Automated SAR Sea Ice Interpretation

by

Qiyao Yu

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Systems Design Engineering

Waterloo, Ontario, Canada, 2006

I hereby declare that I am the sole author of this thesis.

I authorize the University of Waterloo to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize the University of Waterloo to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Abstract

Synthetic aperture radar (SAR) provides an efficient imaging tool for sea ice monitoring, which has found important applications in both scientific and operational activities such as climatic research and ship navigation. Due to the large volume of SAR data acquired, automated computer based interpretation is desired, the goal of which is to segment the image into homogeneous regions and classify each segmented region into the correct ice type. Unfortunately, significant variations exist with respect to the intensity and texture appearance of SAR sea ice due to the numerous imaging parameters and environmental factors, and hence make the task extremely challenging in both the segmentation and classification processes.

The success of the interpretation system is thus largely dependent upon its adaptivity to intensity and texture. On the other hand, models relatively insensitive to intensity and texture are needed for robust descriptions of the ice types. The two issues are manipulated respectively by the two processes of the low level unsupervised segmentation on image pixels and the high level supervised classification on segmented regions. In the latter, features other than the intensity and texture need to be efficiently incorporated, as suggested by the success of human operators in discriminating ice types using additional high level knowledge such as floe shape and existence of fractures. Computations of such features require the low level segmentation to produce regions neither over-segmented nor under-segmented, the balance of which is difficult to achieve due to the complexity of the sea ice scenes and needs to be guided by the high level supervised classification.

This thesis presents a SAR sea ice interpretation system that combines in a bidirectional interactive manner the low level segmentation and high level classification. The segmentation algorithm is a region growing technique and the classification is

based on Markov Random Field (MRF). The two processes are integrated under the Bayesian framework, with both aiming at reducing a defined energy. The overall process keeps refining the segmentation and producing semantic class labels at the same time in an iterative manner. Various features can hence be efficiently combined, and successful results are demonstrated.

Acknowledgments

I would like to express my sincere gratitude to my supervisor Professor David A. Clausi for inviting me to work with his vision and image processing group, for his involvement, patience and insight in my research, and for the help on my life outside school. Without his invaluable support, this work would not have been possible. I am also grateful to my Ph.D committee members, Professor Ed Jernigan, Professor Paul Fieguth, Professor Ed Vrscay, as well as other members in the group, for various enlightening ideas and discussions.

My gratitude should also been given to Dean Flett, Roger DeAbreu, and Matt Arkett in Canadian Ice Service, for their help in my understanding the research problem and providing evaluations on the testing results.

The RADARSAT SAR image is used with the courtesy of Canadian Ice Service, Environment Canada and is copyright Canadian Space Agency. I would also like to thank the support from NSERC and CRYSYS (CRYospheric SYStem in Canada).

Finally, I am deeply indebted to my parents and my wife, Lei, for their love, encouragement and inspiration.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The identification of sea ice types and the evaluation of their properties have found important applications in both scientific and operational activities such as climate research and ship navigation. For consideration of efficiency, expense, and timing requirements, remote sensing techniques are intensively used for monitoring sea ice. During the past decades, much attention has been given to utilizing synthetic aperture radar (SAR) to perform this task. SAR is an active microwave imaging system that transmits a sequence of pulse in a side-looking manner and records the backscatter signals to form a 2-D image. SAR is an appropriate tool for sea ice monitoring for the following three reasons:

- SAR can provide regular imaging of ice fields over extended areas when using satellite platforms.
- SAR has the ability to observe through clouds, and obtain measurements both

day and night.

- The portion of microwave energy returned is largely dependent on the target surface roughness, volume moisture content and electrical properties, which are significantly different among different sea ice types and their snow covers.

Hence the problem arises of how to interpret SAR sea ice imagery accurately. Interpretation of SAR sea ice currently in operation is in the form of ice chart and egg code defined by the World Meteorology Organization (WMO) [111]. In producing the daily ice charts, operators at the Canadian Ice Service (CIS) (or similar agencies of other northern countries) manually select regions visually homogeneous or separated by distinct boundaries. Each selected region may contain a single ice type or several different ice types. The human operator then associates each of the regions with an egg code, which is an oval shape symbol that contains numerical information about the type, concentration (percentage of area coverage), and floe size of the ice components in the specified region. A detailed description can be found at the Canadian Ice Service website [95]. An example SAR sea ice image and the corresponding ice chart and code are given in Figures 1.1 and 1.2 respectively.

Such a form of interpretation has substantial deficiencies. First, the interpretation is not at the pixel level and does not provide location information at the pixel's resolution on the component ice types inside each egg code region. Second, the regions with the egg code are generated subjectively and could be inconsistent among individual operators, and so are the region boundaries. Third, the increasing amount of SAR data demands increasingly higher throughput of the interpretation system, and hence manual procedures are insufficient for future needs.

Therefore, automated computer-based SAR sea ice interpretation is desired. The interpretation system segments the SAR sea ice image into regions of homo-

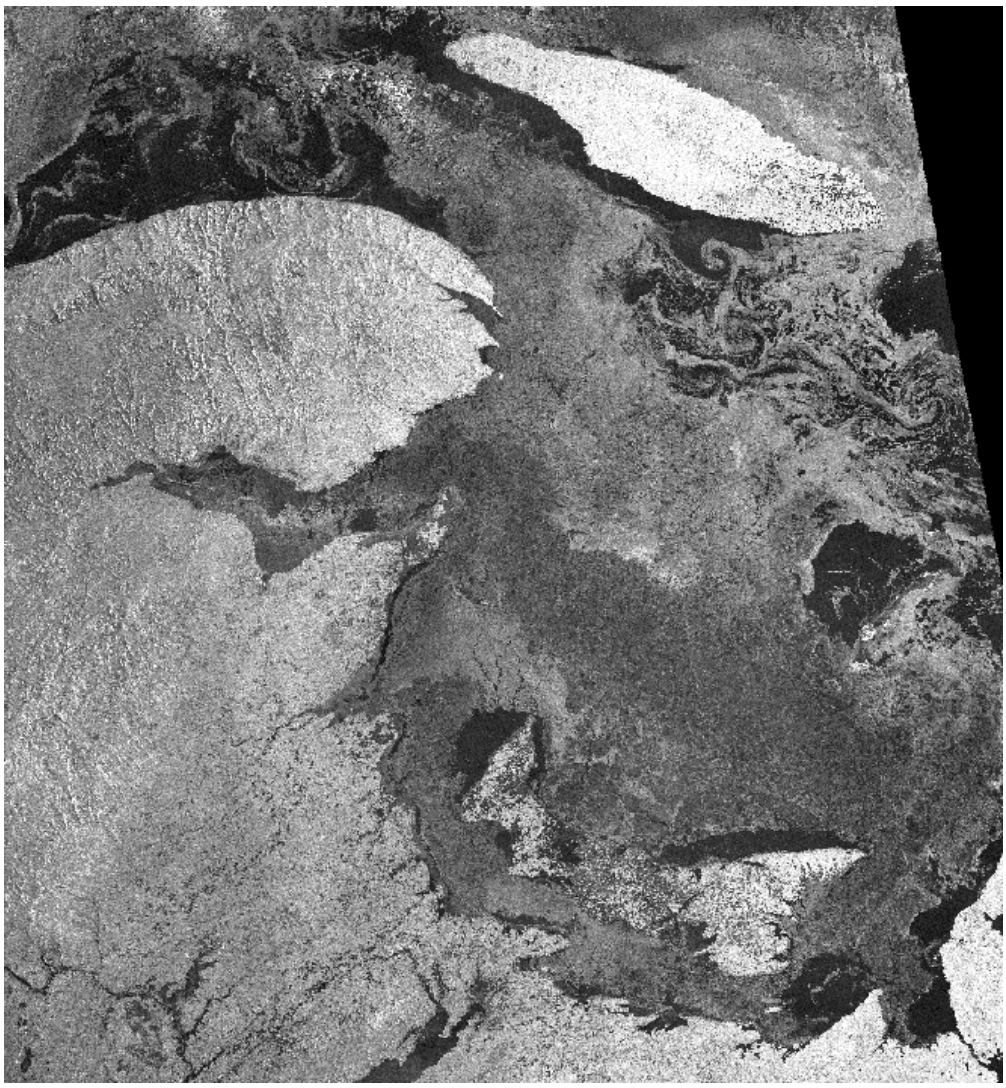


Figure 1.1: A SAR image of Gulf of St. Lawrence captured by RADARSAT ScanSAR C-band mode on Feb 20, 1998, with resolution of $100m$ and size of 6417×5958.

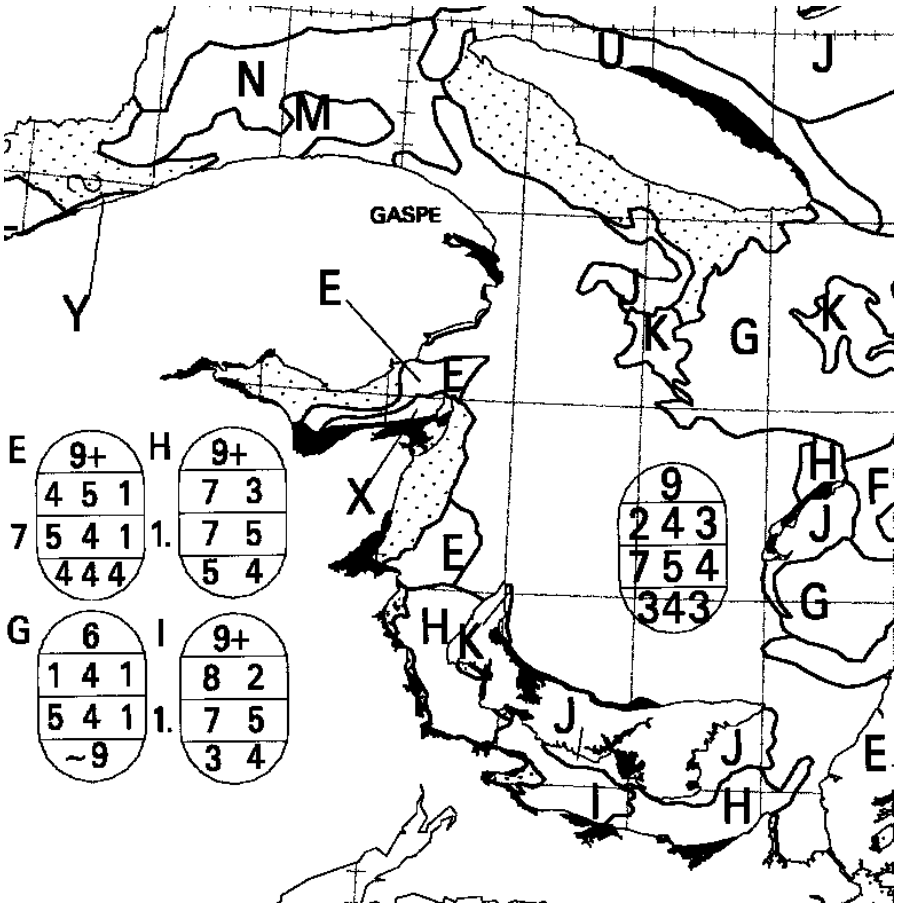


Figure 1.2: Ice chart corresponding to the SAR image in Figure 1.1. Egg codes are defined as per World Meteorological Organization standards [111].

geneous ice and identifies the corresponding ice type for each region. The process should be robust, efficient and as automated as possible. Designing such a system is the ultimate goal of this research. However, before defining the scope and focus of this thesis, it is necessary to look a little deeper into the application and associated problems.

1.2 Implementational Issues

Challenges

A fundamental difficulty with computer-based SAR sea ice image interpretation is the large variation of the backscatter caused by numerous environmental factors and imaging parameters. The same ice types can appear quite different in SAR images of different locations or seasons. Moreover, even within a single image, ice classes can exhibit a varying signature as a function of the incidence angle. Challenges also include the ambiguities of the SAR sea ice appearance among different types, especially in the transition areas. Another situation that disrupts a successful interpretation of SAR sea ice imagery is the speckle noise, which causes significant confusions in extracting the real tone (intensity) and texture features of SAR sea ice. More detailed descriptions about the influence of the environmental factor, imaging parameters, and speckle noise are presented in chapter 2.

Features

In a typical image interpretation system, a feature extraction step is performed first which selects and measures properties that best distinguish the class types. Features should be insensitive to rotation, luminance variation, and any other image specific

factors. As already mentioned, SAR sea ice backscatter has high variation. This causes the extracted low level (pixel-based) features such as tone and texture to be also highly variable. Some more stable features are those high level (region-based) ones such as shape. In the manual production of daily ice charts, human operators identify ice types not only by tone and texture but also by other clues such as floe shape and existence of fractures. It is therefore necessary to incorporate in the interpretation system both low level features and high level knowledge, which include tone, texture, shape and existence of fractures.

Supervised vs. Unsupervised

Based on the extracted features, the interpretation of SAR sea ice can be performed by either classifying each pixel into an ice type directly using knowledge of class distributions, or by grouping the pixels first according to some similarity or dissimilarity measure and then classifying the obtained segments into ice types. In the former, supervised training is required for obtaining the a priori knowledge of the class distributions of low level features, while in the latter pixels are grouped not by pre-trained class distribution but based on current image characteristics in an unsupervised manner. Considering the high variations of tone and texture of the SAR sea ice imagery, a comprehensive training of all possible variability is not feasible, and thus the latter one is more appealing for SAR sea ice due to its better adaptivity.

Segmentation and Classification

The interpretation process is thus split into two tasks: first segmenting the image into disjoint regions, and then classifying the segmented regions to the ice types.

Although the first task, the low level segmentation, is unsupervised, the second task of high level classification has to be trained based on domain knowledge. Models relatively insensitive to the variations of SAR backscatters (intensity and texture) are needed for robust descriptions of the ice types, and this insensitivity is achieved by utilizing high level knowledge which is relatively stable compared to low level features.

In the above, the segmentation and classification are performed separately, with the classification a postprocessing of the segmentation. A substantial deficiency of such a simple unidirectional link between the two processes is the fact that segmented regions may not match the real objects well enough for an accurate subsequent classification. In fact, segmentation is generally not a stand-alone problem but ill-posed if not associated with any implicit or explicit high level classification. Therefore, the classification needs to be able to guide the segmentation, and hence a bidirectional relationship between the two is desired.

For a better understanding of the necessity of such a bidirectional relationship for SAR sea ice interpretation, consider a region with multiple ice floes. The round shape of well defined floes is helpful in identifying some ice types. However, correct shape can only be computed if individual floes are not joined together in the segmentation result. Without knowing (or having an estimate of) what ice type the segments are and whether they are ice floes, it is impossible for a segmentation algorithm to determine if the two segments should be preserved apart or not.

In the literature, the term "interpretation" often refers only to the region-based high level classification, and another term "understanding" represents the overall process. To avoid confusion, both "interpretation" and "understanding" have the same meaning in this thesis. An image interpretation system, or equivalently an image understanding system (IUS), is a system that produces segmented regions

from the image and derives reasonable descriptions on the regions as well.

1.3 Thesis Objectives

None of the existing SAR sea ice interpretation systems has been operationally successful, due to the extremely complex nature of the SAR sea ice interpretation problem. It is thus more practical to start from a easier point, making use of some ancillary information [1] in a semi-automated manner rather than designing a fully automated system all at once. In particular, this thesis makes use of the egg code, which provides for the corresponding region the information on the number of classes for the segmentation purpose and on the candidate ice types for the classification purpose. The incorporation of egg code information not only reduces the solution space, but may also help in finding the correct solution by exploring the correlations between neighboring egg code regions.

Therefore, the goal of this thesis is to design a SAR sea ice imagery understanding system which, with the input of egg code regions, produces a pixel level ice identification for each input egg code region. The approach is not fully automated because it utilizes a limited amount of human provided information. However, incorporating such information is beneficial to the reliability of the system. Such a semi-automated approach is an important and practical step towards a fully automated SAR sea ice interpretation system that is operationally useful.

The thesis contributes in three aspects for SAR sea ice interpretation: an improved unsupervised low level segmentation, a practical high level classification,

[1] Even for human operators in CIS, a single SAR image might not contain enough information for correct ice identification, and additional information such as past ice charts and AVHRR data is required.

and a joint approach that combines the segmentation and classification in a bidirectional and interactive manner.

1.4 Thesis Outline

The organization of the thesis is as follows. Chapter 2 gives an introduction to the application environment, including topics on the SAR imaging system as well as the backscattering characteristics of sea ice. Chapter 3 gives a review of existing image understanding research with emphasis on those related to remote sensing and SAR sea ice field. Chapter 4 proposes a novel image segmentation method efficient for SAR sea ice imagery based on region growing and Markov random field theory. Chapter 5 deals with the high level classification, and combines the low level segmentation with the high level classification to form an overall SAR sea ice image understanding system. Chapter 6 is a summary of the thesis and discusses the future work.

Chapter 2

Application Environment

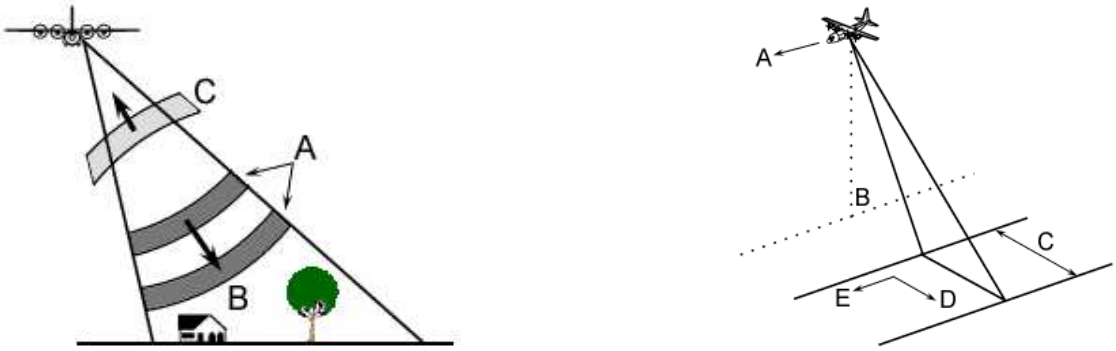
Modern remote sensing techniques make it possible to acquire information about the earth surface in real-time without actually being in contact with it. SAR as a remote sensing technology is especially useful for identifying sea ice types. The active microwave sensors of SAR have the ability to observe through clouds and obtain measurements at night, and the microwave band of SAR can be selected to be sensitive to target surface roughness, moisture content and electrical properties of sea ice. SAR images also have high resolution relative to other non-optical remote sensing technologies. The resolutions could be up to $3m$ for aircraft mounted SAR and $10m$ for space craft mounted SAR.

In this chapter, a brief introduction of the fundamental theory of SAR system is presented in section 2.1, which draws heavily on [38]. Then section 2.2 deals with the SAR signatures of sea ice. Dependence upon the imaging parameters and environmental variables are covered. The last section is a summary.

2.1 Introduction to SAR

2.1.1 Radar Basics

A SAR is a radar system that transmits a sequence of electromagnetic pulses sideways to the ground and records the returned waveforms. Figure 2.1(a) shows the basic mechanism of a side-looking SAR. The transmitter generates pulses (A) at regular intervals. The pulses are focused by the antenna into a beam (B), which illuminates the ground surface obliquely. A portion of the energy (C) is then backscattered and received by the antenna. By measuring the time delay, the distance to a target and hence its location can be determined. With the sensor platform moving forward a 2-dimensional image can be obtained by signal processing of the pulse returns over time and the flight path.



(a) Basics of SAR Imaging (b)Imaging Geometry

Figure 2.1: SAR Basics (from [38])

2.1.2 Viewing Geometry and Spatial Resolution

The imaging geometry of a SAR system is shown in Figure 2.1(b) in which (A), (B), (C), (D), (E) correspond to the **flight direction**, **nadir**, **swath**, **range**, and

azimuth respectively. The range resolution is dependent on the length of the radar pulse, while the azimuth resolution depends on the width of the illumination and the speed of the flight. Detail explanations of the resolution dependency can be found in [38].

In theory, fine range resolution could be achieved by using a very short pulse length. However, this is usually not practical since the peak power may exceed the limitation of the transmitter in providing a good signal-to-noise ratio (SNR). A technique known as pulse compression is therefore used to solve the problem. Wider pulses are transmitted instead of impulse-like waveforms, and the received signal is processed so as to compress the energy into a much narrower pulse [27].

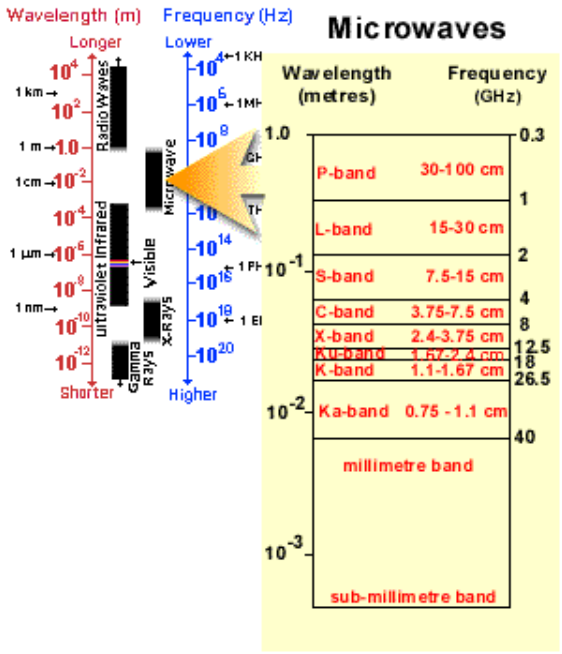


Figure 2.2: Microwave Spectrum (from [38])

Determined by the width of the illumination, azimuth resolution is actually influenced by radar beam angle and the slant range distance. The radar beam angle

is approximately $2\arctan(\lambda/2D)$, where λ is the wavelength and D is the antenna length. This implies that the radar would need an impossibly long antenna to achieve the same level of resolution as that of an optical system, since microwaves have much longer wavelength as shown in Figure 2.2. To overcome this antenna size limitation, a SAR can synthesize the effect of a very long antenna by using the Doppler effect via the motion of the platform (i.e. aircraft or satellite) relative to the ground. For further details, the reader is referred to [27]. Today the resolution provided by RADARSAT (a Canadian spaceborn SAR system) is in the range of 10m-100m, which is sufficient for the study of the structures of sea ice. The CIS primarily uses imagery captured under the ScanSAR model. This $50m$ resolution data is 2×2 block averaged to $100m$ resolution data for operational purpose.

2.2 SAR Sea Ice Imagery

Sea ice is generally a mixture of freshwater ice, brine and air. Based primarily on the duration and thickness of the ice, sea ice categories are defined by WMO [111] and are summarized in Table 2.1. The appearance of those sea ice types in SAR images is determined by the physical properties of the sea ice and the SAR system imaging parameters. The SAR system parameters include wavelength, incident angle, polarization, and etc. The physical property of ice is influenced by numerous environmental variables such as temperature, salinity, wind condition, ocean currents, rate of freezing, history of melting, existence of snow cover and intrusion of sea water.

New Ice:	A general term for recently formed ice. These types of ice are composed of ice crystals which are only weakly frozen together (if at all) and have a definite form only while they are afloat.
Grey:	Young ice 10-15 cm thick.
Grey-white:	Young ice 15-30 cm thick.
Thin first-year:	First-year ice of not more than one winter's growth, 30-70 cm thick.
Medium first-year:	First-year, ice 70-120 cm thick.
Thick first-year:	First-year ice over 120 cm thick.
Old Ice:	Ice which has survived at least one summer melt. Often sub-divided into either second year ice or multi-year ice.

Table 2.1: Ice categories defined by WMO [111]

2.2.1 SAR System Parameters

Whether a surface will appear smooth or rough under the illumination can be indicated by the Rayleigh criterion: $h < \lambda/(8\cos(\theta))$, where h is the root mean square height of the surface, λ is the signal wavelength and θ is the angle of incidence with the surface [109]. If the Rayleigh criterion holds true, the surface will be smooth and reflect away most of the signal like a mirror, resulting in a low returned energy. Similarly, a surface will be considered rough if the Rayleigh criterion is false. Since the smoothness is influenced by the incidence angle, the tone values for areas of the same ice type will decrease as the incidence angle increases across the image.

A suitable microwave band needs to be selected to capture the small scale surface roughness while being able to penetrate well through the atmosphere. The microwave region ranges from $1\mu m$ to $1m$, from the shorter wavelength near the thermal infrared region to the longer wavelengths approaching those used for ra-

dio broadcasts, as shown in Figure 2.2. For monitoring sea ice, the C-band of RADARSAT is suitable. Shorter wavelengths are more attenuated by the atmosphere and snow cover, while longer wavelengths are less suited to discriminating the small scale roughness of new or young ice.

Polarization, another SAR system parameter, also greatly influences the appearance of objects in SAR images. In most radar systems, microwave is transmitted in either horizontal (H) or vertical (V) planes. Similarly the antenna receives either horizontal or vertically polarized backscatters. Thus, there are four combinations (HH, VV, HV and VH) of the mode of transmitting and receiving signals, among which HV and VH are equivalent with respect to the obtained measurement. Depending on the mode, the signals will interact with the surface and be backscattered differently, thus greatly affecting the appearance of the surface.

One important phenomenon of SAR imaging system is that speckle noise is produced. Speckle is caused by the coherent nature of the radar imaging system. When a medium containing many elementary sub-resolution scatterers is illuminated by a radar, the return waves will add either constructively or destructively, forming a random, interference pattern. Such a interference pattern is deemed noise and has been experimentally shown to be predominantly multiplicative [66]. The existence of speckle noise adds considerable difficulty in extracting the real tone and texture features of SAR sea ice. Although there exists many speckle noise filters[66][39][58][88] for suppressing the noise before doing any further processing, such filtering would break down the texture details and blur the boundary of segments. Therefore, no speckle filter will be used in this research.

2.2.2 Physical Properties and SAR Signature of Sea Ice

Sea ice is generally a complex, dynamic and anisotropic mixture material, the constituents of which are freshwater ice crystals, brine and air bubbles. The physical properties of sea ice that influence the corresponding SAR signature include the constituents' dielectric properties, the volume fraction of each constituent, the geometry (shape, size and orientation) of brine pockets in the ice, the surface conditions as well as the property of snow cover. Since these factors are determined by many natural environmental variables including temperature, salinity, wind condition, ocean currents and rate of freezing, studying and utilizing the SAR signature of sea ice is an extremely complex task [36].

The returned signals of sea ice can consist of contributions from both surface scattering and volume scattering (scattering by lower layers). The surface scattering is mainly influenced by the roughness of the surface, and the volume scattering by the complex permittivity (an electrical property defined for high frequencies) as it determines how far the microwaves will penetrate into the ice. Because salt increases the dielectric loss by adding free charge carriers, this complex permittivity of sea ice is largely influenced by the brine volume. For detailed study on complex permittivity of sea ice, the reader is referred to [36].

Since first year ice has a high brine volume and hence is not easily penetrable by microwaves, surface scattering dominates. For first year smooth ice, a large part of the signal energy is reflected away and only a small part is returned in the direction of the radar. As a result the tone values of such an area are quite low, making it appear dark. For first year rough ice, more returns can be obtained and the corresponding area appears brighter.

Multi-year ice has a much lower salinity. Besides limited surface scattering, the

backscatters are mostly contributed by volume scattering, which is influenced by particle (air bubble and ice crystal) size and density. Although multi-year ice often looks apparently brighter than first year ice, diurnal and seasonal variations can sometimes mask backscatter signatures of multi-year ice versus younger ice types. These variations are caused by changes of water content and brine concentrations due to the melting of surface snow cover or occasionally intrusion of sea water.

It is uncertain whether or not young ice can be distinguished from first year smooth ice [19], as they can have a similar range and structure of backscatters. However, young ice is often characterized by leads, which are narrow irregular navigatable cracks consisting of thinner ice and sometimes open water. Identification of young ice can hence be achieved if such well-oriented feature is present. Difference between the backscatter signatures of the two types of young ice, grey and grey-white, can be very subtle. The tone of the two ice types depends largely upon the roughness of the ice surface. Grey ice often has a light grey tone, tear structures, light grey narrow bands indicating rafting and areas of brighter point returns. Grey-white ice often has relatively dark tone, angular/rounded floe shape, light grey ridge lines, and long angular fracture. Ice analysts often use the relative appearances of the lead shapes to distinguish grey and grey-white ice, as opposed to tone and texture.

New ice is generally darker than any older ice, and sometimes are also characterized by rafting as in grey ice.

2.3 Chapter Summary

In this chapter, topics concerning SAR imaging system and SAR sea ice backscattering has been reviewed. Understanding these is important for designing properly

the SAR sea ice imagery interpretation system, which is covered in the following chapters.

Chapter 3

Background

An image understanding system (IUS) aims at constructing a more meaningful representation of the image than its original pixel intensities. Typically, such a "meaningful representation" corresponds to a symbolic description of the classes or objects in the image, and an IUS is defined to be a system that transforms the raw image data into such a symbolic description. Suppose the image space is denoted by S, the image is denoted by $\mathbf{I} = \{I_s | s \in S\}$ where I_s is the image pixel value at site $s \in S$, and the symbolic representation is $\mathbf{L} = \{l_s | s \in S\}$ where l_s is the class (object) label at site $s \in S$. An IUS is a transformation

$$\mathcal{F} : \{I_s | s \in S\} \longrightarrow \{l_s | s \in S\} \tag{3.1}$$

In particular, the IUS that this thesis deals with tries to transform a SAR sea ice image into ice type symbols, with the input of egg code representation of the image which is at a much coarser level.

Direct transformation of the image to the ice type symbolic representation is

pixel-based and is typically a supervised process. In an IUS of such kind, supervised training is performed to obtain the a priori knowledge of the distribution of the computed features for each pixel of the given class or object. Such IUSs have little capability of incorporating region-based features and are often vulnerable to variations of the images. Therefore, they are not considered in this thesis.

More commonly, the image is transformed into an intermediate symbolic representation $\mathbf{x} = \{x_s | s \in S\}$, in which each unique value of x_s does not provide the description of a class or an object, but simply represents a group of pixels that are similar in the computed features. Such a group corresponds to a region if any two pixels of the group can be connected by a path consisting of pixels in this group only, or can be a set of regions (referred as a region cluster). The transformation from image to regions or region clusters is categorized as low level segmentation, while that from regions or region clusters to the final object or cluster labels is the high level classification. Figure 3.1 shows some typical paths that an IUS may follow. Except the red path, all the other three use either regions or region clusters as the intermediate representation, with the overall image understanding task decomposed into the two processes of low level segmentation and high level classification.

The IUS in Figure 3.1 is said to be of bottom-up (data-driven) fashion [73]. In a bottom-up IUS, the relationship between the segmentation and the classification is unidirectional, from the low level to the high level. A substantial deficiency of such a simple unidirectional link between the two processes is the fact that the segmented regions may not match the real objects well enough for an accurate subsequent classification. On the other hand, an IUS can also be top-down (knowledge-driven), which use a priori knowledge of the expected objects to guide the segmentation. Well known examples include Hough transform [65] and deformable template matching [26][50]. A top-down IUS is often computa-

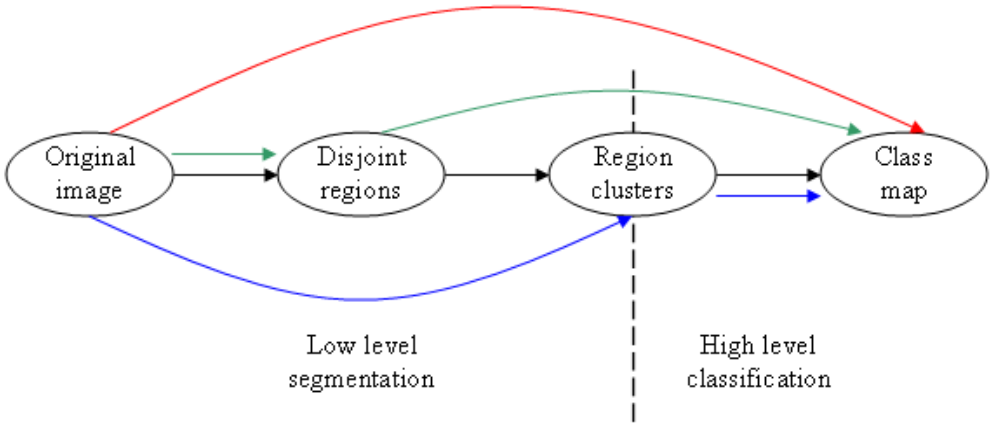


Figure 3.1: Flows of bottom-up interpretation systems.

tionally prohibitive or requires careful initializations. The attractiveness of combining both bottom-up and top-down control strategies have been pointed out by many researchers, and a wide variety of approaches can be found in the literature [48][54][59][102][104][105][106][121].

For a given application, whether the corresponding IUS should choose a bottom-up, top-down, or combined control strategy for the processes depends on the complexity of the image data and the involved knowledge. If the image data is simple, i.e. neither noisy nor ambiguous, image structures (or segmented regions) corresponding to meaningful objects can be reliably obtained, and hence bottom-up approaches are efficient. When the image data is complex but the expected objects can be described by simple models, top-down approaches are more suitable. When both the data and knowledge are complex, the combined bottom-up and top-down approach is preferred for exploring the advantageous of both schemes [103]. For SAR sea ice, both the imagery and the knowledge of sea ice are complex, and a method belonging to this combined control category is expected to be more effective.

This chapter presents a review on image understanding systems. There have been enormous image understanding systems in various application areas such as medical imaging, traffic surveillance, outdoor scene understanding, and remote sensing. Those systems are diverse in control strategy, modeling, knowledge representation and inference. The thesis does not intend to present a comprehensive review of those systems, which is neither necessary nor practical. In particular, those pure top-down approaches are excluded as they are totally different from the work of this thesis. Of interest here are remote sensing systems and emphasis will be put on those systems. Although the low level segmentation is a part of the image understanding process, it is often deemed as a rather distinct field. Thus, a review on image segmentation is given separately in section 3.1, while the overall image understanding systems are reviewed in section 3.2. Section 3.3 provides a summary.

3.1 Low Level Segmentation

An image segmentation is a transformation of the original image into a symbolic representation,

$$\mathcal{G} : \{I_s | s \in S\} \longrightarrow \{x_s | s \in S\} \tag{3.2}$$

by which the image space S is partitioned to n segments $\Omega_1, \ldots, \Omega_n$, such that [80]

$$\begin{aligned} &a) \Omega_i = \{s | x_s = i, s \in S\}, \\ &b) \bigcup_{i=1}^{n} \Omega_i = S, \\ &c) \forall i \neq j : \Omega_i \cap \Omega_j = \emptyset, \end{aligned} \tag{3.3}$$

In the above definition, any Ω_i can be either a single region or a set of disjoint regions (a region cluster). In remote sensing applications, the latter one is most

common. For example, multiple regions that are similar (suggesting of the same ice type) can occur at different locations in a SAR sea ice image, and it is often convenient during the image segmentation to group them together for future ice type identification. Hence, to make different notations for disjoint regions and region clusters is necessary. For the rest of the thesis, a region i is represented by Ω_i and a region cluster i is represented by C_i.

The image segmentation process should produce non-overlapping regions or region clusters meaningful to the subsequent high level classification (ice identification in this thesis). In general, segmentation procedures aim to produce regions that are uniform and homogenous with respect to the extracted features such as tone or texture, and have significant different feature values across region boundaries. Thus, the definition of segmentation in Equation (3.2) can also be represented as

$$\mathcal{G} : \{\mathbf{y}_s | s \in S\} \longrightarrow \{x_s | s \in S\} \tag{3.4}$$

where $\mathbf{y}_s$ is the computed features at site s, and can (but not necessarily) be the original pixel value I_s.

This section presents a review of unsupervised low level segmentation. Unsupervised methods can adapt to the varying tone and texture features of SAR sea ice imagery more effectively, and hence are more suitable compared to supervised approaches which require training. However, considerable research in this field is supervised [5][6][97][99], probably due to the following reasons:

- Successful unsupervised segmentation is not possible without an understanding of the ability of the features. Since unsupervised segmentation is not suitable for the evaluation of features as it is more complex and often uses additional information such as spatial context, supervised classification stud-

ies must come first.

- Supervised classifications provide truth data, making objective performance evaluation possible.

- The success of unsupervised segmentation is often determined by parameters, for example cluster number, initial seeds, edge threshold, etc. Finding appropriate parameters is extremely difficult for the complex SAR sea ice imagery.

Recent years have shown an increased interest in applying unsupervised segmentation methods to SAR sea ice imagery [19][23][30][100][101]. The proposed low level segmentation approach by this thesis is also unsupervised.

In the next, a review is given on related low level feature extraction and segmentation efforts. Although the main interest is in unsupervised approaches, supervised efforts primarily dealing with feature extraction are also included.

3.1.1 Feature Extraction

The feature extraction is a very important stage for both supervised and unsupervised segmentation. The goal of the feature extraction is to generate for each image site a set of values, in the form of feature vector, representative of the behaviors of the image at that site. Efficient features should have significantly different values on different classes (objects), but similar values on the same class (object). Moreover, the various features should be uncorrelated with each other to ensure that the feature vector has been reduced to the fewest possible dimensions.

For the SAR sea ice, tone and texture are the two most commonly used features. Here the word "tone" represents the intensity and the "texture" corresponds to the

spatial variation pattern of the intensities with some defined local region. There are many early examples in the literature that investigate and compare the usefulness of tone and texture information for discrimination of SAR sea ice classes. Shuchman et al. [98] gave an analysis of tone distribution and compare with the texture feature measured by gray level co-occurrence matrix (GLCM). They found that combining tone with variance can perform a satisfactory discrimination among first-year ice, open water and multi-year ice, although nilas and pancake can not be separated from multi-year ice. They also concluded that there is no improvement using texture information (measured by GLCM). Shokr [97] suggested that combining gray tone with one GLCM measure is a suitable choice. This conclusion does not disagree with Shuchman's since the variance measure used in Shuchman's work already contains some texture information. An interesting observation by Shokr is that textures are not capable of separating two ice types if they have a significant overlap in gray tone. Barber and LeDrew [6] found that tonal classifiers are as accurate as textural classifiers but not as robust to range dependent tone variations.

From these efforts, texture features seem to have the capability for discriminating between some sea ice types. But the extent of this capability contributed from the speckle differences caused by tone and the extent from real texture is unclear. Nevertheless, without a doubt, macro textures corresponding to ice structures such as ridges and rafts provide unique information for ice identification. Moreover, textures have the advantage of being more stable to variations caused by incidence angle, as observed by various authors [6][75].

Approaches to texture feature extraction span a wide range of methods. They can be categorized as statistical, structural, signal processing and model-based [108]. Statistical approaches attempt to characterize textures in a probabilistic sense based on definitions such as smooth, coarse, grainy, regular, directional, etc

[19]. Statistical measures include variance, skewness, kurtosis [16], GLCM [44], etc. The structural approaches assume that the texture is a spatial arrangement of basic primitives. The texture features are hence some measurements of the primitives and their spatial arrangements. Since no consistently repeating patterns exist in SAR sea ice textures, structural approaches seem not suitable for SAR sea ice discrimination. Signal processing based approaches use signal processing techniques such as filtering and Fourier transform to obtain texture features. Typical methods include Gabor filtering [16][19] and discrete wavelet transform (DWT) [72][118]. Model-based approaches fit specific models to the textures. The obtained model parameters are then used for texture description. A commonly adopted model is the Markov random field (MRF) [17][119].

In this section, brief reviews are presented on the GLCM texture features, signal processing based texture features, and MRF based texture features.

GLCM Texture Features

A GLCM [44] is a matrix which provides the conditional joint probability of all pairwise combinations of pixels within a computation window (W). These probabilities are conditioned on the sampling distance (δ) and orientation (θ). The (i, j) entry of a GLCM is the number of times the gray level pair i and j occur together with the chosen displacement and direction, normalized by the total number of pixel pairs. Since the computation burden of GLCM is heavy if the full dynamic range is used, quantization of the original image to fewer levels (G) is needed. From the matrix, statistical information can then be generated as features (MAX, UNI, ENT, DIS, CON, INV, IDM, COR), as summarized in Table 3.1.

