protocol, thin structures without any included information cannot be locally estimated.

The topological description provided by the tree structure is determinant to estimate the object's surface elevation, since the neighbors objects presenting different semantics may have interdependencies. As a consequence, all the neighbor regions with 3D semantic content are extracted and added as observations in the least square system. Topological constraints are formulated through slope estimation $d_{thematic_j}$ for each adjacent theme j between the interior and exterior observations

$$d_j = \frac{H_{\text{ext}_j} - H_{\text{int}}}{d_{\text{ext}-\text{int}}} \tag{8.10}$$

The estimated elevation surface is adjusted to respect the external constraints (Table 8.1) carried by all the estimated $d_{thematic_j}$. Once all the neighbor regions observations are ingested in the system, the least square problem is computed[8]. The procedure is repeated until all regions with semantic information are processed.

8.4.3 Objects Integration

To ensure topological consistencies on a regular grid, the elevation data are re-sampled to the complementary image resolution. The interpolation stage is determinant, two approaches are proposed for controlling the shape of the inserted objects;

- using morphological operators,
- using the multilevel B-spline refinement algorithm (see pages 147-152)

Numerous observations already described in section 8.2.2 (fill missing information in DEMs) are re-used to perform the DEM regularization.

Discontinuities with B-spline Interpolation and IWD

To simulate a discontinuity, an exemple is given using a step signal in Fig. 8.9. In general, B-splines are limited to model sharp transitions. Oscillations and smooth behavior along the discontinuity (see the Raw data-set in Fig. 8.9) can be partially removed by densifying the set of observation around the related region. In Fig. 8.9, the densified set of point is close to a "ideal" reconstruction. Since the points are sampled on a grid, the step is completed at about 80 % (between position 7 and 8) and only a small oscilation ($\leq 3\%$) is reported from each side of the transition.

---

[8]An alternative to this approach is to explicitly formulate the inequality constraints as a linear complementary problem [41].
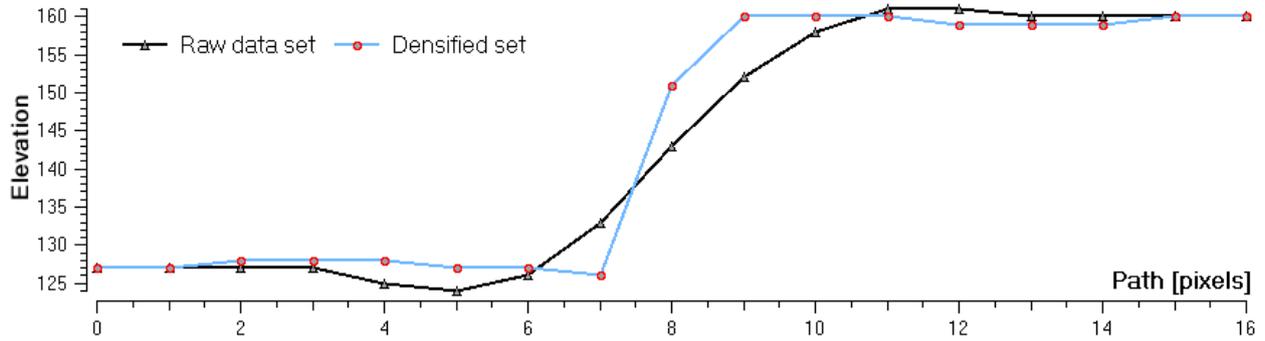
Figure 8.9: B-spline interpolation analysis. To model a step signal (two stages 128 and 160, step at position 8), the initial set of points is sampled every four pixels (each pixel for the densified set). The presented profiles represent the B-spline reconstruction of each data-set.

The simplest form of the inverse weighted distance interpolation [185] follows

$$F(x,y) = \sum_{i=1}^{n} w_i f_i \quad, w_i = \frac{h_i^{-p}}{\sum_{j=1}^{n} h_j^{-p}} \quad, \text{with (p=2):} \quad h_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \qquad (8.11)$$

where i denotes a scatter point of coordinate (x_i, y_i). If the method is popular for its simplicity, it suffers from some artifacts appearing at the data points, and its global formulation implies that all weights w_i need to be recomputed when the set of points changes. Significant improvements are given in [65].

Morphological Operators

In this approach, the enhanced DEM is first interpolated to match the image's resolution. Then morphological operators (dilation-erosion, see page 15) are used to control -according to the thematic of their adjacent regions- transitions and shapes at the border of the inserted regions.

Indeed, similarly as in-painting methods [27] a uniform front propagation around the investigated objects (using morphological operators) is computed and used to perform locally an IWD interpolation. The example of forestry areas emphasizes the interest of this method for integrating discontinuities; because the bare terrain information and the height of canopy is unknown, a regularization is only possible on the forest edge.

B-spline Refinement Algorithm

This approach significantly differs from the previous one. Based on the algorithm presented page 147, interpolation and regularization are performed simultaneously. All the DEM information not integrated in one of the multiple LSP is directly ingested in a set $< P >$ of 3D points to be interpolated. The set $< P >$ is augmented by all the estimated DEM samples included in the regions with semantic contents.

To control the transitions, points are created and then ingested in $< P >$. Such points are constructed using the planimetric information of the region boundaries, given by the tree structure and its estimated height. The difference of resolution between the elevation data and the optical image is exploited to avoid oscillation artifacts: the set of created points is densified by samples computed along the boundary. It enables, as shown in Fig. 8.9, to fit more properly sharp disconti-nuities. The external transitions are integrated in the set $< P >$ using a uniform front propagation around the investigated objects to create the corresponding points.

Indeed, the regularization consists in the interpolation of the scaterred points $< P >$. As a consequence, the enhanced DEM can be sampled to any resolution.

8.4.4 Validation

Experiments have been carried out for a data-set composed of a SPOT 5 "Supermode" color image (2.5 meter resolution) and a X-band SRTM DEM (about 25 meters).

Example with a SPOT 5 Image and a SRTM-X DEM

The tree structure described in chapter 6 models a segmentation of the SPOT 5 image with a relatively important number of regions to avoid the loss of relevant information. The thematic map (Fig. 8.10(b)) corresponds to the classified tree structure using semantics such as "large building", "forest", "bridge", "road", "rivers", "airport", etc. Except the semantic "forest", the 3D models associated with the investigated semantics are either tiled or horizontal plane surfaces (according to Table 8.1).

The region integration in the DEM is emphasized by Fig. 8.10(d) through a gradient visu-alization with North-West illumination. Since this process mainly focuses on specific structures, topographic details are not affected by the DEM regularization. Hydrological networks are more properly integrated. 3D rendering enhancements are presented in Fig. 8.11(a) and Fig. 8.11(b) by using respectively the filtered DEM and the regularized DEM.

Very High Resolution data

While such a model might be too restrictive for the the purpose of urban studies (e.g. 3D modeling of archeological site [78]), it appears well-suited for Earth observation applications with satellite images. As previously mentioned, the semantic information listed in Table 8.1 is scaled to the information content that can be extracted from the complementary data. In this scope, the case of VHR data[9], is not studied here. However, due to the tree structure modeling presented in chapter 6, the decription of elaborated patterns, composed of multiple objects, is feasible.

---

[9]Composed of complex objects, with multiple shapes and discontinuities.

Indeed, as reported in section 8.1, the main challenges of VHR data over urban-areas -with VHR data corrupted by noise and artifacts- rely on the elaboration of robust methods and concepts for the extraction of non-linear features and incomplete 3D primitives. The combination of both spectral and rough elevation data within a unified segmentation algorithm could significantly improve the extraction of these low-level primitives.

Assuming having at disposal a segmentation process able to deal with VHR data-sets, the typical example of a complex roof (more than two symmetric pitches) could be handled with a small library of 3D models to find the one which best fits the observations. Similarly to the approach discussed, the 3D geometry of the selected model would be estimated, and according to its topological properties, either integrated in the data for the purposes of enhancement, or stored in a 3D vector database in the case of information extraction.

Some tests have been conducted to segment an airborne optical image (12 centimeters) using the segmentation approach described in chapter 5. We obtained promising results, but the accuracy of the boundaries between regions and the presence of important shadows spoiled the segmentation quality. The extremely narrow dynamic of the spectral information available also partially explains the results.

Conversely, excelent results are derived from the 1 meter Ikonos images shown in Fig. 5.8 to 5.10 with a much higher dynamic. The "small" resolution increase between the data-sets drastically changes the nature of the considered segmentation problem. Unfortunately, no elevation data were at disposal for the Ikonos data-set.

Summary

Thanks to the image analysis methods presented in chapters 5- 6, characteristical structures which can be associated with a defined elevation modeling can be enhanced. Spatial relationships of regions are taken into account to properly integrate the objects in the DEM. The method assumes that DEM and complementary image are registered with minimal errors. The estimation of the elevation surface relies on the information contained in the data. Having at disposal only extremely disturbed data, the elevation estimation may lead to inaccurate results. In this case, only the external constraint rules are satisfied (consistency between regions, thematics such as "forest").

If good results are obtained with the morphological approach, the method using the multilevel B-spline refinement algorithm appears to be more flexible since it allows to achieve the integration in one stage, and to control the observations taken into account. Finally it does not require to re-sample the elevation data to the complementary image resolution. It should be mentioned that handling sharp transitions is difficult with the B-spline approach; an alternative is to interpolate the bare-terrain and then overlay the object, either directly in the DEM or build a 3D vector database.

The presented system enables to improve accuracy and consistency of digital elevation models, and to generate more realistic 3D visualizations.

