Target Detection in Clutter for Sonar Imagery

Nicolas Erwan Valeyrie

Ocean Systems Laboratory

School of Engineering and Physical Sciences

Heriot-Watt University

A thesis submitted for the degree of

Doctor of Philosophy

February 2014

Abstract

This thesis is concerned with the analysis of side-looking sonar images, and specifically with the identification of the types of seabed that are present in such images, and with the detection of man-made objects in such images. Side-looking sonar images are, broadly speaking, the result of the physical interaction between acoustic waves and the bottom of the sea. Because of this interaction, the types of seabed appear as textured areas in side-looking sonar images. The texture descriptors commonly used in the field of sonar imagery fail at accurately identifying the types of seabed because the types of seabed, hence the textures, are extremely variable. In this thesis, we did not use the traditional texture descriptors to identify the types of seabed. We rather used scattering operators which recently appeared in the field of signal and image processing. We assessed how well the types of seabed are identified through two inference algorithms, one based on affine spaces, and the other based on the concept of similarity by composition. This thesis is also concerned with the detection of man-made objects in side-looking sonar images. An object detector may be described as a method which, when applied to a certain number of sonar images, produces a set of detections. Some of these are true positives, and correspond to real objects. Others are false positives, and do not correspond to real objects. The present object detectors suffer from a high false positive rate in complex environments, that is to say, complex types of seabed. The hypothesis we will follow is that it is possible to reduce the number of false positives through a characterisation of the similarity between the detections and the seabed, the false positives being by nature part of the seabed. We will use scattering operators to represent the detections and the same two inference algorithms to quantify how similar the detections are to the seabed.

To my parents, Jean-Françcois and Catherine, and to my sister, Alice.

Acknowledgements

I should like to thank Yvan, my PHD supervisor, for letting me be part of the Ocean Systems Laboratory, about three years and a half ago. This long-lasting experience has broadened my way of looking at signal processing beyond what I could ever have hopped for. Yvan's sheer enthousiasm towards this field of research and Yvan's ideas — always prolific smart and appealing — helped me understand so much about signal processing and come up through this thesis with a hopefully decent piece of work. I should also like to thank Chris and Yan who also work in the Ocean Systems Laboratory. It has been a pleasure to work next to such people of knowledge. Every time I talked with them through the problems I happened to face during the past three years ended up being enlightening a discussion. I should like to specifically thank Yan for going climbing with me, which kept my mind at peace when, at times, things were hectic in the laboratory. I would also like to thank Rolf and Joel, former students and research associates of the Ocean Systems Labratory, who helped me out a few times with pieces of code. My thanks finally go to Rosie and Kevin, two wonderful dear friends of mine, who, when the time had come to write this thesis, were always there in time of need.

Contents

Chapter 1

Statement of the problems

"'Begin at the beginning' the King said gravely, 'and go on till you come to the end: then stop."'

Lewis Carroll, Alice in Wonderland

1.1 Introduction and context

1.1.1 Sonar imagery

Many physical phenomena are propagation phenomena that can be used for a wide variety of purposes. Electromagnetic waves, for example, propagate through the atmosphere, and proved to be a powerful means of exchanging information, or messages, between remote locations. A transmitter transforms a set of messages into electromagnetic waves which propagate through the atmosphere towards a receiver, as illustrated in Figure 1.1-(a). The waves are transformed back into another set of messages at the receiver. The physical coupling of the transmitter and receiver to the propagation medium, that is, the atmosphere, is achieved by the transmitting and receiving apertures [4, 5]. An aperture corresponds to a physical piece of equipment called the antenna. The transmitted and received messages may be somewhat different because of the transmission, propagation, and reception processes, and because of sources of noise along the propagation paths. Acoustic waves are another example of physical propagation phenomena. Acoustic waves propagate through air, water and solid materials, and what propagates is a structured displacement of the propagation medium. Acoustic waves proved to be a powerful means of exploring the surface of the bottom of the sea because they propagate well through fresh and sea water over a wide range frequencies. A sonar imaging system is, generally speaking, a system which makes use of acoustic waves to create images of the bottom of the sea. A transmitter transforms a signal, that is, a pulse of limited duration, into an acoustic wave which propagates through water towards the sea bottom. The wave is reverberated by the bottom of the sea, propagates through water towards a receiver, and

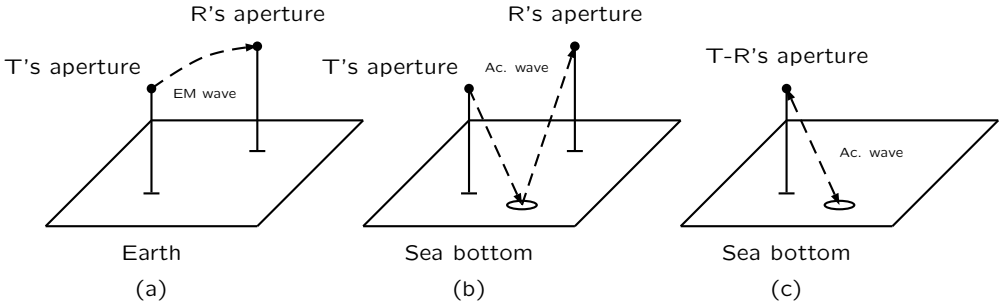


Figure 1.1. Geometry of transmitter (T) and receiver (R) for remote communication systems with electromagnetic (EM) waves and remote imaging systems with acoustic waves (Ac. waves). Left and Middle. Bi-static case. Right. Mono-static case.

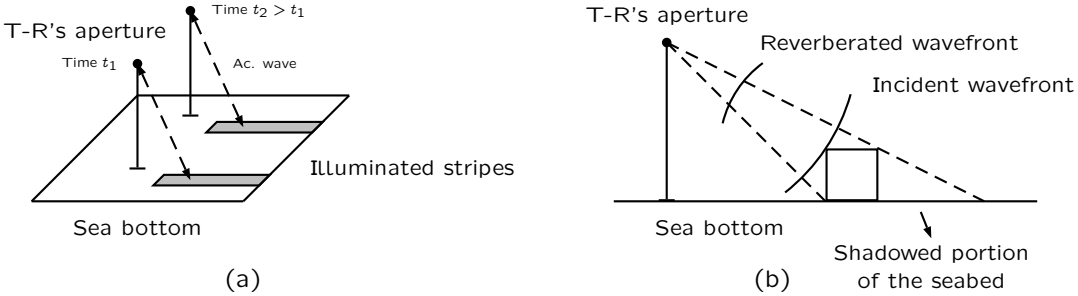


Figure 1.2. Left. Geometry of transmitter (T) and receiver (R) for a side-looking sonar. A narrow stripe of the seabed is illuminated every time a pulse is transmitted. Right. Reverberation of an acoustic wave by an object.

is transformed back into another signal at the receiver, as illustrated in Figures 1.1-(b) and 1.1-(c). The physical coupling of the transmitter and receiver to the propagation medium is, once again, achieved by the transmitting and receiving apertures. The apertures effectively transform time functions, that is, the transmitted and received signals, into space-time phenomena (the acoustic waves), and vice versa [4, 5]. Many sonar imaging systems make use of apertures whose beams exhibit a high directivity in one spatial direction, and are orientated towards the sea bottom. Only a small portion of the bottom of the sea is, consequently, illuminated by the transmitter, a portion of which is viewed by the receiver. An image of the sea bottom may be created when the transmitter and receiver are placed at different locations, so that different portions of surface of the sea bottom are viewed by the imaging system. The imaging setup is mono-static when the transmitter and receiver are at the same place, and bi-static if not, as illustrated in Figures 1.1-(b) and 1.1-(c). Mono-static imaging setups are more common.

A side-looking sonar is a special kind of sonar imaging system. Side-looking sonars often are mounted on underwater vehicles, and can rapidly provide images of wide portions of the sea bottom. Side-looking sonars are mono-static — the

transmitter and receiver are at the same place and correspond to the same physical piece of equipment which is called, as mentioned before, the antenna. The antenna is designed in such a way that only a narrow stripe of the seabed is illuminated every time a pulse (a signal) is transmitted, as illustrated in Figure 1.2-(a). The stripes are perpendicular to the direction of travel of the underwater vehicle. Signals are transmitted at regular time intervals as the vehicle travels through water in a straight path. The received signals are recorded, sampled in time at regular time intervals, and gathered in order to create a digital (discrete) image of the bottom of the sea. A line in the side-looking sonar image of Figure 1.3 corresponds to the recording of one of the received signals. This image was actually obtained with an underwater vehicle equipped with two antennas, one on each side of the vehicle, in order to create an image of what the seabed looks like on both sides of the vehicle. The word "image" is used in a broad sense here because what a side-looking sonar really provides is a two-dimensional representation of the seabed where the two dimensions have the physical dimension of a time. One dimension corresponds to the time between two successive transmissions. The other dimension corresponds to the two-way time of travel of the acoustic waves through water before and after reverberation by the bottom of the sea. A sonar image represents, broadly speaking, the energy of the acoustic waves reverberated by the seabed, and is displayed in grey levels — the higher the grey level, the greater the energy. A representative example of a side-looking sonar image with several types of seabed is presented in Figure 1.3. The types of seabed present in this image are flat areas made of the same type of sediments, areas with sand ripples, and areas with seaweed. A feature common to all side-looking sonar images is the largely black region in the centre of the images. It is called the water column, and corresponds to the two-way time of travel of the acoustic waves through water before and after reverberation by the sea bottom. The vertical white lines, away from the water column, correspond to reverberations by the sea surface, and indicate that this particular image was obtained in a rather shallow water environment. A synthetic aperture sonar is another kind of sonar imaging system. Synthetic aperture sonars provide images of the bottom of the sea with an increased along track resolution with respect to the along track resolution of an image created by a side-looking sonar. The track is the direction of travel of the underwater vehicle. We will not detail the principles of synthetic aperture sonar imagery, and refer to [6] for a review of the current status of synthetic aperture sonar imagery.

Images of the bottom of the sea are of interest for biologists and marine scientists who are interested in being able to identify, in an automatic fashion, the types of seabed that are present in coastal areas. Images of the sea bottom are also of interest for those who wish to detect and identify man-made objects that lay on the bottom of the sea. Man-made objects are distinguished here from natural objects such as rocks and corals. The two problems which will be dealt with in this thesis are the detection (not the identification) of man-made objects (targets) in side-looking sonar images and the identification of the types of seabed in side-looking sonar images. We will further define and discuss those problems

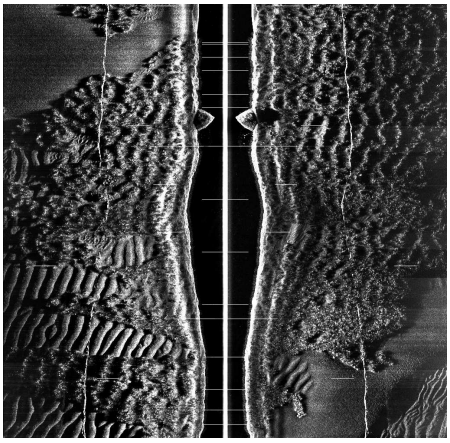


Figure 1.3. Side-looking sonar image with several types of textures arising from significant changes in the topography of the bottom of the sea.

in Section 1.3.1.

1.1.2 Targets and seabed in sonar images

The physical properties of the waves reverberated by the sea bottom depend on the physical properties of the transmitted waves, and on the nature of the sea bottom. Those physical properties (those of the reverberated waves) are passed on to the sonar images. If the transmitted wave is reverberated by a portion of flat seabed, the amount of energy carried by the reverberated wave will mainly depend on the type of sediments the seabed is made of, and on the angle of incidence [7]. Experimental models [8] have been developed over the past years to quantify the amount of energy reverberated by various types of sediments. If the transmitted wave is reverberated by an object (natural or man-made), the reverberated wave will be the result of the interaction, in time and space, between the object and the transmitted wave. The man-made objects (the targets) we consider here are metallic cylinders, truncated cones and wedges, as depicted in Figure 1.4. Being metallic, they all are highly reflective objects. The main visual signature of a man-made object in a sonar image is a highlight region next to a shadow region. The highlight region corresponds to the reflection of the transmitted waves (more than one) on the object. The shadow region corresponds to a lack of acoustic reverberation from the portion of the seabed which is behind the object with respect to the transmitter, as illustrated in Figure 1.2-(b). The shape of the shadow strongly depends on the type of man-made object, and on the angle of view, as illustrated in Figure 1.4. In all cases, the shape of the shadow cast by a man-made object is regular, at least, more regular than the shape of the shadow cast by a natural object, as illustrated in Figure 1.5. A portion of the seabed which is made of a spatial distribution of small rocks, as in Figure 1.5, is called a cluttered area. We will see in Sections 1.2.1 and 1.2.2 that target detection and identification is performed, most of the time, looking for pairs of highlight and shadow regions, and analysing the shape of the shadow.

If the transmitted wave is reverberated by a portion of the sea bottom which is

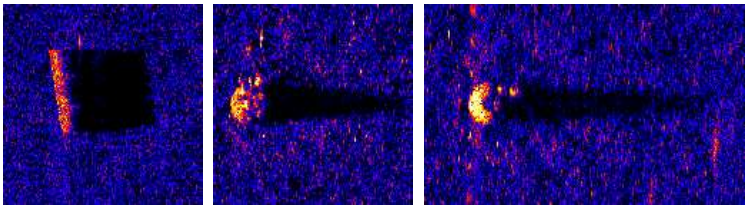


Figure 1.4. Examples of types of targets from a synthetic aperture sonar image. The carrier frequency is 300 kHz, the transducer's beam is 7 degrees wide horizontally and 13 degrees wide vertically. The resolution of the synthetic aperture sonar image is 1.5 centimetres along track (vertically) an 2.5 centimetres across track (horizontally). From left to right, a cylinder, a wedge and a truncated cone. The shape of the shadow differ from target to target.

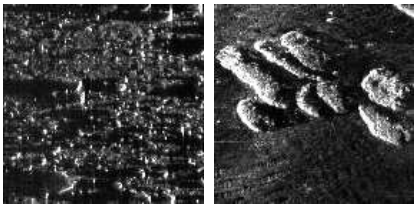


Figure 1.5. Examples of types of sea bottom with rocks from a side-looking sonar image. Left. A cluttered area with small rocks. Right. Big rocks.

significantly not flat, the interaction, in time and space, between the transmitted wave and the changing topography of the bottom around the wave's point of incidence will be critical. For example, should the sea bottom be made of sand ripples, and should the angle of incidence be very low, a noticeable amount of energy will be reverberated by the crests of the ripples whereas no noticeable amount of energy will be reverberated by the troughs of the ripples. Generally speaking, it is this interaction between the waves and the topography of the seabed which makes the seabed appear as textured areas in sonar images. We will see in Section 1.3.2 that texture descriptors which stemmed out of the field of computer vision provide representations well adapted to the seabed. Because one of the problems which will be dealt with in this thesis is the detection of targets in sonar images (not the identification), we will present, in the following section, the useful terminology related to target detection.

1.1.3 Target detection terminology

Target detection aims at deciding whether an area of interest in a sonar image is the result of the reflection of an acoustic wave on a target. A target detector can be generally described as a processing chain, designed to process the sonar image the area of interest comes from in the most advantageous manner, followed by a rule (or set of rules) used to decide whether the area of interest is the result of the reflection of an acoustic wave on a target. The rule can be for example a threshold to be applied to a given value. Four cases can happen:

1. the area of interest is a target and is detected as so,

2. the area of interest is not a target is detected as so,

3. the area of interest is a target and is not detected as a target,

4. the area of interest is not a target and is detected as a target.

A target detector typically processes a collection of sonar images and produces a set of detections, as illustrated in Figure 1.6. The detections can be either true or false positives. The true positives are associated with case (1) while the false positives are associated with case (4). The true positives are therefore the targets (the man-made objects) that were supposed to be detected whereas the false positives are essentially part of the seabed. The performance of a target detector is expressed in terms of the true positive rate and the false positive rate which vary both between 0 and 1. The performance of a target detector is graphically represented by a receiver operating characteristic (ROC) curve which depicts the joint variation of the true positive rate and the false positive rate with the set of rules, as illustrated in Figure 1.7. A conservative set of rules leads to a high positive rate and to a high false positive rate whereas a strict set of rules leads to a comparatively lower false positive rate and to a comparatively lower true positive rate. The objective when designing a target detector is to make the ROC curve as close to the vertical axis as possible in order to obtain, for a given true positive rate, a false positive rate as low as possible. In most cases, the ROC curve does not depict the false positive rate but rather the number of false positives per unit of area (for example, per square kilometre). Target detection and identification is usually performed in two stages. All possible target-like objects are detected first, leading to a high true positive rate at the expense of a high false-positive rate. The detections (the target-like objects) are subsequently identified in the form of a post-processing process, as illustrated in Figure 1.6. The purpose of the identification stage is to make a decision upon the type of the target-like objects, that is to say, to decide whether the detections are cylinders, spheres, wedges, or part of the seabed. In the latter case, the detections are false positives whereas, in the former cases, the detections are true positives. To some extent, the identification stage helps reduce the false positive rate without altering the true positive rate. In the following two sections, we will review the current approaches to target detection and identification with sonar images.

1.2 Current approaches target detection and identification

1.2.1 Target detection

The performance of a target detector strongly depends on the nature of the sea bottom, which can, if need be, be analysed by looking at the types of seabed the false positives belong to. The difficulty in detecting a target (a man-made objet) in a sonar image comes from each type of seabed having a different response to an acoustic wave and a different way of interacting with a target. The purpose

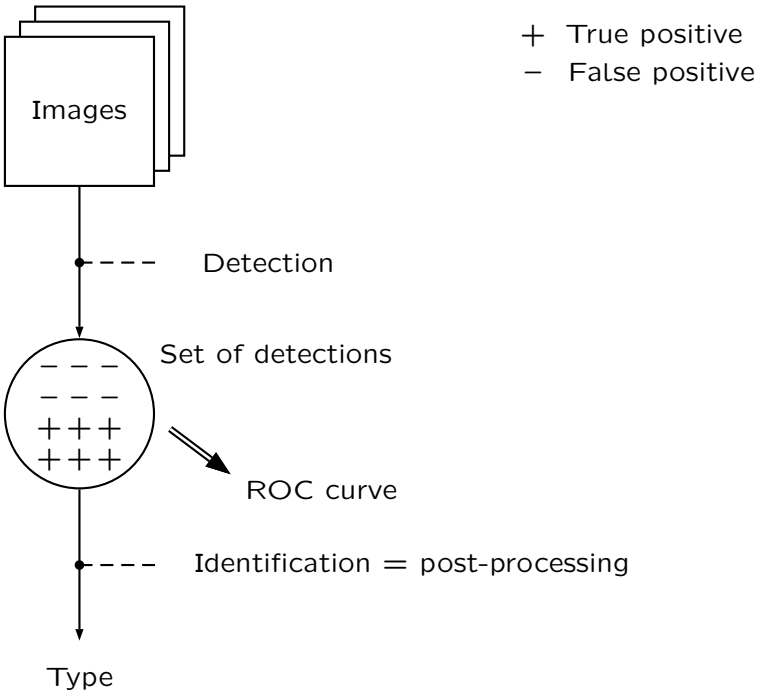


Figure 1.6. Target detection and identification's chain of operations.

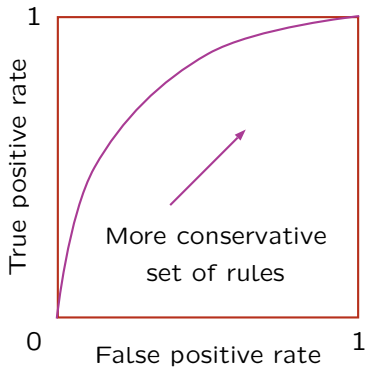


Figure 1.7. Typical receiver operating characteristic (ROC) curve.

of this section is to present the current approaches to target detection with sonar images. Most of these approaches intend to find pairs of highlight and shadow regions because that is the main signature of a highly reflective object present on top of the sea bottom — the shadow region is next to the highlight region and further away from the water column. A matched filter is used in [9–11] to identify areas which closely match the signature of a target. Adaptive thresholding based on local grey level histograms [12–14] and a thresholding with clustering [15, 16] are used to identify areas corresponding to a strong return (the highlight). Those two approaches assume therefore that a target is a lot more reflective than the background and than any natural object. The target detectors we just mentioned are simple and detect all target-like objects. As a result, the number of false positives tends to be high, specifically in complex environments. By complex environments, we mean cluttered areas, and areas with rocks, and textured areas with sand ripples or seaweed. The false positives are removed with further analysis of each detected target-like object using a set of binary classifiers. The classifiers are typically a K-nearest neighbour classifier, a minimum distance classifier, a thresholded classifier or an optimal discrimination filter classifier. High order spectral features in [11] or geometrical features in [9, 10] are used in combination with several binary classifiers, eventually fused to decide whether the target-like object at issue is a target or is not a target. Several simple detection and binary classification schemes are fused in [17, 18] in order to improve the detection performance. Several binary classifiers are fused after a matched filter in [19] to improve the target detection performance as well. The features used in [18] are the object variance and mean, the shadow length and the shadow pixel count. Adaptive thresholding is used in [20, 21] in the mean/standard deviation plane in order to "suppress" the background echoes in a synthetic aperture sonar image. The background of the synthetic aperture sonar images is statistically described by a Weibull distribution. It was shown that the K distribution fit the observations as well as the Weibull distribution does, via a one-sample Kolmogorov-Smirnov test and a χ^2 test. However, the K distribution was not used because of its analytical complexity. More complex detectors use fractals [22], spatial point processes [23] and dual hypothesis theory [24] to detect objects as a local disruption in the texture field. They provide good detection results but heavily depend on large training data sets (they are supervised detectors in this sense), allowing the texture field to be adequately described. Other supervised detectors are based on symbolic pattern analysis [25], principal component analysis [26], and cascades of boosted classifiers [27]. An approach based on active learning in [28] does not require any training set but only the labels (target or clutter) of a few "signatures" in the sonar image of interest (the labels are provided by a diver or an unmanned underwater vehicle). A kernel based algorithm is trained on the labels and eventually used to classify the remaining signatures as a target or not a target. Other detectors are based on mathematical morphology [29], group filters [30], adaptive clutter filtering [18] and canonical correlation analysis [31, 32]. The adaptive clutter filtering is based on an adaptive linear filter exploiting differences between target and clutter correlation, and is, to some extent, related to the

canonical correlation analysis approach. A Bayesian approach is used in [33] in order to incorporate more *a priori* information about the signature of a target than it is done in all the previously mentioned target detectors. The sonar image is segmented via a Markov random field (MRF) where the variables of interest per pixel are the texture class and a binary variable indicating the presence or absence of an object. Only the binary variable is really of interest but the texture labels enable accounting for the background while detecting objects. In [34], target like objects are also detected from the Bayesian segmentation of the sonar images into regions of highlight, shadow and sea bottom reverberation. Markov random fields are efficient when used to incorporate geometrical information about the response of a target and information about the various textured areas. The Bayesian segmentation of images is extremely well coupled with Markov random fields thanks to the Hammersley Clifford theorem, which casts the segmentation problem into an energy minimisation problem [35]. In addition, efficient energy minimisation techniques via graph cuts are presently available [36]. The use of Markov random fields for target detection has significantly reduced the number of false positives but some remain in complex environments [34, 37]. Other approaches to target detection are based on sector scan sonars [38, 39] and high frequency multi beam sonars [40].

1.2.2 Target identification

Target counter measure operations also require the identification of the detected targets, which is mainly done using a set of classifiers. While classification was previously defined as simply deciding whether an target-like object is a target or not a target, geometrical analysis can be used to extract the shape of the object. Since targets are usually regular shaped objects, such as cylinders, truncated cones, spheres or wedges, an target-like object classified as one of these objects is very likely to be a target. A non positive classification as one of these objects leads to the target-like object being identified as not a target. The shape of the shadow is used in [37, 41, 42] for the classification step after extracting the contour of the shadow with a statistical snake. The highlight region is used in [14, 43] but is usually too variable and dependent on the specific sonar conditions to be used as a reliable classification feature. Other approaches use geometrical features [44], correlation analysis [45], an Markov random field [46], a correntropy based matched filtering [47], a voting based scheme [48], wavelet packets and neural networks [49] and canonical correlation analysis [50].

1.2.3 Discussion and perspectives

The purpose of this section is to highlight the limitations of the current approaches to target detection and identification. It appears that target detection and identification are two problems almost entirely solved in simple environments. By simple environments, we mean flat areas, areas with a low clutter density, and areas with small sand ripples — small in the sense that the shadow cast by any

man-made object is readily distinguishable. The way man-made objects appear in sonar images strongly depends on the angle of view, that is to say, on the angle of incidence of the transmitted acoustic pulses which are reverberated by the objects. The dependence is a lot stronger for non-symmetrical objects, such as cylinders and wedges, than for symmetrical objects, such as truncated cones. As a result, the detected objects are, even in simple environments, not always correctly identified with sonar images which only account for one view of the objects. The identification can be improved by fusing multiple views of the same object as in [51] with synthetic aperture sonar images. It is interesting to note that the very same idea can be used to accurately identify the types of seabed. A Bayesian framework is used in [52] to fuse multiple views of the same portion of the sea bottom with synthetic aperture sonar images.

Going back to the limitations of the current approaches to target detection, it appears that those approaches are still challenged by complex environments in the sense that, within a set of detections provided by a target detector, most of the false positives are located in complex environments. By complex environments, we mean areas with vegetation, areas with large sand ripples, and areas with a high clutter density. Most of the false positives are located in complex environments because those environments contain either natural objects which look like man-made objects, or types of seabed which, in places, also look like man-made objects. At this point, it is interesting to realise that most of the current approaches to target detection are strictly speaking image-based, which is why we say that natural objects *look like* man-made objects, or that, in places, the seabed *look like* man-made objects. The acoustic wave reverberated by a natural object or a portion of the seabed would, on the contrary, be different from the one reverberated by a man-made object. There is evidence, in the literature on target detection, that the false positives are mainly located in complex environments. First, let us consider the the detection-orientated Markov random field of [34]. It is said in [34] that "the false [positives] detected all have sizes and signatures comparable to a [target]-like object and are a result of the image containing a lot of object-like clutter." It is also said in [34] that the "false [positives occurred] only in one image where either the seabed or clutter presented a [target]-like signature." The Markov random field of [34] is used in [37] where it is said that "[complex] backgrounds can provide situations in which the [Markov-random-field]-based detection model falsely identify a [target-like object]." Let us consider now the target detector of [27] which is based on a supervised cascade of boosted classifiers and Haar-like features. This detector has false positives mainly located in areas with sand ripples and dense cluttered areas [53].

To some extent, the identification step helps reduce, in the form of a post-processing process, the number of false positives within a set of detections provided by a target detector. What we would like to point out, regarding the identification stage, is that it analyses the detections alone, with respect to models, or templates, and never accounts for the similarity, or lack of similarity, between the detections and the sea bottom. Because a false positive is by nature part of the seabed (the background), one may expect that a false positive is more similar to

the background than a true positive. In this respect, a true positive (the target to be detected) is an anomaly with respect to the sea bottom. There is evidence, in the literature on the classification and fusion of side-looking sonar images, that targets can be looked at as local anomalies with respect to the seabed. By local anomaly, we mean that a target is not similar to the portion of the seabed immediately around it. The results of the work of [54] and [55] are presented in [56] where it is indeed said that "[there] are still miss-classified regions, which have been identified as complex. Many of these regions correspond to objects on the sea floor." It is also said in [56] that there were regions where "the texture field [had] been corrupted by small objects lying on the sea floor." There have also been attempts to detect targets as anomalies with respect to the seabed with the help of fractals or spatial point processes in [22–24]. In those cases, the approach is to model or represent the seabed, using either fractals or spatial point processes, and look for the regions in the sonar images under study that are, with respect to the chosen representation, most different from the background. To some extent, the adaptive clutter filtering of [18] and the canonical correlation analysis of [31] also are attempts to find targets as areas different from the background. At this point, what we would like to point out is that to try and detect targets as anomalies with respect to the seabed does not make any use of the performance already achieved by the currently available target detectors. To be more specific, within a set of detections provided by a target detector, the number of false positives per unit of area (for example, per square kilometre) is extremely low when compared to the area covered by a set of sonar images. It would be more advantageous, at least from a computational perspective, to analyse the similarity between the seabed and the detections provided by a target detection algorithm.

1.3 Aims of the thesis

1.3.1 Statement of the problems

We are now ready to define one of the problems which will be dealt with in this thesis. The problem is that of reducing, in the form of a post-processing process, the number of false positives within a set of detections provided by a target detection algorithm, that being achieved, as illustrated in Figure 1.8, by an analysis of the similarity between the detections and the seabed. This a clutter rejection problem. Clearly, the challenge is to reduce the number of false positives without affecting the number of true positives, which would, if it happened, alter the performance of the target detector alone. The performance of the target detector alone is evaluated with the help of a receiver operating characteristic curve, as explained in Section 1.1.3. The performance of the combined target detector and similarity-based post-processing process will also be evaluated with the help of a receiver operating characteristic curve, as illustrated in Figure 1.8. The detections provided by a target detection algorithm, at least, the ones provided by the algorithm we used in this thesis, correspond to rectangular signals (rectangular

regions) in the sonar images under study, as in Figure 1.9. Being able to analyse the similarity between the detections and the seabed requires therefore being able to analyse the similarity between any two signals extracted from the sonar images under study, or, more generally, being able to analyse the similarity between a signal (the detection) and a collection of signals where the collection of signals represents one or more types of seabed. In this respect, the collection of signals is a *reference* and the signal (the detection) is a *query signal*[1]. We have not yet specified where the collection of signals comes from. The collection of signals may come from a training set, leading to a supervised answer to the clutter rejection problem, or from the the sonar image the query signal itself comes from, leading, this time, to an unsupervised answer to the clutter rejection problem. Possible query signal and reference configurations are given in Figure 1.10. We will not specify where the collection of signals comes from at this point. We have thus far introduced the reference, which represents, through a collection of signals, one or more types of seabed, and a query signal, which is one of the detections provided by the target detection algorithm. Being able to analyse the similarity between the query signal and the reference requires having at our disposal a set of descriptors to represent the signals (the query signal or those in the reference), and an inference algorithm, built upon this representation, which produces a measure of the similarity between the query signal and the reference. The inference algorithm is to be looked at as an anomaly detector because the measured amount of similarity will be used (for example, thresholded) to make a decision of the form "the query signal is or is not similar to the reference." The inference algorithm may not produce a measure of similarity but directly a decision upon whether the query signal is similar to the reference. In the following two sections (Sections 1.3.2 and 1.3.3), we will discuss what can be the set of descriptors (the signal representation) and what can be the inference algorithm. Experimental pieces of work will be introduced in places. They correspond to what was done during the early stages of the thesis and should not be considered as complete pieces of work. We have thus far defined one of the problems which will be dealt with in this thesis, and outlined several intuitive requirements that could be drawn from the problem's definition. We will see in Section 1.3.3 that an inference algorithm used in conjunction with an anomaly detection problem can often be used in conjunction with a classification problem. As a result, should the query signals be regions of interest in a collection of sonar images rather than just the detections provided by a target detector, we may as well consider the problem of identifying, in a supervised fashion, the types of seabed that usually are present in sonar images. We consider the identification as supervised because the query signals are still compared to a reference. To consider this additional problem is all the more indicated should the signal representation (the set of descriptors) be adapted, in one way or another, to the seabed. To summarise, two problems will be dealt with in this thesis, the reduction, in the form of a post-processing process,

[1] We use the terminology of [57] and [58], making use, here, of the words "reference" and "query signal." Actually, the inference algorithm of [57] will be adapted in Chapter 4 to the analysis of side-looking sonar images.

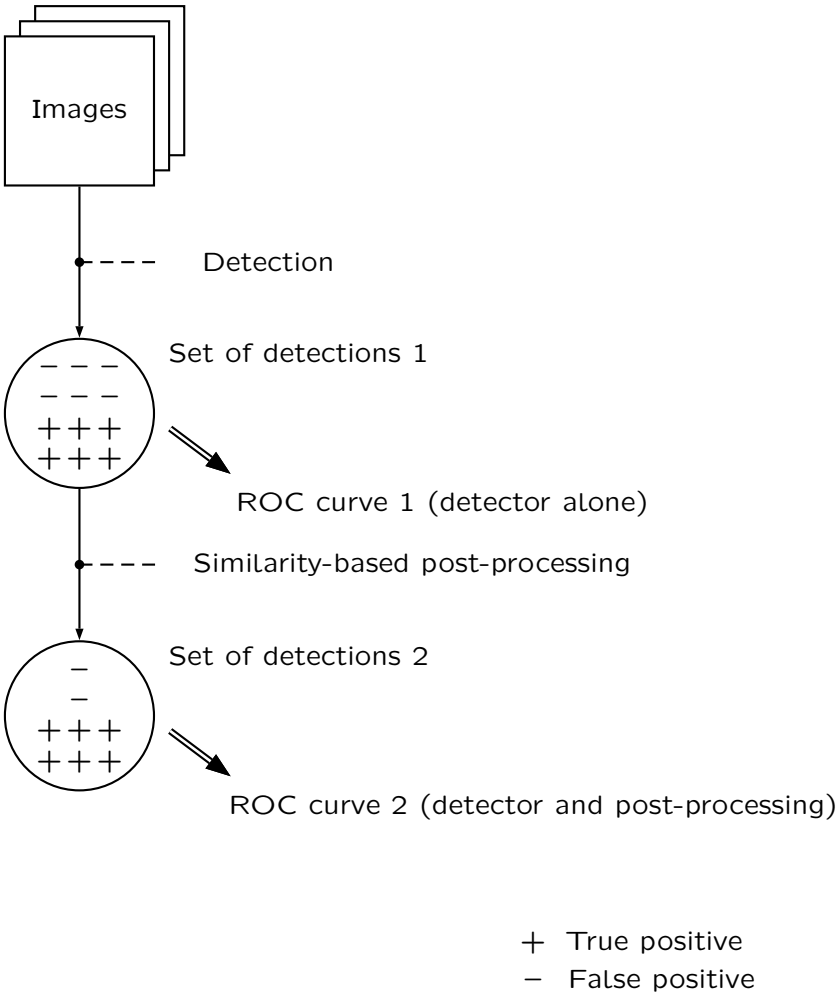


Figure 1.8. Automatic target detection and post-processing chain.

of the number of false positives within a set of detections provided by a target detection algorithm, and the supervised identification of the types of seabed that usually are present in sonar images.

1.3.2 Signal representations

In this section, we will present the possible signal representations. We look at signal representations which are, in one way or another, adapted to the seabed, owing to the fact that a false positive is by nature part of the seabed. Let us start our presentation with simple signal representations. The representation of a signal can be the signal itself, that is to say, the collection of the grey levels of all the pixels in the signal. The representation of a signal can also be the mean and standard deviation of the grey levels of the the pixels which compose the signal. At the end of this section, we will briefly introduce an inference algorithm to illustrate that such simple representations are not well

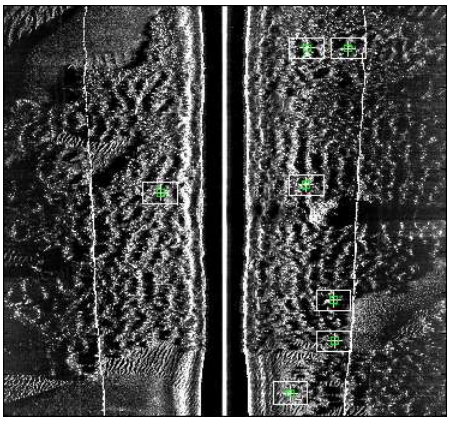


Figure 1.9. Detections provided by the target detector we used in this thesis. The detections are the white rectangular regions. The cross inside a region is the estimated position of the highlight of the detected target-like object.

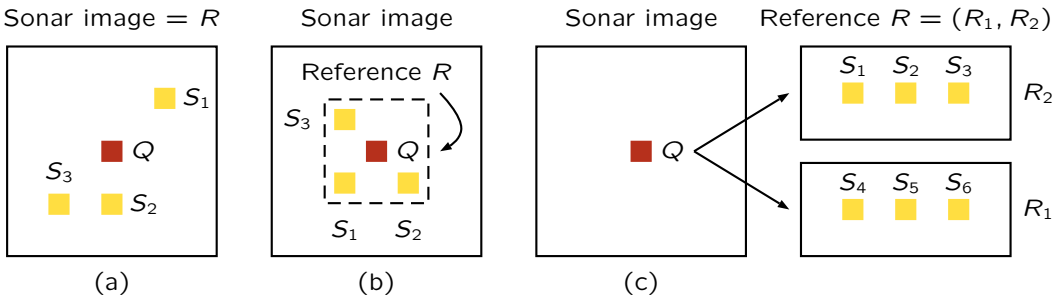


Figure 1.10. Possible query signal and reference configurations. The query signal Q comes from a sonar image and is compared to signals S_i in the reference R. The reference is either (a) the sonar image the query signal comes from, (b) the portion of the sonar image around the query, or (c) an external collection $R = (R_1, R_2)$ of sonar images.

adapted to the seabed[2]. Let us continue our presentation with more advanced signal representations. From an image processing perspective, most types of seabed appear as textured areas in side-looking sonar images. There mainly are two distinct types of textures, the textures which arise from variations in the amount of energy reverberated by flat portions of sediments, and the textures which arise from significant changes in the topography of the bottom of the sea, as illustrated in Figure 1.11. Both types of textures actually arise from the interaction between the transmitting aperture and the topography of the seabed, but this interaction plays a lot greater a role in the generation of the latter type of textures than it does in the generation of the former. This being said, let us review the existing texture representations. Texture representations naturally fall into three paradigms which are the statistical, the structural and the scale-space representations of textures. Sum and difference histograms [59] and co-occurrence

[2]What really is important is the combination of a signal representation and an inference algorithm rather than a signal representation alone, which is why we need an inference algorithm to assess how well any signal representation is adapted to the seabed.

distributions [60, 61] belong to the statistical representations of textures. Textures are considered as random processes in two dimensions and are characterised by the spatial distribution of the grey levels. Seabed identification and segmentation was performed with co-occurrence distributions in [55, 62, 63]. Spatial point processes also lead to statistical representations of textures. The spatial distribution of the pixels corresponding to a certain grey level is modelled by a spatial point process, that is, a stochastic process whose realisations are sets of isolated points in space. A spatial point process was used in [64] to describe the seabed in side-looking sonar images. The structural representations of textures describe complex textures in terms of primitives called texels or textons [65, 66]. The representation of a specific texture requires a vocabulary of textons along with their relationships. Such representations have been mainly used to describe the human perception of textures rather than for automatic texture analysis and identification purposes. Local binary patterns [67] unify the statistical and structural approaches to texture representations. Textures are represented by local spatial patterns and a grey level contrast measure. Extensions to the original local binary pattern operator of [67] now include uniform and rotation invariant binary patterns [68]. Local binary patterns have become of common use in computer vision but were apparently never applied to the analysis of textures in side-looking sonar images. Wavelet transforms of images [69, 70] provide scale-space texture representations that are built upon atoms well localised in space and frequency. Scale-space representations of textures are of interest for the analysis of the types of seabed because the nature of the types of seabed does change with the scale at which they are analysed. Areas with sand ripples, for instance, appear similar to flat areas at fine scales provided that they are made of the same types of sediments. Without entering into details, seabed identification was performed with several types of discrete wavelet transforms in [52, 71, 72]. A fractal is a self-similar mathematical object in the sense that the object as a whole is similar to any part of itself. Fractals are closely related to the wavelet transform [69, 70] for the properties of fractals can be revealed and quantified with the help of a wavelet transform. A parameter of interest is the fractal dimension which can be quantified by means of a continuous wavelet transform. Fractals lead to scale-space representation of textures and were used to describe the seabed in [73].

Traditional texture analysis techniques, such as the wavelet transform and co-occurrence distributions, have difficulty dealing with the significant amount of variability in the appearance of the types of seabed. Scattering operators [74] recently appeared in the field of (one-dimensional) signal and image processing[3]. They provide signal and image representations that are locally translation invariant, stable with respect to deformations, and hence particularly well suited to the identification of textures. Scattering operators iterate over a Littlewood Paley wavelet transform [70] and a complex modulus operator, and have a convolution network type of architecture [75]. Scattering operators were used in [76, 77] to

[3]By one-dimensional signal, we mean the recording in time of a physical quantity such as the voltage drop across a capacitor or a resistor. One-dimensional signals are therefore different from the query signals introduced earlier, the query signals being by nature two-dimensional.

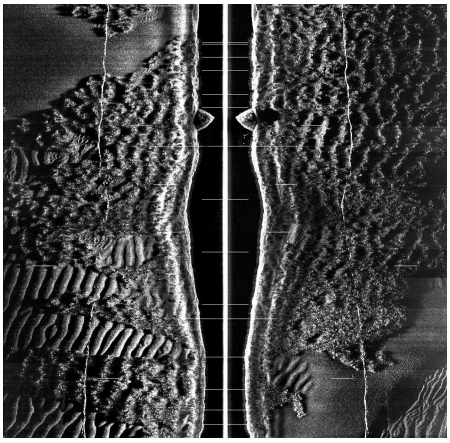


Figure 1.11. Side-looking sonar image with several types of textures arising from significant changes in the topography of the bottom of the sea.

identify textures from the CureT data set, and to identify hand-written digits from the MNIST data set. Scattering operators were also used, in [78] to recognise audio signals and for musical genre classification purposes, and in [79] to represent modulated sounds. It appears therefore that scattering operators are suited not only to the identification of textures but also to the analysis and identification of other kinds of signals and images, which makes scattering operators a very attractive signal representation. Scattering operators were never applied to the identification of the types of seabed in sonar images. We therefore applied scattering operators to identification the types of seabed in a few synthetic aperture sonar images. The identification was strictly speaking distance based — we did not use any inference algorithm such as a neural network or a support vector machine. The results of this piece of work were published in [1] and are depicted through Figures 1.12 to 1.14. We observe that the two types of seabed the sonar images are made of are successfully identified.

The last signal representation we would like to talk about is the one presented in [80]. This signal representation is based on a family of atoms well localised in space and frequency the same way the wavelet transform is based on a family of atoms also well localised in space and frequency. Now, instead of representing a signal with all the atoms in the family, which is what the wavelet transform does, the objective is to represent the signal with a small number of atoms, selected out of the family in an adaptive fashion, in order to obtain a low-level potentially sparse representation of the signal. The matching pursuit algorithm was created in [80] for this purpose. The family of atoms is over-complete, to provide enough freedom and flexibility in the selection of the atoms, and what happens is that the energy of the signal is concentrated in the atoms that are eventually selected by the matching pursuit algorithm[4]. The signal representation induced by the matching pursuit algorithm differs from the signal representations introduced before because it is an adaptive representation. The selected atoms

[4]A proper definition of what is an over-complete family of atoms is given in [70].