The GLCM is a very popular texture feature extraction method in the remote

Maximum Probability (MAX)	$max\{P_{ij}\}\forall(i,j)$
Uniformity (UNI)	$\sum_{i=1}^{G}\sum_{j=1}^{G}P_{ij}^2$
Entropy (ENT)	$-\sum_{i=1}^{G}\sum_{j=1}^{G}P_{ij}\log P_{ij}$
Dissimilarity (DIS)	$\sum_{i=1}^{G}\sum_{j=1}^{G}P_{ij}\lvert i-j\rvert$
Contrast (CON)	$\sum_{i=1}^{G}\sum_{j=1}^{G}P_{ij}(i-j)^2$
Inverse Difference (INV)	$\sum_{i=1}^{G}\sum_{j=1}^{G}\frac{1}{1+\lvert i-j\rvert/G}P_{ij}$
Inverse Difference Moment (IDM)	$\sum_{i=1}^{G}\sum_{j=1}^{G}\frac{1}{1+(i-j)^2/G^2}P_{ij}$
Correlation (COR)	$\sum_{i=1}^{G}\sum_{j=1}^{G}\frac{(i-\mu_x)(j-\mu_y)}{\sigma_x\sigma_y}P_{ij}$

where P_{ij} is the element of the matrix in ith row and jth column, and in all cases (μ_x, μ_y) and (σ_x, σ_y) are means and standard deviations of (row i, column j) respectively

Table 3.1: GLCM texture statistics (from [21])

sensing community. Considerable research has been conducted using GLCM to discriminate SAR sea ice texture [5][6][21][97][99]. Most of those investigate distinct ice categories such as first-year ice and multi-year ice, and the results are satisfactory. Studies were also conducted on the selection of GLCM parameters (δ, θ, G, W) and the performance of each statistic and the combinations of them, as summarized in Table 3.2.

Shokr [97] found that the GLCM statistics are highly correlated and argued that combining gray tone with only one GLCM texture statistic is a suitable choice for SAR sea ice texture classification. Barber and LeDrew [6] suggested, however, that using multiple statistics is more advantageous. Clausi's [21] conclusion agrees with Barber and LeDrew's. He found that the GLCM statistics can be grouped into three categories. Statistics belonging to the same category are highly correlated while those belonging to different categories are relatively independent. Thus, he argued that a combination of three independent statistics (CON, ENT and COR)

Authors	δ	θ	G
Shokr [97]	1*,2,3	Average	16,32
Barber and LeDrew [6]	1*,5,9	0*,45,90	16
Soh and Tsatsoulis [99]	1,2,...,32	Average	8,16,32,64,128,256
Clausi [21]	1	0,45,90,135	8 to 256 (steps of 4)

Statistics	W	data resolution(m)
CON,ENT,IDM,UNI,MAX	5*,7,9	Variety
CON,COR,DIS,ENT,UNI	25	6×6
COR,ENT,IDM,UNI,CON,autocorrelation	64	N.P.
MAX,UNI,ENT*,DIS,CON*,INV,IDM,COR*	16	100×100

*Indicate a preferred choice; N.P. indicates that the information was not provided.

Table 3.2: Summary of selected papers using GLCM statistics as features for SAR sea ice classification (adapted from [21])

is the best choice. For the orientation parameter, Soh et al. [99] averaged the GLCMs of different orientations to make the extracted feature rotation invariant. Barber and LeDrew [6] investigated multiple orientations and conclude that the direction parallel to the look direction of the sensor is significantly better. Clausi [21] supported Barber and LeDrew's observation that independent orientations can produce significantly improved results. For the selection of sampling distance, Barber and LeDrew [6] showed that $\delta = 1$ is the best with the image ($6m$;7-look) used. However, Soh and Tsatsoulis [99] argued that a single displacement value for GLCMs to represent sea ice textural contexts is not advisable, since there is no hint on determining a priori such a value for all samples and range of displacement values is more representative for both local and global details. As to the quantization level parameter, Soh and Tsatsoulis [99] showed that quantization levels fewer than 8 should not be used, and Clausi [21] concluded that a quantization level greater than 24 and less than 64 is often an appropriate choice.

Signal Processing Based Texture Features

Most signal processing based feature extraction approaches assume that textures can be identified by their energy distribution in the spatial-frequency domain. The input image passes through a series of filters and the output responses usually correspond to frequency subband components. They are typically further processed nonlinearly to estimate the local energies. Feature vectors can then be computed with these energies (most commonly smoothed magnitude responses).

Two early filtering approaches are the Laws filter masks [64], and Ring and Wedge filters [25]. Laws suggested using 25 separable filters for the two dimension image, 5 in each dimension. Thus, the spatial frequency domain is almost evenly split into 25 subbands by these filters. Coggins and Jain [25] designed another bank of filters made up of seven dyadically spaced ring filters and four wedge-shaped orientation filters, which extract the frequency and orientation information separately.

Some later approaches, Gabor filtering [16] and the discrete wavelet transform (DWT) [72], have attracted more attention in recent years since they are in some sense similar with the human visual system (HVS). It has been found that the HVS decomposes images by several relative narrow-band filters tuned to different spatial frequency ranges and orientations [15]. Experiments have shown that the frequency bandwidth of each filter is about one octave [85] and the orientation range is not more than 30^o [87]. Both Gabor and wavelet filter banks can be designed to be tuned to octave frequency bands and different orientations. Although the Gabor function is not a wavelet in a strict sense, it can be implemented in a similar manner to that of wavelets.

Another important characteristic of Gabor and DWT is that the narrow banded

signals can be downsampled, providing a multi-resolution representation of the original image. Therefore, combining other techniques with this multi-resolution representation may be helpful in describing large scale features.

Literature in signal processing for SAR sea ice discrimination is quite limited. Clausi [19] gives a comparative study on Gabor filtering and the GLCM for the SAR sea ice texture feature extraction. His experiments showed that Gabor features may be more clusterable than the GLCM features [19, p.151], which makes it advantageous in unsupervised segmentation. Another of his conclusions is that the GLCM features can capture fine boundary details well with a small window size. Since smoothing is often required for the Gabor filtering output, the Gabor technique tends to blur texture boundaries belonging to small regions. Clausi also believed that Gabor filtering is more appropriate for structured textures whose energies are concentrated in low and mid spatial frequency part, while GLCM has stable response to textures with high spatial frequencies and is not so sensitive to noise [22]. Yu [118] demonstrated that the DWT can give a satisfactory classification accuracy, and using GLCM on the multi-resolution DWT representation does not show to be superior.

MRF Based Texture Features

The MRF assumes that any pixel is independent to others outside its defined neighborhood given the configuration of the neighborhood (Markovianity principle). The neighborhood, with appropriate size or order, makes computation of the dependence among sites manageable. By the Hammersley-Clifford theorem [8], the MRF allows the configuration of the sites be described in a probabilistic perspective using energy functions defined on neighborhood cliques. Details will be presented later in Chapter 4.

A commonly seen MRF model for texture feature extraction is the Gaussian Markov random field model (GMRF) [17]. The GMRF assumes that the joint probability of pixels is Gaussian distributed. It has been shown that GMRF pixels can be represented as a linear weighted combination of the neighborhood pixels and additive noise [17][57]. Those weights are computed from the texture image based on the minimum mean square error (MMSE) criteria and can act as texture features.

Yue [119] compared GMRF with GLCM for SAR sea ice segmentation and concluded that GMRF generally performs poorer than GLCM as the model may not fit SAR sea ice textures well. Clausi [20] found that although GMRF lagged Gabor and GLCM, GMRF features tend to be uncorrelated with them, and hence can be fused with either of them to provide supplemental information and improve the SAR sea ice texture features.

3.1.2 Unsupervised Segmentation

The unsupervised segmentation is a segmentation process that does not require the a priori knowledge of the class distributions of features, i.e. without the need of training. Typical unsupervised image segmentation methods include:

- *Histogram thresholding* [83] and *clustering* [35][49]: Histogram thresholding techniques distinguish pixels that have higher feature values from pixels that have lower feature values by thresholds determined from histogram statistics. Clustering methods usually measure the fitness of the extracted features of each pixel into the pre-defined model of each class. Both methods use information only in the feature space and ignore the spatial context of the image

(even though some features themselves can incorporate spatial information), leading to increased sensitivity to noise.

- *Region growing and merging* [1][16][107]: These approaches perform segmentation by merging smaller regions into large regions using statistical similarity tests. The initial regions can be small neighborhoods or even single pixels. Baraldi and Parmiggiani [4] applied region growing to solve the over-segmentation problem for SAR image. Selection of the initial seed regions and the homogeneity criteria are the two major problems of this kind of approach.

- *Edge detection*: These approaches segment by finding discontinuities of features in the image and connecting those edges to form closed boundaries. Various methods have been proposed for the purpose of generating the closed boundaries using edge information, such as directional searching [70], watershed [110][80], and snake contours [51]. The performance of this kind of approach depends largely on how well the edge is defined in the image, and hence it will often encounter difficulty if the image is noisy or has indistinct boundaries.

- *Model-based*: These segmentation approaches take into account the spatial context as well as the homogeneity and discontinuity information by modeling using defined energy functions. The segmentation is obtained as the result of energy minimization. The most popular model in image segmentation is the MRF [9][40][67][116]. Snake contours [51] may also be included in this category. Finding the minimum of energy can be very difficult and computationally complex since the solution space is usually non-convex. As a result, the initial configuration often influences the segmentation significantly, especially for boundary-based models.

The various segmentation approaches can also be categorized as region-based, edge-based and hybrid. Region-based methods explore the homogeneity of features within the region, edge-based methods deal with the dissimilarity across the boundary, while hybrid techniques combine the two. In the recent years, hybrid techniques have drawn more attention and the related research [13][107] span a wide range of class of methods.

Histogram Thresholding and Clustering

Histogram thresholding techniques [83] distinguish pixels that have higher feature values from pixels that have lower feature values by thresholds determined from histogram statistics. Usually, a threshold is selected as a valley between two peaks of the histogram. To avoid the influence from small spikes, a histogram can be smoothed first, or the peak and valley detection can be performed on a multi-scale basis [100]. Such methods will encounter problems when the image does not have distinct histogram modes separated nicely by valleys.

The clustering methods group the pixels based on their feature vectors into clusters. The underlying model of the probability density of the clusters is assumed to be known and the model parameters are estimated from the feature data of the pixels. A typical probability density model is the Gaussian function. Such a Gaussian mixture problem can be formulated by minimizing an objective function

$$\arg \min_{C_1,\ldots,C_n} \left\{ \sum_{i=1}^{n} \sum_{s \in C_i} \left\{ \frac{1}{2} \log(|\Sigma_i|) + \frac{1}{2} (\mathbf{y}_s - \mu_i)^T \Sigma_i^{-1} (\mathbf{y}_s - \mu_i) \right\} \right\}, \tag{3.5}$$

where n is the number of region clusters, μ_i is the mean feature vector of region cluster C_i, and Σ_i is the covariance matrix of region cluster C_i. The expectation-maximization (EM) technique [28] is typically used to estimate the model param-

eters and perform the clustering. The popular K-means clustering [35] is also a simplification of this Gaussian mixture method.

Clustering techniques are related to the thresholding technique with the cluster model parameters determining the corresponding thresholds. In clustering, the number of clusters has to be set a priori or estimated by, for example, the minimum description length [43] principle. Both clustering and histogram thresholding generate region clusters and belong to the category of methods that follow the blue path in Figure 3.1.

Soh and Tsatsoulis [100] detect the peaks of the gray level histogram on a multiscale basis, and thus determine the number of categories as well as the thresholds for segmentation. A cluster aggregation approach named Aggregated Population Equalization (APE) is then applied. Samadani [93] applied the finite Gamma mixture model based on the assumption that each ice class in SAR imagery follows a Gamma distribution. Clausi [19] proposed a clustering algorithm named K-means Iterative Fisher (KIF) and use a divisive binary hierarchical tree to determine the cluster number. It was applied to Gabor and GLCM features of SAR sea ice images with satisfactory accuracy. All the above approaches explore the spatial relationship of image pixels only to a very limited (if any) degree, leading to increased sensitivity to noisy measurements of features.

Spatial Context Model

Spatial context models such as those MRF-based have been popular in image segmentation as they provide an efficient regularization, which is theoretically necessary since image segmentation is generally an ill-posed inverse problem.

From Bayes' rule:

$$P(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{p(\mathbf{y})} \tag{3.6}$$

where $P(\mathbf{x}|\mathbf{y})$ is the posterior conditional probability of the label configuration given the features, $p(\mathbf{y}|\mathbf{x})$ is the distribution of the features conditioned on the label configuration, $P(\mathbf{x})$ is the prior probability of the label configuration, and $p(\mathbf{y})$ is the distribution of the feature vector. Maximizing the posterior $P(\mathbf{x}|\mathbf{y})$ gives a segmentation solution, and it is equivalent to maximizing $p(\mathbf{y}|\mathbf{x})P(\mathbf{x})$ by ignoring the $p(\mathbf{y})$ which is constant for all possible configurations of $\mathbf{x}$. It is thus possible to impose spatial context constraint by modelling $P(\mathbf{x})$ with MRF [30][31][67][116][119].

The multi-level logistic model (MLL) [31] has been a popular spatial context model, and will be described in detail in Chapter 4. Examples of MLL context model applied to SAR sea ice segmentation can be found in [23][30]. Clausi and Yue [23] use the MLL spatial context model together with the GMRF texture features, and the result is not superior compared with the clustering approach on GLCM features. Deng and Clausi [30] focus on weighting the context model adaptively and has achieved better results for SAR sea ice segmentation than the MLL model. Both methods have produced smooth segmentation results relatively insensitive to noise.

Watershed and Region Growing

The watershed algorithm for image segmentation [80][110] simulates the flooding patterns of terrains. Using the gradient magnitude as height, the surface is flooded from local minima of heights and the merging of the waters coming from different sources is prevented. The image is thus segmented into disjoint regions separated by *watershed lines* which represent the locations where water from different sources

meet. The first step in computing the watershed is identifying the minima. Automated approaches typically select them as the local minima from a smoothed gradient image, often leading to over-segmentation results for noisy images.

Region merging is typically used for solving this over-segmentation problem [16][45][82]. In region merging, similar neighboring regions are merged according to a certain criterion, which differentiate various existing merging approaches. Although an inverse operation of splitting inhomogeneous regions into smaller regions can also be integrated with the region merging such as in [107], it is usually computationally prohibitive.

An example that has applied watershed and region growing to SAR sea ice segmentation is ARKTOS [101]. In the ARKTOS system, an initial watershed segmentation is performed, followed by region merging based on criteria of boundary gradient, intensity difference and sizes of the regions. Attributes such as area, average intensity, shape and texture measures are then computed of each obtained region for a Dempster-Shafer rule-based ice classification system. ARKTOS is identified as a complete image understanding system that follows the green path in Figure 3.1.

3.2 High Level Classification and Image Understanding

The high level classification is defined as the labelling process that transforms the region or region cluster representation into the object or class labels:

$$\mathcal{L} : \{x_s | s \in S\} \longrightarrow \{l_s | s \in S\} \tag{3.7}$$

In a bottom-up or combined control IUS, the task of labelling the segmented regions is strongly application dependent and requires application domain knowledge to be known. If the domain knowledge is simple and can be represented by a number of values, feature vector classification can be applied. The feature vector classification has been well understood [35][41], and is the simplest type of high level classification process for an IUS. The features are computed from the segmented regions representing intensity, texture, shape, and other properties of the regions. Such approaches cannot handle cases that are associated with more complex knowledge (e.g. context information), and sophisticated methods are required for encoding and manipulating the knowledge. There are two categories of such methods: knowledge-based and probabilistic-based [33]. The knowledge-based approach overlaps significantly with the subfield of artificial intelligence (AI) called knowledge-based systems. It tries to make the knowledge of how to select, apply and combine various image processing algorithms explicit, and aims at building a system readily extendable, reusable and easy to understand for non-experts on the field of image analysis. The probabilistic-based approach uses formal probabilistic models such as Bayesian belief networks [84] or Markov random fields [78]. They are well grounded in probability theory and are expected to be more robust to minor changes of image characteristics [33].

3.2.1 Knowledge-based System

Defined in the AI field, a knowledge-based system (KBS) customarily refers to a computer based system in which a large part of knowledge is represented symbolically and separated from the inferencing mechanism. Two kinds of knowledge are differentiated: procedure knowledge and declarative knowledge. The procedure knowledge is implicitly incorporated in a model or algorithm, while the declarative

knowledge is represented explicitly and symbolically. The field of KBS mainly deals with the representation and manipulation of declarative knowledge, although procedure knowledge has been an integral part in many KBS for the construction of those relatively simple processing routines used in the KBS. The architecture of a KBS consists of a knowledge base that embodies knowledge, an inferencing engine manipulating the knowledge base, and a working memory where the initial data and intermediate results are stored [42].

The knowledge representation, which defines the data structures for storing knowledge and interpretive procedures for intelligent manipulation of knowledge, is crucial for a KBS since the system's ability to know, update, perceive and understand is restricted by the content and structure of its knowledge base [117]. Some popular knowledge representation techniques include production rule, semantic network, and frames [89]. In this section, a review will be given to some existing image understanding systems, organized by the knowledge representation and inference mechanism. It should be noted that the KBS may not be restricted to the highest level of region-based classification and scene understanding for an IUS. Systems based on image domain knowledge, which is types of and relations among image features as opposed to those application domain knowledge of objects, have been used on lower levels of image segmentation, enhancement and various other image processing tasks [79][73]. Such systems for low level processing are not of interest to this thesis.

Semantic Network

The semantic network is a graph in which nodes represent concepts or instances of objects in the problem domain and the links represent relations and associations between them [86]. Concepts describe the predefined generic model of the objects and

instances are realizations of concepts in the image. The links for the relationships between objects do not have constraints for the naming, but certain classes have developed as standards. For example, the *is-a* link is often used to describe the specialization of an object. Other links typically used are *instance-of*, *part-of*, *con-of* (transformation of an abstract symbol to its concrete realization), and etc. Figure 3.2 shows an example semantic network used in a remote sensing KBS named AIDA for representing landscape objects in aerial images [68]. Another important component is the computation functions that calculate the attribute ranges and values. For example in Figure 3.2, the MRF is a computation function for describing the textured 2D region. The computation functions incorporate procedural knowledge, and have to be selected and defined a priori by experts.

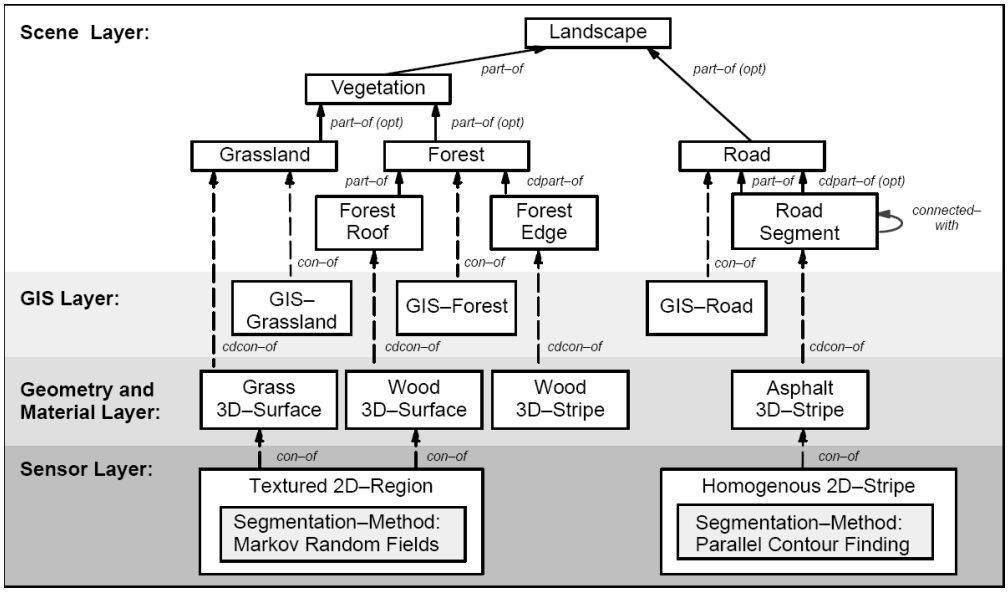


Figure 3.2: Simplified semantic network used in [68].

Different inference procedures can be used for the semantic network. However, the inheritance is always included and implemented by *is-a* and *instance-of* so as to provides a means of dealing with default reasoning. The control of the inference

determines how and in which order the inference is performed on the nodes and links. In the AIDA system, the control is proceeded by executing a sequence of predefined rules order by their associated priorities.

Examples of semantic network based IUS include ERNEST [81], AIDA [68], geoAIDA [7], a digital landscape model based system [56], and etc. Successes have been reported for their selected applications including interpretations of aerial images [7][68], scintigraphic images and MRIs [61][81].

Frames (or Schema)

In a KBS that is based on frame system theory [77], abstract objects in the problem domain are represented by "frames". Each individual frame corresponds to a node in a graph, and the relationships between the frames are the links in the graph. Compared to the nodes in the semantic network, the frames are more complex entities and object oriented. Attached to each frame are a number of "slots", and the slots are associated with attribute-value pairs with each forming a description of the object. A value can be a set or range of numbers, a pointer to another frame, or a function incorporating procedural knowledge.

Frames are arranged in a hierarchical manner via defined relations such as the *is-a* link. As in the semantic network, inheritance is an important part of the inference. Another inference mechanism involved is matching, which match a given frame to a concept frame. The matching process has three components: testing appropriateness, updating slot values, and transferring control to the next appropriate frame [89].

Frames (or Schema) have been used in many KBS image understanding systems such as SIGMA [48], VISION [34], and MESSIE [24]. For example in SIGMA, each

object represented by a frame has three main components: unary properties of the object, relationships with other objects, and associated rules to guide the interpretation process. A rule is composed of condition, hypothesis, and action. Beginning from some hypothesis obtained after an initial segmentation, bottom-up hypothesis generation of high level object instances is performed, which in turn results in the top-down hypothesis generation of missing image structures. The SIGMA system iterates such hypothesis generation processes together with evidence accumulation and hypothesis verification. Success has been demonstrated in understanding scenes such as houses and roads in aerial images [48].

Production Rules

The semantic network and frames have provided representation of objects. Knowledge other than the object models (e.g. how to seek and match) is typically represented by rules. In production rule representations, knowledge is in the form of "if...then..." structure. Such representations have several properties [92]:

- Incorporation of new knowledge can be easily achieved by adding rules to the knowledge base.
- Complex problems are solved by selecting rules and combining the result.
- The result can be explained by reversing the line of reasoning.

Rules are suitable for representing knowledge which occurs as a large number of discrete facts, and are also often combined with object models such as semantic networks and frame for controlling the inference, as in examples of AIDA and SIGMA.

Uncertainty and Dempster-Shafer Theory

Uncertainty arises because of the incompleteness and inaccuracy of the data, knowledge, and understanding. For example, in our application of SAR sea ice analysis, images are noisy, the egg codes are subjective and not necessarily 100% accurate, and transitions from one ice type to another is not discontinuous. A successful KBS should have some mechanism to deal with uncertainty.

There are several means of handling uncertainty for a KBS: certainty factor [46], fuzzy logic [92], Dempster-Shafer theory [96], and probabilistic models [84]. The two commonly used are Dempster-Shafer theory and probabilistic models. Probabilistic models are categorized separately from KBS approaches in section 3.2.2, and this section only reviews Dempster-Shafer based system for image understanding.

The Dempster-Shafer theory is designed to deal with the evidence supporting the proposition rather than computing the probability of a proposition. The following coin flipping example in the AI book by Russell and Norvig shows clear the basic idea of the Dempster-Shafer theory [92](p. 462). "Given that the coin may or may not be fair, what belief should you ascribe to the event of it coming up head? Dempster-Shafer theory says that because you have no evidence either way, you have to say that belief $Bel(Heads) = 0$, and also that $Bel(\neg Heads) = 0$. Now suppose that you have an expert at your disposal who testify with 90% certainty that the coin is fair (i.e. he is 90% sure that $P(Heads) = 0.5$). Then Dempster-Shafer theory gives $Bel(Heads) = 0.9 \times 0.5 = 0.45$ and likewise $Bel(\neg Heads) = 0.45$. There is still a 0.1 gap that is not accounted by the evidence. Dempster's rule shows how to combine evidence to give new values for Bel, and Shafer's work extends this into a complete computational model."

Wesley has applied the Dempster-Shafer theory to a KBS designed to interpret

color images of outdoor natural scenes [113]. Andress and Kak has developed a robot self-location system PSEIKI using the Dempster-Shafer reasoning to resolve conflicting hypothesis [2]. One relevant work to the SAR sea ice interpretation is the ARKTOS system [101]. In ARKTOS, a SAR sea ice image is first segmented into regions by watershed and region merging. Attributes are then computed from the segmented regions. Such attributes include area, average intensity, perimeter, average roughness, elongation, roundness, and etc. Facts are then extracted based on the computed attributes, and rule-based classification is performed using Dempster-Shafter belief theory. The system is reported to be fully automatic, near real-time, and has good performance in high-Arctic areas.

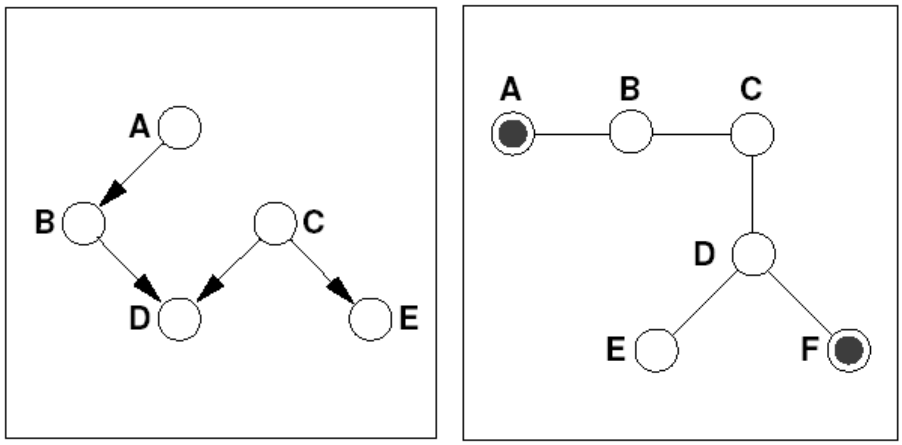
3.2.2 Probabilistic Models

Knowledge-based systems are usually ad hoc. Researchers formulate rules mostly based on intuition or through trial and error. Such a heuristic designing manner not only makes it difficult to compare different KBS, but also cause many systems to be brittle even for very limited variety of images. Probabilistic models such as Bayesian belief networks [84] and Markov random fields [67] have stronger theoretical foundations, and have recently gained more attention for constructing image understanding systems. In those IUSs, the probabilistic models have been used for modelling complex knowledge representations [11][91] or more commonly, applied to the region adjacency graph (RAG) constructed from the segmented regions of the images [53][60][78].

Bayesian Belief Network

A Bayesian belief network (BBN) is a directed acyclic graph (DAG) whose nodes represent propositions (or variables) and whose links represent causal influences [84]. In a Bayesian network, knowledge is represented by the network structure and the conditional probabilities associated with the links. Figure 3.3(a) shows the relational graph example of a BNN. The full probability distribution can be expressed as the product of conditional and prior probabilities. Thus for this example,

$$P(ABCDE) = P(A)P(B|A)P(C)P(D|B,C)P(E|C) \tag{3.8}$$



(a) (b)

Figure 3.3: Relational graphic models (adapted from [112]) (a) A Bayesian network. (b) A singly connected Markov network.

There exists for singly connected networks (i.e. there is only one path between any two given nodes) efficient belief propagation schemes, which obtains posterior probabilities of the unobserved variables given observed variables [84]. In such a belief propagation scheme, information is passed through the network in a distributed manner and, when equilibrium is reached, each proposition will be assigned a mea-

sure of belief consistent with the axiom of probability theory. It should be noted, however, that the region adjacency graph (RAG) typically used for representing objects (segmented regions) in interpretation have loops. The behavior of loopy BBN has been investigated by Weiss [112], and it has been shown that the belief propagation scheme for singly connected nets can still be applied to loopy BBN if it is being used for finding MAP solutions.

Markov Random Fields

Markov random field (MRF) can be applied not only to regular image lattice but also to more general network structures, as shown in Figure 3.3(b). The latter is also named *Markov network*, and is very similar to the BBN. Unlike the BBN which has directed links, the MRF has undirected links representing symmetrical probabilistic dependencies. Clique energies are used to encode the knowledge instead of the conditional probabilities in BBN. The interpretation is then formulated as finding the minimum of the overall energy that corresponds to the maximum of the probability of the interpretation given the domain knowledge and measurements on the image (or image structures obtained).

An early MRF based interpretation work is by Modestino and Zhang [78], in which the clique energy functions is selected using fuzzy set theory. Kim and Yang [53] estimate the clique function by training a back-propagation neural network. In the IUS proposed in this thesis, RAG based MRFs are used for both the region based clustering in Chapter 4 and the SAR sea ice interpretation in Chapter 5.

3.2.3 Others

Some other systems [71][105] are not as rigorously organized as the KBS or probabilistic models, and are designed more specifically for some applications. One system that needs to be mentioned is the map-guided MAGSIC [71] for SAR sea ice interpretation. The MAGSIC first performs the segmentation using the MRF-based approach proposed in [30], and then further labels the obtained region clusters based on the ice chart. Since the ice chart provides the information on how many and what type are the ice classes in a specified region, the labelling for single class regions is straightforward and the beliefs are propagated to other non-single class regions. The adopted propagation scheme in MAGSIC is rule based, implemented by comparing region clusters showing up simultaneously in neighboring egg code regions. Such a propagation mechanism is ad hoc, and allows further improvement using, for example, Bayesian belief network. The MRF based method adopted in this thesis also provides such a possibility.

3.3 Chapter Summary

In this chapter, a review of image understanding systems has been presented. Existing approaches of both low level segmentation and high level classification have been covered, with an emphasis on those related to SAR sea ice field. In the next two chapters, a novel unsupervised low level segmentation method is proposed and a practical high level classification system is designed for SAR sea ice imagery.

Chapter 4

Iterative Region Growing on Semantics

This chapter deals with the unsupervised low level segmentation for SAR sea ice imagery. Although there has been a variety of methods in literature applied to SAR sea ice imagery, those approaches have more or less deficiencies that make them not sufficient for operational use. For example, both the clustering and thresholding approaches [100][93] and MRF approaches [30][119] use stationary models (either feature or spatial context) with the model parameters estimated globally, and thus could result in false partitions for non-stationary images. Moreover, spatial interactions among pixels have not been efficiently utilized in existing approaches. Although the MRF context model adopted in some research [30][119] is able to describe long range (tens or hundreds of pixels) of behaviors by propagating local interactions, such propagations are extremely slow and are often limited within a local range. Another common drawback of previous research is that high level features, for example the shape of ice floes, have not generally played a role in the

segmentation process. Although ARKTOS [101] uses such high level knowledge for classifying the segmented regions into ice types, the process for obtaining the segments (i.e. segmentation) has not benefited from the incorporation of such high level features.

Among the different kinds of methods above, the MRF [30] is an efficient method as it accounts for the nature of SAR speckle noise in a statistically optimal way and at the same time provide an efficient spatial context model for regularization. Moreover, it is convenient for such global approaches to incorporate ancillary information, i.e. the number of classes provided by the egg code, which is beneficial for the reliability of the system. Therefore, this research follows the MRF-based approaches, using a formulation that is derived from the traditional MRF one for the image segmentation problem. Different from existing MRF based approaches, the proposed method has incorporated edge strength in a novel manner to improve on non-stationary situations, and uses a sequence of region merging operations to obtain the solution. Such an iterative region merging procedure also allows the possibility of integrating high level classification to guide the segmentation process, and thus the proposed method is named as *iterative region growing on semantics* (IRGS). Since this chapter only investigates the low level segmentation problem, the integration of the high level classification will not be described until chapter 5.

Section 4.1 briefly reviews the MRF theory and the traditional MRF based image segmentation method. Section 4.2 introduces the proposed method. Results and discussions are presented in section 4.3, while the final section (4.4) provides a summary.

4.1 Markov Random Field Based Segmentation

4.1.1 Markov Random Field

The MRF assumes that any pixel is independent to others outside its defined neighborhood given the configuration of the neighborhood (Markovianity principle). The neighborhood, with appropriate size or order, makes computation of the dependence among sites manageable. Figure 4.1(a) shows the neighborhood of a five-order model defined on a regular image grid. Another definition involved is clique, which is a set of mutually neighboring sites. Figure 4.1(b) gives an example of cliques of a second order model.

The Hammersley-Clifford theorem [8] shows that a random field X is an *MRF* if and only if X is a Gibbs distribution. That is, for any configuration $\hat{x}$ of the state space T of random field X, the joint probability is

$$P(X = \mathbf{x}) = \frac{1}{Z} \exp\{-E(\mathbf{x})\} = \frac{1}{Z} \exp\left\{-\sum_{c \in \mathcal{C}} V_c(\mathbf{x})\right\} \tag{4.1}$$

where $\mathcal{C}$ is the set of all cliques on the entire lattice, $V_c(\mathbf{x})$ is the energy of configuration $\mathbf{x}$ on clique c, $E(\mathbf{x})$ is the total energy of configuration $\mathbf{x}$, and the normalizing constant $Z = \sum_{\mathbf{x} \in T} \exp\{-E(\mathbf{x})\}$ is called the partition function.

The clique energy functions model the interactions among pixels in a neighborhood. Based on different energy functions, various MRFs have been proposed. Popular MRF models include the Gaussian model (GMRF) [17], which is usually applied to feature extraction, and multi-level logistic model (also known as Potts model) [31][116], which is used for modelling context content in segmentation.

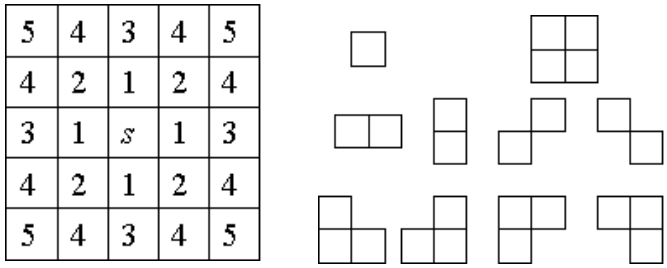


Figure 4.1: Neighborhoods and cliques (a) The neighborhood of a site s with labels of order 5. At order n, the neighborhood contains all the sites that have label less than or equal to n. (b) The 10 clique types of a second order model. Adapted from [67].

4.1.2 MRF Based Segmentation

As an inverse problem defined in (3.4), the image segmentation task can then be formulated as a MAP (maximum a posterior) problem, for which maximizing the a posterior $P(\mathbf{x}|\mathbf{y})$ gives a solution. As already described in Equation (3.6) in section 3.1.2, this is equivalent to maximizing $p(\mathbf{y}|\mathbf{x})P(\mathbf{x})$, and thus it is possible to impose spatial context constraint by modelling $P(\mathbf{x})$ with MRF [30][31][67][116][119]. The multi-level logistic model (MLL) [31] has been a popular spatial context model, whose clique energy is defined as:

$$V(x_s, x_t) = \begin{cases} \beta_0 & \text{if } x_s \neq x_t \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

where s and t are neighboring sites forming a pair-site clique, and β_0 is a positive number. Such a context model makes the prior $P(\mathbf{x})$ large if the local neighborhood region is dominated with one single class and small otherwise, and hence is effective in suppressing noisy configurations of class labels.

Using Equation (4.2) as the spatial context model and assuming the feature model is Gaussian, the objective energy function is thus defined in (4.3), the minimization of which gives a solution to the segmentation problem.

$$\arg \min_{C_1,\ldots,C_n} \sum_{i=1}^{n} \sum_{s \in C_i} \left\{ \frac{1}{2} \log(|\Sigma_i|) + \frac{1}{2} (\mathbf{y}_s - \mu_i)^T \Sigma_i^{-1} (\mathbf{y}_s - \mu_i) + \beta_0 U_i(s) \right\} \qquad (4.3)$$

where $U_i(s)$ is the number of neighboring sites of s that does not belong to region cluster C_i, and other definitions are the same as in (3.5).

An important issue with MRF-based approaches is how to compute the global minimum of the objective function. MRF-based objective functions such as that in (4.3) are highly non-convex [67], and the existence of local minima cause considerable difficulty in finding the global minimum in an intractable vast search space. In practice, a desired label configuration does not necessarily have to correspond to the global minimum of the energy function. However, it is usually the case that a local minimum with lower energy suggests a more accurate result. Methods easily trapped in local minima are therefore not appealing.

Various combinatorial optimization methods have been used, which include iterated conditional mode [9], simulated annealing [40], mean field theory [120], genetic algorithm [3], network inference [18] and graph theoretic techniques [13]. The iterated conditional mode (ICM) [9] is the simplest among all. Utilizing the local dependence among pixels, the ICM iteratively refines the label configurations based on the provisional estimate of those configurations, accepting changes that decrease the energy function the most. Its greedy nature makes it easily trapped in local minima and sensitive to the initial estimate. A better and perhaps the most widely used approach is simulated annealing (SA) [40][55]. Unlike the ICM, SA employs a random search scheme, for example, Metropolis-Hastings sampling [76] which

accepts or rejects changes not deterministically but based on a probability

$$P = \frac{1}{Z} e^{-\delta E/T} \tag{4.4}$$

where δE is the change of energy if the new configuration is accepted, Z is a normalizing factor, and T is a parameter named *temperature* irrespective of the energy function involved. The SA process simulates the physical process of annealing by slowly decreasing the temperature T to force the system into lower energy states. With a certain cooling schedule (annealing schedule), the SA can be guaranteed to find the global minimum [40]. However, the schedules leading to this guarantee are impossibly slow, and there is a tradeoff between the cooling speed and the quality of the solution.

In the SA process, the algorithm allocates an increasing amount of trials to regions in the searching space that have low energy. When the global minimum is in a very small region surrounded by a large population of energetically unfavorable configurations, occurrence of trials that happen to jump into the global minimum region are of very low probability. Such a phenomenon is known as a "golf course" [35], and is difficult to optimize by practical SA methods. Consider an example in Figure 4.2, which deals with the segmentation for two objects (classes) corresponding to the shaded and unshaded respectively. The configuration 4 is the truth with the lowest energy, while the current estimate is configuration 1 corresponding to a local minimum. Let E_i denotes the energy of configuration i and suppose $E_4 < E_1 < E_2 < E_3$. To obtain the correct result from the current estimate, consecutive flips of the three center pixels is required, the occurrence of which is not highly probable. In the search space, the area of energetically unfavorable configurations around the global minimum increases exponentially with the number

of pixels needed to be flipped, which, if large, makes the sampling in the global minimum region almost impossible.

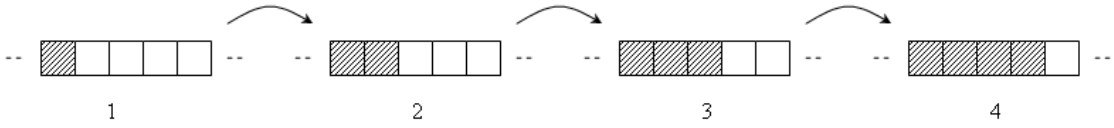


Figure 4.2: An example that SA may fail

For improving on such a "golf course" situation, or relevantly the rate of convergence, large moves in the searching space is necessary. This can be achieved by performing configuration changes on groups of pixels instead of each single pixel during the optimization process. The optimization is thus performed on a hierarchical organization of the pixels, and the grouping structure of pixels can be either fixed [12][52][62][115] or data-adaptive [13][114].

4.2 Proposed Methodology - IRGS

This section presents a novel segmentation method which tries to improve on two aspects compared to the traditional MRF approach: the adaptivity to image non-stationarity and the convergence rate in finding a satisfactory solution. In the traditional MRF approach formulated in (4.3), both the feature model and spatial context model are stationary, with model parameters estimated globally. Such a formulation is limited for non-stationary situations common for SAR sea ice imagery due to incidence and environmental factors. The proposed method incorporates edge strength, which has some capability of describing local behaviors and can improve the spatial context model on its adaptivity to the non-stationarity of the image. The solution is obtained by minimizing a defined energy function iteratively, with each iteration consisting of a region merging process and a classification

process combined in an interactive manner. Since the classification is performed on regions instead of single pixels, long range of spatial interactions is accounted efficiently. Here the classification is a general purpose clustering approach not associated with any meaningful ice identifications, as this chapter deals with the low level segmentation only. However, the proposed method provides the possibility of replacing the clustering process with a SAR sea ice classification process so that various high level domain knowledge can be utilized to guide the low level segmentation, as described in chapter 5. With respect to Figure 3.1, the proposed segmentation process corresponds to the black path from the original image to the region clusters, and future replacement by high level SAR sea ice specific classification corresponds to changing the path to the green one to form a complete image understanding system.