(a) SPOT 5 segmented image

(b) Thematic image

(c) Raw SRTM-X DEM

(d) "Slope" of the regularized DEM
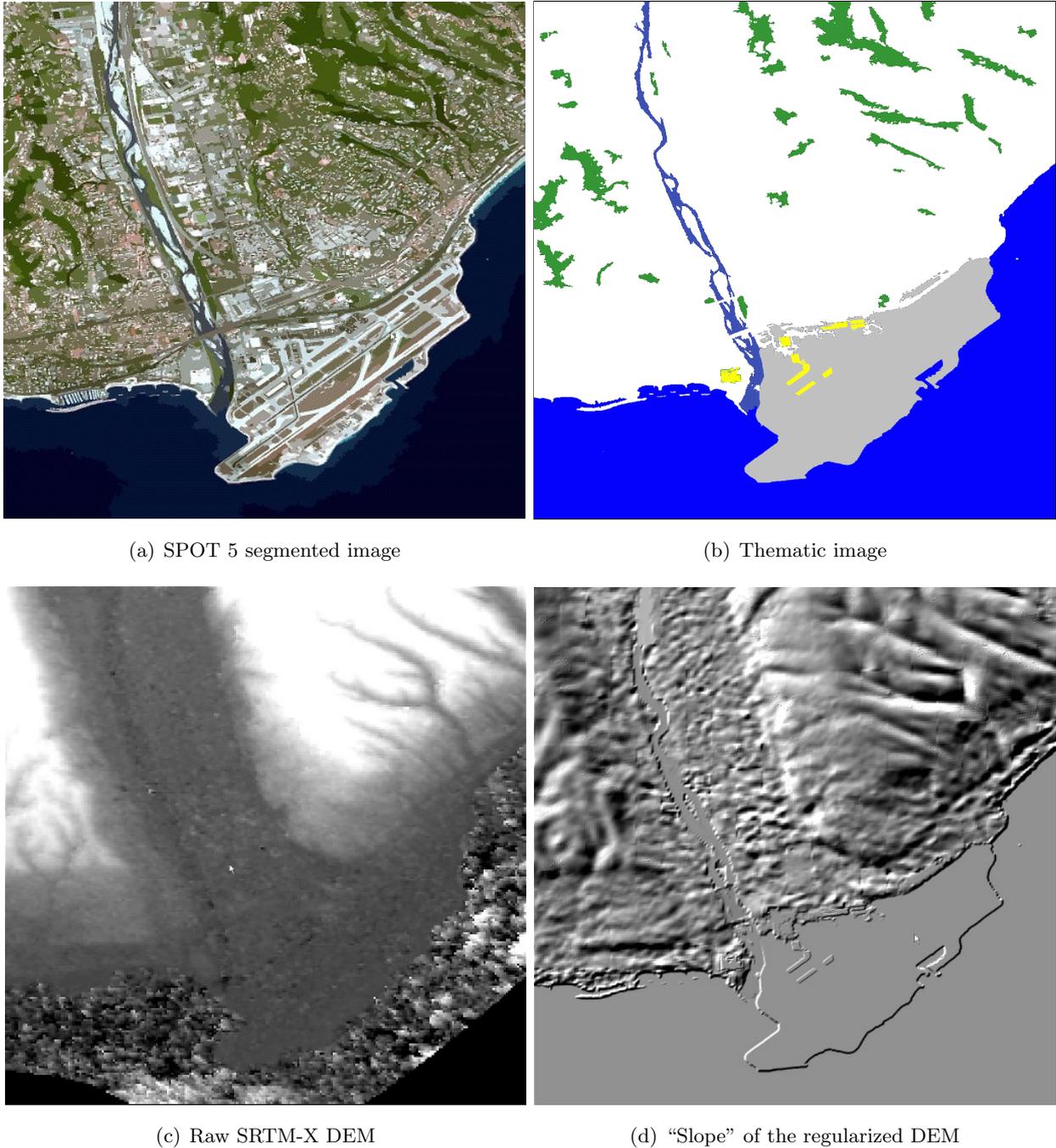
Figure 8.10: Integration of regions extracted (a), (b) from a SPOT 5 "Supermode" image into the SRTM-X DEM data (c) over the Nice (France) test site. Visualization of the regularization using a slope computation (d).
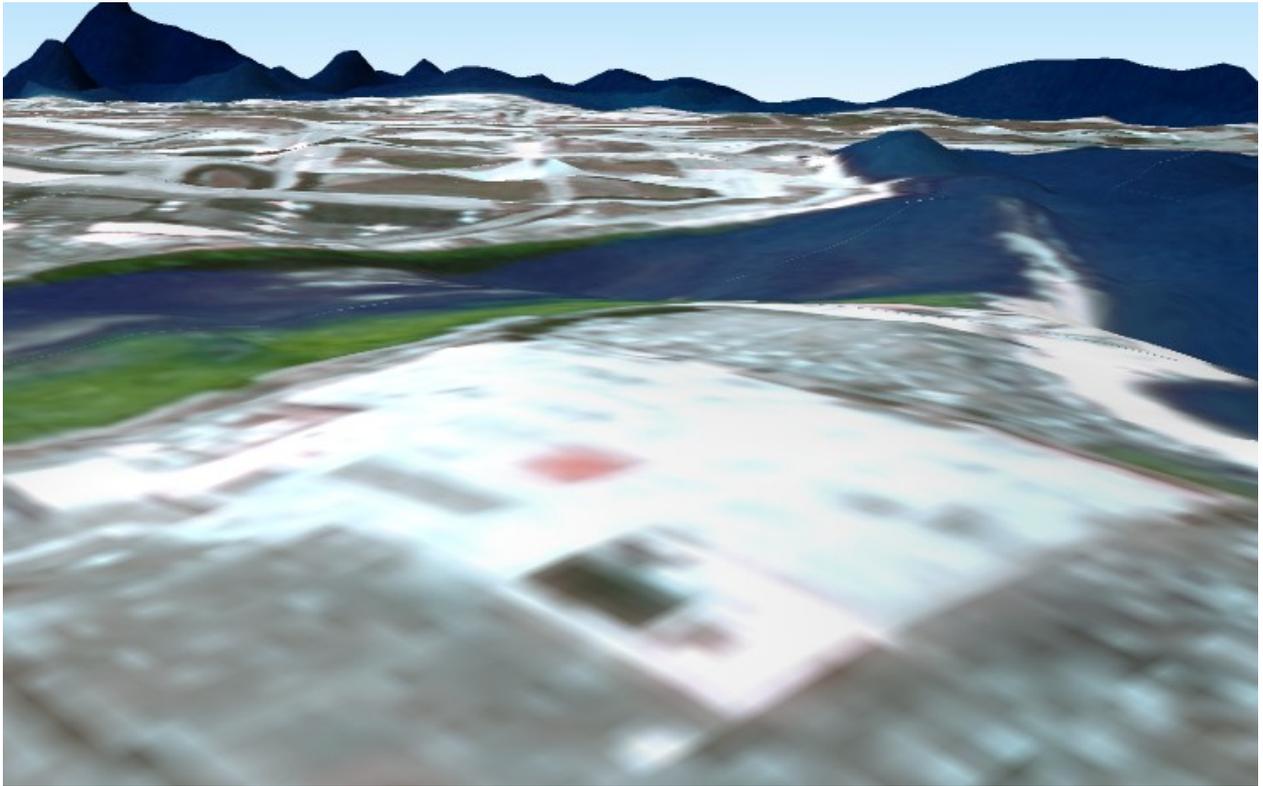
(a) Filtered SRTM-X DEM overlaid by the multispectral SPOT 5 image
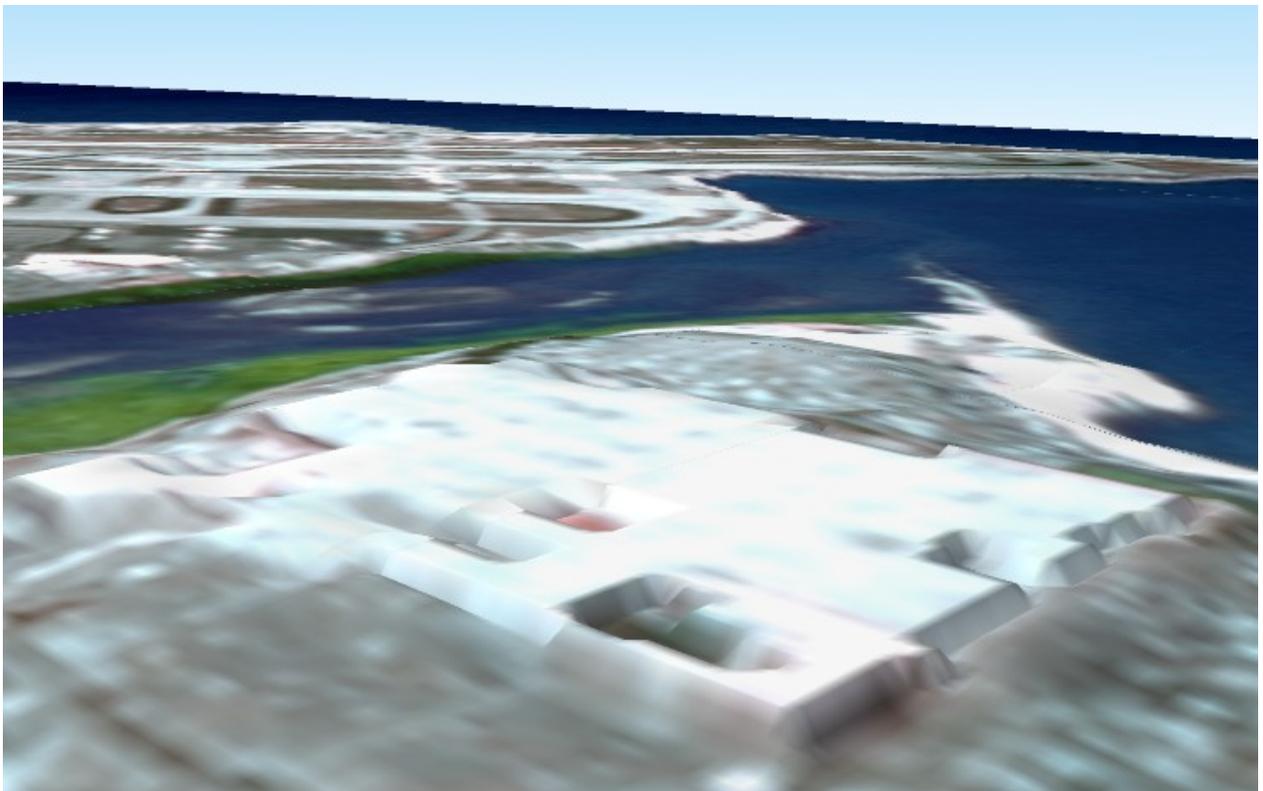
(b) Regularization based on the segmented SPOT 5 image

Figure 8.11: Perspective views using respectively the filtered DEM and the regularized DEM.

Conclusion

With the growing interest for Earth Observation data and technological developments (e.g. increase of resolution and number of potential sources), efficient and adapted methods are required for the extraction of information, its modeling and the enhancement of remote sensing (RS) images. The potential offered by RS techniques to acquire (e.g. Ikonos, QuickBird, etc.) and generate (world homogeneous elevation data-sets SRTM, Cartosat or TanDEM-X) large coverage with important resolution has to be fully exploited by using of suitable methods and tools.

In this scope, after the investigation of the state of the art in computer vision and Earth observation, it was possible to distinguish several major steps for the extraction of meaningful information from remote sensing images:

(1) segmentation to partition images into disjoint and homogeneous regions,

(2) modelization to derive an object-based representation of the extracted region into a set of elementary primitives,

(3) information learning to analyse and discriminate the resulting information according to some similarity criteria or learned instances

To enhance elevation data corrupted by noise, artifacts and inconsistencies, image denoising techniques deliver significant improvements. However, such a stage is not sufficient. A possible application of the presented image processing system relies on the enhancement of elevation data by providing complementary information. Based on a semantical description, derived only from EO data, regularization constraints were applied to integrate subjective knowledge into digital elevation models.

Filtering of Digital Elevation Models

Several denoising methods were presented and compared in the purpose of DEM filtering. Based on statistical, fractal and qualitative analyses, we provided a methodology to quantify and validate DEM filtering accuracy using both optical stereoscopic and InSAR decametric DEMs. Special attention was paid to multi-scale and Bayesian techniques for their ability to deal with non-stationary data such as DEMs. In presence of SAR artifacts or high relief transition, a weak filtering is provided by the soft-thresholding wavelet approach. An adaptive threshold determination should lead to more efficient results. Compared to all other techniques investigated, the Bayesian filter with a Gaussian likelihood and a GMRF as prior information appears to be the most appropriate approach

to filter DEMs. Its locally adaptive noise estimation is well-suited to remove uncorrelated noise from DEM data while preserving structural information. Despite its filtering performances, limitations arise with high resolution data containing structures and/or non-linear features. In such cases, the consistency (related to the semantic of the remotely sensed area) is hardly recovered based only on elevation data. To significantly enhance the DEM, complementary information must be gathered from other sources.

To semantically improve DEMs, we proposed to directly extract information from another EO data instead of using an existing GIS database (e.g. original data employed to generate elevation data). The goal was to link the collected meaningful information with semantic content. The latter is expressed as geometrical regularization constraints and finally integrated into the filtered DEM. We decomposed such an image processing system into three independent problems, applied sequentially; (1) primitive extraction using a robust region-based segmentation, (2) information modeling to generate a complete geometrical and topological description of an image partition, (3) analysis and integration of a priori information to perform the DEM regularization.

Image Segmentation

In a computer vision system, the segmentation is a determinant process since it directly scales the quality and the performances of the overall system. Under the formalism introduced by Mumford and Shah (MS), image segmentation is formulated as an optimization problem and defined simultaneously as a smoothing and as an edge detection problem. The image is approximated by means of piecewise smooth functions, describing disjoint regions bounded by a corresponding discontinuity set. Unfortunately, mainly due to the nature of the unknowns -a function and a curve-, the MS problem is still unresolved. Among its approximations applied to RS images, snakes approaches were not retained since they are mostly used for the extraction of some particular shapes or objects, without the ability to handle topological changes while computing front propagation. If this limitation is resolved for level-sets approaches, their computation is still extremely time-consuming and like snakes, the initialization is difficult and affects quality of the results.

Based on the "cartoon limit" case of the MS functional (piecewise constant functions) that we implement as a region merging algorithm, we obtained promising results in relatively short computational-time. To segment large images, we propose an iterative approach to progressively simplify the processed data in order to maintain a global region competition. Its computational efficiency is scaled to the amount of resources at disposal (memory) to handle and process about hundred millions of boundaries and regions composing a RS image.

To improve the segmentation quality, we investigated additional criteria, regarding the size of regions or their discrimination using second order of statistics. By means of higher function degrees (smooth piecewise polynomial), we dealt with some non-uniform illuminations and tried to segment VHR LIDAR elevation data on urban areas. If interesting results were obtained, the presented framework is not sufficient. Indeed, with the increase of the model complexity, the

initialization is not trivial anymore and further stages are required. The fine to coarse strategy employed is driven by the scale parameter used as decision criterion. It results in a pyramidal sequence of image segmentations presenting coarser and coarser partitions. The determination of a scale to stop the merging process is critical an under-segmentation being always preferred to avoid a definitive loss of information. Like the level of details in a topographic map, the density in regions varies according to the user needs, the application and the studied area (amount of information contained in the image).

To deal with images presenting a large variety of information (e.g. 10-30 centimeter resolution), a single and global scale parameter is not sufficient since patterns or features have their "own" sharpness and roughness, and therefore different merging costs are expected. In this scope, obtaining simultaneously cereal fields discrimination and suitable man-made objects delineation is still an open issue. This remark is however relaxed by the pyramidal evolution of the method. If that is not always the case with classical photographs, the presented segmentation framework appears to be very well-suited for RS images. The algorithm was tested and validated on different image types: panchromatic Cartosat, multispectral Ikonos and SPOT 5, SAR polarimetry. Those experiments on real images emphasized its robustness, velocity and ability to deal with a large amount of data. It can be used as well in the purpose of urban area extraction, as shown by the example given in Fig. 9.1 on a QuickBird multispectral image (resolution of 0.6 meter).

Geometrical and Topological Description

According to the principle stated by Marr [146] to subdivide a computer vision system into operations with an increasing level of descriptions, a stage is proposed to model the regions resulting from a low-level image partition. If the region description as a set of features and labeled pixels is efficient for low-level processing, such a basic modeling is rapidly defeated with higher-level of operations or requests occurring in image interpretation processes. The pixel-wise region definition is replaced by a set of primitives and relations, allowing to describe elaborated data structures, express additional properties and relationships between objects.