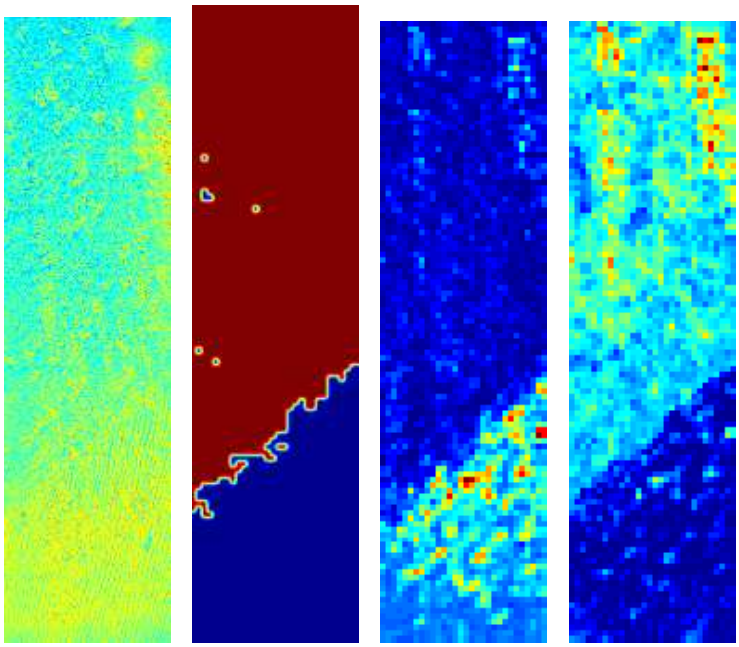


Figure 1.12. Seabed identification in synthetic aperture sonar images. Left. Sonar image. Middle-left. Identification result. Middle-right and right. Distance to each type of seabed.

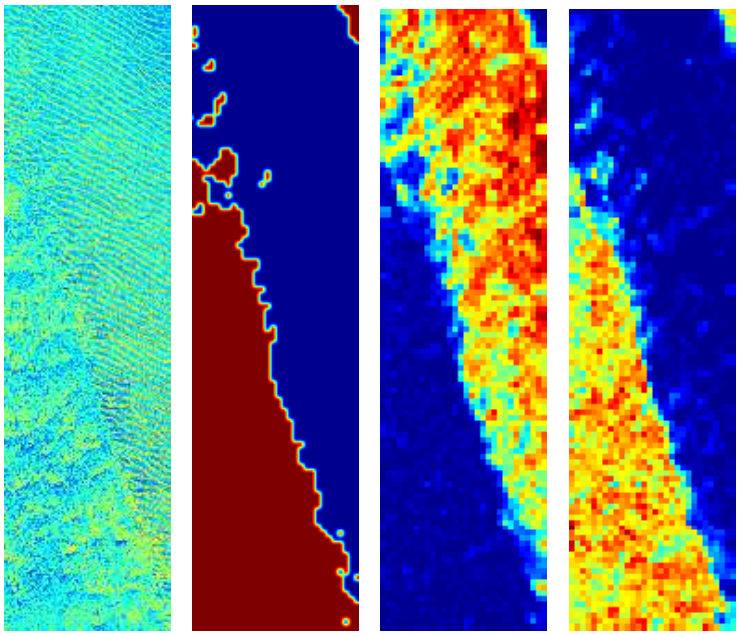


Figure 1.13. Seabed identification in synthetic aperture sonar images. Left. Sonar image. Middle-left. Identification result. Middle-right and right. Distance to each type of seabed.

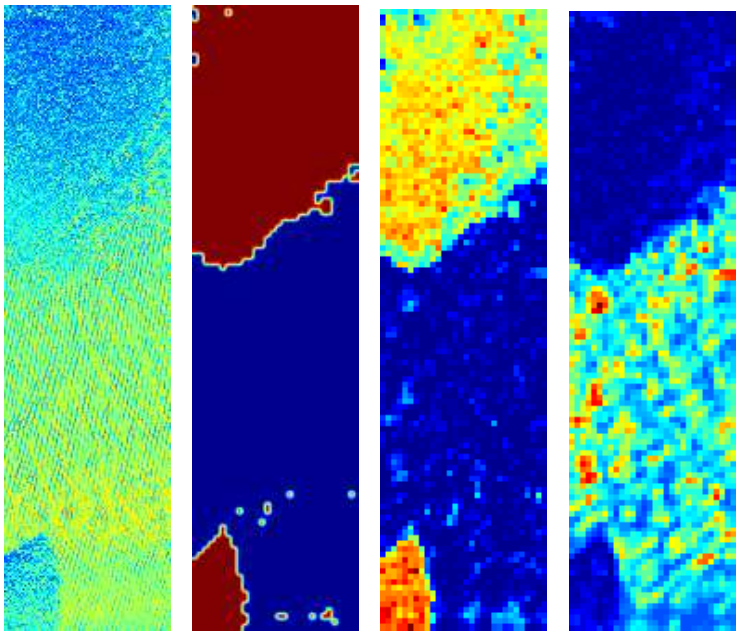


Figure 1.14. Seabed identification in synthetic aperture sonar images. Left. Sonar image. Middle-left. Identification result. Middle-right and right. Distance to each type of seabed.

would therefore differ from query signal to query signal. The signal representation induced by the matching pursuit algorithm is in some ways similar to the sparse representations of [81] and [82].

Let us now introduce an inference algorithm based on the concept of ensemble of patches, as it appears in [83], in order to illustrate that not all signal representations are suited to the analysis of the seabed. The inference algorithm of [83] was used in conjunction with SIFT descriptors [84] to shorten video sequences in time, and to reduce camera images in size. We consider here two signal representations, the one induced by scattering operators, and the one whereby the representation of a signal is the signal itself. The representation induced by scattering operators will be presented in details in Chapter 2. We consider a query signal Q which is, practically, a rectangular region taken around one pixel of the sonar image under study. The query signal is broken into many overlapping patches of small size, as illustrated in Figure 1.15. We denote by N_Q the total number of patches in Q. We do not keep knowledge of the relative arrangement of the patches within the query signal. We consider a reference $R = (R_1, R_2, R_3, \ldots)$ which is a collection of examples of types of seabed. For instance, R_1 can be the example of an area with seaweed and R_2 can be the example of an area with sand ripples, as in Figure 1.16. The R_i's are also broken into many small overlapping patches. As in [83], we say that the query signal Q is similar to the reference R if as many patches of Q are contained in R, which is a simple and yet powerful way of evaluating the similarity between Q and R. The dissimilarity measure between Q and R is

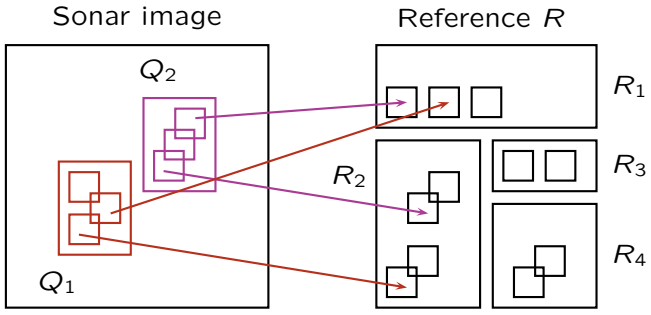


Figure 1.15. Left. Sonar image with two query signals Q_1 and Q_2 both broken into small overlapping patches. Right. Reference $R = (R_1, R_2, R_3, R_4)$ which is a collection of four examples of types of seabed. The examples need not be of the same size. The R_i's are also broken into small overlapping patches.

accordingly defined by

$$\mathcal{D}(Q, R) = \frac{1}{N_Q} \sum_{q \in Q} \min_{r \in R} \| Dq - Dr \| . \tag{1.1}$$

For each patch q in Q, we look for the patch r in R which as similar to q as possible. The similarity between q and r is evaluated in one of the aforementioned two signal representations via Dq and Dr — Dq and Dr represent the set of descriptors associated with q and r[5]. What is to be expected is that the less similar the query signal Q is to the reference R, the greater the dissimilarity $\mathcal{D}(Q, R)$. Let us now apply the inference algorithm to a few side-looking sonar images. The results of this piece of work were published in [2]. The reference is the collection of examples of types of seabed of Figure 1.16. Query signals Q are extracted from two sonar images, leading to two maps of $\mathcal{D}(Q, R)$ for all possible query signals Q. The maps obtained with the signal representation induced by scattering operators are depicted in Figure 1.17. The maps obtained with the signal representation whereby the representation of a signal is the signal itself are depicted in Figure 1.18. We observe that the objects which lay on the bottom of the sea appear as salient regions with respect to the seabed in Figure 1.17 but not in Figure 1.18. We may also acknowledge the potential of describing query signals in terms of ensemble of patches because, provided that the signal representation is well chosen, such a description allows a fine local analysis of the query signals. In this respect, scattering operators appear to be well chosen representation.

We have thus far presented possible signal representations, and emphasised on texture representations because they are particularly well suited to the analysis of the seabed. We also showed, with the help of two experiments, that scattering operators appear to be a texture representation well suited to the analysis of the seabed. In the following section, we will present possible inference algorithms.

[5]We ought to speak of a patch representation instead of a signal representation but, because patches and signals are rectangular regions in sonar images, a signal representation can be used as a patch representation and vice versa.

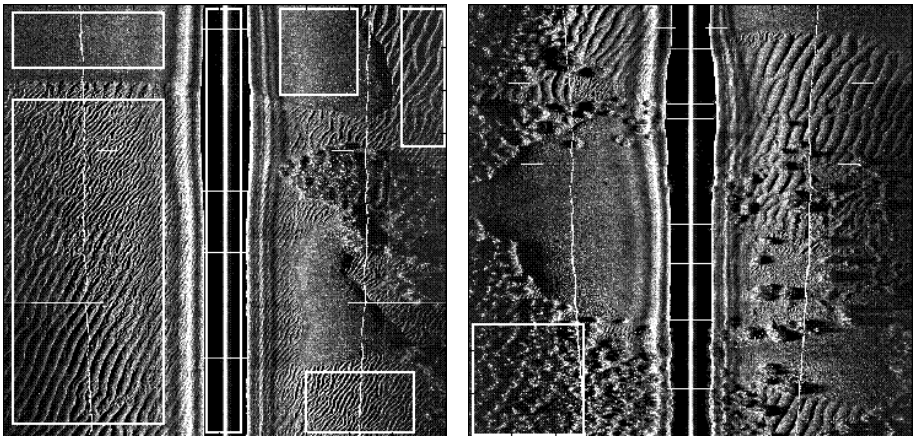


Figure 1.16. Reference $R = (R_1, R_2, \ldots, R_6)$ which is a collection of examples R_i of types of seabed. The types of seabed are areas with sand ripples, areas with seabed and flat areas. The R_i's correspond to the white rectangular regions.

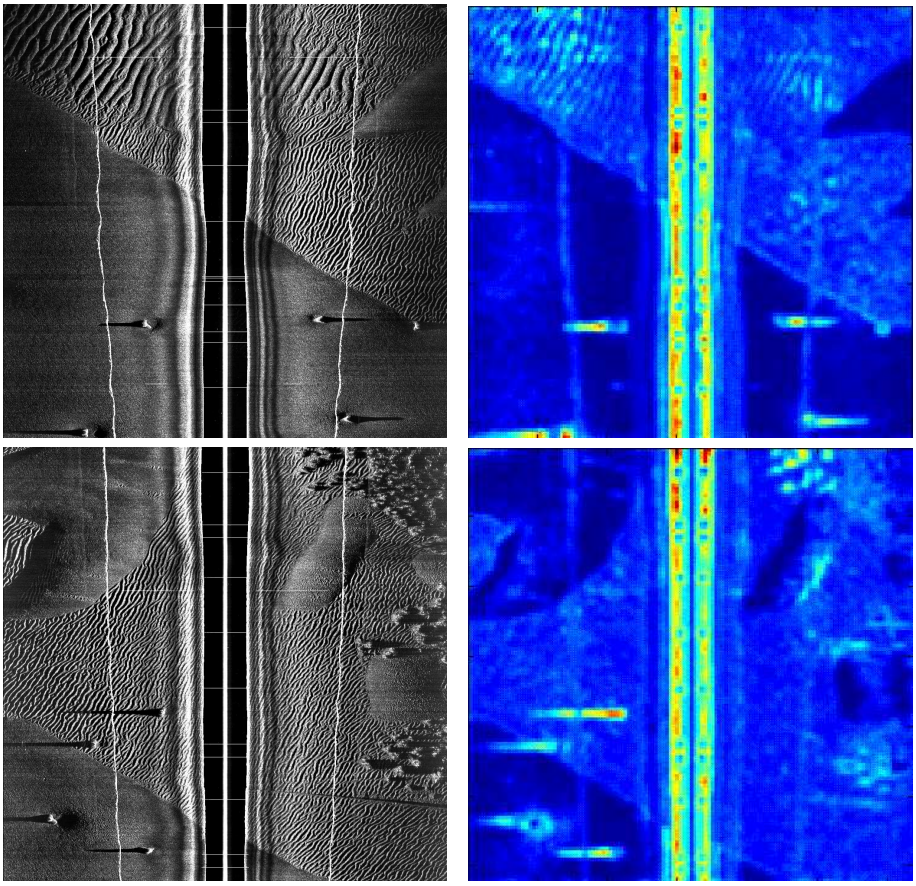


Figure 1.17. Two sonar images on the left hand side and the corresponding saliency maps on the right hand side. The objects present in the two sonar images appear as salient regions.

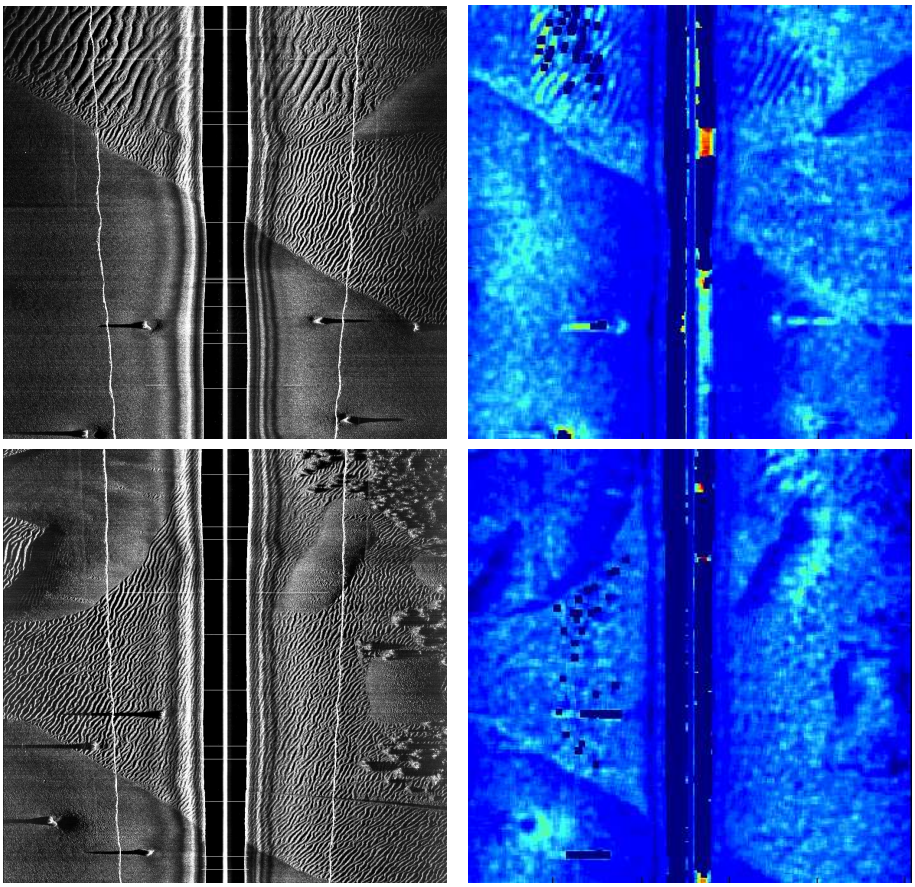


Figure 1.18. Two sonar images on the left hand side and the corresponding saliency maps on the right hand side. The objects present in the two sonar images do not appear as salient regions.

1.3.3 Inference algorithms

In this section, we will present possible inference algorithms which may be used in conjunction with an anomaly detection problem. Most of the time, but not always, will it be possible to use those inference algorithms in conjunction with a classification problem. We look for inference algorithms to analyse the similarity between a query signal and a reference where, as in Section 1.3.1 and as in Figures 1.10 and 1.15, the query signal is either a detection or a region of interest, and where the reference corresponds to a collection of signals which belong to one or more types of seabed. We will assume throughout this section that a signal representation has been defined so that, when needed, we can speak of the space generated by the representation of the signals. Because we will ultimately work with discrete sonar images, as opposed to their continuous counterparts, we assume that this space's dimension is finite. This is not too restrictive an assumption and, besides, this assumption holds for all the signal representations presented in Section 1.3.2.

This being said, let us start our presentation with the inference algorithms derived from the estimation of probability density functions as in [85] and [86]. We consider the signals in the reference and we look at the set of descriptors of those signals as realisations of a multi-dimensional random variable. The probability density function of this random variable is estimated from examples of signals in the reference, hence in a supervised fashion. The density function is then used to decide whether the query signal is similar to the reference. We have thus far looked at the set of descriptors of the signals in the reference as realisations of a random variable. Should we look at the set of descriptors of those signals as points in an affine space, we obtain the inference algorithms derived from the estimation of boundaries around a given set of points. Boundaries may be estimated with the K-centres method [87], the nearest neighbour method [88], the support vector data description method [89] or a one-class support vector machine [90]. The distance between the query signal and the boundaries is used to ultimately decide whether the query signal is similar to the reference. Let us continue our presentation with the inference algorithms derived from the field of data modelling where what is modelled is the distribution, in the space generated by the signal representation, of the set of descriptors of the signals in the reference. A decision upon the similarity between the query signal and the reference is made from an analysis of the distance between the query signal and the model, where the word "distance" is used in a broad sense. Models may arise from the Karhunen-Loève transform [70], also known as the principal component analysis, a K-means clustering and auto-encoders [91]. Auto-encoders are also known as diabolo networks, and are related to neural networks. The inference algorithms we presented up to this point are to be used in conjunction with an anomaly detection problem. All of them can be used, with a few modifications, in conjunction with a classification problem apart from the support vector data description and the one-class support vector machine.

Let us finish our presentation with the inference algorithms derived from the

concept of similarity by composition as in [57]. Those inference algorithms intend to evaluate to what extent the query signal can be composed from pieces taken from the reference, where the pieces are patches of small size with respect to the size of the query signal, as in Figure 1.15. The composition process is similar in mind to that of a puzzle. The complexity of the composition process leads to a measure of the similarity between the query signal and individual signals in the reference. Among those signals, a signal of particular interest is, without entering into details, the one obtained from a maximum a posteriori estimation. In [57], the amount of similarity between the query signal and the reference is defined as the amount of similarity between the query signal and the signal in the reference obtained from a maximum a posteriori estimation. That proved to be well suited to the detections of irregularities in camera images and video sequences when used in conjunction with SIFT descriptors. The work of [57] is extended in [58] where a more advanced measure of the similarity between the query signal and the reference is derived. The work of [58] was used not only to reveal saliency regions in camera images and video sequences but also to classify and cluster camera images, in conjunction with SIFT descriptors, and for speech recognition purposes, in conjunction with the mel-frequency cepstrum descriptors. The inference algorithms of [57] and [58] take both the form of a statistical composition process, which is why we spoke earlier on of a maximum a posteriori estimation. The inference algorithms of [57] and [58] resemble the one presented in Section 1.3.2 because all three consider the query signals as ensemble of patches. In this respect, the size of a patch corresponds to the resolution, above that of a pixel, at which the inference algorithms work. The inference algorithms of [57] and [58] are, however, more advanced than the one of Section 1.3.2 because they account for the spatial arrangement of the patches. From the perspective of detecting targets as anomalies with respect to the seabed, this is an important concept. It is indeed the spatial arrangement of a given number of pixels, or group of pixels, into a highlight and a shadow region which makes a target appear as a target in an image. The pixels — to be more precise, their grey levels — which form the highlight and the shadow regions can be found elsewhere in the image.

1.3.4 Discussion and layout of the thesis

We have thus far presented possible signal representations and inference algorithms to be used in conjunction with either an anomaly detection problem or a classification problem. In this section, we intend to choose one or more signal representations and one or more inference algorithms to answer the two problems defined in Section 1.3.1. As a reminder, those problems are the reduction, in the form of a post-processing process, of the number of false positives in a set of detections provided by a target detector, and the supervised identification of the types of seabed. In the discussion to come, we intend to bear in mind that a signal representation and an inference algorithm should not be considered independently.

Among the inference algorithms presented in the preceding section (Section 1.3.3), the ones based on the concept of similarity by composition naturally stand

out because only those algorithms consider the query signals as geometrical ensembles of patches. Those inference algorithms were used with SIFT descriptors in [57] and [58] to identify salient regions in video sequences and camera images. SIFT descriptors are based on corners and edges appearing in camera images and video frames at different scales. Such descriptors cannot be used with sonar images because edges and corners do not exist in such images. One of the experiments of Section 1.3.2 showed, however, that objects (not necessarily man-made objects) present on the bottom of the sea can be revealed as anomalies with respect to the seabed with the help of the signal representation induced by scattering operators, and an inference algorithm based on the concept of ensemble of patches. This experiment suggests therefore that SIFT descriptors can be replaced by scattering operators in the inference algorithm of [57]. We intend to follow this idea and use the inference algorithm of [57] with scattering operators to try and reduce the number of false positives in a set of detections provided by a target detector (first problem). Because of the flexibility in the definition of the query signals and the reference, we will be able to use the same inference algorithm to identify the types of seabed (second problem). We have not yet spoken of the inference algorithms which are not based on the concept of similarity by composition. The algorithm we are interested in is the one based on the Karhunen-Loève decomposition which can be used to answer an anomaly detection problem or a classification problem. The Karhunen-Loève transform leads to models based on affine spaces. The reason why we are interested in this algorithm is that it was used with scattering operators, in [76, 77] to identify textures and hand-written digits, and in [78] to recognise audio signals. We intend to continue down this line as well and use affine spaces and scattering operators to answer the two problems this thesis is concerned with.

We are now ready to detail the structure of the chapters to come. Chapter 2 will present the signal representation induced by scattering operators, Chapter 3 with present how to analyse side-looking sonar images with affine spaces, and Chapter 4 with present how to analyse side-looking sonar images with a statistical composition process. An answer to the two problems will be given in Chapters 3 and 4 with respectively affine spaces and the statistical composition process.

Chapter 2

Scattering operators for image analysis

"Nothing exists except atoms and empty space. Everything else is opinion."

Democritus

2.1 Introduction

In this chapter, we will present the image representation induced by scattering operators. From the scattering representation of an image, we will be able to obtain the scattering representation of any rectangular region extracted from this image. These regions were called query signals in the preceding chapter. How to obtain the scattering representation of the query signals will be talked through in the following chapter. In this chapter, we will only be interested in the scattering representation of an image. Several problems encountered in the field of image processing require being able to measure or quantify the similarity between images. The underlying notion of similarity depends of course on the problem at stake, and what is then needed, is an image representation which captures the correct notion of similarity. The wavelet transform of images is one possible image representation. This representation is well suited to the image encoding and compression problems but not so much to the recognition of textures or patterns. A texture may indeed be locally deformed and still remain the same texture. A pattern, such as a hand-written digit or a hand-written character, may similarly be locally deformed and still remain the same pattern. Textures and patterns may also undergo rigid transformations such as translations and rotations. The wavelet transform of images leads to an image representation which is not invariant with respect to translations and rotations but rather to a representation which commutes with translations and rotations. The wavelet transform of images leads to an image representation which is not stable with respect to local

deformations because high spatial frequencies are not stable with respect to local deformations. Low spatial frequencies are not, on the contrary, affected as much by local deformations as are high spatial frequencies. What makes two textures or two patterns different is however their respective high frequency content. To be able to recognise textures and patterns, the image representation induced by a wavelet transform traditionally undergoes a smoothing operation. This operation may be achieved with the help of a spatial low-pass filter, the calculation of the histogram of the representation's coefficients, or the estimation of first and second order probability distributions. Such a smoothing operation implies a loss of information which is never recovered from. Scattering operators recently appeared in the field of image processing. Scattering operators iterate over a Littlewood Paley wavelet transform and a modulus operator in order to create an image representation which is stable with respect to local deformations, and locally translation invariant. The Littlewood Paley wavelet transform captures the frequency content of an image in various frequency bands. The effect of the modulus operators is to push the high frequency content of the image towards low frequencies, making eventually the scattering representation stable with respect to local deformations. Because scattering operators iterate over a Littlewood Paley wavelet transform and a modulus operator, we may consider the scattering representation's coefficients obtained after one iteration, those obtained after two iterations, and more generally, those obtained after any number of iterations. The scattering representation's coefficients obtained after one iteration are essentially the result of a smoothing the wavelet transform's coefficients by a low-pass filter. The low-pass filter is the scaling function which appears in the Littlewood Paley wavelet transform. This smoothing operation induces a loss of information which is recovered from by the subsequent iterations. The lost information is indeed contained in the scattering representation's coefficients obtained after more than one iteration. The number of iterations that are needed can be determined by looking at how the image's energy is spread among the scattering representation's coefficients. This is similar to the frame type of condition which ensures that the image representation provided by the Littlewood Paley wavelet transform is complete. The scattering representation has a convolution network type of architecture, a convolution network being a special kind of neural network. Convolution networks iterate over linear and non-linear image transformations, and have been used, in the past, for texture and pattern recognition. The number of iterations that are needed in a convolution network is learnt during a training stage. The number of iterations that are needed in the scattering transform are, on the contrary, specified by the images to be analysed.

This chapter will be organised in the following fashion. We will first present the Littlewood Paley wavelet transform of images. We will emphasise the algorithmic and numerical aspects of the wavelet transform because these aspects have to be considered with great care when working with discrete images, as opposed to their continuous counterparts. We will specifically show how to choose the parameters of the wavelet transform when working with discrete images. We will subsequently present the image representation induced by scattering operators which is based

upon the Littlewood Paley wavelet transform. We will emphasise some of the algorithmic and numerical aspects of the scattering transform because theses aspects also have to be considered with great care when working with discrete images. We will specifically show how to implement the scattering transform of an image with a computer program written in C.

2.2 Wavelets and space-frequency atoms

2.2.1 Introduction

The wavelet transform of images represents images in terms of atoms that are obtained by dilations and rotations of a unique function $\psi(x)$ which is called a *wavelet* and which is well localised in space and frequency [69]. The atoms are the functions

$$\psi_{k,\gamma}(x) = 2^{-2k}\,\psi\big(2^{-k}\,r_\gamma^{-1}(x)\big) \tag{2.1}$$

also well localised in space and frequency. The atoms are obtained from the wavelet by dilations proportional to 2 which is a value commonly chosen for the wavelet transform of images. The dilation progression is said to be dyadic. It is interesting to bear in mind that the analysis of one-dimensional signals, such as the recording of a voltage drop across a capacitor in an electrical circuit, usually require a fine frequency resolution and hence dilations proportional to a with $a \ll 2$. In (2.1) r_γ is the rotation of angle $\gamma \in G$, namely

$$r_\gamma = \begin{bmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{bmatrix}\,, \tag{2.2}$$

and k is a positive or negative integer. Here $G \subset [\,0, 2\pi\,)$ is the finite set of angles defined as follows.

$$G = G_+ \cup G_- \quad \text{with} \quad \begin{cases} G_+ = \big\{\, m\pi/M \,\big\}_{0\le m<M} \\ G_- = \big\{\, \pi + m\pi/M \,\big\}_{0\le m<M} \end{cases} \tag{2.3}$$

where M is a strictly positive natural number. There are $2M$ angles in G, M angles in G_+ and M angles in G_-. G_+ correspond to positive rotations of angles in $[\,0, \pi\,)$ while G_- corresponds to negative rotations of angles in $[\,\pi, 2\pi\,)$. We will see in section 2.3.1 that only positive rotations are needed when analysing real-valued images while both positive and negative rotations are needed when analysing complex-valued images.

2.2.2 Qualitative space frequency localisation

Should the support of the wavelet $\psi(x)$ be centred around x_ψ in space, the support of the atom $\psi_{k,\gamma}(x)$ is centred around $r_\gamma(2^k x_\psi)$ in space. In addition, the extent of the support of $\psi_{k,\gamma}(x)$ is proportional to 2^k. For positive k's, the support of $\psi_{k,\gamma}(x)$ is 2^k times larger than that of $\psi(x)$. For negative k's, the support of

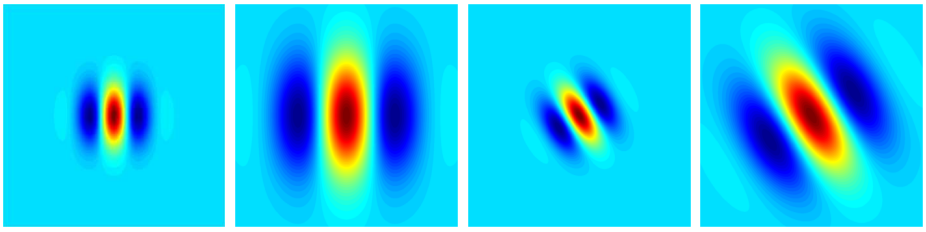


Figure 2.1. Real part of the atoms $\psi_{k,\gamma}(x)$ obtained from the wavelet of Morlet. Left. $\gamma = 0$. $k = 0$ and $k = 1$. Right. $\gamma = \pi/6$. $k = 0$ and $k = 1$. The atoms are depicted for $x \in [-5, 5] \times [-5, 5]$.

$\psi_{k,\gamma}(x)$ is 2^k times smaller than that of $\psi(x)$. The real and imaginary parts of the atoms $\psi_{k,\gamma}(x)$ obtained from the wavelet of Morlet are depicted in Figures 2.1 and 2.2 for $k = 0$ and $k = 1$ and for $\gamma = 0$ and $\gamma = \pi/6$. The atoms are depicted over squares of size 10 centred around the origin $x = 0$. The units associated with the squares are those associated the spatial variable x. Should the image under study be a side-looking sonar image, the units associated with x are seconds. We just discussed the localisation in space of the atoms. We now discuss their localisation in frequency. The Fourier transform $\hat{\psi}(\omega)$ of $\psi(x)$ is defined by [69]

$$\hat{\psi}(\omega) = \int_{\mathbb{R}^2} \psi(x)\, e^{-i\omega x} dx \tag{2.4}$$

with the reconstruction formula

$$\psi(x) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \hat{\psi}(\omega)\, e^{i\omega x} d\omega \ . \tag{2.5}$$

In (2.4) and (2.5) ω is strictly speaking a spatial wave number related to the frequency ν according to $\omega = 2\pi\nu$. However, we will equivalently refer to ω as a wave number or a frequency in this thesis. Upon combining (2.1) and (2.4) we find the expression of the Fourier transform of the atoms, namely

$$\hat{\psi}_{k,\gamma}(\omega) = \hat{\psi}(\, 2^k\, r_\gamma^{-1}(\omega)\,) \ . \tag{2.6}$$

Should the support of $\hat{\psi}(\omega)$ be centred around a frequency ω_ψ in frequency, the support of $\hat{\psi}_{k,\gamma}(\omega)$ is centred around $r_\gamma(\omega_\psi/2^k)$. In addition, the extent of the support of $\hat{\psi}_{k,\gamma}(\omega)$ is proportional to 2^{-k}. For positive k's, the support of $\hat{\psi}_{k,\gamma}(\omega)$ is 2^k times smaller than that of $\hat{\psi}(\omega)$. For negative k's, the support of $\hat{\psi}_{k,\gamma}(\omega)$ is 2^k times larger than that of $\hat{\psi}(\omega)$. The Fourier transforms of the atoms obtained from the wavelet of Morlet are depicted in Figures 2.3 and 2.4 for $k = 0$ and $k = 1$ and for $\gamma \in G_+$. The Fourier transforms of the atoms are depicted in Figures 2.3 and 2.4 for $\omega \in [-\pi, \pi] \times [-\pi, \pi]$.

2.2.3 Scaling function

We know from the preceding section that the localisation in frequency of the atoms changes with the scale parameter k and the angle γ. The atoms are therefore

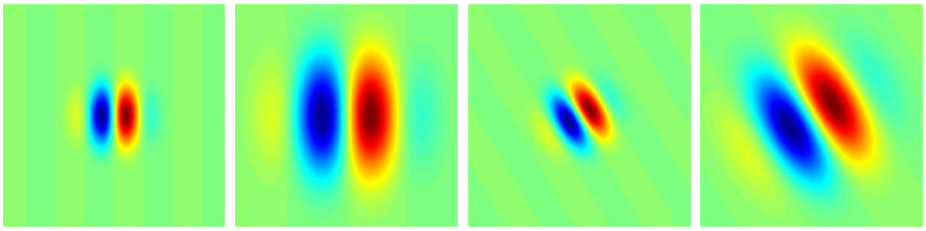


Figure 2.2. Imaginary part of the atoms $\psi_{k,\gamma}(x)$ obtained from the wavelet of Morlet. Left. $\gamma = 0$. $k = 0$ and $k = 1$. Right. $\gamma = \pi/6$. $k = 0$ and $k = 1$. The atoms are depicted for $x \in [-5, 5] \times [-5, 5]$.

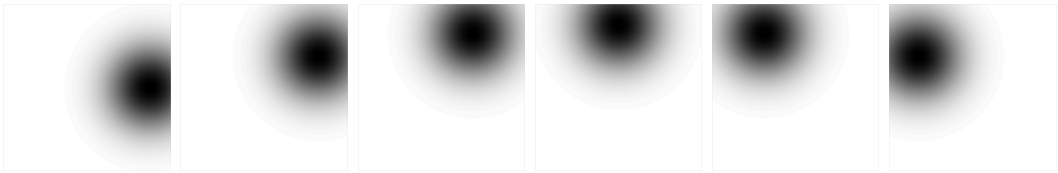


Figure 2.3. Fourier transform of the atoms obtained from the wavelet of Morlet with $k = 0$ and $\gamma \in G_+$. From left to right. $\gamma = 0$ to $\gamma = 5\pi/6$ by steps of $\pi/6$. The Fourier transforms are depicted for $\omega \in [-\pi, \pi] \times [-\pi, \pi]$.

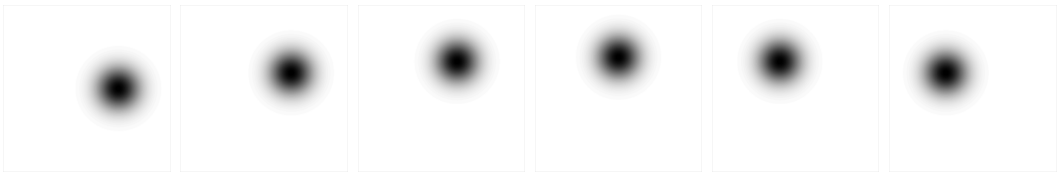


Figure 2.4. Fourier transform of the atoms obtained from the wavelet of Morlet with $k = 1$ and $\gamma \in G_+$. From left to right. $\gamma = 0$ to $\gamma = 5\pi/6$ by steps of $\pi/6$. The Fourier transforms are depicted for $\omega \in [-\pi, \pi] \times [-\pi, \pi]$.

sensitive to various frequency bands. The atoms are specifically sensitive to lower and lower frequencies as the scale parameter k increases. Let K be the maximum value of the scale parameter where K is, for the moment, a positive or negative integer. We will make K a natural number when working with discrete images as opposed to their continuous counterparts. Low frequencies below those captured at the coarsest scale K are captured by a low-pass filter

$$\phi_K(x) = 2^{-2K}\, \phi\big(\, 2^{-K} x \,\big) \tag{2.7}$$

whose Fourier transform is

$$\hat{\phi}_K(\omega) = \hat{\phi}(\, 2^K \omega \,) \; . \tag{2.8}$$

The Fourier transform $\hat{\phi}(\omega)$ of the low-pass filter has to be centred around the origin $\omega = 0$ of the frequency plane. The Fourier transform $\hat{\phi}_K(\omega)$ is also centred around the origin for $\hat{\phi}_K(\omega)$ is obtained from $\hat{\phi}(\omega)$ by a dilation of 2^K. The low-pass filter is also known as the *scaling function*. The introduction of a scaling function is frequent in the derivation of the wavelet transform of one-dimensional signals and images. The scaling function appears indeed in the derivation of the continuous wavelet transform [70] and in the derivation of the Littlewood Paley decompositions, leading, in the latter case, to pseudo quadrature mirror filters [92]. The scaling function also appears in the derivation of translation-invariant wavelet frames [69, 70]. The scaling function is finally at the core of the multi-resolution approximations, leading, for example, to wavelet bases and wavelet packets [70]. The introduction of a scaling function in the derivation of the wavelet transform of signals and images has its benefits from a computational perspective. It sets a lower bound to the scale k of the atoms. We will see, in section 2.3.5, that, when working with discrete images, the spatial sampling period associated with the discrete images sets an upper bound to the scale k of the atoms. As a result, only a finite number of atoms need be considered. We can assume, without loss of generality, that the upper bound is $k = 0$ (included). As a result, K has to be a natural number instead of an integer. The number of atoms to be considered is therefore equal to $2M(K+1)$ if both positive and negative rotations are considered, and equal to $M(K+1)$ if only positive rotations are considered.

2.2.4 Real-valued and complex-valued wavelets

We implicitly assumed in the preceding sections that the wavelet $\psi(x)$ is directional, namely that the support of its Fourier transform $\hat{\psi}(\omega)$ is localised in one and only one region of the frequency plane centred around ω_ψ. A directional wavelet has to be complex-valued. If the wavelet is real-valued, we find indeed that $\hat{\psi}(-\omega) = \hat{\psi}(\omega)^*$ where the superscript $*$ represents the complex conjugate. The support of the Fourier transform of a real-valued wavelet lies therefore on at least two quarters of the frequency plane. Examples of directional wavelets are the wavelet of Morlet and Cauchy wavelets. The wavelet of Morlet is a complex

exponential modulated by a Gaussian function up to a constant correction term, namely

$$\psi(x) = \psi(x_1, x_2) = \kappa \, \exp\left[-\alpha^2 \, x_1^2 - \alpha^2 \, x_2^2\right] \times \left[\exp\left[i \, \xi_0 \, x_1\right] - \eta\right] . \quad (2.9)$$

We will properly introduce this wavelet and its parameters in section 2.3.5. The support of the Fourier transform of Cauchy wavelets [93] strictly lies in a cone whose apex is at the origin of the frequency plane. Inside the cone, the Fourier transform of the wavelet is a decreasing exponential modulated by a polynomial, namely

$$\hat{\psi}(\omega) = \begin{cases} P(\omega) \ \exp\left[-\omega \cdot \varsigma\right] & \text{if} \quad \omega \in \text{ cone} \\ 0 \quad \text{otherwise} \end{cases} \quad (2.10)$$

where $P(\omega)$ is a polynomial in ω and where ς is a parameter which controls the decay of the exponential. We refer to [93] for a detailed presentation of Cauchy wavelets and their parameters. We used the wavelet of Morlet in all numerical experiments, following [76] and [77]. For the wavelet of Morlet and Cauchy wavelets, $\psi(x)$ is complex-valued but $\hat{\psi}(\omega)$ is real-valued. We will see in section 2.3.1 that, because $\hat{\psi}(\omega)$ is real-valued, only positive rotations of angles $\gamma \in G_+$ need be considered when analysing real-valued images.

2.2.5 Quantitative space frequency localisation

We have thus far been using the centre and the extent of the support of either $\psi(x)$ or $\hat{\psi}(\omega)$ without properly defining them. They can be defined by means of the energy distributions of the wavelet in space and frequency as in [94] and [95]. The energy of $\psi(x)$ and $\hat{\psi}(\omega)$ are defined by

$$\| \psi \|^2 = \int_{\mathbb{R}^2} | \psi(x) |^2 \, dx \quad \text{and} \quad \| \hat{\psi} \|^2 = \int_{\mathbb{R}^2} | \hat{\psi}(\omega) |^2 \, d\omega \quad (2.11)$$

with, according to Plancherel formula [69], $\| \hat{\psi} \|^2 = (2\pi)^2 \| \psi \|^2$. As a result,

$$| \psi(x) |^2 \, dx \, / \, \| \psi \|^2 \quad \text{and} \quad | \hat{\psi}(\omega) |^2 \, d\omega \, / \, \| \hat{\psi} \|^2 \quad (2.12)$$

can be interpreted as two-dimensional probability measures which respectively describe the energy distributions of the wavelet in space and frequency. The average values of x and ω along each dimension define the centres $x_\psi = (x_{\psi,1}, x_{\psi,2})$ and $\omega_\psi = (\omega_{\psi,1}, \omega_{\psi,2})$ of the supports of $\psi(x)$ and $\hat{\psi}(\omega)$. Similarly, the standard deviations of x and ω along each dimension define the extents $\Delta x_\psi = (\Delta x_{\psi,1}, \Delta x_{\psi,2})$ and $\Delta \omega_\psi = (\Delta \omega_{\psi,1}, \Delta \omega_{\psi,2})$ of the supports of $\psi(x)$ and $\hat{\psi}(\omega)$. Along the first dimension, we find

$$\begin{cases} x_{\psi,1} = \displaystyle\int_{\mathbb{R}^2} x_1 | \, \psi(x) \, |^2 dx \, / \, \| \psi \|^2 \\ \omega_{\psi,1} = \displaystyle\int_{\mathbb{R}^2} \omega_1 | \, \hat{\psi}(\omega) \, |^2 \, d\omega \, / \, \| \hat{\psi} \|^2 \end{cases} \quad (2.13)$$

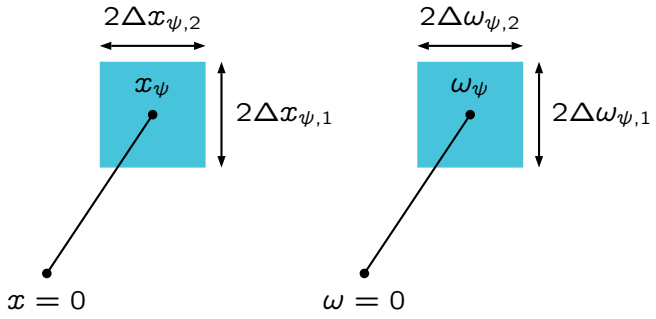


Figure 2.5. Supports of $\psi(x)$ and $\hat{\psi}(\omega)$.

and

$$\begin{cases} (\Delta x_{\psi,1}) = \Big[\int_{\mathbb{R}^2} (x_1 - x_{\psi,1})^2 \,|\, \psi(x) \,|^2 \, dx \,/\, \| \psi \|^2 \Big]^{1/2} \\ (\Delta \omega_{\psi,1}) = \Big[\int_{\mathbb{R}^2} (\omega_1 - \omega_{\psi,1})^2 \,|\, \hat{\psi}(\omega) \,|^2 \, d\omega \,/\, \| \hat{\psi} \|^2 \Big]^{1/2} . \end{cases} \tag{2.14}$$

Along the second dimension, we find

$$\begin{cases} x_{\psi,2} = \int_{\mathbb{R}^2} x_2 |\, \psi(x) \,|^2 dx \,/\, \| \psi \|^2 \\ \omega_{\psi,2} = \int_{\mathbb{R}^2} \omega_2 |\, \hat{\psi}(\omega) \,|^2 \, d\omega \,/\, \| \hat{\psi} \|^2 \end{cases} \tag{2.15}$$

and

$$\begin{cases} (\Delta x_{\psi,1}) = \Big[\int_{\mathbb{R}^2} (x_2 - x_{\psi,2})^2 \,|\, \psi(x) \,|^2 \, dx \,/\, \| \psi \|^2 \Big]^{1/2} \\ (\Delta \omega_{\psi,1}) = \Big[\int_{\mathbb{R}^2} (\omega_2 - \omega_{\psi,2})^2 \,|\, \hat{\psi}(\omega) \,|^2 \, d\omega \,/\, \| \hat{\psi} \|^2 \Big]^{1/2} . \end{cases} \tag{2.16}$$

The supports of $\psi(x)$ an $\hat{\psi}(\omega)$ are consequently localised in the rectangles depicted in Figure 2.5. At this point, it is interesting to realise that we have to quantify the localisation of $\psi(x)$ and $\hat{\psi}(\omega)$ in terms of momenta because $\psi(x)$ and $\hat{\psi}(\omega)$ cannot be both compactly supported. For the wavelet of Morlet and Cauchy wavelets, neither $\psi(x)$ or $\hat{\psi}(\omega)$ is compactly supported. We can similarly define the centre and the extent of the support of either $\phi(x)$ or $\hat{\phi}(\omega)$. We can also deduce the centres and the extents of the supports of the atoms in space and frequency should we combine (2.1) and (2.6) with (2.13) through to (2.16).