There has been a significant amount of work on hybrid region and edge segmentation [13][47][74][107] and energy based region merging [45][94][107]. The proposed IRGS method is different from others in that it uses a sequence of edge penalty functions to approach to the traditional MRF objective function and allows further integration of domain specific classification.

4.2.1 Hybrid Region and Edge Energy Function

The proposed hybrid region and edge energy function is a modification of the traditional MRF function (4.3). Since the term $\beta_0 U_i(s)$ is nonzero only when the site s is at the boundary, Equation (4.3) is equivalent to

$$\arg \min_{C_1,\ldots,C_n} \left\{ \sum_{i=1}^{n} \sum_{s \in C_i} \left\{ \frac{1}{2} \log(|\Sigma_i|) + \frac{1}{2} (\mathbf{y}_s - \mu_i)^T \Sigma_i^{-1} (\mathbf{y}_s - \mu_i) \right\} + \beta_0 \sum_{i=1}^{n} \sum_{s \in \partial C_i} U_i(s) \right\} \tag{4.5}$$

where ∂C_i is the set of boundary sites belonging to region cluster C_i (i.e. those having at least one neighbor not belonging to C_i, shown as the white sites in Figure 4.3(a)).

Instead of penalizing equally for all boundary site pairs in (4.5), a greater penalty can be applied to weak edge and a lesser penalty to strong edge, so that local statistic such as edge strength can be incorporated. Therefore, the penalty term can be replaced with some monotonically decreasing function of edge strength $g(|\nabla|)$ (defined in 4.2.2), where $|\nabla|$ is the strength of the edge between the two neighboring sites astride the boundary. The proposed objective function can thus be defined as

$$\arg \min_{C_1,\ldots,C_n} \left\{ \sum_{i=1}^{n} \sum_{s \in C_i} \left\{ \frac{1}{2} \log(|\Sigma_i|) + \frac{1}{2} (\mathbf{y}_s - \mu_i)^T \Sigma_i^{-1} (\mathbf{y}_s - \mu_i) \right\} + \beta_0 \sum_{i=1}^{n} \sum_{s \in \partial C_i} \sum_{t \in \eta_s, t \notin C_i} g(|I_s - I_t|) \right\} \tag{4.6}$$

where η_s is the neighborhood of site s, and $|I_s - I_t|$ gives the edge strength between the two neighboring sites s and t.

As $\sum_{i=1}^{n} \sum_{s \in \partial C_i} U_i(s)$ in (4.5) is closely related to the length of the boundaries, the role of the spatial context model can be viewed as penalizing the existence of boundaries by their populations. Therefore, the problem can be approximated by

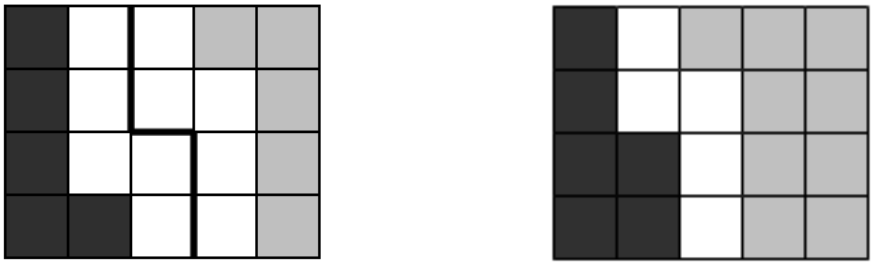
$$\arg \min_{C_1,\ldots,C_n} \left\{ \sum_{i=1}^{n} \sum_{s \in C_i} \left\{ \frac{1}{2} \log(|\Sigma_i|) + \frac{1}{2} (\mathbf{y}_s - \mu_i)^T \Sigma_i^{-1} (\mathbf{y}_s - \mu_i) \right\} + \beta_1 \sum_{i=1}^{n} \sum_{s \in \partial C_i} 1 \right\} \tag{4.7}$$

The ∂C_i here in (4.7) is defined to be the sites separating C_i from other classes instead of the inner boundary in (4.5). That is, for the n region clusters $C_1, \ldots, C_n$

and the image space S,

$$\begin{aligned} &a) \bigcup_{i=1}^{n} C_i \bigcup_{i=1}^{n} \partial C_i = S, \\ &b) \forall i, j : C_i \cap \partial C_j = \emptyset, \\ &c) \forall i \neq j : C_i \cap C_j = \emptyset, \\ &d) \forall i : \partial C_i \subseteq \bigcup_{j=1, j\neq i}^{n} \partial C_j, \end{aligned} \tag{4.8}$$

Figure 4.3(b) gives an example of a boundary defined in such a manner. Such a boundary definition on the regular image lattice is more convenient for boundary manipulations than the double inner boundaries in (4.5), and hence will be used hereafter in this thesis. The β_1 in (4.7) is also a different positive number from the β_0 in (4.5). As the majority part of the boundaries are expected to be smooth, the $U_i(\cdot)$ are mostly 3 [1]. Therefore, the β_1 should be approximately 3 times the β_0.



(a) (b)

Figure 4.3: Boundary definitions: (a) the boundary is the bold black line; (b) the boundary is defined on the image lattice, shown as the white sites.

The proposed objective function corresponding to the new boundary definition

[1] In this thesis, the 8 nearest neighbor system is used

then becomes

$$\arg \min_{C_1,\ldots,C_n} \left\{ \sum_{i=1}^{n} \sum_{s \in C_i} \left\{ \frac{1}{2} \log(|\Sigma_i|) + \frac{1}{2} (\mathbf{y}_s - \mu_i)^T \Sigma_i^{-1} (\mathbf{y}_s - \mu_i) \right\} + \beta_1 \sum_{i=1}^{n} \sum_{s \in \partial C_i} g(|\nabla_s|) \right\} \tag{4.9}$$

where $|\nabla_s|$ is the gradient magnitude at site s. Compared to (4.6), the implementation can be much simpler.

Our method of penalizing differently on the edge strength has the shortcoming of bias on certain classes since boundaries between some classes can be generally weaker than those of others. However, including edge strength is advantageous in describing local behaviors, and as shown in the next section, the bias problem can be alleviated by properly manipulating the edge penalty function $g(\cdot)$.

4.2.2 Choosing Edge Penalty Function

The edge penalty function $g(\cdot)$ can be any monotonically decreasing function, so that the greater the edge strength, the smaller the penalty. Suppose the gradient $|\nabla_s|$ for any site s has been normalized to [0,1]. Then, the penalty function can be formulated as:

$$g(|\nabla_s|) = e^{-(|\nabla_s|/K)^2} \tag{4.10}$$

The parameter K (Fig. 4.4) defines how fast the edge penalty decays with the increase of edge strength. As K increases, the penalty difference between weak and strong edges decreases. When K approaches infinity, all edge penalties are equally 1.

Due to the high variability of the SAR sea ice images, a single value of K may

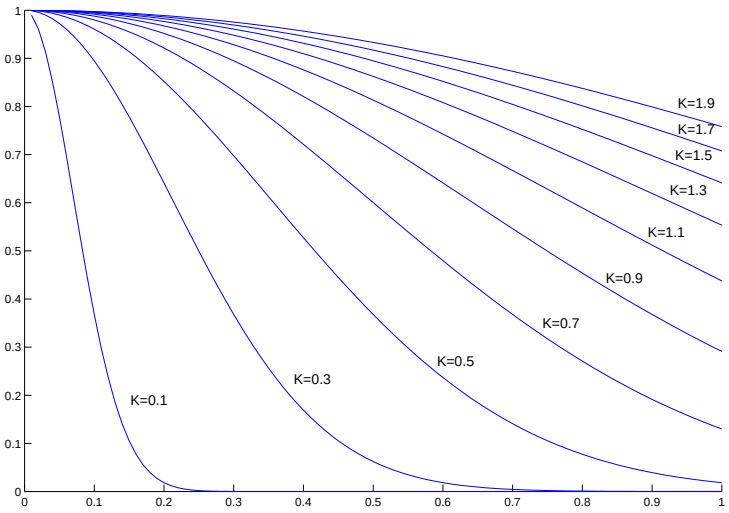


Figure 4.4: Edge penalty functions defined by Equation (4.10)

not work well for all the ice types and scenes. Therefore, the parameter K of the edge penalty function $g(\cdot)$ is allowed to increase during the segmentation, details of which are described later in the next section. The local feature of edge strength is hence utilized, but the bias is significantly reduced.

4.2.3 Optimization

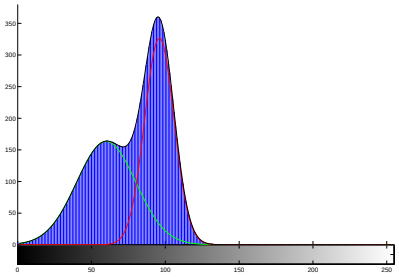
Finding the global minimum of (4.6) or (4.9) is expected to give a satisfactory segmentation. As it is a combinatorial optimization problem, techniques such as simulated annealing [40] can be used. However, due to the impossibility of implementing an ideal optimization process, the solution searching may be trapped into some local minima. If the image is non-stationary, the local minima are easily located far away from the desired solution and convergence to such local minima is intolerable.

Consider the image in Figure 4.5(a), which will also serve as an example in the algorithm testing. Figure 4.5(a) is an image that has two distinct regions, one relatively bright and one relatively dark. Both clusters have Gaussian distributed intensities, but the intensity distribution is non-stationary with a gradual fading effect. Such a fading effect is common in SAR sea ice imagery due to incidence. Figure 4.5(d) gives the segmentation result using classical MLL-based approach defined in (4.3) whose Gaussian function parameters are estimated by the expectation-maximization technique [28] and the β_0 is chosen to be 18. It is quite obvious that the globally estimated statistics for the clusters are not sufficient for the segmentation due to the non-stationarity of the image. Therefore, local information of edge strength is incorporated as in Equation (4.6), where K is 0.5. The simulated annealing is used for finding the solution and the segmentation result is given in Figure 4.5(e). Interestingly however, there is no visible improvement. The region of the global minimum of (4.6) is too small and surrounded by energetically unfavorable configurations ("golf course" [35] as already mentioned before), thus making the SA algorithm inefficient.

Aiming at reducing the relative population of the unfavorable configurations surrounding the global minimum, the perturbation unit is changed from a single pixel to a homogeneous region. The idea is intuitively simple. Pixels or regions that have similar local statistics are grouped together first, and then labelling is performed on the obtained regions using global statistics. Back to the example of Figure 4.5, it is expected that pixels belonging to the left part can be grouped into one region and those belonging to the right part into another region before they are labelled (classified). This leads to the combined region merging and classification approach.

(a) An example image.

(b) The truth. Energy: 65539.2



(c) The histogram.

(d) By the MLL-based approach. Accuracy: 88.2%.

(e) By SA on the objective function in (4.6). Energy: 76235.3; Accuracy: 88.41%

Figure 4.5: Segmentations on an example highly non-stationary image. The left region has a mean grey level of 60 with the standard deviation of 10; The right region has a mean grey level of 96 with the standard deviation of 20. Energies defined by (4.6) are computed with $\beta_0 = 18$ and $K = 0.5$.

Region Merging

For the optimization purpose, the region merging process in our approach should aim at reducing the overall energy function. However, the energy function of (4.9) needs to be modified for the convenience of the merging process. The notation of region clusters is replaced by disjoint regions. That is, the following objective function is used instead of (4.9).

$$\arg\min_{\Omega_1,\ldots,\Omega_n} \left\{ \sum_{i=1}^{n} \sum_{s\in\Omega_i} \left\{ \frac{1}{2}\log(|\Sigma_i|) + \frac{1}{2}(\mathbf{y}_s - \mu_i)^T \Sigma_i^{-1} (\mathbf{y}_s - \mu_i) \right\} + \beta_1 \sum_{i=1}^{n} \sum_{s\in\partial\Omega_i} g(|\nabla_s|) \right\} \tag{4.11}$$

Here n is the number of segmented regions, μ_i is the mean feature vector of region Ω_i, and Σ_i is the feature covariance matrix of region Ω_i. During each merging operation, the overall energy for the resulting configuration is computed and compared with that before merging. If there is a decrease, the merging is justified.

The energy of (4.11) need not be calculated every time when determining whether or not a pair of segments should be merged. There is only a need to compute the difference between the total energy of the two neighboring segments tested and the energy of the one after merging. Suppose two regions Ω_i and Ω_j are being investigated, and let $\Omega_k = \Omega_i \bigcup \Omega_j$. Since

$$\sum_{s\in\Omega_i} \left\{ \frac{1}{2}(\mathbf{y}_s - \mu_i)^T \Sigma_i^{-1} (\mathbf{y}_s - \mu_i) \right\} = \frac{1}{2} M N_i \tag{4.12}$$

where M is the dimension of feature vector and N_i is the number of pixels belonging

to Ω_i, influence of this term is cancelled. The energy difference δE is thus

$$\frac{1}{2}\sum_{s\in\Omega_k}\log(|\Sigma_k|)-\frac{1}{2}\sum_{s\in\Omega_i}\log(|\Sigma_i|)-\frac{1}{2}\sum_{s\in\Omega_j}\log(|\Sigma_j|)-2\beta_1\sum_{s\in\partial\Omega_i\bigcap\partial\Omega_j}g(|\nabla_s|) \quad (4.13)$$

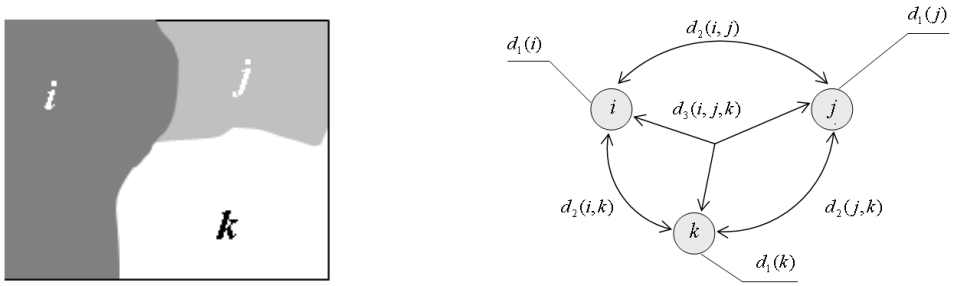
If (4.13) gives a negative value, Ω_i and Ω_j can be merged. If (4.13) gives a positive value, Ω_i and Ω_j are not merged.

A sequential merging order needs to be defined since the merging cannot be performed simultaneously. A natural way is to find the two segments whose energy decrease most if merged, and merge them first. The merging begins on an initial configuration obtained by a watershed algorithm [110], and is iterated until the energy cannot be reduced any more.

Classification

The classification process described here is general purpose without any application specific knowledge, and is usually called "clustering" rather than the term of "classification". With the obtained segments after region merging, a region adjacency graph (RAG) [67] is first built. The RAG is a graph with nodes representing segmented regions and links representing the existence of a common boundary between segments, as the example in Figure 4.6. In this thesis, only unary properties and binary relations are considered.

The MRF is again used to model the dependence among the nodes. As already introduced in section 3.1.1, the summation of clique energies gives a modelling of the probability of the corresponding classification. To make the region based classification generate reasonable solutions to (4.9), the clique energy functions are designed to make the summation of the energies have the same formulation as (4.9).



(a) (b)

Figure 4.6: The segmented regions and the corresponding Region adjacency graph: (a) the segmentation; (b) the RAG with unary properties (d_1), binary relations (d_2), and triple relations (d_3) (from [67]).

The single-node clique energy is thus defined as

$$V_1(x_i = k) = \sum_{s \in \Omega_i} \left\{ \frac{1}{2} \log(|\Sigma_k|) + \frac{1}{2} (\mathbf{y}_s - \mu_k)^T \Sigma_k^{-1} (\mathbf{y}_s - \mu_k) \right\} \tag{4.14}$$

and the pair-node clique energy is

$$V_2(x_i = k, x_j = l) = \begin{cases} 2\beta_1 \sum_{s \in \partial\Omega_i \bigcap \partial\Omega_j} g(|\nabla_s|) & \text{if } k \neq l \\ 0 & \text{otherwise} \end{cases} \tag{4.15}$$

where x_i is the cluster label of region Ω_i, μ_k and Σ_k are the mean vector and covariance matrix of cluster k respectively.

Stochastic methods are then used to minimize the defined energy and assign cluster labels to the regions. The implementation uses the Gibbs sampling [40], which is a statistical sampling technique especially useful for solving combinatorial optimization problems. The Gibbs sampler considers univariate conditional distributions - the distribution when all of the random variables but the current are assigned fix values. Maximizing the complex joint a posterior, or equivalently minimizing the energy of (4.9), is thus decomposed into manipulations of local con-

ditional distributions. The process chooses the class label x_i of any region Ω_i for a value k based on the probability $\frac{1}{Z}e^{-E(x_i=k)/T}$ where Z is a normalizing factor and T is called *temperature* that controls the randomness of the realizations. Unlike the simulated annealing [40] which performs the sampling iteratively while decreasing the temperature T, the Gibbs sampling is applied only once for each region in a classification. Detailed explanation is presented in the next section [p. 67].

Iterative Region Growing Using Semantics (IRGS)

As already mentioned [p. 57], using a single penalty function will produce bias. Different classes that have fuzzy boundaries between could be merged, and a class that has high intra-class variation or noise could be split. Both could result in poor performance of the system, even though in the latter situation the split parts could probably be assigned the same class label in the subsequent classification.

Therefore, the combined merging and classification process is iterative. The diagram for the overall algorithm is shown in Figure 4.7. At first, a watershed segmentation is performed to obtain an initial segmentation, upon which the region adjacency graph (RAG) is built. The iterative process begins with a classification (clustering) on the obtained segments ($\Omega_i, i = 1 \dots n$) using Gibbs sampling. With the classification result, it is possible to determine whether or not two segments Ω_i and Ω_j are allowed to be merged (depending on whether or not they belong to the same class), and then set or clear the corresponding flag accordingly

$$\text{flag}_{ij} = \begin{cases} \text{true}, & x_i = x_j \\ \text{false}, & \text{otherwise} \end{cases}$$

However, the decision that two segments are allowed to be merged does not mean

that they should be merged. Among them, only the merging that will reduce the energy of (4.9) are executed. That is, two neighboring segments Ω_i and Ω_j will be merged only when flag_{ij} is true and δE computed by (4.13) is less than 0. The merging is in a greedy best-first order, i.e. the region pair whose merging will reduce the energy the most is combined first. When all the qualified region pairs have been merged, a new iteration begins with an increased edge penalty. The iteration continues until no merging is possible or a maximum number of iterations has been reached.

In the process, two segments are not allowed to be merged if they belong to different classes, the purpose of which is to suppress the merging between segments of different classes that have weak boundaries in between. This concept is known as semantic region growing [37][102]. Here the merging and classification are iterative and, as such, is referred to as *iterative region growing using semantics* (IRGS). High level knowledge can thus play a role in the segmentation process via classification. Although the classification used in this chapter is based only on low level features, domain specific classifications that using high level knowledge will be integrated into this system in chapter 5.

For the classification in each iteration, the Gibbs sampling is only applied once for each region. The resulting processing speed is much faster than that of using a complete annealing scheme for each classification, which is unnecessary since there is still an outside iteration loop. It should be noted that such a sampling without a complete cooling process causes a limited degree of randomness. This randomness is deliberately allowed because the segmentation is often inaccurate at the first several or tens of iterations. High level features (especially shape) extracted on an inaccurate segmentation result will make the classification not accurate either. It is necessary to let segments have the possibility of being merged for situations where

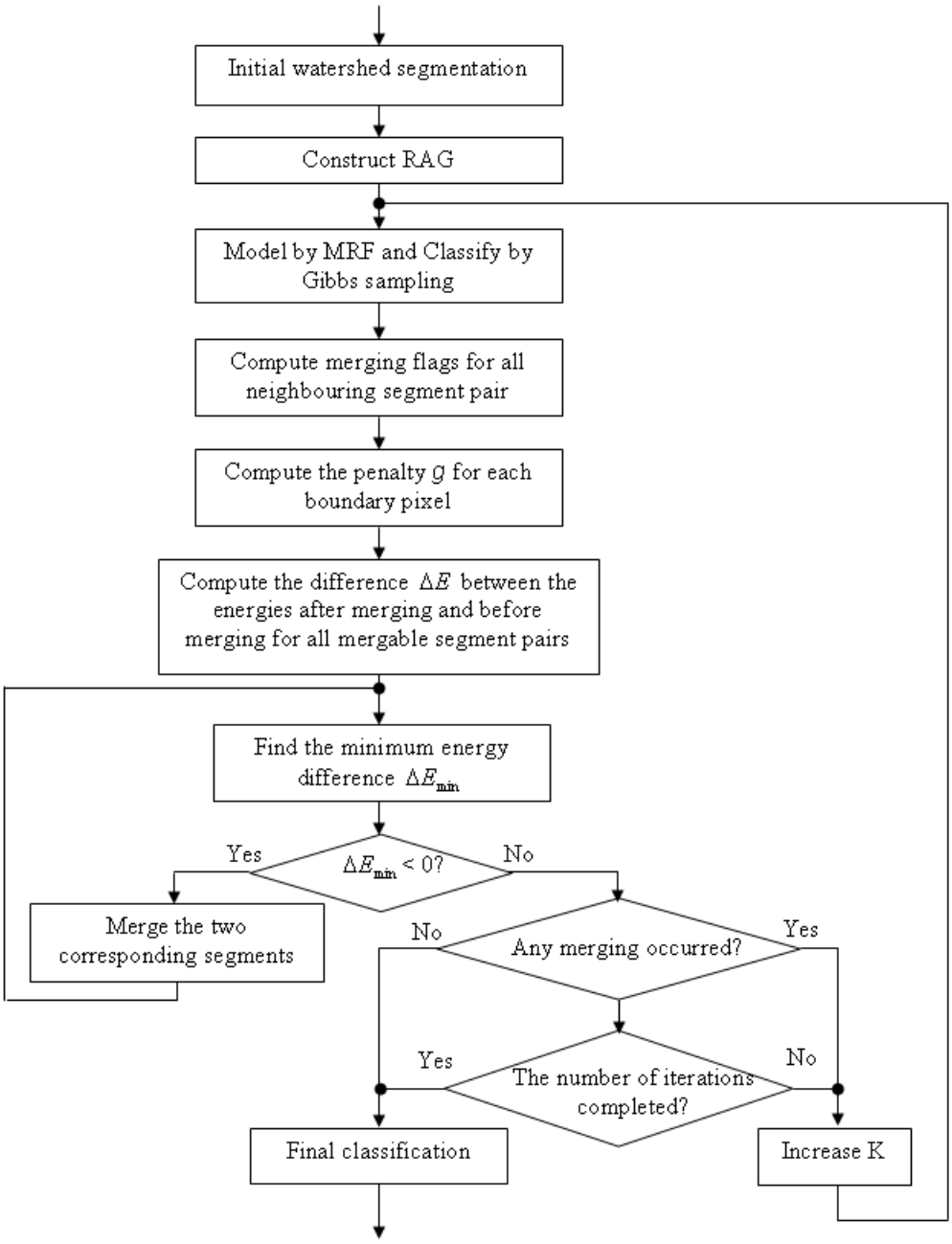


Figure 4.7: System diagram of IRGS

the pair of segments would always be incorrectly labelled as different classes if no randomness is allowed. From the optimization point of view, this randomness can help the solution searching "jump" out of local minima. Although the randomness of classification could probably increase the occurrence of merging between segments actually belonging to different classes, it is not expected to be serious. The randomness decreases quickly when the sizes of regions grow, and hence is relatively high only at the first several iterations of merging where only segments with the weakest boundaries are qualified to be merged.

4.2.4 Another Point of View

To view the proposed method from another point is helpful. Consider the graduated non-convexity (GNC) method which is proposed in [49] for surface reconstruction. The GNC is a deterministic method for computing the global minimum of the non-convexity energy function. It first approximates the energy function with a new convex function. The minimum obtained on the new function thus has the lowest energy globally, but may not correspond to the true global minimum of the original non-convex function. A whole sequence of energy functions is then used to approach, beginning from the convex approximation, to the original non-convex function. For each energy function in the sequence, a local minimum is computed descending from the location obtained in the previous function. The final local minimum is expected to be the true global minimum if any two sequential functions are sufficiently similar. Figure 4.8 illustrate the GNC process.

Similarly, an algorithm can be designed that searches the solution of (4.7) using a sequence of functions defined by (4.9) with the edge penalty changed gradually. The algorithm is summarized as in Table 4.1.

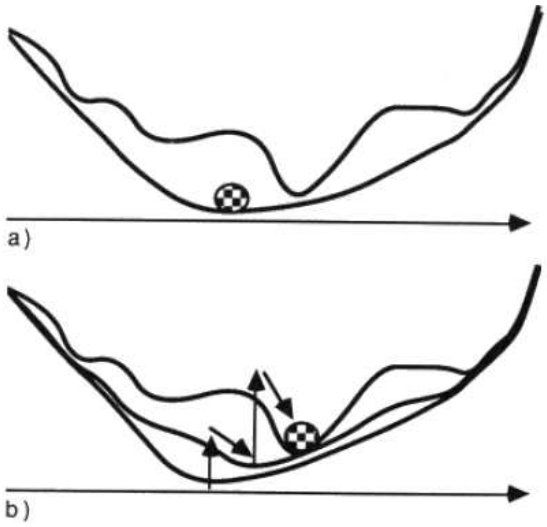


Figure 4.8: The minimum of a non-convex function F may be found by the GNC algorithm, which starting from a convex approximation F^* (a) runs downhill on each of a sequence functions (b), to reach the global minimum. Adapted from [49].

Starting from the extreme case of Gaussian mixture problem [2] in (3.5), the algorithm in Table 4.1 introduces more and more edge penalty during the solution searching. When K is sufficiently large, the energy function is close to (4.7) and an approximate solution to (4.7) is obtained.

Different from the GNC in [49], the initial function of the sequence (the Gaussian mixture) is non-convex. The first minimum obtained, which is the starting position in the search space for subsequent optimization process, may not be globally optimal. Moreover, the sequence of functions in the form of (4.9) does not belong to the cases where GNC is known to be able to find the global minimum (see [49] for details). Regardless of these, the algorithm in Table 4.1 does produce satisfac-

[2]The extreme case corresponds to $K = 0$, where the edge penalty function has a nonzero value only at the zero gradient point. Therefore strictly speaking, it is not an exact Gaussian mixture problem, but one with penalties on situations where neighboring sites with zero edge strength between them belong to different classes. Obviously, such situations can hardly happen in the solutions, and the two problems are approximately equivalent.

1. Construct a RAG. Each node in the graph can be either a pixel or a region obtained by an initial over-segmentation process (for example, watershed segmentation).
2. Based on the RAG, model the regions and their dependencies with MRF defined in (4.14) and (4.15).
3. Classify the regions by finding the minimum of the MRF energy using a greedy method such as the iterated conditional mode (ICM) [9].
4. If the configuration difference between the result and that of the previous iteration is ignorable, or a maximum number of iterations have been reached, the overall process is completed. Otherwise, increase K for the edge penalty function in (4.15) and go back to step 2.

Table 4.1: Algorithm of graduated increased edge penalty segmentation (GIEP).

tory results for many images as shown later in experiments. A similar MRF-based approach proposed by Deng and Clausi [29] has also used a sequence of energy functions during the segmentation process. In their approach, the weight for the context model is increased gradually as shown in Figure 4.9. The main difference is that the shape of edge penalty function used here is a Gaussian function while theirs is constant over the gradient magnitude.

The IRGS in Figure 4.7 can be viewed as a modification of the algorithm in Table 4.1 with an extra merging step added in each iteration. The dimension of the searching space is thus reduced by the merging of similar regions. This could sometimes cause an inaccurate location of the global minimum. However, the resulting larger moves in the search space by region merging is beneficial to the convergence rate and often advantageous on "golf course" situations as already explained earlier.

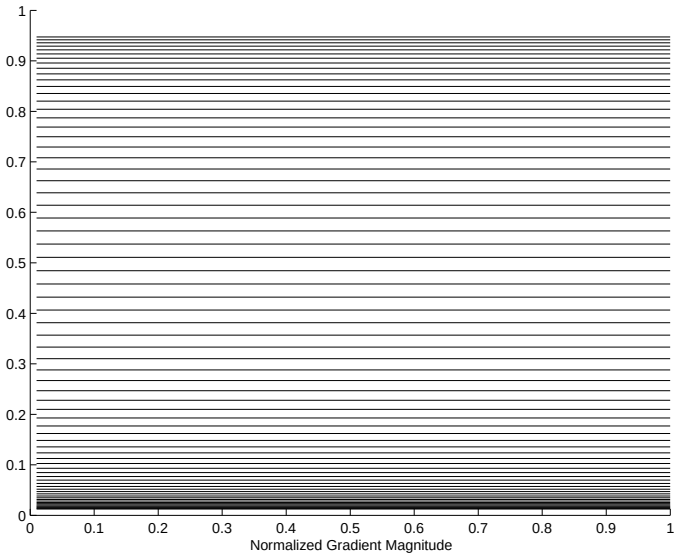


Figure 4.9: Context model weights

4.2.5 Parameter Estimation

The objective function of (4.9) consists of two summation terms, one corresponding to the feature model which is assumed to be Gaussian and the other corresponding to the edge penalty. Parameters involved include the multivariate Gaussian function parameters μ and Σ, the edge penalty weight β_1, and the increment schedule of the edge penalty function parameter K.

Feature Model Parameters

The assumption of the feature model being Gaussian means that the distribution of the feature vector $\mathbf{y}_s$ is a mixture of n Gaussian functions, where n is the number of classes. The model parameters are first estimated based on the equation (3.5) without considering the spatial context (or edge penalty). Such a Gaussian mixture problem is typically solved by the expectation-maximization (EM) algorithm [28].

Let π_i denote the prior for class Ω_i, and w_{si} represent the probability of site s belonging to class Ω_i given the observation data and current estimate of all parameters. The EM algorithm iterates through computing the E step:

$$w_{si} = \frac{p(\mathbf{y}_s|\mu_i, \Sigma_i)\pi_i}{\sum_i p(\mathbf{y}_s|\mu_i, \Sigma_i)\pi_i} \tag{4.16}$$

and the M step:

$$\mu_i = \frac{\sum_s w_{si}\mathbf{y}_s}{\sum_s w_{si}} \tag{4.17}$$

$$\Sigma_i = \frac{\sum_s w_{si}(\mathbf{y}_s - \mu_i)^T(\mathbf{y}_s - \mu_i)}{\sum_s w_{si}} \tag{4.18}$$

$$\pi_i = \frac{\sum_s w_{si}}{\sum_i \sum_s w_{si}} \tag{4.19}$$

The obtained parameters by the above EM are just used initially. They are further updated in each iteration of IRGS, where the parameters are still computed with the M step, but w_{si} of each site s takes value of either 1 or 0 based on whether the provisional class label estimate of s is i or some other class. Although little is known of the influence of the Gaussian mixture solution accuracy on that of (4.7), finding the global minimum of the Gaussian mixture problem is still desirable. It is well known that EM is sensitive to the initial configuration, especially in a multivariate context [10]. In this thesis, a large number (60) of different initial values are chosen randomly. After a small number of iterations, the one with the smallest energy is chosen and refined further by more iterations to obtain the solution.

Edge Penalty Weight β_1

Since the edge penalty weight β_1 is closely related to the MLL model parameter β_0, the estimation of β_0 is consider first. Theoretically sound estimation is difficult and ad hoc schemes that assign experimentally satisfactory values to the parameter are commonly used. For example, Rignot and Chellappa [90] reported that [1.0-1.6] is the best range for β_0 in polarimetric SAR image segmentation. Yue [119] concluded that β_0 could be set between 1 and 3 based on the segmentation experiments on SAR sea ice. The difficulty of the problem is due to the intractability of the normalizing constant of the distribution, known as the partition function, that makes direct maximum likelihood (ML) estimation of β_0 impossible. There are some efforts that try to maximize an approximation of the likelihood [9][63] or use a Monte Carlo scheme [32] to avoid the computation of the partition function.

From intuition, the β_0 should be related to the length of the boundary expected. Larger penalties should be given to images that are expected to be simple, and vice-versa. Descombes et. al. [32] give a mathematical formulation of the relationship between the β_0 and the boundary length using maximum likelihood estimation. This is briefly presented here. Given an image $\mathbf{x}$ representing the random field of labels, let $N(\mathbf{x}) = \sum_{s\in\Omega_i} U_i(s)$, where the definition of $U_i(\cdot)$ is the same as (4.3). Therefore, $N(\mathbf{x})$ is the total number of pairs of neighboring sites that are of different labels, and is approximately proportional to the boundary length. The MLL modelled prior is:

$$P(\mathbf{x}|\beta_0) = \frac{1}{Z(\beta_0)} \exp(\beta_0 N(\mathbf{x})) \tag{4.20}$$

where $Z(\beta_0) = \sum_{\mathbf{x}} \exp(\beta_0 N(\mathbf{x}))$. The maximum likelihood estimate is given by:

$$\frac{\partial \log P(\mathbf{x}|\beta_0)}{\partial \beta_0} = N(\mathbf{x}) - \frac{Z'(\beta_0)}{Z(\beta_0)} = 0$$

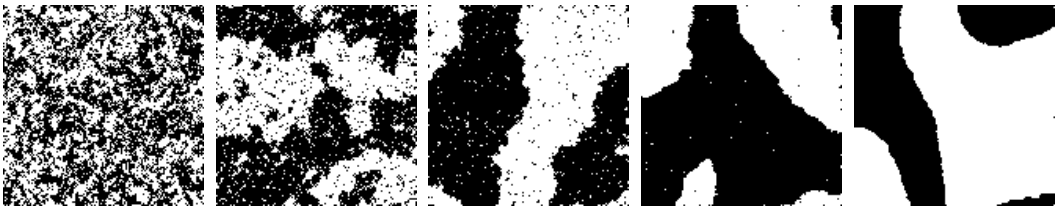
And therefore,

$$N(\mathbf{x}) = \frac{Z'(\beta_0)}{Z(\beta_0)} = \frac{\sum_{\mathbf{x}} N(\mathbf{x}) \exp(\beta_0 N(\mathbf{x}))}{Z(\beta_0)} = \sum_{\mathbf{x}} N(\mathbf{x}) P(\mathbf{x}|\beta_0) \tag{4.21}$$

Equation (4.21) shows that β_0 should choose the value, of which the resulting priors have a mean boundary length (population) equivalent to the expected boundary length (population) of the segmentation. Since β_0 is only related to the boundary length, it can also be concluded that the choice of β_0 has little dependence upon many application specific properties of data such as the noise strength, although the noise strength does influence the accuracy of the boundary length estimation.

Determining the β_0 during the segmentation can thus be achieved by simply estimating the boundary length and looking up the corresponding value in a predetermined graph. Computation of the lookup graph is based on a Monte Carlo scheme since (4.21) is intractable. Twenty five values between 0 and 3 are chosen for β_0, and five hundred 128×128 samples are generated for each value. Figure 4.10 shows several samples obtained from different β_0 values, and Figure 4.11 shows the resulting interpolated graph that gives the correspondence between β_0 and the ratio of the mean boundary length over the image size.

With β_0 known, β_1 is simply set to be $3\beta_0$. Usually β_1/β_0 is a little smaller than 3 since the boundary cannot always be straight. However, such a value is chosen as it is reasonable and encourages uniform regions.



(a) $\beta_0 = 0.3$, $r_{B/R} =$ 0.464172 (b) $\beta_0 = 0.4$, $r_{B/R} =$ 0.245199 (c) $\beta_0 = 0.5$, $r_{B/R} =$ 0.093506 (d) $\beta_0 = 0.8$, $r_{B/R} =$ 0.029073 (d) $\beta_0 = 2.0$, $r_{B/R} =$ 0.020732

Figure 4.10: Samples obtained from different β_0 values. $r_{B/R}$ denotes the ratio of populations of boundary sites over region sites

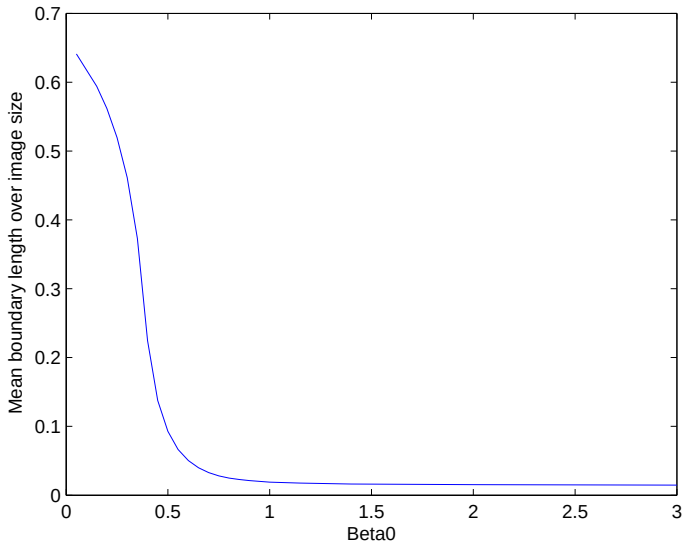


Figure 4.11: Relationship between β_0 and the ratio of mean boundary length over image size

Beginning from a defined initial value, β_0 and β_1 are refined during each iteration based on the boundary length estimated from the provisional segmentation result. In practice, the boundary length estimation is heavily influenced by the noise strength. A noisier image often corresponds to a noisier initial configuration for the optimization process, and if the optimization is easily trapped into a local minimum, an over-segmentation is likely to occur. This situation will cause the β_0 computed from the segmentation result to be smaller than it should be, and will also favor over-segmented configurations in turn. The experiments demonstrated that except the proposed IRGS approach all other tested methods are likely to give such overestimates of expected boundary length. Thus, the β_0 is deliberately set to be much larger, six times the original estimate, for methods other than IRGS.

Edge Penalty Incremental Schedule

The incremental schedule for the penalty function parameter K in (4.10) is desired to be very slow in order to approximate well the approaching from the Gaussian mixture problem of (3.5) to the final objective function of (4.7). On the other hand, the processing time is required to be tolerable for the method to be practically useful. A tradeoff has to be made. The parameter K is simply chosen to be $K^{(t+1)} = 1.02K^{(t)} + 1/255$, where $t = 0, 1, \ldots$ is the sequence number, and $K^{(0)} = 0$.

Gradient Magnitude Computation

In computing the gradient magnitude for (4.9), the original image is convolved with the first derivative of a Gaussian kernel in horizontal and vertical directions respectively to obtain the two directional derivatives. The variance of the Gaussian function influences the smoothness of the boundary obtained to a limited degree,

and for the experiments in this thesis it is set to be 1.0.

The Initial Watershed

The proposed IRGS has an initial watershed segmentation step, implemented as in [110]. The purpose of the watershed segmentation step is to generate regions for constructing the RAG. Although there is no reason why the initial RAG cannot be constructed from individual pixels, such a watershed based RAG construction can significantly speed up the process and reduce the memory requirement. However, this implies that in addition to the overall model for merging and classification a different model has been introduced in the process. It is thus desirable that the initial watershed does not influence the robustness and accuracy of the overall model. This is only possible when the desired segmentation solution belongs to the configuration subspace constrained by the watershed segmented regions, i.e. no pixels belonging to difference ice types have been grouped together by the watershed process. Therefore, over-segmentation that produces very small regions is deliberately chosen. The degree of over-segmentation is determined by the variance of the Gaussian filter for smoothing the gradient magnitude, of which the local minima as used as the seeds. In the implementation, the Gaussian filter has a variance of 1.0, and hence highly over-segmented results are obtained initially.

4.3 Algorithm Testing

The algorithm has a complexity of $O(nr)$, where n is the number of clusters and r is the number of regions in the initial RAG. Obviously, r depends on the complexity of the image contents. In the worst case where each pixel is a region, r is the image

size. To make the execution time acceptable, the algorithm is implemented with C++. For each of the three SAR images (1252×873) tested in this section, the program takes about 4-6 minutes on a P4-1.6G PC with 1G RAM.

4.3.1 Methods Comparison

The focus of the experiments in this section will be put on the comparisons between the IRGS described in Figure 4.7, the region-based graduated increased edge penalty (GIEP) segmentation described in table 4.1 in section 4.2.4, and the classical MLL approach whose objective function is (4.3) and optimization scheme is the simulated annealing (MLL-SA). For comparison, Deng and Clausi's method [30] is also included, which gradually changes the weighting between the feature model and the context model during the optimization process. For all the experiments in this section, only the tone (intensity) feature is used. The annealing schedule, if applied, is always $T^{(t)} = 0.98^t$.