Based on the complex cellular model, the proposed method makes implicit the 2-cells, the resulting regions being modeled only by set of 0- and 1- cells elements, stored in a database $\mathcal{D}$ containing both geometry and adjacency relations. Such a modeling presents a high interest since it eludes all the critical issues encountered with neighborhood graphs.

We proposed to complete the object representation by investigating the inclusion relationships between regions. In this purpose, a tree structure[1] was introduced to store the regions while reflecting their topological properties (region relationships). The overall collected information constitutes a consistent and independent hierarchical database which enables for instance (1) to link the extracted objects to semantic description, (2) to select groups of objects according to spatial or

---

[1]It represents a hierarchical decomposition of the image partition.

Figure 9.1: Cropped area from a full scene segmentation of a QuickBird image (0.6 meter resolution, size: 21163×16487, the 3 first channels are displayed). The average region area is about 350 pixels. The cropped image represents an area of $\approx 1200 \times 526$ meters (image provided by SERTIT).

radiometric dependencies, (3) to derive some complementary information about the region shape (e.g.) or (4) to focus on some specific areas of the image and refine the segmentation.

Integration of Complementary Information

The consistency of elevation data[2] can be improved using a thematic map and some elevation modeling associated to each cover type studied. In this case, we proposed a DEM regularization process achieved in two stages: (i) surface elevation estimation, (ii) transition control, according to the shapes of the inserted objects. The image analysis processing line proposed (segmentation and modeling) is used to "produce" the semantic information. Applying classification techniques

to the extracted and modeled regions, several semantics are trained. It enables the derivation of a cartography which can be further improved through a graphical user interface. Using a learning process, the user's "knowledge"and his specific interests are ingested in the system.

We restricted the object's elevation modeling to planar surfaces. The topological description provided by the tree structure was determinant to estimate the surface elevation from each object, since the neighbor objects presenting different semantics may have interdependencies. To ensure topology consistencies and complete object insertion in the DEM, we re-sampled the elevation data to the image resolution. The importance of the interpolation stage was studied, we tested two approaches for controlling the shape of the inserted objects; (1) using morphological operators, (2) using B-splines. The latter was also tested to fill missing information in stereoscopic DEMs[3].

Evaluation of the Obtained Results and Outlook

Since a filtering stage is not sufficient to properly correct DEMs, an important issue in this dissertation was how to get the thematic map.

The example of the techniques investigated to deal with VHR data (airborne LIDAR) is significant of the difficulties to extract features and patterns from elevation data. With medium to high resolution elevation data (i.e. 25 to 1m), limitations are encountered to the description and the modeling of the information content. The latter being composed of large scale components and some 3D objects and/or primitives. This remark, related to the data content according to the resolution, defines the range of possibilities for the enhancement of DEMs such as SRTM, Cartosat and probably the upcoming TanDEM-X, covering large and homogeneous areas. The use of complementary information appears to be the sole method to semantically correct such data.

For this reason, the initial goal of the thesis -to enhance world coverage SRTM-X DEM and later Cartosat DSM- migrates to the analysis of various types of remote sensing images. Finally, having at disposal a robust segmentation algorithm, very promising results are obtained for the regularization and the enhancement of high resolution DEMs.

We noticed the quality of the segmentation algorithm applied on various types of information and data content and its ability to deal with extremely large amounts of data (for a variational approach). In particular, its behavior on meter resolution color images (Ikonos, QuickBird) or on panchromatic Cartosat images (2.5 meter) is almost optimal. Promising results are expected with the upcoming new high resolution systems (e.g. GeoEyes, Pleiades, Worldview).

If the computer vision system proposed (segmentation, modeling and interpretation using learning methods) can be used for more general purposes, the object modeling description, by its complete geometrical and topological (adjacency, inclusion) description of regions without any assumption about the region-based segmentation previously used, is very appealing to "enter" in cities.

---

[2]To correct errors/inaccuracies of objects having a specific geometry, e.g. assumptions about the lake surface.
[3]Areas where the correspondence problem (image matching) failed.

Complements on Remote Sensing Images

A.1   Radar Imagery

Since the very first day, radar systems have evolved to become an essential and powerful tool in geosciences and remote sensing due to the all time (independence from solar illumination) and weather (cloud penetration) capabilities. Relevant acquisitions are possible in a short time for every area of the Earth. This active and coherent system which operates in the microwave frequencies enables the characterization of the back-scattered surface and the derivation of complementary information thanks to its high spatial resolution using polarization[1] or interferometry techniques.

A.1.1   Side-Looking Radar

Side-Looking Radar (SLR) instruments are used to acquire stripes coverage along one side of the platform track characterized by an azimuth direction at the altitude H. Fig. A.1 presents the configuration of an imaging radar system.
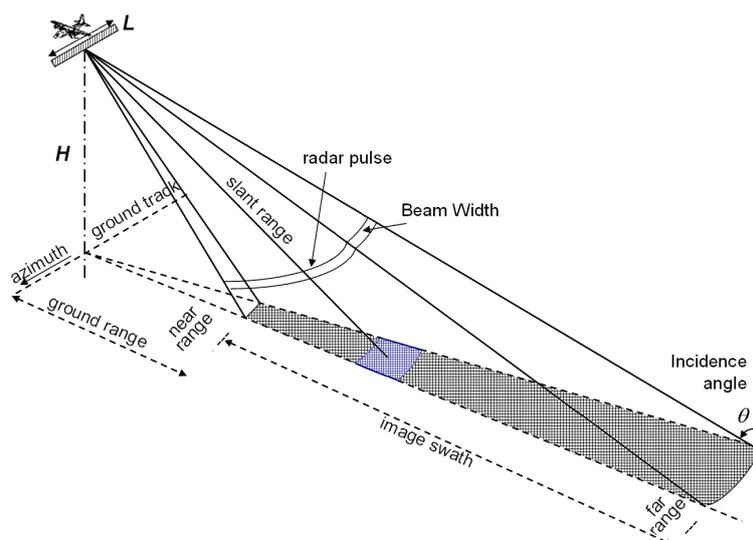
Figure A.1: Imaging Radar. The asymmetry of the side-looking configuration geometry solves the ambiguity by the acquisition of only one side of the trajectory.

---

[1]The polarization of the incident wave has a direct influence on the back-scattered information; a more thorough understanding of radar polarimetry can be found in [216], [201].

The radar antenna is parallel to the orbit and consists of a microwave transmitter and receiver. Short radar pulses are emitted perpendicularly to the satellite track (azimuth) in direction of the ground called "slant range". After a period of time Δt, the back-scattered reflection from the sensed surface is received by the antenna. Assuming the velocity of the electromagnetic waves, Δt contains the distance covered by the signal. In the case of a regular optical path, it corresponds to the distance between the sensor and the ground. This point emphasizes the difference between optical and radar imaging systems; the first one performs angle measures whereas radar computes distances.

Signal Properties and Acquisition Incidences

Radar sensors operate in the microwave domain of the electromagnetic spectrum with typical wavelengths between 1 cm and several meters. Radar imagery content results from the propagation characteristics of microwave and the wavelength of the incident wave is determinant for the interaction with the surface. Thanks to their long wavelength, microwaves can penetrate vegetation and even the ground up to a certain depth [202]. The penetration capabilities depend on the wavelength as well as on the complex dielectric constants, conductivities and densities of the observed targets:

- **Shorter wavelengths** (X, C bands, respectively $\sim$ 3 and 5.6 cm) typically show a high attenuation and are mainly back-scattered on the surface or on the top of the vegetation.

- **Longer wavelengths** (L, P bands) usually penetrate deeper into vegetation and also often into the ground. The back-scattering then contains contributions from the entire volume.

Radar images contain quite different information from images obtained out of optical or infrared sensors. In the optical range, the object reflectivity is mainly due to molecular resonances on the object surfaces. In a radar bandwidth spectrum, the dielectric and geometrical surface properties become relevant for the back-scattering. Because of the sensitivity to dielectric properties, SAR images can theoretically provide relevant information about the vegetation for agricultural and forestry applications. Depending on the incidence angle θ, the back-scatter return characterizes different contents; at small angles, the return signal is mainly dependent on θ and provides information on the surface slope while for large angles it describes small scale structures.

Surface Characteristics Incidences

The radar waves interactions with the surface are mainly driven by both physical and electrical scatterer properties[2]. Various back-scattering mechanisms have been proposed to model the surface wave interactions [58] (see Fig. A.2) and to characterize the amount of energy back-scattered.

- **Specular reflection**: in the case of a perfectly flat surface, the law of reflection states that the re-radiated wave is reflected in one single outgoing direction. Some parts of the wave are absorbed

---

[2]Surface roughness, morphology, slope, density variations, surface cover characterize some physical properties while dielectric constant, absorption or conductivity are typical electrical properties.

| (a) Acquisition Direction | (b) Specular Reflection | (c) Volume Scattering | (d) Multiple Bouncing |
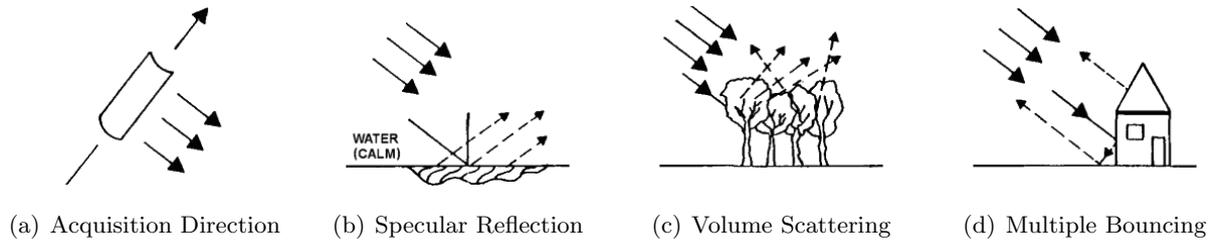
Figure A.2: Some back-scattering models. The corresponding grayscale SAR image representation will appear ranking from dark (b) to bright (d) areas.

by the materials. Except for the case of a wave emission perpendicular to the reflecting surface or double-bounding, flat surfaces such as water or roads are characterized by a low signal-to-noise ratio which leads to extremely "dark" areas in the resulting image.

- **Diffuse scattering**: in the case of rough surfaces, some of the energy is re-radiated in multiple directions and only a part of the energy is back-scattered.

- **Volume scattering** leads to the same effect as the diffuse scattering and mainly affects vegetation coverage such as forest areas; only a part of the wavefront penetrates the scatterers. The depth of penetration depends on the volume scattering characteristics (e.g. canopy density) and the wavelength used.

- **Multiple bouncing**: (generally double) in the case of urban areas (buildings) or corner reflectors; a large part of the emitted signal is received by the antenna.

The different scattering mechanisms and their image interpretation have a variety of geoscientific applications such as oceanography, polar ice or glacier monitoring, etc.

Side-Looking Aperture Radar

A key parameter often used to estimate the quality of a radar image is the resolution. It represents the minimum separation between two objects of equal reflectivity so that they can appear individually in a processed radar image. The most important criterion for establishing resolution is the size of the pulse rectangle projected into the ground at a given instant of time. Indeed, when two or more objects fall within the same pulse rectangle, they are presented as one echo to the radar system since they cannot be resolved as separate entities. Thus, the objects are imaged separately if they are distant from each other by a length exceeding the corresponding dimension of the pulse rectangle. The size of the pulse rectangle is controlled by two independent resolutions: in **Range** and in **Azimuth**, respectively perpendicular and parallel to the ground track in the form of

$$R_{Range} = \frac{c\,\tau}{2\,sin(\theta_v)} \qquad R_{Azimuth} = \frac{\lambda\,H}{L\,cos(\theta_v)} \qquad (A.1)$$

where L is the antenna length which emits in the incidence angle θ_v, pulse of duration τ at the wavelength λ.

The different relations given by Eq. A.1 are fundamental to understand the SLAR principle and its limitations: R_{Range} is independent of the distance between the scatterer and the sensor. Very short pulse durations are necessary to achieve high resolution in range. This resolution is purely a distance approach, controlled by τ. $R_{Azimuth}$ decreases with increasing flight height and the corresponding longer distances to the object. High resolution in azimuth requires large antennas and short object distances. In the azimuth direction, the resolution corresponds to the size of the antenna footprint on the ground. The angular resolution of an antenna in the azimuth direction is limited by the diffraction effects on its aperture, following the Huygens's principle [96].

Applying these relations for a SLAR with the same orbital characteristics as ERS's[3] gives resolutions in range and azimuth of respectively 14.2 and 4.5 km. These results are not acceptable for practical applications.

A.1.2 Synthetic Aperture Radar

The development of the synthetic array radar was limited by its very low resolution. These constraints have been circumvented using two specific techniques: the ***pulse compression*** for the range resolution [136] and the Doppler information to increase the azimuth resolution of the conventional SLAR system [211]; it's the principle of the ***Synthetic Aperture Radar*** (SAR).

