Semantic sensitive satellite image retrieval

Li, Yikun

2008

Li, Y. (2008). Semantic sensitive satellite image retrieval. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/13593>

<https://doi.org/10.32657/10356/13593>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SEMANTIC SENSITIVE SATELLITE
IMAGE RETRIEVAL

SEMANTIC SENSITIVE SATELLITE IMAGE RETRIEVAL

LI YIKUN

LI YIKUN

SCHOOL OF COMPUTER ENGINEERING

2008

2008

SEMANTIC SENSITIVE SATELLITE IMAGE RETRIEVAL

Li Yikun

School of Computer Engineering

A thesis submitted to the Nanyang Technological University

in fulfillment of the requirement for the degree of
Doctor of Philosophy

2008

Acknowledgement

I would like to express my sincere gratitude to my supervisor Dr. Timo Rolf Bretschneider, who has provided valuable instructions and support during my research. He continuously encouraged me to improve my research work and guided me through the composition of this dissertation. Without his help, I would not be able to accomplish my research and dissertation.

I would like to thank the members of Satellite Data Processing (SDP) Group for their time and corporation. In particular, I would like to express my thanks to Bo Liu, Li Wu, Ken Yong Lee and Vladimir Buntilov.

I would like to thank Jasmine Qin, who provided valuable advices for writing and modifying this dissertation.

Finally, I would like to thank my parents. Without their unselfish affection and support, I would not have fulfilled my research and dissertation.

Table of Contents

Acknowledgement	1
Table of Contents	iv
Abstract	viii
List of Acronyms and Abbreviations	x
List of Tables	xii
List of Figures	xv
Chapter 1	1
Introduction	1
1.1 Motivation	1
1.2 Research Objectives	6
1.3 Organisation	8
Chapter 2	9
Literature Review	9
2.1 Content-Based Image Retrieval Systems	9
2.1.1 Selected Systems	9
2.1.2 Summary	12
2.2 Interactive Learning in Content-Based Image Retrieval Systems	13
2.2.1 Relevance Feedback	16
2.2.2 Long-Term Learning	17
2.2.3 Annotation Propagation	18
Chapter 3	21
Remote Sensing Imagery Retrieval Systems	21
3.1 Early Remote Sensing Image Retrieval Systems	21
3.2 Algorithm Development and Mining System	22
3.3 SPIRE system	23
3.4 SemQuery System	24
3.5 Diamond Eye System	25
3.6 Intelligent Satellite Information Mining System	26
3.7 VisiMine System	27
3.8 Knowledge-Driven Information Mining System	28
3.9 Summary	30

Chapter 4.....	32
Framework of the Prototype System.....	32
Chapter 5	36
Feature Extraction and Image Segmentation	36
5.1 Image Characteristic.....	36
5.2 Spectral Feature Extraction	37
5.3 Texture Features Extraction	39
5.4 Image Segmentation	40
Chapter 6	42
Codebook Generation and Image Encoding	42
6.1 Codebook Generation.....	42
6.1.1 Vector Quantisation	43
6.1.2 Tree-Structured Vector Quantisation.....	45
6.2 Image Encoding and Code Representation of Image Regions	46
6.3 Results and Assessment.....	47
Chapter 7	51
Semantic Concept Inference	51
7.1 Semantic Concept Inference via Simple Bayesian Network.....	52
7.2 Interactive Learning via Simple Bayesian Network.....	54
7.3 Semantic Concept Inference via Context-Sensitive Bayesian Network.....	55
7.4 Code Co-occurrence Matrix	59
7.5 Semantic Score.....	60
Chapter 8.....	62
Relevance Feedback.....	62
8.1 Relevance Feedback via the Context-Sensitive Bayesian Network.....	64
8.2 Confusion Code.....	65
8.3 Pre-filtering Function	66
8.4 Composite Relevance Feedback Approach	67
8.5 Results and Assessment.....	68
8.5.1 BNRF and PBNRF Approaches.....	69
8.5.2 Composite Relevance Feedback Approach	72
8.5.3 Heterogeneous Test.....	74
8.5.4 Verification Experiments	78
Chapter 9.....	81

Long-Term Learning.....	81
9.1 Long-term Learning via Context-Sensitive Bayesian Network	81
9.2 Associative Long-Term Learning.....	82
9.3 Results and Assessment.....	84
9.3.1 Comparison of SemQuery and Long-Term Learning Approach	84
9.3.2 Simple Bayesian Network and Context-Sensitive Bayesian Network .	87
9.3.3 Significance of Adjacent Regions.....	89
9.3.4 Learning Ability.....	90
9.3.5 Codebook Size	92
9.3.6 Associative Long-Term Learning Performance.....	94
9.3.7 Thresholding of Semantic Scores	95
9.3.8 Retrieval Example.....	97
9.3.9 Verification Experiments	98
Chapter 10	101
Stepwise Remote Sensing Image Retrieval Scheme.....	101
10.1 Remote Sensing Image Retrieval	101
10.2 Remote Sensing Image Indexing.....	102
10.3 Stepwise Image Retrieval Scheme	102
10.4 Experimental Results.....	103
10.4.1 Computational Cost of the Pre-Selection Step.....	103
10.4.2 Retrieval Performance of the Stepwise Scheme	105
10.4.3 Retrieval Example.....	107
10.4.4 Verification Experiments	109
Chapter 11	110
Further Application of the Developed Framework.....	110
11.1 Application Background.....	110
11.2 VirtSat System.....	111
11.3 Satellite Image Evaluation and Verification.....	115
11.4 Results and Assessment.....	116
Chapter 12	120
Conclusions and Future Directions.....	120
12.1 Conclusions	120
12.2 Contribution.....	122
12.3 Future Directions	123
References.....	125
Publications.....	133

Abstract

In recent years, numerous remote sensing platforms for Earth observation have been developed and together acquire several terabytes of image data per day. However, due to this data volume, straightforward access to the data has become increasingly complex. Most common database systems retrieve the satellite images based on their world-oriented information. Queries not directly related to this type of information, for example, the search for a scene that possesses a similar ground cover characteristic as a query scene, cannot be processed. Therefore, a variety of content-based image retrieval techniques were developed to retrieve images based on abstract automatically extracted features. Applied to remote sensing databases, varying success was reported. The main difficulty is that users think in terms of high-level semantics, which cannot be directly mapped onto extracted low-level features / signals. The result is a gap between the signal classes on the one side, that is, pixels, regions or images that are clustered in the feature space, and semantic concepts on the other side. Although various relevance feedback approaches have been proposed to bridge this gap, most of them can only work well under the assumption that regions belonging to the same or similar signal classes are very likely to share the same semantic concepts as well. Often this assumption is invalid, since there are rarely one-to-one relations among signal classes and semantic concepts. On the one hand, a signal class may be related to multiple semantic concepts under different circumstances. On the other hand, multiple signal classes may be related to the same semantic concept. An example is the change in observed characteristics of the same scene due to changes in sun inclination and season.

In order to bridge the semantic gap, this thesis proposes a specialised framework for content-based retrieval of remotely-sensed optical imagery based on automatic semantic inference using knowledge accumulated from users' feedbacks.

The proposed framework uses an unsupervised vector quantisation algorithm to globally classify the low-level feature vectors of all image regions in the database into a set of codes that forms a so-called codebook. Then each image is encoded based on the code dictionary. The code of a region together with each individual code of its adjacent regions forms a code pair, which reflects the characteristic of the region and its spatial context.

A context-sensitive Bayesian network is proposed to establish multiple-to-multiple linkages to bridge the gap between the low-level code pairs and the semantic concepts. In order to simplify the calculation, a semantic score function set is derived as a special case of the context-sensitive Bayesian network. Each semantic score function is used to measure the extent to which a remote sensing image is related to a semantic concept based on a code co-occurrence matrix.

Three novel relevance feedback approaches are proposed to train the semantic score functions based on user-supplied examples. These approaches have individual advantages in improving the retrieval performance from different perspectives and have shown to be very robust. Moreover, they are able to achieve real-time responses in large remote sensing databases.

A long-term learning approach is suggested to train the semantic score functions based on the history of users' feedbacks with heterogeneous signal characteristics. Moreover, with the objective to improve the performance of the semantic inference, a derivative of the proposed long-term approach also uses a special pre-filtering function to synthesis long-term knowledge from different semantic concepts.

At the retrieval stage, a stepwise image retrieval scheme is adopted to balance effectiveness and efficiency. Firstly, the trained semantic score functions are used to retrieve a candidate image set, which relates to the user's semantic concepts. Next, a computationally more expensive similarity measurement is applied to assess the similarities between the regions of the query image and the regions of the candidate images. The proposed stepwise scheme is able to enhance the retrieval performance as well as to reduce the retrieval time.

Various experiments using actual satellite imagery were carried out in order to highlight the performance characteristics. In particular a comparison with a well-established retrieval system was conducted and showed the advantages of including semantic understanding in the retrieval process.

List of Acronyms and Abbreviations

2D MHMM	Two Dimensional Multiresolution Hidden Markov Model
ALTL	Associative Long-Term Learning
ADaM	Algorithm Development and Mining
BNRF	Bayesian Network Relevance Feedback
BPM	Bayes Point Machine
CBIR	Content-Based Image Retrieval
CBSA	Content-Based Soft Annotation
CRF	Composite Relevance Feedback
CSBN	Context-Sensitive Bayesian Network
FRIP	Finding Region In the Picture
GLA	Generalised Lloyd Algorithm
GMRF	Gibbs-Markov Random Field
GRF	Gibbs Random Field
GUI	Graphical User Interface
IRM	Integrated Region Matching
ISIMS	Intelligent Satellite Information Mining System
ISS	Image Server System
KIM	Knowledge-Driven Information
LTL	Long-Term Learning
MARS	Multimedia Analysis and Retrieval System
ORDBMS	Object-Relational Database Management System
PBNRF	Pre-filtering Bayesian Network Relevance Feedback
PNNA	Pairwise Nearest Neighbour Algorithm
QBIC	Query by Image Content
RBIR	Region-based Image Retrieval
RBRF	Region-Based Relevance Feedback
RF	Relevance Feedback
SBN	Simple Bayesian Network
SDE	Spatial Data Engine
SIMPLIcity	Semantics-Sensitive Integrated Matching for Picturing Libraries
SVM	Support Vector Machine
UFM	Unified Feature Matching

List of Tables

Table I: Summary Table of the Reviewed Systems	31
Table II: Basic characteristic of IKONOS bands.....	37
Table III: Number of codes having non-zero code frequencies with respect to two semantic concepts	49
Table IV: Precision and recall rates using BNRF	69
Table V: Precision and recall rates using PBNRF	70
Table VI: Retrieval time comparison between BNRF and PBNRF	72
Table VII: Precisions of top-100 images using the CRF and the RBRF.	73
Table VIII: Retrieval time comparison between the CRF and the RBRF feedback	74
Table IX: Precisions of BNRF /RBRF approaches for different scenes.....	77
Table X: Precisions of BNRF/RBRF with positive images from different scenes	78
Table XI: Precision and recall rates using BNRF	79
Table XII: Precision and recall rates using PBNRF	80
Table XIII: Mean and standard deviations for precision and recall rates over five separate training and retrieval sessions for semantic classification based on spectral features	86
Table XIV: Precision and recall rates over five separate training and retrieval sessions for semantic classification based on spectral and texture features ..	87
Table XV: Precisions of the semantic inference based on three levels of the tree-structured codebook using simple and context-sensitive Bayesian networks	88
Table XVI: Recall rates of the semantic inference based on three levels of the tree-structured codebook using simple and context sensitive Bayesian networks	89
Table XVII: Precision and recall rates of the LTL approach based on three levels of the tree-structured codebook	92
Table XVIII: Numbers of positive images and positive regions	94
Table XIX: Comparison of precision and recall rates over semantic inference with and without pre-filtering for semantic classification based on spectral features	95
Table XX: Precision and recall rates using different threshold values ν	96
Table XXI: Precision and recall rates over five separate training and retrieval sessions for semantic classification based on spectral and texture features ..	99
Table XXII: Precisions and recall rates of semantic inference based on spectral and texture features using simple and context-sensitive Bayesian networks	99
Table XXIII: Precisions and recall rates of semantic inference based on three levels of the tree-structured codebook using simple and context-sensitive Bayesian networks	100
Table XXIV: The number of pre-selected images for different semantic concepts in the pre-selection stage.....	104
Table XXV: Precision comparison (top-100 images) for IRM with and without pre-selection stages	106