All the above methods perform the feature model parameter estimation simultaneously with the segmentation process. During the segmentation iterations, the spatial context model plays a role in producing the intermediate results, and hence also in the updating of feature model parameters. However, it is unclear under what conditions or situations this role is positive. At the initial stages of iterations where the feature model parameters are far from the true values, incorporation of the spatial context model may produce meaningless segmentation, which in turn can make the resulting updated feature model parameters even worse. Considering that, Deng and Clausi's method gives a very small weight on the spatial context model initially, and then gradually increase the weight during the iterations. The GIEP and IRGS have also used a similar principle, i.e. gradually increasing spatial

context influence, based on the edge strength information. Here, the only method whose feature model parameter estimation needs to be carefully implemented is the MLL-SA. In this thesis, the MLL-SA first uses the feature model parameters estimated without considering the spatial context, as in a pure Gaussian mixture model (GMM) approach. Updating of the parameters is then performed only when a sufficient number of iterations have been completed to achieve low randomness in the segmentation, i.e. the temperature is lower than 0.1.

Testing 1: Non-stationary Synthetic Image

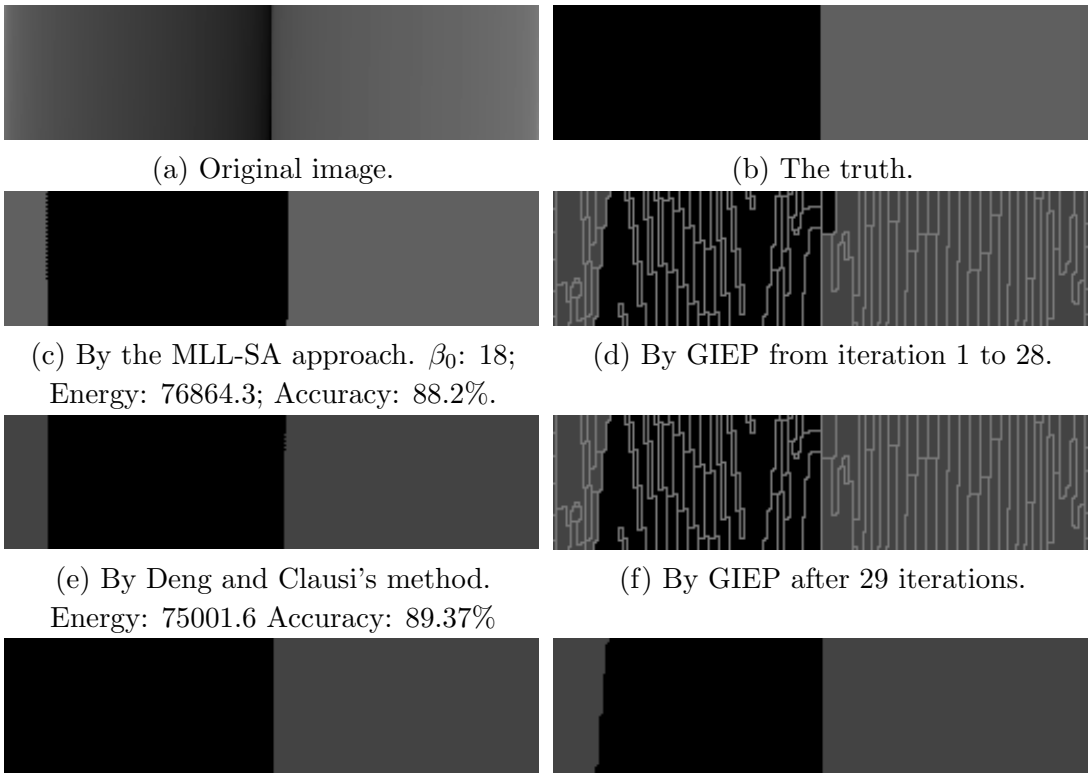
The first example investigated is the one in Figure 4.12(a) which, as already discussed in section 4.2.3, has a fading effect often found in SAR imagery. This example image is a difficult one because of such non-stationarity. Here it is deemed as difficult with respect to approaches that deal with the global behavior of tone or texture of the whole image. It is definitely true that more suitable models or methods, for example a simple edge detection followed by thresholding, can be designed for such a special case. However, preferred techniques are restricted to general global approaches that are relatively insensitive to noise and capable of describing classes in a global manner, which is desired for SAR sea ice discrimination.

As already shown in Figure 4.5(d) in section 4.2.3 and again in Figure 4.12(c), an undesirable result has been obtained by the MLL-SA approach defined in (4.3). This has already been discussed in section 4.2.3. A marginally improved result is produced by Deng and Clausi's method in Figure 4.12(e), where the boundary in the center of the image is not shifted as far as in Figure 4.12(c). The GIEP method is then applied using 200 iterations. The first 28 iterations generate a result shown in Figure 4.12(d). Here in Figure 4.12(d) and as well in 4.12(f), the bright lines are the boundaries of the regions obtained from the initial watershed

algorithm. At iteration 29, there is a change of label for a small region at the center top. Examining that small region, the left part of the region boundary has high gradient magnitude while the right and bottom part of the boundary has low gradient magnitude. The two parts of the boundary are hence penalized differently, but in the first several iterations the difference is too small to be effective. As K grows to 4.04/255 at iteration 29, the right and bottom part is penalized sufficiently larger than the left one so that the region has chosen the same label as its left neighbor. After that, K continues to grow but the penalty difference between the two decreases beginning from a certain point. At last, whether or not the region should be flipped back depends no longer on the gradient magnitude but on the length of the boundaries, and in this example a better configuration of the two is preserved in Figure 4.12(h).

Although an improved result has been obtained by the GIEP approach, there is still a large region of errors in the left of the image. The GIEP is incapable of flipping consecutively all the bright regions in the left end, and larger flipping units are desired. This is achieved by the step of region merging in the IRGS. As a result, the IRGS has produced the best configuration in Figure 4.12(g). Since both the IRGS and GIEP can be viewed as optimization techniques for solving (4.3), as already discussed in section 4.2.4, energies computed by (4.3) are compared among the methods [3]. The computed energies has demonstrated that GIEP's solution corresponds to a local minimum which is a little better than that of the MLL-SA approach, and IRGS locates a minimum with the lowest energy among all. This also justifies the formulation of (4.3) for the segmentation of the image, as the

[3]Since the β_0 are not fixed but estimated adaptively with the data, the estimated values can be quite different among the tested methods. To show the advantages of other methods over the MLL-SA approach in finding a better solution, the β_0 used in energy comparison is fixed as the one optimal to (estimated by) the MLL-SA approach.



(a) Original image.

(b) The truth.

(c) By the MLL-SA approach. β_0: 18; Energy: 76864.3; Accuracy: 88.2%.

(d) By GIEP from iteration 1 to 28.

(e) By Deng and Clausi's method. Energy: 75001.6 Accuracy: 89.37%

(f) By GIEP after 29 iterations.

(g) By IRGS. Energy: 68932; Accuracy: 99.61%.

(h) By GIEP. Region boundaries erased. Energy: 75234; Accuracy: 92.16%.

Figure 4.12: Segmentations on the example in Figure 4.5. Energies defined by (4.3) are computed with $\beta_0 = 18$.

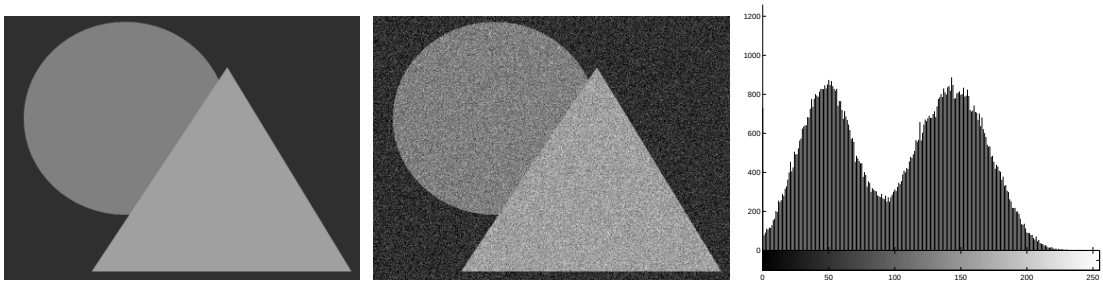
lowest energy obtained corresponds to the configuration with the highest accuracy.

Testing 2: Noisy Synthetic Image

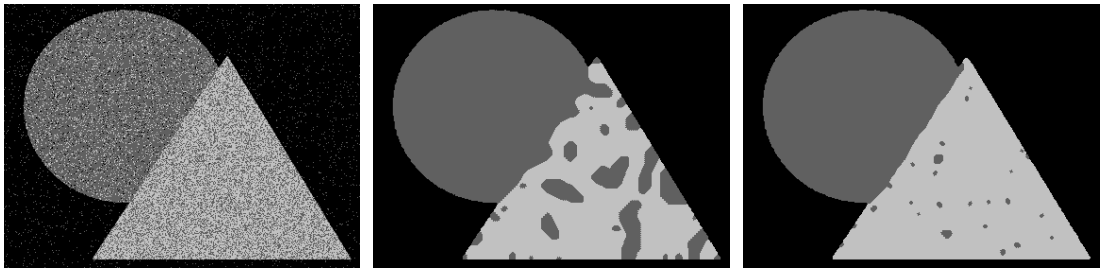
The next experiment is on a synthetic image shown in Figure 4.13(a). Zero mean Gaussian noise of variance 22^2 has been added in Figure 4.13(b), causing a significant overlapping of the intensity ranges between the triangle and circle objects. A Gaussian mixture segmentation based on minimizing the objective function in (3.5) is first applied. The EM technique is used for optimization and the corresponding result shown in Figure 4.13(d) is very noisy. In Figure 4.13(e), a better but still poor result with over-segmentation phenomenon occurring in the triangle has been produced by the MLL-SA approach. Deng and Clausi's method gives an improved result in Figure 4.13(f) but is still not satisfactory. Both methods of Figure 4.13(e) and 4.13(f) update labels pixel by pixel, and without an extremely slow annealing schedule can easily be trapped into a local minimum. Successful segmentations have been achieved by both GIEP in Figure 4.13(g) and IRGS in 4.13(h). The results of the two approaches are exactly the same, with a energy lower than those of others. Compared to the truth, the two methods have achieved 99.64% of accuracy.

Testing 3: Various Noise Strength and Region Size

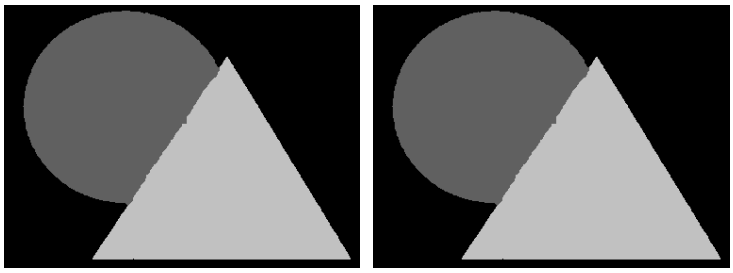
Experiments are performed on a set of synthetic images. In the synthesis of those images, two different intensities are chosen, one relatively dark for the background and one relatively bright for the foreground objects. Those foreground objects have round shape with a limited amount of deformation, and are generated at random locations in the image. With speckle noise added, the resulting images are expected to give a rough simulation of SAR sea ice scenes characterized by water and well



(a) An example image. (b) With zero mean Gaussian white noise ($\sigma = 22$) added. (c) Histogram of the image in (b).



(d) By the Gaussian mixture. Accuracy: 83.64%. (e) By the MLL-SA approach. β_0: 5.008; Energy: 524052; Accuracy: 92.48%. (f) By Deng and Clausi's method. Energy: 499389; Accuracy: 98.32%.

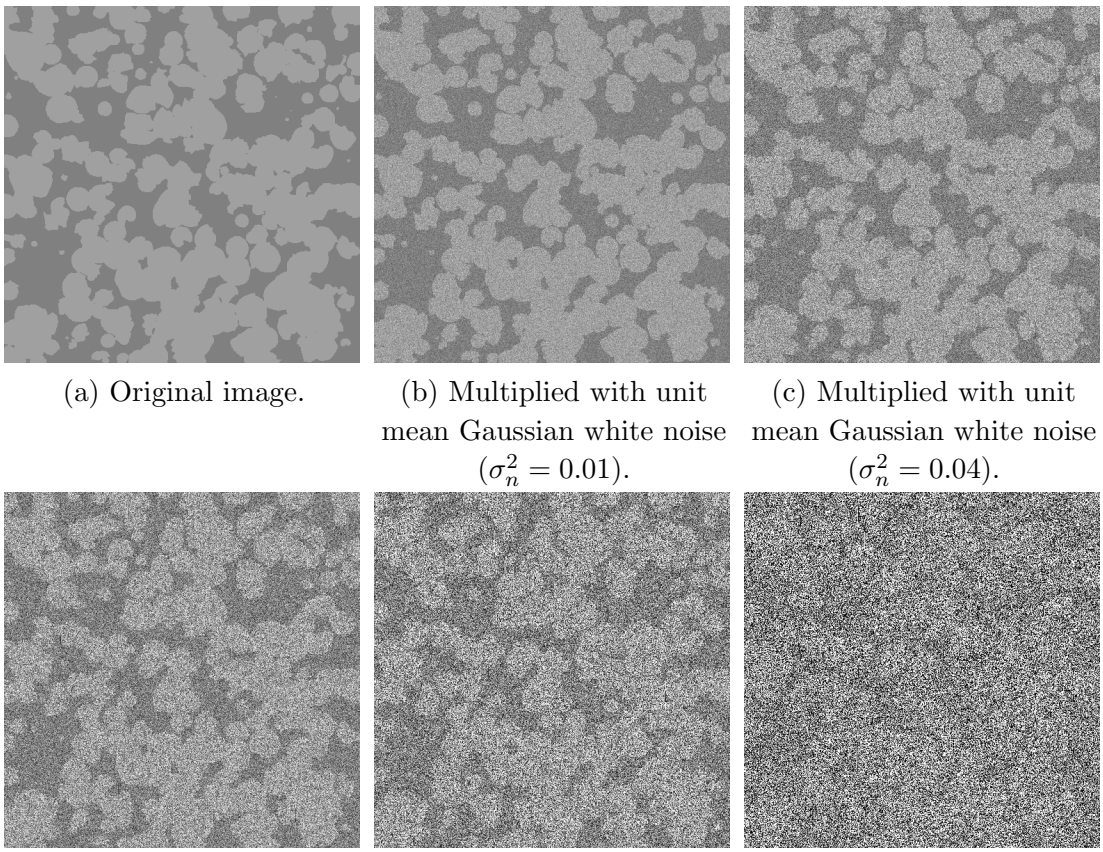


(g) By GIEP. Energy: 487194; Accuracy: 99.64%. (h) By IRGS. Energy: 487194; Accuracy: 99.64%.

Figure 4.13: Segmentations on an example synthetic noisy image. The three grey levels in the original image are 48, 128, 160 respectively.

defined ice floes.

For adding the speckle noise, Gaussian noise with mean of 1.0 is multiplied with the image intensities. Although SAR speckle is generally modelled as a Gamma distribution [93], multiplicative Gaussian noise is much more convenient to use and gives a reasonable approximation to SAR speckle.



(a) Original image.

(b) Multiplied with unit mean Gaussian white noise ($\sigma_n^2 = 0.01$).

(c) Multiplied with unit mean Gaussian white noise ($\sigma_n^2 = 0.04$).

(d) Multiplied with unit mean Gaussian white noise ($\sigma_n^2 = 0.08$).

(e) Multiplied with unit mean Gaussian white noise ($\sigma_n^2 = 0.2$).

(f) Multiplied with unit mean Gaussian white noise ($\sigma_n^2 = 0.8$).

Figure 4.14: Example 512×512 synthetic noisy images. The two grey levels in the original image are 128 and 160. The mean diameter of the simulated ice floes is 20 pixels.

Figure 4.14 shows an example simulated scene and the corresponding synthetic

images with different noise variance (σ_n^2). For the experiment, 16 different levels of noise are chosen, and 20 scenes different in floe shape and location are generated with each having an average floe diameter of 20 pixels. By combining each of the 20 scenes and each of the 16 different levels of noise, a total of 320 synthetic images are available for testing. Segmentations are then performed on these images using each method. Table 4.2 gives a summary of the average percentage accuracies of those methods on each noise level. When the noise variance is relatively low ($\sigma_n^2 \leq 0.03$), all the methods that have utilized spatial context (i.e. all except the Gaussian mixture) give satisfactory results, with Deng and Clausi's method having the highest accuracy. However, as the noise variance increases above 0.03, the performance of Deng and Clausi's method drops quickly. This is due to the difficulty of estimating the correct feature model parameters in the presence of strong noise. The MLL-SA method seems better in such a range ($\sigma_n^2 \in [0.04, 0.08]$) compared to Deng and Clausi's method, probably due to a better local minimum selected by the initial Gaussian mixture based parameter estimation in this experiment. Both of the methods have convergence failure [4] when even higher level noise exists. It can be clearly found that GIEP and IRGS perform well for a wide range of noise strength.

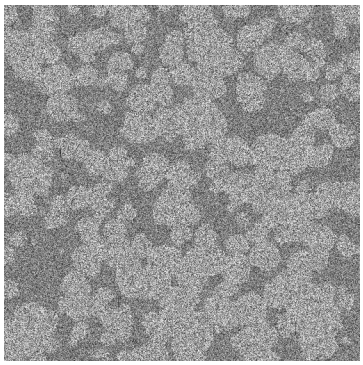
The segmentations on the image with $\sigma_n^2 = 0.08$ are shown in Figure 4.15(b)-(f). At such a noise level, both MLL-SA and Deng&Clausi's method perform poorly due to the inaccurate feature model parameter estimation. In fact, accurate results can be obtained by both methods if the feature model means are fixed with the true values of 128 and 160, as shown in Figure 4.15(g)(h). The GIEP and IRGS results displayed in 4.15(e)(f) have higher accuracy ratios than other methods,

[4]Here the convergence failure means that the method cannot converge to a solution that segments the image into the specified number of classes. That is, all the sites are labelled as one class.

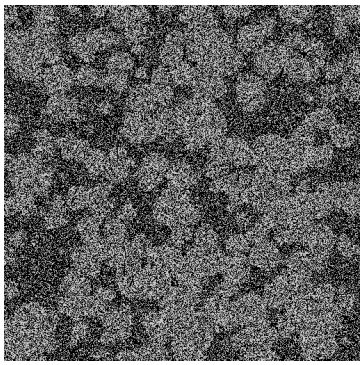
σ_n^2	SNR	Gaussian mixture(p)	MLL-SA(p)	Deng&Clausi's(p)	GIEP(r)	IRGS(r)
0.01	1.9861	75.98%	94.17%	98.65%	98.02%	97.86%
0.02	1.5106	70.82%	92.12%	97.70%	96.32%	95.40%
0.03	1.3522	67.14%	91.30%	95.06%	95.04%	91.00%
0.04	1.2638	64.89%	90.99%	79.44%	93.94%	88.25%
0.05	1.2159	63.77%	91.93%	65.42%	93.06%	88.63%
0.06	1.1809	62.62%	87.66%	66.79%	92.42%	87.34%
0.07	1.1537	61.63%	86.41%	64.86%	91.62%	80.10%
0.08	1.1494	60.76%	78.43%	65.52%	91.35%	80.45%
0.09	1.1283	59.78%	48.72%	63.08%	90.76%	85.50%
0.1	1.1291	58.64%	CF	62.56%	90.28%	81.64%
0.2	1.0912	47.01%	CF	44.15%	87.35%	80.99%
0.3	1.0832	49.44%	CF	CF	84.49%	78.90%
0.4	1.0711	50.76%	CF	CF	81.52%	77.42%
0.6	1.0449	50.96%	CF	CF	75.58%	72.92%
0.8	1.0159	51.09%	CF	CF	70.71%	63.50%
1.0	1.0100	50.82%	CF	CF	68.64%	63.24%

Here p denotes "pixel-based", and r denotes "region-based". CF represents that at least one of the twenty samples has "convergence failure".

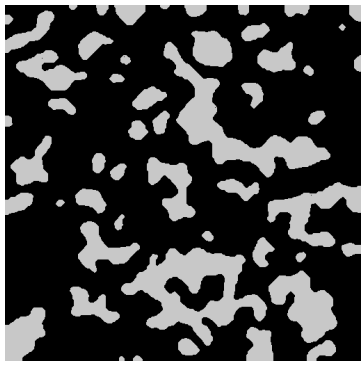
Table 4.2: Summary of segmentation results for the synthetic images. The grey levels in the original image are 128, 160 respectively. The mean diameter of the simulated ice floes is 20 pixels. The image size is 512×512.



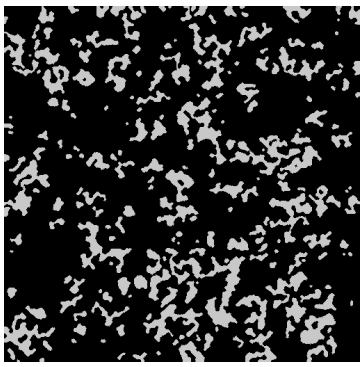
(a) Synthetic noisy image (unit mean multiplicative Gaussian noise, $\sigma_n^2 = 0.08$).



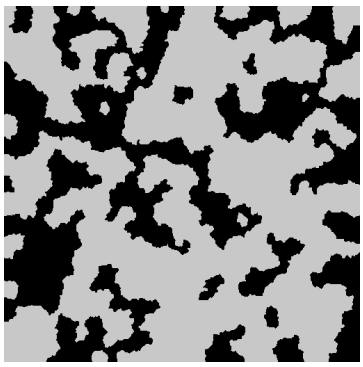
(b) Gaussian mixture. Estimated means: 120.0, 190.1; Accuracy: 59.96%.



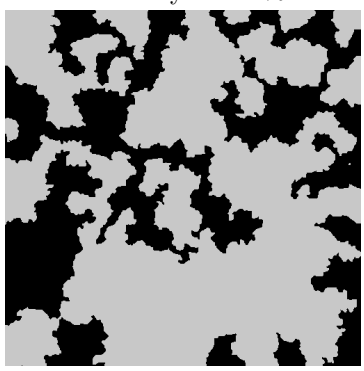
(c) MLL-SA. β_0: 4.558; Estimated means: 140.5, 162.6; Energy: 1.39223e+006; Accuracy: 75.74%.



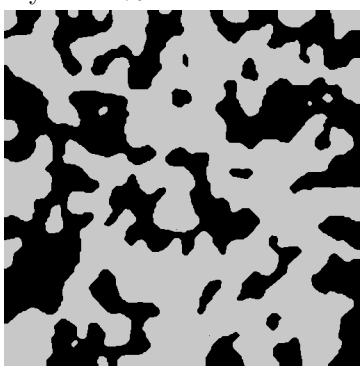
(d) Deng and Clausi's. Estimated means: 140.7, 172.5; Energy: 1.58778e+006; Accuracy: 61.22%.



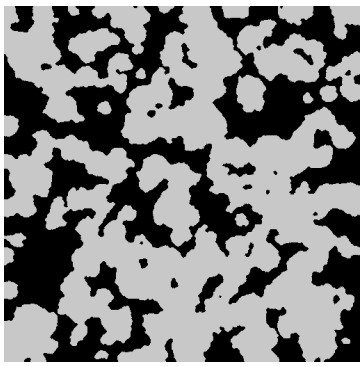
(e) GIEP. Estimated means: 129.5, 158.1; Energy: 1.36575e+006; Accuracy: 91.30%.



(f) IRGS. Estimated means: 130.7, 154.6; Energy: 1.35055e+006; Accuracy: 88.06%.



(g) MLL-SA with true means. Accuracy: 89.81%.



(h) Deng and Clausi's with true means. Accuracy: 95.45%.

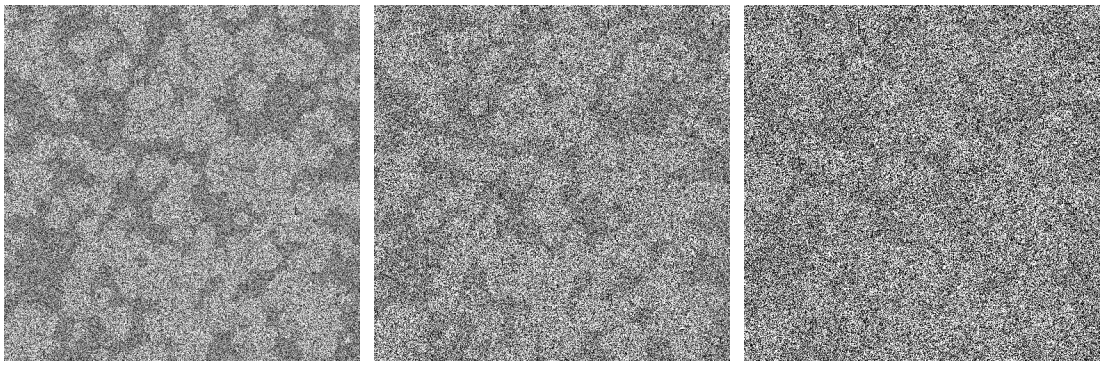
Figure 4.15: Segmentation on Figure 4.14(d)

which implies that the incorporation of edge strength information is quite beneficial to the feature model parameter estimations. GIEP is superior to IRGS in preserving details of the segmentation boundaries. This shortcoming of IRGS is even more serious for images of higher level of noise, as shown in Figure 4.16. In addition to that, the IRGS segmentation results are quite different among Figure 4.16(g)-(i). This is because IRGS is region based and the regions produced by merging can be of large size. For highly noisy and ambiguous images such as Figure 4.16(a)-(c), errors are inevitable and could be amplified by the large regions. From that perspective, GIEP is better since it does not perform merging. In GIEP, there are a sufficient number of regions that have very limited size, and hence the statistical nature of the method allows the error to be controlled in a more limited range.

Compared to GIEP, IRGS is not effective at preserving small details but has obtained a solution with a lower energy, as shown in Figure 4.15(e)(f). Such a phenomenon, as will also be encountered in later experiments, implies that either the Gaussian feature mode or the MLL context model is not appropriate for describing the behaviors of the tested image. For the synthetic images used here, the Gaussian feature model is correct (by the Gaussian nature of the noise), and it is the spatial context model that needs to be investigated.

The MLL context model penalizes the existence of boundary site pairs, and thus favors configurations of large segments with smooth boundaries. Unfortunately, the image in Figure 4.15(a) contains many small segments, and the corresponding boundaries are irregular and have many sharp corners. Therefore, the more energetically favorable solution obtained by IRGS does not correspond a better configuration for the tested image. In the next, segmentations of images having larger floe sizes are investigated.

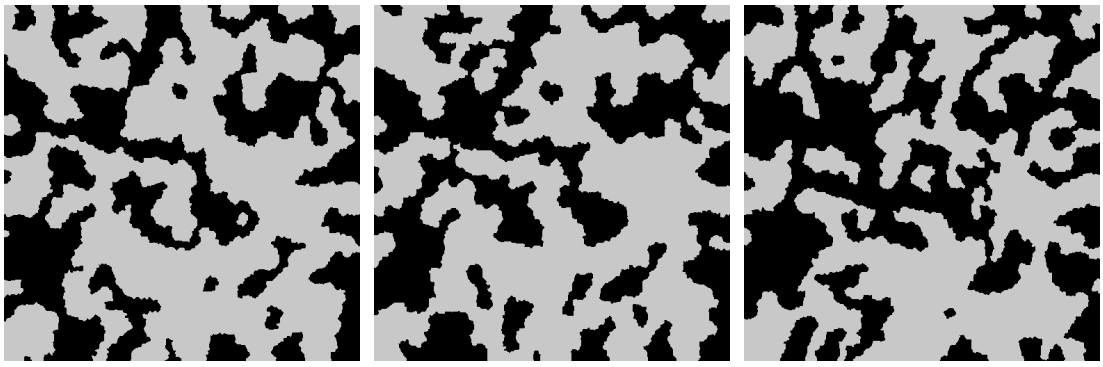
Three more different floe sizes (40, 60, and 80 pixels in diameter) are chosen.



(a) Synthetic image (unit mean multiplicative Gaussian white noise, $\sigma_n^2 = 0.2$).

(b) Synthetic image (unit mean multiplicative Gaussian noise, $\sigma_n^2 = 0.4$).

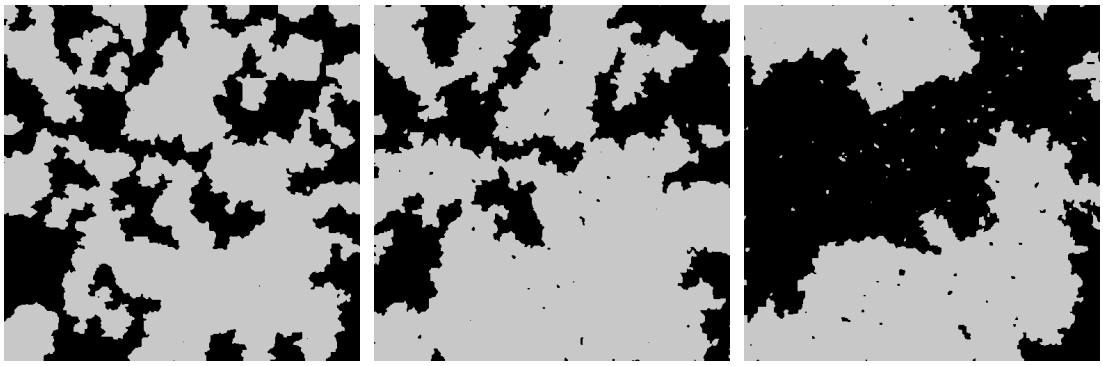
(c) Synthetic image (unit mean multiplicative Gaussian noise, $\sigma_n^2 = 0.8$).



(d) GIEP. Accuracy: 88.02%.

(e) GIEP. Accuracy: 83.52%.

(f) GIEP. Accuracy: 75.70%.

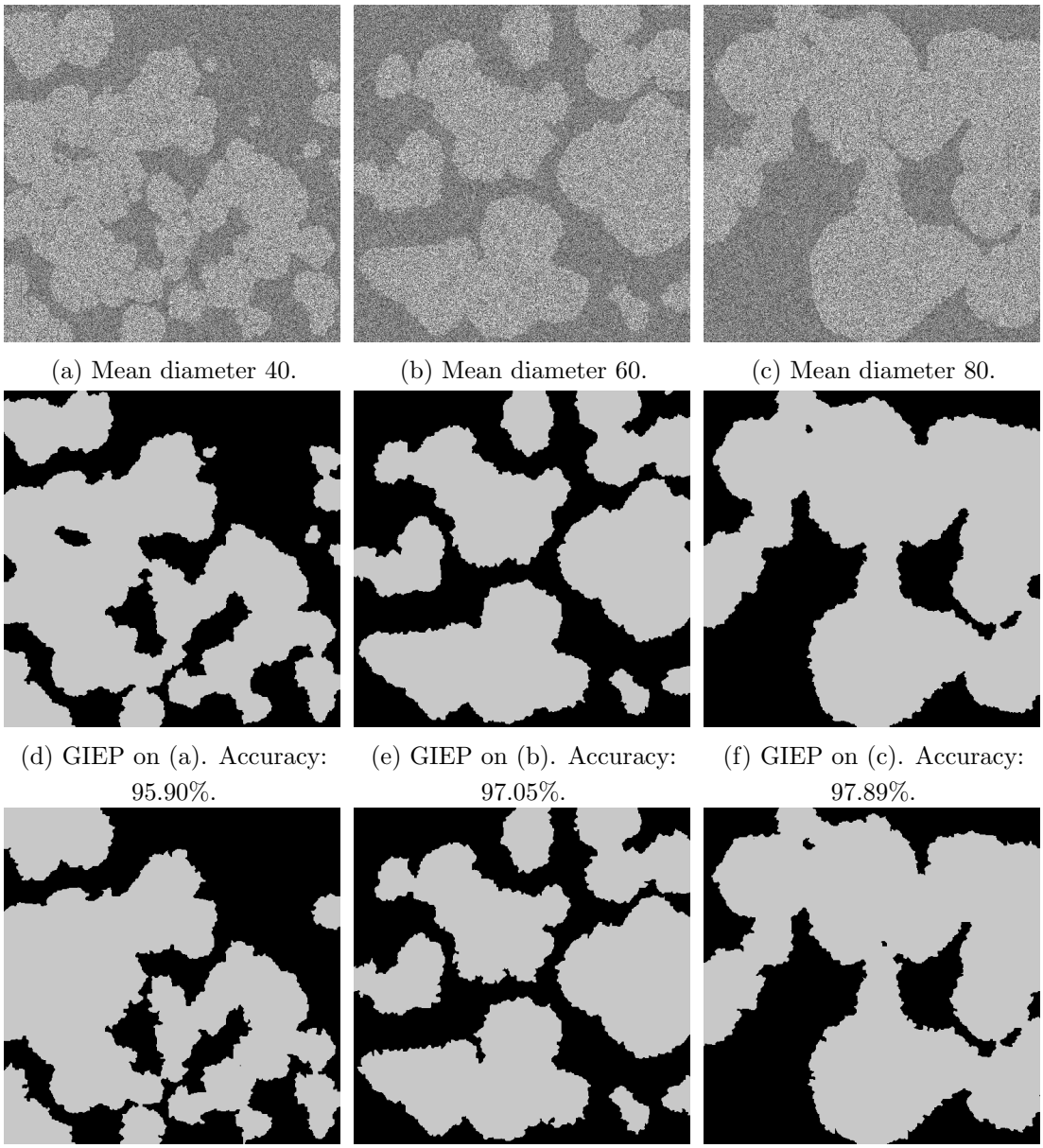


(g) IRGS. Accuracy: 87.60%.

(h) IRGS. Accuracy: 77.80%.

(i) IRGS. Accuracy: 63.50%.

Figure 4.16: Segmentation on synthetic images with different noise strength. The grey levels in the original image are 128, 160 respectively. The mean diameter of the simulated ice floes is 20 pixels. The image size is 512×512.



(a) Mean diameter 40. (b) Mean diameter 60. (c) Mean diameter 80.

(d) GIEP on (a). Accuracy: 95.90%. (e) GIEP on (b). Accuracy: 97.05%. (f) GIEP on (c). Accuracy: 97.89%.

(g) IRGS on (a). Accuracy: 94.21%. (h) IRGS on (b). Accuracy: 97.18%. (i) IRGS on (c). Accuracy: 97.95%.

Figure 4.17: Segmentation on synthetic images with different region size. The grey levels in the original image are 128, 160 respectively. The image size is 512×512. $\sigma_n^2 = 0.08$.

Again, for each of the three sizes 20 scenes are simulated with different levels of noise added. In Figure 4.17, some of the GIEP and IRGS segmentation results on the noise level of $\sigma_n^2 = 0.08$ are presented. As expected, both GIEP and IRGS have higher accuracies approaching to the same value with the growth of region size. Figure 4.18-4.21 gives the segmentation accuracies over the signal to noise ratio (SNR). When noise is relatively low (SNR$>$ 1.4), all methods except the Gaussian mixture model (GMM) method perform well. When the image has heavy noise (SNR$<$ 1.4), the GIEP and IRGS are significantly advantageous over other methods.

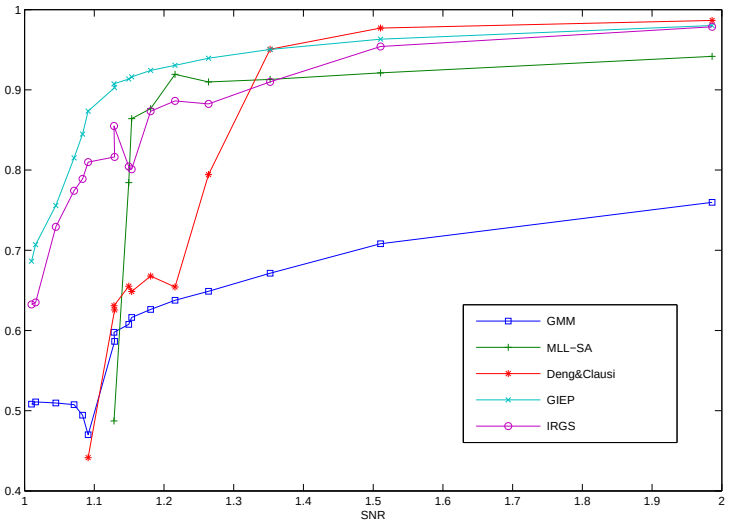


Figure 4.18: Percentage of correctly labelled pixels vs. SNR for images of mean floe diameter 20.

Testing 4: Non-stationary SAR Sea Ice Image

The methods are then applied to SAR sea ice images. Performance evaluations are subjectively based because there is no pixel level ground truth data. All the

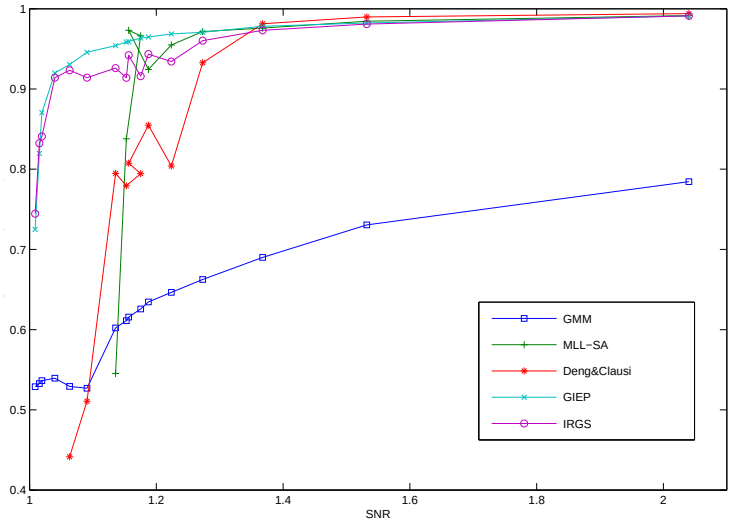


Figure 4.19: Percentage of correctly labelled pixels vs. SNR for images of mean floe diameter 40.

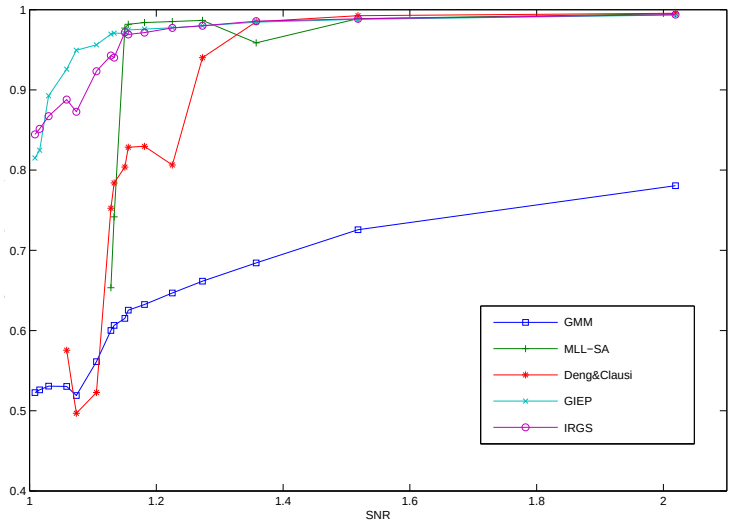


Figure 4.20: Percentage of correctly labelled pixels vs. SNR for images of mean floe diameter 60.

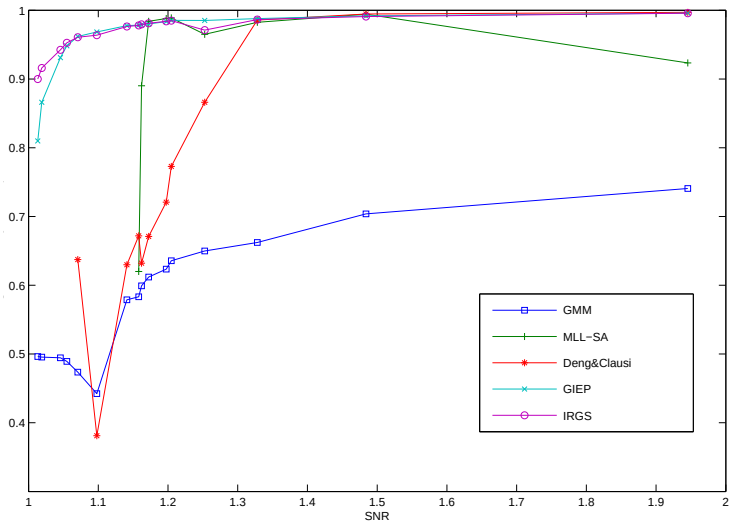


Figure 4.21: Percentage of correctly labelled pixels vs. SNR for images of mean floe diameter 80.