Pulse Compression

The technique called pulse compression enables to get around the dependency of R_{range} with τ. The radar does not emit a rectangular signal modulated by a carrier frequency f_c, but a signal of duration τ, with a linearly modulated frequency called *chirp* [136]. The phase of the emitted signal is limited in time by $\Delta t = \tau$ (Eq. A.2), B being the signal bandwidth,

$$\Phi_{range}(t) = 2\pi \left(f_c \, t + \frac{K \, t^2}{2} \right) \qquad B = K \, \tau \tag{A.2}$$

The received signal is passed through a matched filter [156] which is equivalent to a convolution with an ideal chirp. According to Bamler in [12], the range resolution leads to [4]

$$R_{range} = \frac{c}{2B \, sin(\theta_v)} \tag{A.3}$$

---

[3]The first and second European Remote-Sensing Satellites ERS-1 and ERS-2 were developed by the European Space Agency (ESA) and launched respectively in 1991 and 1995 on a sun synchronous orbit at an altitude H of $780km$ with a 23 degree incidence angle θ_v. A SAR sensor ($L = 10m$) which operates in C-Band ($\lambda = 5.3 \; cm$, $\tau = 37.1 \; s$, bandwidth $B = 15.5MHz$) is on board.

[4]When comparing equations A.1 and A.3, we observe it is equivalent to the emission of a rectangular pulse of duration $\tau = 1/B$. In the case of ERS, R_{range} is increased from 14.2km to 25m.

Synthetic Aperture

The limitation of the azimuth resolution is the antenna's diffraction. Its main lobe width at $-3db$ is given by λ/L; a huge antenna would be necessary to obtain an important resolution. A Synthetic Aperture Radar overcomes these problems and is designed to achieve high resolutions with small antennas over long distances. A SAR system exploits the response of a ground scatterer which contains more than a single radar echo, and shows a typical phase history over the illumination time. A coherent combination of several pulses leads to the formation of a synthetically enlarged antenna - the so-called "Synthetic Aperture". This formation is quite similar to the control of an antenna array, with the difference that only one antenna is used and the different antenna positions are generatedby the movement of the platform. This simulates a larger antenna aperture. As a result, the azimuth resolution is increased and depends only on the antenna length $R_{Azimuth} = L/2$.

Using this technique, the ERS satellite ($L = 10m$) resolution in azimuth grows to 5m. Interestingly, the achieved resolution is now independent of the range distance and is determined only by the size of the real antenna. This phenomenon is explained by the proportional relation between the length of the synthetic aperture and the distance to the object (homothetic triangles).

Contrary to a SLAR, a shorter antenna produces higher resolution thanks to its wider angular radiation characteristics. The problem of small back-scattering power prevents from obtaining extremely high resolution. Therefore, a compromise has to be reached; an antenna long enough is necessary for an adequate focus of the power. Until recently, due to several technical limitations pointed out, only airborne SAR were able to produce metric resolution data. In June 2007, The German Space Agency (DLR) launched the satellite TerraSAR-X, a metric SAR sensor which operates in X-Band with full polarimetry capabilities [9].

Image Geometry

The radar sensor measures distances in the slant-range direction. Even under the hypothesis of a planar illuminated surface, the asymmetry of the system (side-looking) leads to a ground resolution decrease $\Delta y = \Delta x / sin(\theta_i)$ from the near range to the far range, Δx being the constant slant-range resolution. Its high sensitivity to the relief leads to specific geometrical effects which complicate, compared to optical data, the interpretation and the analysis of radar images. Fig. A.3 summarizes the combination of both system acquisition and local surface incidence α_i.

- **Foreshortening**, **Elongation**: $-\theta_i < \alpha_i < \theta_i$, correspond respectively to a compression ($0 < \alpha_i < \theta_i$) and dilation ($-\theta_i < \alpha_i < 0$) of the ground resolution cell.

- **Layover**: $\alpha_i \geq \theta_i$, is an extreme case of foreshortening which causes severe image distortion in inverting the image geometry. It occurs when the top of the mountain is closer to the sensor than its bottom (due to steep terrain slopes)[5].

---

[5]Like foreshortening, layover is most severe for small incidence angles, at the near range of a swath, and in mountainous terrains. The case where $\alpha_i = \theta_i$, resulting from a strong back-scattering, leads to brighter areas.
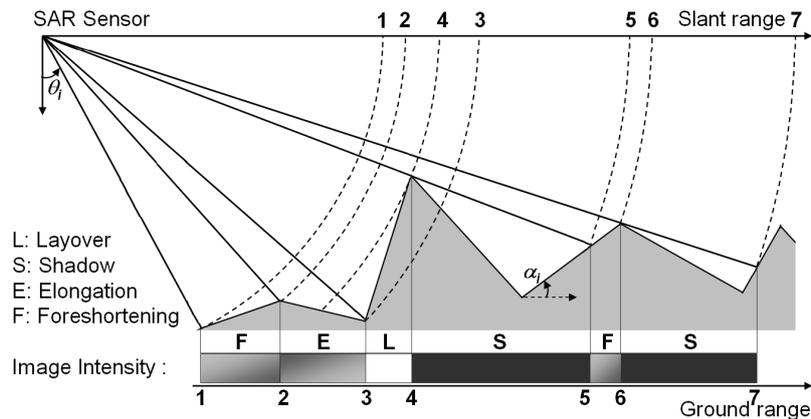
Figure A.3: Geometrical effects. Incidence of the conversion of slant-range to ground-range measures depending on acquisition conditions as well as on local surface variations. Those deformations need to be compensated and re-sampled to a ground range resolution.

- **Shadow**: $\alpha_i \leq \theta_i - \pi/2$, occurs when the radar beam is not able to illuminate the ground surface. These areas appear as dark regions corrupted by thermal noise.

Data radiometry and geometry are affected by the topography of the sensed area and have to be corrected. A post-processing step, called **Geocoding**, is further required to generate SAR images re-sampled to a uniform grid [12]. The use of a DEM allows to take into account the relief and performs the correction (according to the DEM's quality and resolution). The reader is referred in [58] to complementary information about additional problems arising during the SAR processing step, such as the platform motion stability, moving target effects, Earth rotation, range migration, ambiguities resolutions, atmospheric conditions, etc.

A.1.3   SAR Image Statistics

All coherent imaging systems (laser, acoustics, radar images) are affected by a granular and rather strong noise called "speckle" [76] caused by random interferences between the electromagnetic waves reflected from the different scatterers present in the single resolution cell.

Speckle

Scattering properties of the sensed area can only be expressed in terms of statistical parameters; it prevents from considering the SAR raw signal, a random process, as a deterministic signal. Considering the metric-decametric SAR resolution cell and compared to centimetric waves and the roughness and/or diffuse nature of the illuminated surface, the resulting cell constitutes a coherent summation of a large number N of single returned micro-scatterers, each of them being characterized by a complex signal (magnitude E_{i0} and phase Φ_i).

If the number of individual scatterers is important, according to the central limit theorem, the real and imaginary parts of the received signal are Gaussian-distributed with $\mathcal{N}(0, \sigma^2)$. In this case, the speckle is assumed to be fully developed [200]; its considered phases are uniformly distributed between 0 and 2π and magnitudes are statistically independent. This assumption is not valid for the signal returned from specular scatterers. The term $\sigma^2/2$ is equivalent to the so-called radar reflectivity. Neglecting sensor noise, this quantity is proportional to the radar back-scatter coefficient or cross-section. The speckle effect is only visible in the detected amplitude or intensity signal. Indeed, only thermal noise and effects of signal processing corrupt the complex signal.

The correlation of speckle in SAR images induces the granular visual appearance of spatial structures. Assuming a stationary process and statistical independence between the mean intensity μ_i and the speckle noise n_I, the speckle auto-correlation resulting from the SAR process follows

$$R_{n_i}(\Delta x, \Delta y) = 1 + C_{n_i}(\Delta x, \Delta y) \quad , \text{with} \quad C_{n_i}(\Delta x, \Delta y) = \frac{1}{L} \, sinc^2\left(\frac{\Delta x}{r_x}\right) sinc^2\left(\frac{\Delta y}{r_y}\right) \quad (A.4)$$

where r_x, r_y denotes the cell resolution in x (range) and y (azimuth) directions. The correlation C_{n_i} resulting from the SAR processing, i.e. the system's point-spread function, is responsible for the granular aspect of speckle which decreases when the number of looks L grows.

Multi-look Data

To reduce the speckle variance, SAR images are often processed with a technique called multi-looking, either performed in the frequency or in the spatial domain, by averaging respectively different parts of the spectrum (L looks) or neighboring L pixels. The speckle variance is reduced by L and consequently the radiometric resolution is increased, but the spatial resolution is also decreased by the same factor. A compromise has to be found between radiometric and spatial resolution; very often $L = 3$ is used for decametric data. In the case of a uniform cross-section, the following relations can be derived:

- Multi-look intensity speckle adheres to a Gamma distribution characterized by the density

$$p(I = i_0, M_i = \mu_i) = \frac{L^L \, i_0^{L-1}}{\mu_i^L \, \Gamma(L)} \exp\left(-\frac{L \, i_0}{\mu_i}\right) \quad (A.5)$$

where L denotes the number of looks of the data, $\Gamma(L)$ is the Gamma function and μ_i is the expected value. An interesting property of this probability density function around a given mean value μ_i is its interpretation as a multiplicative noise.

- Multilook square root images are instead squareroot Gamma distributed, with μ_a the square root of the expectation value of the amplitude

$$p(A = a | M_a = \mu_a) = 2 \frac{L^L \, a^{2L-1}}{\mu_a^{2L} \, \Gamma(L)} \exp\left(-\frac{L \, a^2}{\mu_a^2}\right) \quad (A.6)$$

Complements on Image Analysis

B.1 Bayesian Image Analysis

B.1.1 Introduction

The interpretation of high complexity RS images by traditional analysis methods is generally incomplete. As proposed by Jeffreys in his book [102], the Bayesian approach is an alternative to frequentist statistics since the notion of probability is regarded as a real number measure - scaled between 0 and 1 - of the plausibility of a proposition when incomplete knowledge does not allow us to establish its truth (1) or falsehood (0) with certainty. For Frequentists, it constitutes a measure of the relative frequency of occurrence that a random event outcomes from a given well-defined random experiment.

By analogy with the comparison presented in introduction between topographic campaigns and remote sensing acquisition for the retrieval of the scene description, the consequences of a certain cause given can be described having a deterministic model and lead, according to our defined models, to a number of possible interpretations. Such a concept is called "deductive logic". For Jaynes in [101], the Bayesian theory can be viewed as the extension of deductive logic to cases where there is uncertainty. The reader is referred in [101] to a deep development about the Bayesian Theory. To come back to our motivations, we are faced with the reverse problem: given a noisy RS image (consequence), some pre-processing operations, retrieve the scene description of the Earth surface encapsulated in the data.

B.1.2 Classical Approach of Estimation Theory

To make the understanding easier, the following conventions are adopted: *random variables* are written in capital letters (X, Y), their realizations in small letters (x). The assignment is expressed by $X = x$. For notation simplification, p is used to represent both probability density and probability distribution functions *pdf*. The pdf $p_X(x)$ of a random variable X, is written $p(X = x)$ as well. The *distribution function* of a random variable X defined on $\mathbb{R}$ is a cumulative function $F(x) = P_r(X \leq x)$, with $F(-\infty) = 0$, $F(\infty) = 1$, and $F(b) - F(a) = P_r(a < X < b)$. The corresponding *probability density function* is defined by

$$p(x) = \frac{dF(x)}{dx} \iff P_r(x < X \leq x + dx) \underset{dx \to 0}{=} p(x)dx \tag{B.1}$$

Having several random variables $\{X, Y\}$, their possible dependencies are expressed using conditional probabilities

$$p(Y = y|X = x) = \frac{p(Y = y, X = x)}{p(X = x)} \tag{B.2}$$

It expresses the probability of some event y, given the occurrence of some other event x. The celebrated Bayes formula is then easily derived from the conditional probability relation (Eq. B.2)

$$p(X = x|Y = y, I) = \frac{p(Y = y|X = x, I)p(X = x|I)}{p(Y = y|I)} \Longleftrightarrow posterior = \frac{likelihood \times prior}{evidence} \tag{B.3}$$

, I being some initial information. If for instance X represents the hypothesis we want to access given some data y, based on some new information Y, we clearly notice through the Bayes theorem, having an existing description of a phenomenon X (the *prior*), the possibility to compute an update of X, the *posterior*. The data prediction $p(Y = y|X = x, I)$ is called the *likelihood*. The *evidence* term $p(Y = y|I)$ which corresponds to the data probability, is computed through the marginalization of the prior model

$$p(Y = y|I) = \int_X p(Y = y|X = x, I) \ p(X = x|I) \ dx \tag{B.4}$$

The Bayesian approach enables consistent *inference* processes by using models in the form of probability distributions to be estimated.

Parameter Estimation

Estimation theory is part of statistical decision theory and deals with estimating the values of parameters X based on measured data corrupted by noise. Based on the observations Y, an *estimator* $\hat{x}$ is an approximation or a guess of the unknown parameter values. X is modeled as a random signal (in our case, an image) and seen as a realization of a stochastic process.