2.3 The Littlewood Paley wavelet transform of images

2.3.1 Introduction and definitions

We will present in this section the Littlewood Paley wavelet transform of images which is the wavelet transform scattering operators are based upon. It is important to bear in mind that this wavelet transform is one among many other types of

wavelet transforms [69, 70, 94]. The Littlewood Paley wavelet transform of an image $f(x)$ is the result of the convolution of the image with every atom and with the low-pass filter. This wavelet transform leads to a redundant image representation as opposed to the image representation induced, for example, by a wavelet basis. The convolutions of the image with the atoms and with the low-pass filter are indeed calculated at every position x and no spatial sub-sampling is performed after the convolutions. Should the image $f(x)$ be real-valued, which is the case for side-looking sonar images, only positive rotations of angles $\gamma \in G_+$ are considered [74]. The wavelet transform of a real-valued image $f(x)$ corresponds therefore to the following set of *wavelet images*

$$\mathcal{W}_K f = \left\{ f * \psi_{k,\gamma}(x) \, , \, f * \phi_K(x) \right\}_{k \leq K \, , \, \gamma \in G_+} \tag{2.17}$$

where $*$ represents the convolution operator. We have

$$f * \psi_{k,\gamma}(x) = \int_{\mathbb{R}^2} f(y) \, \psi_{k,\gamma}(x-y) \, dy \tag{2.18}$$

and

$$f * \phi_K(x) = \int_{\mathbb{R}^2} f(y) \, \phi_K(x-y) \, dy \tag{2.19}$$

or, in terms of the inverse Fourier transform,

$$f * \psi_{k,\gamma}(x) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \hat{f}(\omega) \, \hat{\psi}_{k,\gamma}(\omega) \, e^{i\omega x} \, d\omega \tag{2.20}$$

and

$$f * \phi_K(x) = \frac{1}{(2\pi)^2} \int_{\mathbb{R}^2} \hat{f}(\omega) \, \hat{\phi}_K(\omega) \, e^{i\omega x} \, d\omega \; . \tag{2.21}$$

When the Fourier transform $\hat{\psi}(\omega)$ and the analysed image $f(x)$ are real-valued, we find that $f * \psi_{k,\gamma}(x) = [\, f * \psi_{k,\gamma+\pi}(x) \,]^*$ where the superscript $*$ represents the complex conjugate. As a result, the information contained in the convolutions of the image with the atoms $\psi_{k,\gamma}(x)$ with $\gamma \in G_-$ is already contained in the convolutions of the image with the atoms $\psi_{k,\gamma}(x)$ with $\gamma \in G_+$. This is why only positive rotations of angles $\gamma \in G_+$ are needed when analysing real-valued images. When the analysed image is complex-valued, $f * \psi_{k,\gamma}(x) \neq [\, f * \psi_{k,\gamma+\pi}(x) \,]^*$ and both positive and negative rotations are needed.

2.3.2 Space-frequency image representations

We can assume, without loss of generality, that the wavelet and the low-pass filter are centred around the origin, namely that $x_\psi = 0$ and $x_\phi = 0$. As a result, the coefficient $f * \psi_{k,\gamma}(x)$ at a specific position x contains information about the image around this position x, and information about the frequency content of the image around $r_\gamma(\, \omega_\psi / 2^k \,)$. Similarly, the coefficient $f * \phi_K(x)$ at a specific position x contains information about the image around this position x, and information about the frequency content of the image below the one

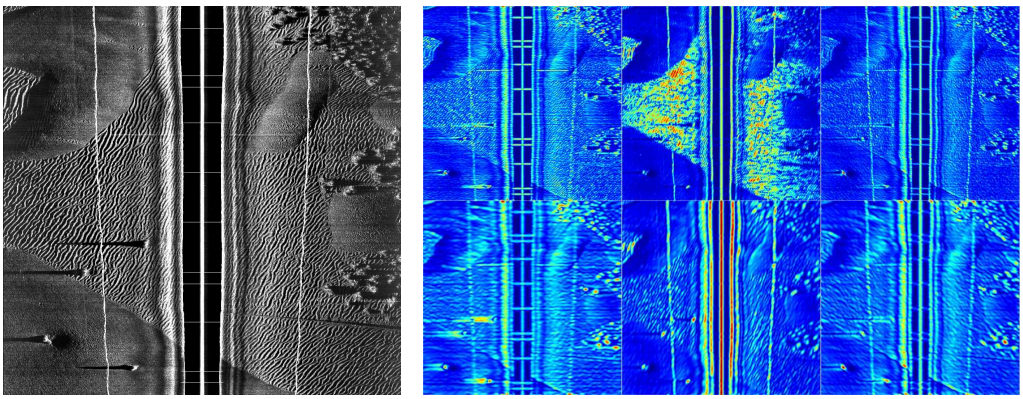


Figure 2.6. Left. Side-looking sonar image A. Right. Modulus of the complex wavelet images $f * \psi_{k,\gamma}(x)$ obtained with the wavelet of Morlet. From left to right. Orientation $\gamma = 0$, $\gamma = 2\pi/6$ and $\gamma = 5\pi/6$ in radians. Top. Scale $k = 2$. Bottom. Scale $k = 3$.

captured at the coarsest scale K. In this respect, the wavelet transform leads to space-frequency representations of images. The wavelet transform transforms a two-dimensional space to a four-dimensional space called the *phase space*. Two dimensions of the phase space correspond to the scale k and the angle γ. The other two dimensions correspond to the position x which appears in the convolutions.

For example, the modulus of the complex wavelet images $f * \psi_{k,\gamma}(x)$ obtained with the wavelet of Morlet out of three side-looking sonar images are depicted in Figures 2.6, 2.7 and 2.8. The wavelet images obtained with $\gamma = 0$ are sensitive to horizontal features, which can be observed around the shadows of the objects in Figure 2.6. The wavelet images obtained with $\gamma = 2\pi/6$ are sensitive to sand ripples. The wavelet images obtained with $\gamma = 2\pi/6$ are all the more sensitive to sand ripples when the scale k is adapted to the the spatial period of the ripples, which appears to be the case when $k = 2$ for the small sand ripples of Figure 2.6, and when $k = 3$ for the large sand ripples of Figure 2.7 and 2.8. The size of the three side-looking images is 1000×1024 pixels (vertically and horizontally). The wavelet images $f * \psi_{k,\gamma}(x)$ were calculated with the help of the FFTW library [96] from (2.20) according to an algorithm which can be found in either [69] or [94]. According to [69], the wavelet images $f * \psi_{k,\gamma}(x)$ which appear in the Littlewood Paley wavelet transform can only be calculated with an algorithm based on the discrete Fourier transform when the wavelet $\psi(x)$ is directional, which is the case for the wavelet of Morlet. The wavelet images $f * \phi_K(x)$ which are the result of the convolution of the side-looking sonar images with the scaling function $\phi_K(x)$ are depicted in Figure 2.9 for $K = 3$. Those wavelet images correspond to the average versions of the original images at the scale K of the scaling function. The wavelet images $f * \phi_K(x)$ were also calculated with the help of the FFTW library [96] from, this time, (2.21).

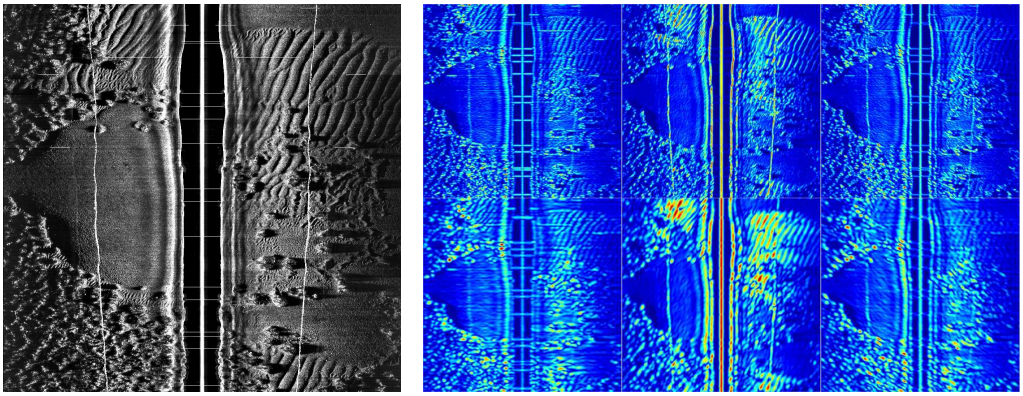


Figure 2.7. Left. Side-looking sonar image B. Right. Modulus of the complex wavelet images $f * \psi_{k,\gamma}(x)$ obtained with the wavelet of Morlet. From left to right. Orientation $\gamma = 0$, $\gamma = 2\pi/6$ and $\gamma = 5\pi/6$ in radians. Top. Scale $k = 2$. Bottom. Scale $k = 3$.

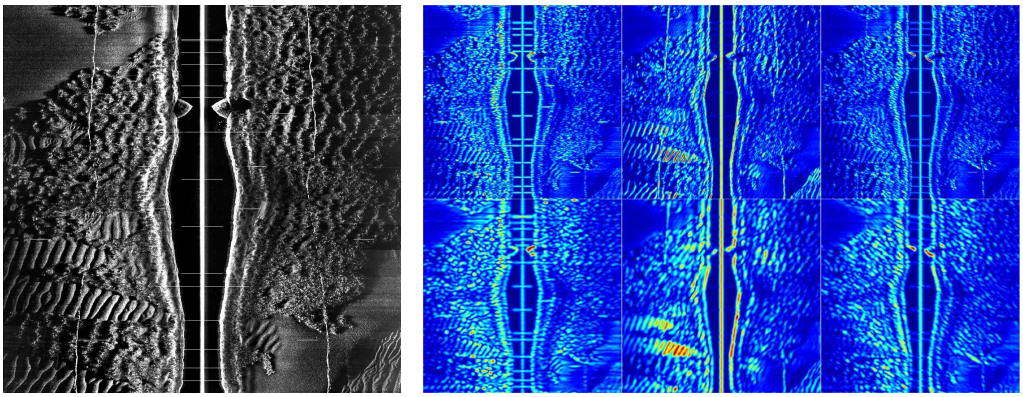


Figure 2.8. Left. Side-looking sonar image C. Right. Modulus of the complex wavelet images $f * \psi_{k,\gamma}(x)$ obtained with the wavelet of Morlet. From left to right. Orientation $\gamma = 0$, $\gamma = 2\pi/6$ and $\gamma = 5\pi/6$ in radians. Top. Scale $k = 2$. Bottom. Scale $k = 3$.

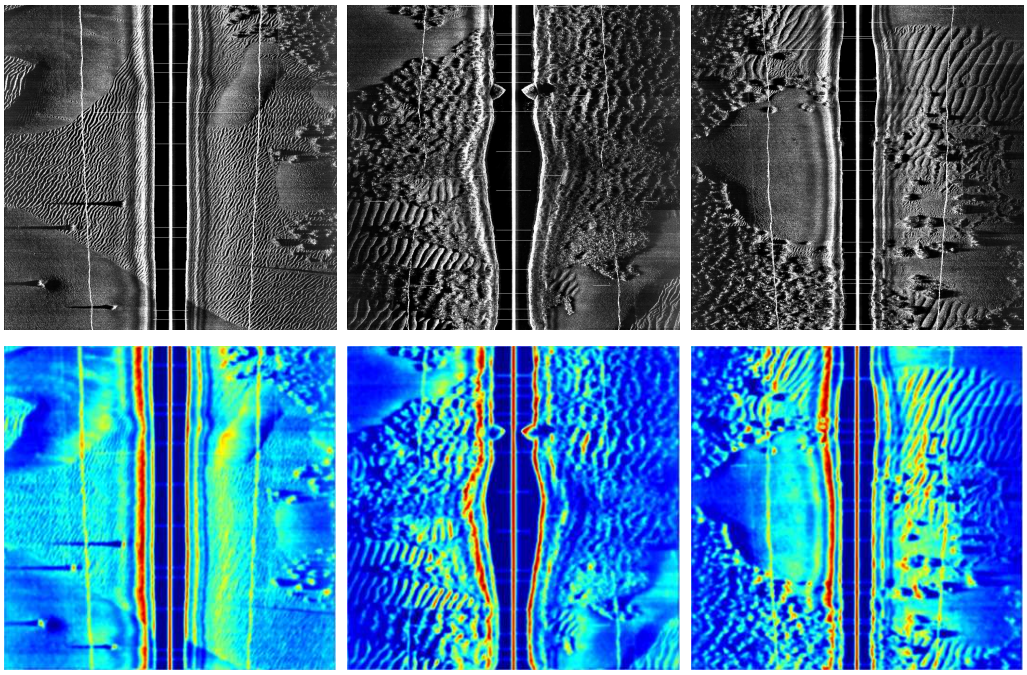


Figure 2.9. Wavelet images resulting of the convolution of the images A, B and C with the scaling function $\phi_K(x)$ for $K = 3$. Top. Side-looking sonar images. Bottom. Wavelet images.

2.3.3 A few properties of the wavelet transform

The Littlewood Paley wavelet transform is a linear operator because the convolution operator is linear. As a result, $\mathcal{W}_K(\alpha f + \beta g) = \alpha \mathcal{W}_K f + \beta \mathcal{W}_K g$ for any scalars α and β and any images f and g. The wavelet transform also commutes with translations because no spatial sub-sampling is performed on the convolutions of the image with the atoms and with the low-pass filter. If $L_\tau f(x) = f(x - \tau)$ is the translation of $f(x)$ in the constant direction τ, then $\mathcal{W}_K L_\tau f = L_\tau \mathcal{W}_K f$. The wavelet transform is therefore not translation invariant, namely $\mathcal{W}_K L_\tau f \neq \mathcal{W}_K f$.

2.3.4 Invertibility and non-expansivity of the wavelet transform

Scattering operators require the Littlewood Paley wavelet transform to be invertible and non-expansive. In order to clarify those two concepts, we need to define the energy of the wavelet transform. The energy of the wavelet transform $\mathcal{W}_K f$ is defined as the sum of the energy of the images in the set $\mathcal{W}_K f$, namely

$$\| \mathcal{W}_K f \|^2 = \| f * \phi_K \|^2 + \sum_{k=-\infty}^{K} \sum_{\gamma \in G_+} \left\{ \| f * \psi_{k,\gamma} \|^2 \right\} \tag{2.22}$$

with

$$\| f * \psi_{k,\gamma} \|^2 = \int_{\mathbf{R}^2} | f * \psi_{k,\gamma}(x) |^2 \, dx \tag{2.23}$$

and

$$\| f * \phi_K \|^2 = \int_{\mathbf{R}^2} | f * \phi_K (x) |^2 \, dx \tag{2.24}$$

in (2.22). The Littlewood Paley wavelet transform is invertible and non-expansive if there exists a real number ϵ with $0 \leq \epsilon < 1$ such that

$$(1 - \epsilon) \, \|f\|^2 \leq \| \mathcal{W}_K f \|^2 \leq \|f\|^2 \tag{2.25}$$

or equivalently [70] such that

$$1 - \epsilon \leq | \hat{\phi}(\omega) |^2 + \frac{1}{2} \sum_{k=-\infty}^{0} \sum_{\gamma \in G} \Big\{ | \hat{\psi}_{k,\gamma} (\omega) |^2 \Big\} \leq 1 \, . \tag{2.26}$$

(2.25) indicates that the energy of the wavelet transform is equivalent but not necessarily equal to the energy of the image. Equality arises when $\epsilon = 0$ in which case the wavelet transform is unitary and $\| \mathcal{W}_K f \| = \| f \|$. (2.25) also indicates that the wavelet transform is non-expansive, namely

$$\| \mathcal{W}_K f - \mathcal{W}_k g \| \leq \| f - g \| \tag{2.27}$$

for any images f and g. (2.25) finally indicates that the wavelet transform is invertible, in other words, that it possible to reconstruct the image $f(x)$ from the images in the set $\mathcal{W}_K f$. It is indeed possible to show [70] that there exists a set of reconstruction atoms $\tilde{\psi}_{k,\gamma} (x)$ and a reconstruction low-pass filter $\tilde{\phi}(x)$ such that

$$f(x) = f * \phi_K * \tilde{\phi}_K (x) + \sum_{k=-\infty}^{K} \sum_{\gamma \in G_+} \Big\{ f * \psi_{k,\gamma} * \tilde{\psi}_{k,\gamma} (x) \Big\} \tag{2.28}$$

with

$$\tilde{\phi}_K(x) = 2^{-2K} \, \tilde{\phi}\big(2^{-K} x \big) \tag{2.29}$$

If we define

$$S(\omega) = | \hat{\phi}(\omega) |^2 + \frac{1}{2} \sum_{k=-\infty}^{0} \sum_{\gamma \in G} \Big\{ | \hat{\psi}_{k,\gamma} (\omega) |^2 \Big\} \, , \tag{2.30}$$

one possible choice for the reconstruction atoms and the reconstruction low-pass filter is

$$\hat{\tilde{\psi}}_{k,\gamma} (\omega) = \hat{\psi}_{k,\gamma} (\omega)^* \, / \, S(\omega) \quad \text{and} \quad \hat{\tilde{\phi}}(\omega) = \hat{\phi}(\omega)^* \, / \, S(\omega) \, . \tag{2.31}$$

The reconstruction atoms and the reconstruction low-pass filter are all well defined because, according to (2.26), $S(\omega)$ is bounded above and below by 1 and $1 - \epsilon$, respectively. If the wavelet transform is not unitary, optimisation algorithms have to be used to obtain the reconstruction atoms and the reconstruction low-pass filter [70]. If the wavelet transform is unitary (namely, if $S(\omega) = 1$) the reconstruction atoms and the reconstruction low-pass filter are readily obtain

from the atoms and the low-pass filter according to

$$\hat{\bar{\psi}}_{k,\gamma}(\omega) = \hat{\psi}_{k,\gamma}(\omega)^* \quad \text{and} \quad \hat{\bar{\phi}}(\omega) = \hat{\phi}(\omega)^* \; . \tag{2.32}$$

Constraints on the wavelet $\psi(x)$ and the low-pass filter $\phi(x)$ can be drawn from (2.26). (2.26) states that $S(\omega)$ has to be bounded at all frequencies ω. Upon combining (2.6) and (2.30) we find that $S(0) = +\infty$ unless $\hat{\psi}(0) = 0$. As a result, the Fourier transform $\hat{\psi}(\omega)$ of the wavelet has to vanish at the origin of the frequency plane. The wavelet and the atoms strictly act therefore as band-pass filters. We also find from (2.30) that the Fourier transform $\hat{\phi}(\omega)$ of the low-pass filter has to satisfy

$$1 - \epsilon \leq |\, \hat{\phi}(0) \,|^2 \leq 1 \; . \tag{2.33}$$

In the derivation of scattering operators, the low-pass filter is chosen such that

$$\hat{\phi}(0) = \int_{\mathbb{R}^2} \phi(x)\, dx = 1 \tag{2.34}$$

in which case (2.33) is satisfied. Finally, it is important to notice that the sum in (2.25) runs over angles $\gamma \in G_+$ while the sum in (2.26) runs over angles $\gamma \in G$. We now show why this is the case, proving the implication (2.26) $\Rightarrow$ (2.25). We start from from (2.30) from which we have

$$\begin{aligned} S(2^K\omega) &= |\, \hat{\phi}(2^K\omega) \,|^2 + \frac{1}{2} \sum_{k=-\infty}^{0} \sum_{\gamma \in G} \Big\{ |\, \hat{\psi}_{k,\gamma}(2^K\omega) \,|^2 \Big\} \\ &= |\, \hat{\phi}(2^K\omega) \,|^2 + \frac{1}{2} \sum_{k=-\infty}^{K} \sum_{\gamma \in G} \Big\{ |\, \hat{\psi}_{k,\gamma}(\omega) \,|^2 \Big\} \; . \end{aligned} \tag{2.35}$$

Plancherel formula applied to (2.35) multiplied by $|\, f(\omega) \,|^2$ leads to

$$\begin{aligned} &\int_{\mathbb{R}^2} S(2^K\omega)\, |\hat{f}(\omega)\,|^2\, d\omega \\ &\qquad = (2\pi)^2 \,\|\, f * \phi_K \,\|^2 \; + \; \frac{(2\pi)^2}{2} \sum_{k=-\infty}^{K} \sum_{\gamma \in G} \Big\{ \|\, f * \psi_{k,\gamma} \,\|^2 \Big\} \end{aligned} \tag{2.36}$$

where the summation runs over angles in $\gamma \in G$. We assumed earlier on that the image $f(x)$ real-valued, which implies that $\hat{f}(-\omega) = \hat{f}(\omega)^*$. This leads in turn to $\|\, f * \psi_{k,\gamma} \,\| = \|\, f * \psi_{k,\gamma+\pi} \,\|$ for any scale k and any angle γ. We then find, from

the definition (2.3) of G, that

$$\int_{\mathbb{R}^2} S(2^K\omega)\,|\hat{f}(\omega)\,|^2\,d\omega$$
$$= (2\pi)^2\,\|\,f * \phi_K\,\|^2 \;+\; (2\pi)^2 \sum_{k=-\infty}^{K} \sum_{\gamma\in G_+} \Big\{\,\|\,f * \psi_{k,\gamma}\,\|^2\,\Big\}\,. \quad (2.37)$$

where the summation now runs over angles $\gamma \in G_+$. If (2.26) is satisfied, we have

$$(1-\epsilon)\int_{\mathbb{R}^2} |\hat{f}(\omega)\,|^2\,d\omega \le \int_{\mathbb{R}^2} S(2^K\omega)\,|\hat{f}(\omega)\,|^2\,d\omega \le \int_{\mathbb{R}^2} |\hat{f}(\omega)\,|^2\,d\omega \quad (2.38)$$

and then

$$(1-\epsilon)\,\|f\|^2 \le \;\|\,\mathcal{W}_K\,f\,\|^2 \;\le \|f\|^2\,. \quad (2.39)$$

The implication (2.26) $\Rightarrow$ (2.25) is then satisfied. We refer to [74] for a proof of the implication (2.26) $\Leftarrow$ (2.25).

2.3.5 The wavelet of Morlet for image analysis

We will present, in this section, the wavelet of Morlet which is a complex exponential modulated by a Gaussian function, up to a constant correction term. We will explain how to set the parameters of the wavelet of Morlet when working with discrete sonar images. We will also explain that, together with a Gaussian low-pass filter, or scaling function, the wavelet of Morlet leads to a Littlewood Paley wavelet transform which is invertible and non-expansive. The wavelet of Morlet is defined by

$$\psi(x) = \psi(x_1,x_2) = \kappa\,\exp\left[-\alpha^2\,x_1^2 - \alpha^2\,x_2^2\right] \times \left[\exp\left[\,i\,\xi_0\,x_1\,\right] - \eta\,\right] \quad (2.40)$$

where κ is a normalisation constant, where ξ_0 represents a spatial wave number, and where η is a constant correction term. The Fourier transform of the wavelet is

$$\hat{\psi}(\omega) = \hat{\psi}(\omega_1,\omega_2)$$
$$= \frac{\kappa\pi}{\alpha^2} \times \left\{\,\exp\left[-\frac{(\omega_1-\xi_0)^2+\omega_2^2}{4\alpha^2}\right] - \eta\,\exp\left[-\frac{\omega_1^2+\omega_2^2}{4\alpha^2}\right]\,\right\} \quad (2.41)$$

from which we find

$$\hat{\psi}(0) = \frac{\kappa\pi}{\alpha^2} \times \left\{\,\exp\left[-\xi_0^2/(4\alpha^2)\right] - \eta\,\right\}. \quad (2.42)$$

The constant correction term $\eta = \exp\left[-\xi_0^2\,/\,(4\alpha^2)\right]$ ensures therefore that $\hat{\psi}(0) = 0$ which is a requirement drawn from condition (2.26) in section 2.3.4. If we assume [69, 70, 77] that $\eta \ll 1$, we find that

$$\psi(x) = \psi(x_1,x_2) \simeq \kappa\,\exp\left[-\alpha^2\,x_1^2 - \alpha^2\,x_2^2\right] \times \exp\left[\,i\,\xi_0\,x_1\,\right] \quad (2.43)$$

and

$$\widehat{\psi}(\omega) = \widehat{\psi}(\omega_1, \omega_2) \simeq \frac{\kappa\pi}{\alpha^2} \times \exp\left[-\frac{(\omega_1 - \xi_0)^2 + \omega_2^2}{4\alpha^2} \right] . \tag{2.44}$$

From (2.43) we find that $x_\psi \simeq (0,0)$ and $\Delta x_\psi \simeq (1/\alpha, 1/\alpha)$ and $\|\psi\| = (\kappa/\alpha)\sqrt{\pi/2}$. From (2.44) we find that $\omega_\psi \simeq (\xi_0, 0)$ and $\Delta \omega_\psi \simeq (\alpha, \alpha)$ and $\|\widehat{\psi}\| \simeq \kappa\,\pi\sqrt{2\pi}\,/\,\alpha$. If η is in the order of 1 ($\eta \sim 1$) the expressions we just found for the centres and the extents of the supports of $\psi(x)$ and $\widehat{\psi}(\omega)$ have to be ever so slightly corrected.

Wavelet's parameters We choose the parameters ξ_0 and α of the wavelet of Morlet in order to create a *dyadic partition* of the portion of the frequency plane which is available when working with discrete sonar images. We can assume, without loss of generality, that the spatial sampling period associated with the discrete sonar images equals one in both directions. According to Shannon's sampling theorem, the available portion of the frequency plane is the square $[-\pi, \pi] \times [-\pi, \pi]$ where π is Nyquist's limit. We then set $\xi_0 = 3\pi/4$ and $\alpha = \pi - \xi_0 = \pi/4$ so that the support of the Fourier transform $\widehat{\psi}(\omega)$ lies in the annulus defined by the radii $\pi/2$ and π, which is illustrated in Figure 2.10. The atoms are obtained from the wavelet by dilations proportional to 2. As a result, the support of the Fourier transform $\widehat{\psi}_{k,\gamma}(\omega)$ lies in the annulus defined by the radii $\pi/2^{k+1}$ and $\pi/2^k$. The supports of the Fourier transforms of the atoms create therefore a dyadic partition of upper part of the square $[-\pi, \pi] \times [-\pi, \pi]$ when k varies from 0 to K and when $\gamma \in G_+$. Returning to what was said in section 2.2.3, the spatial sampling period associated with the discrete sonar images sets an upper bound to the scale parameter k of the atoms. The upper bound is 0 and the corresponding scale, the finest, can be interpreted as the scale of a pixel. We can verify *a posteriori* that $\eta = 0.11 \ll 1$ when $\xi_0 = 3\pi/4$ and $\alpha = \pi/4$. The normalisation constant $\kappa = \alpha/(\pi\sqrt{2\pi})$ is chosen so that $\|\widehat{\psi}\| = 1$.

One observation If the atoms were obtained from the wavelet by dilations proportional to a with a not necessarily equal to 2, an a-adic partition of the square $[-\pi, \pi] \times [-\pi, \pi]$ would be obtained in a similar fashion. The support of the Fourier transform $\widehat{\psi}_{k,\gamma}(\omega)$ now lies in the annulus defined by the radii π/a^{k+1} and π/a^k. As a result, the parameters ξ_0 and α have to satisfy

$$\xi_0 + \alpha = \pi \quad \text{and} \quad \xi_0/a + \alpha/a = \xi_0 - \alpha \tag{2.45}$$

which leads to

$$\xi_0 = \pi\,(a+1)/(2a) \quad \text{and} \quad \alpha = \pi\,(a-1)/(2a)\ . \tag{2.46}$$

If $a = 2$ we find $\xi_0 = 3\pi/4$ and $\alpha = \pi/4$ as before.

Scaling function's parameters The circle of radius $\pi/2^{K+1}$ centred around the origin $w = 0$ is not covered by any of the supports of the Fourier transforms of

the atoms but rather by the support of $\hat{\phi}_K(\omega)$. $\hat{\phi}_K(\omega)$ is obtained from $\hat{\phi}(\omega)$ by a dilation of 2^{-K}. What is then required is that the support of $\hat{\phi}(\omega)$ covers the circle of radius $\pi/2$ centred around the origin, which is illustrated in Figure 2.11. We use the Gaussian low-pass filter $\phi(x)$ defined by

$$\phi(x) = \phi(x_1, x_2) = \kappa' \exp\left[-\beta^2 x_1^2 - \beta^2 x_2^2 \right] \tag{2.47}$$

where κ' is another normalisation constant. The Fourier transform of the low-pass filter is

$$\hat{\phi}(\omega) = \frac{\kappa'\pi}{\beta^2} \times \exp\left[-\frac{\omega_1^2 + \omega_2^2}{4\beta^2} \right] \tag{2.48}$$

from which we obtain

$$\omega_\phi = (0,0) \quad \text{and} \quad \Delta\omega_\phi = (\beta, \beta) . \tag{2.49}$$

The support of $\hat{\phi}(\omega)$ covers therefore the circle of radius $\pi/2$ centred around the origin when $\beta = \pi/2$. The normalisation constant $\kappa' = \beta^2/\pi$ is chosen so that $\hat{\phi}(0) = 1$, which relates to condition (2.34) in section 2.3.4.

Littlewood Paley wavelet transform The wavelet of Morlet and the Gaussian scaling function lead together to a Littlewood Paley wavelet transform which is invertible and non-expansive if

$$1 - \epsilon \leq S(\omega) \leq 1 \quad \text{with} \quad S(\omega) = |\hat{\phi}(\omega)|^2 + \frac{1}{2} \sum_{k=-\infty}^{0} \sum_{\gamma \in G} \left\{ |\hat{\psi}_{k,\gamma}(\omega)|^2 \right\} \tag{2.50}$$

We consequently have to find a lower bound and an upper bound for $S(\omega)$. In order to account for the fact that the images we analyse are discrete, we estimated the bounds over the circle of radius ξ_0 [97]. Indeed, the atoms corresponding to $k < 0$ no longer take part in the wavelet transform. We numerically find that (2.50) is satisfied for $\epsilon = 0.47$. At this point, we know how to calculate the Littlewood Paley wavelet transform of an image. We know how to set the parameters of the wavelet and the scaling function. We also know that the chosen parameters lead to a wavelet transform which is invertible and non-expansive. We are now ready to introduce the scattering transform of images.

2.4 Scattering operators

2.4.1 Introduction to scattering operators

We will introduce in this section the scattering transform of images [74]. We start with the introduction of a slight change of notations. We denote by $\lambda = (k, \gamma)$ any pair of scale and orientation and we denote by $\Lambda = \{ (k, \gamma) : k \leq K \text{ and } \gamma \in G_+ \}$ the set of all possible pairs of scales and orientations. Only real-valued images are analysed so that $\gamma \in G_+$ in Λ (instead of $\gamma \in G = G_+ \cup G_-$). A path

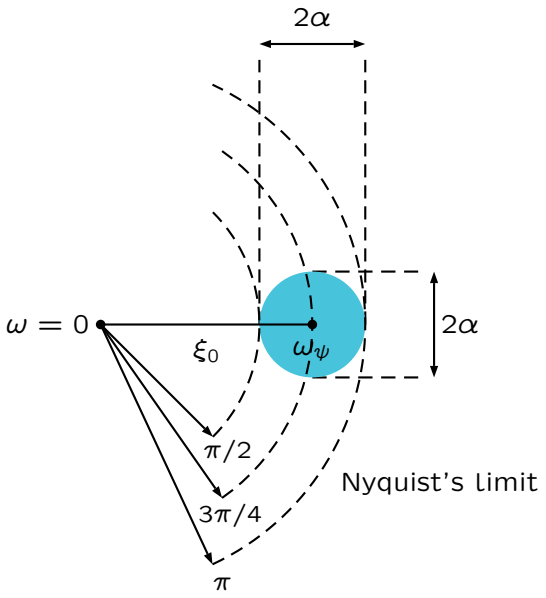


Figure 2.10. Supports of $\hat{\psi}(\omega)$ for the wavelet of Morlet.

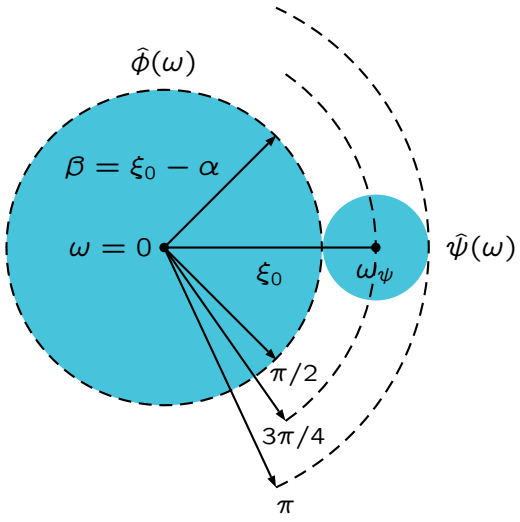


Figure 2.11. Supports of $\hat{\psi}(\omega)$ and $\hat{\phi}(\omega)$ for the wavelet of Morlet and the Gaussian scaling function.

$p = (\lambda_1, \ldots, \lambda_n) \in \Lambda^n$ of length $n \geq 0$ is defined as an ordered set of pairs of scales and orientations. The path of length zero is denoted by the symbol $\emptyset$ of the empty set. The scattering image calculated along the path p of length n is defined by

$$S_p f(x) = \underbrace{|\cdots|}_{n \text{ times}} f * \psi_{\lambda_1}| * \psi_{\lambda_2}| \cdots * \psi_{\lambda_n}| * \phi_K(x) \tag{2.51}$$

and is the result of n iterations over (1) the convolutions with the atoms and (2) the complex modulus operator followed eventually by a convolution with the scaling function $\phi_K(x)$. The scattering image calculated over the path of length zero is defined by

$$S_\emptyset f(x) = f * \phi_K(x) \; . \tag{2.52}$$

Scattering images $S_\emptyset f(x) = f * \phi_K(x)$ calculated with $K = 3$ over the path of length zero are depicted in Figure 2.12. We can see that the scattering images calculated over the path of length zero correspond the average versions of the original images at the coarsest scale K of the wavelet transform. The spatial resolution associated with the images $S_\emptyset f(x)$ is therefore that of the scaling function $\phi_K(x)$ (see Section 2.2.5 for a quantitative evaluation of the spatial resolution of the scaling function). The scattering images $S_\lambda f(x) = |f * \psi_\lambda| * \phi_K(x)$ calculated with $K = 3$ over paths $\lambda = (k, \gamma)$ of length one are depicted in Figures 2.13, 2.14 and 2.15 for the side-looking sonar images A, B and C. Those scattering images correspond to the averaged versions of the amplitude of the wavelet images $f * \psi_\lambda(x)$ at the coarsest scale K of the wavelet transform. The spatial resolution associated with the images $S_\lambda f(x)$ is consequently also that of the scaling function $\phi_K(x)$. Averaging the amplitude of the wavelet images induces a loss of information. The larger K, the stronger the averaging, and the greater the loss of information. The lost information can be recovered should we observe that $|f * \psi_\lambda| * \phi_K(x)$ is the low-frequency component of the wavelet transform of $|f * \psi_\lambda(x)|$. We have indeed

$$\mathcal{W}_K |f * \psi_\lambda| = \left\{ |f * \psi_\lambda| * \psi_{\lambda'}(x) \, , \, |f * \psi_\lambda| * \phi_K(x) \right\}_{\lambda' \in \Lambda} . \tag{2.53}$$

Because the Littlewood Paley wavelet transform is invertible, the lost information is contained in the wavelet images $|f * \psi_\lambda| * \psi_{\lambda'}(x)$ with $\lambda' \in \Lambda$. Those wavelet images lead in turn to a new set of scattering images calculated this time over paths of length two, namely

$$S_{\lambda,\lambda'} f(x) = ||f * \psi_\lambda| * \psi_{\lambda'}| * \phi_K(x) \; . \tag{2.54}$$

The spatial resolution associated with the scattering images (2.54) is once again that of the scaling function $\phi_K(x)$. At this point, it is important to observe that the λ's are frequency-related parameters which indicate the localisation in frequency of the atoms, whereas K is a spatially-related parameter which indicates the intrinsic spatial resolution of the scattering images. The scattering images calculated over paths of length two provide co-occurrence information about pairs (λ, λ') of scale and orientation for those scattering images involve the convolutions

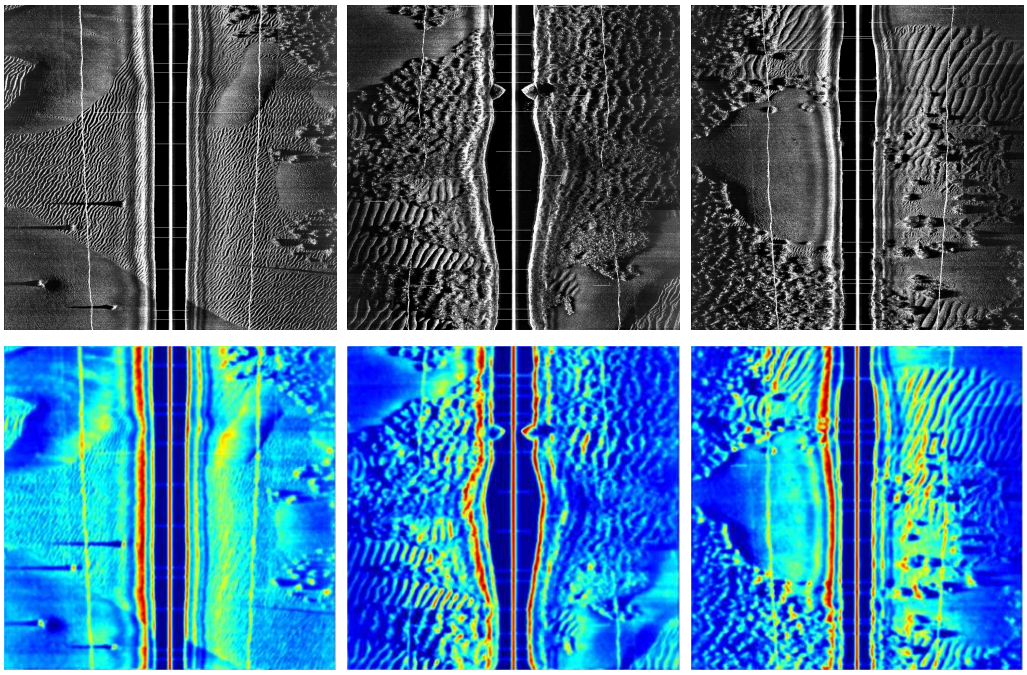


Figure 2.12. Scattering images $S_\emptyset f(x)$ corresponding to the path of length zero with $K = 3$. Top. Side-looking sonar images A, B and C. Bottom. Scattering images.

with the atoms $\psi_\lambda(x)$ and $\psi_{\lambda'}(x)$. Averaging the amplitude of the wavelet images $|f * \psi_\lambda| * \psi_{\lambda'}(x)$ by $\phi_K(x)$ induces another loss of information which can be recovered from by another wavelet transform, leading to a new set of scattering images calculated this time over paths of length three. When we repeat this mechanism, we obtain the scattering images (2.51) calculated over paths of length $n \geq 0$. At this point, it is important to realise that all scattering images have the same spatial resolution, which is the resolution, above that of a pixel, at which the scattering transform operates. We will make use of this fact in Chapter 4.

2.4.2 Scattering transform

We have thus far defined what is the scattering image calculated over a path p of any length, and highlighted some of the properties of the scattering images. We now define what is the scattering transform of an image. The scattering transform of an image $f(x)$ corresponds to the set of all scattering images calculated along all possible paths of any length. The scattering transform of the image $f(x)$ corresponds therefore to the set

$$Sf(x) = \left\{ S_p f(x) \ : \ p \in \mathcal{P} \right\} \tag{2.55}$$

where $\mathcal{P} = \cup_{n \geq 0} \Lambda^n = \emptyset \cup \Lambda \cup \Lambda^2 \cup \cdots \cup \Lambda^n \cup \cdots$ is the set of all possible paths of any length. The set (2.55) contains an infinite but countable number of elements. We will see in Section 2.4.5 that, when working with discrete sonar images, as opposed to their continuous counterparts, we can define a discrete

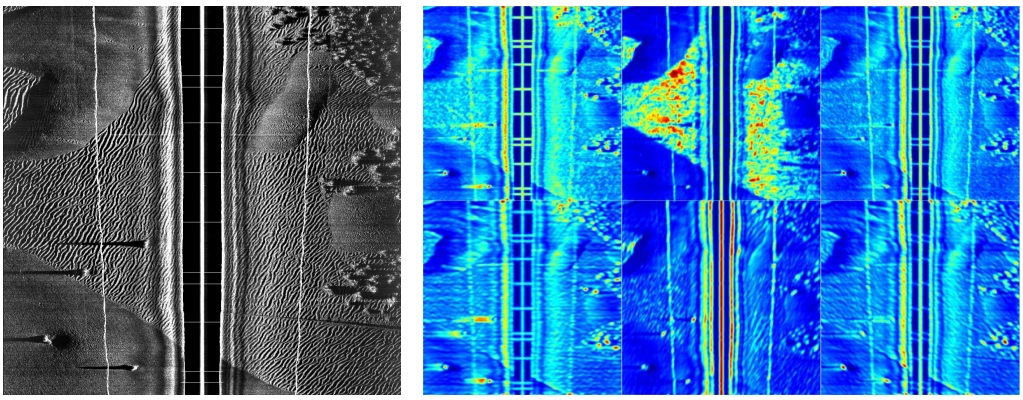


Figure 2.13. Scattering images $S_p f(x)$ with $K = 3$ corresponding to paths $p = (k, \gamma)$ of length one. Left. Side-looking sonar image A. Right. Scattering images. From left to right. Orientation $\gamma = 0$, $\gamma = 2\pi/6$ and $\gamma = 5\pi/6$ in radians. Top. Scale $k = 2$. Bottom. Scale $k = 3$.

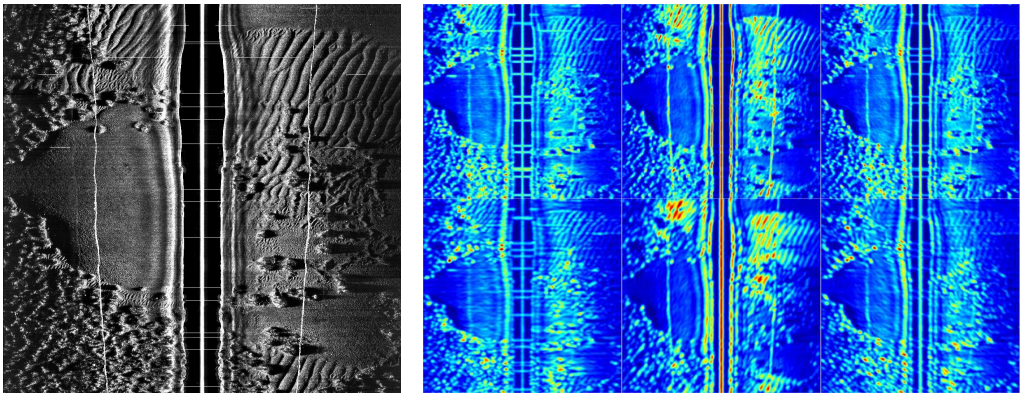


Figure 2.14. Scattering images $S_p f(x)$ with $K = 3$ corresponding to paths $p = (k, \gamma)$ of length one. Left. Side-looking sonar image B. Right. Scattering images. From left to right. Orientation $\gamma = 0$, $\gamma = 2\pi/6$ and $\gamma = 5\pi/6$ in radians. Top. Scale $k = 2$. Bottom. Scale $k = 3$.