SAR sea ice images are acquired by the RADARSAT ScanSAR C-band mode with the resolution of $100m$. The first SAR sea ice image is captured over Baffin Bay on June 24, 1998, and presented in Figure 4.22. A two class segmentation (ice vs. water) is performed on the image. In the image, the water is dark and the ice is bright. It should be noted that there is a region of rough water towards the top right part of the image that appears a little brighter than the rest of the water, and hence the image is not stationary with respect to the statistics of water. A Gaussian mixture segmentation is shown in Figure 4.23. The result image is "spotty", caused by the heavy speckle noise. With the spatial context model incorporated, the MLL-SA approach has produced a much smoother segmentation in Figure 4.24. However, due to the stationary model assumed, many sites belonging to rough water are incorrectly grouped into the other cluster (see the center right part of the image). Such phenomenons also occur in Deng and Clausi's result in Figure

4.25. Similar to the example in **Testing 1**, both the GIEP and IRGS improves on such non-stationary situations, as shown in Figure 4.26 and 4.27 respectively. The rough water region is much cleaner with very few sites incorrectly labelled. It is interesting to note that the MLL-SA approach gives the lowest energy while that of the IRGS is slightly higher. This is not surprising as the chosen β_0 for the energy computation is only optimal with respect to the MLL-SA approach. Moreover, since some part of the water is relatively bright compared to the rest, one Gaussian function is not well suited for describing such a cluster having two sub-clusters. In fact, it is often unclear whether or not the global minimum of the defined objective function corresponds to the desired configuration for a complex image such as Figure 4.22. Therefore, unlike the simple synthetic examples in **Testing 1** and **2**, the computed energy in complex images may not indicate the accuracy of the obtained configuration. An objective function with sub-optimal solutions that are consistent with the intuition is often acceptable, and this example is such a case.

Testing 5: Noisy SAR Sea Ice Image with Indistinct Ice Types

In this experiment, a SAR sea ice image of the Gulf of St. Lawrence captured on Feb 19, 1997 is used. The image is shown in Figure 4.29 consisting of two different ice types besides the water. The left and right part of the image is relatively dark corresponding to grey ice, and some brighter regions in the center are grey-white ice. Figure 4.30-4.35 show the results by various methods. Similar to the example in **Testing 2**, due to the large overlapping of the intensity range between the grey and grey-white ice, all methods except IRGS perform quite poor. Although a subjective evaluation is difficult for such an image because of the subtle intensity difference between the two ice types and the heavy noise, it is still obvious that IRGS is advantageous in producing large homogeneous regions consistent with large

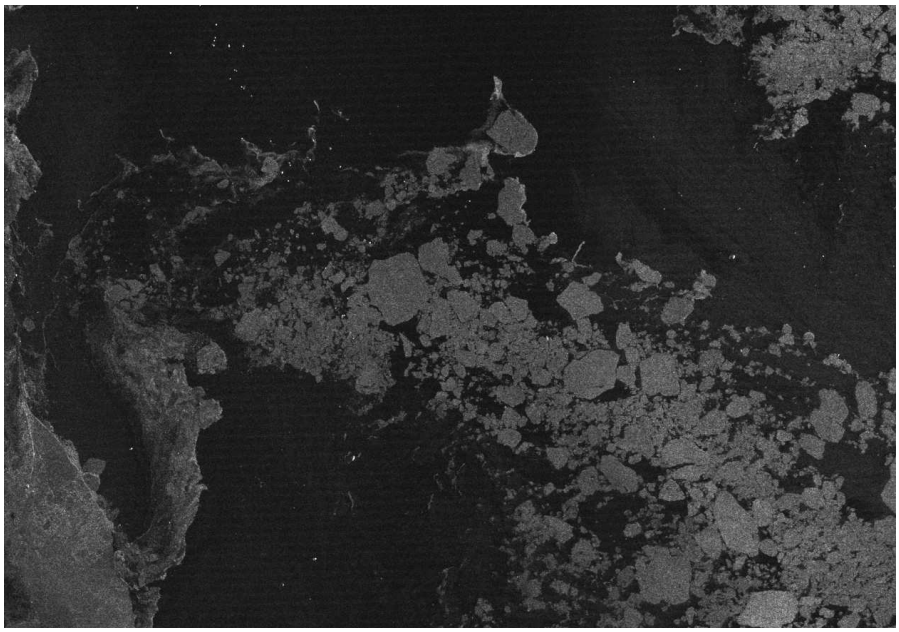


Figure 4.22: A SAR image (1252×873) captured over Baffin Bay on June 24, 1998.

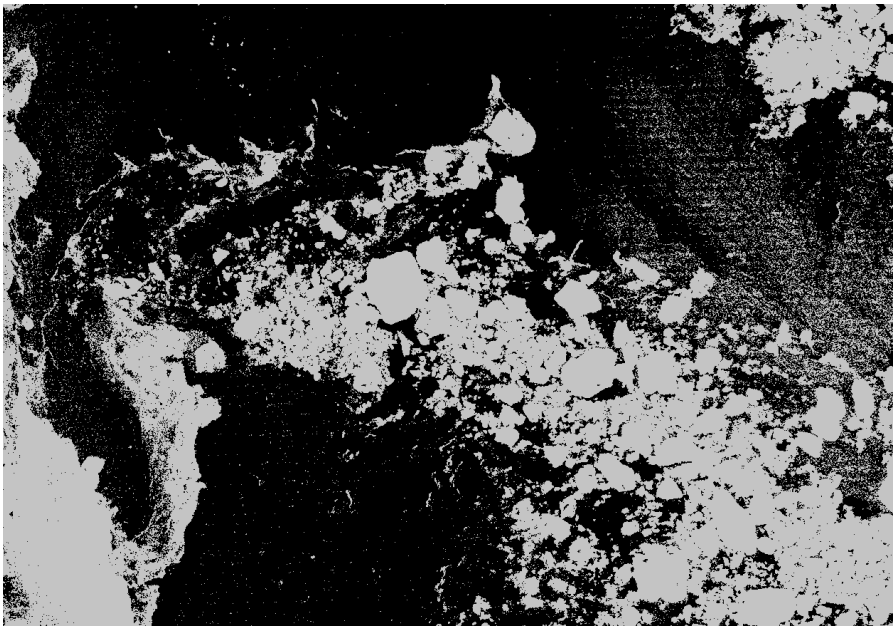


Figure 4.23: Gaussian mixture segmentation on Fig. 4.22.

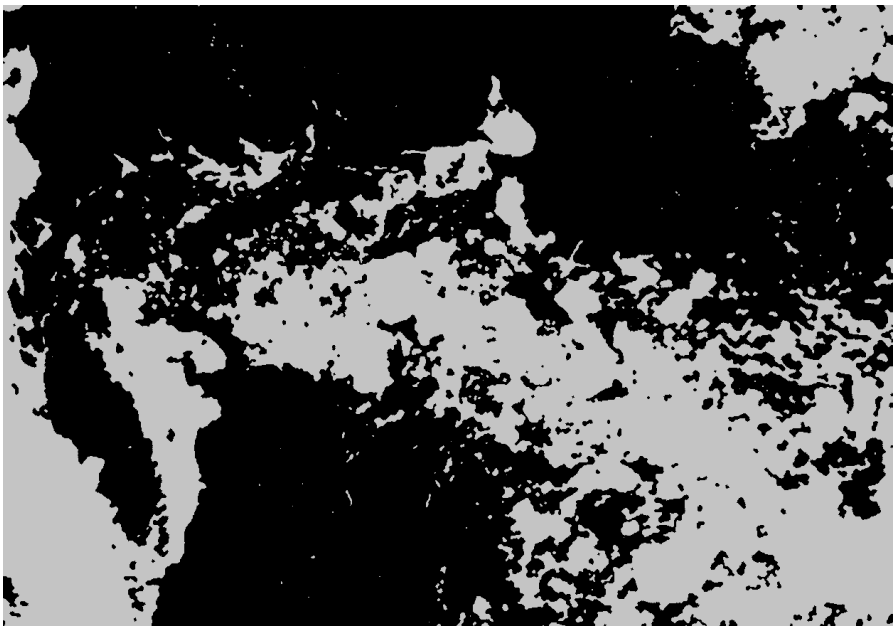


Figure 4.24: MLL-SA segmentation on Fig. 4.22. β_0: 4.18785; Energy: 4.26153e+006.

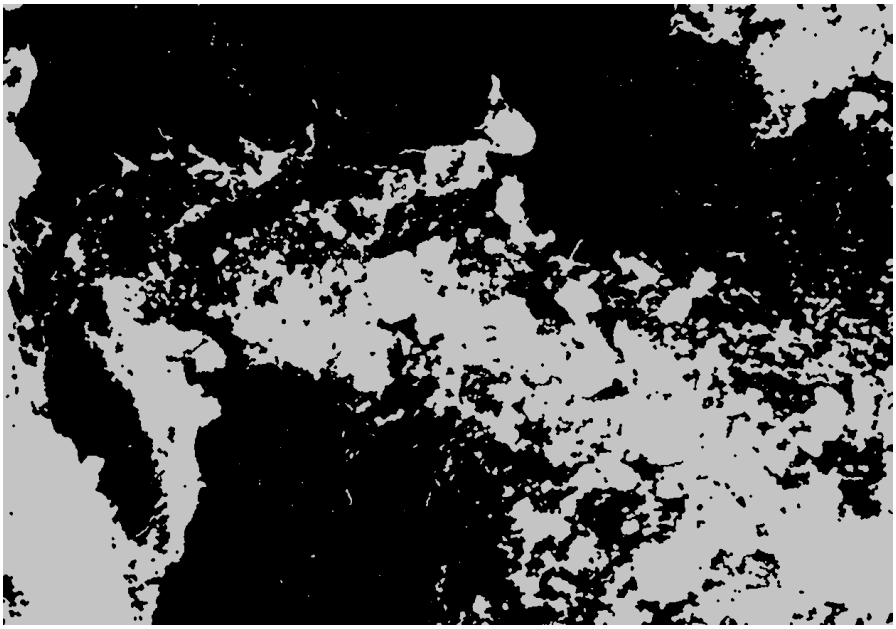


Figure 4.25: Segmentation on Fig. 4.22 using Deng and Clausi's method. Energy: 4.34642e+006.

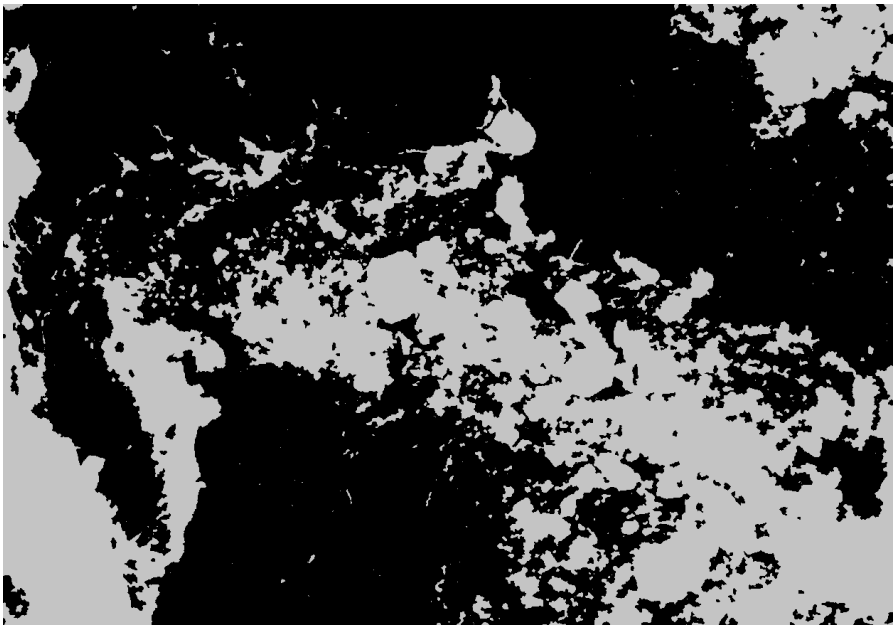


Figure 4.26: GIEP segmentation on Fig. 4.22. Energy: 4.3394e+006.

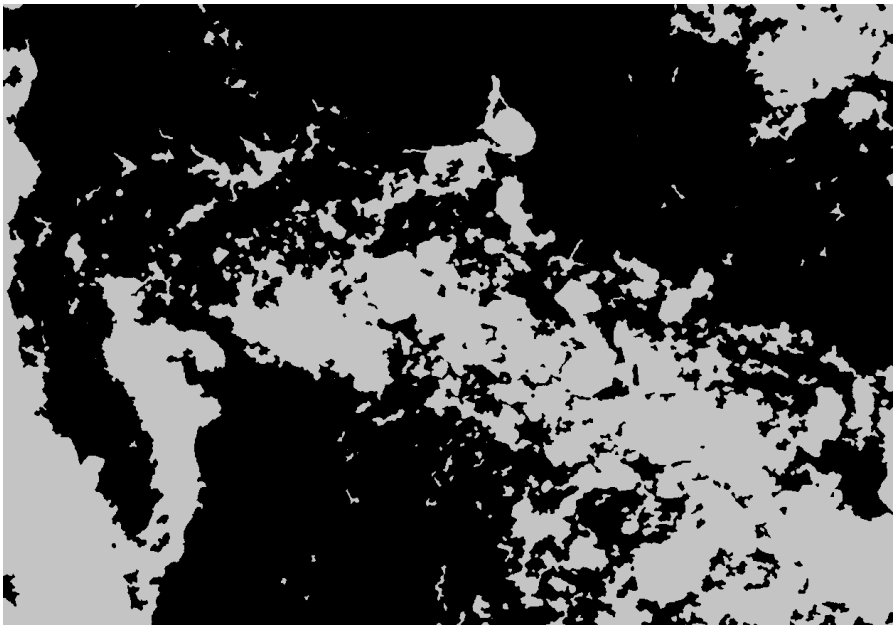


Figure 4.27: IRGS segmentation on Fig. 4.22. Energy: 4.26885e+006.

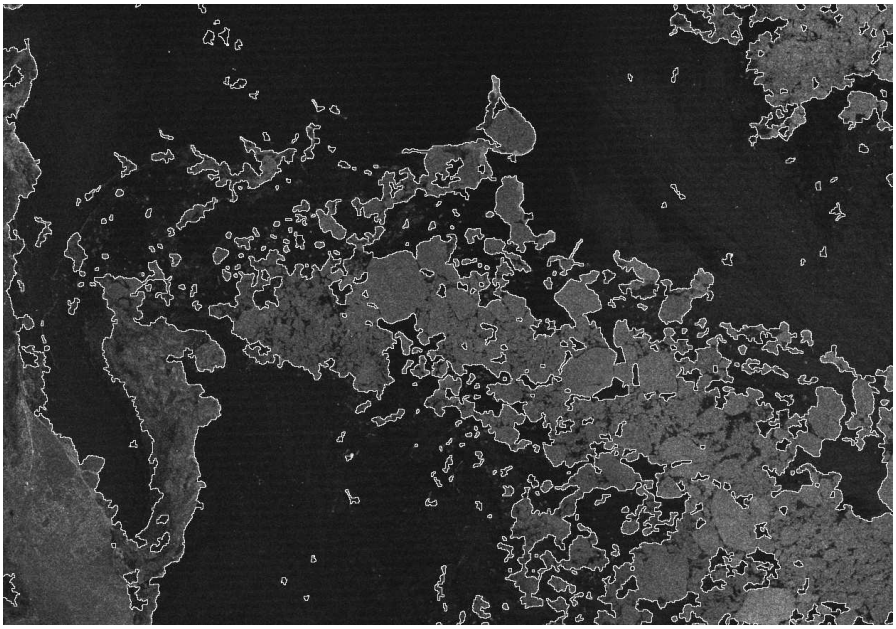


Figure 4.28: IRGS segmentation on Fig. 4.22 with boundary overlayed.

ice floes, as shown in Figure 4.34. The GIEP method lags behind the IRGS in generating uniform regions, but seems to be better than other methods as shown in Figure 4.33, probably due to its region-based nature and its advantage in the feature model estimation under heavy noise condition, as already discussed in **Testing 3**.

The estimated SNR of this image, excluding water regions, is 2.2653. Referring back to Figure 4.18-4.21, such a SNR is in the range where GIEP, ISRG, and Deng&Clausi's methods should perform similarly. This implies the existence of large differences between the synthetic images and the true SAR sea ice imagery in noise, texture, or some intrinsic image structures.

Testing 6: SAR Sea Ice Image with High Intra-class Variations

Figure 4.36 is another image of the Gulf of St. Lawrence acquired on Feb 20, 1998. In the image, the center of the bottom is water and land. Regions relatively dark

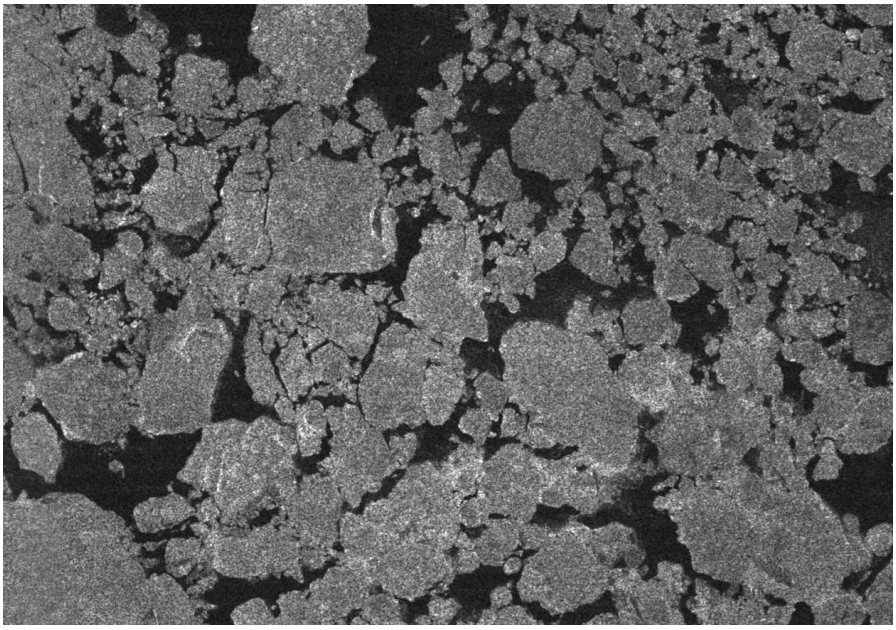


Figure 4.29: A SAR image (1252×873) captured over the Gulf of St. Lawrence on Feb 19, 1997.

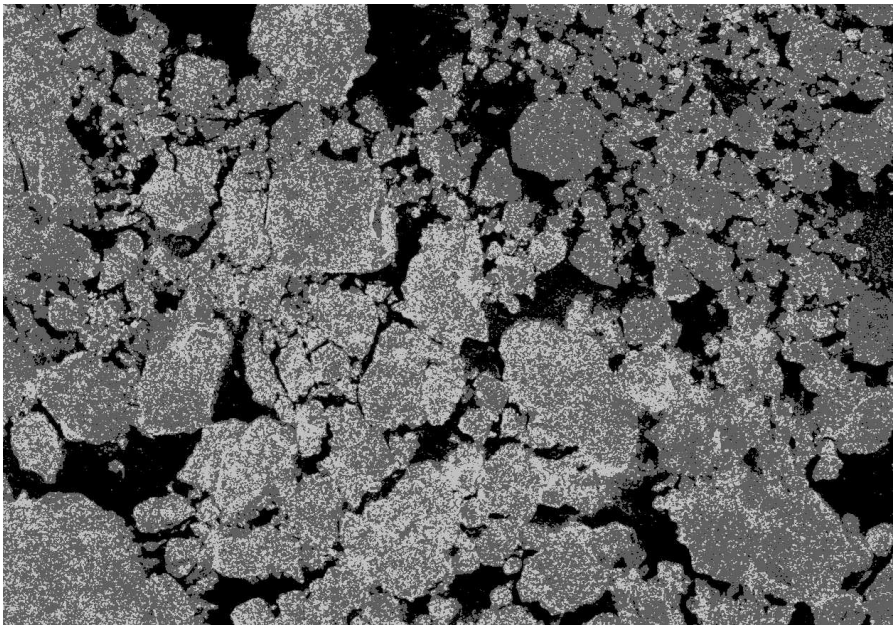


Figure 4.30: Gaussian mixture segmentation on Fig. 4.29.

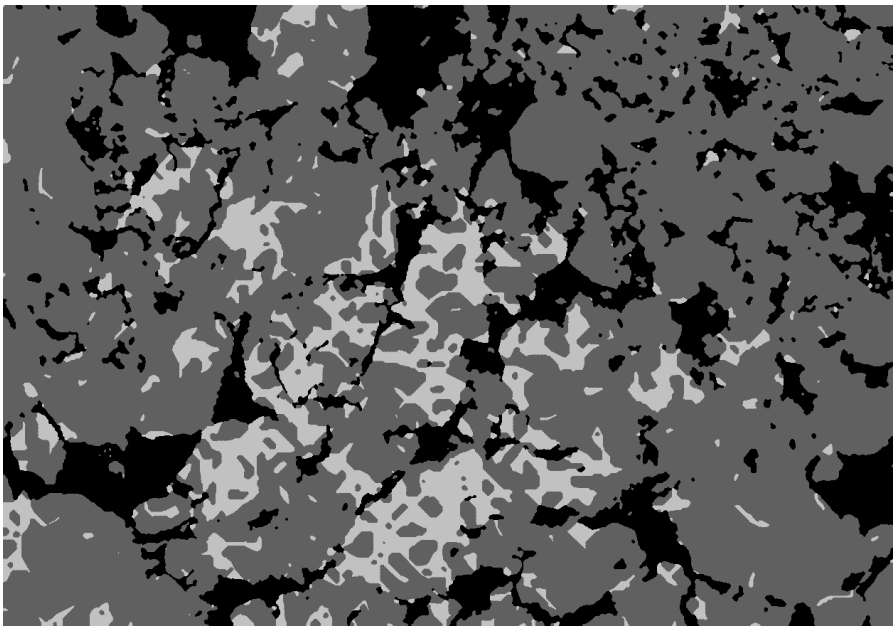


Figure 4.31: MLL-SA segmentation on Fig. 4.29. β_0: 3.60617; Energy: 5.78134e+006.

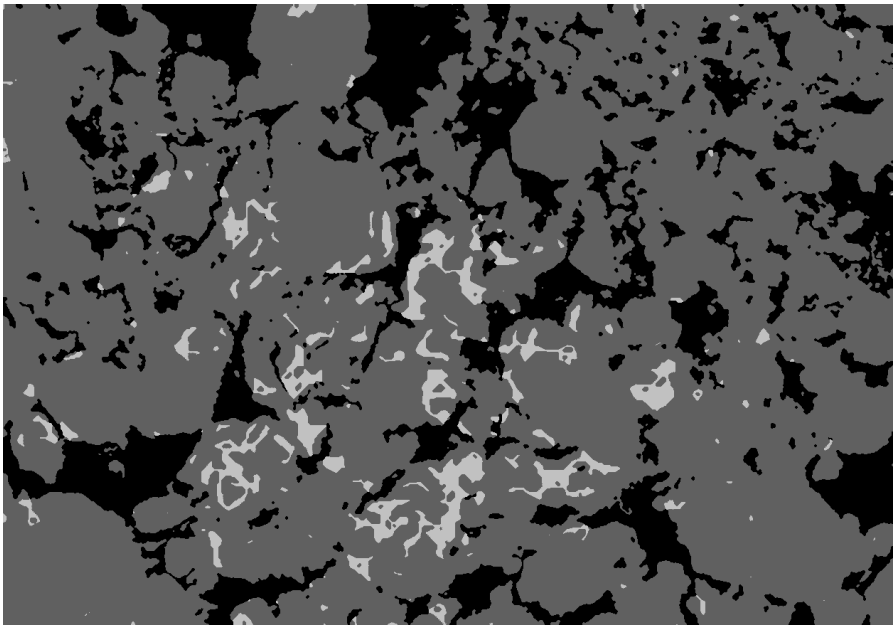


Figure 4.32: Segmentation on Fig. 4.29 using Deng and Clausi's method. Energy: 5.59624e+006.

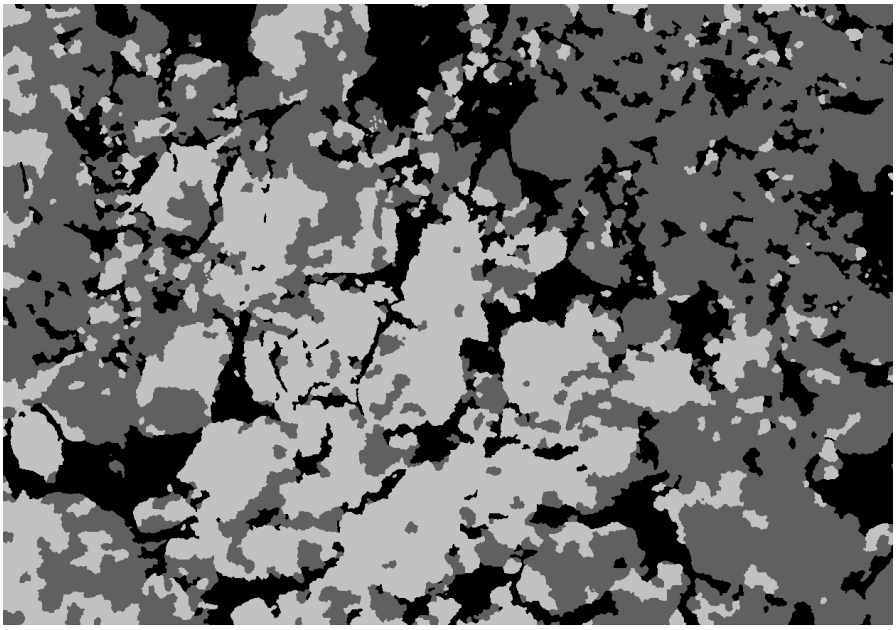


Figure 4.33: GIEP segmentation on Fig. 4.29. Energy: 5.59507e+006.

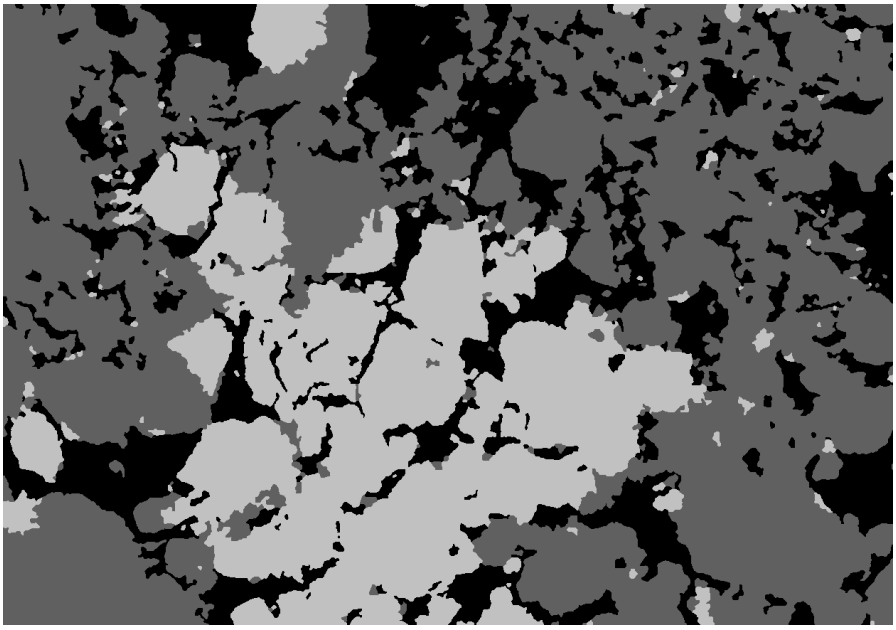


Figure 4.34: IRGS segmentation on Fig. 4.29. Energy: 5.54939e+006.

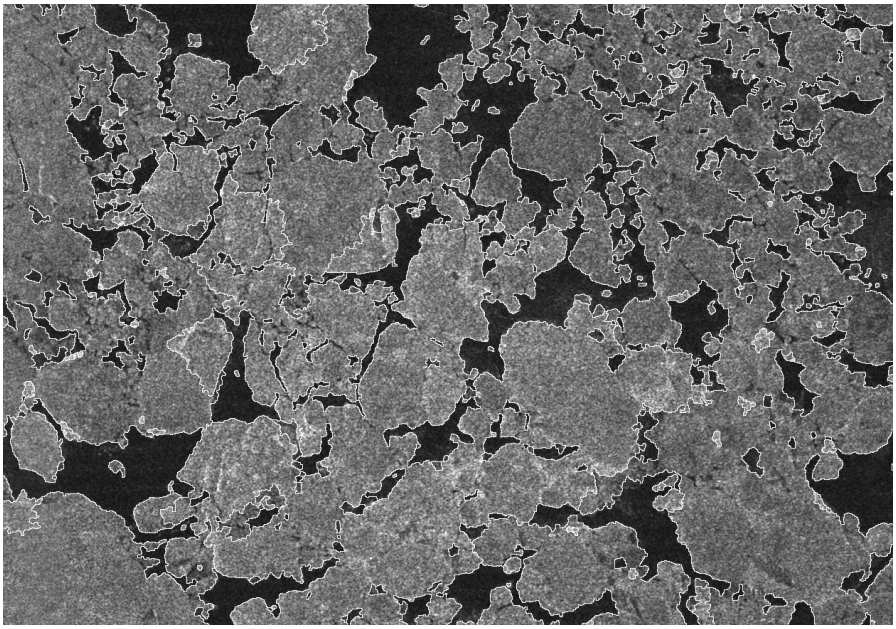


Figure 4.35: IRGS segmentation on Fig. 4.29 with boundary overlayed.

with brighter ridges inside are grey-white ice, and the rest are grey ice. The image is difficult to segment because of the large intra-class variations of the grey-white ice. Again, over-segmentation is obvious for the grey-white ice in the results by MLL-SA approach as shown in Figure 4.38 and by Deng and Clausi's method as shown in Figure 4.39. In Figure 4.40, the GIEP improves for the over-segmentation phenomenon by using large moves in search space. However, the moving paces are still limited by the regions obtained from the initial watershed segmentation.

The IRGS gives larger homogeneous regions in Figure 4.41. On the other hand, similar to the results in **Testing 3**, it can also be found that the IRGS result is likely to be an under-segmentation. Regions corresponding to leads (the dark long cracks) in the grey ice have disappeared, which is undesirable from an operational perspective. Some brighter regions in the right of the image, which could possibly be grey ice, are also merged with surrounding regions. This is because the MLL model

tends to generate blob-like configurations, and is not well suited for modelling the prior for scenes where the shape of objects are far from blob-like. Since the IRGS can be viewed as a method trying to find the global minimum of the objective function with MLL as the prior, it is possible that the minimum found is not a desired configuration due to the inappropriateness of the prior model. In the image, both the leads and the bright regions in the right do not have blob-like shapes, and thus the scene is not a favored configuration by the MLL prior model. Possible improvement can be achieved by using a domain specific model as the prior instead of the generic MLL, which will be presented in chapter 5.

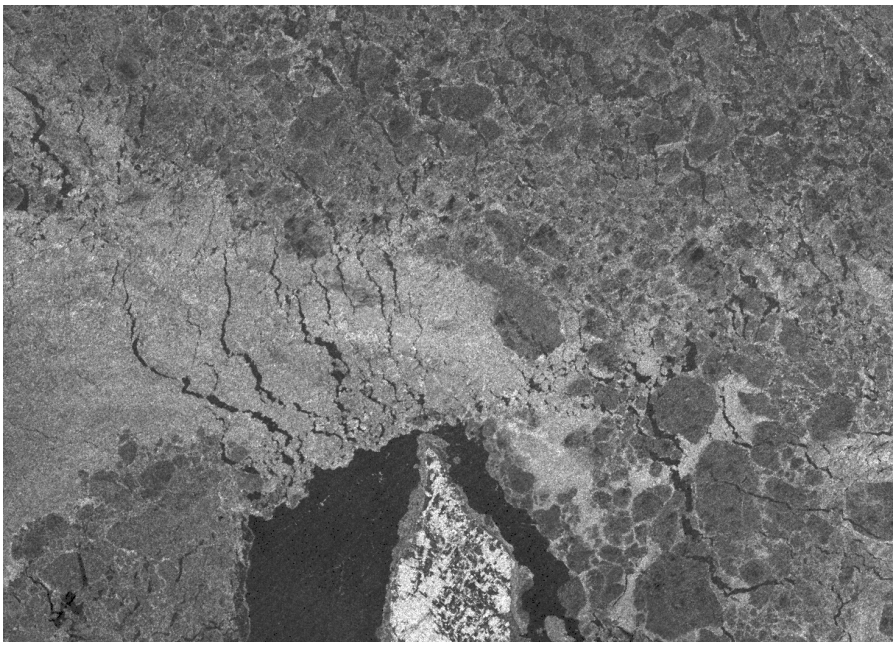


Figure 4.36: A SAR image (1252×873) captured over the gulf of St. Lawrence on Feb 20, 1998.

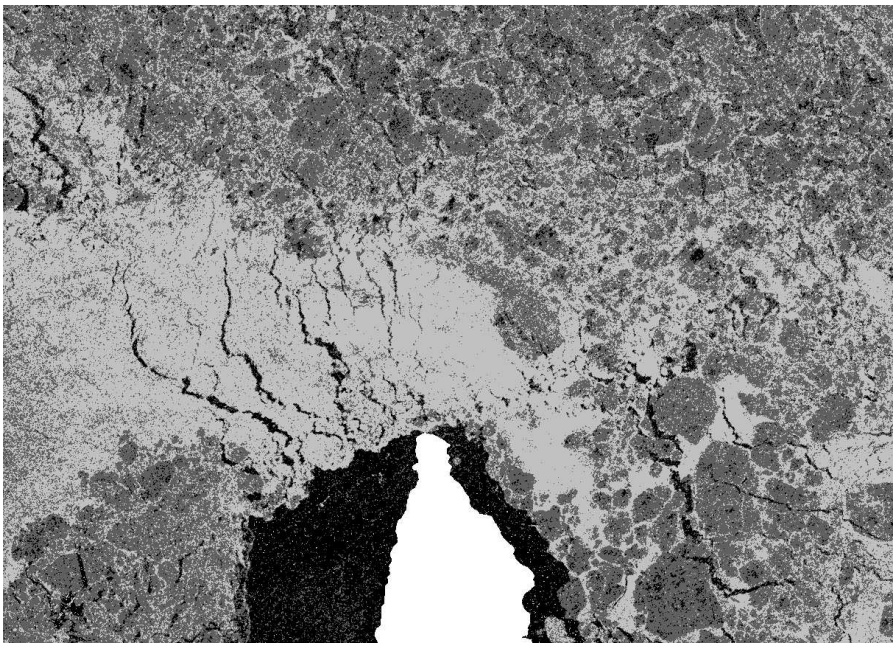


Figure 4.37: Gaussian mixture segmentation on Fig. 4.36.

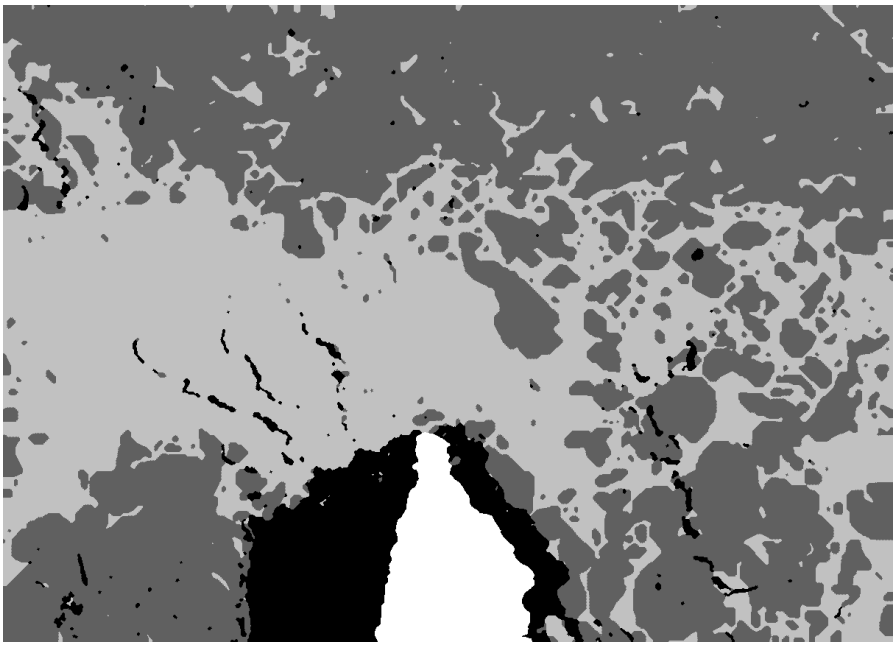


Figure 4.38: MLL-SA segmentation on Fig. 4.36. β_0: 4.1526; Energy: 5.23105e+006.

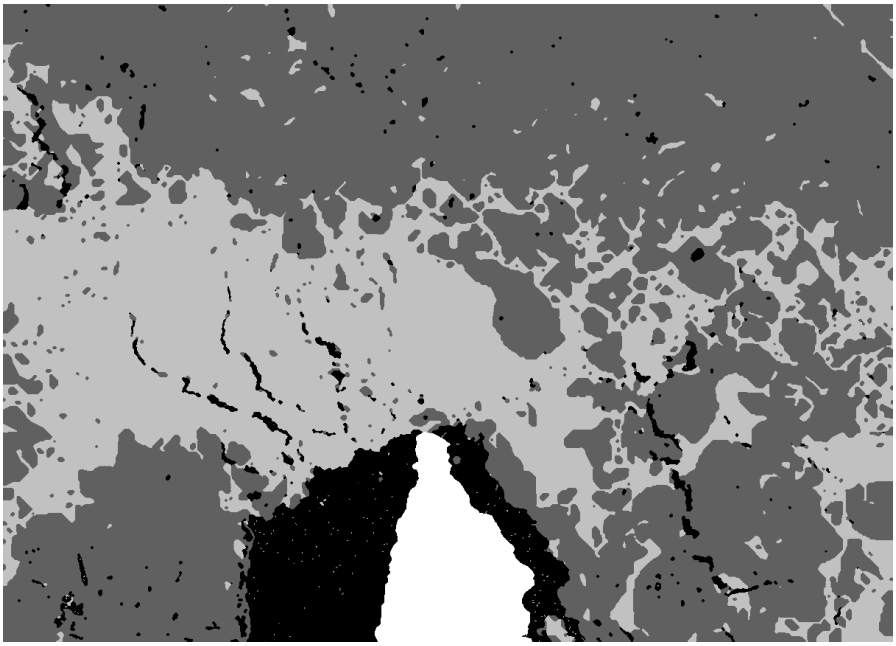


Figure 4.39: Segmentation on Fig. 4.36 using Deng and Clausi's method. Energy: 5.29724e+006.

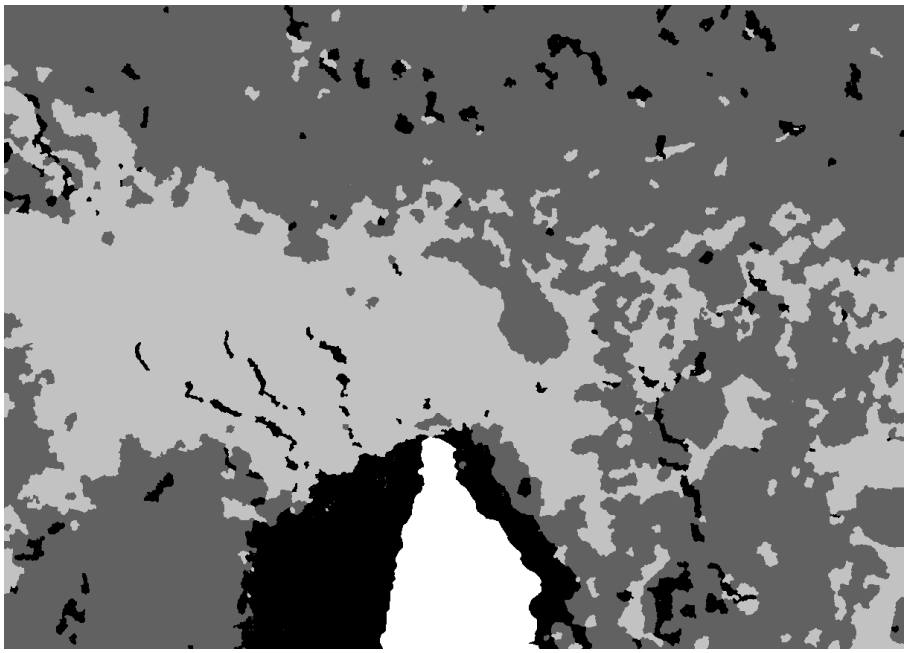


Figure 4.40: GIEP segmentation on Fig. 4.36. Energy: 4.93387e+006.