Type of estimator	Equation $c(x, \varepsilon_x) =$	Estimate
Min. Mean Square Error	$(x - \hat{x})^2$	$\hat{x}_{MMSE} = \int_X x \ p(x|y) \ dx$
Maximum Likelihood		$\hat{x}_{ML} = \arg \max_x \log p(x|y)$
Maximum *a posteriori*	$\begin{cases} 0 : & |x - \hat{x}| \leq \frac{\delta}{2} \\ 1 : & else \end{cases}$	$\hat{x}_{MAP} = \arg \max_x \{\log p(x|y) + \log p(x)\}$

Table B.1: Classical estimators and their corresponding cost functions.

Since X is unknown, the estimation error $\varepsilon_X = x - \hat{x}$ cannot be found directly. A measure of ε_x is derived using mainly uniform or quadratic cost functions $c(X = x, \varepsilon_X)$. Some of them are described in Table B.1 where $p(Y = y)$ is written $p(y)$. In this scope, estimation parameter problems are formulated by minimizing the Bayes Risk R (Eq. B.5). R denotes the expectation of the cost function defined over the joint space of the pdf $p(X = x, Y = y)$

$$R = E[c(X = x, \varepsilon_x)]p(Y = y|I) = \int_X \int_Y c(X = x, \hat{x}_x)p(X = x, Y = y) \ dx dy \tag{B.5}$$

Several remarks are pointed out. In classical estimation theory, the cost function is seen as modeling some prior information. The evidence term from the Bayes's formula defined in Eq. (B.3) is neglected and used as normalization factor. The likelihood distribution encapsulates both the forward model of the specific sensor and its noise. Since the posterior distribution is inferred from the likelihood distribution and the *a priori* model of the desired parameter, the MAP estimator is the most complete approach in information extraction. The use of different noise and prior models allows to retrieve all the other presented estimators.

Estimator Quality

This section presents roughly some notions required to measure the goodness of an estimator. An excellent description of Estimation Theory is dressed by Papoulis in [160]. Two celebrated information measures are *Shannon's entropy* and *Kullback-Leibler divergence*. The entropy $S(X)$ represents the average information contained in a random variable X with the pdf $p(x)$

$$S(X) = -\int_X p(X = x) \log p(X = x) dx$$

A generalization of the entropy is the Kullback-Leibler divergence. It denotes a function of two pdfs potentially characterizing a random variable X

$$L(p, q) = \int_X p(X = x) \log \frac{p(X = x)}{q(X = x}} dx$$

An estimator $\hat{x}(y)$ is itself a random variable, thus its quality can be evaluated by its conditional mean and by an even more important quality criterion, its conditional variance $\sigma_{\hat{x}}^2$ (the smaller $\sigma_{\hat{x}}^2$, the better $\hat{x}$). In addition, the optimum quality limit is given by the Cramer Rao bound. For a vector parameter, the Fisher information matrix measures the lower bound of the accuracy and the covariance matrix gives its quality.

Both the Cramer Rao bound and the Fisher information matrix are expressed as a function of the Kullback-Leibler (K-L) divergence. The K-L divergence shows that the bounds of a quality estimator are given by the likelihood fitting to a desired true likelihood. This can be understood as a comparison of the likelihood distributions allowing a decision about the identity of the data. It constitutes the basis for describing the information content of a stochastic model.

B.1.3   Gibbs and Markov Random Fields

To properly formulate prior pdf's of high complexity images (both spectral and structural), a solution is the application of a set of models with local properties.

Stochastic and Markov Processes

Let's define a one-dimensional discrete signal with N samples as a realization of a stochastic process. We then consider a set $X = \{X_1, ...X_N\}$ of N stochastic variables and its corresponding

set of values $x = \{x_1, ...x_N\}$. A stochastic process is fully determined by the knowledge of all pdfs; we can thus describe X according to its joint probability $p(X = x) = p(X_1 = x_1, ...X_N = x_N)$ Based on the definition of the conditional probability, the joint probability is expressed

$$p(X = x) = p(X_1 = x_1)p(X_2 = x_2|X_1 = x_1)...p(X_N = x_N|X_1 = x_1, ...X_{N-1} = x_{N-1})$$

The resulting decomposition is called *chain rule*. In practice, for analytical tractability, such formalism is not applied since p(X=x) is decomposed in 2^N terms, which is unexploitable for a large amount of data from a combinatorial point of view. To overcome this problem, a solution is the application of Markov processes. A first-order Markov process is defined by two conditions:

positivity: $\quad p(X = x_k) > 0 \quad \forall\, k \in \{1, N\}$

Markovianity: $\quad p(X_{k+1} = x_{k+1}|X_1 = x_1, ...X_k = x_k) = p(X_{k+1} = x_{k+1})|X_k = x_k)$

In other words, the Markovianity property means that the state of the process X_{k+1} only depends on the preceding state X_k. We can therefore rewrite $p(X = x)$ in the form

$$p(X = x) = p(X_1 = x_1) \prod_{k=2}^{N} p(X_k = x_k|X_{k-1} = x_{k-1})$$

Cliques and neighborhood system

Let S be an image defined on a regular lattice $\mathcal{L}$ whose sites are pixels and $\mathcal{N}$ a neighborhood system (see section 2.1.3, page 12). The definition of a neighborhood system enables to introduce a proper sub-graph of a neighborhood graph on $\mathcal{L}$ called *clique*. A clique C is defined as a subset of the lattice $\mathcal{L}$ which satisfies either of the following conditions:

- C consists of a single site
- all pairs of distinct sites in C are neighbors $\{s, r \in C, r \neq s\} \Leftrightarrow s \in \mathcal{N}_r, r \in \mathcal{N}_s$

Based on this definition, the *local clique set* C_s for the site s represents the set of cliques of S which contains s. Depending on the neighborhood system employed, several cliques are defined.

Markov Random Fields

The concept of Markov random fields is a generalization of Markov processes to two-dimensional signals (e.g. an image). X, defined on the set S, is a Markov random field on S with respect of $\mathcal{N}$ if it satisfies the Markovianity condition

$$p(X_k = x_k|X_{S\setminus\{k\}} = x_{S\setminus\{k\}}) = p(X_k = x_k)|X_{\mathcal{N}_k} = x_{\mathcal{N}_k})$$

where $S \setminus \{k\}$ is the set difference and $x_{\mathcal{N}_k} = \{x_l|l \in \mathcal{N}_k\}$ The Markovianity depicts the local characteristics and limit the direct interactions to $\mathcal{N}$. Markov random fields are suitable to describe correlations of dependencies between neighborhood pixels. The interest of such a mathematical model is motivated by the similar statistical or structural properties locally observed in images.

Gibbs Random Fields

A Gibbs random field (GRF) is a random field which follows a Gibbs distribution

$$p(X = x) = \frac{1}{Z} \exp\left(-\frac{1}{T}E(x)\right) \quad , \text{with} \quad Z = \sum_x \exp\left(-\frac{1}{T}E(x)\right) \tag{B.6}$$

where the constant T, called *temperature*, controls the distribution sharpness in the same way as the variance for Gaussian distribution[1]. Z is a normalization term defined over the entire configuration space, therefore its computation is almost unreachable. E is an energy function defined as the sum of potential functions V_c over the set of cliques C:

$$E(X = x) = \sum_{c \in C} V_c(x) \tag{B.7}$$

The potential function V_c is a parametric model $\{\theta_c\}$ which quantifies the strength of statistical dependence between pixels (pixel interaction) within a clique c. The Gibbs distribution $p(X = x)$ appears as a probability measure of a particular configuration or structure in the image. Such properties are extremely interesting for texture modeling.

Markov-Gibbs Equivalence

The Markov property describes conditional dependencies of image signals within a local neighborhood system, which is proved to be useful in capturing and modeling highly localized structures or textures. However, computing the full joint pdf from the conditional pdfs is almost intractable. GRFs have global properties and enable to describe in a parametric way the structural patterns in an image. The equivalence between MRF and GRF is established by the *Hammersley-Clifford* theorem [73]. It states that a random field is a Markov random field if and only if the corresponding joint probability distribution is a Gibbs distribution. The conditional probability of a MRF is therefore rewritten in terms of clique potential V_c

$$p(X_k = x_k | X_{S-\{k\}} = x_{S-\{k\}}) = \frac{\exp^{-\sum_{c \in C} V_c(x)}}{\sum_{x_{S-\{k\}}} \exp^{-\sum_{c \in C} V_c(x)}} \tag{B.8}$$

The Markov-Gibbs equivalence allows to use global parameter models which are locally estimated and provides a theoretical basis for modeling prior pdfs within the Bayesian framework.

B.1.4   Classical Pdf Models

Parameter estimation requires the knowledge of the likelihood and the prior pdfs. The quality of the resulting estimate is conditioned by the applied models and their capacity to describe the image content. We focus our attention to the image and noise models used in [46], [209], [208] and further improved in [171] for image information extraction and denoising problems.

---

[1]The Gaussian distribution belongs to the Gibbs distribution family.

Let's consider X and Y as stochastic processes. The associated realizations x, y denote respectively noise-free images and observations. Given the image y_i, corrupted by a noise which follows an assumed pdf model, retrieve the noise-free image x_i. In the Bayesian approach, the likelihood acts to represent the noise model.

Gauss Markov Random Fields as Prior Pdf

In [46], Gauss Markov random fields (GMRF) models are used to represent the prior model, which comes in the form of a textured model, as depicted in Eq. (B.9):

$$p(X_s = x_s | X_{s+r} = x_{s+r}, s + r \in C, \theta_{s+r}) = \frac{1}{\sqrt{2\pi\sigma_s^2}} \exp\left(-\frac{(x_s - \sum_{r+s\in C} \theta_{s+r}.x_{s+r})^2}{2\sigma_s^2}\right) \quad \text{(B.9)}$$

, where x_s is the pixel value centered on the neighborhood of cliques C and normally distributed (variance σ_s^2) around its weighted mean $\bar{x}_s$. As introduced in [30], $\bar{x}_s$ is computed using a symmetrical set of scalar parameters $\theta = \{\theta_1, ...\theta_m\}$ associated to the neighborhood of cliques C around the site s. The set of hyper-parameters θ denotes the symmetric parametric model which represents the GMRF. The local neighborhood interaction configurations are expressed by the model order. High model orders allow the description of complex data structures defined on wider vicinities.

A GMRF, also known as *Auto-Normal* model, is a special case of GRF whose no contextual interaction occurs. The set of cliques C is therefore only constituted of single site clique potentials which enable, combined with the model linearity, relatively fast computational implementations. For parameter estimation, a technique is proposed in [29] to represent the model in the frequency domain and in [123] in the wavelet domain. Depending on the image characteristics and the corresponding noise modeling associated, suitable likelihood pdfs have to be used.

Likelihood Pdf Models

In the case of data corrupted by an additive *Gaussian* noise, the relation between noise-free pixels and their noisy instances is expressed using a **Gaussian likelihood**

$$p(Y_i = y_i | X_i = x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - x_i)^2}{2\sigma^2}\right) \quad \text{(B.10)}$$

where the variance σ_2 denotes the strength of the distribution centered around x_i.

The phenomenon of fully developed speckle noise in SAR intensity images is generally modeled as a multiplicative noise which follows a Gamma distribution (see chapter A.1.3, page 173). To filter square-root SAR images, a **Gamma likelihood** is expressed as a square root Gamma distribution

$$p(Y_i = y_i | X_i = x_i) = \frac{L^L \ y_i^{L-1}}{x_i^L \ \Gamma(L)} \exp\left(-\frac{L \ y_i}{x_i}\right) \quad \text{(B.11)}$$

where L denotes the number of looks of the data and $\Gamma(L)$ is the Gamma function.

B.2   Some Pseudo Code

Two pseudo codes are presented for region or object labeling given a connectivity system $\mathcal{N}$. An incremental region growing is presented in Alg. 6. It can be extended to image segmentation by operating multiple passes and relaxing a predicate function $\mathcal{P}()$. Alg. 7 illustrates the proposed method to label features or regions.

Algorithm 6: An incremental Region growing method.

Input: Image I to be labeled (m pixels) using $\mathcal{P}()$ and $\mathcal{N}$

Output: IndexMap and a list of region $S = \{R_1...R_n\}$, with $I = \bigcup_{i=1}^{n} R_i$ and $n \leq m$

begin

    Initialize label i to 1 and set all pixels of indexMap to -1

    for $p \leftarrow 1$ **to** m **do**

        if $indexMap(p) = -1$ **then**

            Create R_i and label indexMap$(p) = i$

            Initialize list of pixels $\in R_i$: listPos$(1) = p$, $nb = 1$

            for $k \leftarrow 1$ **to** nb **do**

                if $indexMap(listPos(k)) = \text{-1 or } k = 1$ **then**

                    forall $\mathcal{N}_j$ *of* $\mathcal{N}$ **do**

                        Scan neighbourhood: $q =$listPos$(k)_{\mathcal{N}_j}$

                        if $indexMap(q) = -1 \text{ and } \mathcal{P}(q,p) = true$ **then**

                            Label: indexMap$(q) = i$

                            Add to the list of pixel: listPos$(nb) = q$, $nb = nb + 1$

                        end

                  end

                end

            end

            Add listPos toR_i

            Increment label: $i = i + 1$

        end

    end

end

Algorithm 7: An image labeling algorithm.