Table XXVI: Precision comparison (top-100 images) for IRM with and without pre-selection stages	109
Table XXVII: Categories of non-satellite images	116

List of Figures

Figure 1: Scene of Hong Kong (Ikonos, © Space Imaging)	2
Figure 2: Scene of flooding in New Orleans (Ikonos, © Space Imaging)	3
Figure 3: Diamond Eye system architecture according to [14]	26
Figure 4: Architecture of the Intelligent Satellite Information Mining System according to [26]	27
Figure 5: VisiMine system architecture [61]	28
Figure 6: Offline component of KIM system according to [32]	29
Figure 7: Online component of KIM system according to [32]	30
Figure 8: First stage and second stage of the semantic inference and retrieval ...	33
Figure 9: Third stage of the semantic inference and retrieval	34
Figure 10: Stepwise image retrieval scheme	35
Figure 11: Procedure diagram for the equalised quantisation approach	37
Figure 12: Equalised quantisation of remote sensing image: (a) Image before quantisation, (b) Image after quantisation (Ikonos, © Space Imaging 2003)	38
Figure 13: Hierarchical wavelet decomposition: (a) Original remote sensing image, (b) Three-level wavelet decomposition (Ikonos, © Space Imaging, 2003)	40
Figure 14: Tree-structured codebook with three levels	45
Figure 15: Image encoding procedure	47
Figure 16: Spectral code frequencies of the experimental remote sensing database	48
Figure 17: Spectral code frequency of the remotely sensed images that are related to semantic concept ‘bare soil’	48
Figure 18: Spectral code distribution of the remotely sensed images that are related to semantic concept ‘rock’	49
Figure 19: Simple Bayesian network	53
Figure 20: Context-sensitive Bayesian network	57
Figure 21: Performance of BNRF and PBNRF: (a) Average precision, (b) Average number of correct images, (c) Precision vs. Number of correct images	71
Figure 22: Average retrieval time comparison of BNRF and PBNRF	72
Figure 23: Retrieval performance of CRF and RBRF: (a) Average precision, (b) Average retrieval time	74
Figure 24: Water body in different scenes	76
Figure 25: Precisions of the BNRF/.RBRF approaches with positive images from different sources	77
Figure 26: Retrieval performance of CSBN and SBN at three levels of the tree- structured codebook: (a) Average precision, (b) Average recall	89
Figure 27: Retrieval performance for areas partially and totally covered by forest with respect to the size of the training image set: (a) Precision, (b) Recall, (c) Precision vs. recall	91
Figure 28: Retrieval of forest areas based on different codebook levels in the LTL approach in comparison with SemQuery: (a) Precision, (b) Recall, (c) Precision vs. recall	93
Figure 29: Average precision of LTL and ALTL with various threshold values	95

Figure 30: Precision vs. recall rate for different threshold values v	97
Figure 31: Retrieval example based on the query term ‘urban area’	98
Figure 32: Average retrieval precision of IRM with and without pre-selection stages.....	106
Figure 33: Average retrieval time of IRM with and without pre-selection stages	107
Figure 34: Retrieval example using IRM.....	108
Figure 35: Retrieval example using the stepwise retrieval scheme	108
Figure 36: VirtSat system architecture	112
Figure 37: Flowchart of a crawler thread parsing a web page	114
Figure 38: Number of correctly classified satellite images based on different codebook sizes and different training set sizes	117
Figure 39: Precision of the classified satellite images with respect to the codebook size and the size of the image training set	118
Figure 40: Recall rate of the classified satellite images with respect to the codebook size and the size of the image training set	118
Figure 41: Precisions and recall rates based on different levels of the codebook: (a) 100 codes, (b) 200 codes, (c) 300 codes.....	119

Chapter 1

Introduction

1.1 Motivation

Research on content-based image retrieval (CBIR) developed significantly during the last decade. The impetus behind the accelerated development of CBIR techniques is that the amount of image data is continuously increasing. The images are widely applied in different areas, such as meteorology, astronomy, medicine, and remote sensing, museums, forensics, etc. Facing the enormous amount of information contained in today's image databases, users experience great difficulties in exploring and retrieving relevant image data by means of world-oriented information, pre-entered keywords and / or other accompanied text information. Although text information retrieval techniques have been fully developed, there are three basic problems [34] that prevent these techniques from being straightforwardly applied in the domain of image retrieval. Firstly, low-level visual information of images cannot be precisely described by characterising the colour, texture, shape, etc. Secondly, manually annotating each stored image is an almost infeasible solution due to time and cost constraints especially for large image databases. Thirdly, text information retrieval techniques often rely on a particular vocabulary, which the users may not be familiar with.

The listed problems gave major impetus to the development of the early CBIR systems, including QBIC (Query by Image Content) [34], MARS (Multimedia Analysis and Retrieval System) [77], PicHunter [25], VisualSEEK [99], and PicToSeek [37]. The general concept of these systems is to describe the image content by a-priori automatically extracted descriptors, called feature vectors, which are compared with a correspondingly extracted vector from the query image by similarity measurements. These CBIR systems successfully overcame the above highlighted problems: Firstly, the extracted features are capable of describing low-level visual information contained in the images more accurately and are especially useful in case of queries based on visual information. Secondly, the low-level features are

automatically extracted from the images without any manual work, thus, CBIR techniques are cheaper and faster alternatives compared to manual annotation based solutions. Thirdly, CBIR systems have no special requirements with regard to the user's background, knowledge, language, vocabulary, etc.

Figure 1: Scene of Hong Kong (Ikonos, © Space Imaging)

Since the CBIR techniques are successfully applied in generic image retrieval, they are used to solve the problems of remote sensing image retrieval. The major problem in remote sensing image retrieval is the amount of remotely sensed imagery. In recent years, numerous remote sensing platforms for Earth observation have been developed and together acquire several terabytes of image data per day. Due to this data volume, the straightforward access is becoming increasingly complex. Most commonly databases retrieve the satellite images based on their world-oriented information, such as the geographical location, the spectral and spatial characteristics of the imaging instrument, and the acquisition date. Queries not directly related to this type of information, that is, the search for a scene that possesses a similar ground cover characteristic as a query scene, cannot be processed. Possible scenarios are:

- Ecologist: A certain region reveals symptoms of salification and a corresponding satellite image of the area was purchased. It is of interest to find other regions which suffered from the same phenomenon and which were successfully recovered or preserved, respectively. Thus, the applied strategies in these regions could help to develop effective counteractions for the specific case under investigation.

- Sociologist: Urban growth has an impact on the social behaviour of a city's inhabitants, e.g. Hong Kong is one of the most densely populated cities in the world like can be seen in Figure 1. In order to avoid an undesirable development, it is advisable to study parallel and anti-parallel developments in other cities that have the similar appearance in the satellite images. In this context satellite imagery is a powerful tool for comparing spatial changes in urban areas on a large scale.
- Geography teacher: Figure 2 shows a recently published scene of the flooding in New Orleans. It is an example for the influence of natural disasters on the economy of a society. Students can improve their understanding if they are able to put areas that endured similar incidences.

Figure 2: Scene of flooding in New Orleans (Ikonos, © Space Imaging)

For the purpose of solving these types of content-related queries, one solution is to manually annotate the remote sensing imagery with a set of keywords, which refer to the scene content. However, in the real world this is an almost infeasible solution, due to time and cost constraints caused by the amount of data.

Previously, a variety of generic CBIR techniques were applied to remote sensing databases and a varying degree of success was reported. It was observed that these techniques cannot always meet the user's expectation, for example, if the retrieval intentions are more inclined to the semantic aspects of the imagery. The main difficulty is that the user has a high-level semantic understanding of the desired query

result, that is, “forest surrounded by agricultural areas”, which cannot be directly mapped onto the extracted low-level features. From the remote sensing perspective, the difficulty is caused by the semantic gap between signal classes on one side, that is, pixels, regions or images that are clustered in the feature space, and semantic concepts on the other side.

One reason for of the semantic gap is the utilisation of global features, which cannot comprehensively reflect the important properties of individual objects and ground cover types. Thus, region-based image retrieval (RBIR) systems use local features to represent remote sensing imagery at the object level and reduce the width of the semantic gap to some extent. However, they fail to bridge it completely. One reason is that remote sensing images are complex and contain a huge variety of information, which cannot be described sufficiently and precisely by the generally automatically extracted feature vectors. Therefore, at least in the near future, remote sensing image retrieval systems will need interactive communications with the users to collect more subjective information which can be incorporated in the query processing to improve the image retrieval results.

The interactive communication ability can be provided by relevance feedback (RF) techniques, which are very popular in some CBIR systems and some remote sensing image retrieval systems. The RF techniques are able to interactively capture the semantic concepts of the user’s query. They free the user from the direct modification of internal parameters in order to optimise the retrieval precision. Instead, the user only evaluates the retrieval result iteratively. Then, the techniques gradually adjust the query parameters to improve the retrieval accuracy. The success of the RF approaches can be attributed to their real-time response and their sensitivity to the end user’s subjectivity. However, four problems of the RF techniques [42] limit their application in remote sensing image retrieval. Firstly, most RF approaches cannot completely bridge the semantic gap since they work under the assumption that feedback examples belonging to the same or similar signal classes are very likely to share the same semantic concept as well. Often, this assumption is not valid since there are rarely one-to-one matches between signal classes and semantic concepts. On the one hand, a signal classes may be related to multiple semantic concepts under different circumstances. From the regional perspective, Human beings infer the semantic concept of a region based on not only the region itself but also its contexts i.e. the adjacent regions. For

example, a bright / saturated image region may be a part of a technical installation on the rooftop of a skyscraper, a cloud, or a rocky area in the mountains. On the other hand, multiple signal classes may be related to the same semantic concept, which is also known as sensory gap [97]. An example is the variation in observed characteristics of the same scene due to the changes in sun inclination and season. It should be acknowledged that such variation contains rich information that can be used to discriminate ground cover classes. However, the utilisation of such variation is out of the scope of this dissertation, and therefore is not extended. Secondly, the performances of RF techniques are often limited by the asymmetrical and sparse feedback examples. The sparseness is caused by the fact that the user often has limited patience with providing sufficient feedback examples while the asymmetry refers to the fact that negative examples often outnumber positive examples. Thirdly, the real-time response of most RF approaches cannot be guaranteed especially for large remote sensing databases. Fourthly, most RF techniques lack the capability to memorise, that is, the knowledge obtained during a query session is discarded at the end of a query session.

In summary, it is necessary to design novel RF techniques that are able to capture the multiple-to-multiple matches between signal classes and semantic concepts based on the asymmetrical and sparse feedback examples. At the same time, these RF approaches have to achieve real-time response in large databases.