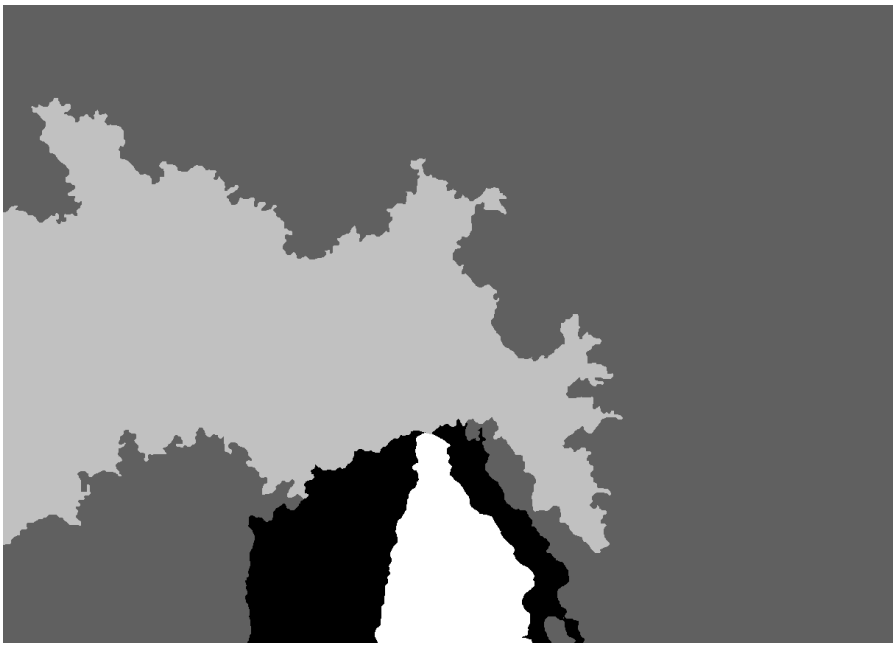


Figure 4.41: IRGS segmentation on Fig. 4.36. Energy: 4.57517e+006.

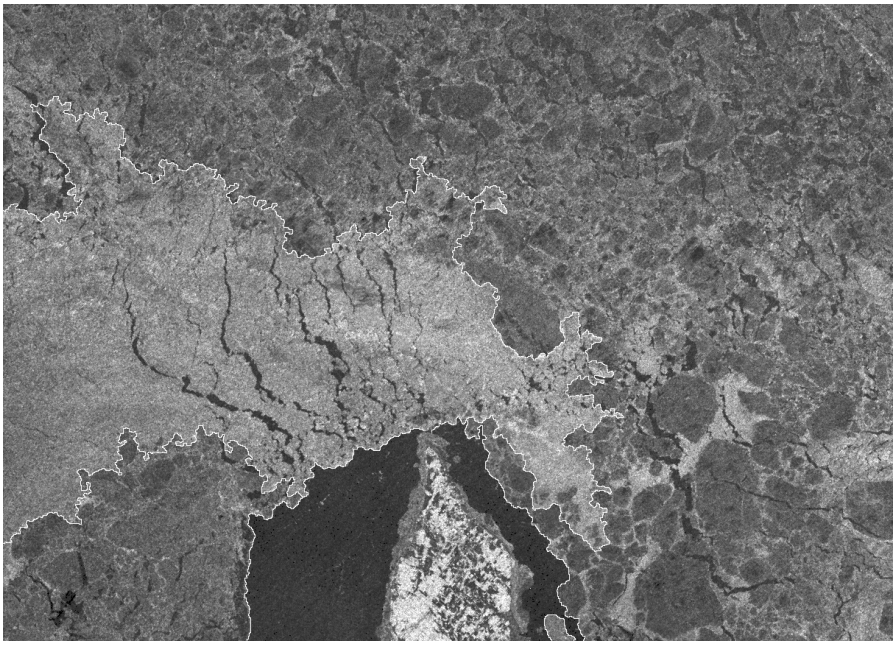


Figure 4.42: IRGS segmentation on Fig. 4.36 with boundary overlayed.

Summary of Results

For most cases, the IRGS generates results with the lowest energy, which correspond to configurations consisting of large homogeneous regions. Both IRGS and GIEP methods can improve on non-stationary and over-segmentation situations. The power of IRGS and GIEP on alleviating the over-segmentation problem is due to their region-based implementation and their incorporation of edge information. Although IRGS is region-based in nature, there is no reason why the GIEP cannot be implemented as pixel-based. Pixel-based GIEP provides an alternative to the traditional optimization techniques, such as the iterated conditional mode and simulated annealing, for the MLL-based image segmentations.

4.3.2 Applied to Texture Features

In this section, GLCM texture features are used for testing the efficiency of the IRGS approach. The GLCM statistics are currently the preferred texture features for SAR sea ice identification [6][99][23]. The GLCM requires many parameters (δ,W,θ,G) to be set. There is no known method to automatically select these parameters, and typically the parameters are determined based on supervised trials. The parameters should be dependent upon texture specific characteristics. For example, if a texture is coarse (i.e. formed by primitives that are large in scales) the sampling distance δ and the window size W should also be relatively large. For the experiments in this section, two kind of textures are used. The Brodatz textures [14] are generally coarse while SAR sea ice textures can have both large scale characteristics (macro texture) and fine details (micro texture). Therefore, the window size and sampling distance are different among the experiments depending on the image used, while the other two parameters are the same for both textures: the quantization level

G is 32, and the average of four orientations (0^o,45^o,90^o,135^o) are adopted for the extraction of rotation invariance features.

Testing 7: Brodatz Texture Mosiac

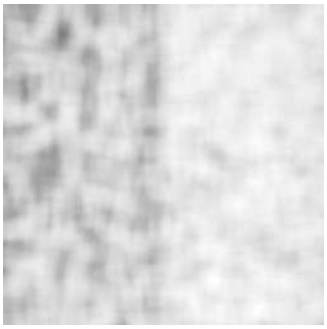
The image used is a mosaic of two Brodatz textures [14] - bark (D12) and straw (D15). It is of 256×256 and is shown in Figure 4.43(a). With a 17×17 window selected and the sampling distance set to be 2, the GLCM ENT statistic (defined in Table 3.1), is computed on the image and the Gaussian mixture clustering is applied to the obtained feature. Figure 4.43(b) gives the ENT feature normalized to [0 1], and Figure 4.43(c) is the Gaussian mixture clustering result. It can be found that small patches of errors exist in the segmentation result. This is expected to be solvable by either increasing the window size W or using a spatial context model. Increasing the window size can make the ENT feature image more smooth, but at the same time will weaken the discontinuities at the boundaries and hence cause inaccurate boundary locations. An MLL context model is incorporated as before for the tone feature, and the result is shown in Figure 4.43(d). Significant reduction of over-segmentation phenomenon has been achieved by the MLL-SA approach. In Figure 4.43(e), the IRGS gives the best result. Again, the IRGS has exhibited its advantage over traditional MLL-SA approaches in improving on the over-segmentation situations. The boundary location obtained by IRGS is not precise. Due to the implicit smoothing in computing the texture features, the ENT image in Figure 4.43(b) does not have a distinct boundary but rather a vague and broad boundary region. One might think of using the original image instead of the ENT image for computing the edge penalty term in the model. However, for textures having similar intensities, edge strength inside the homogeneous texture regions are often much higher than that on the boundaries. Figure 4.43(g) shows

the normalized gradient computed on the original image, and Figure 4.43(i) is the mean gradient over columns computed respectively on the original image and the ENT one. The figures show that the gradient magnitude of both original image and ENT image at the boundary is smaller than some of those inside the texture region, and it is unpredictably whether including such edge information is beneficial or not to the segmentation process. The IRGS result using the edge strength computed on the original image is shown in Figure 4.43(h), and the classification accuracy is only marginally lower than that in Figure 4.43(e).

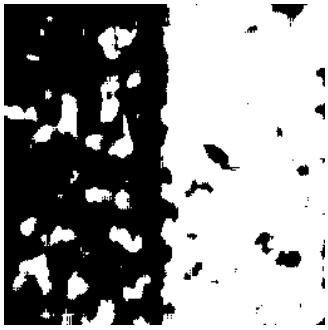
Testing 8: SAR Sea Ice Texture

In this experiment, the IRGS is applied to the GLCM texture features computed on the SAR sea ice image of Figure 4.29. There has already been considerable research investigating GLCM parameters and features for the SAR sea ice texture classification [6][21][97][99], and the experiment here does not intend to perform a thorough and comprehensive study on selecting the optimal parameters and statistics. Here W is simply chosen to be 9. If using larger windows such as 17×17, failure of segmentation has been noted in the experiments (not presented), where the two clusters of grey and grey-white ice are merged as the same class and boundary regions are erroneously identified as a separate class. The sampling distance δ is set as 1, and two statistics of ENT and DIS (defined in Table 3.1) are used. Figure 4.44 shows the Gaussian mixture segmentation on the two GLCM texture features. Similar as the case using tone feature in **Testing 5**, the result is very noisy. The MLL-SA approach produces an improved but still highly over-segmented result in Figure 4.45. Such over-segmentation phenomenon is greatly reduced by the IRGS in Figure 4.47, where the edge penalty term is computed on the original image. Comparing this result with Figure 4.34, it is difficult to say which is better since no

(a) The texture mosaic image (256×256).



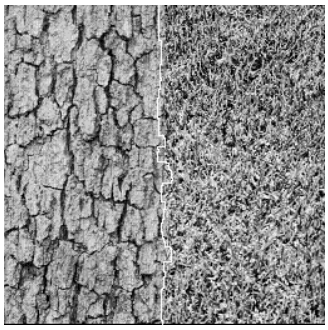
(b) The ENT feature ($W = 17, \delta = 2.$)



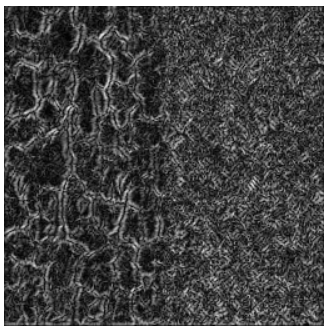
(c) Gaussian mixture segmentation on (b). Accuracy: 89.60%

(d) The MLL-SA segmentation on (b). $\beta = 18$; Accuracy: 96.63%.

(e) The IRGS segmentation on (b). Accuracy: 98.90%.

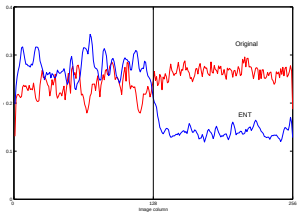


(f) The IRGS segmentation on (b) with boundary overlayed.



(g) The normalized gradient computed on (a).

(h) The IRGS segmentation on (b) with edge penalty computed from (a). Accuracy: 98.75%.



(i) The averaged gradient over columns computed on (a) and (b) respectively.

Figure 4.43: Segmentation on a texture mosaic image.

ground truth data is available. However, it can be concluded IRGS is advantageous for improving on over-segmentation situations. An experiment that use the DIS for computing the edge strength has also been conducted in Figure 4.46. Unlike Figure 4.47, the IRGS is not successful this time in producing large homogeneous regions. A possible reason is the implicit smoothing in the computation of texture features, which has reduced significantly not only the edge strength at the boundary locations but also the variance of features inside the homogeneous regions, making the merging criterion stronger.

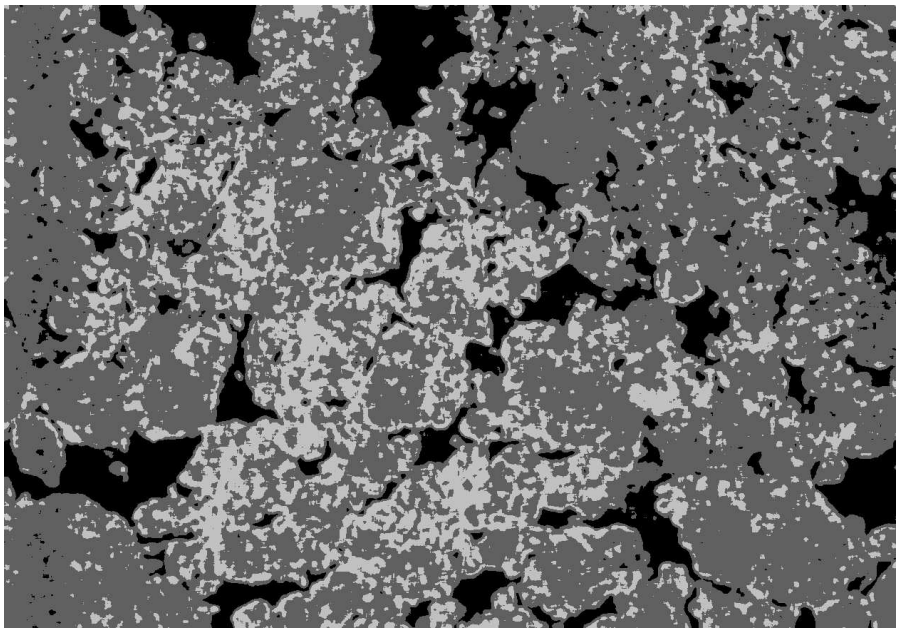


Figure 4.44: The Gaussian mixture segmentation on the GLCM features (ENT and DIS) for Figure 4.29.

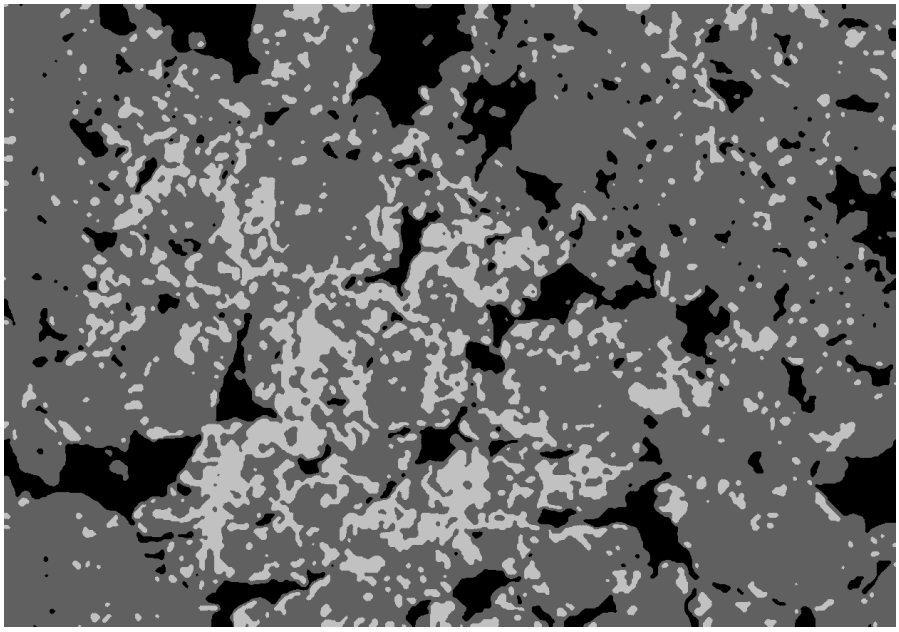


Figure 4.45: The MLL-SA segmentation on the GLCM features (ENT and DIS) for Figure 4.29.

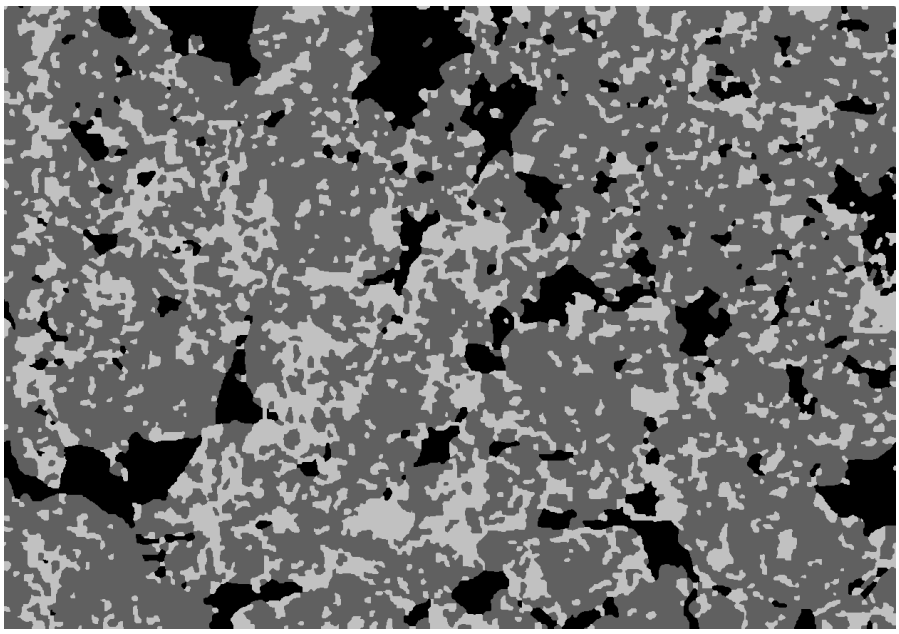


Figure 4.46: The IRGS segmentation on the GLCM DIS feature for Figure 4.29. The edge penalty is computed on the DIS feature.

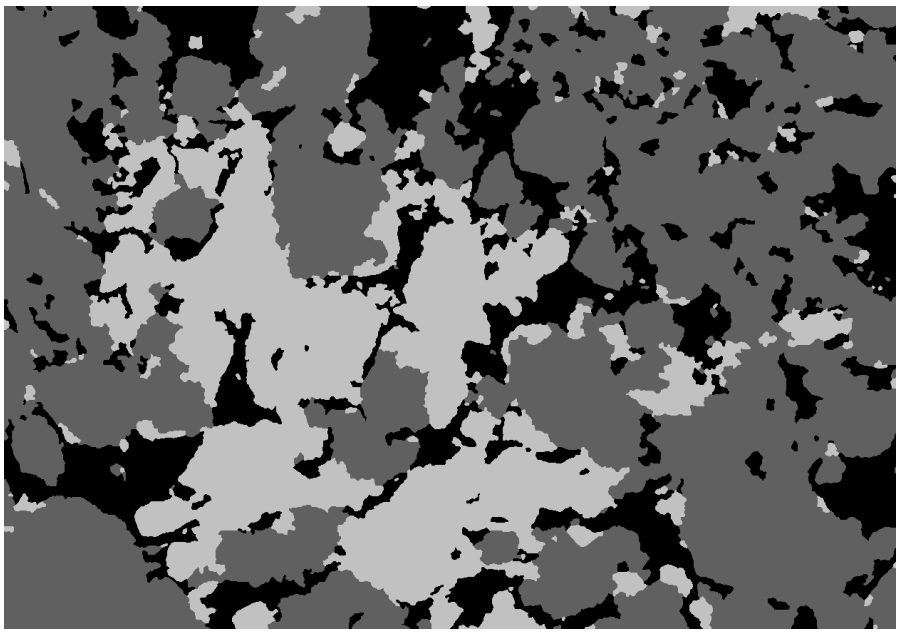


Figure 4.47: The IRGS segmentation on the GLCM features (ENT and DIS) for Figure 4.29. The edge penalty is computed on Figure 4.29.

4.4 Chapter Summary

In this chapter, a new unsupervised segmentation approach has been proposed. The method is a region merging method with the merging criteria being an energy function constructed in a statistical optimal way under the Bayesian framework. The method is advantageous in locating the boundaries accurately and producing large homogeneous regions. However, these two advantages are not so obvious for textures as for tone features. The proposed method has also provided a framework for incorporating domain specific knowledge based classification, which is expected to be beneficial to both the low level segmentation and the high level region labelling process.

Chapter 5

High Level Classification and Overall System

The low level segmentation transforms the raw image data into the segmented regions, allowing a much reduced but sufficient description of the intrinsic properties (e.g. mean grey level, shape, size, etc) and relationships (e.g. proximity, connectivity, etc) of the objects that the regions correspond to. Further object recognitions can then be performed with such properties and relationships by a high level classification process, whose goal is to attach semantic symbols to the segmented regions. By combining the low level segmentation and the high level classification, an image understanding system (IUS) is thus constructed.

The IUS implemented in this chapter is based on the IRGS system proposed in chapter 4. In chapter 4, the IRGS method has used a domain independent labelling (clustering) to influence the region growing for the low level segmentation. There is no reason why the general purpose labelling cannot be replaced by a domain specific classification. The domain specific classification can still be MRF based, but the

clique energy functions are a modified version on (4.14) and (4.15) to incorporate domain knowledge.

The resulting IUS combines both bottom-up and top-down control strategies in the system. From that perspective, the proposed method is advantageous relative to pure bottom-up systems such as ARKTOS [101], since in a bottom-up system regions obtained by the low level segmentation may not sufficiently match the real objects for an accurate subsequent high level classification. In fact, segmentation is generally not a stand-alone problem but ill-posed if not associated with any implicit or explicit interpretation. The high level classification process in the IRGS system has some degree of control on the low level segmentation, and hence a bidirectional relationship between the two processes has been achieved.

This concept of controlling region growing by interpretation result is known as semantic region growing [37][102]. A number of other works have also demonstrated success to various degrees in incorporating region (object) information into the segmentation process [54][59][104]. However, their high level classification results directly generate the region growing, which hence can be viewed as part of the high level classification process. From that perspective, this work is substantially different in that the high level classification and the low level segmentation are still two distinct processes and are iteratively refined in a cooperative manner. A hierarchical organization of the segmented regions can be formed during the iterations, which makes it possible to establish a reasonable match between the segmented regions and the objects. Such a match is crucial in incorporating object features such as shape.

Compared to the knowledge-based image understanding systems, the MRF based high level classification system is simpler and easier to build. Moreover, the formulation of the MRF based classification can be combined with that of low

level segmentation to give a well formulated MAP problem under the Bayesian framework.

The MRF based high level classification method applied here is based on Modestino and Zhang's work [78], to which a brief introduction is given in section 5.1. Section 5.2 describes the implementation for SAR sea ice. Experiments and discussions are presented in section 5.3. Section 5.4 is the chapter summary.

5.1 MRF Based Classification

In section 4.2.3, a MRF based clustering method has been designed on the RAG structure, with the related single and pair node clique energies formulated by (4.14) and (4.15) respectively. Similarly, the high level domain specific classification can be designed here using the clique energy functions to encode domain knowledge.

The design rule of the clique energy function is simple. If the interpretation of the regions in a clique tends to be consistent with the measurement and the domain knowledge, the energy function should decrease, resulting in a relatively low value, and vice versa. Modestino and Zhang defines the clique functions for single node cliques as follow [78].

Let c be an arbitrary single node clique with node i representing the segmented region Ω_i. Let the corresponding clique function be denoted by $V_c(x_i : \mathcal{K}, \mathbf{y}_i)$, which depends on the configuration of labelling x_i, the measurements of the unary properties $\mathbf{y}_i$, as well as the domain knowledge represented by $\mathcal{K}$. Suppose that $\mathbf{y}_i$ has m components $y_i^{(1)}, y_i^{(2)}, \ldots, y_i^{(m)}$ representing measurements of m defined features of Ω_i. With the assumption that they are independent, a single node

clique function is defined as

$$V_c(x_i : \mathcal{K}, \mathbf{y}_i) = \sum_{j=1}^{m} w_c^{(j)}(x_i, \mathcal{K}) B_c^{(j)}(x_i : \mathcal{K}, y_i^{(j)}) \tag{5.1}$$

where $B_c^{(j)}(\cdot), j = 1, 2, \ldots, m$ are called *basis functions*, and the $w_c^{(j)}(x, \mathcal{K})$ are a set of non-negative weights for the basis functions with

$$\sum_{j=1}^{m} w_c^{(j)}(x, \mathcal{K}) = 1 \tag{5.2}$$

$w_c^{(j)}(x, \mathcal{K})$ not only depends on the selected feature j, but also upon the labelling configuration x and knowledge $\mathcal{K}$. However, for simplicity $w_c^{(j)}(x, \mathcal{K})$ is chosen to be a constant. The basis function should give low values when the measurements $\mathbf{y}$ are consistent with the knowledge $\mathcal{K}$ for the specified interpretation x. There are a number of different basis functions that can be used, as presented in [78]. For example, the function can choose the form of the negative logarithm of the measurement distribution given the interpretation:

$$\begin{aligned} B_c^{(j)}(x : \mathcal{K}, y^{(j)}) &= -a_c^{(j)} \log P_c^{(j)}(x | \mathcal{K}, y^{(j)}) \\ &\propto -a_c^{(j)} \log P_c^{(j)}(y^{(j)} | \mathcal{K}, x) \end{aligned} \tag{5.3}$$

And if Gaussian distribution is assumed for the basis function, it results in a square form of distance function as shown in Figure 5.1(a). Such a function is not robust to outliers. Considering that the number of training samples are often limited for the high level classification task and the Gaussian assumptions may not be valid, more robust distance functions such as Figure 5.1(b) are favored. Other possible functions include the thresholding in Figure 5.1(c) and the piecewise linear function

in Figure 5.1(d) related to fuzzy set theory. One special basis function that needs to be mentioned is a constant function, whose value is always 1. The constant function is especially useful for incorporating rule based knowledge. For example, if an interpretation x_i is not valid according to some rule, the constant basis function increases the clique energy regardless of the feature values.

Without loss of generality, the basis functions in Figure 5.1 are all positive and assume that the mean feature value of the class is 0. The basis function is symmetric if there is no reason to manipulate differently the larger values and smaller values of the features. Sometimes, to use a one sided version of those functions is necessary. For example, ellipse fitting error can be used as a shape feature for round ice floes. Such a feature is always positive, and the smaller it is, the lower value of the basis function. The single-sided version of the piecewise linear function in Figure 5.1(d) is a suitable choice, and is used in the thesis.

The multiple node clique functions are also in the form of the weighted sum of basis functions. The design principle of multiple node clique energies is identical to that of single node clique energy, except that binary or higher order relations are used as features instead of unary properties. This thesis limits to pair node cliques without considering higher order cliques.

5.2 Implementation of SAR Sea Ice Classifier

5.2.1 Selected Features and Clique Energy

The domain knowledge involved in SAR sea ice interpretation includes those related to tone, texture, shape, and existence of leads (long narrow ice fractures). Incorporation of the domain knowledge into the system and investigation of the

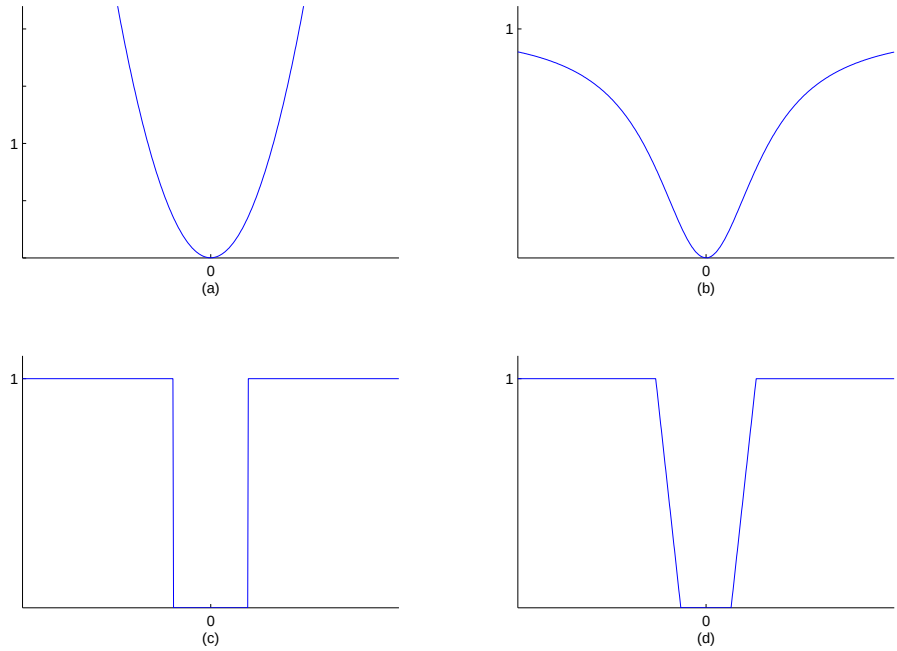


Figure 5.1: Example of different basis function

possible choice of clique energy functions requires training samples or other form of data provided by experts. For example, Mohammed Shokr at the Canadian Atmospheric Environment Service is leading a project of analyzing tone statistics of SAR sea ice based on past images and ice records. Such statistics are possibly valuable for constructing the knowledge base for the system. In this thesis, training samples are extracted from 8 SAR sea ice scenes in addition to Figures 4.29 and 4.36, the details of which will be presented later in section 5.2.3. Based on the selected training samples, various features have been investigated on whether they are suitable for ice identification and how they are encoded in the MRF clique energy function.

Single node clique energy

Single node clique energies model the unary properties of the objects. For the clustering in section 4.2.3, the single node energy (4.14) describes how similar the measured features of the segmented region is to the mean feature vector of a region cluster. For the SAR sea ice application, it is the mean feature vector of the ice types that should be matched against, and preferred features for the domain should be measured from the segmented regions. Here tone, texture features, and shape descriptors are considered.

The simplest unary property of a segmented region is the mean of the tone. However, tone is sensitive to the environmental factors and imaging parameters. Figure 5.2 shows the scatter plot of the tone of various ice types computed from the selected training samples. The fact that many ice types overlap significantly in the tone values is obvious. For example, water has a mean of tone around 50, while new ice, grey ice and thin first year ice all have quite a few samples located within that range. Similar situations have occurred for the two GLCM texture features CON and ENT ($G = 32$, $\delta = 1$, average of 0^o,45^o,90^o,135^o), as shown in Figure

5.3 and 5.4. As observed from all three figures, none of the classes have a distinct range of feature values relative to the other classes. There does exist some useful information in the figures such as the separability between water and multi-year ice, and the tendency of increase in the feature values from water to thin ice, and then to thick ice. However, such information is associated with the relative difference of the features rather than the feature values themselves, and hence is more easily represented by pair node clique functions.

It can be concluded that both the tone and GLCM texture features are not reliable for describing the ice types, and hence direct incorporation of them into the single node clique functions is not adopted. However, it has been observed that grey-white ice is often characterized by relatively dark ice floes with brighter ridges existing between. Thus, the following measure named *normalized CON* is used for capturing such texture features.

$$y_i^{(tex)} = \frac{\mathrm{CON}_i}{\mu_i^{(tone)} \mu_i^{(tone)}} \tag{5.4}$$

Here CON_i is the CON measured from region Ω_i, and $\mu_i^{(tone)}$ is the mean tone value of region Ω_i. The reason that CON is normalized by the square of mean tone is due to the multiplicative nature of speckle noise. The division is expected to be able to cancel the part of the texture strength contributed by speckle noise, while preserving and enhancing the texture of grey-white ice. The defined measure is closely related to the square of the coefficient of variation, which is commonly utilized in speckle filtering research [69]. Figure 5.5 shows that the normalized CON feature separates grey-white ice well from the two others for the sample image of Figure 4.36. Similar observations can be found from the computed feature values of the sample image of Figure 5.9(h) as well. However, for some other images where grey-white ice is

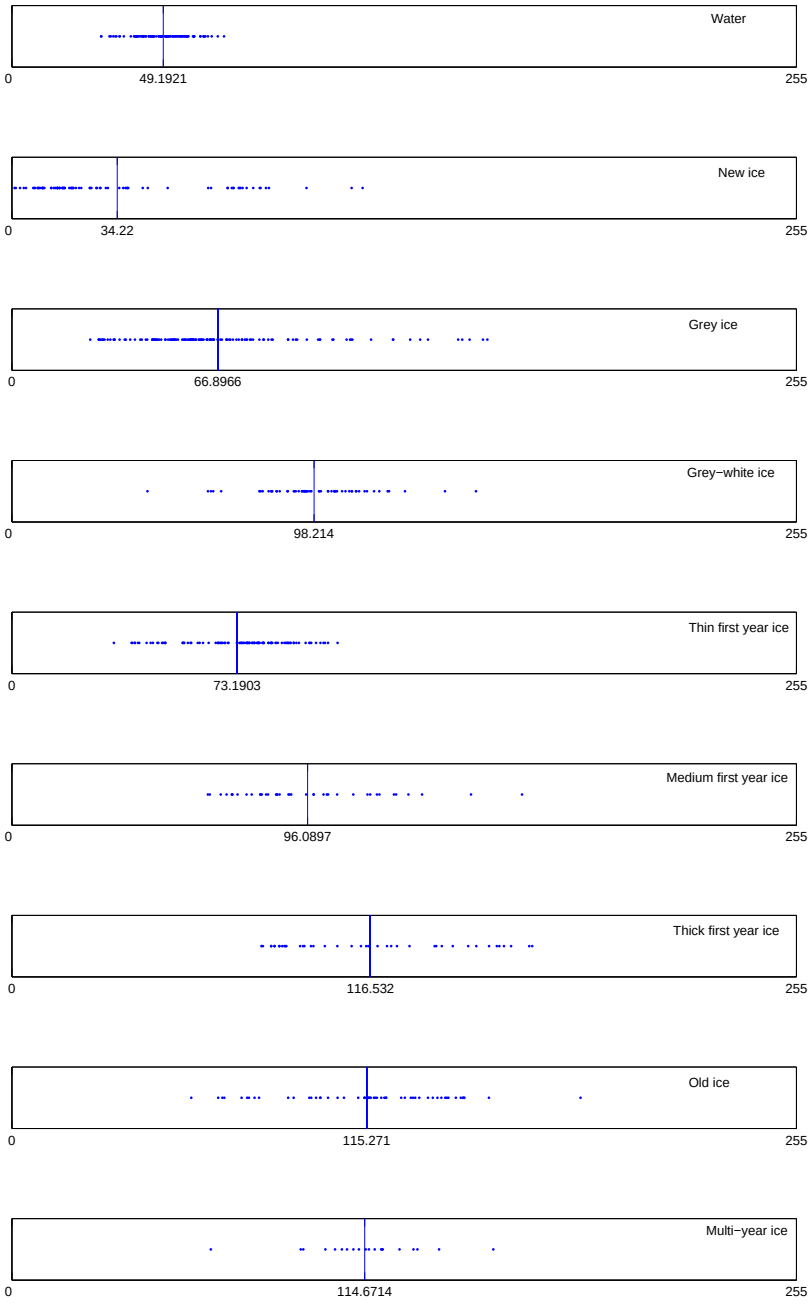


Figure 5.2: Distributions of the tone of water and ice classes.

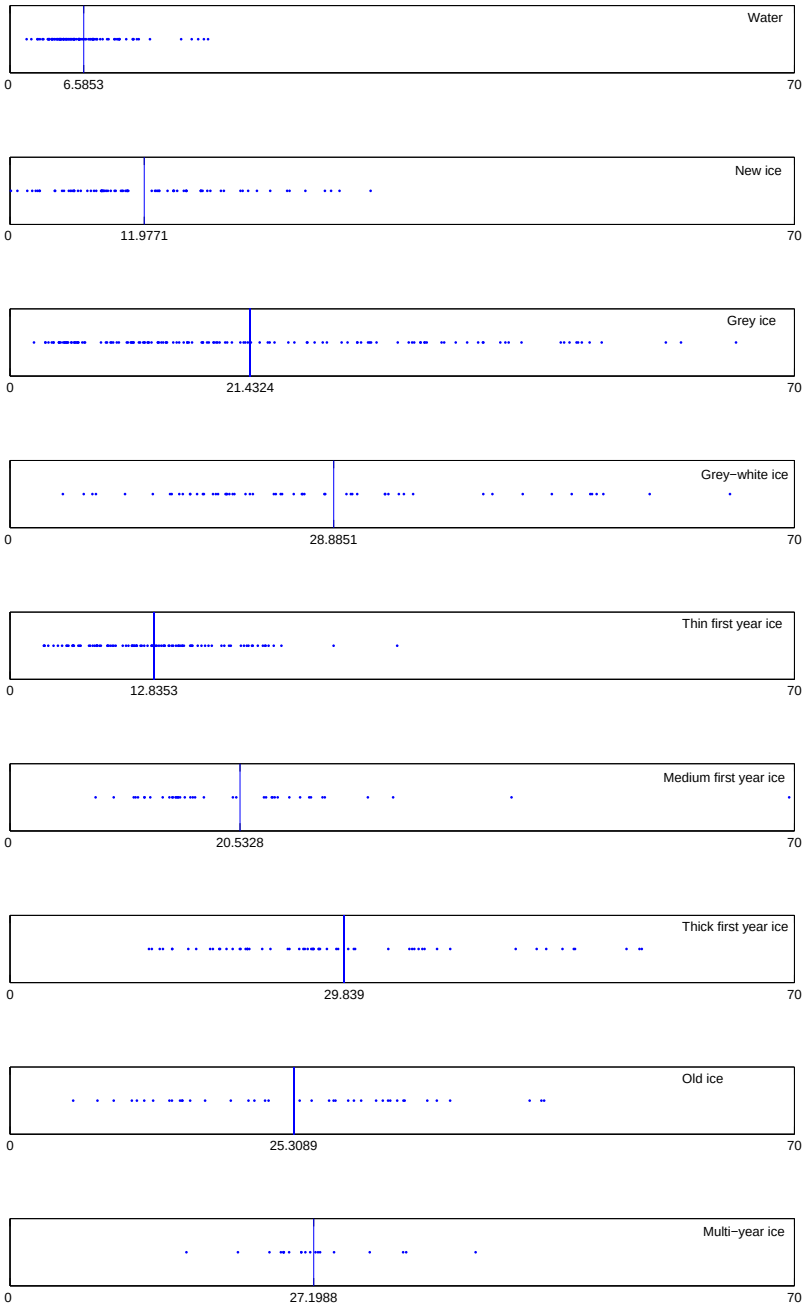


Figure 5.3: Distributions of the GLCM CON feature of water and ice classes ($\delta = 1$,$G = 32$,average of 0^o,45^o,90^o,135^o).

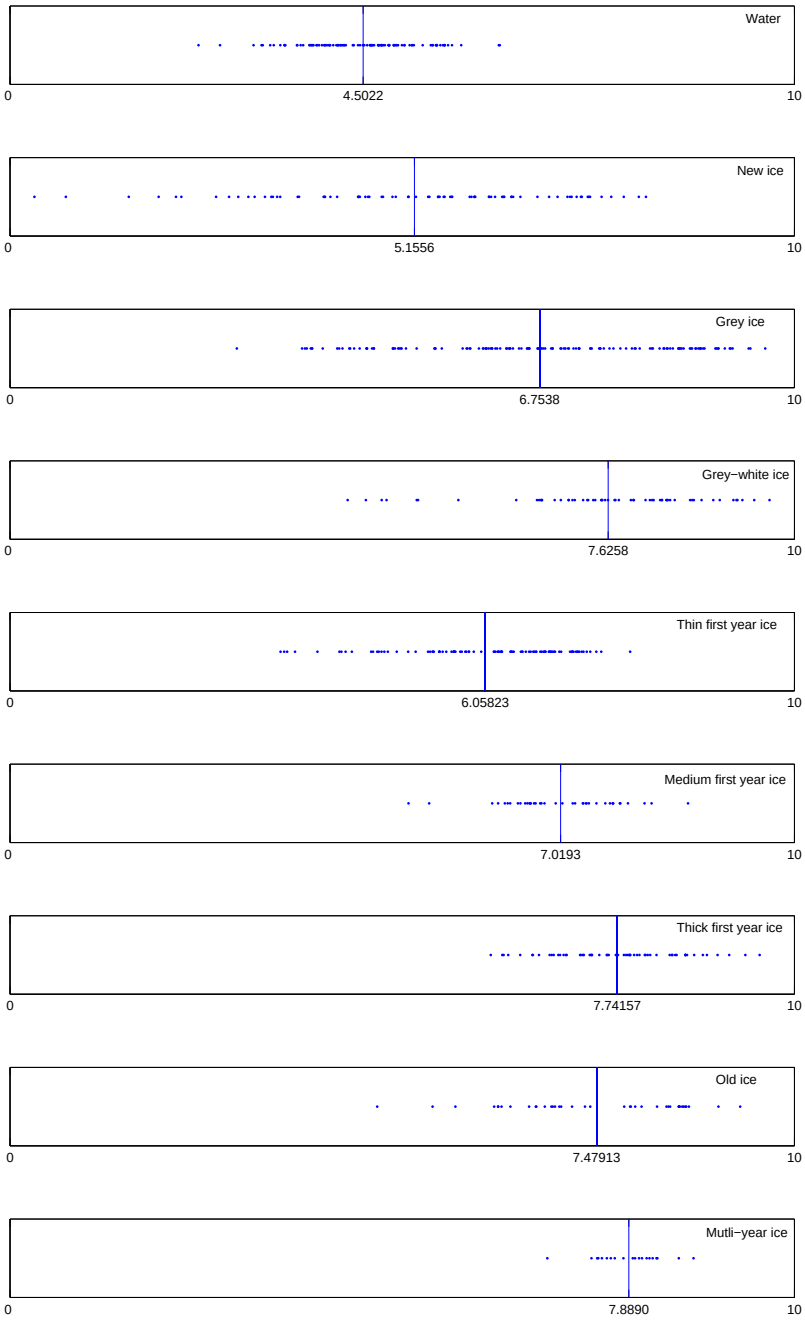


Figure 5.4: Distributions of the GLCM ENT feature of water and ice classes ($\delta = 1$,$G = 32$,average of 0^o,45^o,90^o,135^o).

relatively bright and do not exhibit obvious ridges, such as in Figure 4.29, 5.9(b), and 5.9(n), there are significant overlaps between the grey-white ice and other ice types with respect to the defined feature.