Input: Image to be labeled (m pixels) using $\mathcal{P}()$ and $\mathcal{N}$

Output: IndexMap and a list of region $S = \{R_1...R_n\}$ with $I = \bigcup_{i=1}^{n} R_i$ and $n \leq m$

begin

    Initialize label i to 1 and set all pixels of indexMap to -1

    for $p \leftarrow 1$ **to** m **do**

        if $indexMap(p) = -1$ **then**

            forall $\mathcal{N}_j$ *of* $\mathcal{N}$ **do**

                if $indexMap(p_{\mathcal{N}_j}) > -1$ *and* $\mathcal{P}(p, p_{\mathcal{N}_j}) = true$ **then**

                    indexMap$(p) = $ indexMap$(p_{\mathcal{N}_j})$

                    break

                end

            end

            if $indexMap(p) = -1$ **then**

                indexMap$(p) = i$

                Increment label: $i = i + 1$

            end

        forall $\mathcal{N}_j$ *of* $\mathcal{N}$ **do**

            if $indexMap(p_{\mathcal{N}_j}) = -1$ *and* $\mathcal{P}(p, p_{\mathcal{N}_j}) = true$ **then**

                indexMap$(p_{\mathcal{N}_j}) = $ indexMap(p)

            end

        end

    end

    Build list of region $S = \{R_1...R_i\}$

    for $p \leftarrow 1$ **to** m **do**

        $a = $ indexMap(p)

        Add p to R_a

    end

    for $p \leftarrow m$ **to** 1 **do**

        forall $\mathcal{N}_j$ *of* $\mathcal{N}$ **do**

            $a = $ indexMap(p), $b = $ indexMap$(p_{\mathcal{N}_j})$

            if $a \neq b$ *and* $\mathcal{P}(R_a, R_b) = true$ **then**

                merge R_a and R_b to R_a

                update indexMap

                delete R_b

            end

        end

    end

end

Bibliography

[1] R. Adams and L. Bischof. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6):641–647, 1994. 60

[2] L. Ambrosio. Variational problems in SBV and image segmentation. *Acta Appl. Math.*, 17(1):1–40, 1989. 62

[3] L. Ambrosio and V. M. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via Γ-convergence. *Communications on Pure and Applied Mathematics*, 43(8):999–1036, 1990. 62

[4] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations (second edition)*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, 377 pages, 2006. 62

[5] I. Ben Ayed, A. Mitiche, and Z. Belhadj. Multiregion level-set partitioning of synthetic aperture radar images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):793–800, 2005. 65

[6] I. Ben Ayed, A. Mitiche, and Z. Belhadj. Polarimetric image segmentation via Maximum-Likelihood approximation and efficient multiphase level-sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1493–1500, 2006. 65

[7] G.P. Babu and M.N. Murty. A near-optimal initial seed value selection in K-Means algorithm using a genetic algorithm. *Pattern Recognition Letters*, 14(10):763–769, 1993. 56

[8] G.H. Ball and D.J. Hall. ISODATA, a novel method of data analysis and pattern classification. *Technical Report, No4, Menlo Park California Stanford Research Institute*, 1965. 56

[9] R. Bamler and M. Eineder. The pyramids of Gizeh seen by TerraSAR-X a prime example for unexpected scattering mechanisms in SAR. *IEEE Geoscience and Remote Sensing Letters*, 5(3):468–470, 2008. 2, 42, 172

[10] R. Bamler, M. Eineder, B. Kampes, H. Runge, and N. Adam. SRTM and beyond: current situation and new developments in spaceborne SAR and InSAR. In *Proceedings of the ISPRS, Workshop High Resolution Mapping from Space, on CD-ROM*, 2003. 41, 125

[11] R. Bamler and P. Hartl. Synthetic aperture radar interferometry. *Inverse Problems*, 14(3):R1–R54, 1998. 38

[12] R. Bamler and B. Schaettler. SAR data acquisition and image formation. In G. Schreier, editor, *SAR Geocoding: Data and Systems*, pages 53–102. Wichmann Verlag, Karlsruhe, 1993. 171, 173

[13] A. Barron, J. Rissanen, and B. Yu. The minimum description length principle in coding and modeling.

IEEE Transactions on Information Theory, 44(6):2743–2760, 1988. 61

[14] J.M. Beaulieu and M. Goldberg. Hierarchy in picture segmentation : A stepwise optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(2):150–163, 1989. 60

[15] P.J. Besl and R.C. Jain. Segmentation through symbolic surface descriptions. *International Conference in Computer Vision and Pattern Recognition*, 86:77–85, 1986. 80

[16] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, 35:99–109, 1943. 18

[17] A. Blake and S. Zimmerman. *Visual Reconstruction*. MIT Press, Cambridge, MA 225 pages, 1987. 61

[18] A. Bonnet. On the regularity of edges in image segmentation. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 13(4):485–528, 1996. 62

[19] C.A. Bouman and M. Shapiro. A multiscale random field model for Bayesian image segmentation. *IEEE Transactions on Image Processing*, 3(2):162–177, 1994. 4, 57, 60, 121

[20] P.S. Bradley and U. M. Fayyad. Refining initial points for K-Means clustering. In *Proc. 15th International Conf. on Machine Learning*, pages 91–99. Morgan Kaufmann, San Francisco, CA, 1998. 56

[21] J.P. Braquelaire and J.P. Domenger. Representation of region segmented images with discrete geometric maps. *Image and Vision Computing*, 17(10):715–735, 1999. 92

[22] C.R. Brice and C.L. Fennema. Scene analysis using regions. *Artificial Intelligence*, 1(3):205–226, 1970. 88, 90

[23] T. Brox and D. Cremers. On the statistical interpretation of the piecewise smooth mumford-shah functional. In *Scale Space and Variational Methods in Computer Vision (SSVM07)*, pages 202–213, 2007. 64

[24] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986. 3, 24, 54

[25] V. Caselles, B. Coll, and J.M. Morel. Topographic maps and local contrast changes in natural images. *International Journal of Computer Vision*, 33(1):5–27, 1999. 24, 100

[26] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997. 3, 63

[27] T.F. Chan and J. Shen. *Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods*. Society for Industrial and Applied Mathematics (SIAM) Philadelphia, PA, 400 pages, 2006. 4, 62, 152, 158

[28] T.F. Chan and L.A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001. 65

[29] R. Chellappa. Two dimensional discrete Gaussian Markov random field models for image processing. In L. N. Kanal and A. Rosenfeld, editors, *Progress in Pattern Recognition*, pages 79–112. North Holland Pub. Co., 1985. 180

[30] R. Chellappa, S. Chatterjee, and R. Bagdazian. Texture synthesis and compression using Gaussian-Markov random field models. *IEEE Transactions on Systems, Man, and Cybernetics*, 15(2):298–303, 1985. 180

[31] P.C. Chen and T. Pavlidis. Image segmentation as an estimation problem. In A. Rosenfeld, editor, *Image Modelling*, pages 9–28. Academic Press, 1981. 60

[32] P. Cignoni, F. Ganovelli, E. Gobbetti, F. Marton, F. Ponchio, and R. Scopigno. BDAM – batched dynamic adaptive meshes for high performance terrain visualization. In *Proc. Eurographics 2003*,

volume 22(3) of *Computer Graphics Forum*, pages 505–514, 2003. 32

[33] P. Cignoni, F. Ganovelli, E. Gobbetti, F. Marton, F. Ponchio, and R. Scopigno. Adaptive tetra-puzzles — efficient out-of-core construction and visualization of gigantic polygonal models. In *Proc. SIGGRAPH 2004*, volume 23(3) of *ACM Transactions on Graphics*, pages 796–803, 2004. 32

[34] M. Ciucu, P. Héas, M. Datcu, and J.C. Tilton. Scale space exploration for mining image information content. In *Mining Multimedia and Complex Data, Revised Papers from KDD Workshop MDM/KDD 2002*, volume 2797 of *Lecture Notes in Computer Science*, pages 119–133. Springer, 2003. 55

[35] L.D. Cohen. On active contour models and balloons. *Computer Vision, Graphics and Image Processing*, 53(2):211–218, 1991. 64

[36] D. Comaniciu. An algorithm for data-driven bandwith selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(2):281–288, 2003. 58

[37] D. Comaniciu and P. Meer. Robust analysis of feature spaces: Color image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition, (CVPR'97)*, pages 750–755, 1997. 4, 54, 57, 115, 116

[38] D. Comaniciu and P. Meer. Mean Shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002. 57, 58

[39] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994. 23

[40] M. Cord, M. Jordan, and J.P. Cocquerez. Accurate building structure recovery from high resolution aerial imagery. *Computer Vision and Image Understanding*, 82:138–173, 2001. 156

[41] R.W. Cottle and G.B. Dantzig. Complementary pivot theory of mathematical programming. *Linear Algebra and its applications*, 1:103–125, 1968. 157

[42] T.R. Crimmins. Geometric filter for reducing speckle. *Optical Engineering*, 25(5):651–654, 1986. 119

[43] G. Damiand, Y. Bertrand, and C. Fiorio. Topological model for two-dimensional image representation: Definition and optimal extraction algorithm. *Computer Vision and Image Understanding*, 93(2):111–154, 2004. 92, 93

[44] P. D'Angelo, M. Lehner, T. Krauss, D. Hoja, and P. Reinartz. Towards automated DEM generation from high resolution stereo satellite images. In *ISPRS Proceeding, Commission IV, WG IV/9*, 2008. 34, 38, 45, 125

[45] M. Datcu, D. Luca, and K. Seidel. Wavelet-based digital elevation model analysis. In *Proceedings of the 16th EARSeL Symposium, Malta*, pages 283–292, 1996. 138

[46] M. Datcu, K. Seidel, and M. Walessa. Spatial information retrieval from remote sensing images: - part 1: Information theoretical perspectives. *IEEE Transactions on Geoscience and Remote Sensing*, 36(5):1431–1445, 1998. 123, 179, 180

[47] I. Daubechies. *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, USA, 357 pages, 1992. 12, 129

[48] R. Deriche. Using Canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision*, 1(2):167–187, 1987. 3, 25, 54, 144

[49] A. Desolneux, L. Moisan, and J.M. Morel. From gestalt theory to image analysis a probabilistic approach. In *Interdisciplinary Applied Mathematics (vol. 34)*. Springer-Verlag New York, Inc., 276 pages, 2008. 9, 24

[50] Centre National d'Études Spatiales. Méthodes de fusion multi-résolution pour la synthèse d'images. In *Centre de Comptéences Techniques en Traitement du Signal et des Images, Seminar FUSION, CNES*

Toulouse, December 16-17 2002, 2002. 43

[51] D.L. Donoho. De-noising by soft thresolding. *IEEE Transactions on Information Theory*, 41(3):613–627, 1995. 5, 122, 129

[52] D.L. Donoho and I.M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 8:425–455, 1994. 129

[53] D.L. Donoho and I.M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, 1995. 122, 129

[54] R.O. Duda, P.E. Hart, and D.G. Stork. *Pattern Classification*. Wiley-Interscience, 654 pages, second edition, 2000. 22, 26, 109

[55] J.F. Dufourd and F. Puitg. Functional specifications and prototyping with oriented combinatorial maps. *Journal of Computational Geometry*, 16(2):129–156, 2000. 92

[56] M. Durupt and F. Taillandier. Automatic building reconstruction from a digital elevation model and cadastral data: An operational approach. In *ISPRS Commission III Symposium on Photogrammetric Computer Vision (PCV06), Bonn*, 2006. 146

[57] M. Eineder. Problems and solutions for InSAR digital elevation models generation of mountainous terrain. In *Proceedings of FRINGE Workshop, Frascatti, 1-5 December 2003, ESA SP-550*, 2003. 49

[58] C. Elachi. *Spaceborne Radar Remote Sensing: Applications and Techniques*. IEEE, 288 pages, 1988. 169, 173

[59] S. Oude Elberink and G. Vosselman. Adding the third dimension to a topographic database using airborne laser scanner data. *International Archieves of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(3):92–97, 2006. 146

[60] M. Elmqvist. Ground surface estimation from airborne laser scanner data using active shape models. *International Archieves of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34:114–118, 2002. 143

[61] C. Fiorio. A topologically consistent representation for image analysis: the frontiers topological graph. In S. Miguet, A. Montanvert, and Ubeda S., editors, *6th International Conference on Discrete Geometry for Computer Imagery (DGCI'96), Lyon*, number 1176 in Lecture Notes in Computer Sciences, pages 151–162, 1996. 92

[62] M.A. Fischler and R.C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communication of the ACM*, 24(6):381–395, 1981. 144, 156

[63] E.W. Forgy. Cluster analysis of multivariate data: efficiency vs. interpretability of classification. *Biometrics*, 21(3):768–769, 1965. 56

[64] G. Forlani, C. Nardinocchi, C. Scaioni, and P. Zingaretti. Building reconstruction and visualization from LIDAR data. *International Archieves of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34:151–156, 2003. 144

[65] R. Franke and G.M. Nielson. Scattered data interpolation of large sets of scattered data. *International Journal for Numerical Methods in Engineering*, 15:1691–1704, 1980. 158