The short-term learning behaviour of the RF approaches has two disadvantages. Firstly, a user has to repeat the same tedious feedback iterations in order to obtain an identical result achieved earlier [5]. Secondly, if users share similar retrieval interest, then individual experience cannot be utilised across user queries. Facing these problems, various long-term learning (LTL) approaches were proposed to add memory capability to CBIR systems. These approaches retrieve images based on the long-term knowledge learned from previous users' feedbacks. A typical LTL approach infers semantic similarities of the images solely based on the historical logs of the users' feedbacks, which are more reliable to represent the semantic relationships than the low-level features. However, it is very difficult to collect sufficient logs that cover all the images in the databases. Therefore, most LTL approaches have limited applications in remote sensing databases due to the rapidly increasing data volume. Moreover, compared to the RF approaches, LTL approaches have to deal with more feedback examples. These examples come from different users

and, thus, have very inhomogeneous signal characteristics. Therefore, an effective and efficient LTL approach is needed to learn the knowledge from a large number of feedback examples with inhomogeneous signal characteristics.

For large remote sensing databases, it is necessary to retrieve and index the remote sensing images in an effective and efficient way. Traditionally, most CBIR systems index the contained images mainly based on the extracted low-level features. However, detailed descriptions of the remote sensing images require high-dimensional feature vectors, which increase computational costs [9]. Thus, it is almost infeasible to perform a linear search within large remote sensing databases with real-time response. In fact, the indexing problem is even more severe in RBIR systems since the number of regions is far larger than the number of images. Moreover, region-based similarity measurements calculate the similarity between two images based on the individual region-to-region similarity. Hence, the corresponding computational cost is significantly larger than for global similarity measurements [57]. Although various indexing approaches have been proposed to reduce the number of images to be searched, they are automatically performed merely based on the low-level features and incorporate little or no semantic information. Thus, the retrieved images are similar in the low-level feature space but often semantically irrelevant. Therefore, a stepwise image retrieval scheme is needed to narrow the search scope according to the user's query intention. As a result the image retrieval can be carried out more effectively and efficiently.

1.2 Research Objectives

The goal of this research is to develop a specialised framework for content-based retrieval of remotely sensed optical imagery based on automatic semantic inference using knowledge gathered from users' feedbacks. In order to deal with the complexity of remote sensing images, the framework has to infer the semantic concepts from local features of segmented images. However, large volumes of local features significantly limit the image retrieval speed. Thus, an appropriate solution is required to compress the number of low-level features into compact cross-image representations, that is, signal classes for practical applications in remote sensing databases.

Human beings are able to interpret an image region based on its spectral and textural characteristics as well as its context, i.e. adjacent regions. Therefore, the research needs to develop an appropriate statistical model, which is able to approximate the context-sensitive semantic inference ability of human beings. Moreover, the semantic concepts of remote sensing images cannot be described sufficiently and precisely by the generally automatically extracted features. Thus, the semantic inference cannot accurately infer the semantic concepts merely based on the extracted features. Therefore, the framework needs to incorporate interactive communication with the users to collect more subjective information which can be used to improve the image retrieval results. In order to interactively learn from the user, RF and LTL approaches are required to cope with short-term learning and long-term learning, respectively.

RF approaches are able to capture the user's query intentions based on the user-supplied examples. Since the user-supplied examples may contain ground cover types with very inhomogeneous signal characteristics, the intended RF approach has to be robust. In order to reduce the inaccuracy caused by the compact representations of images, the RF approaches need to improve the retrieval precision based on the compact representations and the low-level features. Finally, real-time response even for large remote sensing databases is desirable.

In order to enable the framework to possess memory capacity, a LTL approach has to be designed to learn the subjective information from the accumulated users' feedbacks. Note that the feedback examples are supplied from different users and often have heterogeneous signal characteristics. Thus, the LTL approach has to be robust to the heterogeneous signal characteristics. In order to comprehensively exploit the semantic information contained in the users' feedbacks, the long-term approach has to include the long-term knowledge from training sample of different semantic concepts in order to improve the performance of the semantic inference.

For large remote sensing databases, it is necessary to retrieve and index remote sensing images in an effective and efficient way. In order to improve the retrieval precision and reduce the response time, a stepwise image retrieval scheme has to be developed in order to pre-select a group of the remote sensing images related to the user's query intention. Then, a more computational expensive region-based similarity measurement can be used to reorder the pre-selected images according to the query

image. Therefore, the retrieval time is significantly reduced and the final top-ranking images are visually similar and semantically relevant.

1.3 Organisation

The remaining part of this thesis is organised as follows: Chapter 2 provides a literature review with respect to CBIR systems and related work in interactive learning. Chapter 3 reviews seven remote sensing image retrieval systems. In order to bridge the semantic gap between signal classes and semantic concepts, an extended framework is proposed in Chapter 4, whereby the individual parts are illustrated in details in the following chapters. Firstly, Chapter 5 describes the processes of image segmentation, low-level feature extraction and vector representation. Secondly, these feature vectors are classified into signal classes by an appropriate vector quantisation technique in Chapter 6. In Chapter 7, a context-sensitive Bayesian network is proposed to link signal classes with semantic concepts. In order to simplify the calculation, a set of semantic score functions is derived as a special case of the context-sensitive Bayesian network. With the objective to interactively train the semantic score function, three RF approaches are proposed in Chapter 8. Since the RF approaches lack memory capability, an LTL approach is presented in Chapter 9. In Chapter 10, the trained semantic score functions are used to reduce the retrieval time as well as to enhance the retrieval precision for a more expensive similarity measurement. In Chapter 11, the developed framework is applied in the VirtSat system, which is a virtual satellite working in the World Wide Web environment. Eventually, Chapter 12 provides a summary of the conducted research work and draws conclusions. Future research proposals are discussed.

Chapter 2

Literature Review

This chapter reviews the related work of CBIR systems, interactive learning approaches and remote sensing image retrieval systems.

2.1 Content-Based Image Retrieval Systems

2.1.1 Selected Systems

The IBM QBIC system [34] is a commercial system, which supports queries on large image and video databases. It retrieves images/videos based on the content information such as colour, texture, shape, location. It supports multiple types of queries: Sample images, sketches and drawings, selected colour and texture patterns. The functionalities of the system are supported by two components, namely, the database population and the database query. During the database population, features are extracted and stored in a database for the similarity comparison. In order to support queries at object level, the system is able to extract features from the detected objects within the images, whereby the objects are manually, semi-automatically, or automatically identified. During the database query, the different kinds of distance functions are used to measure the similarities between the automatically extracted features.

As an early CBIR system, the aim of the MARS system [77] is to apply existing text information retrieval techniques in content-based retrieval of multimedia databases. The system represents an image by multiple types of low-level features, such as colour, texture, shape and layout. In order to make use of the text information retrieval techniques, the system extends a Boolean retrieval model to retrieve images. For example, a user is able to specify a query by a Boolean expression to retrieve images that are similar to an image A in colour and an image B in texture. Based on the basic Boolean retrieval model, the system further develops fuzzy Boolean retrieval and probabilistic Boolean retrieval to access the image databases. In order to improve the

retrieval performance, the system designs a RF approach [87] to capture the user's query intention.

VisualSeek [98] is a highly functional prototype image retrieval system, which supports content-based queries based on automatically extracted features. VisualSeek is a RBIR system that can automatically detects and indexes salient colour regions of the stored images. The system automatically detects the regions using colour set back-projection, whereby the colour sets are able to compactly represent colour histograms. The dominant characteristic of the system is that the user may formulate his/her queries by diagramming spatial arrangements of the colour regions. Then, the system retrieves the images according to the colour sets and the spatial arrangements. Thereby, the system is able to support a wide variety of complex joint colour/spatial queries.

The PicToSeek [37] image search system is designed for searching images on the World Wide Web environment. It uses colour features and shape features to automatically collects, indexes and catalogues the stored images. The system supports interactive query formulation to retrieve partially identical images. It allows the users to choose the desired classes of features and measures the importance of each individual feature by the weight. The system also develops a query-point movement RF approach to improve the retrieval performance.

Netra [72] is a prototype RBIR system, which utilises colour texture, shape and spatial information of detected regions to acquire similar regions in the database. It incorporates a robust segmentation algorithm to automatically detect the regions from the stored images. Once an image is segmented into a set of regions at the ingestion time, features are automatically extracted. Since the features of the regions distribute sparsely in the colour space, a subset of the features can be used to represent the regions without affecting the perceptual quality. Therefore, Netra utilised the unsupervised vector quantisation algorithm to compress colour, texture and shape features into compact representations, which can be used to index large image databases quickly. In order to retrieve images based on heterogeneous features, the system intersects the retrieval results of each individual features and sorts the final result using a weighted similarity measurement. Moreover, the system is able to cope with queries that incorporate the spatial arrangements of the regions and access the images that have multiple complex objects.

Blobworld [15] is an online RBIR system, which retrieves images based on coherent regions. The system automatically extracts colour and textural features from the segmented regions (blobs). It retrieves images based on regional features, which represent images at object level, and achieves high performance. In order to support large-scale retrieval, the system uses a tree to index the regional features. Furthermore, it reduces the computational cost of high-dimensional indexing by projecting the high-dimensional feature vectors into the low-dimensional feature space.

SIMPLIcity [110] (Semantics-sensitive Integrated Matching for Picture LIBraries) is an experimental RBIR system, which uses the integrated region matching (IRM) algorithm to compute the similarities between two segmented images. The SIMPLIcity system uses a wavelet-based approach to segment each image into a set of regions and represents the segmented regions by colour, texture, shape and location. The accuracy of the segmentation is not crucial for the SIMPLIcity system since the IRM algorithm is designed to tolerate inaccurate segmentation. In order to capture the semantic concepts, the system also uses the IRM algorithm to categorise the stored images into different semantic classes and thus narrow down the search range to a subset of the database for the fast retrieval. The unified feature matching (UFM) is the other similarity measurement in the SIMPLIcity system. It measures the similarity based on a set of fuzzy features, which naturally characterise the gradual transition between regions. Thus, the UFM algorithm is able to incorporate the uncertainties of the segmentation algorithm into the final similarities. Therefore, the UFM algorithm is more robust than the IRM algorithm.

Finding Region in the Picture (FRIP) [60] is a RBIR tool, which is able to tolerate region scaling, rotation and translation in the image retrieval. The system uses adaptive circular filters to accurately segment the stored images. Then, colour, texture, scale, shape, and location features are extracted from the detected regions by an optimal approach, which reduces the computational complexity without affecting the retrieval accuracy. The FRIP system designs a sophisticated feature matching technique to measure the similarities between two images. In order to increase the retrieval efficiency of large image database, a stepwise scheme is applied to add an inexpensive quick-and-dirty search step before the feature matching. Since the added step coarsely selects a small number of candidate images, the computational cost is significantly reduced.

2.1.2 Summary

All the reviewed CBIR systems use multiple types of features to represent the image contents. Among these systems, the most commonly used features are colour, texture and shape. Since various features are extracted using different approaches, these features may need different similarity measurements. Some systems use weighted linear formulas to integrate the similarity measurements of the features. However, the features may not be linearly related. Moreover, users may have difficulties in assigning the weights. Therefore, some systems retrieve images based on each individual features and merge the retrieved results.