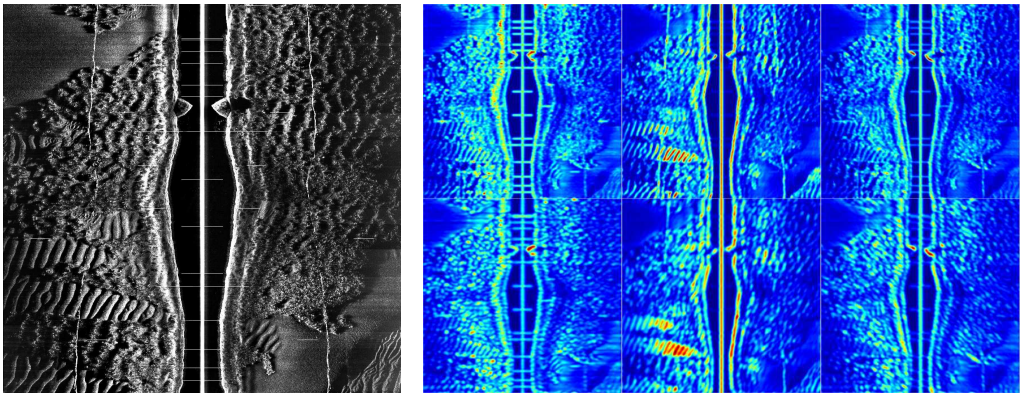


Figure 2.15. Scattering images $S_p f(x)$ with $K = 3$ corresponding to paths $p = (k, \gamma)$ of length one. Left. Side-looking sonar image C. Right. Scattering images. From left to right. Orientation $\gamma = 0$, $\gamma = 2\pi/6$ and $\gamma = 5\pi/6$ in radians. Top. Scale $k = 2$. Bottom. Scale $k = 3$.

scattering transform whereby only a finite number of scattering images need be considered. The energy of the scattering transform is defined as the sum of the energy of the scattering images the same way the energy of the Littlewood Paley wavelet transform is defined as the sum of the energy of the wavelet images. We have

$$\| Sf \|^2 = \sum_{p \in \mathcal{P}} \| S_p f \|^2 \tag{2.56}$$

where

$$\| S_p f \|^2 = \int_{\mathbf{R}^2} | S_p f(x) |^2 \, dx \tag{2.57}$$

is the energy of the scattering images. The energy of the scattering transform is defined by a summation over an infinite number of paths. However, the energy of the scattering images decreases exponentially as the length of the paths increases [74]. As a result, the summation can run over the set of paths of length below or equal to a natural number N, namely

$$\| Sf \|^2 \approx \sum_{p \in \emptyset \cup \Lambda \cup \Lambda^2 \cup \cdots \cup \Lambda^N} \| S_p f \|^2 \, , \tag{2.58}$$

leading to a "fair" approximation of the energy of the scattering transform. Only the scattering images defined over paths of length below or equal to N need be calculated, which reduces greatly the computational requirements of the scattering transform. In all numerical experiments, we used $N = 1$ or $N = 2$, because of what was just mentioned, and also because of memory requirements. The energy of the scattering transform defines a metric which can be used to compare two images f and g. The metric is

$$\| Sf - Sg \| = \Big[\sum_{p \in \mathcal{P}} \| S_p f - S_p g \|^2 \Big]^{1/2} \tag{2.59}$$

where

$$\| S_p f - S_p g \|^2 = \int_{\mathbf{R}^2} | S_p f(x) - S_p g(x) |^2 \, dx \, . \tag{2.60}$$

We refer to [74] for a comprehensive presentation of the properties of this metric. We just mention that, because the wavelet transform and the complex modulus operator are contractive, the metric is also contractive, that is $\| Sf - Sg \| \leq \| f - g \|$. Because of the exponential decay of the energy of the scattering images, the metric is approximated in all numerical experiments by

$$\| Sf - Sg \| \approx \Big[\sum_{p \in \emptyset \cup \Lambda \cup \Lambda^2 \cup \cdots \cup \Lambda^N} \| S_p f - S_p g \|^2 \Big]^{1/2} \, . \tag{2.61}$$

2.4.3 Scattering operators and convolution networks

The architecture of the scattering transform of the image $f(x)$ is similar to the architecture of a convolution network [75], which is illustrated in Figure 2.16. The root of the network corresponds to the original image $f(x)$. The first layer of

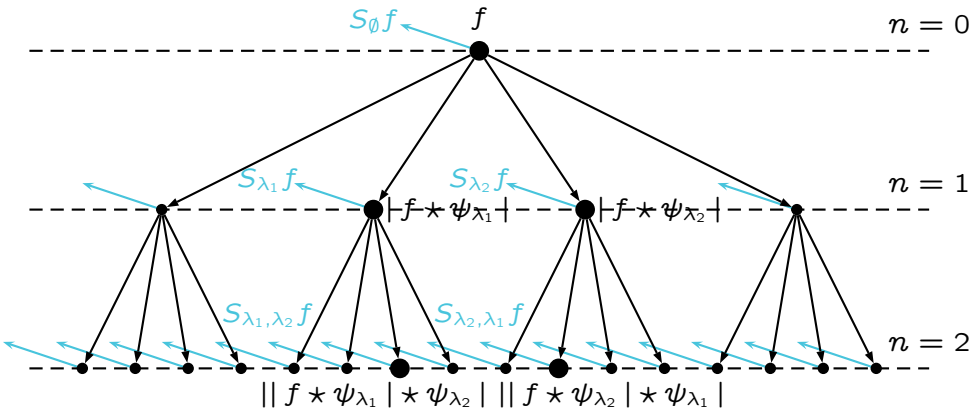


Figure 2.16. Architecture of a scattering operator. The root of the network corresponds to the original image. The nodes of the first layer of the network correspond to the images $|f * \psi_\lambda|$ for all $\lambda \in \Lambda$. The nodes of the second layer correspond to the images $||f * \psi_\lambda| * \psi_{\lambda'}|$ for all $\lambda, \lambda' \in \Lambda$. Each node output a scattering image from the action of the low-pass filter. The root of the network corresponds to interferences of order $n = 0$, the first layer to interferences of order $n = 1$, and the second layer to interferences of order $n = 2$.

the network is obtained from the root by one iteration over the wavelet transform and the complex modulus operator. The nodes of the first layer correspond therefore to the images $|f * \psi_\lambda|$ for all $\lambda \in \Lambda$. The second layer of the network is obtained from the first layer by another iteration over the wavelet transform and the complex modulus operator. The nodes of the second layer correspond therefore to the images $||f * \psi_\lambda| * \psi_{\lambda'}|$ for all $\lambda, \lambda' \in \Lambda$. Such a process is repeated until the maximum length N of the paths is reached. Each node of the network outputs a scattering image from the action of the low-pass filter $\phi_K(x)$. The root node outputs $S_\emptyset f$, the nodes of the first layer output $S_\lambda f$ for all $\lambda \in \Lambda$, and the nodes of the second layer output $S_{\lambda,\lambda'} f$ for all $\lambda, \lambda' \in \Lambda$. The nodes of a layer n with $0 \leq n \leq N$ output $S_p f$ for all paths $p \in \Lambda^n$.

2.4.4 Energy decay of the scattering images

We mentioned in Section 2.4.2 that according to [74] the energy of the scattering images decays exponentially as the length of the paths increases. We will now confirm that this is the case for the side-looking sonar images A, B and C. We denote the length of a path p by $l(p)$. Experimentally, we observe that we should have for all paths p

$$\log \| S_p f \|^2 / \| f \|^2 = \alpha - 2\, l(p) \tag{2.62}$$

where α is a finite constant, and where the constant 2 seems to relate to the dyadic scale progression of the atoms. The quantity $\log \|S_p f\|^2 / \|f\|^2$ is depicted in Figure 2.17 as a function of $l(p)$ for the scattering images $S_p f(x)$ obtained out

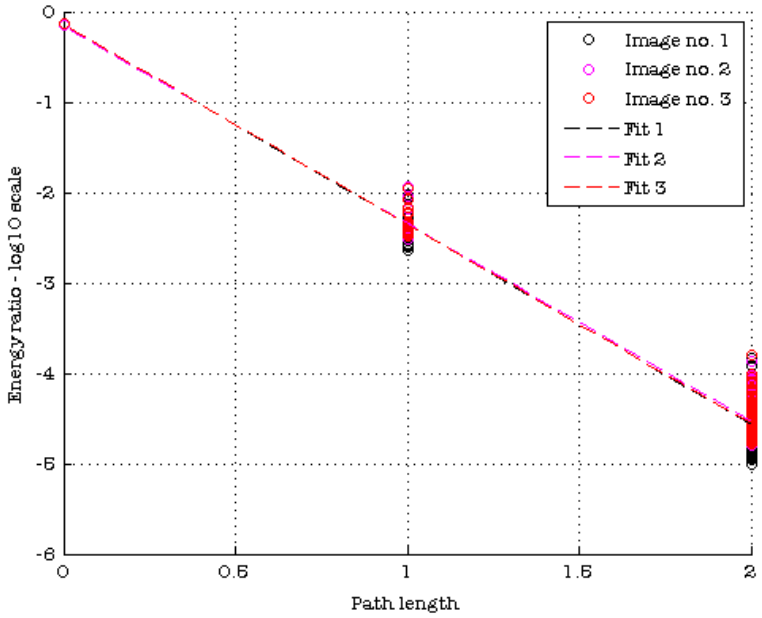


Figure 2.17. Energy decay of the scattering images $S_p f(x)$ obtained out of three sonar images. The circles represent the quantity $\log \|S_p f\|^2 / \|f\|^2$ as a function of the length $l(p)$ of the paths. Each circle corresponds to a scattering image. The dotted lines correspond to a linear fitting.

of the side-looking sonar images A, B and C. A linear fitting leads to

$$\begin{cases} \log \|S_p f\|^2 / \|f\|^2 = -0.15 - 2.2\, l(p) & \text{(image A)} \\ \log \|S_p f\|^2 / \|f\|^2 = -0.16 - 2.2\, l(p) & \text{(image B)} \\ \log \|S_p f\|^2 / \|f\|^2 = -0.14 - 2.2\, l(p) & \text{(image C)}\ , \end{cases} \tag{2.63}$$

suggesting that the energy of the scattering images does decay exponentially as the length of the paths increases.

2.4.5 Fast scattering operators

We introduced in section 2.4.2 the energy $\| Sf \|^2$ of the scattering transform of an image $f(x)$ as the sum of the energy of all possible scattering images. We explained that, because the energy of the scattering images decreases exponentially as the length of the paths increases, the energy of the scattering transform can be approximated by

$$\| Sf \|^2 \approx \sum_{p \in \emptyset \cup \Lambda \cup \Lambda^2 \cup \cdots \cup \Lambda^N} \| S_p f \|^2 \tag{2.64}$$

where N, a natural number, is the maximum length of the paths. Only the scattering images calculated over paths of length below or equal to N need be considered. In this section, we will explain that the approximation (2.64) can be

refined by considering the scattering images calculated over *frequency-decreasing paths* of length below or equal to N. We first consider two pairs of scale and orientation $\lambda = (k, \gamma)$ and $\lambda' = (k', \gamma')$ and the scattering image

$$S_{\lambda,\lambda'} f(x) = || f * \psi_\lambda | * \psi_{\lambda'} | * \phi_K (x) . \tag{2.65}$$

The frequency content of the image $f * \psi_\lambda (x)$ lies in the annulus defined by the radii $\pi/2^{k+1}$ and $\pi/2^k$ because the atom $\psi_\lambda(x)$ strictly acts as a band-pass filter. The width of the annulus is equal $\pi/2^{k+1}$. As a result, the frequency content of $| f * \psi_\lambda (x) |$ lies in the circle of radius $\pi/2^{k+1}$ centred around the origin of the frequency plane. If we write $f_\lambda (x) = f * \psi_\lambda (x)$, the Fourier transform of $| f * \psi_\lambda (x) | = | f_\lambda(x) |$ is indeed

$$\widehat{| f_\lambda |}(\omega) = \int_{\mathbf{R}^2} \widehat{f_\lambda} (\omega') \, \widehat{f_\lambda} (\omega' - \omega)^* \, d\omega' \tag{2.66}$$

and behaves like an auto-correlation function. This function of ω is therefore maximum around the origin of the frequency plane. The extent of this function, around the origin, corresponds to the extent of the support of $\widehat{f_\lambda} (\omega)$ which is equal to $\pi/2^{k+1}$. The frequency content of $| f * \psi_\lambda (x) |$ does therefore lie in the circle of radius $\pi/2^{k+1}$ centred around the origin of the frequency plane. The effect of the modulus operator is to "push" the energy of the wavelet coefficients $f * \psi_\lambda (x)$ towards low-frequencies. As a result, the coefficients $| f * \psi_\lambda | * \psi_{\lambda'} (x)$ only are non-negligible if the support of $\widehat{\psi}_{\lambda'} (\omega)$ lies in the circle of radius $\pi/2^{k+1}$, that is to say, if $k' > k$. In such a case, (λ, λ') is a frequency-decreasing path. Such a reasoning can be applied the scattering images calculated along paths of arbitrary length, from which we conclude that the scattering images need only be calculated over frequency-decreasing paths. We define $\Lambda_\downarrow^n = \{ (\lambda_1, \ldots, \lambda_n) \in \Lambda^n : k_1 < k_2 < \ldots < k_n \}$ the set of frequency decreasing paths of length n. We have $\Lambda_\downarrow^0 = \emptyset$ and $\Lambda_\downarrow^1 = \Lambda$. As a result, the approximation (2.64) can be refined according to

$$\| Sf \|^2 \approx \sum_{p \in \emptyset \cup \Lambda \cup \Lambda_\downarrow^2 \cup \cdots \cup \Lambda_\downarrow^N} \| S_p f \|^2 . \tag{2.67}$$

We know that when working with discrete sonar images, the spatial sampling period associated with the discrete images sets an upper bound to the scale k of the atoms. The upper bound is $k = 0$ (included) as explained in Section 2.3.5. Among all frequency-decreasing paths $p = (\lambda_1, \ldots, \lambda_n) \in \Lambda_\downarrow^n$ of length n, the paths that can be and are considered when working with discrete sonar images are the ones for which $k_i \geq 0$ for all indexes i between 1 and n. We therefore define the following subset of frequency-decreasing paths of length n.

$$\bar{\Lambda}_\downarrow^n = \{ (\lambda_1, \ldots, \lambda_n) \in \Lambda_\downarrow^n : k_1 \geq 0 , \, k_2 \geq 0 , \, \ldots , \, k_n \geq 0 \} . \tag{2.68}$$

We have for example $\bar{\Lambda}_\downarrow^0 = \emptyset$ and $\bar{\Lambda}_\downarrow^1 = \{ (k, \gamma) : 0 \leq k \leq K \text{ and } \gamma \in G_+ \}$ and $\bar{\Lambda}_\downarrow^2 = \{ (\lambda, \lambda') : 0 \leq k, k' \leq K \text{ and } \gamma, \gamma' \in G_+ \text{ and } k' > k \}$. We can now define the discrete scattering transform $\bar{S}$ of an image as the set of all scattering images

than can possibly be calculated when working with discrete images. The discrete scattering transform of an image $f(x)$ is therefore defined by

$$\bar{S}f(x) = \left\{ S_p f(x) \ : \ p \in \emptyset \cup \cdots \cup \bar{\Lambda}_{\downarrow}^n \cup \cdots \cup \bar{\Lambda}_{\downarrow}^N \right\} . \tag{2.69}$$

If N is the maximum length of the paths, M the number of angles in G_+, and K the depth of the wavelet transform, the number of scattering images involved the set (2.69) is

$$\sum_{n=0}^{N} M^n \binom{K+1}{n} = 1 + M(K+1) + M^2(K+1)K/2 + \cdots + M^N \binom{K+1}{N} . \tag{2.70}$$

where $\binom{K+1}{n}$ is the binomial coefficient between $K+1$ and n.

2.4.6 A C implementation of fast scattering operators

We will present in this section a way to calculate the discrete scattering transform of an image in C. The emphasis will be on the data structure we used to calculate scattering images along frequency-decreasing paths, and to store those images in memory. The data structure we used is a linked list [98] which corresponds to a root and an ordered collection of nodes, as illustrated in Figure 2.18. A node is essentially composed of a link (or reference) to an image, ultimately, a scattering image, and of a link (or reference) to the following node. The links are pointer variables containing the address in memory of either a node or an image. The last node is linked to a NULL node, which is a standard way to terminate a linked list. The root is essentially composed of a link to the first node, and of a link to the last node. It is therefore possible to add a node either at the beginning or at the end of the list. It is also possible to traverse through the list from the first node to the last one, and hence gain access to the images associated with each node. The linked list is constructed sequentially in several steps, the objective being to calculate the scattering images

$$S_p f(x) = |\cdots| f * \psi_{\lambda_1}| * \psi_{\lambda_2}| \cdots * \psi_{\lambda_n}| * \phi_K(x)$$

along paths $p = (\lambda_1, \ldots, \lambda_n) \in \bar{\Lambda}_{\downarrow}^n$ of length n with $0 \leq n \leq N$. The steps are the following. First, the root of the linked list is created. A node is created, added to the list and linked to the original image $f(x)$. Then, the original image $f(x)$ is used to calculate the images

$$|\cdots| f * \psi_{\lambda_1}| * \psi_{\lambda_2}| \cdots * \psi_{\lambda_n}(x)|$$

for all $(\lambda_1, \ldots, \lambda_n) \in \bar{\Lambda}_{\downarrow}^n$ and $1 \leq n \leq N$. Nodes are created added to the list, and linked to every one of those images. This is done by traversing through the list from the first node to the last one. Finally, the list is traversed through from the first node to the last one one more time. In this process, the convolutions between the images the nodes are linked to and the scaling function $\phi_K(x)$ are

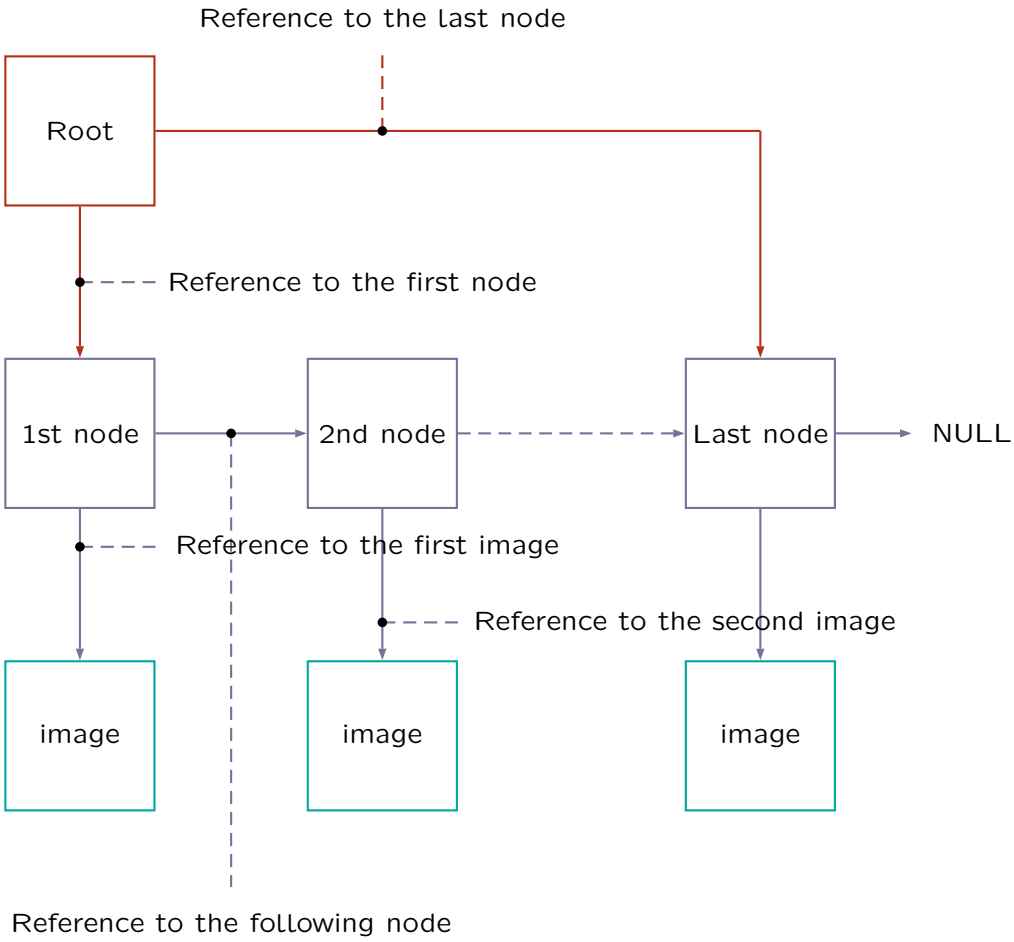


Figure 2.18. Linked list.

calculated, leading eventually to the scattering images $S_p f(x)$ for $p \in \bar{\Lambda}^n_{\downarrow}$ and $0 \leq n \leq N$.

2.5 Conclusions and perspectives

This chapter was mainly concerned with the scattering representation of images. The beginning of the chapter dealt with the Littlewood Paley wavelet transform of images which is one among many other kinds of two-dimensional wavelet transforms. The Littlewood Paley wavelet transform of images was considered from a mathematical perspective and from an algorithmic perspective. Both perspectives are indeed needed to fully understand and use any sort of wavelet transform. The end of the chapter dealt with the scattering representation of images. This representation, derived from the Littlewood Paley wavelet transform, possesses many advanced mathematical properties. One property which is important for the rest of the thesis is that the scattering representation possesses a constant intrinsic spatial resolution. This resolution is related to the depth of the Littlewood Paley wavelet transform, and makes the scattering representation stable with respect to local image deformations. We shall see in Chapter 3 that such a stability makes the scattering representation very well adapted to the textures commonly encountered in side-looking sonar images. We shall also make use of the intrinsic spatial resolution of the scattering representation in Chapter 4 where an inference algorithm based on the concept of similarity by composition will be presented.

The achievements of the present chapter are the following. We successfully implemented the Littlewood Paley wavelet transform of images in C with the help of an algorithm based on the discrete Fourier transform. This algorithm made use of the FFTW library which is a collection of fast C routines for computing the discrete Fourier transform. We provided this chapter with the Littlewood Paley wavelet transforms of a few side-looking sonar images with several types of seabed, and showed that specific aspects of the types of seabed of these images were "captured" by the wavelet transform. At this point, it is important to mention that the algorithm based on the discrete Fourier transform is the only one we could use to implement the Littlewood Paley wavelet transform of images. Had we been considering the Littlewood Paley wavelet transform of one-dimensional signals, such as the recording of the voltage drop across a capacitor, we could have used other algorithms that are faster than the one based on the discrete Fourier transform. Some of these algorithms are exact and others are approximate. This distinction depends on the wavelet.

In this chapter, we decided to use the wavelet of Morlet, and we explained how to chose the parameters of this wavelet, which is, as far as we know, something that is never explained in the literature. We explained that the parameters of the wavelet are obtained in order to create a partition of the frequency plane that is adapted to the scale progression of the atoms. To be more precise, the scale progression of the atoms was chosen dyadic and, according to Shannon's regular sampling theorem, only a portion of the frequency plane (below Nyquist's frequency) is available when working with discrete images. This portion of the frequency plane is hence partioned in a dyadic fashion. We finally showed that when the parameters of the wavelet of Morlet are chosen in such a way, this wavelet leads to a Littlewood Paley wavelet transform which is non-expansive

and invertible.

In this chapter, we also successfully implemented the scattering transform of images in C, making use of a specific C data structure called a linked list. We provided this chapter with the scattering transforms of a few side-looking sonar images, and explained how different they were from the Littlewood Paley wavelet transforms of the same images. At this point, we ought to be discussing the algorithmic complexity of the scattering transform of images. It is undeniable from what was written in this chapter that the scattering representation of images is algorithmically very complex. The representation of a *single* image is indeed made of between one hundred and two hundred scattering images (the exact number of scattering images depends on the values given to the parameters of the scattering transform). Besides, the scattering images cannot be obtained independently from one another, which is why we had to use a linked list in our C implementation of the scattering transform, that is, a rather advanced data structure. We find, however, that the algorithmic complexity of the scattering representation is balanced by the fact that it only depends on three parameters. The image representations presented in Chapter 1 depend on a comparatively larger number of parameters. Sum and difference histograms and co-occurrence distributions (of common use in the field of sonar imagery) and local binary patterns notably do so. When an image representation is used for illustrative purposes, which sometimes happens, the number of parameters upon which the representation depends are not of much importance. When an image representation is used in conjunction with an inference algorithm, which shall be the case in the following chapters, the number of parameters upon which the representation depends are of a much greater importance. The parameters have indeed to be set carefully according to the purpose of the inference algorithm, and the more parameters are to be set, the harder a task that is. In this respect, there is a clear advantage in using scattering operators to represent side-looking sonar images.

At this point of the thesis, we are ready to provide an answer to the two problems defined in Chapter 1. These are the supervised identification of the types of seabed that are present in side-looking sonar images, and the reduction, in the form of a post-processing process, of the number of false positives in a set of detections provided by a target detector. This will be done in Chapters 3 and 4 with the help of two inference algorithms, one based on affine spaces, and the other based on the concept of similarity by composition. Both inference algorithms make use of the scattering representation of images though in a different way.

Chapter 3

Image analysis with affine spaces

"In the space between chaos and
shape there was another chance."

Jeanette Winterson, The World
and Other Places.

3.1 Inference algorithms with affine spaces

3.1.1 Introduction

In the previous chapter, we presented the scattering transform of images. We showed that the scattering transform of an image corresponds to a set of scattering images obtained out of a Littlewood Paley wavelet transform. We emphasised the algorithmic and numerical aspects of the scattering transform because these aspects have to be considered with great care when working with discrete images. In this chapter, we will address the two problems defined in Chapter 1 with the help of the signal representation induced by the scattering transform. These problems are the supervised identification of the types of seabed present in side-looking sonar images, and the reduction, in the form of a post-processing process, of the number of false positives in a set of detections provided by a target detector. Let us consider first the supervised identification of the types of seabed. To solve this problem, we consider a query signal in one of a sonar images under study, that is to say, a rectangular region in one of the sonar images under study. We also consider a set of classes corresponding to several types of seabed. The members of the classes also are rectangular signals. The query signal and the member of the classes are all described by the scattering transform. A model based on an affine space is associated with each class. A decision is made upon the type of seabed of the query signal by looking at the distance between the query signal and each affine space. The distance is calculated in the space generated by the scattering transform. The affine spaces are created out of a training set, which is

why we say that the identification of the types of seabed is supervised. We created the training set by extracting signals out of a few sonar images. We also created a testing set by extracting signals out of another set of sonar images in order to quantitatively evaluate the performance of the identification of the types of seabed. Our approach is based on the signal representation induced by scattering operators but could have well been based on another signal representation, such as the one induced by local binary patterns, co-occurrence distributions, or even Gabor atoms. A study of the seabed identification performance obtained with several signal representations is always valuable. We therefore compared the seabed identification performance obtained with the signal representation induced by scattering operators with the seabed identification performance obtained with the signal representation induced by local binary patterns. We chose local binary patterns because they are one of the best texture representations available at the moment, unifying the statistical and structural representations of textures, namely co-occurrence distributions and Gabor atoms.

Let us consider now the problem of reducing the number of false positives in a set of detections provided by a target detector. This is a clutter rejection problem. The objective here is to look at the similarity between the detections and the seabed, and remove, from the initial set of detections, the detections which are similar to the seabed. These detections are meant to be false positives because a false positive is by nature part of the seabed. We used, in this chapter, a target detector based on a cascade of boosted classifiers and Haar-like features because this detector is one of the best target detectors available at the moment. This detector specifically has a low number of false positives per square kilometre. As a result, the detections to be removed from the initial set of detections are not obvious false positives, as would be portions of flat seabed, but are rather similar to the true positives. The similarity between the true and the false positives makes the clutter rejection problem challenging, the objective being indeed to ultimately remove as many false positives as possible without affecting the number of true positives. A detection provided by the cascade of boosted classifiers corresponds to a rectangular region in a sonar image. We then look at a detection as a query signal. We also consider a set of classes corresponding to several types of seabed, the members of the classes being rectangular signals as well. A model based on an affine space is associated with each class. We say that a query signal (a detection) is not similar to the seabed if it is far away from all the affine spaces, that is to say, from all the considered types of seabed. To be more precise, we define a boundary around each affine space. We then say that the query signal (the detection) is not similar to the seabed if, for every affine space, the distance between the query signal and the affine space falls beyond the affine space's boundary. If this is the case, the detection remains in the initial set of detections. If this is not the case, the detection is similar to at least one type of seabed and is hence removed from the initial set of detections.

This chapter will be organised in the following fashion. We will first present the construction of the affine spaces and the derivation of the boundaries. We will then address the supervised identification of the types of seabed and the clutter

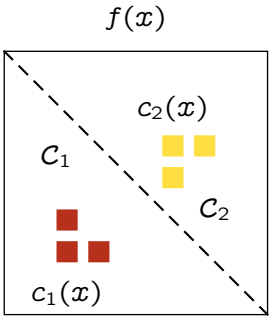


Figure 3.1. Signals $c_1(x)$ and $c_2(x)$ extracted from an image $f(x)$ with two classes $\mathcal{C}_1$ and $\mathcal{C}_2$. Three signals are extracted from each class and are depicted by coloured squares.

rejection problem. The work on the identification of the types of seabed was published in [3] and this paper is currently being reviewed.

3.1.2 Scattering transforms of signals

Let us consider an inference problem with Q classes $\mathcal{C}_q$ with $1 \leq q \leq Q$ and $Q \geq 1$. Each class will ultimately correspond to a type of seabed. Each class is represented by a real-valued random variable $C_q(x)$ whose realisations are rectangular signals $c_q(x)$ extracted from the sonar images under study, as illustrated in Figure 3.1. We know from chapter 2 that the scattering transform $Sf(x)$ of an image $f(x)$ corresponds to the set of all scattering images calculated along all possible paths of any length, that is to say

$$Sf(x) = \{ S_p f(x) \ : \ p \in \mathcal{P} \} \tag{3.1}$$

where $\mathcal{P} = \cup_{n \geq 0} \Lambda^n = \emptyset \cup \Lambda \cup \Lambda^2 \cup \cdots \cup \Lambda^n \cup \cdots$ is the set of all possible paths of any length. Should a signal $c_q(x)$ correspond to a region $\mathcal{R}$ in the image $f(x)$, the scattering transform $S_q(x)$ of $c_q(x)$ is obtained by extracting the same region from the scattering transform of $f(x)$, that is to say

$$Sc_q(x) = \{ S_p f(x) \ : \ p \in \mathcal{P} \text{ and } x \in \mathcal{R} \} \ . \tag{3.2}$$

$Sc_q(x)$ corresponds to a set of signals with an infinite but countable number of elements the same way $Sf(x)$ corresponds to a set of scattering images with an infinite but countable number of elements. We know from chapter 2 that when working with discrete sonar images, as opposed to their continuous counterparts, only a finite number of scattering images need be considered. Those images are the ones calculated along frequency-decreasing paths $\bar{\Lambda}^n_\downarrow$ with $0 \leq n \leq N$ where N is the maximum length of the paths. The discrete scattering transform $\bar{S}f(x)$ of the image $f(x)$ is

$$\bar{S}f(x) = \{ S_p f(x) \ : \ p \in \emptyset \cup \cdots \cup \bar{\Lambda}^n_\downarrow \cup \cdots \cup \bar{\Lambda}^N_\downarrow \} \tag{3.3}$$

from which we obtain the discrete scattering transform $\bar{S}c_q(x)$ of $c_q(x)$,

$$\bar{S}c_q(x) = \left\{ S_p f(x) \,:\, p \in \emptyset \cup \cdots \cup \bar{\Lambda}^n_{\downarrow} \cup \cdots \cup \bar{\Lambda}^N_{\downarrow} \text{ and } x \in \mathcal{R} \right\} . \tag{3.4}$$

$\bar{S}c_q(x)$ corresponds to a finite set of signals. We know from chapter 2 that this set contains $\sum_{n=0}^{N} M^n \binom{K+1}{n}$ elements where M and K are respectively the number of orientations in G_+ and the depth of the Littlewood Paley wavelet transform. Should the region $\mathcal{R}$ be made of $L_1 L_2$ pixels, the discrete representation $\bar{S}c_q(x)$ of $c_q(x)$ is made of

$$L = L_1 L_2 \times \sum_{n=0}^{N} M^n \binom{K+1}{n} \tag{3.5}$$

coefficients. The number L of coefficients may be reduced should we perform a sub-sampling operation by 2^{K+1} on the signals in the set (3.4). In such a case, L_1 and L_2 in (3.5) have to be replaced by $L_1/2^{K+1}$ and $L_2/2^{K+1}$. The sub-sampling operation does not induce a loss of information because the frequency content of the scattering signals (or images) lies in the circle of radius $\pi/2^{K+1}$. Figures 3.2, 3.3 and 3.4 depict the discrete scattering transform of three signals, a signal extracted from a region with sand ripples, a signal extracted from a region with seaweed, and a signal extracted from a flat portion of the seabed. The size of the signals is $L_1 \times L_2 = 32 \times 32$. We used $K = 2$, $N = 2$ and $M = 6$ to calculate the discrete scattering transform of the signals. We then find that

$$\sum_{n=0}^{N} M^n \binom{K+1}{n} = 1 + M \times (K+1) + M^2 \times \frac{(K+1)K}{2} = 127 . \tag{3.6}$$

The representation of each signal is therefore made of $L = 127 \times 4 \times 4 = 2\,032$ coefficients after a sub-sampling operation by $2^{K+1} = 8$. Had the sub-sampling operation not been performed, the representation of each signal would have been made of $L = 127 \times 32 \times 32 = 130\,048$ coefficients. From a computational perspective, there is therefore a clear advantage in performing a sub-sampling operation. The representation of each signal was normalised so that its energy (that of the representation and not the signal) is equal to 1. The discrete scattering transform $\bar{S}$ transforms the space of the signals which has $L_1 L_2$ dimensions into the space of the scattering transforms of the signals which has L dimensions. We are considering an inference problem with Q classes $\mathcal{C}_q$ represented by random variables $C_q(x)$ whose realisations are rectangular signals $c_q(x)$. A solution to this inference problem (not fully stated yet) is found in the space of the scattering transform of the signals. Each class $\mathcal{C}_q$ is described by an affine space of low dimension (with respect to L) constructed out of the Karhunen-Loève decomposition of the discrete scattering transform $\bar{S}C_q(x)$ of the random variable $C_q(x)$.

3.1.3 Karhunen-Loève decomposition

The discrete scattering transform $\bar{S}C_q(x)$ of the real-valued random variable $C_q(x)$ is a also real-valued random variable. The dimension of $\bar{S}C_q(x)$ is L whereas the

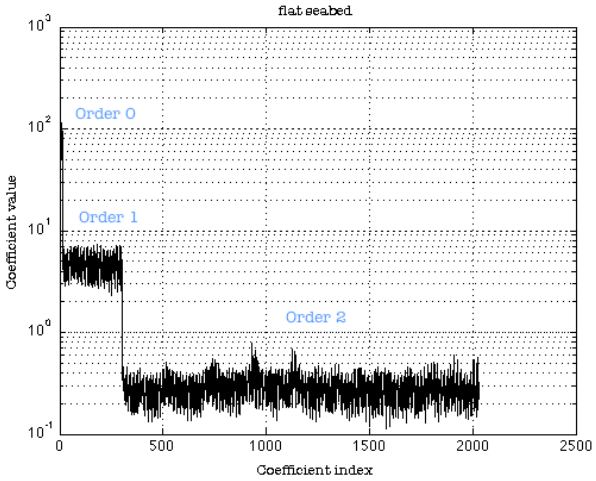


Figure 3.2. Discrete scattering transform of a signal extracted from a portion of flat seabed. The size of the signal is $L_1 \times L_2 = 32 \times 32$. We used $K = 2$, $N = 2$ and $G = 6$ to calculate the discrete scattering transform.

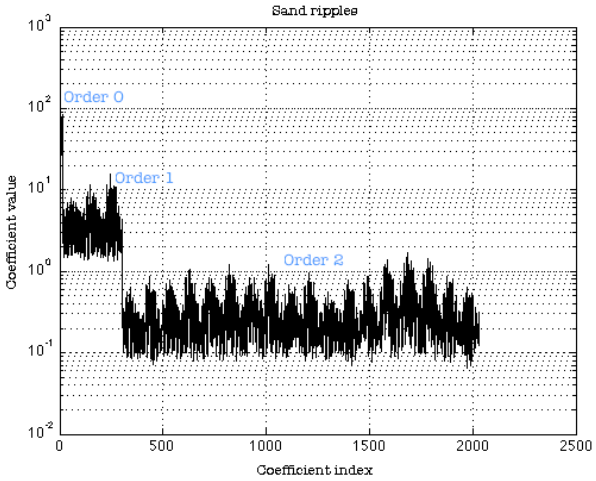


Figure 3.3. Discrete scattering transform of a signal extracted from a region with sand ripples. The size of the signal is $L_1 \times L_2 = 32 \times 32$. We used $K = 2$, $N = 2$ and $G = 6$ to calculate the discrete scattering transform.

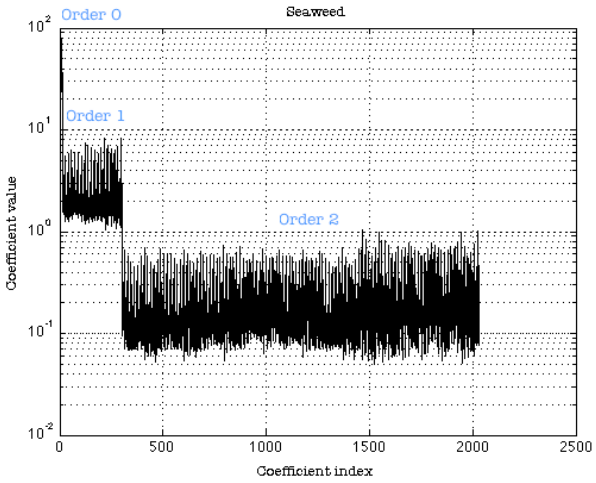


Figure 3.4. Discrete scattering transform of a signal extracted from a region with seaweed. The size of the signal is $L_1 \times L_2 = 32 \times 32$. We used $K = 2$, $N = 2$ and $G = 6$ to calculate the discrete scattering transform.

dimension of $C_q(x)$ is $L_1 L_2$. The mean M_q and covariance matrix Σ_q of $\bar{S}C_q(x)$ are

$$\begin{cases} M_q = \mathbb{E}[\,\bar{S}C_q(x)\,] \\ \Sigma_q = \mathbb{E}[\,(\bar{S}C_q(x) - M_q)\,(\bar{S}C_q(x) - M_q)^T\,] \end{cases} \tag{3.7}$$

where $\mathbb{E}$ represents the statistical expectation operator, and where the superscript T represents the matrix transposition operator. We look at $\bar{S}C_q(x) - M_q$, which appears in the expression of the covariance matrix Σ_q, as a random variable whose realisations belong to the $\mathbb{R}$-linear space $\mathbb{R}^L$ equipped with the scalar product $\langle \cdot, \cdot \rangle$ defined by $\langle v, w \rangle = \sum_{l=1}^{L} v_l w_l$ for any elements $v = [\,v_1 \cdots v_L\,]^T$ and $w = [\,w_1 \cdots w_L\,]^T$ of $\mathbb{R}^L$. In matrix form, $\langle v, w \rangle = v^T w = w^T v$. Two elements v and w of $\mathbb{R}^L$ are orthogonal if $\langle v, w \rangle = 0$. The norm $\|v\|$ of an element v of $\mathbb{R}^L$ is defined by $\|v\| = \langle v, v \rangle^{1/2}$. The L complex roots $\rho_{q,1}, \ldots, \rho_{q,L}$ of the characteristic polynomial of Σ_q are the eigenvalues of Σ_q [99]. An element v of $\mathbb{R}^L$ is an eigenvector of Σ_q associated to the eigenvalue $\rho_{q,l}$ if

$$\Sigma_q\, v = \rho_{q,l}\, v\ . \tag{3.8}$$

The eigenvectors of Σ_q are not unique. The eigenvalues of Σ_q are not unique either. Σ_q is symmetric and real from which we know [99] that the eigenvalues of Σ_q are all real. We also know [99] that there exists a set $\mathcal{B}_q = (\xi_{q,1}, \ldots, \xi_{q,L})$ of eigenvectors of Σ_q which forms an orthonormal basis of $\mathbb{R}^L$. The eigenvector $\xi_{q,l}$ is associated with the eigenvalue $\rho_{q,l}$ for all indexes l. We have therefore

$$\Sigma_q\, \xi_{q,l} = \rho_{q,l}\, \xi_{q,l} \tag{3.9}$$

and

$$\langle \xi_{q,l}, \xi_{q,l'} \rangle = \begin{cases} 1 & \text{if} \quad l = l' \\ 0 & \text{if} \quad l \neq l' \end{cases} . \tag{3.10}$$

An element v of $\mathbb{R}^L$ can be decomposed onto $\mathcal{B}_q$ as follows.

$$v = \sum_{l=1}^{L} \langle v, \xi_{q,l} \rangle \, \xi_{q,l} \tag{3.11}$$

where $\langle v, \xi_{q,l} \rangle$ are the coordinates of v in $\mathcal{B}_q$. The decomposition of the random variable $\bar{S}C_q(x) - M_q$ onto $\mathcal{B}_q$ is

$$\bar{S}C_q(x) - M_q = \sum_{l=1}^{L} \underbrace{\langle \bar{S}C_q(x) - M_q, \xi_{q,l} \rangle}_{\zeta_{q,l}} \, \xi_{q,l} \tag{3.12}$$

which leads to the Karhunen-Loève decomposition of $\bar{S}C_q(x)$, namely

$$\bar{S}C_q(x) = M_q + \sum_{l=1}^{L} \underbrace{\langle \bar{S}C_q(x) - M_q, \xi_{q,l} \rangle}_{\zeta_{q,l}} \, \xi_{q,l} \, . \tag{3.13}$$

The coordinates $\zeta_{q,l}$ of $\bar{S}C_q(x) - M_q$ in $\mathcal{B}_q$ are one-dimensional real-valued random variables. It can be shown in [70] that the coordinates are decorrelated, namely $\mathbb{E}[\, \zeta_{q,l} \zeta_{q,l'} \,] = 0$ if $l \neq l'$. It can also be shown in [70] that $\mathbb{E}[\, \zeta_{q,l} \,] = 0$ and $\mathbb{E}[\, \zeta_{q,l}^2 \,] = \rho_{q,l}$. The Karhunen-Loève decomposition of $\bar{S}C_q(x)$ has a geometrical interpretation. The realisations of the random variable $\bar{S}C_q(x)$ generate a cloud of points in an L-dimensional affine space[1]. The cloud of points is centred centred around the point M_q. The shape of the cloud of points (relatively to M_q) is characterised by the statistical distribution of $\bar{S}C_q(x) - M_q$. The elements $\xi_{q,l}$ of $\mathcal{B}_q$ give the directions of the principal axes of the cloud. Large eigenvalues $\rho_{q,l}$ correspond to the directions along which the cloud is highly elongated.