[66] H. Freeman. On encoding arbitrary geometric configurations. *IRE Transactions on Electronic Computers*, 10:260–268, 1961. 12, 88

[67] H. Freeman. Computer processing of line-drawing images. *ACM Computing surveys*, 6(1):57–97, 1974. 12, 14, 98

[68] V.S. Frost, J.A. Stiles, K.S. Shanmugan, and J.C. Holtzman. A model for radar images and its

application to adaptive filtering of multiplicative noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 4(2):157–166, 1982. 121

[69] K.S. Fu and J.K. Mui. A survey on image segmentation. *Pattern Recognition*, 13(1):3–16, 1981. 54

[70] K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1):32–40, 1975. 57

[71] N.C. Gallagher and G.L. Wise. A theoritical analysis of the properties of median filters. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 29(6):1136–1141, 1981. 119

[72] F. Galland, N. Bertaux, and Ph. Refregier. Minimum description length SAR image segmentation. *IEEE Transactions on Image Processing*, 12(9):995–1006, 2003. 54, 61

[73] S. Geman and D. Geman. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984. 3, 61, 179

[74] E. Gobbetti, F. Marton P. Cignoni, M. Di Benedetto, and F. Ganovelli. Compressed batched dynamic adaptive meshes for terrain rendering. In *Proc. Eurographics 2006*, volume 25(3) of *Computer Graphics Forum*, pages 333–342, 2006. 32

[75] M.F. Goodchild. The fractional Brownian process as a terrain simulation model. In *Proceedings of the Thirteenth Annual Pittsburgh Conference in Modelling and Simulation*, volume 13, pages 1133–1137, 1982. 11

[76] J.W. Goodman. Statistical properties of laser speckle patterns. In In J.C.Dainty, editor, *Laser Speckle and Related Phenomena*, volume 9, pages 9–75. Springer Berlin, Heidelberg, 1975. 173

[77] L. Graham. Synthetic interferometer radar for topographic mapping. In *Proceedings IEEE*, volume 62(6), pages 763–768, 1974. 38

[78] P. Grussenmeyer, M. Koehl, and M. Nour el Din. 3D geometric and semantic modelling in historic sites. In *XVII CIPA International Symposium, October 3-6, 1999, Olinda, Brazil*, 1999. 159

[79] L. Guigues, J.P. Cocquerez, and H. Le Men. Scale-sets image analysis. *International Journal of Computer Vision*, 68(3):289–317, 2006. 78

[80] N. Haala, C. Brenner, and C. Statter. An integrated system for urban model generation. In *ISPRS Commission II Symposium, Cambridge*, 1998. 146

[81] R.M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, 1979. 25, 26

[82] R.M. Haralick, K. Shanmugam, and I. Dinstein. Texture features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6):610–621, 1973. 3, 25

[83] R.M. Haralick and L.G. Shapiro. Image segmentation techniques. *Computer Vision, Graphics and Image Processing*, 29(1):100–132, 1985. 54, 56, 58, 60

[84] R.M. Haralick, S.R. Sternberg, and X. Zhuang. Image analysis using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):532–550, 1987. 16

[85] P.S. Heckbert. Color image quantization for frame buffer display. *ACM Computer Graphics (SIGGRAPH '82 Proceedings)*, 16(3):297–307, 1982. 55

[86] C. Heipke, A. Koch, and P. Lohmann. Analysis of SRTM DTM - methodology and practical results. *Journal of the Swedish Society for Photogrammetry and Remote Sensing*, 2002(1):69–80, 2002. 45

[87] F. Heitz. Traitement du signal bidimensionnel. Cours de l'Ecole Nationale Supérieure de Physique de Strasbourg (ENSPS), France, 2001. 12

[88] S. Hinz, R. Bamler, and U. Stilla. Airborne and spaceborne traffic monitoring. *International Society*

for Photogrammetry and Remote Sensing, 61(3-4):135–136, 2006. 50

[89] S. Hinz and A. Baumgartner. Automatic extraction of urban road networks from multi-view aerial imagery. *International Society for Photogrammetry and Remote Sensing*, 58(1-2):83–98, 2003. 25

[90] H. Hirschmüller. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:328–341, 2007. 36

[91] T.H. Hong and A. Rosenfeld. Compact region extraction using weighted pixel linking in a pyramid. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(2):222–229, 1984. 60

[92] M. Honikel. Strategies and methods for the fusion of digital elevation models from optical and SAR data. In *Proceedings of the Joint ISPRS/EARSeL Workshop, Fusion of Sensor Data, Knowledge Sources and Algorithms for Extraction and Classification of Topographic Objects*, pages 83–89, 1999. 45

[93] S.L. Horowitz and T. Pavlidis. Picture segmentation by a tree traversal algorithm. *Journal Association for Computing Machinery*, 23(2):368–388, 1976. 60

[94] L. Hubert. Approximate evaluation techniques for the single-link and complete-link hierarchical clustering procedures. *Journal of the American Statistical Association*, 69(347):698–704, 1974. 55

[95] M.F. Hutchinson and J.C. Gallant. Representation of terrain. In J.P. Wilson and J.C. Gallant, editors, *Geographical Information Systems: Principles and Technical Issues*, pages 105–124. John Wiley and Sons: New York, NY, 1999. 32

[96] C. Huygens. *Traité de la Lumière*. Springer-Verlag, Berlin-Heidelberg-New York, 256 pages, 1690. 171

[97] J. Illingwork and J. Kittler. A survey of the hough transform. *Computer Vision, Graphics and Image Processing*, 44(1):87–116, 1988. 156

[98] A.K. Jain. *Fundamentals of Digital Image Processing*. Prentice Hall, 592 pages, 1988. 9, 14, 16, 17, 25, 54

[99] A.K. Jain and R.C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall, Inc., Upper Saddle River, NJ, 334 pages, 1988. 24, 25, 26, 55

[100] A.K. Jain, M.N. Murty, and P.J. Flynn. Data clustering: a review. *ACM Computing Surveys*, 31(3):264–323, 1999. 26

[101] E.T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, 2003. 175

[102] H. Jeffreys. *Theory of Probability*. Oxford University Press, Oxford (3d edition 1961), 2005. 175

[103] M.F. Jiang, S.S. Tseng, and C.M. Su. Two-phase clustering process for outliers detection. *Pattern Recognition Letters*, 22(6-7):691–700, 2001. 56

[104] J.A. Lozano J.M Pea and P. Larraaga. An empirical comparision of four initialization methods for the K-Means algorithm. *Pattern Recognition Letters*, 20(10):1027–1040, 1999. 56

[105] T. Joachims. Making large-scale SVM learning practical. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, chapter 11, pages 169–184. MIT Press, Cambridge, MA, 1999. 112

[106] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988. 3, 63

[107] M.G. Kendall. Discrimination and classification. In P.R. Krishnaiah, editor, *Multivariate Analysis*, pages 165–185. Academic Press, Inc., New York, 1966. 55

[108] T.J.M. Kennie and G. Petrie. Introduction to terrain modelling - application fields and terminology. In G. Petrie and T.J.M Kennie, editors, *Terrain Modelling in Surveying and Civil Engineering*, pages 1–3. McGraw-Hill: New York, NY, 1990. 30

[109] E. Khalimsky, R. Kopperman, and P.R. Meyer. Computer graphics and connected topologies on finite ordered sets. *Topology and its Applications*, 36(1):1–17, 1990. 92

[110] S.S. Khan and A. Ahmad. Cluster center initialization algorithm for K-Means clustering. *Pattern Recognition Letters*, 25(11):1293–1302, 2004. 56

[111] J. Kilian, N. Haala, and M. Englich. Capture and evaluation of airborne laser scanner data. *International Archieves of Photogrammetry, Remote Sensing and Spatial Information Sciences, Part B3.*, 31:383–388, 1996. 143

[112] R. Kimmel and A. Bruckstein. On edge detection, edge integration and geometric active contours. In *In Proceedings of Int. Symposium on Mathematical Morphology, ISMM 2002*, 2002. 64, 144

[113] A. Koch. An approach for the semantically correct integration of a digital terrain model and a 2D GIS vector data set. In *ISPRS Proceeding, Commission IV, WG IV/6*, 2004. 146, 154

[114] A. Koch and C. Heipke. Semantically correct 2.5D GIS data- the integration of a DTM and topographic vector data. *International Society for Photogrammetry and Remote Sensing*, 61(1):23–32, 2006. 2, 5, 146, 154

[115] G. Koepfler, C. Lopez, and J.M. Morel. A multiscale algorithm for image segmentation by variational method. *SIAM Journal on Numerical Analysis*, 31(1):282–299, 1994. 3, 67, 68, 70, 73, 81

[116] T.Y. Kong and A. Rosenfeld. Digital topology: Introduction and survey. *Computer Vision, Graphics and Image Processing*, 48(3):357–393, 1989. 4, 12, 88, 94

[117] V.A. Kovalevsky. Finite topology as applied to image analysis. *Computer Vision, Graphics and Image Processing*, 46(2):141–161, 1989. 4, 12, 90, 94, 95, 97

[118] V.A. Kovalevsky. Algorithms and data structures for computer topology. In *Digital and image geometry: advanced lectures*, pages 38–58. Springer-Verlag New York, Inc., 2001. 90

[119] K. Kraus. *Photogrammetrie - Band 1 Geometrische Informationen aus Photographien und Laserscanneraufnahmen*. Walter de Gruyter Verlag, Berlin - New York, 516 p, 2004. 35

[120] K. Kraus and N. Pfeifer. Determination of terrain models in wooded areas with airborne laser scanner data. *International Society for Photogrammetry and Remote Sensing*, 53:193–203, 1998. 143

[121] T. Krauss, P. Reinartz, and M. Lehner. Comparison of DSM generation methods of urban areas from Ikonos images. *Revue Française de Photogrammetrie et de Télédétection*, 2006(184):101–106, 2006. 34

[122] G. Krieger, A. Moreira, H. Fiedler, I. Hajnsek, M. Werner, M. Younis, and M. Zink. TanDEM-X: a satellite formation for high-resolution SAR interferometry. *IEEE Transactions on Geoscience and Remote Sensing*, 45(11):3317–3341, 2008. 39

[123] S. Krishnamachari and R. Chellappa. GMRF models and wavelet decomposition for texture segmentation. In *Proceedings of the IEEE International Conference in Image Processing (ICIP'95)*, volume 3, pages 568–571, 1995. 180

[124] D.T. Kuan, A.A. Sawchuk, T.C. Strand, and P. Chavel. Adaptive restauration of images with speckle. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35(3):373–383, 1987. 120

[125] S.M. LaValle and S.A. Hutchinson. A Bayesian segmentation methodology for parametric image models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):211–217, 1995. 57

[126] Y.G. Leclerc. Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, 3(1):73–102, 1989. 4, 61, 62

[127] J.S. Lee. Digital image enhancement and noise filtering by use of local statistics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(2):165–168, 1981. 120

[128] J.S. Lee. Speckle supression and analysis for synthetic aperture radar images. *Optical Engineering*,

25(5):636–643, 1986. 120

[129] S. Lee, G. Wolberg, and S.Y. Shin. Scattered data interpolation with multilevel B-splines. *IEEE Transactions on Visualization and Computer Graphics*, 3(3):228–244, 1997. 147

[130] M. Lehner, P. D'Angelo, R. Müller, and P. Reinartz. Stereo evaluation of CARTOSAT-1 data summary of DLR results during CARTOSAT-1 scientific assessment program. In *ISPRS Proceeding, Commission I, SS-11*, pages 1295–1300, 2008. 34, 45, 125

[131] M. Lehner and R.S. Gill. Semi-automatic derivation of digital elevation models from stereoscopic 3-line scanner data. In *IAPRS, Vol. 29, Part B4 Washington, USA*, pages 68–75, 1992. 36, 45, 50, 125

[132] Z. Li, Q. Zhu, and C. Gold. *Digital Terrain Modeling: Principles and Methodology.* CRC Press: Boca Raton, FL, 2005. 30, 32

[133] A. Lopes, E. Nezry, R. Touzi, and H. Laur. Structure detection and statistical adaptive speckle filtering in SAR images. *International Journal of Remote Sensing*, 14(9):1735–1758, 1993. 122

[134] A. Lopes, R. Touzi, and E. Nezry. Adaptive speckle filters and scene heterogeneity. *IEEE Transactions on Geoscience and Remote Sensing*, 28(6):992–1000, 1990. 122

[135] D. Lu and Q. Weng. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5):823–870, 2007. 108

[136] H.D. Lüke. *Korrelationssignale. Korrelationsfolgen und Korrelationsarrays in Nachrichten- und Informationstechnik, Meßtechnic und Optik.* Springer Verlag, Berlin, 308 pages, 1992. 38, 171

[137] J.B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability; held at the Statistical Laboratory, University of California*, pages 282–297, 1967. 18, 56

[138] P.C. Mahalanobis. On the generalised distance in statistics. *Proceedings National Institute of Science. India*, 2(1):49–55, 1936. 18

[139] C. Maire. Évaluation de l'apport de donnés optique et radar de très haute résolution pour la cartographie de dégâts de tempête en milieu forestier. *Bulletin XYZ (French Association of Topography journal)*, 93:28–35, 2003. 32