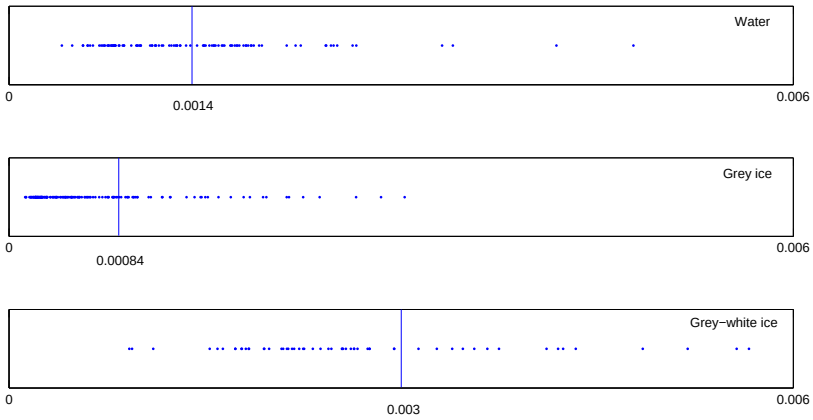


Figure 5.5: Distributions of the texture feature defined in (5.4) ($\delta = 8$, $G = 32$, average of 0^o,45^o,90^o,135^o), extracted from Figure 4.36.

Therefore, a hypothesis is made of the grey-white ice that deals with whether or not the grey-white ice has dark floes with bright ridges. The hypothesis is tested by investigating the histogram for the *normalized CON* of the image being analyzed. Suppose the region cluster having the highest mean *normalized CON* is C_h. That is, $\forall j, \bar{y}_{C_h}^{(tex)} \geq \bar{y}_{C_j}^{(tex)}$. If C_h is well separated from any other region cluster, i.e.

$$\forall j, \bar{y}_{C_h}^{(tex)} - \bar{y}_{C_j}^{(tex)} \geq \sigma_{C_h}^{(tex)} + \sigma_{C_j}^{(tex)}$$

where $\sigma_{C_j}^{(tex)}$ is the standard deviation of $y_s^{(tex)}$, $s \in C_j$, then C_h is said to have coarse texture and very likely corresponds to the grey-white ice. The corresponding clique energy of the region cluster C_h for the grey-white ice configuration is decreased by a constant [1]. The clique function thus is in the form of constant basis function,

[1]Here the nodes correspond to region clusters rather than disjoint regions. For an explanation

with the weight $w_c^{(tex)}$ being either a negative number (-0.3 by trials) or 0 according to whether the hypothesis is true or not. Moreover, a threshold $T^{(tex)}$ is set as $\bar{y}_{C_h}^{(tex)} - \sigma_{C_h}^{(tex)}$, and the corresponding clique energy for any region Ω_i is defined as follows:

$$B_c^{(tex)}(x_i : \mathcal{K}, y_i^{(tex)}) = \begin{cases} 1 & y_i^{(tex)} > T^{(tex)} \text{ and ice } x_i \text{ is grey-white ice} \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

Another unary property that can be incorporated into the single node clique function is the shape. Two shapes are of particular interest. The ellipse shape can be used for describing ice floes, and the elongated shape can be used for describing cracks and leads. The egg code has provided information on whether or not a specified ice type in the current region has well defined floe shape (i.e. the ice floe size is larger than the resolution of the image), and thus the corresponding clique basis function is defined as

$$B_c^{(ell)}(x_i : \mathcal{K}, y_i^{(ell)}) = \begin{cases} 0 & y_i^{(ell)} < C_1^{(ell)} \text{ and ice } x_i \text{ has well defined floes} \\ \frac{y_i^{(ell)} - C_1^{(ell)}}{C_2^{(ell)} - C_1^{(ell)}} & C_1^{(ell)} \leq y_i^{(ell)} < C_2^{(ell)} \text{ and ice } x_i \text{ has well} \\ & \text{defined floes} \\ 1 & C_2^{(ell)} \leq y_i^{(ell)} \text{ and ice } x_i \text{ has well defined floes} \\ 0 & \text{otherwise} \end{cases} \tag{5.6}$$

Here, $C_1^{(ell)}$ and $C_2^{(ell)}$ are two parameters for the piece-wise linear basis function as depicted in Figure 5.6, and are experimentally set as 4 and 12 respectively. $y_i^{(ell)}$ is

see section 5.2.2.

the ellipse fitting error normalized by the boundary length, computed by

$$y_i^{(ell)} = \frac{\sum_{s \in \partial\Omega_i} D(s, \mathbf{e})}{\int ds}$$

where $D(s, \mathbf{e})$ is the Euclidean distance between boundary site s and the nearest site of ellipse $\mathbf{e}$.

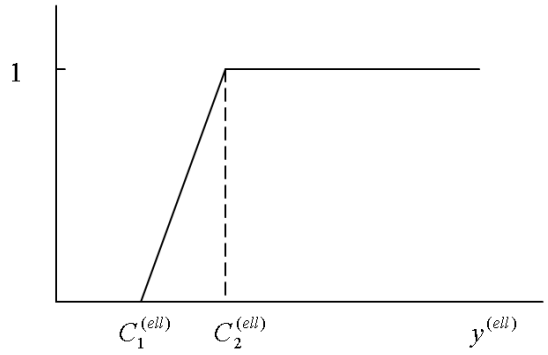


Figure 5.6: Basis function for ellipse shape energy.

The degree that a shape is elongated is measured by $y^{(elon)} = l_{cross}/l_{max}$, where l_{max} is the length of the minimum bounding rectangle, and l_{cross} is the minimum crossing length of the segmented region in the direction normal to the long axis of the rectangle. They are depicted in Figure 5.7.

The basis function for the elongated shape energy is defined in (5.7), and the corresponding shape is illustrated in Figure 5.8. Here, the basis function is designed in a robust manner. When $y_i^{(elon)}$ is much smaller than $C^{(elon)}$, the function approximates the square Euclidean distance from the feature value to the zero. If $y_i^{(elon)}$ is much larger, it gradually approaches the upper limit of 1. This kind of function is more robust to outliers than the commonly used square distance function. Here

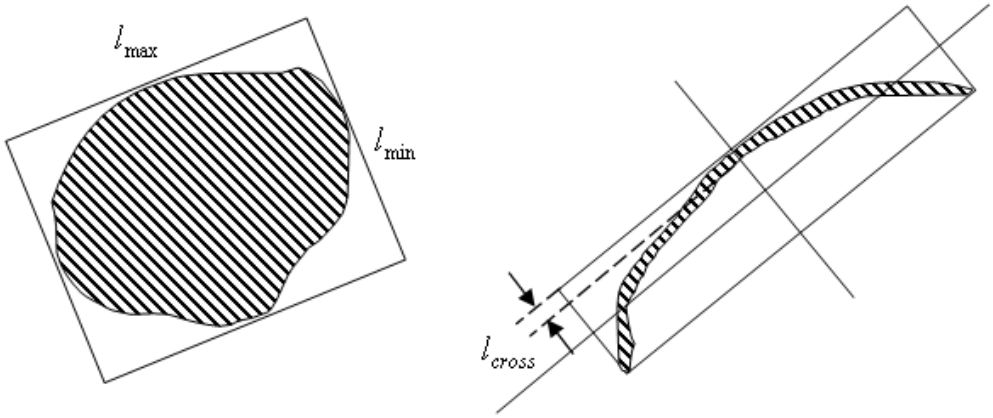


Figure 5.7: Minimum bounding rectangle and shape parameters.

in the system, $C^{(elon)} = 0.2$.

$$B_c^{(elon)}(x_i : \mathcal{K}, y_i^{(elon)}) = \frac{\left(\frac{y_i^{(elon)}}{C^{(elon)}}\right)^2}{1 + \left(\frac{y_i^{(elon)}}{C^{(elon)}}\right)^2} \tag{5.7}$$

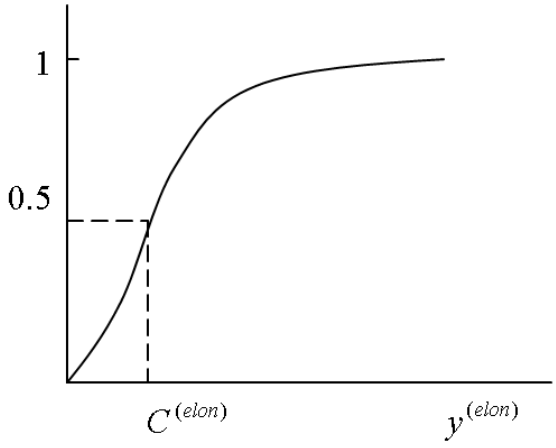


Figure 5.8: Basis function for elongated shape energy.

There are also many other features that can be possibly considered as candidates. For example, Soh *et. al* [101] has listed more than 20 features applied in the ARKTOS system. This thesis tries to design a system that is relatively simple and general, and hence selects a limited number of essential features.

Multiple node clique energy

Tone difference is a useful feature for describing binary relations. This is reasonable as observations can often be made on situations such as "new ice looks darker than thicker ice in an image". The distribution of the tone difference can be measured from training samples, and assuming that the feature is Gaussian distributed the corresponding basis function can be defined as the negative logarithm of this Gaussian pdf. However, for the consideration of outliers again, the robust scheme is adopted resulting in the following basis function:

$$B_c^{(td)}(x_i, x_j : \mathcal{K}, y_i^{(td)}, y_j^{(td)}) = \frac{\left(\frac{y_i^{(td)} - y_j^{(td)} - \mu_{x_i x_j}^{(td)}}{\sigma_{x_i x_j}^{(td)}}\right)^2}{C^{(td)} + \left(\frac{y_i^{(td)} - y_j^{(td)} - \mu_{x_i x_j}^{(td)}}{\sigma_{x_i x_j}^{(td)}}\right)^2} \tag{5.8}$$

where $\mu_{x_i x_j}^{(td)}$ is the mean of tone difference between class x_i and x_j, $\sigma_{x_i x_j}^{(td)}$ is the standard deviation of tone difference between class x_i and x_j, and $C^{(td)}$ is a constant (set as 1 here).

The spatial context of ice types is another important binary feature. For example, leads can often be found inside grey ice floes. Because of such a observation, the system considers the lead as a sperate class with the shape defined as elongated.

The basis function corresponding to spatial context is as follows

$$B_c^{(sc)}(x_i, x_j : \mathcal{K}, y_i^{(sc)}, y_j^{(sc)}) = \begin{cases} -1 & \text{ice } x_i,\ x_j \text{ are often spatially adjacent} \\ 0 & \text{otherwise} \end{cases} \tag{5.9}$$

5.2.2 Overall System and Optimization

In addition to a different classification process, the low level segmentation also uses a slightly different objective function from that in chapter 4. The objective function for region merging is as follows:

$$\arg \min_{\Omega_1,\ldots,\Omega_n} \left\{ \sum_{i=1}^{n} \sum_{s\in\Omega_i} \left\{ \frac{1}{2}\log(|\Sigma_i|) + \frac{1}{2}(\mathbf{y}_s - \mu_i)^t \Sigma_i^{-1}(\mathbf{y}_s - \mu_i) \right\} \right.$$
$$\left. +\beta_1 \sum_{i=1}^{n} \sum_{s\in\partial\Omega_i} g(|\nabla_s|) + \alpha \sum_{i,\forall x_i \in X_f} \sum_{s\in\partial\Omega_i} e_i(s) \right\} \tag{5.10}$$

Compared to the objective function in (4.11), there is an additional term corresponding to a shape related energy. Here, $e_i(s)$ is the error of site s in the ellipse fitting for the boundary $\partial\Omega_i$, α is a positive constant set to be 0.01, x_i denotes the class (object) type for segment Ω_i, and X_f is the set of ice types that are characterized by well defined floes for which ellipse shape segments, which correspond to individual ice floes, are favored.

The optimization of the overall IUS follows the diagram of Figure 4.7 and the description in 4.2.3. However, for the high level classification step in each iteration, classification is performed first on the region clusters and then on individual regions. The reason is similar to why region merging is adopted in IRGS - using a large group of pixels or regions as the unit for labelling allows reduced computations to reach a solution. The subsequent classification on individual regions acts as a refining

process in a top-down manner. Also, some features or clique functions have to be computed on a region cluster base (e.g. the distribution of the normalized CON), and exhaustive classification methods can be applied to region clusters since the number of clusters is no more than 5 (in addition to water, the egg code definition allows at most 4 different ice types).

5.2.3 Training

SAR sea ice images of 8 different scenes have been used for preparing the training samples, covering various regions such as Baffin Bay, Gulf of St. Lawrence, and Beaufort Sea, and in various seasons as well. They are all acquired by RADARSAT in ScanSAR C-band mode and have resolution of $100m$ (images of $50m$ resolution 2×2 block averaged by CIS). Based on the ice chart, polygon regions that have relatively clear interpretations (i.e. only a single ice type is contained or the ice mixture can be visually separated without much difficulty) are extracted from those images. Table 5.1 provides a list of those images, and Figure 5.9 are the thumbnails.

The selected polygon regions are further segmented (without ice identification) by the IRGS presented in chapter 4. The obtained segmentation maps are then manually refined, and ice identification is performed based on visual interpretation of the segmented regions. By doing that, an ice interpretation map has been prepared manually with the aid of computer for each training sample image. It should be mentioned that during the manual preparation of the interpretation map, any region that has low confidence in being identified as one ice type rather than others will be assigned a "unsure" label and excluded from future training process.

Among all the features incorporated, difference of tone is the only one that needs to be trained. That is, the two parameters associated with the clique energy

Figure	Date	Location	Number of Classes	Ice types (see Table 2.1 for details)
5.9(a)	Feb 07, 1998	Baffin Bay	2	thin first-year, thick first-year
5.9(b)	Feb 19, 1997	Gulf of St. Lawrence	4	thin first-year, grey-white, grey, water
5.9(c)	Jun 24, 1998	Baffin Bay	2	thin first-year, thick first-year
5.9(d)	Oct 13, 1997	Beaufort Sea	3	grey, new, water
5.9(e)	Oct 13, 1997	Beaufort Sea	2	multi-year, grey
5.9(f)	Jun 27, 1998	Beaufort Sea	3	thin first-year, thick first-year, water
5.9(g)	Jun 24, 1998	Baffin Bay	1	thick first-year
5.9(h)	Feb 20, 1998	Gulf of St. Lawrence	4	thin first-year, grey-white, grey, water
5.9(i)	Jun 24, 1998	Baffin Bay	2	thin first-year, grey
5.9(j)	Mar 01, 2003	Gulf of St. Lawrence	1	thick first-year
5.9(k)	Mar 08, 2003	Gulf of St. Lawrence	2	medium first-year, thin first-year
5.9(l)	Mar 01, 2003	Gulf of St. Lawrence	2	medium first-year, thin first-year
5.9(m)	Mar 01, 2003	Gulf of St. Lawrence	2	grey, new
5.9(n)	Mar 01, 2003	Gulf of St. Lawrence	3	grey-white, grey, new
5.9(o)	Mar 01, 2003	Gulf of St. Lawrence	2	grey, new

Table 5.1: Details of the training images. All images are RADARSAT in ScanSAR C-band mode with the resolution of $100m$. The ice types are defined in Table 2.1.

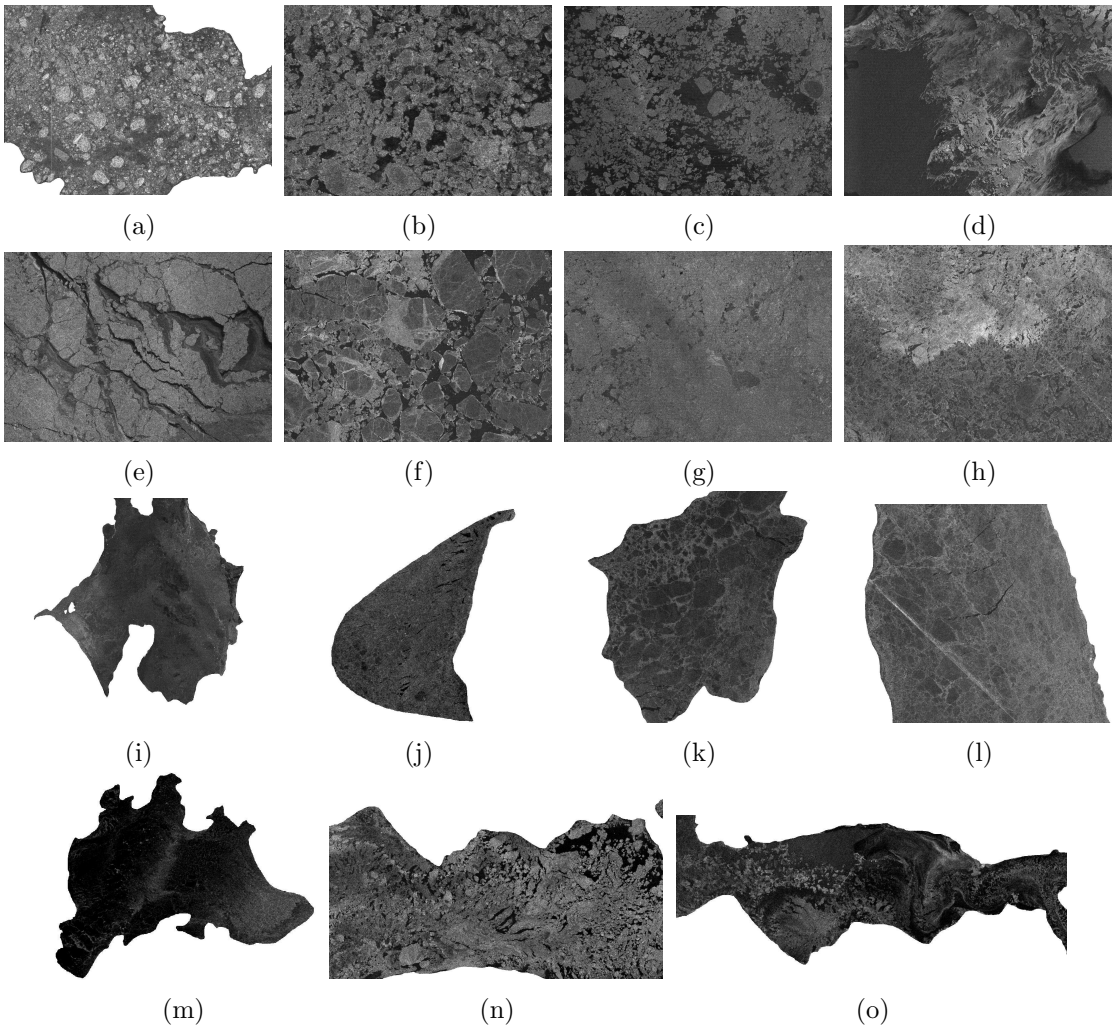


Figure 5.9: Selected training sample images. See Table 5.1 for image details.

corresponding to tone difference, $\mu_{x_i x_j}$ and $\sigma_{x_i x_j}$ in (5.8), needs to be estimated from the training samples. Since the available SAR sea ice images are limited, the interpretation maps obtained manually do not provide sufficient disjoint regions for any pair of ice types. Smaller regions from intermediate IRGS results are thus randomly selected as training samples. Based on the prepared interpretation map, the mean and variance of tone difference is computed from the selected training samples.

Other parameters include the $w_c^{(tex)}$ (the weight associated with normalized CON), $w_c^{(ell)}$ (the weight associated with ellipse shape), $C_1^{(ell)}$ and $C_2^{(ell)}$ (the ellipse shape basis function parameters), $w_c^{(elon)}$ (the weight associated with elongated shape), $C^{(elon)}$ (the elongated shape basis function parameter), $w_c^{(td)}$ (the weight associated with tone difference energy), $C^{(td)}$ (the tone difference basis function parameter), and $w_c^{(sc)}$ (the weight associated with spatial context energy). The values of those parameters are selected by trials, and are summarized in Table 5.2.

$w_c^{(tex)}$	$w_c^{(ell)}$	$C_1^{(ell)}$	$C_2^{(ell)}$	$w_c^{(elon)}$	$C^{(elon)}$	$w_c^{(td)}$	$C^{(td)}$	$w_c^{(sc)}$
-0.3	0.1	4	12	6	0.2	1	1	1

Table 5.2: Summary of the parameters for the high level classification.

5.3 Algorithm Testing

Experiments have been performed on the training set and several other SAR sea ice images. All are acquired by RADARSAT ScanSAR C mode with $100m$ resolution. Some of the experiment results are shown in Figure 5.10-5.17. Due to the difficulty of obtaining pixel level ground truth, the results are evaluated subjectively. An

online meeting with experts in Canadian Ice Service (CIS) was held on April 26, 2006 for such an evaluation. The experiment results agree (at least roughly) with the human expert interpretations, and the system has performed successfully.

In Figure 5.10, the red lines are region boundaries before the final classification. The system has shown some degree of capability of preserving individual ice floes, which is useful in identifying ice types. This justifies the incorporation of ellipse shape energy in the objective function of (5.10) for segmentation.

Experiments are then performed on the image of Figure 4.36, which has already be included in the segmentation algorithm testings in chapter 4. The result is presented in Figure 5.11. Compared to the segmentation result of Figure 4.41, a number of leads and some grey ice regions in the right have been preserved. Thus, the system has achieved a good balance between preserving details and producing large homogeneous regions using the domain specific model.

The roles of various features used in the classification model are then investigated by examining the results obtained without using the investigated features in the classification. In Figure 5.12, two experiments are performed excluding respectively the spatial context feature, which is modelled by Equation (5.9), and the coarse texture feature, which is modelled by Equation (5.5). In Figure 5.12(a) where the spatial context feature is excluded, small regions of leads and grey ice are not preserved. Without the spatial context feature which favors the co-occurrence of leads and grey ice, those leads and grey ice are more easily to be classified as other ice types, and subsequent merging of those regions with surrounding regions is likely to occur. In Figure 5.12(b), the labelling of grey and grey-white ice is flipped. This is because that the tone difference feature used here assumes that grey-white ice is generally brighter than grey ice (may not be true, but is the result of the preliminary training). Without using the coarse texture knowledge, there is

no reason why the system should choose the relatively bright region in the image as grey ice.

In Figure 5.13, the system has identified successfully large regions of homogeneous ice, but on the other hand has lost some small details such as the gaps among the floes at the bottom of the image. At the bottom right of the image, an open water region has been identified by CIS experts but misclassified as new ice by the system. Such an error occurs due to the relatively bright appearance of the rough water in this region caused by wind, and the relatively strong edge between the smooth water and rough water.

The interpretation in Figure 5.14 has also lost some details such as narrow cracks. Errors have also occurred at the bottom right of the image, where some multi-year ice has been misclassified as grey ice due to the relatively dark appearance caused by snow cover, as indicated by CIS experts. Nevertheless, the system is again successful.

The three images of Figure 5.15-5.17 are difficult because there are no distinct boundaries and the appearance of some ice types are very close. The system performs surprisingly well and has generated results that look meaningful.

5.4 Chapter Summary

In this chapter, a complete image understanding system has been constructed. The system has combined bottom-up and top-down approaches, and is simple and extendable. The performance of the system is satisfactory based on preliminary evaluations on the image set currently available. Further validation by remote sensing and ice experts on more data sets are needed.

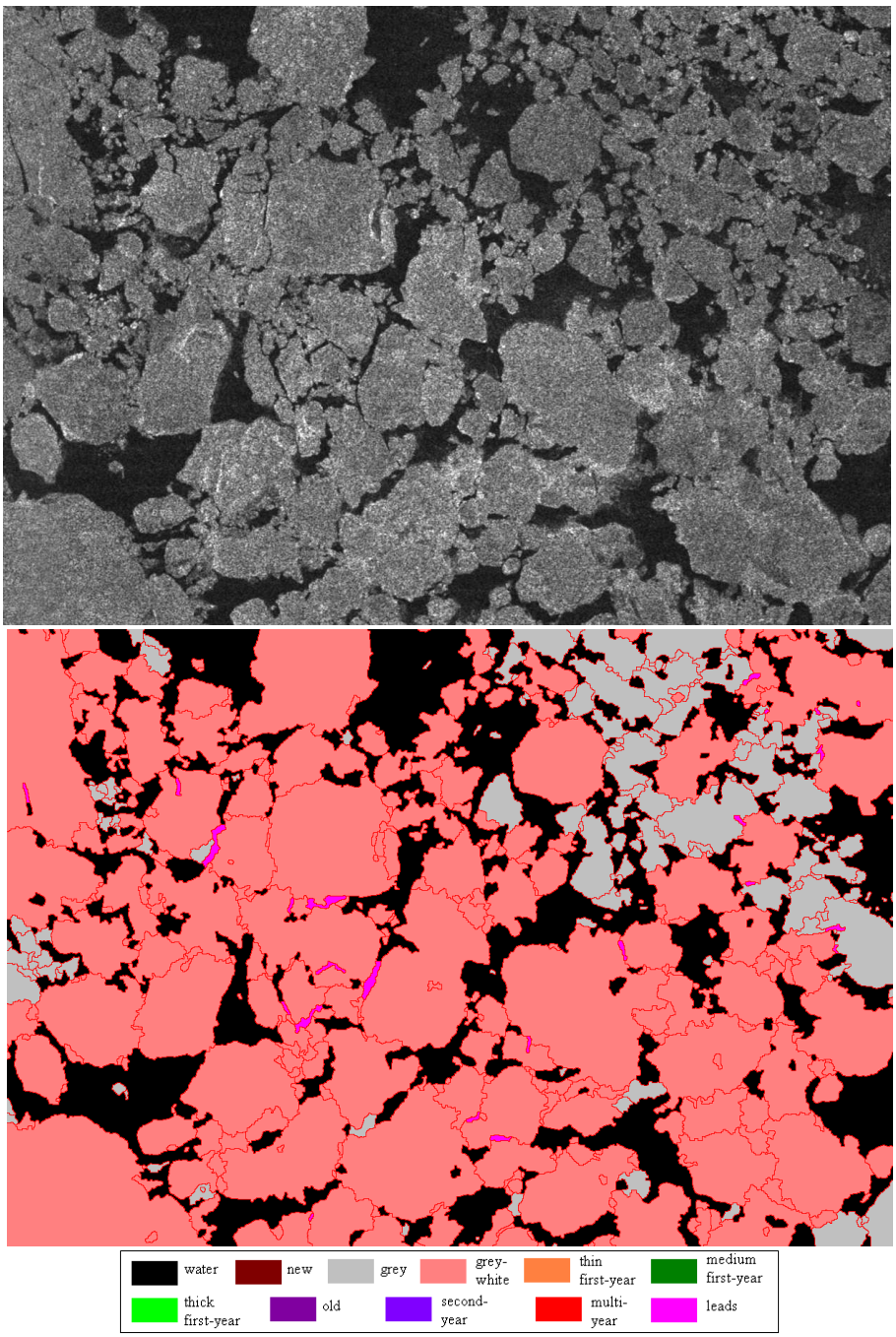


Figure 5.10: Interpretation result on Figure 4.29 [p. 98].

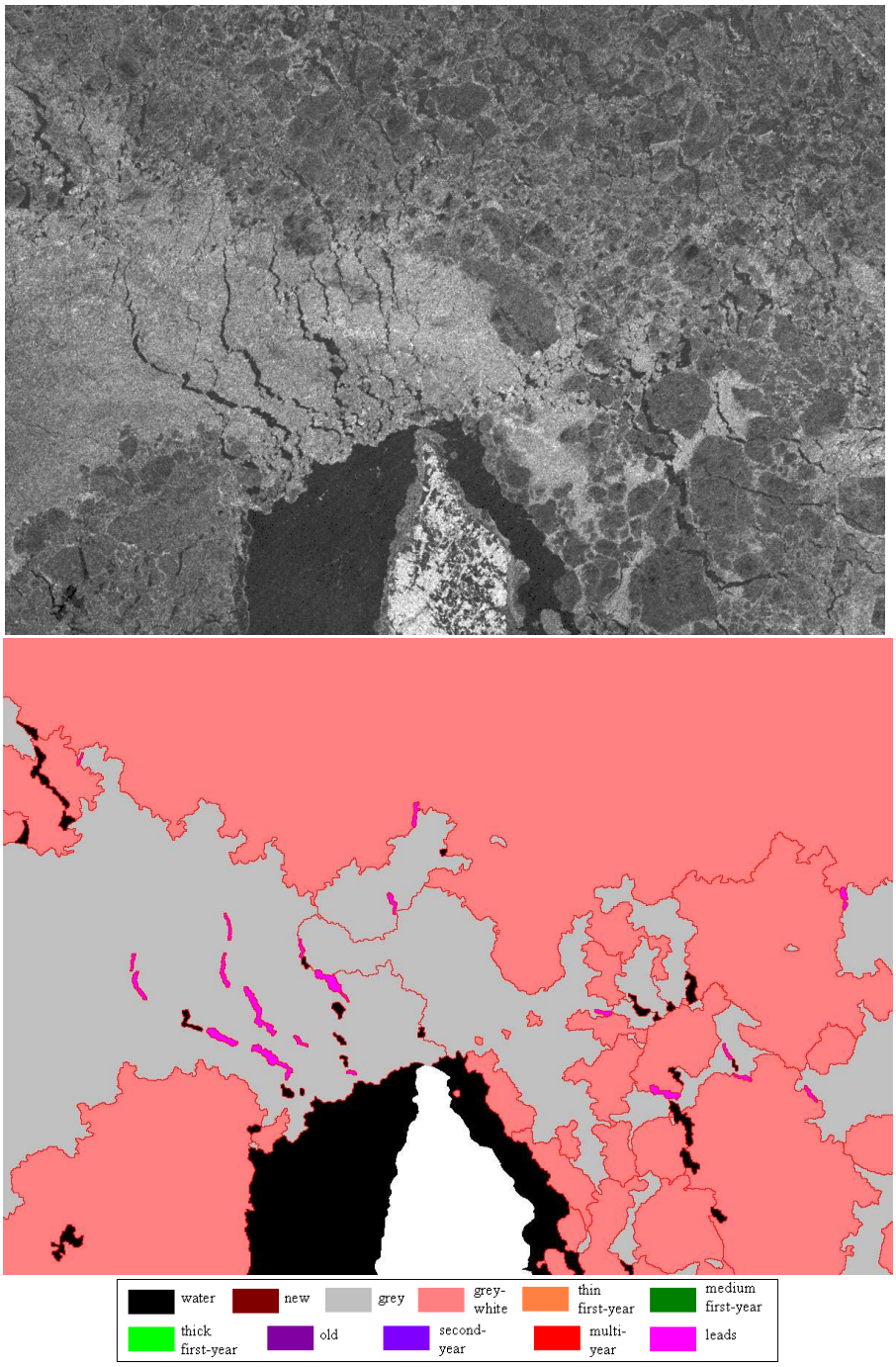


Figure 5.11: Interpretation result on Figure 4.36 [p. 102].

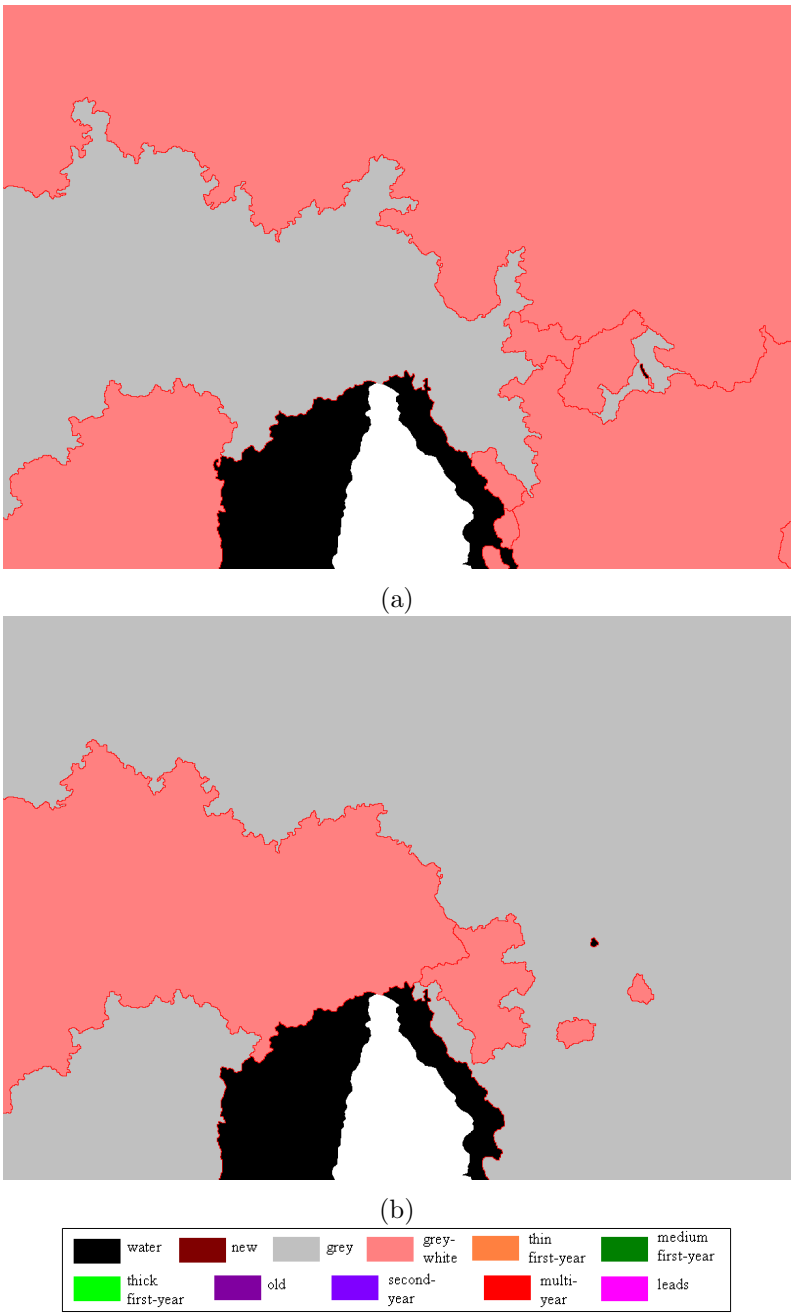


Figure 5.12: Interpretation result on Figure 4.36[p. 98] without using some features. (a) Spatial context modelled by Equation (5.9) is excluded, resulting an under-segmentation in which many leads and grey ice regions disappear; (b) Coarse texture indicator modelled by Equation (5.5) is excluded, causing complete wrong ice identification between grey ice and grey-white ice.

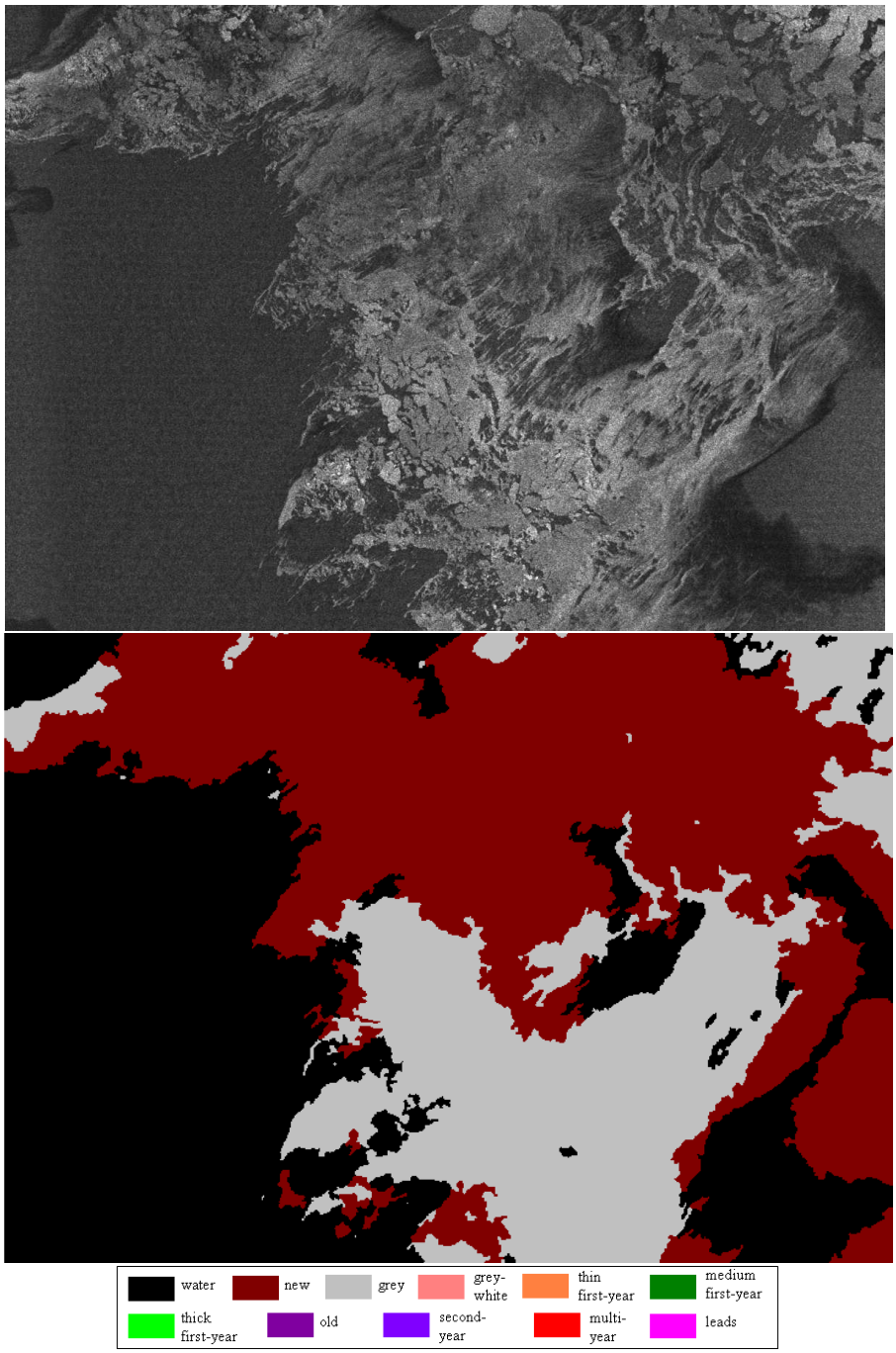


Figure 5.13: Interpretation result on Figure 5.9(d) [p. 132]. Boundary lines removed.