Most of the reviewed CBIR systems are RBIR systems, which represent images by local features. The local features are extracted from the automatically detected regions and can reflect the important properties of individual objects within the images. Thus, they are very useful if the users' query intentions are at object level. However, inaccurate segmentation may deteriorate the retrieval performance. In order to solve the problem, some systems use a robust segmentation algorithm to improve the segmentation accuracy while other systems design the region-based similarity measurements that are robust to inaccurate segmentation. The other problem of the RBIR systems is that the region-based similarity measurements often have high computational complexity. Thus, the systems require long time to accomplish the linear search within a large database. This problem can be solved by compressing the local features into compact representations or reducing the dimensionality of the feature vectors. It can also be solved by appropriate indexing techniques or inexpensive quick-and-dirty pre-selection approaches. However, these solutions may reduce the system accuracy.

Among the reviewed CBIR systems, only MARS and PickToSeek systems use the RF approaches to capture the users' query intentions. Therefore, both systems are able to interactively learn from the users to bridge the semantic gap between low-level features and high-level semantic concepts. The work related to interactive learning is reviewed in the next subsection.

2.2 Interactive Learning in Content-Based Image Retrieval Systems

A variety of content-based image retrieval techniques were developed to retrieve images based on the automatically extracted features. The main difficulty is that the user has a high-level semantic understanding of the desired query result, which can hardly be directly mapped onto the extracted low-level features. The result is a semantic gap between the low-level features and high-level semantic concepts. The semantic gap is reflected by two aspects:

1. *Inconsistency between low-level features and high-level semantic concepts* [63]: Low-level feature space cannot accurately describe the semantic concept. From the semantic perspective, images that locate closely in the feature space may be semantically irrelevant while images that are scattered in the feature space may be semantically relevant.
2. *The subjectivity of the user perception* [88]: Different users may make different similarity judgement about the same images. Even the same user may make different similarity judgement about the same images under different circumstances.

Given the above two aspects, it is a misleading to improve the retrieval performance merely based on the automatically extracted features, which cannot sufficiently and precisely describe the semantic concepts. Therefore, CBIR systems need interactive communication with the users to collect more subjective information that can be used to improve the image retrieval results.

Motivated by the semantic gap, many interactive learning approaches have been developed. Most of them can be categorised into three branches: RF, LTL and image annotation propagation (IAP).

The concept of RF was firstly invented in the domain of text information retrieval [92]. Then, it was transformed into CBIR during early and mid 1990's [113]. RF approaches are very popular in CBIR systems since they are able to interactively capture the semantic concepts of the users' queries. The approaches free the user from the direct modification of internal parameters in order to optimise the retrieval precision. Instead, the user only evaluates the retrieval result iteratively. Then, the system gradually adjusts the query parameter to improve the retrieval accuracy. This

can be achieved in three different ways: query point movement [87], re-weighting of different feature dimensions [88] and updating of the probability distribution [25]. The success of RF can be attributed to its real-time response and its sensibility to the end-user's subjectivity. However, the three drawbacks [42] of RF limited their potential.

- A. RF cannot completely solve the inconsistency between low-level features and high-level semantic concepts. Derived from the domain of text information retrieval, most RF approaches assume that low-level features of images are as powerful as keywords in representing the semantic concepts [42]. However, this assumption is often not valid. Furthermore, most RF approaches were performed based on locally inferable assumption [112], which means that in the low-level feature space, neighbouring images should be semantically similar, or be able to confer knowledge from each other. Obviously, this assumption is conflicted with the first aspect of the semantic gap. Therefore, most RF approaches lack the capability of associating non-neighbouring images in the feature space with similar semantic concepts.
- B. RF lacks the memory capability. Most RF approaches cannot retain the obtained knowledge over a long period of time in order to continuously improve the retrieval performance. Therefore, a user has to repeat the same feedback iterations only to get the identical results he/she has obtained before [5]. Moreover, most RF approaches give a user no opportunity to use the knowledge accumulated from other users' feedbacks to satisfy his/her own query intention.
- C. RF is performed based on asymmetric and sparse feedback examples. Since users have limited patience with experiencing tedious feedback iterations, the number of positive and negative feedback examples is often far less than the number of feature dimensions. Furthermore, the negative feedback examples often outnumbered the positive feedback examples.

Facing these problems, various long-term learning (LTL) approaches were proposed to enable the interactive learning with memory capacity. These approaches retrieve images based on the long-term knowledge accumulated from previous users' feedbacks. Most LTL approaches can be grouped into two categories [42]: the LTL approaches in the first categories [67] [46] totally discard the assumption that

knowledge is locally inferable [112]. The concept is that semantic similarities of images are not inferable based on the similarities of images in the low-level features space. Instead semantic similarities are inferred solely based on historical logs of users' feedbacks, which are more reliable in representing semantic information than low-level features. Generally, LTL approaches achieve better performances than most RF techniques. However, it is very difficult to collect sufficient feedback information which covers all imagery in the databases. Therefore, these LTL approaches have limited applications in remote sensing databases due to the rapidly increasing data volume. The second category of the LTL approaches [63] [35] keeps the locally inferable assumption. The approach proposed by [63] initialises the semantic similarities between image clusters according to their feature similarities, and then gradually modifies the semantic similarities based on the accumulated users' feedbacks. In [35], semantic similarity between two arbitrary images is initialised to the same value, and the final similarity is multiplication of semantic similarities and low-level feature similarity. For both LTL approaches, the semantic similarities approximate the feature similarities in the situation that users' feedbacks are very sparse.

Manual annotations of images achieves very high accuracy. However, it is an almost infeasible solution for large image databases due to time and cost constraints. The alternative solution is to perform IAP, which annotates a small number of images and then propagates the annotations to a large number of unlabelled images based on similarity measurements or classification algorithm. IAP approaches can be categorised into four classes [42]: annotation by feature similarities [84] [90], active learning [112] [17], RF based annotation [71] [100] and machine learning based annotation [66] [4]. In particular, the RF based annotation techniques are different from the RF techniques discussed before because they are designed to perform IAP. Among these IAP approaches, RF based annotation is quite intriguing due to its economising and adaptability. The economising is caused by the fact that the annotation is accomplished by the users instead of professional annotators. The adaptability refers to the fact that the system is able to continuously improve its annotation performance through learning from the users. However, as most IAP approaches, RF based annotation cannot map the images located far way in the feature space into the same or similar semantic concepts.

2.2.1 Relevance Feedback

We will now discuss the different kinds of RF approaches. The re-weighting of features is inspired by the fact that only a subset of the feature components is important for describing the user's query intention. The importance of feature components can be reflected by the weights, which are updated based on the user's feedbacks. The re-weighting approach [89] re-weights multiple kinds of features hierarchically. First, the approach re-weights components of each individual feature. Then, it assigns weight to each kind of features.

Inspired by the similar approach [91] in the text information retrieval area, query point movement approach [87] approximates the optimal query point by moving the query point towards the relevant images and far away from the irrelevant images in the feature space.

The typical approach in the probability distribution category is the PicHunter framework [25]. The framework predicts the user-desired images based on the probability distribution of the accumulated users' feedbacks. In order to maximise the obtained information, the framework develops an entropy-minimising display algorithm to display the top-ranked images that provide the system with maximum information. The other approach [100] in this category has progressive learning capability. It uses a Bayesian classifier to improve the retrieval performance based on the positive examples and the negative examples in three steps. Firstly, a Gaussian distribution is estimated in the feature space to describe the user's query intention based on the positive examples. Then, the estimated Gaussian distribution is used to measure the ranks of which the stored images are related to the user's query intention. Finally, the negative examples are used to modify the ranks of the retrieved images. In order to reduce the redundancy, the principal component analysis (PCA) is used to detect the lower dimensional feature subspace based on the feedback examples. Then, the features are projected onto the detected feature subspaces. Thus, the retrieval speed, retrieval accuracy, and storage requirement are significantly optimised.

Some approaches consider RF as a learning or classification problem. These approaches train classifiers to classify the stored images into relevant and irrelevant images based on the user-supplied examples. The typical approaches include decision tree learning [73] and boosting [104]. The decision tree learning approach trains a decision tree based on the user's feedback while the boosting approach combines a

collection of simpler classifiers into a strong classifier. The combined strong classifier is able to classify the stored images based on a small number of selected features, which are selected from a very large number of features.

In order to capture the user's query intention at the object level, some region-based relevance feedback (RBRF) approaches have been developed. Inspired by the query point move approach [89], Jing, et al., [56] [57] combines the regions of the initial query image with the positive images into a pseudo query image. The pseudo query image is used as the optimal query at the next iteration of the retrieval. However, the regions of the pseudo image increase rapidly with the iterations. Thus, the retrieval speed gradually slows down. To solve the problem, the regions in the pseudo image are merged into a small number of larger regions. In order to detect the regions of interest to the user, the approach re-weights the regions of the pseudo images based on the user's feedback. Recently, a novel region-based RF approach [64] re-estimates region features and region importance based on attribute graphs. The approach converts images into the attribute graphs and thereby takes the spatial arrangements into consideration. A comprehensive survey of latest developments in RF can be found in [113].

2.2.2 Long-Term Learning

The information embedding system [63] splits and merges image clusters to make each image cluster a semantically related subspace based on the accumulated users' feedbacks. The system uses a semantic correlation matrix to store the correlation weights, which reflect the semantic relationships among different image clusters. In fact, the correlation matrix is able to link image clusters located far away in the feature space with the same semantic concept. Although the information embedding is an effective framework, the high computational complexity and long convergence time prevent it from being applied in large databases.

The FeedbackByPass system [5] is able to learn the query mapping from the past users' feedbacks. The query mapping is able to map initial query parameters to optimal query parameters, which include optimal query point and optimal distance function. Due to the query mapping, the system is able to completely bypass the feedback iterations for memorised queries and predict optimal query parameters for

new queries. However, the system requires a lot of users' feedbacks to achieve the desired performance.

The statistical correlation analysis [67] approach uses a bigram correlation model to capture the semantic relationships among images. It bridges the semantic gap solely based on historical logs of users' feedbacks other than low-level features. Therefore the approach is able to precisely capture the users' subjectivity. However, it is very difficult to collect sufficient feedback information to cover all imagery in the database. The LTL approach proposed by [35] learns the semantic similarities among images that are consistent with the users' subjectivity. Then, the learned long-term similarities are combined with the content-based similarities to facilitate the image retrieval. Therefore, the learned semantic similarities converge to content-based similarities if the feedback examples are sparse.

He et al. [46] incrementally construct a semantic space according to the user-supplied examples. The proposed framework assumes that relevant images of a query belong to the same hidden semantic class, which has no explicit meaning unless the user specifies it. Based on the assumption, the semantic meaning of a given image can be represented by the learned hidden semantic classes. Since different queries may belong to the same hidden semantic class, the framework utilises singular vector decomposition (SVD) to form a compact semantic space with less hidden semantic classes.

2.2.3 Annotation Propagation

The early approach proposed by Picard and Minka [84] performs annotation propagation based on image texture in two steps. Firstly, the images are decomposed into square subimages, which are pre-classified into tree structures based on the low-level features. Then, through the constructed tree structures, the annotations are propagated to the unlabeled subimages.

Saber and Tekaple [90] proposes a system that propagates annotations based on the contour of the detected regions in three steps. Firstly, regions are detected by the adaptive Bayesian colour segmentation approach. Secondly, the detected regions are refined according to the edge information. Finally, the contour of the one/multiple regions is matched with a set of annotated template to discover the potential

annotation. However, the system simply failed when the annotations cannot be precisely represented by the shapes.

Some approaches propagate annotation according to active learning and classification. An active learning system proposed in [112] uses a predefined tree of semantic attributes to propagate annotations. The system uses the biased kernel regression to estimate the attribute probabilities of unlabeled objects based on their annotated neighbours in the feature space. In order to maximise the obtained knowledge, the system selects the objects with the highest uncertainty with regard to the semantic attributes.