[140] C. Maire and M. Datcu. Earth observation image and DEM information aggregation for realistic 3-D visualization of natural landscapes. *IEEE Transactions on Geoscience and Remote Sensing*, 43(11):2676–2683, 2005. 4, 5, 100, 101

[141] C. Maire and M. Datcu. Object and topology extraction from remote sensing images. In *Proceedings of the IEEE International Conference on Image Processing (ICIP'05)*, volume 2, pages 196–200, 2005. 4, 100, 101

[142] C. Maire, M. Datcu, and P. Audenino. SAR DEM filtering by mean of Bayesian and multi-scale, non-stationary methods. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS'03)*, volume 6, pages 3916–3918, 2003. 5, 124

[143] C. Maire and H. Yésou. Storm forest damage mapping based on very high resolution InSAR data. In *Proceeding Fringe 2003*, 2003. 32, 47, 144

[144] S. Mallat. A theory for multiresolution signal decomposition: The Wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989. 11, 12, 121, 139

[145] B.B. Mandelbrot. *The Fractal Geometry of Nature.* W.H. Freeman and Company, 468 pages, 1982. 11

[146] D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* W.H. Freeman and Company, San Francisco, 397 pages, 1982. 3, 23, 87, 165

[147] D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London. Series*

B, Biological Sciences, 207(1167):187–217, 1980. 23, 24

[148] J.L. Marroquin and F. Girosi. Some extensions of the K-Means algorithm for image segmentation and pattern classification. *Report, AIM-1390 (21 pages)*, 1993. 56, 57

[149] D.F. Maune, S.M. Kopp, C.A. Crawford, and C.E. Zervas. Introduction. In D.F. Maune, editor, *Digital Elevation Model Technologies and Applications: The DEM Users Manual*, pages 1–34. American Society for Photogrammetry and Remote Sensing: Bethesda, MD, 2001. 30

[150] F. Meyer. Colour image segmentation. In *IEEE International Conference on Image Processing and its Applications (ICIP'92)*, pages 303–306, 1992. 54

[151] F. Meyer and S. Beucher. Morphological segmentation. *Journal of Visual Communication and Image Representation*, 1(1):21–46, 1990. 4, 54

[152] C.L. Miller and R.A. Laflamme. The digital terrain model - theory and application. *Photogrammetric Engineering*, 24(3):433–442, 1958. 30

[153] J.M. Morel and S. Solimini. *Variational Methods in Image Segmentation: With Seven Image Processing Experiments*, volume 14 of *Progress in Nonlinear Differential Equations and Their Applications*. Birkhauser, Boston, 225 pages, 1995. 24, 62, 68

[154] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989. 3, 61, 62, 63, 66

[155] P.M. Narendra. A separable median filter for image smoothing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3(1):20–29, 1981. 119

[156] J.R. Ohm and H.D. Lüke. *Signalübertragung. Grundlagen der digitalen und analogen Nachrichtenübertragungssysteme*. Springer Verlag, Berlin, 444 pages, 2007. 171

[157] M. Ortner, X. Descombes, and J. Zerubia. Building outline extraction from digital elevation models using marked point processes. *International Journal of Computer Vision*, 72(2):107–132, 2007. 144

[158] S. Osher and J.A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of Computation Physics*, 79:12–49, 1988. 64

[159] N.R. Pal and S.K. Pal. A review on image segmentation techniques. *Patern Recognition*, 26(9):1277–1294, 1993. 54

[160] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Inc (3d edition 1965), 1991. 177

[161] T.N. Pappas. An adaptive clustering algorithm for image segmentation. *IEEE Transactions on Image Processing*, 40(4):901–914, 1992. 57

[162] T.N. Pappas and N.S. Jayant. An adaptive clustering algorithm for image segmentation. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, (ICASSP'89)*, volume 3, pages 1667–1670, 1989. 57

[163] N. Paragios and R. Deriche. Geodesic active regions: A new paradigm to deal with frame partition problems in computer vision. *Journal of Visual Communication and Image Representation*, 13(1-2):249–268, 2002. 62

[164] N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *International Journal of Computer Vision*, 46(3):223–247, 2002. 62

[165] R. Passini, D. Betzner, and K. Jacobsen. Filtering of digital elevation models. In *ASPRS Annual Convention, Washington*, 2002. 143

[166] T. Pavlidis. *Structural Pattern Recognition*. Springer-Verlag, Berlin-Heidelberg-New York, 302 pages,

1977. 3, 54, 59, 88, 89, 91, 101

[167] T. Pavlidis and Y. Liow. Integrating region growing and edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3):225–233, 1993. 60, 61

[168] G. Peters. Outliers in rough K-Means clustering. In *Pattern Recognition and Machine Intelligence*, pages 702–707, 2005. 56

[169] L. Polidori. *Validation de Modèles Numériques de Terrain. Application à la Cartographie des Risques Géologiques*. PhD thesis, Université de Paris VII, 1991. 32

[170] C. Prati and F. Rocca. Improving slant range resolution of stationary objects with multiple SAR surveys. *IEEE Transactions on AES*, 29(1):135–144, 1993. 40

[171] M. Quaturlli and M. Datcu. Information fusion for scene understanding from interferometric SAR data in urban environments. *IEEE Transactions on Geoscience and Remote Sensing*, 41(9):1976–1985, 2003. 179

[172] S.J. Redmond and C. Heneghan. A method for initialising the K-Means clustering algorithm using Kd-Trees. *Pattern Recognition Letters*, 28(8):965–973, 2007. 56

[173] P. Reinartz, R. Müller, M. Lehner, and M. Schröder. Accuracy analysis for DSM and orthoimages derived from SPOT HRS stereo data using direct georeferencing. *International Society for Photogrammetry and Remote Sensing*, 60(3):160–169, 2006. 35, 37, 45, 125

[174] J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1978. 61

[175] J. Rissanen. Minimum description length principle. In S. Kotz and N. L. Johnson, editors, *Encyclopedia of Statistical Sciences*, volume 5, pages 523–527. John Wiley and Sons, New York, 1987. 61

[176] A. Rosenfeld. Connectivity in digital pictures. *Journal of the ACM*, 17(1):146–160, 1970. 13, 88

[177] A. Rosenfeld. Arcs and curves in digital pictures. *Journal of the Association for Computing Machinery*, 20(1):81–87, 1973. 88

[178] A. Rosenfeld. Adjacency in digital pictures. *Information and Control*, 26(1):24–33, 1974. 13, 88, 89, 91

[179] C. Samson, L. Blanc-Féraud, G. Aubert, and J. Zerubia. A level set model for image classification. *International Journal of Computer Vision*, 40(3):187–197, 2000. 65

[180] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42, 2002. 36

[181] G. Schreier. Geometrical properties of SAR images. In G. Schreier, editor, *SAR Geocoding: Data and Systems*, pages 103–134. Wichmann Verlag, Karlsruhe, 1993. 40

[182] M. Schröder, H. Rehrauer, K. Seidel, and M. Datcu. Interactive learning and probabilistic retrieval in remote sensing images achives. *IEEE Transactions on Geoscience and Remote Sensing*, 38(5):100–119, 2000. 109

[183] S.Z. Selim and M.A. Ismail. K-Means-type algorithms: A generalized convergence theorem and characterization of local optimality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(1):81–86, 1984. 56

[184] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, Inc., 610 pages, 1983. 15, 16

[185] D. Shepard. A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference*, pages 517–524. ACM, 1968. 151, 158

[186] R.N. Shepard and P. Arabie. Additive clustering representations of similarities as combinations of discrete overlapping properties. *Psychological Review*, 86(2):87–123, 1979. 27

[187] J. Shewchuk. Delaunay refinement algorithms for triangular mesh generation. *Computational Geome-*

try: Theory and Applications, 22(1):21–74, 2002. 31

[188] J. Shewchuk. General-dimensional constrained Delaunay and constrained regular triangulations: Combinatorial properties. *Discrete and Computational Geometry*, page to appear, 2008. 31

[189] G. Sithole and G. Vosselman. Experimental comparison of filter algorithms for bare-earth extraction from airborne laser scanning point clouds. *International Society for Photogrammetry and Remote Sensing*, 59(1-2):85–101, 2004. 144

[190] G. Sithole and G. Vosselman. Filtering of airborne laser scanner data based on segmented point clouds. In *ISPRS WG III/3, III/4, V/3 Workshop Laser scanning, Enschede, the Netherlands*, 2005. 143

[191] W. Skarbek and A. Koschan. Colour image segmentation - a survey. *Report, Technical University Berlin*, 1994. 17, 54

[192] T. Smith and J. Guild. The C.I.E. colorimetric standards and their use. *Transactions of the Optical Society*, 33(3):73–134, 1931. 17

[193] M. Spann and R. Wilson. A quad-tree approach to image segmentation which combines statistical and spatial information. *Pattern Recognition*, 18(3-4):257–269, 1985. 60

[194] G.W. Stewart. On the early history of the singular value decomposition. *Report, CS-TR-2855, University of Maryland*, 1992. 21

[195] T. Toutin. Elevation modelling from satellite visible and infrared (VIR) data. *International Journal of Remote Sensing*, 22(5):1097–1125, 2001. 33, 36

[196] T. Toutin. Geometric processing of remote sensing images: models, algorithms and methods. *International Journal of Remote Sensing*, 25(10):1893–1924, 2004. 34

[197] T. Toutin and L. Laurence. State-of-the-art of elevation extraction from satellite SAR data. *International Society for Photogrammetry and Remote Sensing*, 55(1):13–33, 2000. 33, 39, 41

[198] R. Touzi, A. Lopes, J. Bruniquel, and P.W. Vachon. Coherence estimation for SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 37(1):135–149, 1999. 40

[199] B. Tso and P.M. Mather. *Classification Methods for Remotely Sensed Data.* CRC Press, 2001. 108

[200] M. Tur, K.C. Chin, and J.W. Goodman. When is speckle noise multiplicative? *Applied Optics*, 21(7):1157–1159, 1982. 174

[201] F.T. Ulaby and C. Elachi. *Radar Polarimetry for Geoscience Applications.* Artech House Publishers, 388 pages, 1990. 33, 168

[202] F.T. Ulaby, R.K. Moore, and A.K. Fung. *Microwave remote sensing - Active and Passive, Volume I.* Addison-Wesley, 456 pages, 1981. 169

[203] V. Vapnik. *The Nature of Statistical Learning Theory.* Springer-Verlag New York, Inc., 188 pages, 1995. 109, 110

[204] L.A. Vese and T.F. Chan. A multiphase level set framework for image segmentation using the Mumford and Shah model. *International Journal of Computer Vision*, 50(3):271–293, 2002. 62, 65

[205] R.F. Voss. Fractals in nature: from characterization to simulation. In D. Saupe H.O. Peitgen, editor, *The Science of Fractal Images*, pages 21–70. Springer-Verlag New York, Inc., 1988. 138

[206] G. Vosselman. Slope based filtering of laser altimetry data. *International Archieves of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XXXIII, Amsterdam, 2000. 143

[207] M. Walessa. *Bayesian Information Extraction from SAR Images.* PhD thesis, Universität Siegen, Germany, 2001. 123, 124

[208] M. Walessa and M. Datcu. Enhancement of interferometric DEMs by spatially adaptive model-based filtering in non-stationary noise. In *Proc. USAR 2000*, pages 695–698, 2000. 5, 124, 129, 179

[209] M. Walessa and M. Datcu. Model based despeckling and information extraction from SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 38(5):2258–2269, 2000. 123, 124, 179

[210] J.H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963. 55

[211] C. Wiley. *Pulsed Doppler Radar Method and Means*. US Patent No.3.196.436, 1954. 38, 171

[212] Y. Wong and E.C. Posner. A new clustering algorithm applicable to multispectral and polarimetric SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 31(3):634–644, 1993. 55

[213] X. Wu and K. Zhang. A better tree-structured vector quantizer. In *Proceedings Data Compression Conference, Snowbird, Utah*, pages 392–401, 1991. 55

[214] H. Yésou, S. Clandillon, C. Meyer, B. Allenbach, C. Maire, P. de Fraipont, J.C. Favard, and J. Bequignon. Operational mapping of storm forest damage and assessment of future vhr sar and optical sensors. In *Proc. EARSeL Symposium 2003*, pages 457–464, 2003. 32

[215] H.A. Zebker and R. Goldstein. Topographic mapping from interfermetric SAR observations. *Journal of Geophysical Research*, 91(B5):4993–4999, 1986. 38

[216] H.A. Zebker and J.J. Van Zyl. Imaging radar polarimetry: a review. In *Proceedings of the IEEE*, volume 79(11), pages 1583–1606, 2005. 33, 168

[217] K. Zhang, S. Cheng, D. Whitman, M. Shyu, J. Yan, and C. Zhang. A progressive morphological filter for removing non-ground measurements from airborne LIDAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 41(4):872–882, 2003. 143

[218] S.C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing and Bayes / MDL for multiband image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(9):884–900, 1996. 4, 62, 64

[219] S.W. Zucker. Region growing: Childhood and adolescence. *Computer Vision, Graphics and Image Processing*, 5(3):382–399, 1976. 54, 58, 60