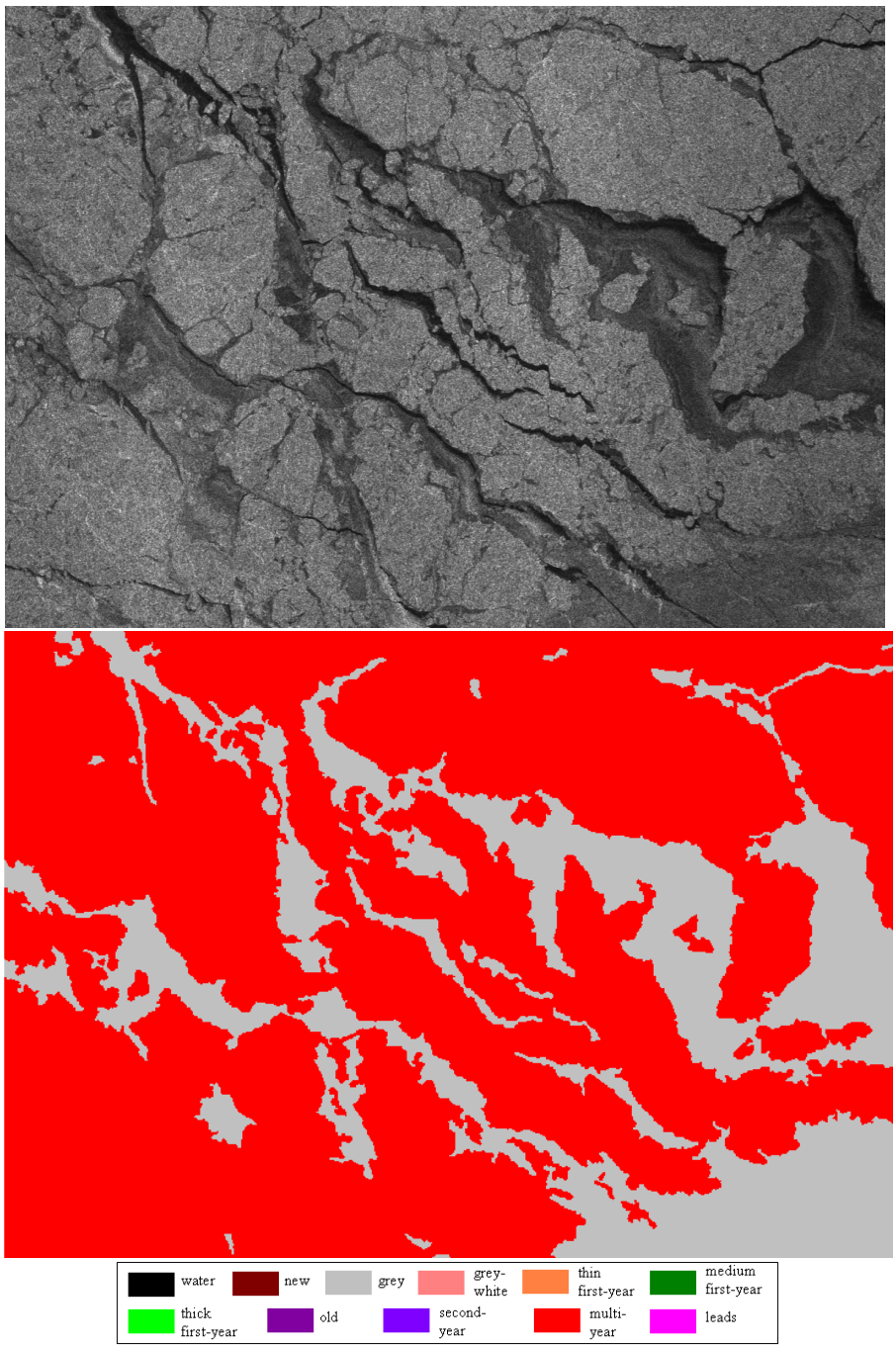


Figure 5.14: Interpretation result on Figure 5.9(e) [p. 132]. Boundary lines removed.

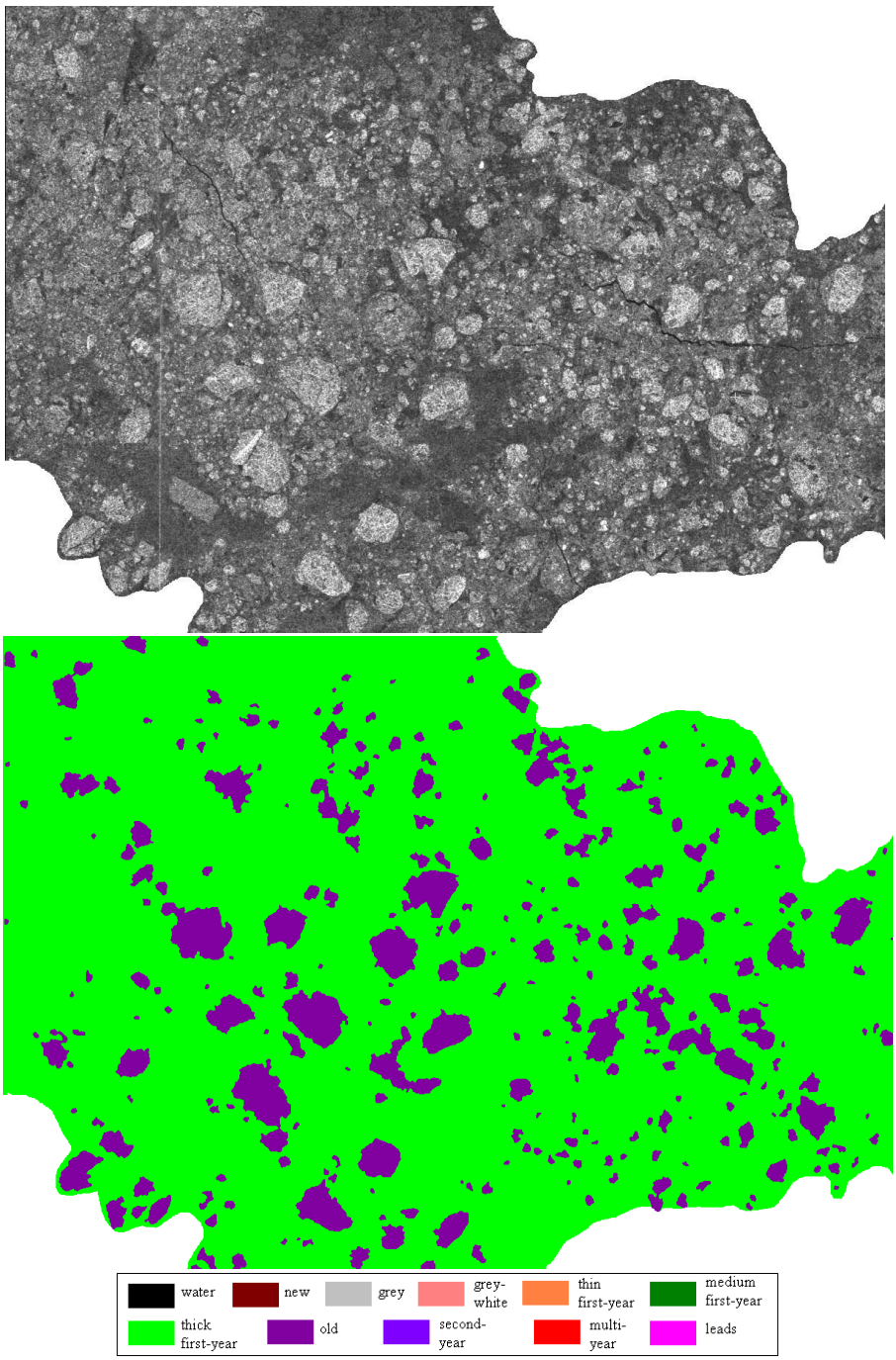


Figure 5.15: Interpretation result on a Baffin bay image acquired on Feb 7, 1998. The region has old ice and thick first-year ice. Boundary lines removed.

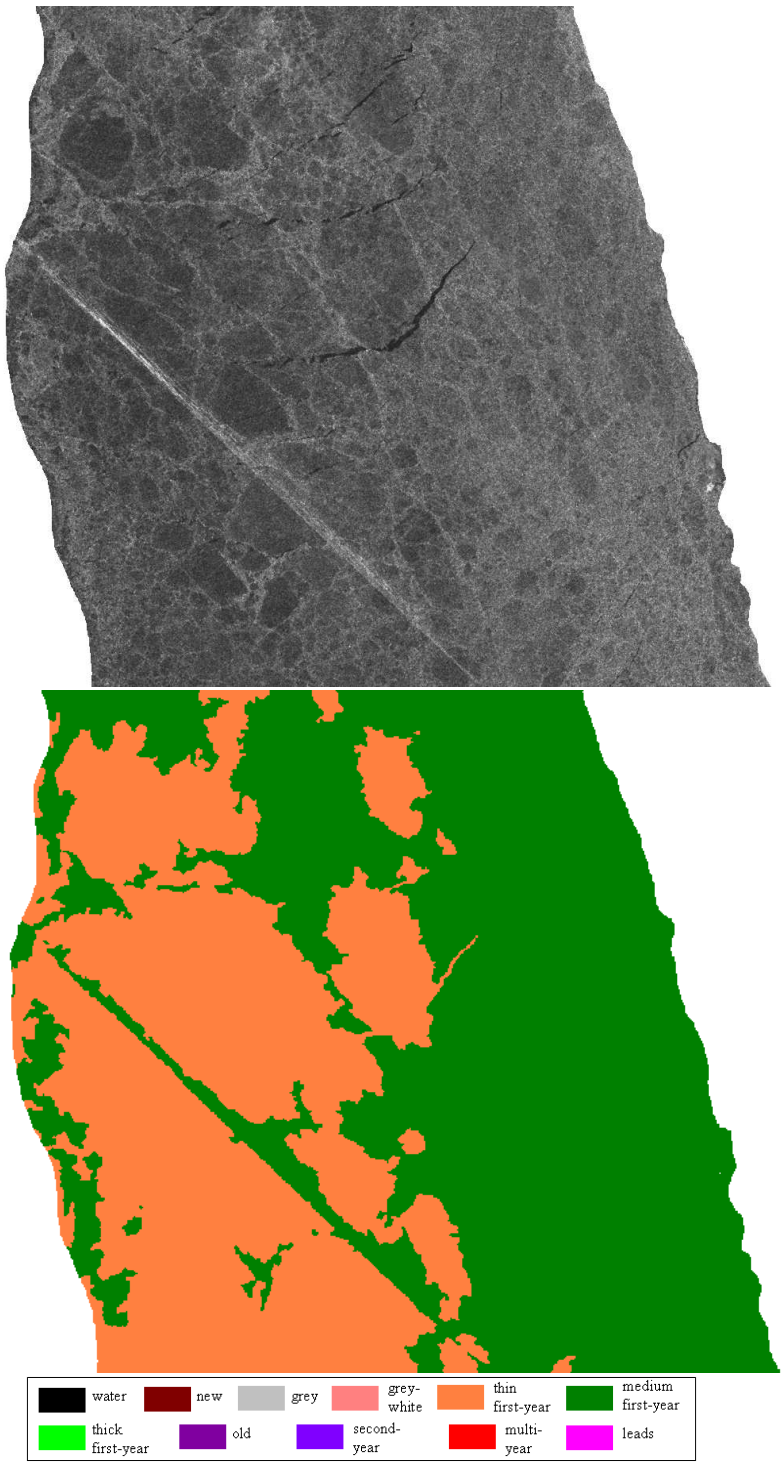


Figure 5.16: Interpretation result on Figure 5.9(l) [p. 132]. Boundary lines removed.

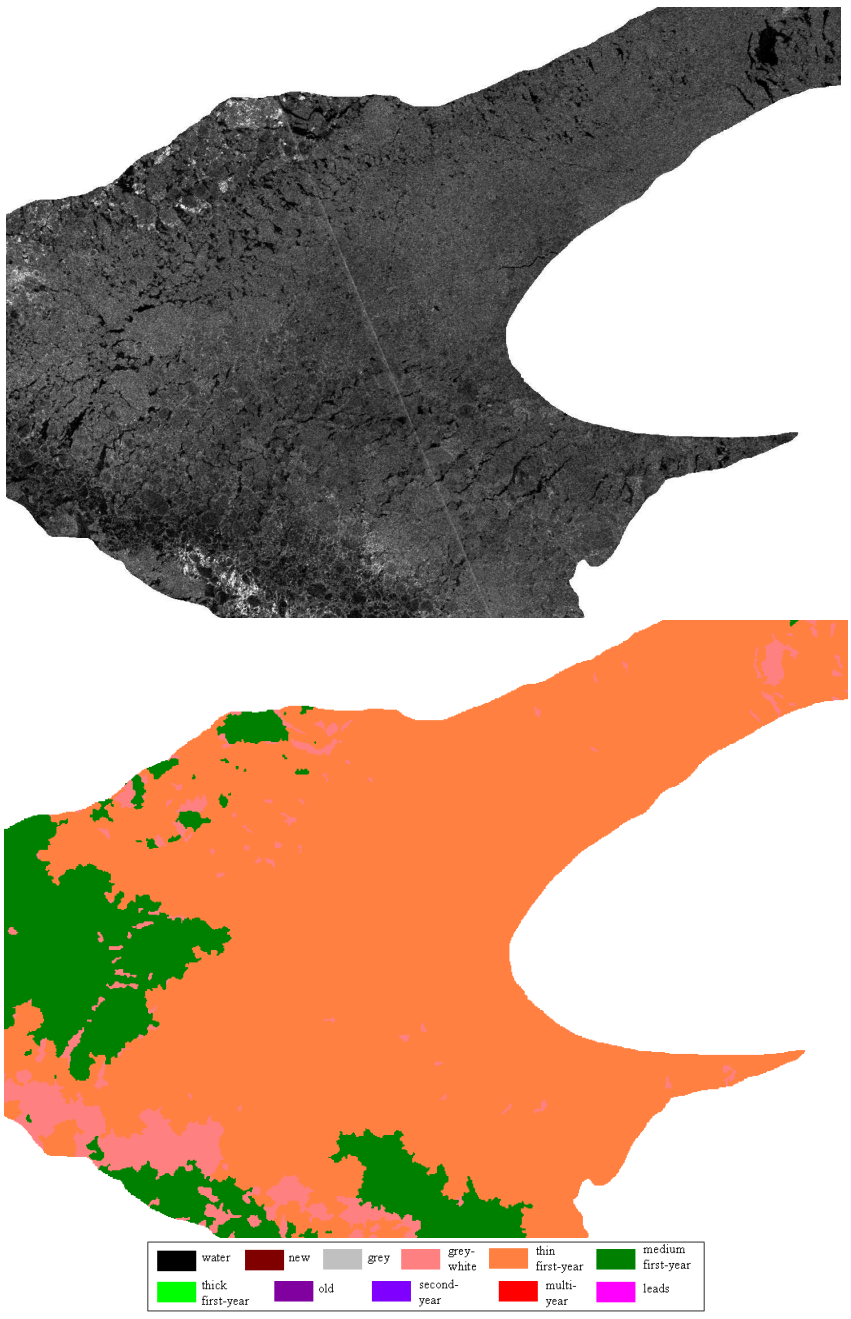


Figure 5.17: Interpretation result on a Gulf of St. Lawrence image acquired on Mar 1, 2003. The region has thin first-year, medium first-year, and grey-white ice. Boundary lines removed.

Chapter 6

Conclusion

6.1 Summary and Research Contribution

This thesis deals with the task of developing an automated computer-assisted SAR sea ice image understanding system. A novel Markov random field (MRF) based image segmentation and interpretation method called iterative region growing on semantics (IRGS) is proposed, and its applicability to operational SAR sea ice image analysis is investigated. Experimental results have shown that the IRGS algorithm is efficient for both general segmentation and the SAR sea ice interpretation task.

The IRGS algorithm consists of two separate but cooperative processes, corresponding respectively to the image segmentation in the low level and the region labelling at a high level. Both of the two processes aim at locating the global minimum of a corresponding MRF-based energy function.

In the low level segmentation, the defined MRF-based energy function is approached by a series of different functions that are characterized by gradually increased edge penalty. The searching is performed by an iterative region growing

method, with each iteration a process for finding a local minimum of the current function approximation. Such an optimization method tends to obtain configurations that are composed of large homogeneous regions which, for most images investigated, correspond to lower values of the defined energy than those by other techniques such as the simulated annealing. From that perspective, IRGS is more advanced. It has been observed that IRGS may lag in preserving small details behind classical MRF based approaches, which are pixel-based and use the simulated annealing. However, this does not imply a weakness of IRGS in optimization but rather the inaccuracy of the MLL prior model.

A more suitable model for SAR sea ice imagery is then designed, also based on MRF modeling techniques. Various features such as tone, texture, shape and spatial context have been combined using the MRF clique potential formation, thus not only providing a better prior model for the low level segmentation but at the same time allow interpretations performed on the segmented regions. Such an MRF-base high level classification approach is simple and has been demonstrated to be able to produce meaningful results for SAR sea ice imagery.

The low level segmentation and the high level classification has been integrated into a complete image understanding system in a cooperative manner. The high level classification performs the labeling on the regions obtained by the low level segmentation, and on the other hand it influences the segmentation process in turn by resisting the merging operations between any two regions not belonging to the same class. Considered under the Bayesian framework, the overall image understanding system tries to solve a MAP (maximum a posterior) problem, with the low level image features and high level domain knowledge associated respectively with a feature model and a prior model. The summation of the energies of the two models gives the overall objective function, and both processes try to reduce one

part of the overall energy without changing greatly the other.

In summary, the contributions of this thesis include: 1) the design and implementation of a novel image segmentation algorithm and 2) the investigation and development of a practical image understanding system for SAR sea ice imagery analysis. To the knowledge of the author, there is only one system [101] that have incorporated various high level knowledge in developing an image understanding system for SAR sea ice, and none of the existing approaches have integrated the low level segmentation and the high level classification based on the MRF energy formulation under the Bayesian framework. The work in this thesis is not limited to the SAR sea ice field but can possibly be applied to applications in other domains.

6.2 Future Research Directions

6.2.1 Refine with region splitting

In the system developed in this thesis, the high level classification is not always accurate. Otherwise, there is no need to do the iterative refinements. Two regions that should belong to different classes may be assigned with the same class label, and thus have the possibility of being merged incorrectly. Since there is no reverse operation (i.e. region splitting) of the region merging in the system, errors caused by the merging cannot be corrected later.

Therefore, it is desirable to add region splitting steps in the solution searching. The splitting of a region into two arbitrary shape regions is usually very computationally expensive. A possible scheme is to limit the configuration space of a splitting to the one before merging. That is, the splitting does not generate new regions but only rolls back to a previous configuration. Thus, it is necessary to pre-

serve a record of the merging process, probably in the form of binary trees. At the top level is the roots of a set of binary trees with each root node corresponding to an object in the image. Top-down refinements can then be performed by changing the structures of the binary trees, computing the resulting energies of the region labeling configurations, and accepting or rejecting the change by deterministic or stochastic techniques.

6.2.2 Model the Incidence Effect

The segmentations and interpretations of the SAR sea ice imagery sometimes suffer heavily from the fading effect caused by the graduated change in incidence angle. The patterns of the tone and texture vary spatially. It has been observed that the relationship between the mean of tone in decibel scale and the incidence angle can be approximated by a linear model, of which the slope depends on the roughness and moisture content of snow or ice [75]. If an assumption can be made about the homogeneity of any segmented region on the roughness and moisture content of ice, the segmentation (region merging) process can possibly be improved by replacing the stationary Gaussian feature model with a spatially varying Gaussian feature model. To be specific, equation(5.10) can be modified as

$$\arg \min_{\Omega_1,\ldots,\Omega_n} \left\{ \sum_{i=1}^{n} \sum_{s\in\Omega_i} \left\{ \frac{1}{2}\log(|\Sigma_i|) + \frac{1}{2}(\mathbf{y}_s - \mu_i(s))^t \Sigma_i^{-1} (\mathbf{y}_s - \mu_i(s)) \right\} \right.$$
$$\left. + \beta_1 \sum_{i=1}^{n} \sum_{s\in\partial\Omega_i} g(|\nabla_s|) + \alpha \sum_{i,\forall x_i \in X_f} \sum_{s\in\partial\Omega_i} e_i(s) \right\} \tag{6.1}$$

where $\mu_i(s)$ is different from the μ_i in (5.10) in that it is the expected feature value of the region i at site s and varies with the coordinates of s. With a linear model,

$\mu_i(s)$ and the model coefficients can be computed using least square error fitting based on all feature values and coordinates of the sites in region i.

Moreover, if an assumption can be made about the homogeneity of a given ice type in a given egg code region on the roughness and moisture content, the slope of the linear model may also be used as a feature in the high level classification. Either applying a linear model to the segmentation or to the region labelling can be an interesting future work.

6.2.3 Incorporate Gabor Texture Features

Some experiments on including the GLCM texture features in IRGS have already been performed in chapter 4. Unfortunately, the implicit smoothing operation of the GLCM computation has made the MRF context model not quite necessary. The larger the window is, the less effect the context model has. Unlike GLCM, the computation of the Gabor features are not necessarily associated with smoothing, although a post-processing of smoothing is usually applied. Gabor features are able to provide different descriptions of textures than GLCM features [22], and are expected to be more efficient in identifying macro textures. Incorporating unsmoothed Gabor features in the IRGS system can be a future work.

Bibliography

[1] R. Adams and L. Bischof. Seeded region growing. *IEEE Trans. Pattern Anal. Machine Intell.*, 16(6), 1994.

[2] K. M. Andress and A. C. Kak. Evidence accumulation and flow of control in a hierarchical reasoning system. *AI Magazine*, 9(2):75–94, 1988.

[3] P. Andrey and P. Tarroux. Unsupervised segmentation of Markov random field modeled textured images using selectionist relaxation. *IEEE Trans. Pattern Anal. Machine Intell.*, 20(3):252–262, 1998.

[4] A. Baraldi and F. Parmiggiani. SAR image segmentation exploiting no background knowledge on speckled radiance: A feasibility study. Technical Report TR-98-018, International Computer Science Institute, Berkeley, CA, USA.

[5] A. Baraldi and F. Parmiggiani. An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *IEEE Trans. Geosci. Remote Sensing*, 33(2):293–304, 1995.

[6] D. G. Barber and E. F. LeDrew. SAR sea ice discrimination using texture statistics: A multivariate approach. *Photogrammetric Engineering and Remote Sensing*, 57(4):385–395, 1991.

[7] J. Bückner, M. Pahl, O. Stahlhut, and C. E. Liedtke. geoAIDA - a knowledge based automatic image data analyser for remote sensing data. In *Proceedings, 2nd International ICSC Symposium on Computational Intelligence Methods & Applications*, Bangor, Wales, UK, June 2001.

[8] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *J. of the Royal Statistical Society, Series B*, 36(2):192–236, 1974.

[9] J. Besag. On the statistical analysis of dirty pictures. *J. of the Royal Statistical Society, Series B*, 48(3):259–302, 1986.

[10] C. Biernacki, G. Celeux, and G. Govaert. Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate gaussian mixture models. *Computational Statistics & Data Analysis*, 41(3-4):561–575, 2003.

[11] T. O. Binford and T. S. Levitt. Evidential reasoning for object recognition. *IEEE Trans. Pattern Anal. Machine Intell.*, 25(7):837–851, 2003.

[12] C.A. Bouman and M. Shapiro. A multiscale random field model for bayesian image segmentation. *IEEE Trans. Image Processing*, 3(2), 1994.

[13] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Machine Intell.*, 23(11):1222–1239, 2001.

[14] P. Brodatz. *Textures: A Photographic Album*. New York: Dover, 1966.

[15] F. W. Campbell and J. G. Robson. Application of fourier analysis to the visibility of gratings. *J. Physiology*, 197:551–565, 1968.

[16] K. R. Castleman. *Digital Image Processing*. Prentice Hall, 1996.

[17] R. Chellappa. Two-dimensional discrete Gaussian Markov random field models for image processing. *Journal of the Institution of Electronics and Telecommunication Engineers*, 35(2):114–120, 1989.

[18] L. Cheng and T. Caelli. Unsupervised image segmentation: A bayesian approach. In *Proceedings, 16th International Conference on Vision Interface*, Halifax, Canada, June 2003.

[19] D. A. Clausi. *Texture Segmentation of SAR Sea Ice Imagery*. PhD thesis, Dept. of System Design Engineering, University of Waterloo, Waterloo, ON, Canada, 1996.

[20] D. A. Clausi. Comparison and fusion of co-occurrence, gabor, and MRF texture features for classification of SAR sea ice imagery. *Atmosphere and Oceans*, 39(4):183–194, 2001.

[21] D. A. Clausi. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of Remote Sensing*, 28(1):45–62, 2002.

[22] D. A. Clausi and H. Deng. Design-based texture feature fusion using gabor filters and co-occurrence probabilities. *IEEE Trans. Image Processing*, 14(7):925–936, 2005.

[23] D. A. Clausi and B. Yue. Comparing co-occurrence probabilities and Markov random fields for texture analysis of SAR sea ice imagery. *IEEE Trans. Geosci. Remote Sensing*, 42(1):215–228, 2004.

[24] V. Clement, G. Giraudon, S. Houzelle, and F. Sandakly. Interpretation of remotely sensed images in a context of multisensor fusion using a multispecialist architecture. *IEEE Trans. Geosci. Remote Sensing*, 31(4):779–791, 1993.

[25] J. M. Coggins and A. K. Jain. A spatial filtering approach to texture analysis. *Pattern Recognition Letters*, 3(3):195–203, 1985.

[26] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[27] Jr. D. C. Munson and R. L. Visentin. A signal processing view of strip-mapping synthetic aperture radar. *IEEE Trans. Acoust., Speech, Signal Processing*, 37:2131–2147, December 1989.

[28] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

[29] H. Deng and D. A. Clausi. Unsupervised image segmentation using a simple MRF model with a new implementation scheme. *Pattern Recognition*, 37(12):2323–2335, 2004.

[30] H. Deng and D. A. Clausi. Unsupervised segmentation of synthetic aperture radar sea ice imagery using a novel Markov random field model. *IEEE Trans. Geosci. Remote Sensing*, 43(3):528–538, 2005.

[31] H. Derin and H. Elliott. Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE Trans. Pattern Anal. Machine Intell.*, 9(1):39–55, 1987.

[32] X. Descombes, R. D. Morris, J. Zerubia, and M. Berthod. Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood. *IEEE Trans. Image Processing*, 8(7):954–963, 1999.

[33] B. A. Draper. From knowledge bases to Markov models to PCA. In *2003 Workshop on Computer Vision System Control Architectures*, Graz, Austria, March31 2003.

[34] B. A. Draper, R. T. Collins, J. Brolio, A. R. Hanson, and E. M. Riseman. The schema system. *International Journal of Computer Vision*, 2:209–250, 1989.

[35] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, Inc., 2001.

[36] F. D. Carsey et al. *Microwave Remote Sensing of Sea Ice*. American Geophysical Union, Washington DC, 1992.

[37] J. A. Feldman and Y. Yakimovsky. Decision theory and artificial intelligence I: Semantics-based region analyzer. *Artificial Intelligence*, 5(4):349–371, 1974.

[38] Canada Center for Remote Sensing web site. A tutorial of remote sensing. http://www.ccrs.nrcan.gc.ca/ccrs/eduref/tutorial/indexe.html.

[39] V. S. Frost, J. A. Stiles, K. S. Shanmugan, and J. C. Holtzman. A model for radar images and its application to adaptive digital filtering of multiplicative noise. *IEEE Trans. Pattern Anal. Machine Intell.*, 4:157–166, 1982.

[40] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, 6(6):721–741, 1984.

[41] R. C. Gonzalez and R. E. Woods. *Digital Image Processing (2nd Edition)*. Prentice Hall, 2002.

[42] D. Graham and A. N. Barrett. *Knowledge-Based Image Processing Systems*. Springer, 1997.

[43] M. Hansen and B. Yu. Model selection and minimum description length principle. *J. American Statistic Association*, 96:746–774, 2001.

[44] R. M. Haralick, K. Shanmugam, and I. Dinstein. Texture features for image classification. *IEEE Trans. Syst., Man, Cybern.*, 3(6):610–621, 1973.

[45] K. Haris, S. N. Efstratiadis, N. Maglaveras, and A. K. Katsaggelos. Hybrid image segmentation using watersheds and fast region merging. *IEEE Trans. Image Processing*, 7(12):1684–1699, 1998.

[46] D. Heckerman. Probabilistic interpretation for MYCIN's certainty factors. In L. N. Kanal and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 167–196. Elsevier, 1986.

[47] R. Huang, V. Pavlovic, and D. N. Metaxas. A graphical model framework for coupling MRFs and deformable models. In *Proceedings, 2004 Computer Vision and Pattern Recognition*, volume 2, pages 739–746, June27-July2 2004.

[48] S. S. Hwang, L. S. Davis, and T. Matsuyama. Hypothesis integration in image understanding systems. *Computer Vision, Graphics, and Image Processing*, 36(2-3):321–371, 1986.

[49] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice Hall, 1988.

[50] A. K. Jain, Y. Zhong, and S. Lakshmanan. Object matching using deformable templates. *IEEE Trans. Pattern Anal. Machine Intell.*, 18(3):267–278, 1996.

[51] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int'l J. Computer Vision*, 1(4):321–331, 1987.

[52] Z. Kato, M. Berthod, and J. Zerubia. A hierarchical Markov random field model and multitemperature annealing for parallel image classification. *Graphic Models and Image Processing*, 58(1):18–37, 1996.

[53] I. Y. Kim and H. S. Yang. Efficient image labelling based on markov random field and error backpropagation network. *Pattern Recognition*, 26(11):1695–1707, 1993.

[54] I. Y. Kim and H. S. Yang. An integration scheme for image segmentation and labelling based on Markov random field model. *IEEE Trans. Pattern Anal. Machine Intell.*, 18(1):69–73, 1996.

[55] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi Jr. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[56] H. Koch, K. Pakzad, and R. Tnjes. Knowledge based interpretation of aerial images and maps using a digital landscape model as partial interpretation. In *Workshop on Semantic Modeling of Topographic Information from Images and Maps*, Bonn, Germany, May21-23 1997.

[57] S. Krishnamachari and R. Chellappa. Multiresolution gauss-markov random field models for texture segmentation. *IEEE Trans. Image Processing*, 6(2):251–267, 1997.

[58] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel. Adaptive restoration of images with speckle. *IEEE Trans. Acoust., Speech, Signal Processing*, 35(3):373–383, 1987.

[59] K. S. Kumar and U. B. Desai. Joint segmentation and image interpretation. In *Proceedings, International Conference on Image Processing*, pages 853–856, September16-19 1996.

[60] V. P. Kumar and U. B. Desai. Image interpretation using Bayesian networks. *IEEE Trans. Pattern Anal. Machine Intell.*, 18(1):74–77, 1996.

[61] F. Kummert, H. Niemann, R. Prechtel, and G. Sagerer. Control and explanation in a signal understanding environment. *Signal Processing*, 32:111–145, 1993.

[62] J. Laferte, P. Perez, and F. Heitz. Discrete Markov image modeling and inference on the quadtree. *IEEE Trans. Image Processing*, 9(3), 2000.

[63] S. Lakshmanan and H. Derin. Simultaneous parameter estimation and segmentation of gibbs random field using simulated annealing. *IEEE Trans. Pattern Anal. Machine Intell.*, 11(8):799–813, 1989.

[64] K. I. Laws. Rapid texture identification. In *Proceedings, the SPIE Conference on Image Processing for Missile Guidance*, pages 376–380, 1980.

[65] V. F. Leavers. Survey: Which Hough transform? *CVGIP: Image Understanding*, 58(2):250–264, 1993.

[66] J. S. Lee. Speckle analysis and smoothing of SAR images. *Computer Graphics and Image Processing*, 17:24–32, 1981.

[67] S. Z. Li. *Markov Random Field Modeling in Image Analysis*. Springer, 2001.

[68] C. E. Liedtke, J. Bückner, O. Grau, S. Growe, and R. Tönjes. AIDA: A system for the knowledge based interpretation of remote sensing data. In *Proceedings, 3rd International Airborne Remote Sensing Conference & Exhibition*, volume 2, pages 313–320, Copenhagen, Denmark, July 1997.

[69] A. Lopes, R. Touzi, and E. Nezry. Adaptive speckle filters and scene heterogeneity. *IEEE Trans. Geosci. Remote Sensing*, 28(6):992–1000, 1990.

[70] W. Y. Ma and B. S. Manjunath. Edgeflow: A technique for boundary detection and image segmentation. *IEEE Trans. Image Processing*, 9(8), 2000.

[71] P. Maillard, D. A. Clausi, and H. Deng. Operational map-guided classification of SAR sea ice imagery. *IEEE Trans. Geosci. Remote Sensing*, 43(12):2940–2951, 2005.

[72] S. G. Mallat. A theory for multi-resolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Machine Intell.*, 11(7):674–693, 1989.

[73] T. Matsuyama. Expert systems for image processing - knowledge-based composition of image analysis processes. *Computer Vision, Graphics, and Image Processing*, 48(1):22–49, 1989.

[74] B. McCane and T. Caelli. Multi-scale adaptive segmentation using edge and region based attributes. In *Proceedings, 1st International Conference on Knowledge-Based Intelligent Electronic Systems*, volume 1, pages 72–80, May 1997.

[75] M. P. Mäkynen, A. T. Manninen, M. H. Similä, J. A. Karvonen, and M. T. Hallikainen. Incidence angle dependence of the statistical properties of C-band HH-polarization backscattering signatures of the baltic sea ice. *IEEE Trans. Geosci. Remote Sensing*, 40(12):2593–2605, 2002.

[76] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calcuations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

[77] M. Minsky. A framework for representing knowledge. In P. H. Winston, editor, *The Psychology of Computer Vision*, pages 211–277. McGraw-Hill, 1975.

[78] J. W. Modestino and J. Zhang. A Markov random field model based approach to image interpretation. *IEEE Trans. Pattern Anal. Machine Intell.*, 14(6):606–615, 1992.

[79] A. M. Nazif and M. D. Levine. Low level image segmentation: An expert system. *IEEE Trans. Pattern Anal. Machine Intell.*, 6(5):555–577, 1984.

[80] H. T. Nguyen, M. Worring, and R. van den Boomgaard. Watersnakes: Energy-driven watershed segmentation. *IEEE Trans. Pattern Anal. Machine Intell.*, 25(3), 2003.

[81] H. Niemann, G. F. Sagerer, S. Schroder, and F. Kummert. Ernest: A semantic network system for pattern understanding. *IEEE Trans. Pattern Anal. Machine Intell.*, 12(9):883–905, 1990.

[82] R. Nock and F. Nielsen. Statistical region merging. *IEEE Trans. Pattern Anal. Machine Intell.*, 26(11):1452–1458, 2004.

[83] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. Syst., Man, Cybern.*, 9:62–66, 1979.

[84] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[85] D. A. Pollen and S. F. Ronner. Visual cortical neurons as localized spatial frequency filters. *IEEE Trans. Syst., Man, Cybern.*, 13(5):907–916, 1983.

[86] M. R. Quillian. Semantic memory. In M. Minsky, editor, *Semantic Information Processing*, pages 354–402. M.I.T. Press, 1968.

[87] D. G. Albrecht R. L. De Valois and L. G. Thorell. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 22:545–559, 1982.

[88] G. Ramponi and C. Moloney. Smoothing speckled images using an adaptive rational operator. *IEEE Signal Processing Lett.*, 4(3), 1997.

[89] A. R. Rao and R. Jain. Knowledge representation and control in computer vision systems. *IEEE Expert*, 3(1):64–97, 1988.

[90] E. Rignot and R. Chellappa. Segmentation of polarimetric synthetic aperture radar data. *IEEE Trans. Image Processing*, 1(3):281–300, 1992.

[91] R. Rimey and C. Brown. Control of selective perception using Bayes nets and decision theory. *International Journal of Computer Vision*, 12:173–207, 1994.

[92] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 1995.

[93] R. Samadani. A finite mixture algorithm for finding proportions in SAR images. *IEEE Trans. Image Processing*, 4(8):1182–1186, 1995.

[94] A. Sarkar, M. K. Biswas, and K. M. S. Sharma. A simple unsupervised MRF model based image segmentation approach. *IEEE Transaction on Image Processing*, 9(5):801–812, 2000.

[95] Canadian Ice Service. http://ice-glaces.ec.gc.ca, Last accessed: Dec. 20, 2005.

[96] G. Shafer. *A Mathematical Theory of Evidence*. Princeton Univ. Press, Princeton, NJ, USA, 1976.

[97] M. E. Shokr. Evaluation of second-order texture parameters for sea ice classification from radar images. *Journal of Geophysical Research*, 96(C6):10625–10640, 1991.

[98] R. A. Shuchman, C. C. Wackerman, A. L. Maffett, R. G. Onstott, and L. L. Sutherland. The discrimination of sea ice types using SAR backscatter statistics. In *Geoscience and Remote Sensing Symposium*, pages 381–385, Vancour, B.C., 1989.

[99] L. K. Soh and C. Tsatsoulis. Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Trans. Geosci. Remote Sensing*, 37(2):780–795, 1999.

[100] L. K. Soh and C. Tsatsoulis. Unsupervised segmentation of ERS and radarsat sea ice images using multiresolution peak detection and aggregated population equalization. *Int. J. Remote Sensing*, 20(15&16):3087–3109, 1999.

[101] L. K. Soh, C. Tsatsoulis, D. Gineris, and C. Bertoia. ARKTOS: An intelligent system for SAR sea ice image classification. *IEEE Trans. Geosci. Remote Sensing*, 42(1):229–248, 2004.

[102] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision*. Thomson Course Technology, 1998.

[103] P. Suetens, P. Fua, and A. Hanson. Computational strategies for object recognition. *ACM Computing Surveys*, 24(1):5–62, 1992.

[104] J. M. Tenenbaum and H. G. Barrow. Experiment in interpretation-guided segmentation. *Artifical Intelligence*, 8(3):241–274, 1977.

[105] J. Ton, J. Sticklen, and A. K. Jain. Knowledge-based segmentation of landsat images. *IEEE Trans. Geosci. Remote Sensing*, 29(2):222–232, 1991.

[106] Z. Tu, X. Chen, A. L. Yuille, and S. C. Zhu. Image parsing: Unifying segmentation, detection, and recognition. *International Journal of Computer Vision*, 63(2):113–140, 2005.

[107] Z. Tu and S. C. Zhu. Image segmentation by data-driven Markov chain monte Carlo. *IEEE Trans. Pattern Anal. Machine Intell.*, 24(5):657–673, 2002.

[108] M. Tuceryan and A. K. Jain. *Handbook of Pattern Recognition and Computer Vision*, chapter Chapter 2: Texture Analysis. World Scientific, Singapore, 1993.

[109] F. T. Ulaby, R. K. Moore, and A. K. Fung. *Microwave Remote Sensing, Volumes 1, 2, and 3*. Massachusetts: Addison-Wesley, 1982.

[110] L. Vincent and P. Soille. Watershed in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Machine Intell.*, 13(6):583–598, 1991.

[111] World Meteorology Organization web site. http://www.wmo.ch/index-en.html, Last accessed: Dec. 17, 2005.

[112] Y. Weiss. Correctness of local probability propagation in graphical models with loops. *Neural Computation*, 12(1):1–41, 2000.

[113] L. P. Wesley. Evidential knowledge-based computer vision. *Optical Engineering*, pages 363–379, 1986.

[114] S. Wesolkowski and P. Fieguth. Hierarchical regions for image segmentation. In *Proceedings, International Conference on Image Analysis and Recognition*, Porto, Portugal, September29-October1 2004.

[115] R. Wilson and C. Li. A class of discrete multiresolution random fields and its application to image segmentation. *IEEE Trans. Pattern Anal. Machine Intell.*, 25(1):42–56, 2002.

[116] C. S. Won and H. Derin. Unsupervised segmentation of noisy and textured images using Markov random fields. *CVGIP: Graphical Models and Image Processing*, 54(4):308–328, 1992.

[117] W. A. Woods. What's important about knowledge representation. *IEEE Trans. Comput.*, 16(1):22–27, 1983.

[118] Q. Yu. Texture classification of SAR sea ice using the wavelet transform. Master's thesis, Dept. of Electrical and Computer Engineering, Memorial University of Newfoundland, St.John's, NF, Canada, May 2002.

[119] B. Yue. SAR sea ice recognition using texture methods. Master's thesis, Dept. of System Design Engineering, Univeristy of Waterloo, Waterloo, ON, Canada, 2001.

[120] J. Zhang. The mean field theory in EM procedures for Markov random fields. *IEEE Trans. Signal Processing*, 40(10):2570–2583, 1992.

[121] N. Zlatoff, B. Tellez, and A. Baskurt. Image understanding and scene models: A generic framework integrating domain knowledge and Gestalt theory. In *Proceedings, 2004 International Conference on Image Processing*, pages 2355–2358, Singapore, October24-27 2004.