Content-based soft annotation [17] (CBSA) system uses active learning to perform the soft annotation propagation. It trains K binary classifiers based on the manually labelled training images, which belong to K semantic labels. The binary classifier is implemented in two learning methods, support vector machines (SVMs) and Bayes point machines (BPMs). Each classifier is used to evaluate a confidence score, which measures the extent to which an unlabelled image is related to the corresponding label. As a result, each unlabeled image is annotated by a K -dimensional label-vector.

Some annotation propagation approaches are performed based on users' feedbacks. One early attempt is the semi-automatic image annotation framework [71]. In this framework, the user starts a query by submitting a query keyword and provides his/her feedback with respect to the initial result. The query keyword is used to annotate or increase the annotation confidence for the positive images, and delete annotation or reduce annotation confidence for the negative images. In order to propagate annotations, the annotated images are used to predict the annotation of unlabelled images based on the low-level features. Therefore, the proposed framework is able to progressively improve the annotation in the database based on accumulated users' feedbacks. The other framework in this category is the progressive learning [100], which has been discussed in Section 2.2.1.

Recently, Li et al. [66] presents an automatic linguistic indexing system, which measures the extent to which an image is related to semantic categories. Training images belonging to different semantic categories are used to train a dictionary of two dimensional multiresolution hidden Markov models (2D MHMMs). Then, each model is used to evaluate the extent to which an image belongs to the corresponding semantic category. Although the system is highly accurate, it cannot be directly used for on-line training due to the high computational complexity.

Some approaches use the statistical models to represent the relationships among the stored images and the words in the description documents. According to the trained statistical models, these approaches are able to link the images with the words. Barnard et al. [4] uses multiple statistical models to find the relationships among image regions and words, whereby the multiple statistical models are able to describe the distribution in the joint space of image regions and words. Mori et al. [79] uses a simplified statistical model to link words with vector quantised representation of image regions. The approach can be used in large image databases because it has low computational complexity.

Chapter 3

Remote Sensing Imagery Retrieval Systems

In this Chapter, several remote sensing image retrieval systems are reviewed. These systems provide valuable information that is very useful for the framework proposed in this research.

3.1 Early Remote Sensing Image Retrieval Systems

In this section, four early remote sensing image retrieval systems are reviewed.

The Titan [19] system is developed to address two major problems. Firstly, in order to support real-time response for large remote sensing databases, the system partitions the data into coarse-grained data blocks and uses a simplified R-tree to index the partitioned blocks. Moreover, the system declusters the data blocks onto a large disk farm to maximise the disk parallelism and linearises the data blocks in each disk to reduce the number of disk seeks. Secondly, for the purpose of reducing the data volume, the system coordinates the computation of data compression to process the data on the machine that stores it. The system supports four kinds of queries: temporal bounds, spatial bounds, sensor type, number and resolution of output images.

The Microsoft TerraServer [6] is designed to store and index aerial images, satellite images and topographical images of the Earth. In order to support millions of on-line users simultaneously, the TerraServer adopts the typical 3-tier architecture. The first tier is a graphical web browser or any software that supports HTTP 1.1 protocols and HTML 3.2 document structure. The second tier consists of web server applications that are able to accomplish the users' requirements and access the database system in the third tier. The third tier is a SQL server 7.0 relational database system, which supports SQL stored procedures. The system supports four kinds of queries: coverage map query, place query, coordinates query and famous place query.

The Image Server System [108] (ISS) is mainly developed to access the Magellan data set, which contains the remote sensing data of Venus planet. The system can also be used to access the data set of other planets. In order to access the remotely distributed data archives provided by various suppliers, the system is composed of

one central server and multiple local servers, whereby the central server updates the inventory database, duplicates data cached on the local servers and deals with the requests that cannot be handled by the local servers. The local servers receive requests from users and communicate with the central server to retrieve data that are unavailable in the local caches. To save the memory capacity, they delete the cached data that are seldom used. The system utilises retrieval client to give the user access to the database. The retrieval client consists of two major parts, the query definition dialog and the map browser. The query definition dialog provides a set of menus and forms to support the composition of user queries. The map browser allows users to specify coordinates with mouse pointer and traverse through the map pyramid stored in the inventory database.

GeoBrowse [76] is an integrated prototype system, which is designed to manage, retrieve, process and analyse remotely sensed images. GeoBrowse consists of three distributed components: a graphical user interface (GUI), an object-relational database management system (ORDBMS) and a scientific problem solving environment (S-PLUS). These components can be installed on separate platforms since they are able to remotely cooperate with each other. The ORDBMS stores and manages large volumes of remote sensing data and derived statistical objects. The S-PLUS provides a scientific data mining environment, which offers more than 2000 analysis and graphic function. Together with the S-PLUS, the system is able to extract various features, trends and statistical attributes from the stored images. In the system, the extracted features are content-based indices of remotely sensed images. In fact, Geobrowse is one of the early systems that support content-based queries on large remote sensing databases.

3.2 Algorithm Development and Mining System

Algorithm Development and Mining [85] (ADaM) system was developed to detect geophysical phenomena and extract features from large volumes of earth science data. The system is able to handle a variety of data sets, which have different formats, types and structures. It supports various operations to extract information for different applications in Earth science. The applications range from general-purpose applications, such as image processing and statistical analysis, to highly specialised science algorithms for specific data sets. The system adopts data pipeline architecture

to guarantee the portability, network accessibility and extensibility. The pipeline architecture is composed of input module, processing module, and output module, which break the data mining process into a series of consecutive steps. The input and the output modules isolate the processing module from reading and writing different data formats. This makes the processing module focus on phenomena detection and feature extraction. The processing module mainly deals with three kinds of problems. Firstly, the module is able to detect known phenomena with a set of fixed algorithms. Secondly, the module uses interactive learning algorithms to detect known phenomena. Thirdly, the module is able to detect anomalies in the data sets, whereby the anomalies may be spurious values or rare phenomena that can hardly be detected in large volumes of data.

3.3 SPIRE system

The SPIRE [10] is a digital library framework, which is used to access scientific data, particularly imagery. It focuses on support for content-based search. In the SPIRE system, the search and retrieval paradigm is implemented in a progressive framework, which integrate data representations with search operations. This framework is especially useful for scientific data since the richness of non-structured data such as images and video requires more advanced methodologies for content specification than most current image retrieval systems. The framework organizes data in an abstract pyramid called InfoPyramid, which permits query specification at four abstraction levels including pixel or raw data level, feature level, semantic level and metadata level.

The SPIRE system supports two ways of querying the abstraction levels of the InfoPyramid. The user can specify a query through a semantic label, or providing positive examples and negative examples. In order to specify information transformation within the pyramid, the user can define new features by designing sequences of data algebra and processing operators to be applied to the raw data. Moreover, the user can categorize features and attach labels to individual categories.

In the SPIRE system, similarity search is conducted at different abstraction levels of the InfoPyramid. At the metadata level, searches are usually sharp, and matches are not ranked according to similarity. Since at the semantic level objects are extracted by either classical or fuzzy classifiers, similarity between sharply extracted objects is

defined using a zero-one metric, and similarity between fuzzy labels is defined by membership function. In order to refine searches iteratively, the computational mechanism is designed to refine example-based queries by providing additional positive and negative examples to the query engine.

In order to support a scalable and extensible search framework based on combination of date representation and data manipulation, The SPIRE system enable a user to manipulate predefined and pre-extracted data at the semantic and feature level. Further more, the system can extract new content at any level of the InfoPyramid during the query for a specialized user group. In particular, the search engine can access a library of signal/image processing operators that utilise properties of the compression algorithm to produce significant increases in execution speed over their classical counterparts. Such algorithms extract features and semantic information from the raw data, and semantic information from pre-extracted features.

In order to effectively search multimedia data, the SPIRE describes content by object-based representation. The system create simple object by a pipeline. Since simple objects are not sufficient to support an extensible framework, more complex definitions are required to define compound objects, which consist of simpler objects and a set of relationships. In order to search a compound object, the object is firstly parsed into a relation graph. Then, a minimum spanning tree of the graph is produced. A dynamic programming algorithm is used to search the tree in a depth-first, greedy way.

In order to support query specification and object definition, a novel user-interface model, DanDE(short for drag-and-drop English) is designed to provide a highly flexible general-purpose environment.

3.4 SemQuery System

The SemQuery system [94], [95] is developed to support visual queries based on heterogeneous features. It is successfully applied in remote sensing image database [94]. In the system, the images are grouped into different semantic clusters, for example, water, grass, residence, and agriculture. Each semantic cluster is represented by two sets of images that belong to colour and texture feature classes, respectively. For each feature class, each semantic cluster consists of a set of subclusters, and its scope is the union of the sub-cluster scopes. The scope of a semantic cluster based on

both color and texture feature classes is determined by intersecting the scopes of the clusters of the two feature classes. An image is assigned to a semantic cluster if it belongs to the scope of all the heterogeneous clusters of the semantic cluster. During the query session, if a query image is provided with a keyword, only images contained in the corresponding semantic cluster indicated by the keyword will be considered in the retrieving process. If no keyword is provided, the retrieving process will find the matched semantic clusters for the query image based on its features. The system also utilises a multilayer neural network to merge the different similarity measurements of different feature classes.

3.5 Diamond Eye System

Diamond Eye [14] system adopts a distributed software architecture to enable users to mine and analyse the information contained in large image collections. Users communicate with one or multiple data mining servers through a graphic user interface which is implemented as a Java applet. Each data mining server is integrated with an objected-oriented database and a computational engine. The computational engine may be formed by a network of high-performance workstations, which execute parallel distributed algorithms for expensive image processing, object recognition and content-related queries. The objected-oriented database stores and manages the user profiles, images, image meta-data, image annotations, content information, recognition models and algorithms. The client software is composed of four components: (1) The IMAGEBROWSER is used to load, display and annotate images. (2) The PLOTBROWSER performs standard 2-D plotting functions. (3) The QUERYBUILDER enables the user to construct various visual models of query objects and executes the query by comparing a query model with the stored images. (4) The THUMBNAILBROWSER is used to view and interact with mosaics of thumbnail images, which are results of a query. The system architecture is shown in Figure 3.

In general, the Diamond Eye architecture has three advantages. Firstly, the potential users may evaluate advanced data mining and machine learning techniques on the World Wide Web. Secondly, In order to guarantee the reusability, the system factors out software infrastructure that is commonly used in scientific mining applications.

Thirdly, the system allows the close cooperation between algorithm developer and domain experts.

Figure 3: Diamond Eye system architecture according to [14]

3.6 Intelligent Satellite Information Mining System

The Intelligent Satellite Information Mining System [26], [27] (ISIMS) is a next generation architecture, which enables user to rapidly retrieve high resolution satellite images. It develops a tool to update and manage large volumes of historical and newly acquired remote sensing archives and allows experts to directly access these archives in a user-friendly way. It also provides friendly user interfaces for image query and browsing.

The system contains two levels. The first level is responsible for interactively learning from the users based on the extracted features. The second level is responsible for dealing with content-related queries. In order to facilitate online learning, a signal content index is generated based on the cluster description, the scale information and

the type of stochastic model used to extract the features. The ISIMS system uses Bayesian hierarchical learning to interactively learn the prior knowledge of image structures from the user. Thereby, the user is able to annotate similar image structures occurring in the stored remote sensing images. The resulting annotations can be used to facilitate other users' queries. The architecture of ISIMS is shown in Figure 4.