3.1.4 Approximation with affine spaces

The exact decomposition of $\bar{S}C_q(x) - M_q$ onto $\mathcal{B}_q$ is given by (3.12). An approximation of $\bar{S}C_q(x) - M_q$ can be obtained when considering, in the decomposition, some of the elements of $\mathcal{B}_q$, instead of all of them. To be more precise, let $\mathbf{V}_{q,l}$ be the linear space spanned by the first l elements of $\mathcal{B}_q$ ($1 \leq l < L$) where the elements of $\mathcal{B}_q$ are listed in eigenvalue-decreasing order. The approximation of $\bar{S}C_q(x) - M_q$ by $\mathbf{V}_{q,l}$ is defined as the orthogonal projection $P_{\mathbf{V}_{q,l}}(\bar{S}C_q(x) - M_q)$

[1] An affine space is loosely speaking a linear space "attached" to a point. Elements of an affine space are called points whereas elements of a linear space are called vectors. The difference between two points in affine space creates a vector which belongs to the linear space the affine space is built upon.

of $\bar{S}C_q(x) - M_q$ onto $\mathbf{V}_{q,l}$. We have

$$\bar{S}C_q(x) - M_q \simeq P_{\mathbf{V}_{q,l}}(\,\bar{S}C_q(x) - M_q\,) = \sum_{l'=1}^{l} \zeta_{q,l'}\,\xi_{q,l'}\ . \tag{3.14}$$

The corresponding expected quadratic error is

$$\begin{aligned}\mathbb{E}\left[\,\|\,\bar{S}C_q(x) - M_q - P_{\mathbf{V}_{q,l}}(\,\bar{S}C_q(x) - M_q\,)\,\|^2\,\right] \\ = \mathbb{E}\left[\,\|\sum_{l'=l+1}^{L}\zeta_{q,l'}\,\xi_{q,l'}\,\|^2\right] = \mathbb{E}\left[\sum_{l'=l+1}^{L}\zeta_{q,l'}^2\right] \\ = \sum_{l'=l+1}^{L}\mathbb{E}\left[\,\zeta_{q,l'}^2\right] = \sum_{l'=l+1}^{L}\rho_{q,l'}\ .\end{aligned} \tag{3.15}$$

It can be shown in [70] that for any dimension l such that $1 \leq l < L$, $\mathbf{V}_{q,l}$ is the linear space of dimension l which leads to the smallest expected quadratic error. To be more precise, if $\mathcal{B}$ is another orthonormal basis of $\mathbb{R}^L$, $\mathbf{U}_l$ the linear space spanned by the first l elements of $\mathcal{B}$, and $P_{\mathbf{U}_l}$ the orthogonal projection onto $\mathbf{U}_l$,

$$\begin{aligned}\mathbb{E}\left[\,\|\,\bar{S}C_q(x) - M_q - P_{\mathbf{U}_l}(\,\bar{S}C_q(x) - M_q\,)\,\|^2\,\right] \geq \\ \mathbb{E}\left[\,\|\,\bar{S}C_q(x) - M_q - P_{\mathbf{V}_{q,l}}(\,\bar{S}C_q(x) - M_q\,)\,\|^2\,\right]\ .\end{aligned} \tag{3.16}$$

The approximation obtained out of the Karhunen-Loève decomposition is therefore optimal in the expected quadratic error sense. Approximating $\bar{S}C_q(x) - M_q$ by the linear space $\mathbf{V}_{q,l}$ is equivalent to approximating $\bar{S}C_q(x)$ by the affine space

$$\mathbf{A}_{q,l} = M_q + \mathbf{V}_{q,l}\ . \tag{3.17}$$

The approximation of $\bar{S}C_q(x)$ by $\mathbf{A}_{q,l}$ is defined as the orthogonal projection of $P_{\mathbf{A}_{q,l}}(\,\bar{S}C_q(x)\,)$ of $\bar{S}C_q(x)$ onto $\mathbf{A}_{q,l}$. We have

$$\bar{S}C_q(x) \simeq P_{\mathbf{A}_{q,l}}(\,\bar{S}C_q(x)\,) = M_q + P_{\mathbf{V}_{q,l}}(\,\bar{S}C_q(x) - M_q\,) \tag{3.18}$$

$$= M_q + \sum_{l'=1}^{l}\zeta_{q,l'}\,\xi_{q,l'} \tag{3.19}$$

as illustrated in Figure 3.5. A measure of how coarse the approximation of $\bar{S}C_q(x)$ by $\mathbf{A}_{q,l}$ is can be obtained through an analysis of the distance $d_{\mathbf{A}_{q,l}}(\,\bar{S}C_q(x)\,)$ between $\bar{S}C_q(x)$ and the affine space $\mathbf{A}_{q,l}$. We have

$$\begin{aligned}d_{\mathbf{A}_{q,l}}(\,\bar{S}C_q(x)\,) = \|\,\bar{S}C_q(x) - P_{\mathbf{A}_{q,l}}(\,\bar{S}C_q(x)\,)\,\| \\ = \|\,\bar{S}C_q(x) - M_q - P_{\mathbf{V}_{q,l}}(\,\bar{S}C_q(x) - M_q\,)\,\| \\ = \|\sum_{l'=l+1}^{L}\zeta_{q,l'}\,\xi_{q,l'}\,\| = \left[\sum_{l'=l+1}^{L}\zeta_{q,l'}^2\right]^{1/2}\ .\end{aligned} \tag{3.20}$$

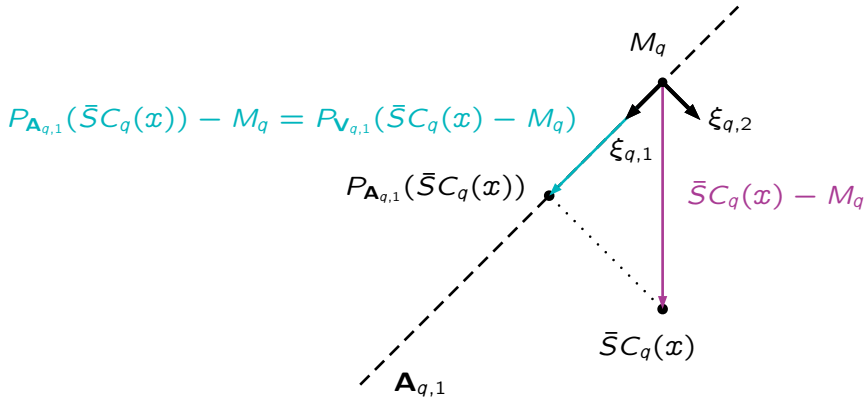


Figure 3.5. Linear approximation of the random variable $\bar{S}C_q(x)$ by the affine space $\mathbf{A}_{q,1} = M_q + \mathbf{V}_{q,1}$ of dimension 1 where $\mathbf{V}_{q,1}$ is the linear space spanned by the first eigenvector $\xi_{q,1}$ of Σ_q. In the figure, $L = 2$ and the two eigenvectors $\xi_{q,1}$ and $\xi_{q,2}$ of Σ_q form an orthonormal basis of $\mathbb{R}^L = \mathbb{R}^2$.

This distance is a one-dimensional real-valued positive random variable for the coordinates $\zeta_{q,l}$ are themselves one-dimensional real-valued random variables. Upon combining $\mathbb{E}[\,\zeta_{q,l}^2\,] = \rho_{q,l}$ and Markov inequality [100] applied to (3.20), we find that

$$\mathbb{P}(\, d_{\mathbf{A}_{q,l}}(\,\bar{S}C_{q,l}(x)\,) \geq a\,) \leq \frac{1}{a^2} \times \mathbb{E}[\, d_{\mathbf{A}_{q,l}}(\,\bar{S}C_{q,l}(x)\,)^2\,] = \frac{1}{a^2} \times \sum_{l'=l+1}^{L} \rho_{q,l'} \quad (3.21)$$

for any positive non-zero real number a. In (3.21) $\mathbb{P}(\, d_{\mathbf{A}_{q,l}}(\,\bar{S}C_{q,l}(x)\,) \geq a\,)$ is the probability that $d_{\mathbf{A}_{q,l}}(\,\bar{S}C_{q,l}(x)\,)$ is greater than a. Markov inequality (3.21) states therefore that for at most ε % of the realisations $c_q(x)$ of $C_q(x)$ (for example 5 % if $\varepsilon = 0.05$) the distance $d_{\mathbf{A}_{q,l}}(\,\bar{S}c_q(x)\,)$ between $\bar{S}c_q(x)$ and the affine space $\mathbf{A}_{q,l}$ is greater than

$$\tau_{q,l}\,(\varepsilon) = \left[\, \frac{1}{\varepsilon} \times \sum_{l'=l+1}^{L} \rho_{q,l'} \right]^{1/2} . \quad (3.22)$$

In other words, for at least $(1-\varepsilon)$ % of the realisations $c_q(x)$ of $C_q(x)$ (for example 95 % if $\varepsilon = 0.05$) the distance $d_{\mathbf{A}_{q,l}}(\,\bar{S}c_q(x)\,)$ between $\bar{S}c_q(x)$ and the affine space $\mathbf{A}_{q,l}$ is smaller than $\tau_{q,l}\,(\varepsilon)$, as illustrated in Figure 3.6. At this point, we can outline the behaviour of $\tau_{q,l}\,(\varepsilon)$ with respect to l. We know from Section 3.1.3 that $\mathbb{E}[\,\zeta_{q,l}^2\,] = \rho_{q,l}$ from which we can conclude that the eigenvalues $\rho_{q,l}$ are all positive. The eigenvalues are also listed in order of decreasing amplitude. As a result, $l \mapsto \tau_{q,l}\,(\varepsilon)$ is a decreasing function. This has a geometrical interpretation. An affine space $\mathbf{A}_{q,l}$ of low dimension l coarsely approximates $\bar{S}C_q(x)$ and, loosely speaking, many realisations of $\bar{S}C_q(x)$ are far away from $\mathbf{A}_{q,l}$. An affine space $\mathbf{A}_{q,l'}$ of dimension $l' > l$ approximates more accurately $\bar{S}C_q(x)$ than $\mathbf{A}_{q,l}$ does.

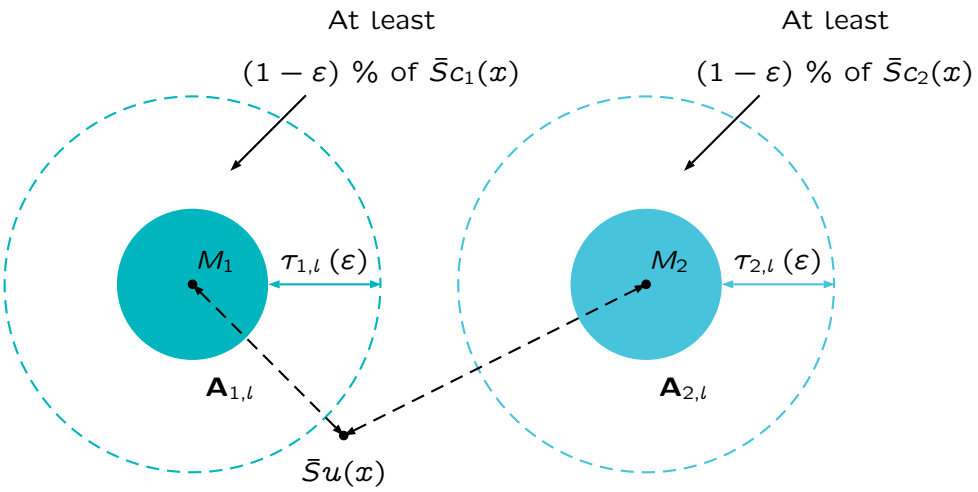


Figure 3.6. Statistical analysis of the distance to the affine spaces.

Comparatively fewer realisations of $\bar{S}C_q(x)$ lie far away from $\mathbf{A}_{q,l'}$, leading to $\tau_{q,l'}(\varepsilon) < \tau_{q,l}(\varepsilon)$ with ε remaining constant. Bearing in mind that each class $\mathcal{C}_q$ corresponds to a type of seabed, and should ε be close to 0, $\tau_{q,l}(\varepsilon)$ can be used to decide whether a test signal $u(x)$ (which is not necessarily a realisation $c_q(x)$ of a class $\mathcal{C}_q$) belongs to any of the types of seabed. For example, if $\varepsilon = 0.05$, we know that 95 % of the realisations of $\bar{S}C_q(x)$ lie at most $\tau_{q,l}(0.05)$ away from $\mathbf{A}_{q,l}$. If the distance between $\bar{S}u(x)$ and $\mathbf{A}_{q,l}$ is greater than $\tau_{q,l}(0.05)$ for all class indexes q, we can conclude that the test signal does not belong to any of the types of seabed.

3.1.5 Analysis of signals

We have thus far considered an inference problem with $Q \geq 1$ classes $\mathcal{C}_q$ where each class corresponds to a type of seabed, and presented how to describe the classes $\mathcal{C}_q$ with affine spaces $\mathbf{A}_{q,l}$ of low dimension l. We have yet to fully state what is the inference problem at stake. Let $u(x)$ be a rectangular signal extracted from one of the sonar images under study as in Figure 3.1. The distance between the discrete scattering transform $\bar{S}u(x)$ of the signal $u(x)$ and the affine space $\mathbf{A}_{q,l}$ is

$$d_{\mathbf{A}_{q,l}}(\,\bar{S}u(x)\,) = \|\,\bar{S}u(x) - P_{\mathbf{A}_{q,l}}(\,\bar{S}u(x)\,)\,\| \tag{3.23}$$

$$= \Big[\sum_{l'=l+1}^{L} \langle\,\bar{S}u(x) - M_q, \xi_{q,l'}\,\rangle^2\Big]^{1/2}. \tag{3.24}$$

We consider two inference problems.

1. In the first inference problem, the objective is to make a decision upon the type of seabed the signal belongs to, that is to say, to make a decision upon the class $\mathcal{C}_q$ the signal belongs to. This is a classification problem and we

K	Depth of the wavelet transform
M	number of orientations in G_+
N	Maximum length of the paths
l	Dimension of the affine spaces $\mathbf{A}_{q,l}$
$L_1 L_2$	Size of the signal $u(x)$

Table 3.1. Parameters the classification algorithm depends upon.

K	Depth of the wavelet transform
M	number of orientations in G_+
N	Maximum length of the paths
l	Dimension of the affine spaces $\mathbf{A}_{q,l}$
$L_1 L_2$	Size of the signal $u(x)$
$1-\varepsilon$	Percentage of the data described

Table 3.2. Parameters the anomaly detection algorithm depends upon.

have to have $Q > 1$. The estimated class index of the signal $u(x)$ is the index of the affine space which minimises $d_{\mathbf{A}_{q,l}}(\bar{S}u(x))$, namely

$$\text{estimated class index of } u(x) = \arg\min_q d_{\mathbf{A}_{q,l}}(\bar{S}u(x)) \ . \tag{3.25}$$

2. In the second inference problem, the objective is to decide whether the signal is part of the seabed, that is to say, to decide whether the signal belongs to any of the classes $\mathcal{C}_q$. This is an anomaly detection problem and $Q \geq 1$ (as opposed to $Q > 1$). The signal is declared normal if it is part of the seabed, and abnormal otherwise. To be more precise, the signal $u(x)$ is declared abnormal if for all class indexes q the distance between the discrete scattering transform $\bar{S}u(x)$ of the signal $u(x)$ and the affine space $\mathbf{A}_{q,l}$ is greater than $\tau_{q,l}(\varepsilon)$, namely

$$\text{decision upon } u(x) = \begin{cases} \text{abnormal if } d_{\mathbf{A}_{q,l}}(\bar{S}u(x)) > \tau_{q,l}(\varepsilon) \quad \text{for all} \quad q \\ \text{normal otherwise .} \end{cases} \tag{3.26}$$

The parameters the classification algorithm depends upon are listed in table 3.1. The parameters the anomaly detection algorithm depends upon are listed in table 3.2. In Figure 3.6, the signal belongs to $\mathcal{C}_1$ (inference problem 1) but is anomaly with respect to the two classes $\mathcal{C}_1$ and $\mathcal{C}_2$ (inference problem 2).

3.2 Supervised seabed identification

3.2.1 Seabed identification related work

Most of the work carried on either the supervised or unsupervised identification of the types of seabed commonly present in side-looking sonar images is based on texture descriptors. We opted for a rather chronological presentation of the work that was carried on this subject in the past. The types of seabed are described by their fractal dimension in [73] and identified with Fisher's linear and quadratic discriminants. Fractal analysis and spatial point processes are used in [101]. A spatial point process model is also used in [64] to describe the types of seabed. The types of seabed are described by a set of features derived from grey level co-occurrence distributions in [102] and identified with a hybrid neural network which consists of a supervised network driven by an unsupervised one. Spectral features and grey level run-length features are used in [103] along with a back-propagation neural network to identify the types of seabed. Kohonen self-organising maps and a set of features derived from grey level co-occurrence distributions are introduced in [104] as two means of describing the types of seabed. An unsupervised identification scheme that relies on another Kohonen map is further employed.

A more recent approach is the following. The types of seabed are identified in [62] with a statistical analysis of the similarity between grey level co-occurrence distributions. Such an approach differs from some of the aforementioned ones because it does not rely on a set of features derived from co-occurrence distributions, but rather on the co-occurrences distributions themselves. The similarity between the distributions is measured by the Kullback-Leibler divergence. The work presented in [62] is generalised in [105] in order to combine several texture descriptors by way of a weighted sum of Kullback-Leibler divergences. The considered texture descriptors are the sonar image grey levels, and descriptors which come from Gabor and wavelet image expansions.

We also ought to mention seabed segmentation methods with side-looking sonar images as well as seabed identification methods. Active contours are used for seabed segmentation in [55] together with a set of features derived from grey level co-occurrence distributions [60]. A fast implementation of the set of features which uses sum and difference histograms is used [59]. This approach can only handle two types of seabed. Two seabed segmentation approaches are studied in [63]. The first approach is a Bayesian approach that relies on the maximum marginal probability criterion, as opposed to the more traditional maximum a posteriori criterion. The second approach is based on active contours, and can handle more than two types of seabed. The two seabed segmentation approaches rely of the weighted sum of Kullback-Liebler divergences presented in [105] as well as on the same set of texture descriptors.

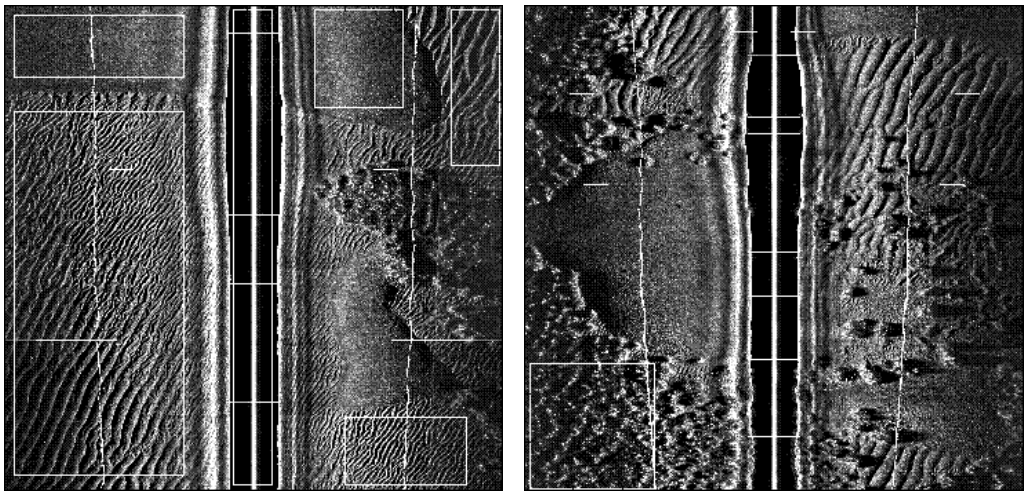


Figure 3.7. Training regions used for the supervised identification of the types of seabeds. The training regions are the white rectangular areas.

3.2.2 Presentation of the data set

The data set under study is composed of two hundred and sixty side-looking sonar images with three types of seabed. The dataset was obtained from a survey of the sea bottom off the coast of Italy by an autonomous underwater vehicle. The types seabed present in the data set are flat sandy areas made of a given type of sediments, areas with sand ripples of various widths and shapes, and areas with seaweed. Each type of seabed corresponds to a class $\mathcal{C}_q$. The identification of the types of seabed was performed with the help of the first inference algorithm (see Section 3.1.5). Each class $\mathcal{C}_q$ is described by affine space $\mathbf{A}_{q,l}$ of low dimension l constructed out of a set of training signals $c_q(x)$. The Karhunen-Loève decomposition was performed in C with the help of the LAPACK library [106]. The training regions, from which the training signals are extracted, are the white rectangular areas depicted on Figure 3.7. It is important to notice that the training set is very small compared to the entire data set. It is possible to use such a small training set because of the stability of scattering operators with respect to deformations.

3.2.3 Quantitative and qualitative results

The performance of the classification algorithm was analysed in details over various sets of parameters (see Table 3.1). We set $M = 6$ which is a value commonly chosen for the wavelet transform of images [70]. M is the number of orientations of the wavelet transform. K varies from 2 to 5, N from 1 to 2, and l from 5 to 25. K and N are respectively the depth of the wavelet transform and the maximum length of the paths of the scattering transform. The size $L_1 \times L_2$ of the signals are 4×4, 8×8, 16×16 and 32×32. In order to quantitatively evaluate the performance of the algorithm, six out of the two hundred and sixty images were carefully manually ground truthed. Those six images constitute the testing set. The error rates we obtained for each set of parameters are given through Figures 3.8 to 3.11. The error rates are also given in Tables 3.3 to 3.6. The following

patterns of behaviours are observed.

1. For a given dimension l of the affine spaces, the error rate decreases with an increase of the depth K of the wavelet transform and the maximum order of interferences N. Such a pattern is to be expected. When K increases, the spatial variability of the scattering images is reduced because of the smoothing effect of the low-pass filter. The seabed representation becomes more stable, which results in a decrease of the error rate. When K increases, the spatial variability of the scattering images is reduced, which implies that some information is lost. Such a loss is compensated for when the maximum length N of the paths is increased.

2. For a given depth K of the wavelet transform and a given order N of interferences, the error rate either increases or decreases with an increase of the dimension l of the affine spaces. Affine spaces of low dimension provide a coarse representation of the types of seabed. The representation improves as the dimension increases, which results in a decrease of the error rate. When the dimension increases further more, the affine spaces begin to merge, which results in an increase of the error rate. We can therefore assume that there exist a *critical dimension* below which the error rate decreases as the dimension increases, and beyond which the error rate increases as the dimension increases. The classification results can all be interpreted in terms of this critical dimension. For example, for signals of size 4×4 and $(K, N) = (3, 1)$, the error rate increases when the dimension l increases. We can therefore conclude that the critical dimension is below 5. For signals of size 32×32 and $(K, N) = (2, 2)$, the error rate decreases when the dimension l increases. We can therefore conclude that the critical dimension is above 25. The critical dimension provides an estimate of the compactness of the representations of the types of seabed *with respect to each other*.

We have thus far looked at the behaviour of the seabed identification algorithm with respect to the depth of the Littlewood Paley wavelet transform, the maximum length of the paths of the scattering transform, and the dimension of the affine spaces. Let us now look at performance of the seabed identification algorithm. A minimal error rate of 3.59 % was reached when $K = 3$, $N = 2$, $l = 15$ and with signals of size 32×32, which corresponds to an overall excellent characterisation of the types of seabed. The seabed characterisation is all the more excellent should we account for the extreme small size of the training set. The training set is indeed made of portions of only two side-looking sonar images, which demonstrates the generalisation capabilities of the signal representation induced by scattering operators. The confusion matrix which corresponds to the minimal error rate of 3.59 % is given in Table 3.7. The columns of the confusion matrix correspond to the estimated types of seabed, the lines, to the true types of seabed. As a result, the lines of the confusion matrix sum up to one, up to minor round-off errors. Here are two examples of how to read the confusion matrix. The probability that the estimated type of a signal is ripples whereas its true type is seaweed equals 0.0255. The probability that the estimated type of a signal is flat whereas its true

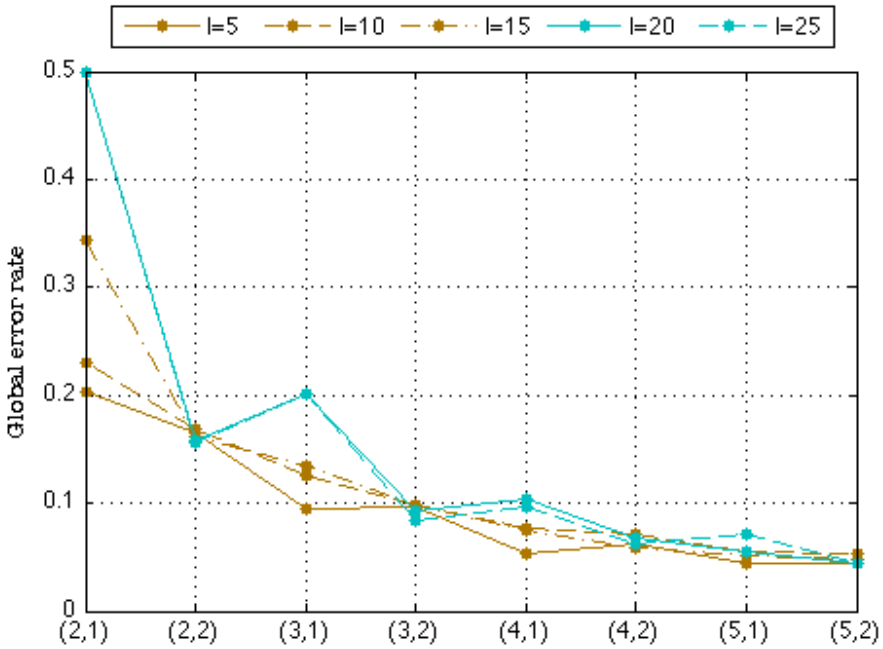


Figure 3.8. Global error rate obtained with signals of size 4×4*. The various pairs* (K, N) *are indicated at the bottom. The various dimensions* l *are indicated at the top.*

type is seaweed equals 0.0498. The six sonar images that compose the testing set are depicted in Figures 3.12 and 3.13 along with the classification maps which correspond to the minimal error rate of 3.59 %. On Figures 3.12 and 3.13 we observe that some of the wrongly classified regions correspond to objects on the sea bottom, which could be because the objects induce a local perturbation of the texture field. This is one fact in favour of the hypothesis stated in Chapter 1 according to which objects can be revealed as local anomalies with respect to the sea bottom.

To obtain an excellent seabed characterisation with one set of parameters is one thing. It is even better is the seabed characterisation remains as good across a wide range of parameters. From Figures 3.8 to 3.11 we observe that characterisation of the seabed remains excellent across a wide range of parameters. The seabed characterisation only deteriorates when the chosen parameters are unreasonable. By an unreasonable choice of parameters, we mean a very low affine space dimension, a Littlewood Paley wavelet transform of limited depth, and signals of very small sizes. The fact that the characterisation of the seabed remains excellent across a wide range of parameters is, to our opinion, something in favour of the seabed identification algorithm we presented in this chapter. From our experience, it is not often the case that an inference algorithm is not very sensitive to variations in the parameters it is based upon.

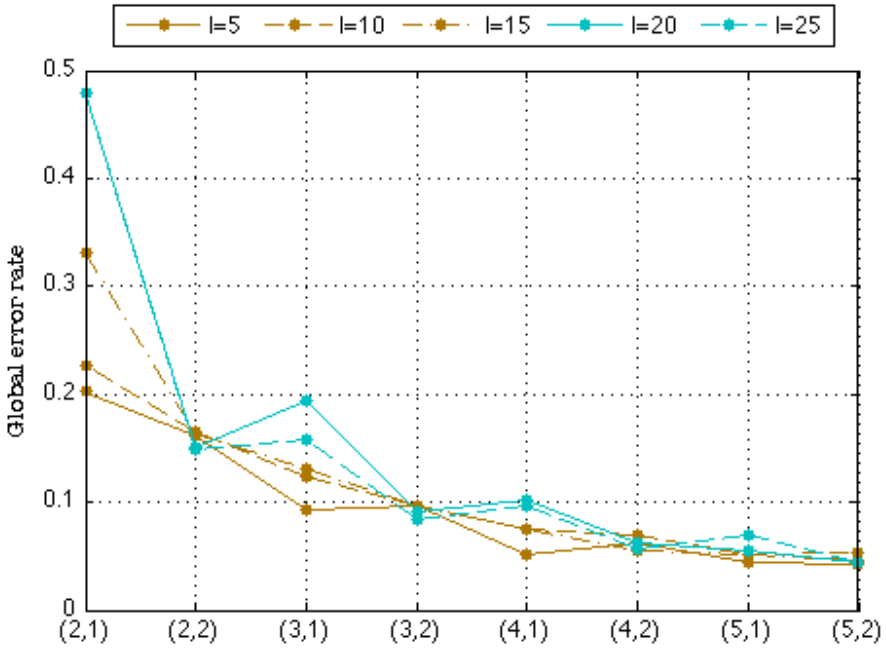


Figure 3.9. Global error rate obtained with signals of size 8×8. *The various pairs* (K, N) *are indicated at the bottom. The various dimensions* l *are indicated at the top.*

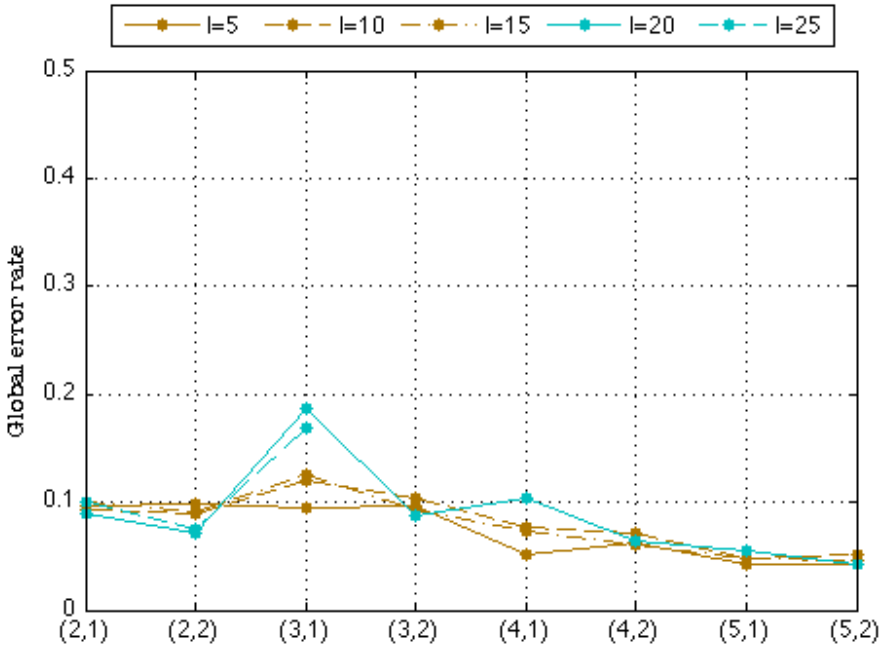


Figure 3.10. Global error rate obtained with signals of size 16×16. *The various pairs* (K, N) *are indicated at the bottom. The various dimensions* l *are indicated at the top.*

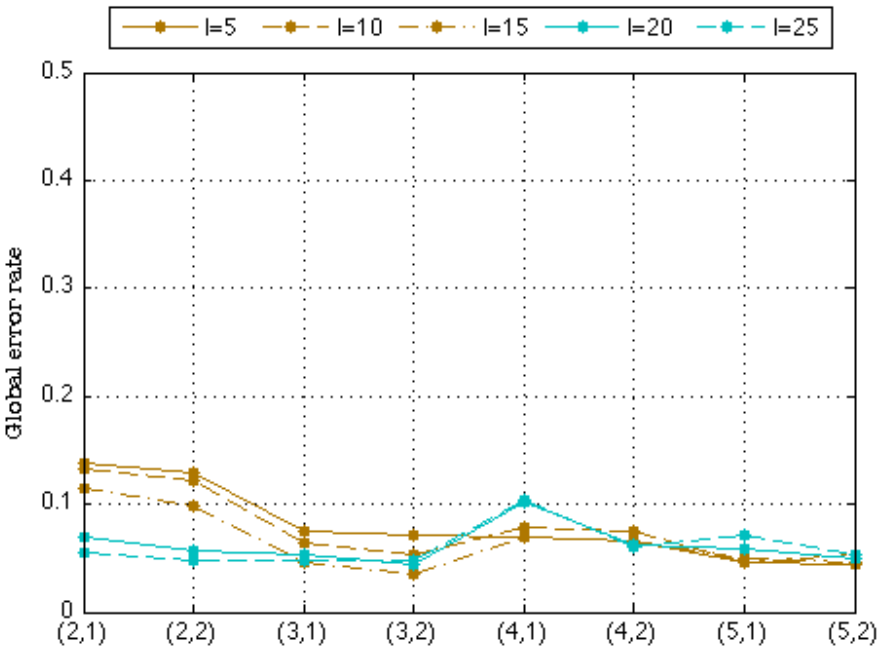


Figure 3.11. Global error rate obtained with signals of size 32×32. The various pairs (K, N) are indicated at the bottom. The various dimensions l are indicated at the top.

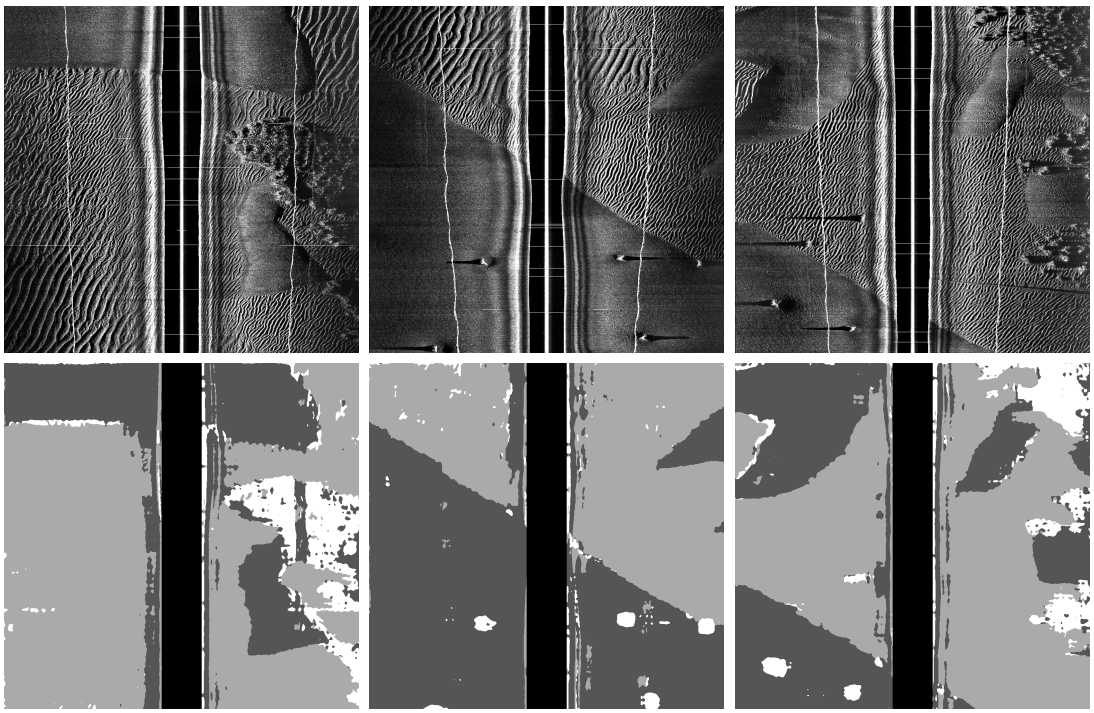


Figure 3.12. Top. Three out of the six sonar images that compose the testing set. The images are of size 1000×1024 pixels. Bottom. Classification maps which correspond to the minimal error rate of 3.59 %. The colour coding is the following. Black corresponds to flat areas, grey to areas with sand ripples, and white to areas with seaweed.

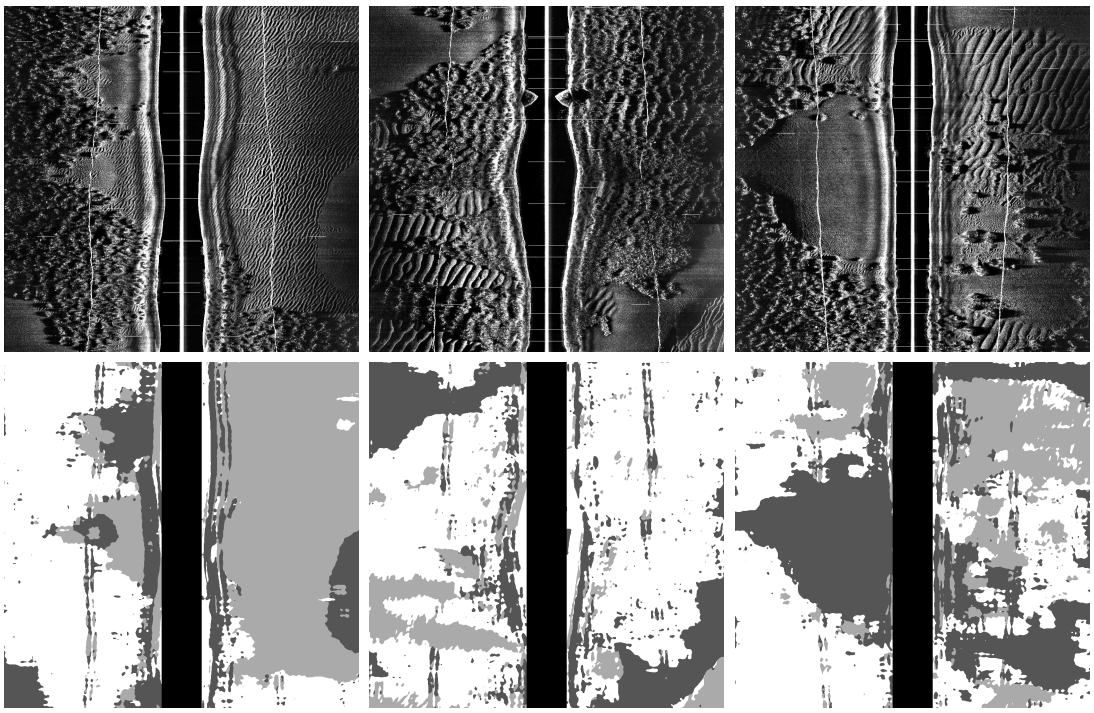


Figure 3.13. Top. Three out of the six sonar images that compose the testing set. The images are of size 1000 × 1024 pixels. Bottom. Classification maps which correspond to the minimal error rate of 3.59 %. The colour coding is the following. Black corresponds to flat areas, grey to areas with sand ripples, and white to areas with seaweed.

K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	5	5	5	5	5	5	5	5
Error rate	0.204	0.166	0.095	0.097	0.053	0.063	0.044	0.044
K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	10	10	10	10	10	10	10	10
Error rate	0.231	0.169	0.125	0.099	0.076	0.072	0.055	0.054
K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	15	15	15	15	15	15	15	15
Error rate	0.344	0.162	0.135	0.099	0.075	0.059	0.052	0.047
K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	20	20	20	20	20	20	20	20
Error rate	0.498	0.157	0.202	0.093	0.103	0.068	0.054	0.045
K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	25	25	25	25	25	25	25	25
Error rate	0.498	0.157	0.201	0.083	0.096	0.063	0.071	0.044

Table 3.3. Seabed classification performance – signal size 4×4

K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	5	5	5	5	5	5	5	5
Error rate	0.202	0.161	0.094	0.097	0.052	0.062	0.043	0.043
K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	10	10	10	10	10	10	10	10
Error rate	0.227	0.166	0.123	0.096	0.075	0.070	0.052	0.054
K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	15	15	15	15	15	15	15	15
Error rate	0.332	0.164	0.131	0.096	0.075	0.055	0.051	0.046
K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	20	20	20	20	20	20	20	20
Error rate	0.479	0.150	0.194	0.091	0.103	0.062	0.054	0.044
K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	25	25	25	25	25	25	25	25
Error rate	0.479	0.149	0.159	0.083	0.096	0.057	0.070	0.044

Table 3.4. Seabed classification performance – signal size 8×8

K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	5	5	5	5	5	5	5	5
Error rate	0.097	0.098	0.095	0.097	0.052	0.063	0.043	0.043
K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	10	10	10	10	10	10	10	10
Error rate	0.093	0.090	0.120	0.103	0.076	0.070	0.048	0.052
K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	15	15	15	15	15	15	15	15
Error rate	0.101	0.092	0.126	0.093	0.073	0.061	0.049	0.045
K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	20	20	20	20	20	20	20	20
Error rate	0.089	0.071	0.187	0.087	0.104	0.063	0.055	0.043
K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	25	25	25	25	25	25	25	25
Error rate	0.101	0.074	0.168	0.082	0.103	0.052	0.068	0.045

Table 3.5. Seabed classification performance – signal size 16×16

K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	5	5	5	5	5	5	5	5
Error rate	0.137	0.129	0.076	0.071	0.070	0.067	0.046	0.045
K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	10	10	10	10	10	10	10	10
Error rate	0.133	0.122	0.064	0.053	0.078	0.075	0.046	0.053
K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	15	15	15	15	15	15	15	15
Error rate	0.115	0.098	0.046	0.036	0.070	0.067	0.049	0.047
K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	20	20	20	20	20	20	20	20
Error rate	0.069	0.058	0.054	0.045	0.103	0.063	0.059	0.049
K	2	2	3	3	4	4	5	5
N	1	2	1	2	1	2	1	2
l	25	25	25	25	25	25	25	25
Error rate	0.056	0.047	0.047	0.047	0.104	0.060	0.070	0.053

Table 3.6. Seabed classification performance – signal size 32×32

3.2.4 Local binary patterns

The results of our classification algorithm were compared with another classification algorithm based on the signal representation induced by local binary patterns [68]. We will now explain briefly what is the signal representation induced by local binary patterns. Let us consider a pixel P with grey level $G(P)$ in one of the sonar images under study. Let us also consider a circle of radius r drawn around the pixel P, and n points $P_0, \ldots, P_{n-1}$ uniformly distributed on the circle, as illustrated in Figure 3.14. The grey level of the point P_i is $G(P_i)$. If one of the n points does not correspond to a pixel in the sonar image, the grey level of this point is obtained by interpolation. The LBP code of the pixel P is defined by

$$LBP_{r,n}(P) = \sum_{i=0}^{n-1} 2^i \times H\left[G(P_i) - G(P)\right] \tag{3.27}$$

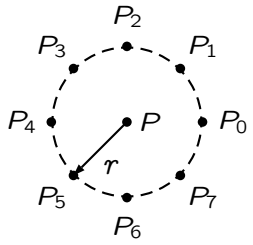
where H is the unit step function

$$H(x) = \begin{cases} 1 & \text{if} \quad x \geq 0 \\ 0 & \text{otherwise} . \end{cases} \tag{3.28}$$

Consequently, the LBP code of the pixel P is obtained by thresholding the neighbourhood of P, the result of which is considered as a binary number. LBP codes may be defined for various radii r and number of points n. Let us consider now a query signal, that is to say, a rectangular region in one of the sonar images under study. The query signal is made of a certain number of pixels for which LBP codes can be calculated. The representation of the query signal is defined as the histogram of the LBP codes of the pixels in the query signal. For the identification of the types of seabed, we used a combination of three LBP codes obtained with 8 points on a circle of radius 1, 16 points on a circle of radius 3, and 24 points on a circle of radius 5. We also used the same training and testing sets. We finally used Fisher's quadratic discriminant classifier. Other classifiers such as support vector machine were tested with very small differences in the results. Again other parameters combinations were used and we chose the optimal ones in terms of error rate. Signals of size 32×32 were chosen for optimal performance. An error rate of 13.07 % was obtained which is almost two and a half times larger than the minimal error rate obtained with the affine spaces. The confusion matrix for

	Flat	**Ripples**	**Seaweed**
Flat	0.9696	0.0066	0.0238
Ripples	0.0054	0.9831	0.0115
Seaweed	0.0498	0.0255	0.9247

Table 3.7. Confusion matrix that corresponds to the minimal error rate of 3.59% *obtained with* $K = 3$, $N = 2$, $l = 15$ *and signals of size* 32×32. *The columns correspond to the estimated class and the lines to the true class. For example, the probability that the estimated class of a signal is seaweed whereas its true class is flat equals 0.0238.*



$$LBP_{r,8}(P) = \sum_{i=0}^{7} \; 2^i \times H\,[\,G(P_i) - G(P)\,]$$

Figure 3.14. Local binary pattern of a pixel P with grey level $G(P)$.

	Flat	**Ripples**	**Seaweed**
Flat	0.7712	0.1626	0.0663
Ripples	0.0009	0.9787	0.0204
Seaweed	0.0192	0.1578	0.8230

Table 3.8. Confusion matrix obtained with local binary patterns. The columns correspond to the estimated class and the lines to the true class. For example, the probability that the estimated class of a signal is ripples whereas its true class is seaweed equals 0.1578.

the local binary pattern algorithm is given in Table 3.8.

3.3 Clutter rejection

3.3.1 Presentation of the problem

We consider, as a starting point, a set of detections provided by a target detector, as illustrated in Figure 3.15. The detections can be true or false positives. An analysis of this set of detections leads to a receiver operating characteristic (ROC) curve which represents the performance of the target detector alone. In order to reduce the number of false positives, within the initial set of detections, and hence improve the performance of the target detector, the similarity between the detections and the seabed is analysed with the help of the second inference algorithm (see Section 3.1.5). This is a post-processing process based on an anomaly detection algorithm, as illustrated in Figure 3.15. The detections are declared abnormal if they are not part of the seabed, and normal otherwise. It is implicitly assumed here that the true positives (the targets) are not as similar to the seabed as are the false positives, the false positives being by nature part of the seabed. The detections that are declared normal are discarded from the initial set of detections, leading to a second set of detections. An analysis of the second set of detections leads to a collection of receiver operating characteristic curves (one curve or more) which represents the performance of the combined target detector and post-processing process. It is also possible to analyse, from the second set of detections, the performance of the post-processing process alone

should we analyse the number of true positives that are declared normal, and the number of false positives that are declared abnormal. The issue here is to reduce the number of false positives without reducing the number of true positives. We will see in Section 3.3.3 that it is possible to reduce the number of false positives within the initial set of detections without reducing the number of true positives. Be that as it may, we will see in the same section that this does not improve the performance of the target detector, which suggests that the false positives that are discarded by the post-processing process are the ones already discarded by the target detector. The details of the post-processing process are given in Figure 3.16. A detection corresponds to a rectangular signal in one of the sonar images under study. Signals of the same size as the detection are extracted from the portion of the sonar image around the detection, providing information about the local background. We assume that the background is made of more than one type of seabed each type being referred to as a class $\mathcal{C}_q$. The types of seabed are obtained by a K-means clustering algorithm in the space of the discrete scattering transform of the signals. The affine spaces $\mathbf{A}_{q,l}$ and the limits $\tau_{q,l}(\varepsilon)$ are obtained for each class $\mathcal{C}_q$ so as to decide whether the detection is part of the seabed, as illustrated in Figure 3.16. All computations were performed in C with the help of the LAPACK library [106].

3.3.2 Presentation of the data set

We used a set of 155 Marine Sonics side-looking sonar images with three types of seabed, flat sandy areas made of a given type of sediments, areas with sand ripples of various widths and shapes, and areas with seaweed. We used a set of 155 simulated targets. One target was inserted in each image according to [56]. In brief, the side-looking sonar images are used to obtain an estimate of the topography and reflectivity of the sea bottom, as well as an estimate of the antenna's beam pattern. This is an inversion process. The topography and reflectivity are locally changed where the targets are to be added. The locally changed topography and reflectivity of the sea bottom, combined with the estimated beam pattern, are used to create sonar images with simulated targets. The resolution of a pixel is equal to 12.0 cm $\times$ 5.8 cm (vertical resolution $\times$ horizontal resolution). We then find that the total area covered by the images is equal to 1.1 km^2. The automatic target detector we used is based on a cascade of boosted classifiers and Haar-like features [27]. We chose such a detector because it is one of the best target detectors available at the moment. The structure of the target detector is given in Figure 3.17. A set of Haar-like features are calculated for a query signal, that is to say, a region in one of the sonar images under study. Let us assume that the cascade is made of n classifiers $C_1, \ldots, C_n$ organised sequentially. A classifier C_k uses one Haar-like feature f_k. This feature is thresholded in order to accept of reject the query signal. If the query signal is accepted by the classifier C_k, the query signal is passed on to the following classifier C_{k+1}. This classifier uses another Haar-like feature f_{k+1} which is thresholded in order to accept or reject the query signal. A detection is a query signal which has been accepted by all

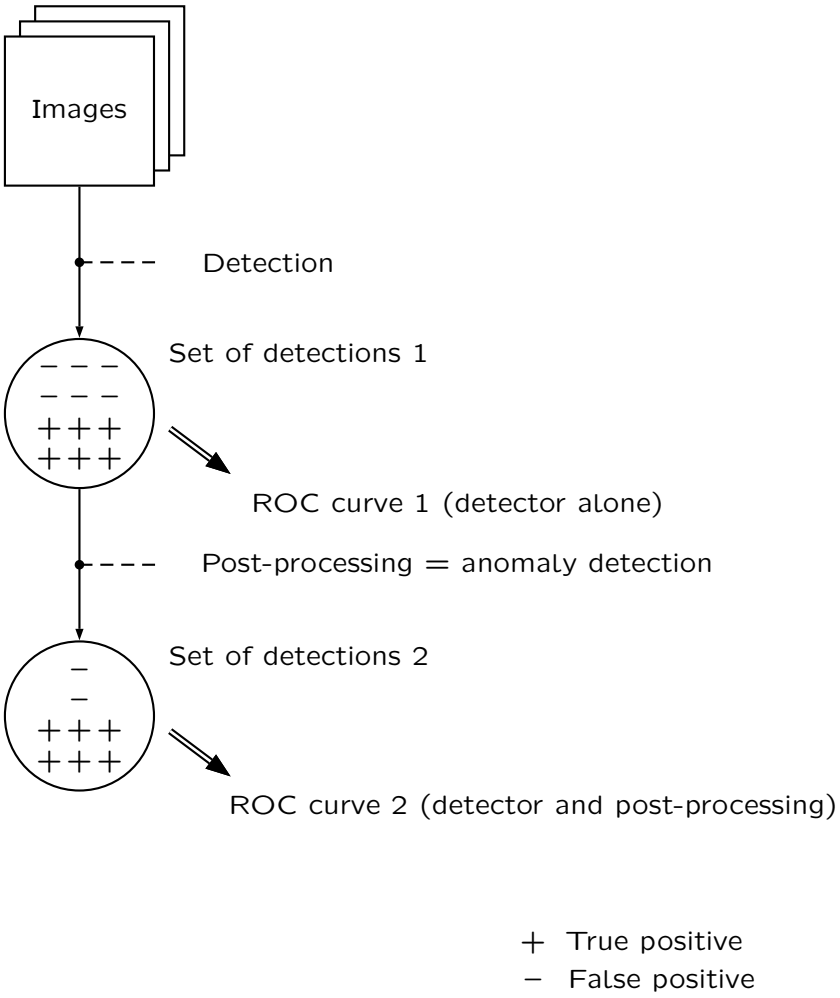


Figure 3.15. Automatic target detection and post-processing chain.

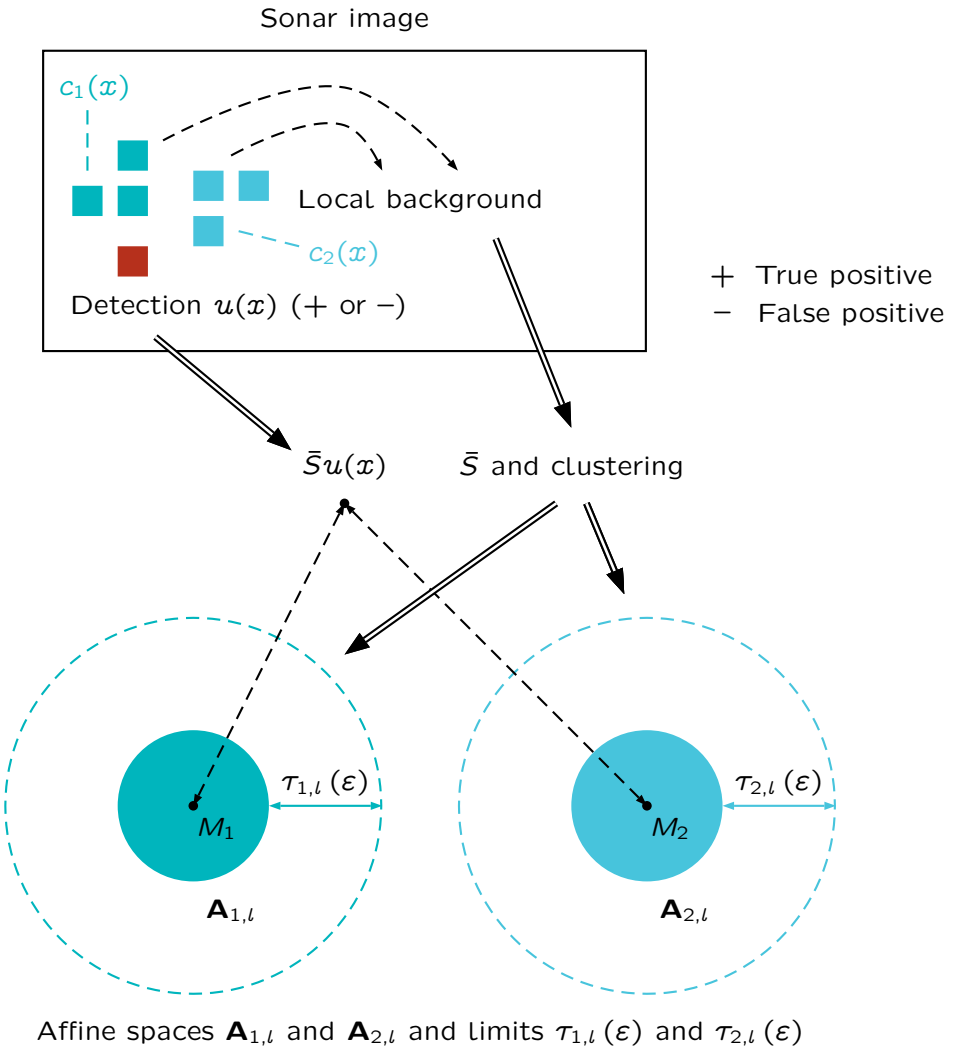


Figure 3.16. Details of the post-processing process.

Number of simulated targets	155
Number of detections	282
Number of true positives	128
Number of false positives	154
Number of false negatives	27

Table 3.9. Performance of the target detector alone

classifiers. The features $f_1, \ldots, f_n$ used by the classifiers and the corresponding thresholds are learnt during a training stage. A score between 0 and 1 is associated with each detection at the end of the cascade of classifiers. A detection with a score close to 1 is likely to be a target whereas a detection with a score close to 0 is likely to be a portion of the seabed. When thresholded at various points between 0 and 1, the score leads to the receiver operating characteristic curve of the target detector alone, which is given in Figure 3.18. The curve depicts the combined variation of the true positive rate and the number of false positives per square kilometre. The total number of detections provided by the target detector is 282 among which 128 detections are true positives and 154 detections are false positives — 27 targets are definitely lost, and will not be recovered by the post-processing process. We can conclude that the maximum true positive rate is equal to $128/155 \simeq 0.83$, which is the value of the true positive rate obtained when the score is thresholded at 0. We can similarly conclude that the maximum number of false positives per square kilometre is equal to $154/1.1 = 140$, which is the value of the number of false positives per square kilometre obtained when the score is thresholded at 0. Details about the set of detections produced by the target detector are given in Table 3.9. Examples of true positives are given in Figure 3.19. The true positives are either cylinders or truncated cones. Examples of false positives are given in Figure 3.20. We observe that some of the false positives are target-like objects whose signature is a highlight region next to a shadow region. We also observe that some of the false positives correspond to regions with sand ripples, regions with seaweed, or regions with sand ripples and seaweed. The false positives produced by the target detector are therefore the regions in the sonar images under study that are the hardest to reject.

3.3.3 Quantitative results

The performance of the post-processing process was analysed in details over various sets of parameters (see Table 3.2). We set $\varepsilon = 0.05$ and $M = 6$ and $N = 1$. K varies from 2 to 6 and l from 1 to 50. The size $L_1 L_2$ of the signals is given by the target detector. The size of the signals is approximately 40×25 (horizontally and vertically). First, we present the performance of the post-processing process alone by considering, within the initial set of detections, the number of true positives that are declared normal, and the number of false positives that are declared abnormal. From those numbers, we obtain the true positive error rate and the false positive error rate (which have however nothing to do with the true posi-

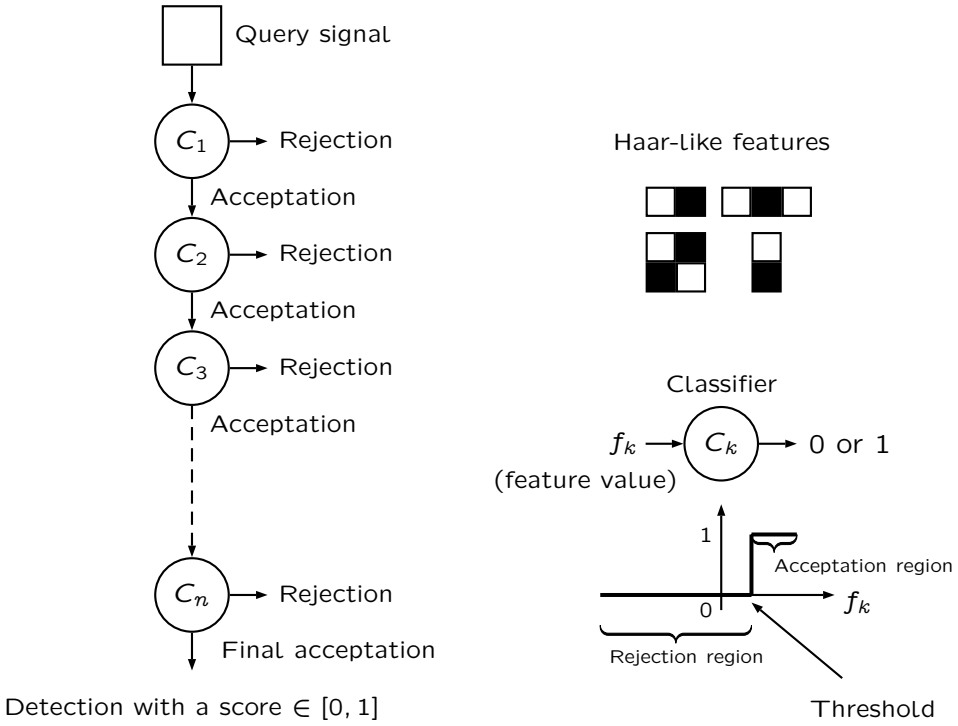


Figure 3.17. Target detector based on a cascade of boosted classifiers.

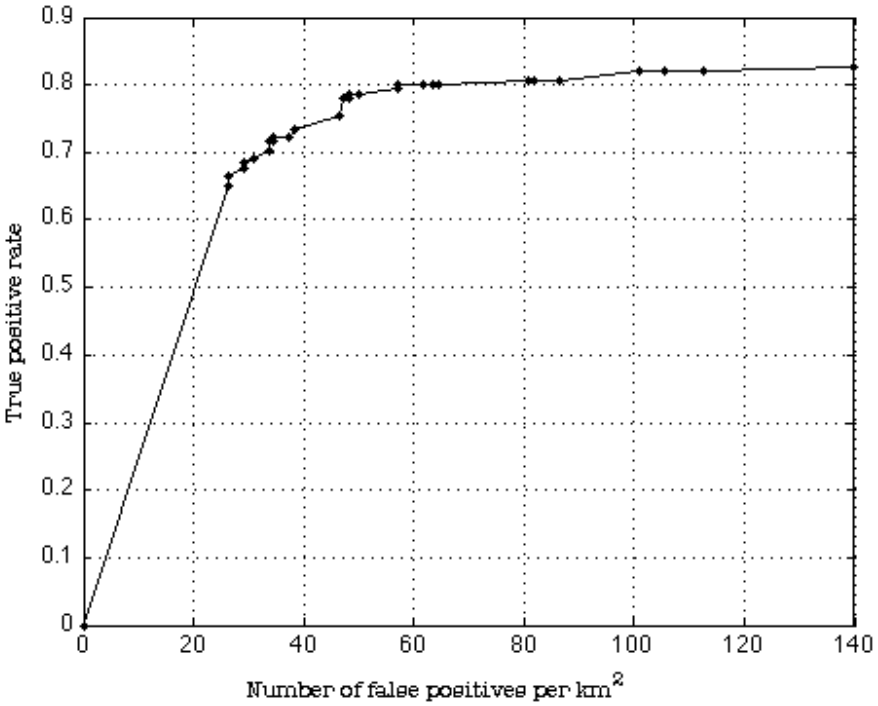


Figure 3.18. Receiver operating characteristic (ROC) curve of the automatic target recognition (ATR) algorithm alone. The true positive rate is depicted as a function of the number of false positives per square kilometre.

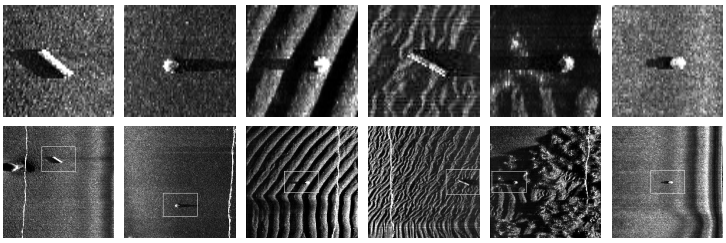


Figure 3.19. Examples of true positives. Top row. True positives alone. Bottom row. True positives in context.

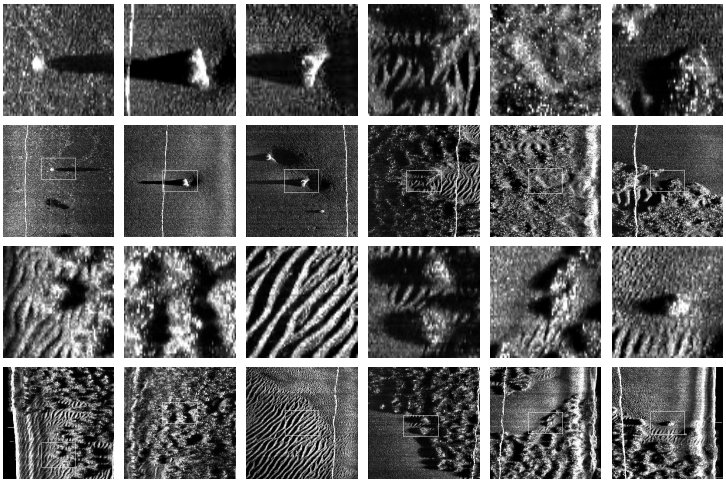


Figure 3.20. Examples of false positives. From top to bottom. Row 1 and 3. False positives alone. Rows 2 and 4. False positives of rows 1 and 3 in context.

tive rate and the false positive rate depicted by a receiver operating characteristic curve). The error rates are depicted through Figures 3.21 to 3.25. We observe one pattern of behaviour. When the dimension ℓ of the affine spaces $\mathbf{A}_{q,\ell}$ is small, the affine spaces provide a coarse representation of the types of seabed, leading to high limits $\tau_{q,\ell}(\varepsilon)$. As a result, many true positives, which are truly far away from the affine spaces, are declared normal, resulting in a high true positive error rate. When the dimension ℓ increases, the representation of the types of seabed improves, resulting in comparatively smaller values of $\tau_{q,\ell}(\varepsilon)$. Comparatively fewer true positives are declared normal, which results in a decrease of the true positive error rate. The false positive error rate starts increasing at this point. When the dimension ℓ increases further more, the true positive error rate keeps decreasing until it becomes equal to zero. The false positive error rate keeps increasing until it be comes equal to one[2]. We observe that the dimension corresponding to the intersection point of the two curves decreases as K increases. This shows that the seabed representation induced by scattering operators does become more stable as the depth of the wavelet transform increases. We also observe that there exists a dimension for which the true positive error rate is equal to zero, and for which the false positive error rate is strictly below one, and maximum. This dimension is for example equal to 19 in Figure 3.22, to 10 in Figure 3.23, and to 8 in Figure 3.24. In this case, it is possible to reduce the number of false positives, within the initial set of detections, without affecting the number of true positives. This suggests that the performance of the target detector can be improved by the post-processing process. The receiving operating characteristic curves obtained for the combined target detection algorithm and post-processing process are depicted through Figures 3.26 to 3.29. We observe that, on the whole, the target detector is not improved, which can be interpreted in the following fashion. The

[2]The false positive error rate is actually not equal to zero in Figure 3.25 when $\ell \geq 43$. The values of the false positive error rate were not available when $\ell \geq 43$ because the running of the post-processing process did not fully come to an end.We can safely assume that the false positive error rate would have been equal to one when $\ell \geq 43$.

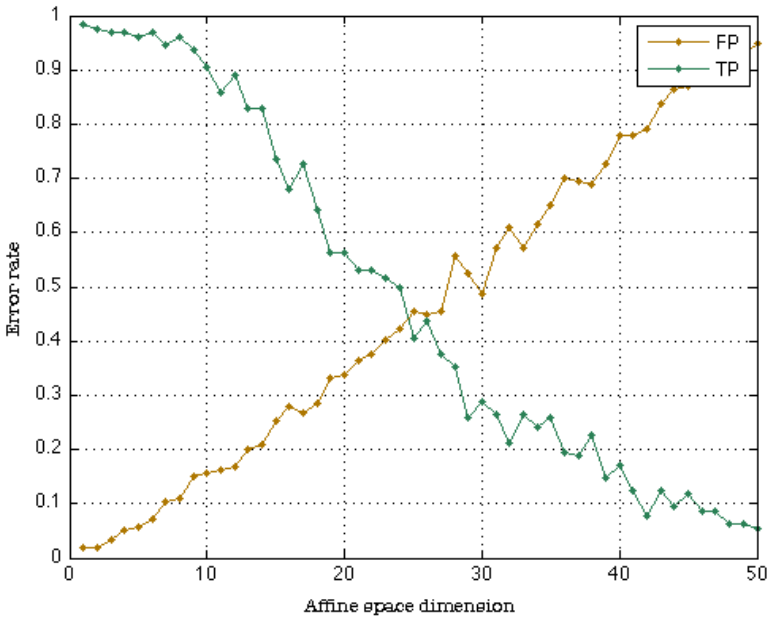


Figure 3.21. True positive (TP) error rate and false positive (FP) error rate as a function of the affine space dimension l for $K = 2$ and $N = 1$.

false positives that are discarded by the post-processing process are the ones already discarded by score of the target detection algorithm. Those false positives are the obvious false positives, and not the "hard" ones we would have liked to discard in the first place. We observe a little improvement in Figure 3.27 when the true positive rate is between 70 % and 80 %. This corresponds to $K = 4$ and $N = 1$. We therefore run the post-processing process for $K = 4$, $N = 2$ and $N = 3$ to see whether the little improvement could be made greater. We observe in Figures 3.30 to 3.33 that this is not the case.

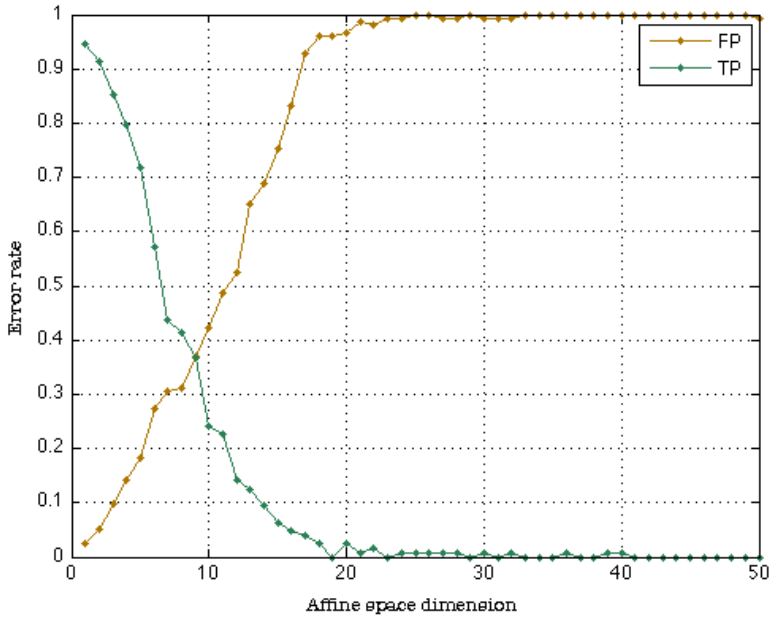


Figure 3.22. True positive (TP) error rate and false positive (FP) error rate as a function of the affine space dimension l for $K = 3$ and $N = 1$.

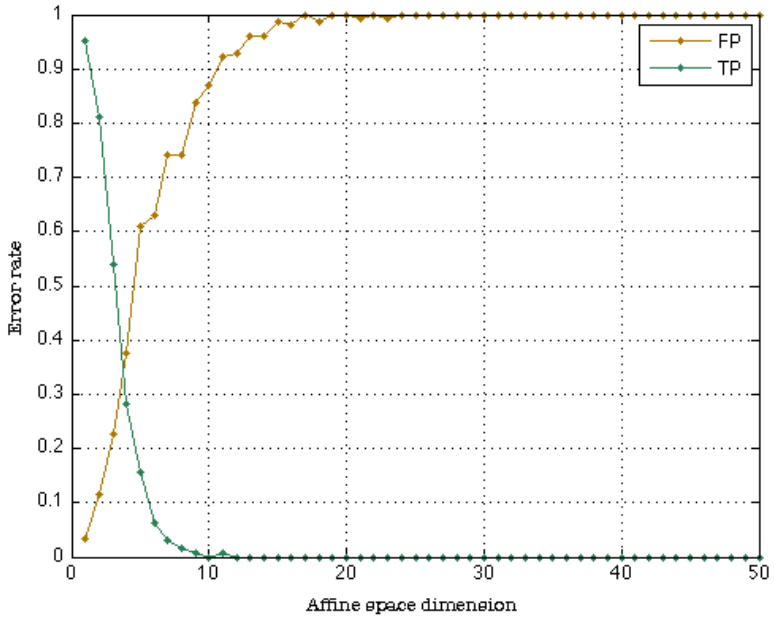


Figure 3.23. True positive (TP) error rate and false positive (FP) error rate as a function of the affine space dimension l for $K = 4$ and $N = 1$.

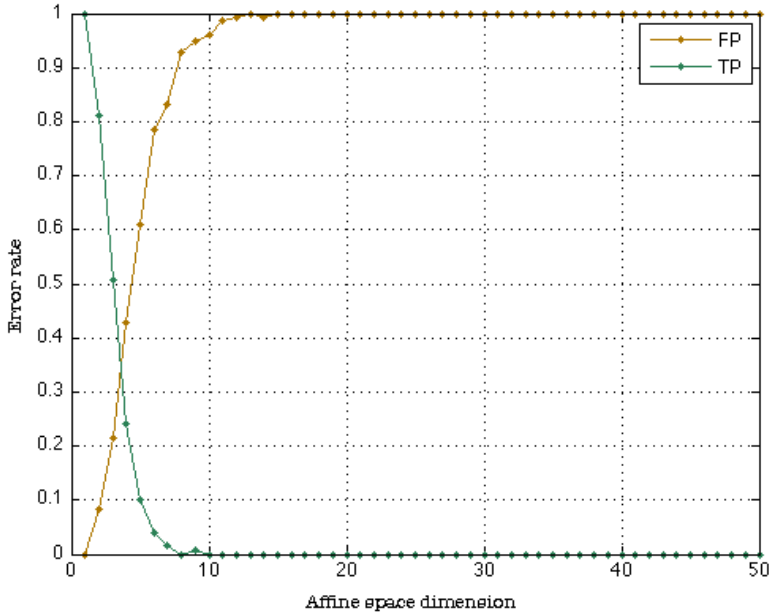


Figure 3.24. True positive (TP) error rate and false positive (FP) error rate as a function of the affine space dimension l for $K = 5$ and $N = 1$.

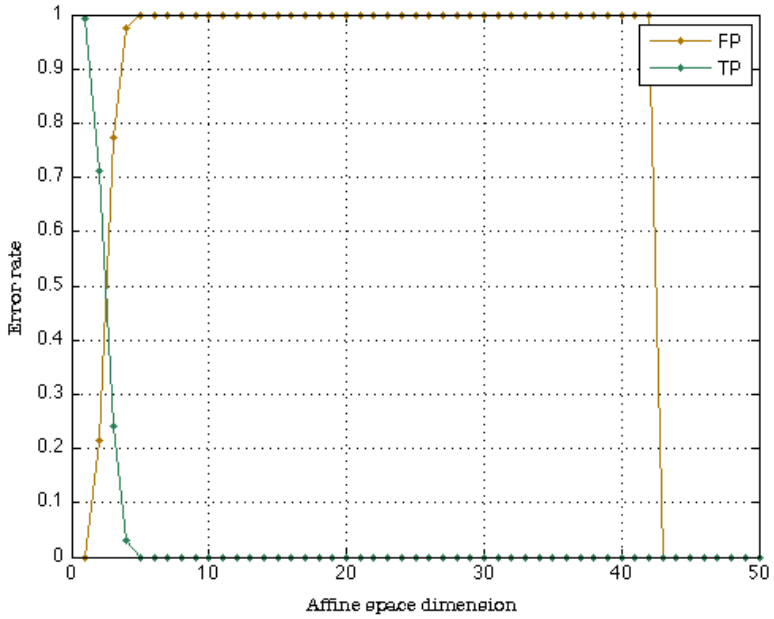


Figure 3.25. True positive (TP) error rate and false positive (FP) error rate as a function of the affine space dimension l for $K = 6$ and $N = 1$.

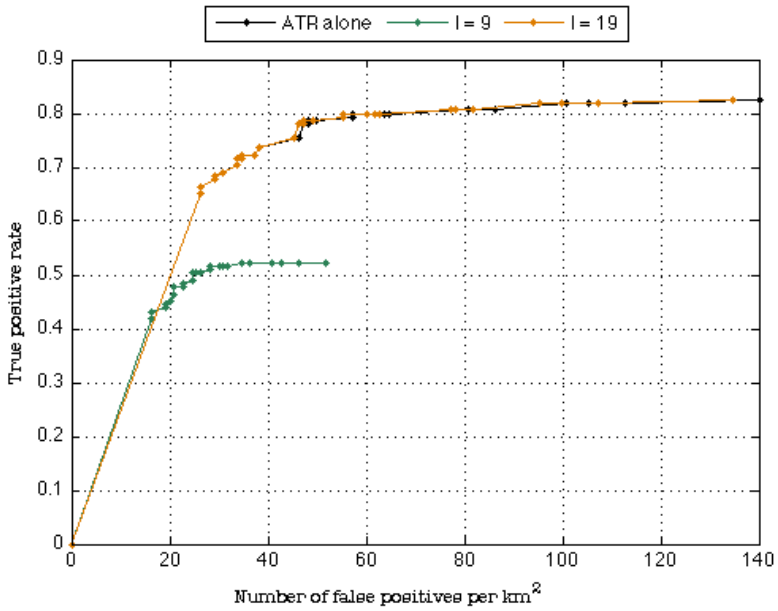


Figure 3.26. Receiver operating characteristic (ROC) curves of the automatic target recognition (ATR) algorithm alone and of the combined target detector and post-processing process for $K = 3$ and $N = 1$.

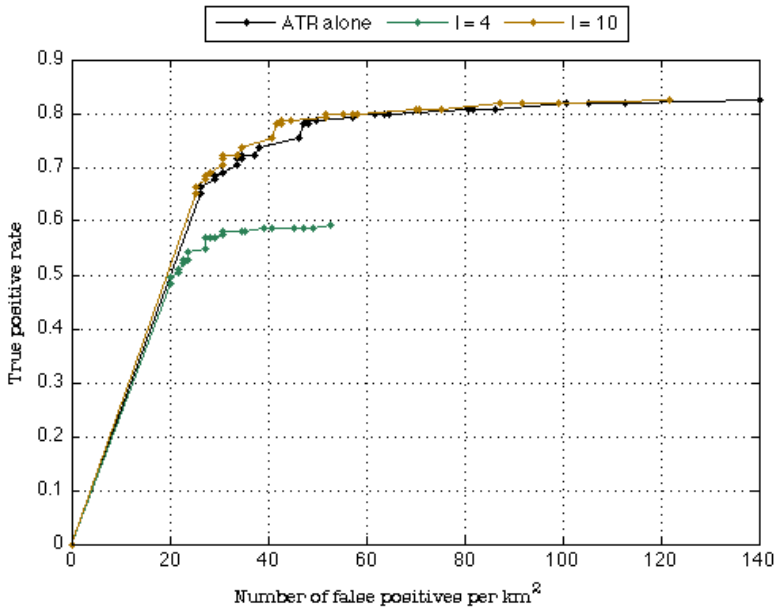


Figure 3.27. Receiver operating characteristic (ROC) curves of the automatic target recognition (ATR) algorithm alone and of the combined target detector and post-processing process for $K = 4$ and $N = 1$.

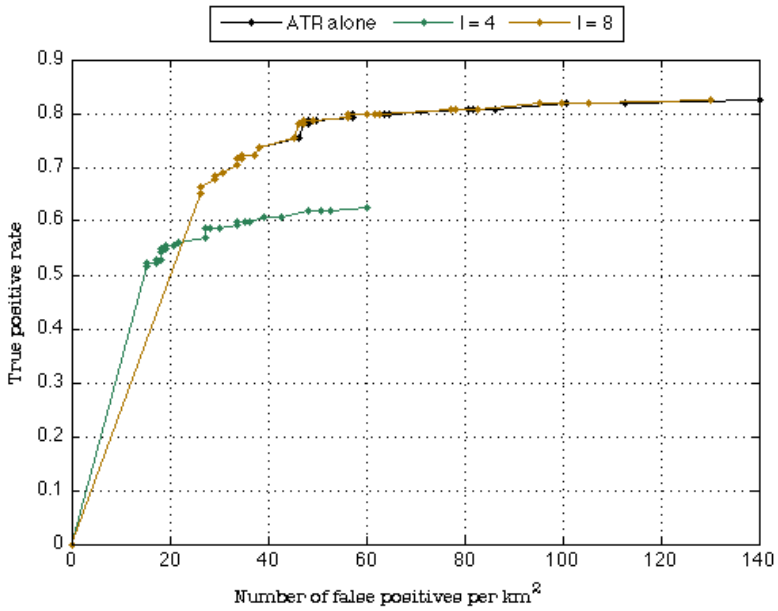


Figure 3.28. Receiver operating characteristic (ROC) curves of the automatic target recognition (ATR) algorithm alone and of the combined target detector and post-processing process for $K = 5$ and $N = 1$.

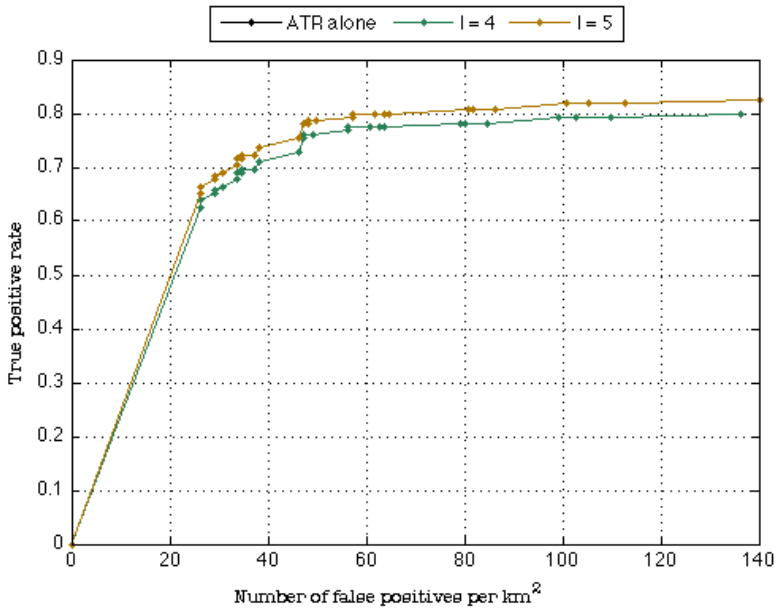


Figure 3.29. Receiver operating characteristic (ROC) curves of the automatic target recognition (ATR) algorithm alone and of the combined target detector and post-processing process for $K = 6$ and $N = 1$.

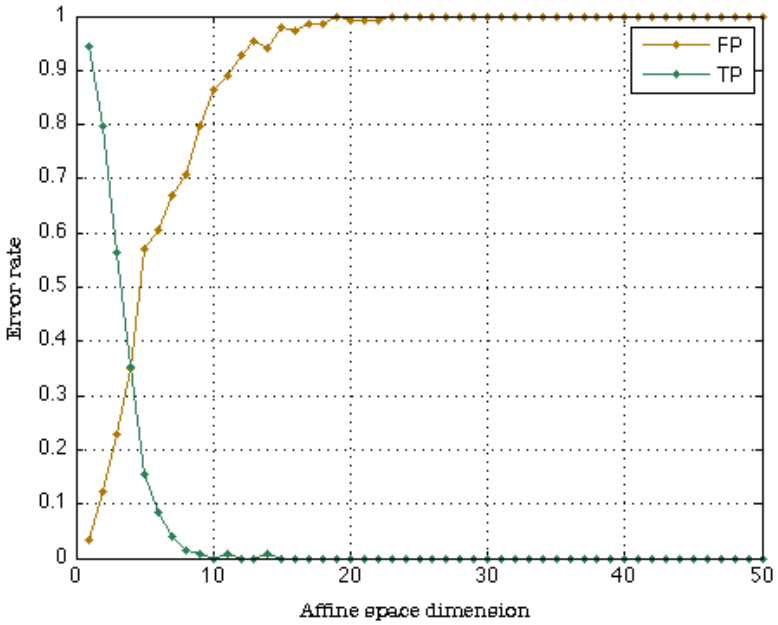


Figure 3.30. True positive (TP) error rate and false positive (FP) error rate as a function of the affine space dimension l for $K = 4$ and $N = 2$.

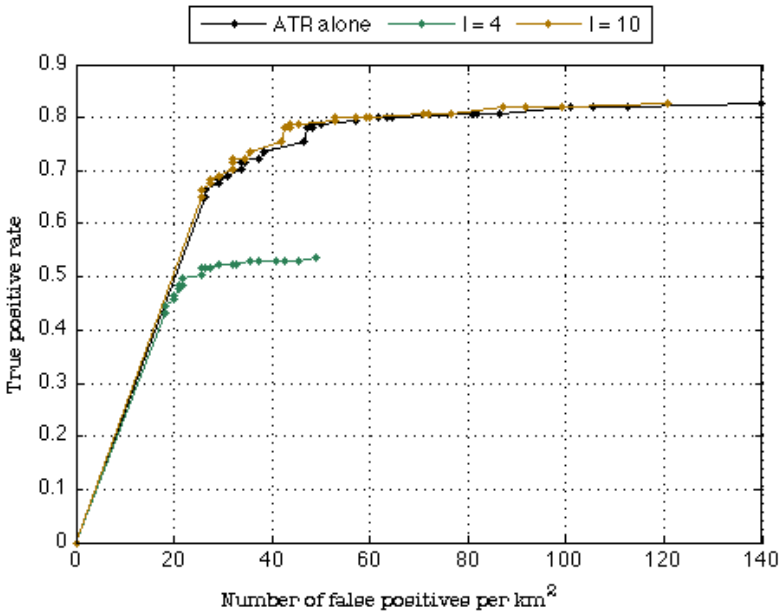


Figure 3.31. Receiver operating characteristic (ROC) curves of the automatic target recognition (ATR) algorithm alone and of the combined target detector and post-processing process for $K = 4$ and $N = 2$.

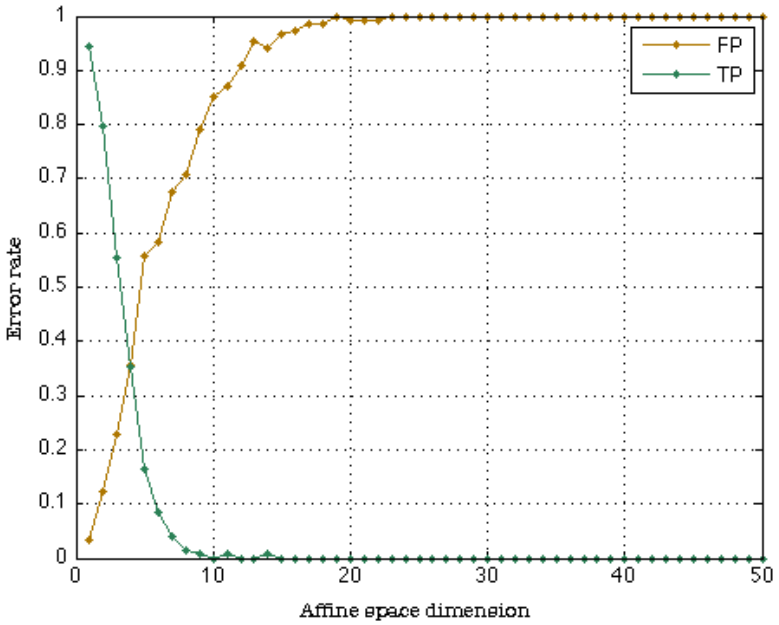


Figure 3.32. True positive (TP) error rate and false positive (FP) error rate as a function of the affine space dimension l for $K = 4$ and $N = 3$.