Figure 4: Architecture of the Intelligent Satellite Information Mining System according to [26]

3.7 VisiMine System

VisiMine [61] is an interactive mining system, which is able to perform statistical analysis for large volumes of remote sensing images. The system architecture is shown in Figure 5. The system extracts pixel-level features, region level features and tile-level features from the remote sensing imagery. The remote sensing images and the extracted features are stored in the Oracle database system. In particular, region features are stored in the ESRI Spatial Data Engine (SDE). VisiMine provides various data mining functionalities, for example, similarity searches, clustering of image tiles, classification, and regression analysis. In order to improve the data mining ability, VisiMine connects to S-PLUS to invoke more than 2000 functions for scientific data analysis. VisiMine uses a SQL-like query language to supports different queries, which include specifying the data mining tasks and selecting features for data mining.

Figure 5: VisiMine system architecture [61]

VisiMine supports a label training method, which employs naïve Bayesian classifiers to interactively label the remote sensing archives. The naïve Bayesian classifiers are trained based on textural or spectral properties of positive and/or negative examples. In order to reduce the cost of pixel-wised labelling, VisiMine also uses decision tree models to perform labelling base on regional features.

3.8 Knowledge-Driven Information Mining System

The knowledge-driven information mining [28], [32] (KIM) system is developed as a next-generation architecture to access state-of-the-art remote sensing archives. It is able to interactively communicate with the user to retrieve images according to the semantic concept of the user's query. For this purpose, the system hierarchically links image content with the semantic concepts through five levels: image data (level 0), primitive image features (level 1), meta-features (level 2), signal classes (level 3) and user's query intentions (level 4). The linkages between image contents and semantic concepts are established in three consecutive steps: Firstly, primitive features and meta-features are extracted from remote sensing images using a group of prior Gibbs random field (GRF) models. Secondly, in the second level, an unsupervised clustering algorithm is used to classify image primitive features and meta-features into a vocabulary of signal classes. Finally, the system uses a simple Bayesian network to link the detected signal classes with the semantic concepts.

From the system perspective, the KIM system manages and explores the remote sensing images through an online component and an offline component. The offline component is used to perform computational intensive operations, that is, primitive features extraction, data reduction, and the detection of signal classes, while the on-line component interactively communicates with users to links semantic concepts with signal classes using the simple Bayesian network [32]. Based on a small number of user-supplied examples, the online component is able to navigate the users to their desired remote sensing images. The architectures of the offline component and online component are shown in Figure 6 and Figure 7, respectively.

Figure 6: Offline component of KIM system according to [32]

Figure 7: Online component of KIM system according to [32]

3.9 Summary

Early remote sensing imagery retrieval systems access remote sensing images mainly based on geographical coordinates, time of acquisition and sensor type. Only few of images acquired by these systems can be used if the user prefers to the content aspects, for example, spectral characteristics and textural characteristics. Therefore, various remote sensing information mining systems were developed to access remotely sensed images based on the image features. These systems are not simply combinations of the traditional remote sensing image retrieval systems and the CBIR systems. In fact, they are able to discover hidden patterns from large volumes of remote sensing data and retrieve remotely sensed images according the users' query intentions. Nevertheless, these information mining systems have different emphasises with regard to remote sensing databases. AdaM system focuses on detecting geophysical phenomena and extracting features from large volumes of earth science data. Based on extracted features, it is able to detect known phenomena and anomalies through fixed algorithms and interactive learning algorithms. Diamond Eye system enables users to mine and analyse the information contained in the object orientated database, which stores and manages the images, imager annotations, content information, and recognition algorithms and models. In the system, users can construct various visual models of query objects and compare a query model with the stored images. The ISIMS system uses the Bayesian hierarchical models to interactive learn the

knowledge of image structure from the user. Thereby, the user is able to annotate similar image structure in the database. The resulting annotations can be user to facilitate other users' queries. The VisiMine is an interactive mining system, which provides various data mining functionalities such as similarity search, clustering of image tiles, classification and regression analysis. It employs naïve Bayesian classifiers to interactively annotate the remote sensing images based on features of positive and/or negative examples. The KIM system is able to interactively communicate with user to retrieve images according to the semantic concept of the user's query. It establishes the linkages between image contents and semantic concepts through a simple Bayesian network.

Table I: Summary Table of the Reviewed Systems

System name	Features	Method
ADaM	—	Pipeline architecture
Diamond Eye	—	Distributed software architecture
ISIMS	—	Bayesian hierarchical model
Visimine	—	Naïve Bayesian classifier
KIM	Gibbs random field	Simple Bayesian network
SPIRE	—	Abstraction pyramid
SemQuery	Color histogram and wavelet texture	Semantic cluster
Proposed System	Color histogram and wavelet texture	Context-sensitive Bayesian network

Chapter 4

Framework of the Prototype System

In the previous chapter, various remote sensing image retrieval systems were reviewed. Among these systems, the KIM system interactively links image contents with semantic concepts through a simple Bayesian network. Inspired by the KIM system, a context-sensitive Bayesian network is designed to approximate semantic understanding ability of human being. Based on the context-sensitive Bayesian network, this framework is able to interactively communicate with the users to collect more subjective information which can be used to infer semantic concepts from the images and improve the image retrieval results. The framework gains this semantic inference ability through a three-stage process:

Firstly, every image is segmented into disjoint regions using a simple segmentation algorithm and low-level features are extracted from each region. Although the chosen segmentation algorithm is by no means optimal, it provides sufficiently homogenous regions and demonstrates the insensitivity of the proposed approach to possible missegmentation.

In the second step, the region descriptors, that is, multi-dimensional vectors, are classified into a finite number of signal classes/codes with homogeneous spectral or textural characteristics to facilitate the later semantic learning process. To implement this task, the system firstly utilises a vector quantisation algorithm, which creates a compact representation of possible region descriptors. This crucial step reduces the large volumes of the region descriptors for practical use. Then, the images in the database are encoded by codes based on the result of the previous quantisation. In fact, codes are compact representations of image regions and can be regarded as labels of the “vocabulary” of signal classes, which are valid across all the stored remotely sensed images. The first and the second step are shown in Figure 8.

Figure 8: First stage and second stage of the semantic inference and retrieval

Afterwards the regions' semantic concepts are inferred through additional consideration of the adjacent regions. The code of a region together with code of each of its adjacent regions forms code pairs, which reflect the spectral / textural characteristic of the region and its spatial context. A code pair provides more information regarding the corresponding region's semantic concept and better resembles the perception of human beings who often interpret an image region in a satellite scene by taking its adjacent regions into account. Therefore, the context-sensitive Bayesian network is used to establish multiple-to-multiple linkages among code pairs and semantic concepts. The key characteristics of the context-sensitive Bayesian network is that it links semantic concepts of image regions based on the spectral and textural characteristics of the regions themselves as well as their context, that is, the adjacent regions. To simplify the calculation, a set of semantic score functions is derived as a special case of the context sensitive Bayesian network. Each semantic score function is used to measure the extent to which a given remote sensing image is related to a

semantic concept based on the so-called code co-occurrence matrix. Figure 9 shows the flow of the semantic score functions training. In this framework, RF and LTL approaches are used to train the semantic score function.

Figure 9: Third stage of the semantic inference and retrieval

Three novel RF approaches are proposed to train the semantic score functions based on user-supplied examples. These approaches have respective advantages in improving the retrieval performance from different perspectives, that is, precision and recall rate. Due to the semantic score functions and code representations, these approaches are robust to user-supplied images containing ground cover types with very inhomogeneous signal characteristics. In order to reduce the inaccuracy caused by the quantisation error, one of the RF approaches uses a so-called pre-filtering function to improve the retrieval precision based code representations as well as low-level features. Moreover, these approaches are able to achieve real-time response in large remote sensing databases.

Since RF approaches lack memory capacity, a LTL approach is proposed to train the semantic score functions from accumulated users' feedbacks, which have heterogeneous signal characteristics. Moreover, the proposed long-term approach also

uses the pre-filtering function to synthesise the long-term knowledge of different semantic concepts to improve the performance of the semantic inference.

Figure 10: Stepwise image retrieval scheme

At the retrieval stage, a stepwise image retrieval scheme like the one shown in Figure 10 is adopted to balance effectiveness and efficiency. Firstly, the trained semantic score functions as shown in Figure 9 are used to retrieve a set of candidate images that are related to the user's semantic concepts. Next, a more expensive region-based similarity measurement IRM is applied to measure the similarity between the query image and the candidate images found in the first stage. Since the number of candidate images is smaller by order of magnitudes than the number of stored images in the database and due to the candidate images' strong relations with the user's query concepts, the proposed stepwise scheme enhances the retrieval performance and reduces the retrieval time.

Chapter 5

Feature Extraction and Image Segmentation

This chapter introduces the utilised image pre-processing in order to extract spectral and textural features from remote sensing images. Then the extracted features are used for image segmentation.

5.1 Image Characteristic

Remote sensing imagery data is heterogeneous in nature since different data sets are acquired by various remote sensing platforms, which have different spatial, spectral and temporal resolutions. Early low-resolution remote sensing platforms could hardly acquire detailed information on ground cover types. Often a large amount of classes are mixed together in low-resolution remote sensing images. Thus, it is difficult for feature extraction approaches to extract useful information from these low-resolution images. For instance [103], an urban area acquired by Landsat TM (30m spatial resolution) is composed of complex mosaics. Small objects such as buildings, roads and trees can hardly been identified within the area. But these objects can be identified in meter-resolution images. In recent years, numerous high-resolution systems have been launched, which make meter-resolution remote sensing images widely available. High-resolution images are able to record highly complex ground information in detail. Feature extraction approaches can easily extract useful information from high-resolution images. Therefore, this study chooses IKONOS image data to test the performance of the developed framework. Space Imaging's IKONOS earth imaging satellite [70], launched in September 1999, is the first 1-meter commercial remote sensing satellite in the World. The IKONOS panchromatic images are 1-meter resolution, whereby the spectral range is $0.526 - 0.929\mu\text{m}$. The IKONOS multispectral data has four spectral bands with 4-meter resolution. The basic characteristics of the four bands are summarised in Table II. The combination of IKONOS panchromatic and multispectral data can produce 1-meter true or false colour images. Each pixel of the panchromatic and multispectral image is described by 8 and 11 bits data per band, respectively.

Table II: Basic characteristic of IKONOS bands

Bands	Spectral	Wavelength (μm)
1	Blue	0.445 – 0.516
2	Green	0.506 – 0.595
3	Red	0.632 – 0.698
4	Near infrared	0.757 – 0.853
Panchromatic	–	0.526 – 0.929

5.2 Spectral Feature Extraction

The spectral information in remote sensing imagery is captured by the different intensities in multiple spectral bands, whereby the intensities are determined by various electromagnetic radiations reflected or emitted from the ground area. Since human beings are very sensitive to distinguish different colours, the multiple spectral bands together provide more information on the corresponding ground areas than each individual band. Therefore, multispectral data provide humans with more information for interpreting and recognising various ground cover types in remotely sensed images. Various spectral extraction techniques have been used to extract spectral features, which are able to precisely describe the multispectral contents of remote sensing images.

Figure 11: Procedure diagram for the equalised quantisation approach

In this study, colour histogram [109] is extracted to describe the spectral distributions of remote sensing images. For tri-spectral remote sensing imagery, if each spectral band is represented by 2^8 intensities, there will be 2^{24} possible colours, which result in 2^{24} -dimensional colour histogram. Such high dimensionality prevents colour

histogram from being applied in practical applications due to high indexing cost, storage requirement, and long execution time. Therefore, it is necessary to reduce the dimensionality to balance the storage requirement and the indexing cost on one hand, and loss of system accuracy on the other hand [109]. It is also necessary to reduce the execution time as much as possible, since the short execution time is a very critical factor for remote sensing image retrieval due to large volumes of data in remote sensing archives. In this research, an equalised quantisation [109] approach is implemented to reduce the dimensionality. Since the remote sensing images are represented as true or false colour images in the RGB colour space, the equalised quantisation approach equally divides each of red, green and blue bands into 4 levels and results in 64 cubic cells in the RGB colour space like shown in Figure 11. The colour loss caused by the quantisation is demonstrated in Figure 12. In this way, the original 2^{24} -dimensional colour histogram is reduced into 64 dimensions. Comparing to other quantisation approaches such as hierarchical clustering approach and CNS-supervised clustering approach [109], the equalised quantisation approach achieves moderate image retrieval performance using the least execution time.

Figure 12: Equalised quantisation of remote sensing image: (a) Image before quantisation, (b) Image after quantisation (Ikonos, © Space Imaging 2003)

Large remote sensing imagery often contains more than one ground cover types, the spectral distributions of which can hardly be precisely and sufficiently described by a global colour histogram. Thus, each stored remote sensing image is decomposed into a group of non-overlapping 30×30 subimages to capture the local spectral

distributions, from which 64-dimensional colour histograms are extracted.

5.3 Texture Features Extraction

Large remote sensing imagery often contains more than one ground cover types, the textures of which can hardly be precisely and sufficiently described by a global texture feature vector. Thus, each stored remote sensing image is decomposed into non-overlapping 30×30 subimages to capture the local texture, from which 10-dimensional texture feature vectors are extracted.

Texture in remote sensing imagery conveys the information of spatial arrangements and interrelationships of image pixels [24], [82]. Although non universal descriptive terms are accepted for texture, texture is often termed with smoothness, roughness, regularity, contrary, fineness, and likeness [81], [93]. During the last few decades, various approaches (e.g. co-occurrence matrices, first-order statistics, second-order statistics) have been proposed to extract texture features from remote sensing imagery ([7], [16], [43]–[45], [47], [52], [62]). For instance, grey level co-occurrence matrices (GLCM) together with different second-order texture statistics are often used to extract texture features in a moving window, and have achieved varying success ([43]–[45], [62]). However, these approaches are restricted to the analysis of local spatial relationships within relatively small neighbourhoods [75]. Therefore, wavelet-based approaches have been proposed to extract texture features in the spatial-frequency domain, whereby the extracted features describe the textures at multiple resolutions ([33], [74], [75], [86]). This research applies the wavelet-based approach [89], which utilises Daubechies wavelets [1] to extract texture features in remotely sensed images. For this purpose a subimage is decomposed into three wavelet levels, that is, ten de-correlated sub-bands. By decomposing the sub-image into three wavelet levels, the texture characteristics of each decomposed sub-image then were analysed at 1-, 2-, 4- and 8-meter resolutions, which are sufficient to describe all the ground cover features contained in the experimental datasets. Coefficients of the ten sub-bands form ten coefficients sets as shown in Figure 13. Afterwards, in order to capture the characteristics of a certain scale and orientation of the original subimage, the standard deviation of the wavelet coefficients in each set is calculated and utilised to form a ten-dimensional feature vector [89].

Figure 13: Hierarchical wavelet decomposition: (a) Original remote sensing image, (b) Three-level wavelet decomposition (Ikonos, © Space Imaging, 2003)

5.4 Image Segmentation

In this section, an image segmentation procedure that decomposes a remote sensing image into a set of regions is described, whereby a region refers to a group of connected pixels sharing uniform or near-similar texture or spectral characteristics [20], [114], [12]. The aim of region-based segmentation is to detect the regions in an image based on spectral/textural similarity as well as spatial proximity [83], [11]. However, the state-of-the-art segmentation techniques still cannot meet the requirement of accuracy. Moreover, highly precise segmentation requires high computational cost. Therefore, in this research, a simple but efficient segmentation procedure is implemented and afterwards semantic concept inference and interactive learning approaches are designed to be robust against inaccurate segmentation.

The segmentation procedure is based on the k -means algorithm, which clusters the features of the subimages to determine the dominant spectral/textural characteristics and thereby detect the potential regions. Since the segmentation is based on subimages rather than pixels, it significantly reduces the computational cost. After the segmentation, the number of region-based features is much smaller than the number

of subimage-based features, which limit the number of feature vectors and reduce the computational load of subsequent code generation. However, the number of distinct regions in an image is not fixed and thus the image segmentation requires k to be determined dynamically [106]. In this research, a maximum number of expected clusters k_{max} is pre-determined to 8, which is assumed to be no more than the number of ground cover types contained in one image. Then the k -means algorithm assigns the feature vectors into v clusters Q_i ($i \in \{1 \dots v\} \wedge 1 \leq v \leq k_{max}$). According to the Davies–Bouldin validity index [106], the optimal number of clusters is given by

$$k_{opt} = \min_v \frac{1}{v} \sum_{l=1}^v \max_{l \neq k} \frac{S(Q_k) + S(Q_l)}{d(Q_k, Q_l)}, \quad (1)$$

where $S(\cdot)$ is the within-cluster distance defined as

$$S(Q_k) = \frac{\sum_j \|f_t(j) - f_t^c(k)\|}{N_k} \quad (2)$$

using the actual feature vectors $f_t(j)$ and the centres $f_t^c(k)$ of the clusters Q_k for a given image I_t . In order to balance the granularity of the clustering, that is, avoiding over-segmentation, the distance $d(\cdot, \cdot)$ between clusters is also considered:

$$d(Q_k, Q_l) = \|f_t^c(k) - f_t^c(l)\|. \quad (3)$$

After all the clusters are assigned according to the k_{opt} -means algorithm, the subimages corresponding to each cluster j ($1 \leq j \leq k_{opt}$) are unified into the region R_t^j . Hence, the image I_t is decomposed into a set of regions $R_t^1, \dots, R_t^k$. The feature of R_t^j is indicated as $F_t(R_t^j)$, which is the centroid $f_t^c(j)$ of the cluster j . F_t is a feature function that maps the region to its feature vector. The feature function set $F = \{F_1, \dots, F_m\}$ contains different feature functions corresponding to the different kinds of features. In this research, the feature function set F contains two feature functions F_1 and F_2 , where F_1 is the colour histogram feature function and F_2 is the wavelet feature function.

Chapter 6

Codebook Generation and Image Encoding

In this chapter, an unsupervised vector quantisation algorithm is used to globally partition the low-level feature vectors of all image regions in the database into a set of codes that forms a so-called codebook. Then, each image is encoded by an individual subset of codes based on the low-level features of its regions.

6.1 Codebook Generation

In order to capture the multiple-to-multiple matches between signal classes and semantic concepts, the first step is to accurately detect the signal classes in the feature space, and then an appropriate learning approach is performed to capture the multiple-to-multiple matches. The RF approaches and the supervised learning approaches cannot detect the signal classes accurately due to the insufficient training data, which is a limitation factor of many supervised learning algorithms in the CBIR field [42]. Therefore, this research uses an unsupervised vector quantisation algorithm, that is, the generalised Lloyd algorithm (GLA) [21], to globally partition the low-level feature vectors of all image regions in the database into a set of codes called codebook. Each image is encoded by an individual subset of codes based on the low-level features of its regions. In fact, the codes can be regarded as labels of the “vocabulary” of signal classes, which are valid across all the stored remotely sensed images [32]. In general, five major advantages are obtained by using the codes to represent regional contents:

1. Unsupervised vector quantisation is performed based on all remote sensing images in the database. Thus, the detected signal classes are reliable due to sufficient training data.
2. The codes are cross-image representations [32] and each code is related to multiple images in the database. If semantic inference is performed based on the code representations, the knowledge obtained from a small number of training

images is able to cover a large number of the stored images. Thus, the learning efficiency is significantly increased.

3. Semantic inference can be performed to establish the multiple-to-multiple matches between codes and semantic concepts and thus bridge the gap between signal classes and semantic concepts. This is the focus of the following four chapters.
4. Unsupervised vector quantisation significantly compresses a large number of regional features for practical use.
5. Since the unsupervised vector quantisation algorithm compresses the n -dimensional regional feature vectors into scalar-valued codes, comparing code representations requires much less computational load compare code representations than comparing high-dimensional feature vectors. Consequently, the comparison speed is significantly increased.

It should be noticed that unsupervised vector quantisation achieves these advantages at the cost of decreased system accuracy due to the quantisation error [32], [58]. Thus, this research developed a filtering function to compensate the system accuracy in the low-level feature space, which is discussed in Chapter 7.

6.1.1 Vector Quantisation

The vector quantisation technique [36] is used to classify image features into a set of signal classes. Let all the image regions in the database denoted by $D = \{R_d^1, \dots, R_d^j, \dots, R_d^N\}$, where N is the number of image regions. Let F be a set of m different feature functions $F_i (1 \leq i \leq m)$, where m is the number of the feature functions. In the feature space F_i , the vector quantizer Q is used to partition all the image regions D into L signal classes $\lambda = \{\lambda_u\} (1 \leq u \leq L)$ according to the n -dimensional regional features $F_i(D) = \{F_i(R_d^j)\} (1 \leq j \leq N)$. Hence, each signal class λ_u is a cluster of feature vectors and is expressed as:

$$\lambda_u = \{f \mid f \in F_i(D) \quad \text{and} \quad M_i(f) = c_u\}, \quad (4)$$

where c_u is the centre feature vector of the signal class λ_u and f is a feature vector in set $F_i(D)$. M is a mapping:

$$M : R^n \rightarrow CB_i = \{c_1, \dots, c_u, \dots, c_L \mid c_u \in R^n\}, \quad (5)$$

where R^n is the Euclidean space of dimension n and c_u is the centre feature vector of the signal class λ_u . CB_i is called the codebook in the feature space F_i . The integers $(1 \leq u \leq L)$ are called codes, and are the indices of the centroid feature vectors c_u and signal classes λ_u ($1 \leq u \leq L$). The vector quantizer Q should satisfy the following two conditions:

$$\bigcup_u \lambda_u = F_i(D), \quad 1 \leq u \leq L, \quad (6)$$

$$\lambda_u \cap \lambda_v = \emptyset \quad \forall u \neq v. \quad (7)$$

Given a distortion function $d(f_j, c_u)$ that measure the distance between feature vector f_j and signal class c_u , that is, the Euclidean distance, the optimal mapping M minimises the average distortion $\bar{D}$:

$$\bar{D} = \frac{1}{N} \sum_{u=1}^L \sum_{f_j \in \lambda_u} d(f_j, c_u) \quad (8)$$

In order to be optimal, the mapping M has to satisfy the necessary Lloyd conditions:

- (i) *Nearest Neighbour Condition*: if $M(f_j) = c_u$ then $d(f_j, c_u) < d(f_j, c_v) \quad \forall u \neq v$
- (ii) *Centroid Condition*: the centroid c_u of each signal class λ_u is its mean vector (note that this is only true for the e^2 distortion):

$$c_u = \frac{1}{N_u} \sum_{f_j \in \lambda_u} f_j \quad 1 \leq u \leq L, \quad (9)$$

where N_u is the number of feature vectors in signal class λ_u .

Although there is no optimal solution to the problem of quantisation, various algorithms were developed to reduce the average distortion $\bar{D}$. These algorithms include the pairwise nearest neighbour algorithm (PNNA), the splitting algorithm, and the generalised Lloyd algorithm (GLA). Among these algorithms, only the GLA is able to fulfil the Lloyd conditions for optimality in an iterative way. However, the GLA is easily attracted by the nearest local minimum with respect to the initial codebook. Therefore, other algorithms, such as the PNNA for instance, are used to generate a suitable initial codebook to improve the performance of the GLA. In this research, the implementation follows the same strategy where the main steps are summarised as follows:

- (i) Predetermine the number of the signal classes L , and initialise the centroids c_u of the signal classes λ_u ($1 \leq u \leq L$).
- (ii) Assign all feature vectors $F_i(D)$ to each signal classes λ_u based on the *Nearest Neighbour Condition*.
- (iii) Update the centroids c_u of the signal classes λ_u by computing the mean vectors of the signal classes λ_u .
- (iv) Go to step (ii) if the algorithm has not converged, otherwise stop.