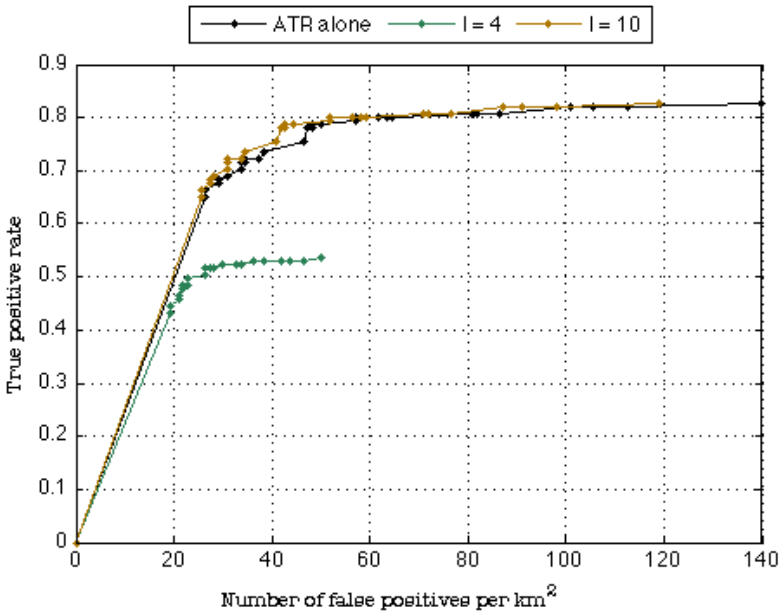


Figure 3.33. Receiver operating characteristic (ROC) curves of the automatic target recognition (ATR) algorithm alone and of the combined target detector and post-processing process for $K = 4$ and $N = 3$.

3.4 Conclusions and perspectives

This chapter was concerned with the analysis of side-looking sonar images with affine spaces. Signals (namely, rectangular regions in the images under study) are described by means of the discrete scattering transform, introduced in Chapter 2, and compared to a collection of classes. Each class represents a type of seabed, and is modelled by an affine space of low dimension with respect to that of the representation of the signals. To be more precise, the dimension of the signal representation is more than a hundred times greater than that of the affine spaces. The kind of comparison that is made between the signals and the classes depends on the inference problem at stake. In a classification problem, the objective is to make a decision upon the type of seabed (the class) of the signals. In an anomaly detection problem, the objective is to decide whether the signals belong to any of the classes, that is to say, whether the signals are at all part of the seabed. These two inference problems are the mathematical formulation of the the two problems that are dealt with in this thesis.

The achievements of this chapter are the following. We showed that the combined use of the scattering transform and an inference algorithm based on affine spaces led to an excellent characterisation of the seabed in side-looking sonar images. The types of seabed we considered are flat areas and areas made of sand ripples or vegetation. We quantitatively showed that the characterisation of the seabed remains excellent across a wide range of parameters (five parameters are to be set in total). The characterisation only deteriorates when the chosen parameters are unreasonable. By an unreasonable choice of parameters, we mean a very low affine space dimension, a Littlewood Paley wavelet transform of limited depth, and signals of very small sizes. The fact that the characterisation of the seabed remains excellent across a wide range of parameters is, to our opinion, something in favour of the seabed identification algorithm we presented in this chapter. From our experience, it is not often the case that an inference algorithm is not very sensitive to variations in the parameters it is based upon. In this chapter, we also showed that the characterisation of the seabed is not affected by the presence of artefacts in the side-looking sonar images. For instance, the types of seabed are not mistaken from one another around the vertical white lines which correspond to the acoustic returns from the surface of the water. We finally showed that only a small training set was needed to obtain such an excellent characterisation of the seabed. By small training set, we mean that portions of only two images were used as a training set to classify more than two hundred images. This demonstrates the generalisation capabilities of the signal representation induced by scattering operators. It could be that another inference algorithm based on, for instance, a neural network or a support vector machine, performs better than the one based on affine spaces. This was not assessed in this thesis, mainly because the inference algorithm based on affine spaces was found computationally inexpensive, which balanced the computational complexity of the scattering transform alone.

In this chapter, we quantitatively compared the sea bottom characterisation

performance obtained with two signal representations, the one induced by the scattering transform and the one induced by local binary patterns. Local binary patterns are of common use in the field of computer vision but, as far as we know, have never been used in the field of sonar imagery. We chose this alternative signal representation over the ones derived from co-occurrence matrices or sum and difference histograms because it was shown in the literature that these representations are encompassed by local binary patterns. We showed that scattering operators outperform local binary patterns by about 10 %. Specifically, sand ripples are not mistaken for seaweed as much, which is a recurrent issue when one intends to identify the types of seabed. For such a purpose, we therefore recommend the use of the scattering representation over the representation that have been used in the past.

This chapter was also concerned with the reduction, in the form of a post-processing process of the number of false positives in a set of detections provided by a target detector. Such a task is commonly performed through the extraction of the highlight and the shadow of the detections, and the analysis of these two with respect to templates. Our original hypothesis, as stated in Chapter 1, is that it is possible to reduce the number of false positives without affecting the number of true positives should we quantify the similarity between the detections and the seabed, the false positives being by nature part of the seabed. This was cast in this chapter as an anomaly detection problem. We showed that our original hypothesis was wrong when the detections are represented by scattering operators, and when the inference algorithm is based on affine spaces. However, one would have to combine other signal representations and other inference algorithms to assess whether our original hypothesis can ever be true. We will do so in the following chapter. We will indeed combine the same signal representation with another inference algorithm based on the concept of ensemble of patches.

Chapter 4

Image analysis by composition

"Nothing goes by luck in composition. It allows no tricks. The best you can write is the best you are."

Henry David Thoreau

4.1 Composition process

4.1.1 Introduction

Several problems encountered in the field of sonar imagery require a good measure of the similarity between regions of the sonar images under study. We will present, in this chapter, an inference algorithm which is based on the concept of visual composition [57]. This inference algorithm originally appeared in the field of computer vision, and is adapted, in this chapter, to the analysis of side-looking sonar images. We will show that the inference algorithm is potentially applicable to the two inference problems stated in the previous chapter. We consider a query signal Q, that is to say, a rectangular region in one of the sonar images under study. The query signal is compared to a reference R which can be, for example, the rest of the sonar image the query signal comes from, an external collection of other sonar images, or an external collection of portions of sonar images, as illustrated in Figure 4.1. In this figure, the reference is the collection $R = (R_1, R_2, R_3, R_4)$ of portion of sonar images representing each a type of seabed. The inference algorithm intends to evaluate to what extent the query signal can be composed from pieces taken from the reference. The composition process is similar in mind to that of a puzzle. The complexity of the composition process leads to a measure of the similarity between the query signal and signals S in the reference. The signals S are rectangular regions in the reference of the same size as that of the query signal. A signal of particular interest is the signal

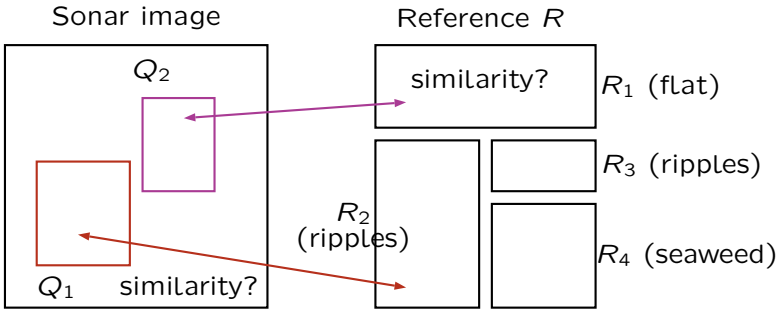


Figure 4.1. Query signal Q and reference R.

$S^\star$ which is most similar to the query signal. We will explain in Section 4.1.2 in what sense $S^\star$ is most similar to Q. Knowledge of what the signal $S^\star$ is with respect to the reference leads to a variety of interpretations, depending on the problem at issue. We will explain in Section 4.1.2 what interpretation is to be followed when answering the two inference problems. The pieces involved in the composition process are rectangular patches of small size with respect to the size of the query signal, as illustrated in Figure 4.2. The size of the patches can be interpreted as the spatial resolution, above that of a pixel, at which the inference algorithm works. The patches in the query signal are compared to the patches in the reference, and then combined in order to compose the query signal. This is where the visual composition process comes into place. A set of descriptors is associated to every patch to make the inference algorithm sensitive subtle local changes in side-looking sonar images. The composition process takes therefore into account the geometric arrangement of the patches as well as their descriptors. The set of descriptors comes from the image representation induced by scattering operators, as presented in Chapters 2 and 3. If a patch correspond to a region $\mathcal{R}$ ($\mathcal{R}$ is not R) in a sonar image, the descriptors of the patch are obtained by extracting the same region from the discrete scattering transform of the image. We mentioned in Chapter 2 that image representation induced by scattering operators possesses an intrinsic constant spatial resolution. This constant spatial resolution is of high interest because we can make it match the resolution at which the inference algorithm work. Doing so, we combine the inference algorithm and the image representation in an attractive way. We will see in Sections 4.1.2 and 4.1.3 that the inference algorithm takes the form of a maximum a posteriori estimation over a Bayesian network. The network statistically models the dependencies between the relative positions of the patches, and their descriptors. In this respect, the visual composition process is also a statistical composition process. The maximum a posteriori estimation is practically and efficiently performed with a belief propagation algorithm [107].

4.1.2 Statistical inference over a Bayesian network

We present, in this section, the inference algorithm based on the concept of visual composition [57]. We consider a query signal Q in the sonar image under study,

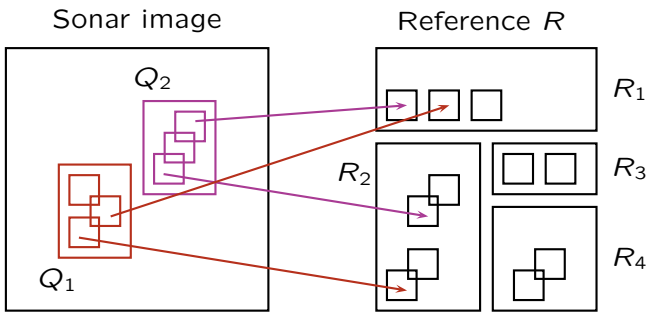


Figure 4.2. The query signals and the reference are broken into patches of small size which are compared to each other.

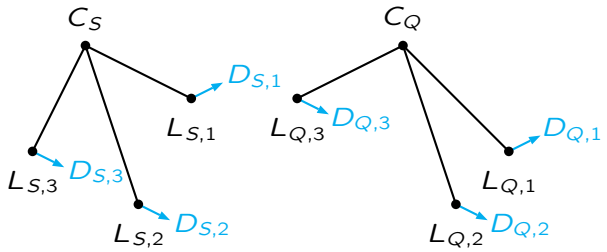


Figure 4.3. Ensembles of patches S and Q. The centres of the ensembles are C_S and C_Q. The positions of the patches in the ensembles are $L_{S,k}$ and $L_{Q,k}$. The descriptors of the patches in the ensembles are $D_{S,k}$ and $D_{Q,k}$.

and a signal S in the reference R, as we did in the introduction. The two signals are represented each by an ensemble of patches. A patch is represented by a position and a descriptor. Patches in S are denoted by $P_{S,k}$ and patches in Q by $P_{Q,k}$ as illustrated on Figure 4.3. In both cases, k is a natural number that corresponds to the index of the patch in either S or Q. We denote by $L_{S,k}$ and $L_{Q,k}$ the positions of the patches $P_{S,k}$ and $P_{Q,k}$. We denote by $D_{S,k}$ and $D_{Q,k}$ the descriptors of the patches $P_{S,k}$ and $P_{Q,k}$. We then write

$$\begin{aligned} P_{S,k} &= (L_{S,k}, D_{S,k}) \\ P_{Q,k} &= (L_{Q,k}, D_{Q,k}) \end{aligned} \tag{4.1}$$

The positions of the patches in S are defined with respect to a reference point C_S, and the positions of the patches in Q are defined with respect to a reference point C_Q. The reference points are the global positions of the ensembles of patches. As a result, C_S is the global position of S in the reference, and C_Q, the global position of Q in the sonar image under study. The signals S and Q are ultimately represented by the collection of random variables

$$\begin{aligned} S &= (C_S, P_{S,1}, \ldots, P_{S,k}, \ldots) \\ Q &= (C_Q, P_{Q,1}, \ldots, P_{Q,k}, \ldots) \end{aligned} \tag{4.2}$$

The similarity between S and Q is measured by the joint probability distribution

$\mathbb{P}(S, Q)$ which quantifies how likely it is to find two ensembles of patches at C_S and C_Q for which the relative arrangement of the patches as well as the descriptors of the patches are similar. The signal $S^\star$ in the reference which is most similar to Q is obtained with a maximum a posteriori estimation, namely

$$\begin{aligned} S^\star &= \arg\max_S \mathbb{P}(S \mid Q) \\ &= \arg\max_S \mathbb{P}(Q \mid S)\, \mathbb{P}(S) \end{aligned} \tag{4.3}$$

The conditional probability distribution $\mathbb{P}(Q \mid S)$ describes how to generate a query Q from a signal S. The generation process allows for changes in the positions and the descriptors of the patches. The following assumption is made. Given an ensemble of patches S and the global position C_Q of the ensemble of patches Q, the patches $P_{Q,k}$ in Q are independent of each other, and only depend on their corresponding patches $P_{S,k}$ in S. Such an assumption is similar to the memoryless channel assumption in hidden Markov models [108]. The conditional probability distribution $\mathbb{P}(Q \mid S)$ can therefore be written

$$\begin{aligned} \mathbb{P}(Q \mid S) &= \mathbb{P}(C_Q, P_{Q,1}, \ldots, P_{Q,k}, \ldots \mid C_S, P_{S,1}, \ldots, P_{S,k}, \ldots) \\ &= \mathbb{P}(C_Q \mid C_S, P_{S,1}, \ldots, P_{S,k}, \ldots) \\ &\quad \times \mathbb{P}(P_{Q,1}, \ldots, P_{Q,k}, \ldots \mid C_S, C_Q, P_{S,1}, \ldots, P_{S,k}, \ldots) \\ &= \mathbb{P}(C_Q) \times \mathbb{P}(P_{Q,1}, \ldots, P_{Q,k}, \ldots \mid C_S, C_Q, P_{S,1}, \ldots, P_{S,k}, \ldots) \\ \mathbb{P}(Q \mid S) &= \mathbb{P}(C_Q) \times \prod_k \Big\{ \mathbb{P}(P_{Q,k} \mid C_S, C_Q, P_{S,k}) \Big\} \end{aligned} \tag{4.4}$$

We can further express this conditional probability distribution in terms of the positions and the descriptors of the patches.

$$\begin{aligned} \mathbb{P}(Q \mid S) &= \mathbb{P}(C_Q) \times \prod_k \Big\{ \mathbb{P}(P_{Q,k} \mid C_S, C_Q, P_{S,k}) \Big\} \\ &= \mathbb{P}(C_Q) \times \prod_k \Big\{ \mathbb{P}(L_{Q,k}, D_{Q,k} \mid C_S, C_Q, L_{S,k}, D_{S,k}) \Big\} \\ &= \mathbb{P}(C_Q) \times \prod_k \Big\{ \mathbb{P}(L_{Q,k} \mid C_S, C_Q, L_{S,k}, D_{S,k}) \\ &\quad \times \mathbb{P}(D_{Q,k} \mid C_S, C_Q, L_{S,k}, D_{S,k}) \Big\} \end{aligned} \tag{4.5}$$

$$\mathbb{P}(Q \mid S) = \mathbb{P}(C_Q) \times \prod_k \Big\{ \mathbb{P}(L_{Q,k} \mid C_S, C_Q, L_{S,k}) \times \mathbb{P}(D_{Q,k} \mid D_{S,k}) \Big\} \tag{4.6}$$

The prior probability distribution $\mathbb{P}(S)$ describes all knowledge available about the signals S in the reference. The following assumption is made. The global position C_S of the ensemble of patches S, and the patches $P_{S,k}$ of that ensemble are independent of each other. As a result,

$$\mathbb{P}(S) = \mathbb{P}(C_S) \times \prod_k \Big\{ \mathbb{P}(P_{S,k}) \Big\} \tag{4.7}$$

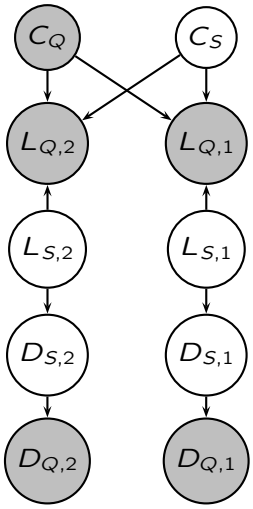


Figure 4.4. Bayesian network representing the statistical dependencies between the random variables introduced in the inference algorithm.

which is further reduced to

$$\mathbb{P}(S) = \mathbb{P}(C_S) \times \prod_k \Big\{ \mathbb{P}(L_{S,k})\, \mathbb{P}(D_{S,k} \mid L_{S,k}) \Big\} \tag{4.8}$$

We have defined thus far statistical dependencies within the ensemble of patches S by means of the prior probability distribution $\mathbb{P}(S)$. We also have defined statistical dependencies across the ensembles of patches S and Q by means of the conditional probability distribution $\mathbb{P}(Q \mid S)$. A Bayesian network provides a convenient graphical representation of those statistical dependencies as illustrated on Figure 4.4. Each node in the Bayesian network represents one of the aforementioned random variables. An arrow between two nodes defines a statistical dependency between the corresponding two random variables. A conditional probability distribution is then associated with each arrow. The conditional probability distributions which appear in (4.6) and (4.8) are the ones needed in the Bayesian network associated with S and Q. A prior probability distribution is associated with the nodes in the Bayesian network with no arrow directed towards them. The prior probability distributions which appear appear in (4.6) and (4.8) are the ones needed in the Bayesian network associated with S and Q.

4.1.3 Belief propagation algorithm

Upon combining (4.3) with (4.6) and (4.8) we find the equation of the maximum a posteriori estimation.

$$S^\star = \arg\max_S \Big\{ \mathbb{P}(C_S) \times \prod_k \Big\{ \mathbb{P}(L_{S,k})\, \mathbb{P}(D_{S,k} \mid L_{S,k}) \times\ \mathbb{P}(D_{Q,k} \mid D_{S,k})\, \mathbb{P}(L_{Q,k} \mid C_S, C_Q, L_{S,k}) \Big\} \Big\} \tag{4.9}$$

What is to be performed in (4.9) is the maximisation of a product of probability distributions over all possible values of S. The maximisation is efficiently performed with the help of a belief propagation algorithm. This algorithm is based on messages being exchanged and combined between the nodes of the Bayesian network. The messages carry information about what should be the state of the nodes. A belief propagation algorithm is essentially sequential. Messages are exchanged and combined between the nodes at the edge of the Bayesian network before other messages are exchanged and combined between the nodes at the core of the Bayesian network. We will now highlight the main steps of the belief propagation algorithm. Let us consider the branch of the Bayesian network which corresponds to the patches $P_{S,k}$ and $P_{Q,k}$. A message $\eta_{D_{S,k} \to L_{S,k}}(L_{S,k})$ is first passed from node $D_{S,k}$ to node $L_{S,k}$ about what should be the value of $L_{S,k}$ as illustrated on Figure 4.5. This message is readily obtained from (4.9) by considering the maximisation over $D_{S,k}$ only.

$$\eta_{D_{S,k} \to L_{S,k}}(L_{S,k}) = \max_{D_{S,k}} \Big\{ \mathbb{P}(D_{Q,k} \,|\, D_{S,k})\, \mathbb{P}(D_{S,k} \,|\, L_{S,k}) \Big\} \tag{4.10}$$

A message $\eta_{L_{S,k} \to C_S}(C_S)$ is then passed from node $L_{S,k}$ to node C_S about what should be the value of C_S. This message is readily obtained from (4.9) and (4.10) by considering the maximisation over $L_{S,k}$ in (4.9) and bearing in mind that the maximisation over $D_{S,k}$ is included in the previous message.

$$\eta_{L_{S,k} \to C_S}(C_S) = \max_{L_{S,k}} \Big\{ \mathbb{P}(L_{S,k})\, \eta_{D_{S,k} \to L_{S,k}}(L_{S,k}) \times \mathbb{P}(L_{Q,k} \,|\, C_S, C_Q, L_{S,k}) \Big\} \tag{4.11}$$

The messages passed to C_S by the branches of the Bayesian network are eventually combined in order to find the maximum a posteriori estimate of C_S. The estimate is readily obtained from (4.9) and (4.11) by considering the maximisation over C_S in (4.9) and bearing in mind that the maximisations over $D_{S,k}$ and $L_{S,k}$ are included in the previous messages.

$$C_S^\star = \arg\max_{C_S} \Big\{ \mathbb{P}(C_S) \times \prod_k \Big\{ \eta_{L_{S,k} \to C_S}(C_S) \Big\} \Big\} \tag{4.12}$$

In addition

$$\mathcal{S}_Q^\star = \max_{C_S} \Big\{ \mathbb{P}(C_S) \times \prod_k \Big\{ \eta_{L_{S,k} \to C_S}(C_S) \Big\} \Big\} \tag{4.13}$$

provides a measure of the similarity between $S^\star$ and Q. The belief propagation algorithm provides therefore the position of $S^\star$ in the reference and a measure of the similarity between $S^\star$ and Q. It is also possible to obtain the positions and the descriptors of the patches in $S^\star$ by *following the sequence of maximisations backwards*, as illustrated in Figure 4.6. We are however only interested in $C_S^\star$ and $\mathcal{S}_Q^\star$ in this chapter. In Figure 4.6, the vertical arrows represent the range of the possible values of the random variables C_S, $L_{S,k}$ and $D_{S,k}$. Those values can be either discrete or continuous, which will be talked through in Section 4.1.4. The

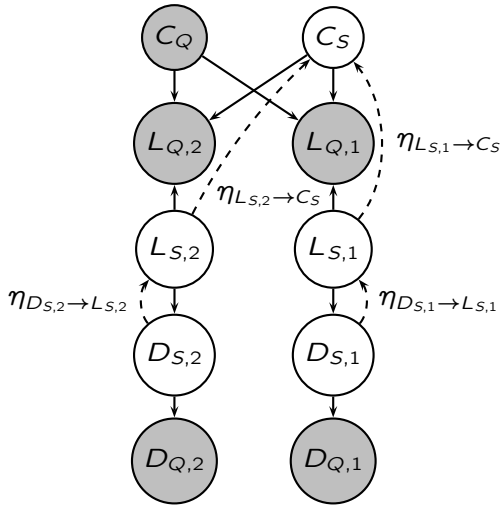


Figure 4.5. Messages being exchanged and combined between the nodes of the Bayesian network associated with S and Q.

x's, y's and z's represent specific values of the random variables. For example, $x_{D_{S,1}}$ and $y_{D_{S,1}}$ are two possible values of $D_{S,1}$ while x_{C_S}, y_{C_S} and z_{C_S} are three possible values of C_S. The plain arrows directed towards the centre of the figure represent the messages being exchanged and combined between the nodes during the belief propagation algorithm. The dashed arrows directed towards the edges of the figure represent the sequence of maximisations being followed backwards from one possible value x_{C_S} of C_S, which is the position $C_S^\star$ of $S^\star$ found from (4.12).

4.1.4 Statistical models

We have yet to specify what are the prior and conditional probability distributions which appear in (4.9) and hence what are the possible values of the random variables which appear in the Bayesian network. The reference points C_S and C_Q of the ensembles of patches S and Q are discrete random variables because sonar images come in discrete form. The possible values of C_S are the positions of the pixels of the sonar images the reference is composed of while the possible values of C_Q are the positions of the pixels of the sonar image under study. We assume a uniform prior probability distribution for C_S, which leads to

$$\mathbb{P}(C_S) = \text{constant} \tag{4.14}$$

The positions $L_{S,k}$ and $L_{Q,k}$ of the patches in the ensembles of patches S and Q are likewise discrete random variables. The possible values of $L_{S,k}$ are the positions of the pixels of the sonar images the reference is composed of while the possible values of $L_{Q,k}$ are the positions of the pixels of the sonar image under study. We assume a uniform prior probability distribution for $L_{S,k}$, which leads to

$$\mathbb{P}(L_{S,k}) = \text{constant} \tag{4.15}$$

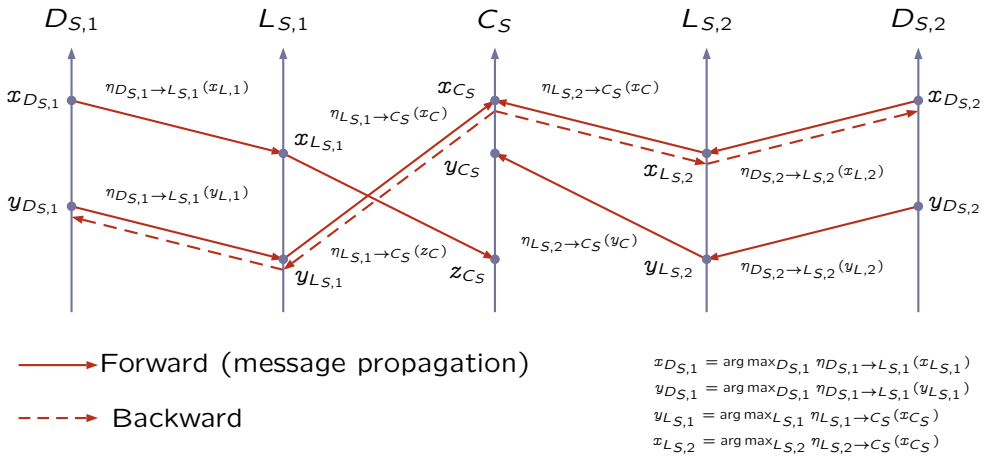


Figure 4.6. Following of the sequence of maximisations backwards from x_C (dashed lines).

We also assume that

$$\mathbb{P}(L_{Q,k} \,|\, C_S, C_Q, L_{S,k}) = \begin{cases} \text{constant} & \text{if} \quad L_{Q,k} \in \mathcal{R} \\ 0 & \text{if} \quad L_{Q,k} \notin \mathcal{R} \end{cases} \tag{4.16}$$

where $\mathcal{R}$ is a rectangular region drawn about $C_Q + L_{S,k} - C_S$. Here, $L_{S,k} - C_S$ is the difference of two points and hence the vector drawn from C_S to $L_{S,k}$. As a result, $C_Q + L_{S,k} - C_S$ is the point obtained by drawing the same vector $L_{S,k} - C_S$ from C_Q as illustrated on Figure 4.7. This conditional probability distribution is locally uniform, and models the allowed geometrical deformations when a signal Q is generated from a signal S. The extent of the allowed geometrical deformations is controlled by the extent of $\mathcal{R}$. The descriptors $D_{S,k}$ and $D_{Q,k}$ of the patches in S and Q are multivariate real-valued random variables obtained from the discrete scattering transform (see Section 4.1.1 and (3.5) in Section 3.1.2). We assume that

$$\mathbb{P}(D_{Q,k} \,|\, D_{S,k}) \propto \exp\big[-\frac{1}{2}(D_{S,k} - D_{Q,k})^T \times \Sigma^{-1} \times (D_{S,k} - D_{Q,k})\big] \tag{4.17}$$

where Σ is the a pre-defined covariance matrix (the size of the covariance matrix is given by (3.5) in Section 3.1.2 in which $L_1 L_2$ becomes the size of the patches). This conditional probability distribution models the allowed distortions in the values of the descriptors when a signal Q is generated from a signal S. We finally assume that

$$\mathbb{P}(D_{S,k} \,|\, L_{S,k}) = \begin{cases} 1 & \text{if} \quad (L_{S,k}, D_{S,k}) \in R \\ 0 & \text{if} \quad (L_{S,k}, D_{S,k}) \notin R \end{cases} \tag{4.18}$$

Note that we use a rectangular region $\mathcal{R}$ in (4.16) and the reference R in (4.18). This conditional probability distribution models the statistical relation between the descriptor of a patch and its position within the reference R. This probabil-

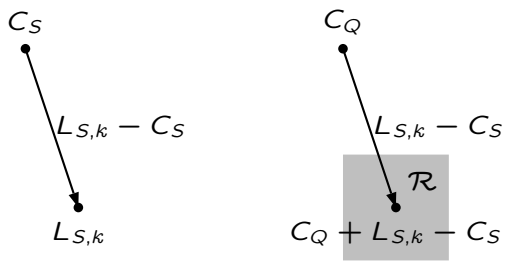


Figure 4.7. Geometrical setting associated with one of the conditional probability distribution of the Bayesian network.

ity distribution does not depend on any parameter and is nonetheless perfectly adapted to the reference R. The inference algorithm benefits from this non-parametric probability distribution because no parameters need be estimated for specific references.

4.1.5 Behaviour of the belief propagation algorithm

We will illustrate, in this section, the behaviour of the belief propagation algorithm. We will specifically illustrate what are the messages exchanged and combined between the nodes of the Bayesian network. We consider three side-looking sonar images and three query signals Q (one per image). We define the reference R as the rest of the sonar images the query signals comes from. The reference is therefore different for each query signal. The sonar images and the query signals are depicted in Figures 4.8, 4.10 and 4.12. We consider one patch $P_{Q,k_\bullet}$ in the query Q. The selected patches are depicted in Figures 4.8, 4.10 and 4.12. Messages are exchanged and combined within the branch $k_\bullet$ of the Bayesian network. The message

$$\eta_{D_{S,k_\bullet} \to L_{S,k_\bullet}}(L_{S,k_\bullet})$$

is depicted in Figures 4.9, 4.11 and 4.13 for every possible value of $L_{S,k_\bullet}$. This message indicates the most likely positions of the patch $P_{S,k_\bullet}$. We can see that the type of seabed the patch is made of is selected by this message. The message

$$\eta_{L_{S,k_\bullet} \to C_S}(C_S)$$

is depicted in Figures 4.9, 4.11 and 4.13 for every possible value of C_S. This message indicates the most likely positions of the ensemble of patches S as inferred from the patch $P_{S,k_\bullet}$ only. The spatial effect of the extent of the rectangular region $\mathcal{R}$ in (4.16) can be observed at this point, especially in Figure 4.11. The messages passed to C_S by the branches of the Bayesian network, namely by the branch $k_\bullet$ that was just considered and all the others, are combined in order to find the maximum a posteriori estimate of C_S as in (4.12).

$$\log_{10} \Big\{ \mathbb{P}(C_S) \times \prod_k \Big\{ \eta_{L_{S,k} \to C_S}(C_S) \Big\} \Big\} \tag{4.19}$$

is depicted in Figures 4.9, 4.11 and 4.13 for every possible value of C_S. The maximum value of (4.19) is $\log_{10}(\mathcal{S}_Q^\star)$ and leads to the position of $S^\star$ in the reference. We can see that $S^\star$ belongs to the same type of seabed as the query. Here are the parameters we used to create the figures. The size of a query signal is 32×32, the size of a patch is 4×4, and the size of the region $\mathcal{R}$ in (4.16) is 4×4. The patches do not overlap each other in the query. The covariance matrix Σ in (4.17) was chosen diagonal, the diagonal elements being all equal to 0.1. We chose this value for the diagonal elements because it is within the range of values of the descriptors of the patches. We also used $K = 4$, $M = 6$ and $N = 1$ to compute the discrete scattering transform. We chose $N = 1$ so that the computational complexity of the composition process is not prohibitive. From the figures referenced to in this section, we can make three observations.

1. We can see that the statistical composition process can be used to answer the first inference problem defined in Section 3.1.5 of Chapter 3, that is to say, to identify the types of seabed present in side-looking sonar images. There is indeed a difference of approximately 25 dB between the correct type of seabed (that of the query signal) and the other ones in Figures 4.9 and 4.11[1]. We observe that the composition process distinguishes the small and the large sand ripples of Figure 4.10.

2. The query signal in Figure 4.12 is the highlight of an object. We observe in Figure 4.13 that, compared to the background, that is to say, to the sea bottom, the other objects in the image are most similar to the query signal by approximately 35 dB. This suggests the that the similarity $\log_{10}(S_Q^\star)$ between the query signal and the reference can be used to answer the second inference problem defined in Section 3.1.5 of Chapter 3, that is to say, to reduce the number of false positives in a set of detections provided by an automatic target recognition algorithm.

3. The query signal in Figure 4.12 is the highlight of an object. We observe in Figure 4.13 that, should the selected patch be part of the highlight, the patch is found most similar to the highlight of the other objects in the image. We observe in Figure 4.14 that, should the selected patch be part of the flat area around the highlight, the patch is found most similar to the flat areas in the image, and not to the highlight of the objects. It is therefore the geometrical arrangement of the patches that made the query signal of Figure 4.12 most similar to the objects in the image, and not to the sea bottom.

4.2 Seabed identification

In this section, we will qualitatively illustrate how the composition process can be used to identify the types of seabed. We consider a set of nine side-looking sonar

[1] We use dB as the unit here because what is displayed is $\log_{10}\{\mathbb{P}(C_S) \times \prod_k \{\eta_{L_{S,k} \to C_S}(C_S)\}\}$ instead of $\mathbb{P}(C_S) \times \prod_k \{\eta_{L_{S,k} \to C_S}(C_S)\}$. We have therefore 1 dB $= \log_{10}(10)$.

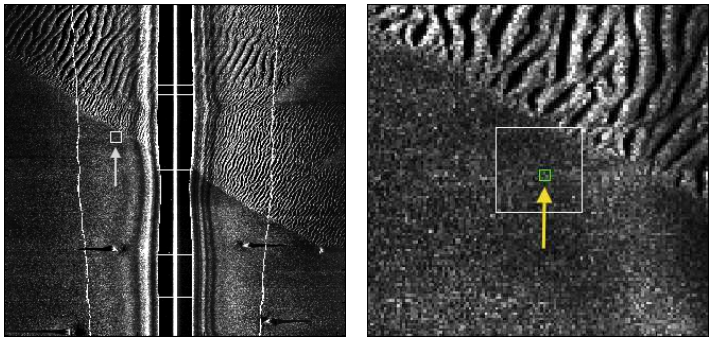


Figure 4.8. Left. Side-looking sonar image and query signal Q. Right. Selected $P_{Q,k_\bullet}$ patch inside the query signal Q.

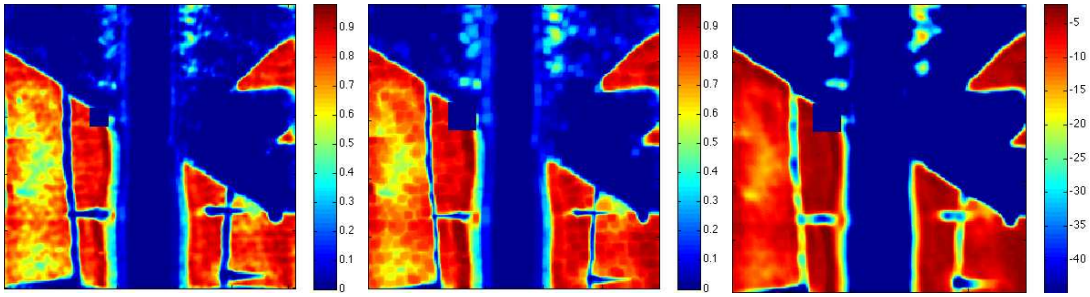


Figure 4.9. Left. Message $\eta_{D_{S,k_\bullet} \to L_{S,k_\bullet}}(L_{S,k_\bullet})$ passed from node $D_{S,k_\bullet}$ to $L_{S,k_\bullet}$ for all possible values of $L_{S,k_\bullet}$. Middle. Message $\eta_{L_{S,k_\bullet} \to C_S}(C_S)$ passed from node $L_{S,k_\bullet}$ to C_S for all possible values of C_S. Right. Value of (4.19) for all possible values of C_S.

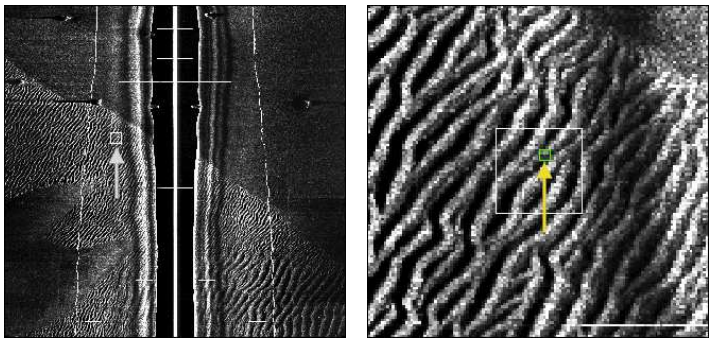


Figure 4.10. Left. Side-looking sonar image and query signal Q. Right. Selected $P_{Q,k_\bullet}$ patch inside the query signal Q.

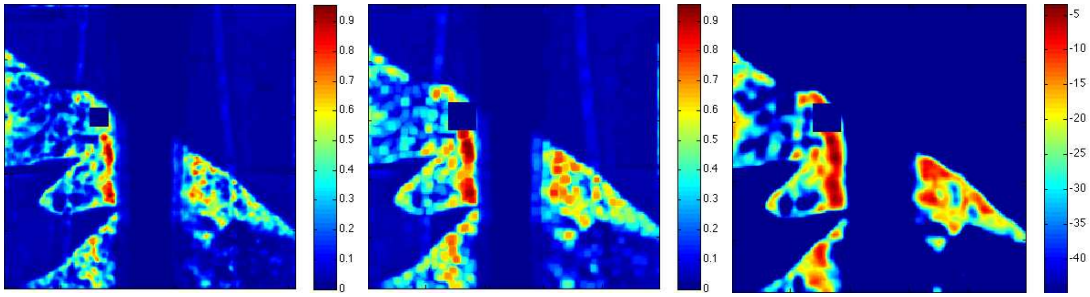


Figure 4.11. Left. Message $\eta_{D_{S,k_\bullet} \to L_{S,k_\bullet}}(L_{S,k_\bullet})$ passed from node $D_{S,k_\bullet}$ to $L_{S,k_\bullet}$ for all possible values of $L_{S,k_\bullet}$. Middle. Message $\eta_{L_{S,k_\bullet} \to C_S}(C_S)$ passed from node $L_{S,k_\bullet}$ to C_S for all possible values of C_S. Right. Value of (4.19) for all possible values of C_S.

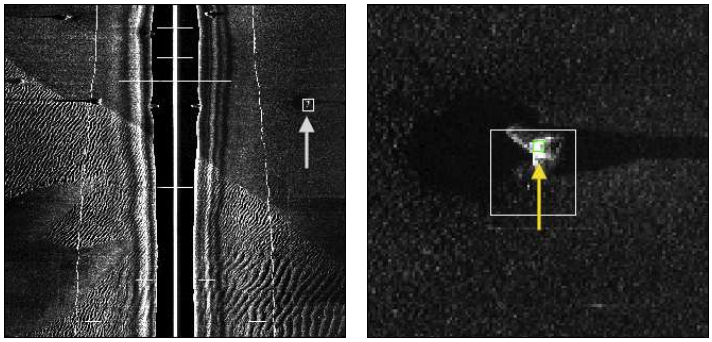


Figure 4.12. Left. Side-looking sonar image and query signal Q. Right. Selected $P_{Q,k_\bullet}$ patch inside the query signal Q.

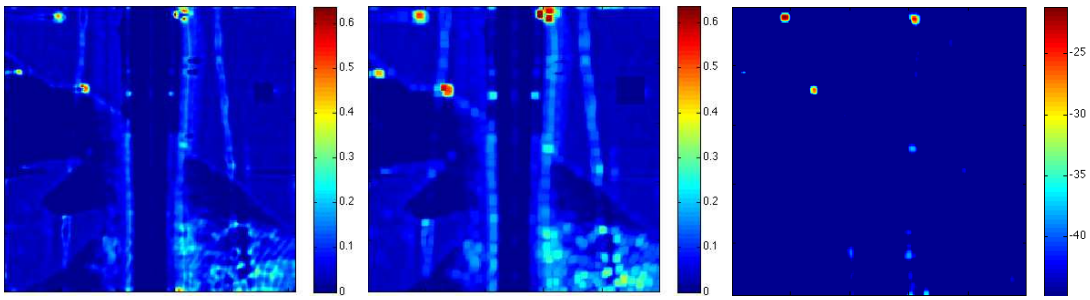


Figure 4.13. Left. Message $\eta_{D_{S,k_\bullet} \to L_{S,k_\bullet}}(L_{S,k_\bullet})$ passed from node $D_{S,k_\bullet}$ to $L_{S,k_\bullet}$ for all possible values of $L_{S,k_\bullet}$. Middle. Message $\eta_{L_{S,k_\bullet} \to C_S}(C_S)$ passed from node $L_{S,k_\bullet}$ to C_S for all possible values of C_S. Right. Value of (4.19) for all possible values of C_S.

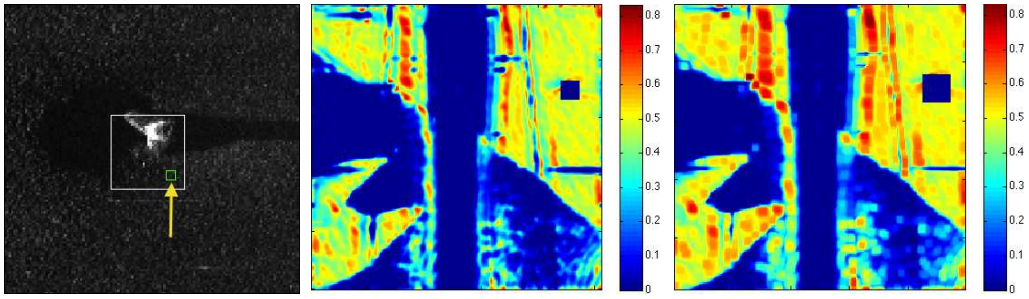


Figure 4.14. Left. Selected $P_{Q,k_\bullet}$ patch inside the query signal Q. Middle. Message $\eta_{D_{S,k_\bullet} \to L_{S,k_\bullet}}(L_{S,k_\bullet})$ passed from node $D_{S,k_\bullet}$ to $L_{S,k_\bullet}$ for all possible values of $L_{S,k_\bullet}$. Right. Message $\eta_{L_{S,k_\bullet} \to C_S}(C_S)$ passed from node $L_{S,k_\bullet}$ to C_S for all possible values of C_S.

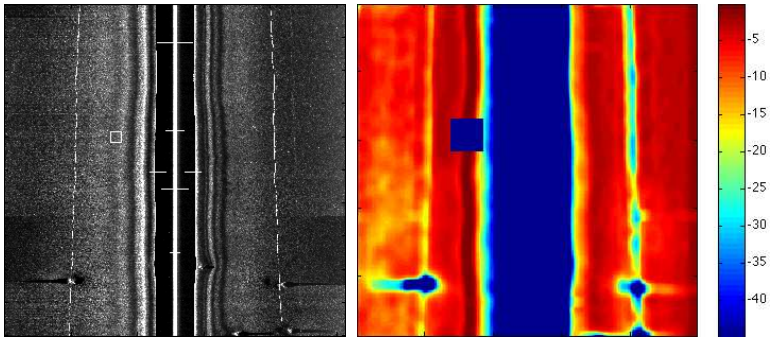


Figure 4.15. Left. Query signal. Right. Value of (4.19) *for all possible values of* C_S.

images and as many query signals Q (one per image). The reference R is defined as the rest of the sonar images the query signals come from. The reference is therefore different for each query signal. The query signals are depicted through Figures 4.15 to 4.23. The value of (4.19) for all possible values of C_S is also depicted through Figures 4.15 to 4.23. We used the same parameters as those listed in Section 4.1.5. The size of a query signal is 32×32, the size of a patch is 4×4, and the size of the region $\mathcal{R}$ in (4.16) is 4×4. The patches do not overlap each other in the query. The covariance matrix Σ in (4.17) was chosen diagonal, the diagonal elements being all equal to 0.1. We also used $K = 4$, $M = 6$ and $N = 1$ to compute the discrete scattering transform. We observe that flat areas and areas with sand ripples are well identified by the composition process with a separation of approximately 20 to 25 dB. We specifically observe in Figure 4.17 that the composition process distinguishes between small and large sand ripples, the large sand ripples being located on the bottom-left part and on the top-right part of the sonar image. We did not quantitatively evaluate the seabed identification performance obtained with the composition process the same way we evaluated, in Chapter 3, the seabed identification performance obtained with the classification algorithm based on affine spaces. The composition process is computationally very expensive and cannot be run in a reasonable amount of time with single-threaded computer programs. It takes indeed more than a few days to analyse a side-looking sonar image made of 1000×1024 pixels. At this point, it is important to realise that the composition process can be run on multi-threaded computer programs because of its algorithmic structure. We successfully wrote a computer program based on the GPU (the computer's graphics processing unit) which implements the composition process. The GPU-based computer program is multi-threaded, and uses the CUDA computing architecture.

4.3 Clutter rejection

In this section, we consider the problem of reducing, in the form of a post-processing process, the number of false positives within a set of detections provided by a target detection algorithm, that being achieved by analysing the amount of

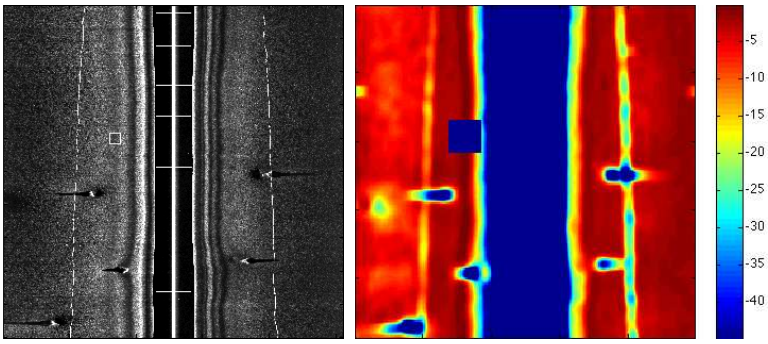


Figure 4.16. Left. Query signal. Right. Value of (4.19) *for all possible values of* C_S.

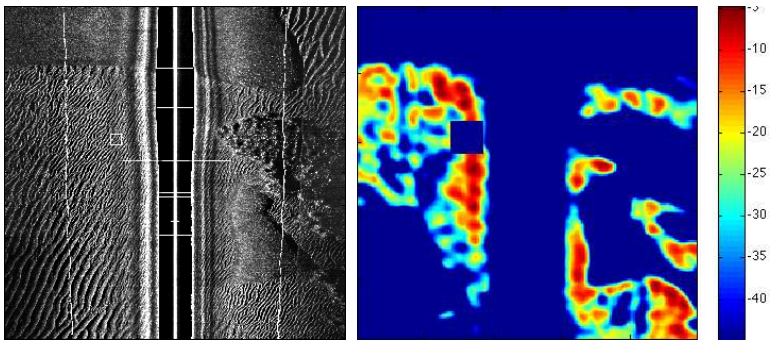


Figure 4.17. Left. Query signal. Right. Value of (4.19) *for all possible values of* C_S.

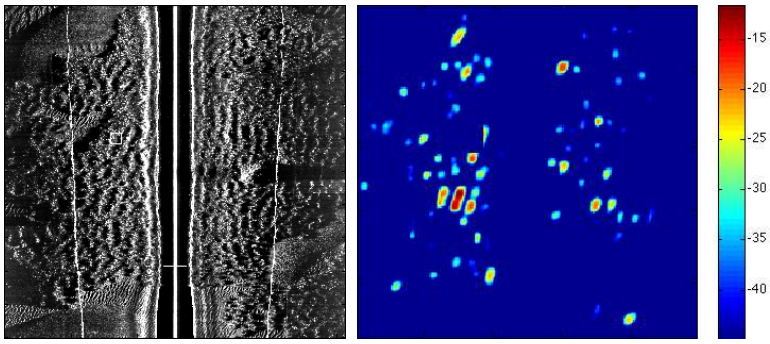


Figure 4.18. Left. Query signal. Right. Value of (4.19) *for all possible values of* C_S.

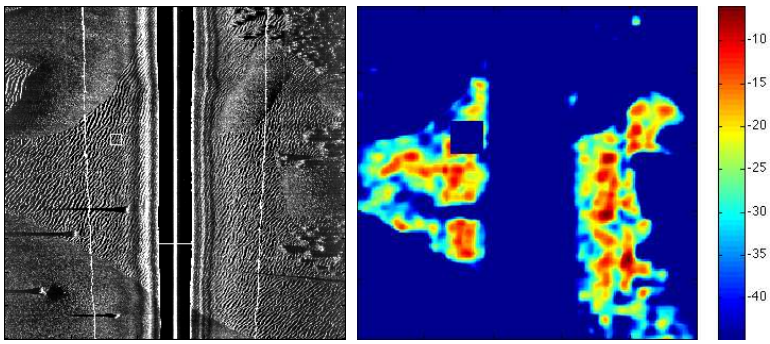


Figure 4.19. Left. Query signal. Right. Value of (4.19) *for all possible values of* C_S.

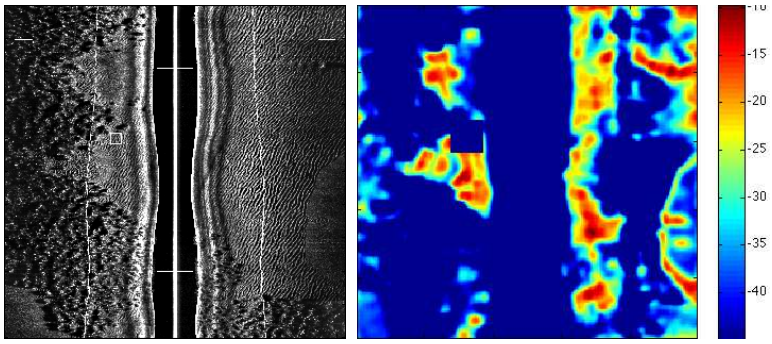


Figure 4.20. Left. Query signal. Right. Value of (4.19) *for all possible values of* C_S.

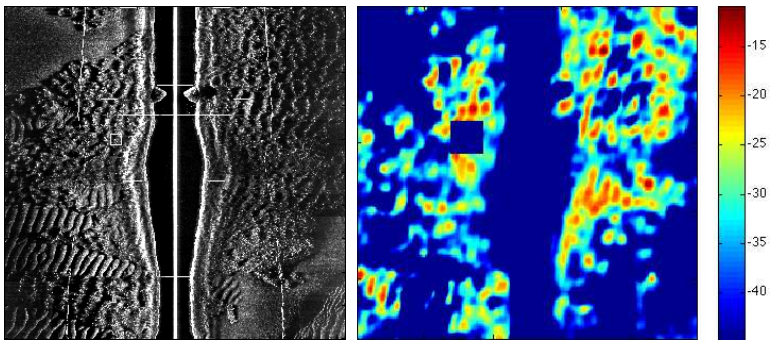


Figure 4.21. Left. Query signal. Right. Value of (4.19) *for all possible values of* C_S.

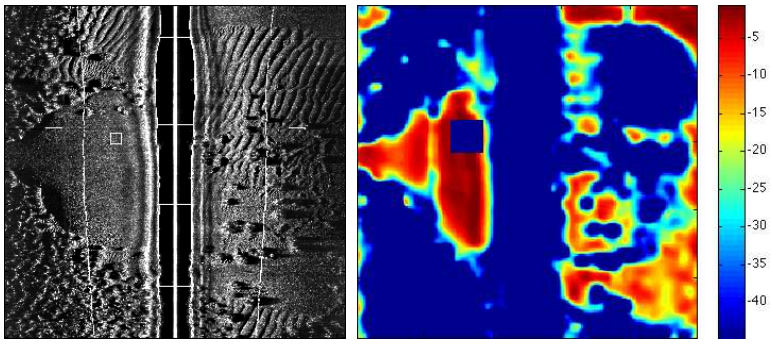


Figure 4.22. Left. Query signal. Right. Value of (4.19) *for all possible values of* C_S.

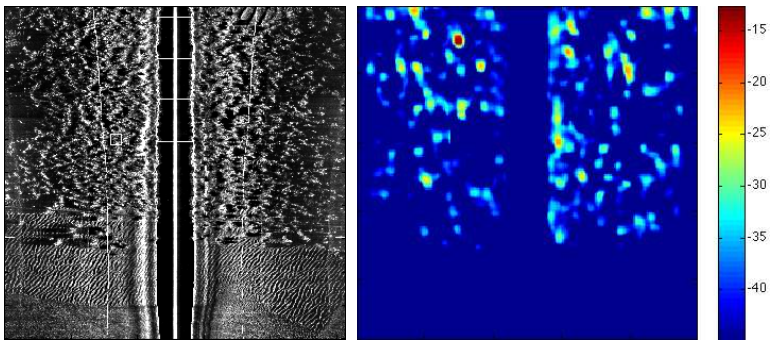


Figure 4.23. Left. Query signal. Right. Value of (4.19) *for all possible values of* C_S.

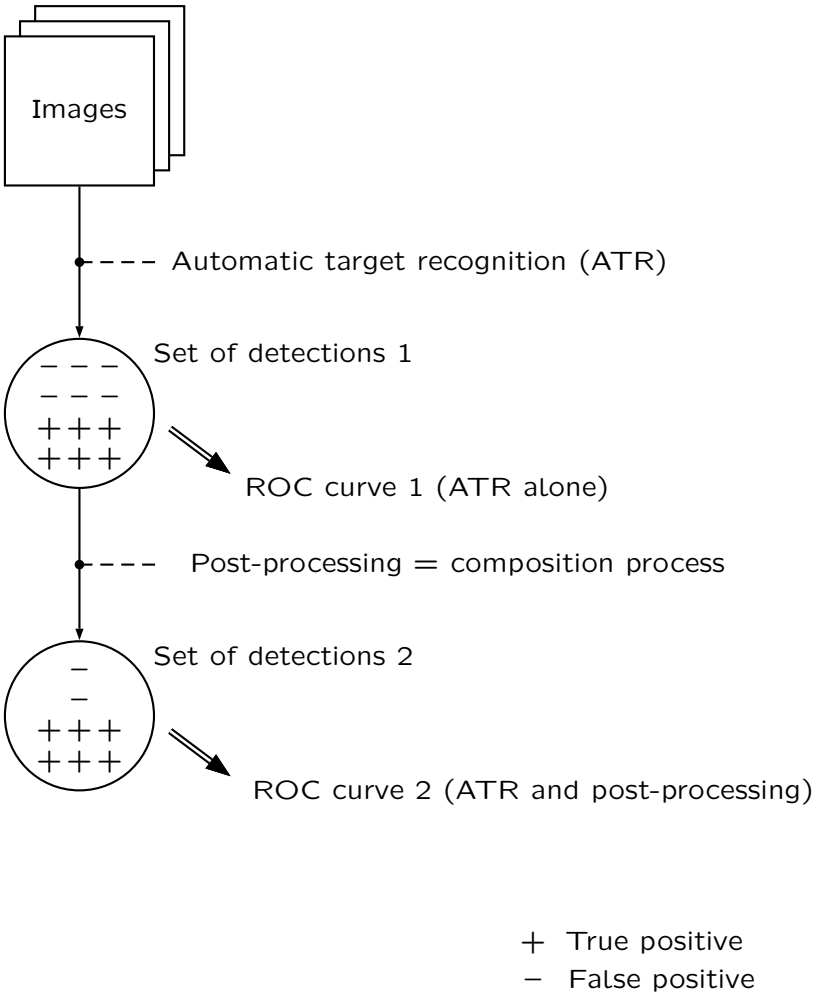


Figure 4.24. Automatic target detection and post-processing chain.

similarity between the detections and the seabed. The amount of similarity is measured with the help of the composition process, as illustrated in Figure 3.15. We used the same data set of 155 side-looking sonar images which was presented in Section 3.3.2 of Chapter 3. The reference associated to a detection is the rest of the sonar image the detection comes from. The amount of similarity between the detection and the reference is defined as the maximum value of (4.19) which is thresholded in order to decide whether the detection is similar to the seabed or not. We used the same parameters as those listed in Section 4.1.5. The detections that are declared similar to the seabed are discarded from the initial set of detections. The receiver operating characteristic curves of the combined target detector and post-processing process are depicted in Figure 4.25. We observe that the target recognition algorithm is not improved by the post-processing process. The reason why this is the case is because there is no separation between the similarity of the true positives and the false positives, as illustrated in Figure 4.26.

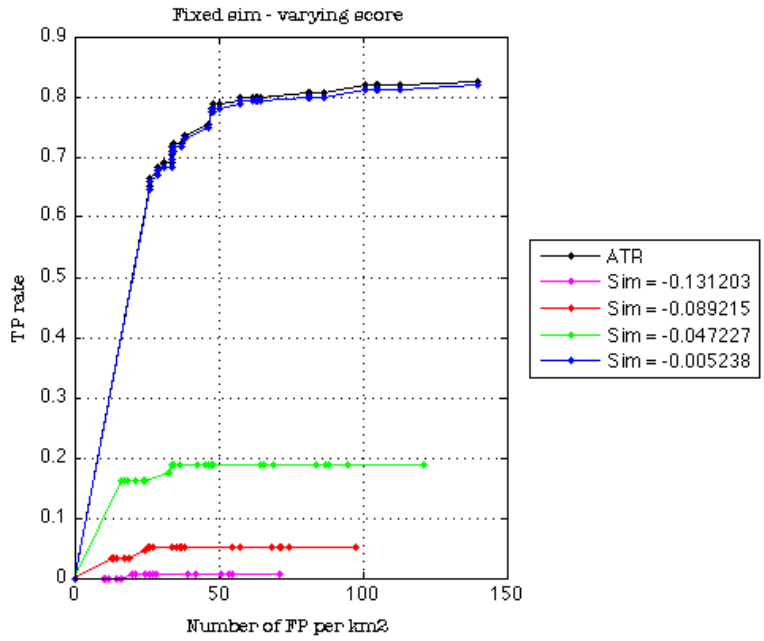


Figure 4.25. Receiver operating characteristic curves of the combined target detector and post-processing process.

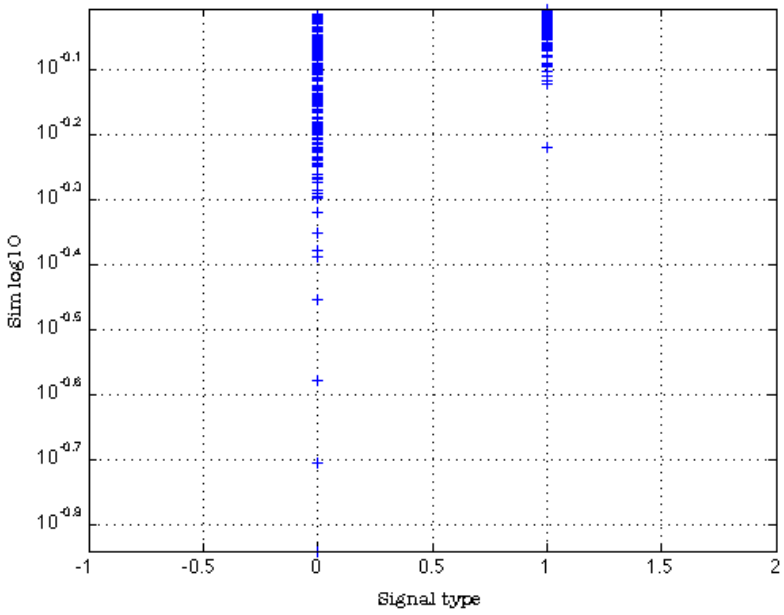


Figure 4.26. Similarity for the true and the false positives.

4.4 Conclusions and perspectives

This chapter was concerned with the analysis of side-looking sonar images with an inference algorithm based on the concept of visual composition. Query signals are represented by a geometrical arrangement of patches. The query signals, in other words, the arrangements of patches themselves, are compared to a reference. The nature of the reference and the sort of comparison that is made between the query signals and the reference depend on the inference problem at stake. In a classification problem, the objective is to make a decision upon the type of seabed of the query signals. In such a case, the reference can be an external collection of know types of seabed. In this chapter, we rather chose to analyse the query signals with respect to the sonar images they come from, so as to demonstrate the how flexible the inference algorithm can be. In an anomaly detection problem, the objective is to decide whether the query signals are at all part of the seabed. Our original hypothesis is that the objects can be revealed as local anomalies with respect to the seabed. We hence defined the reference as the portion of the sonar images immediately around the query signals.

The achievements of this chapter are the following. We re-wrote the theory behind the statistical composition process to make it look like the theory behind hidden Markov models for the mathematical formalism of the two are very much alike. We did so because we found that the leading papers on the composition process were written in a rather informal fashion. In this chapter, we also successfully implemented the composition process in C in two separate ways. One way and the first makes use of a single-threaded computer program which runs on the CPU (the computer's central processing unit). The other way makes use of a multi-threaded computer program which runs on the GPU (the computer's graphics processing unit). We find it important to mention that to write the multi-threaded computer program from the single-threaded computer program was straightforward. We used the C CUDA architecture for this purpose. We had to write a multi-threaded computer program because of the computational complexity of the composition process.

In this chapter, we applied the inference algorithm to the characterisation of the seabed. We only did so in a qualitative way rather than in a quantitative way because of time constraints. We draw three conclusions for this piece of work. First, we acknowledge once more the potential of scattering operators for the description of the types of seabed. The comparison of the descriptors of the patches during the first step of the composition process lead to neat segmentations of the side-looking sonar images when they are made of flat areas and areas with sand ripples. Second, we observe that the inference algorithm presented in this chapter is good at describing flat types of seabed and sand ripples. In the latter case, small sand ripples are not mistaken from large sand ripples because their representation differ. Should we go back to what was written in Chapter 2 about the Littlewood Paley wavelet transform, we may remember that small and large sand ripples respond to atoms at two different scales. This is why small and large sand ripples are identified as different by the composition process. We ought to mention

that such a result differs from the one of Chapter 3 where small and large sand ripples were considered as part of the same class and identified as such. Third, we cannot say for certain that the composition process is good at characterising areas with vegetation. It seems to be the case but only a quantitative analysis such as the one conducted in Chapter 3 would be able to tell for certain. In spite of the good seabed characterisation obtained with the composition process, we would still recommend the inference algorithm based on affine spaces for such a purpose. The latter approach is indeed a lot less computationally expensive, and very stable with respect to changes in the parameters it depends upon.

Moving beyond the identification of the types of seabed, we believe that the introduction of query signals, references, and ensemble of patches opens the door to new ways of analysing side-looking sonar images in an unsupervised fashion. Very often one has to assess whether two portions of the same sonar image are similar without relying on a training set. Very often as well one has to adjust any sort of processing to the types of seabed that are present in the sonar images. This can be easily handled once the query signals, the references, and the ensemble of patches are properly defined. Besides, if in this chapter the query signals are represented by a geometrical arrangement of patches, the query signals can be represented by other sorts of ensemble of patches, most of which are detailed in the leading papers on the composition process. For instance, the algorithm presented in Chapter 1 does not account for the relative arrangement between the patches, and still provides satisfactory results.

This chapter was also concerned with the reduction, in the form of a post-processing process of the number of false positives in a set of detections provided by a target detector. Such a task is commonly performed through the extraction of the highlight and the shadow of the detections, and the analysis of these two with respect to templates. Our original hypothesis, as stated in Chapter 1, is that it is possible to reduce the number of false positives without affecting the number of true positives should we quantify the similarity between the detections and the seabed, the false positives being by nature part of the seabed. This was cast in this chapter as an anomaly detection problem. We showed that our original hypothesis was wrong when the detections are represented by a geometrical ensemble of patches.

Chapter 5

Conclusions

"'Begin at the beginning' the King said gravely, 'and go on till you come to the end: then stop.'"

Lewis Carroll, Alice in Wonderland

We have now reached the last chapter of the thesis. It is hence time to return to the problems we looked at as they were stated in the very first chapter of the thesis. The two problems which were dealt with in the thesis are the supervised identification of the types of seabed that are present in side-looking sonar images, and the detection of man-made objects in side-looking sonar images. These images are, broadly speaking, the result of the physical interaction between acoustic waves and the bottom of the sea. Because of this interaction, the types of seabed appear as textured areas in side-looking sonar images. The texture descriptors commonly used in the field of sonar imagery fail at accurately identifying the types of seabed because the types of seabed, hence the textures, are extremely variable. In this thesis, we used scattering operators over more traditional texture descriptors to represent the types of seabed. We assessed how well the types of seabed are identified with the help of two inference algorithms, one based on affine spaces, and the other based on the concept of similarity by composition. What really is important is indeed the combination of a seabed representation and an inference algorithm rather than a seabed representation alone, which is why we needed at least one inference algorithm to assess how well scattering operators represent the types of seabed. The thesis was also concerned with the detection of man-made objects in side-looking sonar images. An object detector may be described as a method or a succession of processes which, when applied to a certain number of sonar images, produces a set of detections. Some of these are true positives, and correspond to real objects while others are false positives, and do not correspond to real objects. The false positives are therefore part of the seabed by nature. The present object detectors suffer from a high false positive rate in complex environments, that is to say, complex types of seabed. The hypothesis we followed in the thesis is that it is possible to reduce the number of false positives through a characterisation of the similarity between the detections and the seabed, the

false positives being by nature part of the seabed. We used scattering operators to represent the detections and the same two aforementioned inference algorithms to quantify how similar the detections are to the seabed.

In the thesis, we showed that scattering operators represent extremely well the types of seabed which usually are present in side-looking sonar images. The types of seabed that were considered were flat areas and areas made of sand ripples or vegetation. With the inference algorithm based on affine spaces, which only depends on a few parameters, we showed that the characterisation of the seabed is excellent and that it remains as good across a wide range of parameters. Such a behaviour is of much interest for our experience tells us that most of the time, inference algorithms are very sensitive to the values assigned to the parameters they depend upon. We also showed that the characterisation of the seabed is not affected by the presence of artefacts in the side-looking sonar images. The types of seabed are for instance not mistaken one from another around the vertical white lines which correspond to the acoustic returns from the surface of the water. We finally showed that only a small training set was needed to obtain such an excellent characterisation of the seabed. By small training set, we mean that portions of only two images were used as a training set to classify more than two hundred images. This demonstrated the generalisation capabilities of the signal representation induced by scattering operators. Scattering operators were compared to local binary pattern of which the texture descriptors commonly used in the field of sonar imagery are a special case. We showed that scattering operators outperform local binary patterns by about 10 %. Specifically, sand ripples are not mistaken for seaweed as much, which is a recurrent issue when one intends to identify the types of seabed. With the inference algorithm based on the concept of similarity by composition, the characterisation of the seabed proved to be good, mostly over flat types of seabed and sand ripples, although that was only addressed in a qualitative fashion. The inference algorithm based on the concept of similarity by composition is a lot more algorithmically complex than the one based on affine spaces. So should we were to continue the work on the characterisation of the seabed with the former algorithm, we would keep the concept of ensemble of patches, maybe the geometrical arrangement of the patches, but make the ensemble computationally lighter by considering adaptive representations. What we have in mind is something along the lines of the sparse representations or the basis pursuit algorithm. What we are saying is that we would completely think the concept of ensemble of patches over.

In the thesis, we also showed that our original hypothesis, according to which it is possible to reveal objects as anomalies with respect to the seabed, did not work with the inference algorithm based on affine spaces and with the one based on the concept of similarity by composition. The target detector we used is one of the best available at the moment so that what we are saying is that we could not, following our hypothesis, do better than one of the best target detectors available at the moment. We may at this point wonder whether our hypothesis can ever work and improve the best target detectors. We believe that this is the case but that to find the right signal representation and inference algorithm for

this purpose is an extremely hard task. So should we were to find these two, we would consider adaptive compact representations in order to find the key features of either the detections or the seabed. We once again think of something along the lines of the sparse representations or the basis pursuit algorithm.

List of publications

[1] N. Valeyrie, Y. Pailhas, and Y. Petillot, "Texture recognition in synthetic aperture sonar images with scattering operators," in *Underwater Acoustic Measurements: Technologies and Results 4th International Conference and Exhibition*, 2011.

[2] N. Valeyrie and Y. Petillot, "Target Detection in Side-Looking Sonar Images With Scattering Operators," in *Proceedings of the European Conference on Underwater Acoustics*, 2012.

[3] ——, "Side-Looking Sonar Supervised Seabed Classification With Scattering Operators," *Submitted to IEEE Transactions on Geoscience and Remote Sensing*, 2013.

Bibliography

[4] D. Middleton, *An Introduction to Statistical Communication Theory*. Wiley-Interscience, A John Wiley & Sons, Inc., Publication, 1996.

[5] ——, "A Statistical Theory of Reverberation and Similar First-Order Scattered Fields. Part I: Waveforms and the General Process," *IEEE Transactions on Information Theory*, vol. 13, no. 3, pp. 372–392, July 1967.

[6] M. Hayes and P. Gough, "Synthetic aperture sonar: A review of current status," *Oceanic Engineering, IEEE Journal of*, vol. 34, no. 3, pp. 207–224, July 2009.

[7] R. J. Urick, *Principles of Underwater Sound*, 3rd ed. McGraw-Hill, 1983.

[8] APL-UW, "High-frequency environmental acoustic models handbook," University of Washington, Tech. Rep., 1994.

[9] G. Dobeck, J. Hyland, and L. Smedley, "Automated detection and classification of sea mines in sonar imagery," in *Proceedings of SPIE, the International Society for Optical Engineering*, vol. 3070, 1997, pp. 90–110.

[10] G. Dobeck, "Fusing sonar images for mine detection and classification," in *Proceedings of SPIE, the International Society for Optical Engineering*, vol. 3710, 1999, pp. 602–614.

[11] V. Chandran, S. Elgar, and A. Nguyen, "Detection of mines in acoustic images using highorder spectral features," *Oceanic Engineering, IEEE Journal of*, 2002.

[12] M. Doherty, J. Landowski, P. Maynard, G. Uber, D. Fries, and F. Maltz, "Side scan sonar object classification algorithms," in *Unmanned Untethered Submersible Technology, 1989. Proceedings of the 6th International Symposium on*, 1989.

[13] C. Ciany and J. Huang, "Computer aided detection/computer aided classification and data fusion algorithms for automated detection and classification of underwater mines," in *Proceedings of the Oceans conference*, vol. 1, 2000, pp. 277–284.

[14] S. Johnson and M. Deaett, "The application of automated recognition techniques to side-scan sonar imagery," *Oceanic Engineering, IEEE Journal of*, vol. 19, no. 1, pp. 138–144, Jan 1994.

[15] T. Aridgides, M. Fernandez, and G. Dobeck, "Adaptive 3-dimensional range-crossrange-frequency filter processing string for sea mine classification in side scan sonar imagery," in *Proceedings of the SPIE Conference*, 1997.

[16] C. Ciany and W. Zurawski, "Performance of computer aided detection/computer aided classification of underwater mines," in *CAD/CAC conference*, 2001.

[17] ——, "Performance of fusion algorithms for computer aided detection and classification of bottom mines in the shallow water environment," in *Proceedings of the Oceans conference*, vol. 4, Oct. 2002, pp. 2164–2167.

[18] T. Aridgides, M. Fernandez, and G. Dobeck, "Fusion of adaptive algorithms for the classification of sea mines using high resolution side scan sonar in very shallow water," in *Proceedings of the Oceans conference*, vol. 1, 2001, pp. 135–142.

[19] G. Dobeck, "Algorithm fusion for automated sea mine detection and classification," in *Proceedings of the Oceans conference*, vol. 1, 2001, pp. 130–134.

[20] F. Maussang, J. Chanussot, A. Hetet, and M. Amate, "Mean–standard deviation representation of sonar images for echo detection: Application to sas images," *IEEE Journal of Oceanic Engineering*, vol. 32, no. 4, pp. 956–970, Oct. 2007.

[21] F. Maussang, M. Rombaut, J. Chanussot, A. Hétet, and M. Amate, "Fusion of local statistical parameters for buried underwater mine detection in sonar imaging," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, pp. 1–19, 2008.

[22] L. Linnett, S. Clarke, C. Reid, and A. Tress, "Monitoring of the seabed using sidescan sonar and fractal processing," in *Proceedings of the Underwater Acoustics Group*, I. of Acoustics, Ed., vol. 15, 1993, pp. 49–64.

[23] L. Linnett, D. Carmichael, S. Clarke, and A. Tress, "Texture Analysis of Sidescan Sonar Data," in *Texture analysis in radar and sonar, IEE Seminar on*, November 1993, pp. 1–6.

[24] B. Calder, L. Linnett, and D. Carmichael, "Spatial stochastic models seabed object detection," in *Proceedings of SPIE, the International Society for Optical Engineering*, 1993.

[25] C. Rao, K. Mukherjee, S. Gupta, A. Ray, and S. Phoha, "Underwater mine detection using symbolic pattern analysis of sidescan sonar images," in *Proceesings of the American Control Conference*, 2009.

[26] P. Saisan and S. Kadambe, "Shape normalized subspace analysis for underwater mine detection," in *Image Processing, International Conference on*, Oct. 2008, pp. 1892–1895.

[27] J. Sawas and Y. Petillot, Y. Pailhas, "Target detection and classification using cascades of boosted classifiers," in *European Conference on Underwater Acoustics*, 2010.

[28] E. Dura, Y. Zhang, X. Liao, G. Dobeck, and L. Carin, "Active learning for detection of mine-like objects in side-scan sonar imagery," *IEEE Journal of Oceanic Engineering*, vol. 30, no. 2, pp. 360–371, 2005.

[29] R. Grasso and F. Spina, "Small bottom object density analysis from side scan sonar data by a mathematical morphology detector," in *Information Fusion, International Conference on*, 2006.

[30] M. An, T. TCobb, B. Shenefelt, and R. Tolimieri, "Advances in group filter applications to sea mine detection," in *Proceedings of the Oceans Conference*, Sept. 2006, pp. 1–5.

[31] J. D. Tucker, M. Azimi-Sadjadi, and G. Dobeck, "Coherent-based method for detection of underwater objects from sonar imagery," p. 65530U, 2007.

[32] Y. Zhao, N. Wachowski, and M. Azimi-Sadjadi, "Target coherence analysis using canonical correlation decomposition for sas data," in *Proceedings of the Oceans conference*, 2009.

[33] B. Calder, L. Linnett, and D. Carmichael, "A bayesian approach to object detection in sidescan sonar," in *Proceedings of the International Conference on Image Processing and Its Applications*, vol. 2, Jul 1997, pp. 857–861 vol.2.

[34] S. Reed, Y. Petillot, and J. Bell, "An automatic approach to the detection and extraction of mine features in sidescan sonar," *Oceanic Engineering, IEEE Journal of*, vol. 28, no. 1, pp. 90–105, Jan 2003.

[35] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, "A comparative study of energy minimization methods for markov random fields with smoothness-based priors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, no. 6, pp. 1068–1080, June 2008.

[36] V. Kolmogorov and R. Zabin, "What energy functions can be minimized via graph cuts?" *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 2, pp. 147–159, Feb. 2004.

[37] S. Reed, Y. Petillot, and J. Bell, "Model-based approach to the detection and classification of mines in sidescan sonar," *Appl. Opt.*, vol. 43, no. 2, pp. 237–246, 2004.

[38] S. Perry and L. Guan, "Pulse-length-tolerant features and detectors for sector-scan sonar imagery," *IEEE Journal of Oceanic Engineering*, vol. 29, no. 1, pp. 138–156, 2004.

[39] B. Stage and B. Zerr, "Detection of objects on the sea bottom using backscattering characteristics dependent on the observation point," *Oceanic Engineering, IEEE Journal of*, 1997.

[40] M. Brissette, J. Hughes-Clarke, J. Bradford, and B. MacGowan, "Detecting small seabed targets using a high frequency multibeam sonar: geometric models and test results," in *Proceedings of the Oceans Conference*, 1997.

[41] S. Reed, Y. Petillot, and J. Bell, "Automated approach to classification of mine-like objects in sidescan sonar using highlight and shadow information," *Radar, Sonar and Navigation, IEE Proceedings*, vol. 151, no. 1, pp. 48–56, Feb 2004.

[42] E. Dura, J. Bell, and D. Lane, "Superellipse fitting for the recovery and classification of mine-like shapes in sidescan sonar images," *IEEE Journal of Oceanic Engineering*, vol. 33, no. 4, pp. 434–444, Oct. 2008.

[43] P. Pidsley and M. Way, "Processing for cad and image based cac," in *Proceedings of the CAD/CAC conference*, 2001.

[44] J. Del Rio Vera, E. Coiras, J. Groen, and B. Evans, "Automatic target recognition in synthetic aperture sonar images based on geometrical feature extraction," *EURASIP Journal on Advances in Signal Processing*, pp. 1–9, 2009.

[45] J. Groen, E. Coiras, J. Del Rio Vera, and B. Evans, "Model-based sea mine classification with synthetic aperture sonar," *Radar, Sonar and Navigation, IET*, 2010.

[46] R. Laterveer, D. Hughes, and S. Dugelay, "Markov random fields for target classification in low frequency sonar," in *Proceedings of the Oceans Conference*, 1998, pp. 1274–1278.

[47] E. Hasanbelliu, J. Principe, and C. Slatton, "Correntropy based matched filtering for classification in sidescan sonar imagery," in *Systems, Man and Cybernetics, 2009, IEEE International Conference on*, 2009, pp. 2757–2762.

[48] G. Foresti, V. Murino, C. Regazzonu, and A. Trucco, "A voting-based approach for fast object recognition in underwater acoustic images," *Oceanic Engineering, IEEE Journal of*, vol. 22, no. 1, pp. 57–65, 1997.

[49] M. Azimi-Sadjadi, D. Yao, Q. Huang, and G. Dobeck, "Underwater target classification using wavelet packets and neural networks," *Neural Networks, IEEE Transactions on*, vol. 11, no. 3, pp. 784–794, May 2000.

[50] A. Pezeshki, M. Azimi-Sadjadi, and L. Scharf, "Undersea target classification using canonical correlation analysis," *Oceanic Engineering, IEEE Journal of*, vol. 32, no. 4, pp. 948 –955, Oct. 2007.

[51] J. Groen, D. Williams, and W. Fox, "A Model-Bases Multi-View Image Registration Method For SAS Images," in *European Conference on Underwater Acoustics*, 2011.

[52] D. Williams, "Bayesian data fusion of multiview synthetic aperture sonar imagery for seabed classification," *IEEE Transactions on Image Processing*, vol. 18, no. 6, 2009.

[53] J. Sawas, "Personal communication," 2010.

[54] S. Reed, I. Ruiz, C. Capus, and Y. Petillot, "The Fusion of Large Scale Classified Side-scan Sonar Image Mosaics," *Image Processing, IEEE Transactions on*, vol. 15, no. 7, pp. 2049–2060, July 2006.

[55] M. Lianantonakis and Y. Petillot, "Sidescan Sonar Segmentation Using Texture Descriptors and Active Contours," *Oceanic Engineering, IEEE Journal of*, vol. 32, no. 3, pp. 744–752, July 2007.

[56] Y. Petillot, S. Reed, and E. Coiras, "A framework for evaluating underwater mine detection and classification algorithms using augmented reality," in *Proceedings of the Undersea Defence Technology*, 2006.

[57] O. Boiman and M. Irani, "Detecting irregularities in images and in video," *International Journal of Computer Vision*, vol. 74, no. 1, pp. 17–31, August 2007.

[58] O. Boiman, "Inference by composition," Ph.D. dissertation, Faculty of mathematics and computer science, the Weizman institute of science, March 2009.

[59] M. Unser, "Sum and Difference Histograms for Texture Classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 1, pp. 118–125, Jan. 1986.

[60] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *Syst., Man, Cybern., IEEE transactions on*, vol. 3, no. 6, pp. 610–621, Nov. 1973.

[61] R. Haralick, "Statistical and Structural Approaches to Texture," *Proc. IEEE*, vol. 67, no. 5, pp. 786–804, May 1979.

[62] I. Karoui, J.-M. Boucher, R. Fablet, and J.-M. Augustin, "Multiscale Segmentation of Textured Sonar Images Using Co-Occurrence Statistics," in *Image Processing, International Conference on*, vol. 3, Oct. 2004, pp. 1691–1694.

[63] I. Karoui, R. Fablet, J.-M. Boucher, and J.-M. Augustin, "Seabed Segmentation Using Optimized Statistics of Sonar Textures," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, no. 6, pp. 1621–1631, June 2009.

[64] L. Linnett, D. Carmichael, and S. Clarke, "Texture Classification Using a Spatial-Point Process Model," in *Vision, Image and Signal Processing, IEE Proceedings*, vol. 142, no. 1, 1995, pp. 1–6.

[65] Julesz, "Textons, the Element of Texture Perception, and Their Interactions," *Nature*, vol. 290, pp. 91–97, 1981.

[66] Beck, J. and Hope, B. and Rosenfeld, A., Ed., *Human and Machine Vision*. Academic Press, 1983.

[67] T. Ojala, M. Pietikäinen, and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Feature Distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[68] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971 –987, Jul. 2002.

[69] B. Torréssani, *Analyse Continue par Ondelettes*. EDP Sciences, 1995.

[70] S. Mallat, *A Wavelet Tour of Signal Processing*, 3rd ed. Academic Press, 2009.

[71] D. Williams, "Unsupervised Seabed Segmentation of Synthetic Aperture Sonar Imagery via Wavelet Features and Spectral Clustering," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, Nov. 2009, pp. 557–560.

[72] E. Coiras, V. Myers, and B. Evans, "Reliable Seabed Characterization for MCM Operations," in *Proceedings of the Oceans Conference*, October 2007, pp. 1–4.

[73] D. Carmichael, L. Linnett, S. Clarke, and B. Calder, "Seabed Classification Through Multifractal Analysis of Sidescan Sonar Imagery," *Radar, Sonar and Navigation, IEE Proceedings*, vol. 143, no. 3, pp. 140–148, June 1996.

[74] S. Mallat, "Group Invariant Scattering," *To appear in Communications in Pure and Applied Mathematics*, Apr. 2012.

[75] Y. LeCun, K. Kavukvuoglu, and C. Farabet, "Convolutional Networks and Applications in Vision," in *Proc. International Symposium on Circuits and Systems*, 2010.

[76] J. Bruna and S. Mallat, "Classification with Scattering Operators," in *Computer Vision and Pattern Recognition, IEEE Conference on*, June 2011.

[77] ——, "Invariant Scattering Convolution Network," *Submitted to IEEE Trans. Pattern Anal. Mach. Intell.*, Feb. 2012.

[78] Andèn, J. and Mallat, S.G., "Multiscale Scattering for Audio Classification," in *Proceedings of the International Society for Music Information Retrieval*, 2011.

[79] ——, "Scattering Representation of Modulated Sounds," in *Proceedings of the Digital Audio Effects (DAFx) conference*, 2012.

[80] S. Mallat and F. Bergeaud, "Matching Pursuit of Images," *Computational and Applied Mathematics*, vol. 15, no. 2, 1996.

[81] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *Information Theory, IEEE Transactions on*, vol. 47, no. 7, pp. 2845–2862, 2001.

[82] D. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[83] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summurazing Visual Data Using Bidirectional Similarity," in *CVPR'08*, 2008.

[84] D. Lowe, "Object Recognition from Local Scale-Invariant Features," in *Proceedings of the International Conference on Computer Vision*, vol. 2, Sept. 1999, pp. 1150–1157.

[85] V. Barnett and T. Lewis, *Outliers in Statistical Data*. Wiley series in probability and mathematical statistics, 1978.

[86] L. Tarassenko, P. Hayton, N. Cerneaz, and M. Brady, "Novelty Detection for the Identification of Masses in Mammograms," in *Proceedings of the Fourth International Conference on Artificial Neural Networks*, 1995, pp. 442–447.

[87] A. Ypma and R. Duin, "Support Objects for Domain Approximation," in *Proceedings of the eighth International Conference on Artificial Neural Networks*, 1998.

[88] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. John Wiley & Sons, 1973.

[89] D. Tax and R. Duin, "Support Vector Data Description," *Journal of Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004.

[90] L. Manevitz and M. Yousef, "One-Class SVMs for Document Classification," *Journal of Machine Learning*, pp. 139–154, 2001.

[91] N. Japkowicz, C. Myers, and M. Gluck, "A Novelty Detection Approach to Classification," in *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, 1995, pp. 518–523.

[92] M. Muschietti and B. Torréssani, " Pyramidal Algorithms for Littlewood-Paley Decompositions," *SIAM J. Math. Anal.*, vol. 26, pp. 925–943, 1995.

[93] J.-P. Antoine, "The continuous wavelet transform in image processing," Institut de Physique theorique, universite catholique de Louvain, Tech. Rep., 1998.

[94] J.-P. Kahane and P.-G. Lemarié-Rieusset, *Fourier Series and Wavelets*. Gordon and Breach, 1995.

[95] L. Cohen, *Time-Frequency Analysis*. Prentice Hall Signal Processing Series, 1995.

[96] M. Frigo and S. G. Johnson, "The design and implementation of FFTW3," *Proceedings of the IEEE*, vol. 93, no. 2, pp. 216–231, 2005, special issue on "Program Generation, Optimization, and Platform Adaptation".

[97] J. Bruna, "Personal communication."

[98] K. Barclay, *ANSI C, Problem Solving and Programming*. Prentice Hall, 1990.

[99] G. Golub and C. Van Loan, *Matrix Computations*, 3rd ed. The John Hopkins Universty Press, 1996.

[100] J. Jacod and P. Protter, *Probability Essentials*, 2nd ed. Springer, 2004.

[101] L. Linnett, D. Carmichael, S. Clarke, and A. Tress, "Texture Analysis of Sidescan Sonar Data," in *Texture Analysis in Radar and Sonar, IEE Seminar on*, June 1993.

[102] B. Zerr, E. Maillard, and D. Gueriot, "Sea-Floor Classification by Neural Hybrid System," in *Oceans'94 Conference Proceedings*, vol. 2, 1994, pp. 239–243.

[103] W. Stewart, M. Jiang, and M. Marra, "A Neural Network Approach to the Classification of Sidescan Sonar Imagery from a Midocean Ridge Area," *Oceanic Engineering, IEEE Journal of*, vol. 19, no. 2, pp. 214–224, 1994.

[104] N. Pican, E. Trucco, M. Ross, D. Lane, Y. Petillot, and I. Tena Ruiz, "Texture Analysis for Seabed Classification: Co-occurrence Matrices vs. Self-Organizing Maps," in *Oceans'98 Conference Proceedings*, vol. 1, Sept. 1998, pp. 424–428.

[105] I. Karoui, R. Fablet, J.-M. Boucher, W. Pieczynski, and J.-M. Augustin, "Fusing of Textural Statistics using a Similarity Measure: Application to Texture Recognition and Segmentation," *Journal of Pattern Analysis and Applications*, vol. 11, no. 3, pp. 425–434, 2008.

[106] E. Angerson, Z. Bai, J. Dongarra, A. Greenbaum, A. McKenney, J. Du Croz, S. Hammarling, J. Demmel, C. Bischof, and D. Sorensen, "LAPACK: A Portable Linear Algebra Library for High-Performance Computers," in *Supercomputing '90. Proceedings of*, Nov. 1990, pp. 2–11.

[107] J. Yedidia, W. Freeman, and Y. Weiss, "Understanding Belief Propagation and its Generalizations," in *International Conference on Artificial Intelligence*, 2001.

[108] L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.