6.1.2 Tree-Structured Vector Quantisation

The GLA generates a codebook of fixed size, but a suitable size cannot be determined straightforwardly. On the one hand to improve the semantic inference accuracy, a larger codebook is desirable in order to allow individual code associate even to relatively infrequently occurring regions. However, this specialisation counteracts retrievals that are represented by a not perfectly suited query. On the other hand, if the codebook size is decreased, then less suitable queries are tolerated, but a mixing of actually disjoint semantic concepts, and hence less accuracy, can be observed.

To solve the problem the used vector quantisation technique was extended in the form of a tree-structured approach [36] by applying the GLA recursively. The resulting tree-structure codebook CB_i is composed of codes at different level CB_i^l , where l indicates the level of the codebook. An example for such a tree-structured codebook is shown in Figure 14. Each code in one of the lower levels of the codebook is associated with a signal class that contains a larger number of feature vectors of the regions. Contrarily, in the higher levels of the codebook a signal class is associated with a smaller number of feature vectors. Thus, flexibility and efficiency can be guaranteed by enabling the user to choose codes at different levels CB_i^l .

Figure 14: Tree-structured codebook with three levels

The construction of the tree starts from the root, that is, all the image regions D are partitioned into L signal classes according to their features $F_i(D)$. Then, at the root level the regions belonging to the same signal class are again classified into M sub-signal classes in the next level. This procedure is repeated until the predefined level is met.

6.2 Image Encoding and Code Representation of Image Regions

If the codebook CB_i is not tree-structured, then each region R_i^j in the image I_i is encoded as the code $C_i(R_i^j)$, which refers to the signal class nearest to the region's feature vector $F_i(R_i^j)$. $C_i(\cdot)$ is a code function that maps the region to its code. The code function set $C = \{C_1, \dots, C_n\}$ contains the different code functions corresponding to the different kinds of feature functions. The formal definition of $C_i(\cdot)$ with respect to the region R_i^j follows the *Nearest Neighbour Condition*:

$$C_i(R_i^j) = \arg \min_{1 \leq u \leq L} (d(F_i(R_i^j), c_u)) \quad (10)$$

Accordingly, each image is represented by a list of codes. This procedure is illustrated in Figure 15.

If the codebook is tree-structured, the code function is redefined as $C_i^l(\cdot)$, where the superscript l indicates the level of the tree-structured codebook. The encoding procedure is summarised as the follows:

- (i) Given a tree-structured codebook CB_i with m levels, a region R_i^j of an image I_i , and its feature vector $F_i(R_i^j)$. The current tree-structure level l is set to 1.
- (ii) At the l th level of the codebook, find the signal class λ_u , whose centroid c_u is nearest to the region's feature vector $F_i(R_i^j)$. Then, the region R_i^j is encoded as the code u . Thus, the code function $C_i^l(R_i^j)$ maps the region R_i^j to the code u at the l th level of the codebook. Afterwards increase l by 1.
- (iii) If $l > m$, stop. Otherwise go to step (ii).

Finally, a given region R_i^j is encoded as a list of codes, each of which refers to a signal class at the different levels of the codebook. Thus, the image I_i is represented by multiple code representations at different codebook levels which balances the problem of predetermining a suitable codebook size.

Figure 15: Image encoding procedure

6.3 Results and Assessment

In the experiments, 26 multispectral satellite scenes acquired by IKONOS in 2003 were utilised. Each scene was rotated by 16 different angles and the interim results were cut into image tiles. After removing unsuitable image tiles, i.e. subimages that incorporated undefined areas from the resampling process, a total set of 25'827 image tiles with a size of 300×300 pixel were available for the database. The dataset covers more than 37'190 km² with a variety of different landscape types, while the rotated versions of the original data ensure that a sufficiently large amount of similar images exists.

Prior to the retrieval experiments, the stored images were manually classified according to seven semantic concepts, that is, cloud (981), water (7913), forest (4398), buildings (7781), farmland (1873), bare soil (1750) and rock (1328). The numbers in brackets indicate how many related images were associated with the corresponding semantic, whereby a given image may belong to multiple categories. The types of categories were chosen to cover a broad spectrum of ground covers and are only to be seen as an example. Note that the manual classification provides the foundation for the performance assessment. In the experiment, a three-level tree-

structured codebook was generated based on spectral features as defined in Section 5.2, whereby the numbers of the codes at three different levels are 300, 600 and 1200, respectively. For convenience, the codes in the spectral codebook are called spectral codes. This experiment only uses the codes at the first level of the spectral codebook.

Figure 16: Spectral code frequencies of the experimental remote sensing database

Figure 17: Spectral code frequency of the remotely sensed images that are related to semantic concept ‘bare soil’

In order to proven that the codes are cross-image representations, experiments were carried out to illustrate the spectral code frequencies of the entire database, whereby the code frequency measures the number of images that are related to the individual code. In Figure 16 each code is related more than one image in the database. Thus, it

can be proven that the codes are cross-image representations, which index the signal classes that are valid across the entire database.

Figure 18: Spectral code distribution of the remotely sensed images that are related to semantic concept ‘rock’

Table III: Number of codes having non-zero code frequencies with respect to two semantic concepts

	Cloud	Water	Forest	Urban	Farm	Soil	Rock
Cloud	—	128	113	116	73	72	68
Water	—	—	251	265	134	95	87
Forest	—	—	—	235	134	93	87
Urban	—	—	—	—	132	96	87
Farm	—	—	—	—	—	69	65
Soil	—	—	—	—	—	—	81
Rock	—	—	—	—	—	—	—

In order to prove that there exist multiple-to-multiple matches between codes and semantic concepts, the code frequencies of two different semantic concepts, that is, ‘bare soil’ and ‘Rock’, are shown in Figure 17 and Figure 18, respectively. The two figures are analysed from two perspectives. Firstly, for each individual semantic concept, more than one code has non-zero frequency. This shows that one semantic concept can be related to more than one signal classes. Moreover, if a code has high frequency correspondence to a semantic concept, the corresponding signal class is strongly related to the semantic concept. Inversely, if a code has low or zero

frequency with respect to a semantic concept, the corresponding signal class is weakly related to the semantic concept. Therefore, the code frequencies of a semantic concept can be used to measure the extents to which the codes and the corresponding signal classes are related to the semantic concept. Secondly, comparing the code frequencies of two arbitrary semantic concepts, Table III shows the number of codes having non-zero frequencies with respect to both semantic concepts. For example, there are 81 codes having non-zero frequencies with respect to ‘Soil’ and ‘Rock’. This shows that the same signal class can be interpreted as different semantic concepts under different circumstances. Integrating the above two perspectives, it can be proven that there exist multiple-to-multiple matches between signal classes and semantic concepts. In order to effectively and efficiently capture the multiple-to-multiple matches, a context-sensitive Bayesian network is proposed in the next chapter.

Chapter 7

This chapter introduces a simple Bayesian network as well as a context-sensitive Bayesian network in order to link signal classes with semantic concepts. To simplify the calculation, a set of semantic score functions is derived from the context-sensitive Bayesian network.

Semantic Concept Inference

The previous chapter has established the existence of multiple-to-multiple matches between signal classes and semantic concepts, which can be represented by the multiple-to-multiple linkages between codes and semantic concepts. In this research, the goal of the semantic concept inference is to measure the extent to which a remotely sensed image is related to a set of predefined semantic concepts. In order to fulfil this aim, it is essential to establish the multiple-to-multiple linkages between signal classes and semantic concepts. From the perspective of probability, the stochastic linkages within the joint space of codes and semantic concepts can be modelled by Bayesian networks. Recently, the simple Bayesian network has been adopted by several remote sensing information mining systems (ISIMS [26], [27], VisiMine [61], KIM system [32], [28]) to bridge the semantic gap. However, the simple Bayesian network does not take into account the spatial information in the semantic inference process. This drawback significantly deteriorates the performance of the semantic inference especially when the automatically extracted features incorporate insufficient or no spatial information, in our case, the colour histogram. Therefore, this research proposes a context-sensitive Bayesian network to incorporate spatial information in the semantic inference process. The key characteristics of the context-sensitive Bayesian network is that it infers semantic concepts of image regions based on the spectral and textural characteristics of the regions themselves as well as their context, that is, the adjacent regions. This characteristic not only significantly improves the performance of the semantic concepts inference, but also allows more flexibility in choosing the low-level features.

In order to illustrate the advantages of the context-sensitive Bayesian network, this chapter firstly provide a complete description of the simple Bayesian network in the context of the code representations of remotely sensed imagery. Then the context-sensitive Bayesian network is presented.

7.1 Semantic Concept Inference via Simple Bayesian Network

The simple Bayesian network [31] has been adopted by Schröder, et al., (2000), who developed an interactive learning approach to link remote sensing archives to the user's interests by mapping extracted image features within a Bayesian framework to the so-called meta-features, that is, semantic concepts. Later, Datcu, et al., (2003) developed a knowledge-driven information mining system [32] based on Schröder's work. A comprehensive description can be found in Datcu, et al., (2005) [28].

In Chapter 5 and Chapter 6, the images are segmented and encoded; the corresponding features are extracted. Based on the encoded images, the semantic concepts are inferred through the Bayesian network. Let a set of predefined semantic concepts be $SC = \{SC_1, \dots, SC_Z\}$. Let the image I_t contains a set of region $\{R_t^j\}$, where the individual code of the region R_t^j is denoted as $C_t^l(R_t^j)$. Regional codes $C_t^l(R_t^j)$ belongs to the l level of a tree-structured codebook $CB_t^l = \{c_1, \dots, c_u, \dots, c_L | c_u \in R^p\}$, whereby each code u is a positive integer indexing a signal class λ_u and the corresponding centroid c_u in the feature space of F_i , and L is the total number of codes. In order to perform the semantic inference, the stochastic links between codes u ($1 \leq u \leq L$) and semantic concepts SC_k ($1 \leq k \leq Z$) have to be built and measured by the conditional probability $P(u|SC_k)$. Then, the extent to which the image I_t is related to the semantic concept SC_k is measured by the posterior probability $P(SC_k|I_t)$. Figure 19 shows the simple Bayesian network, which in this case represents the probabilistic relations among a remotely sensed image, image regions, codes, and semantic concepts.

Figure 19: Simple Bayesian network

According to the simple Bayesian network, the posterior probability is calculated as

$$\begin{aligned}
 P(SC_k | I_t) &= \sum_{R_t^j \in I_t} P(SC_k | R_t^j) P(R_t^j | I_t) \\
 &= \sum_{R_t^j \in I_t} P(R_t^j | I_t) \sum_{u \in CB_t^j} P(SC_k | u) P(u | R_t^j)
 \end{aligned} \tag{11}$$

According to Bayes' formula, for $P(u) > 0$,

$$P(SC_k | u) = \frac{P(u | SC_k) P(SC_k)}{P(u)}. \tag{12}$$

Then, Equation (11) can be expressed as

$$\begin{aligned}
 P(SC_k | I_t) &= \sum_{R_t^j \in I_t} P(R_t^j | I_t) \sum_{u \in CB_t^j} \frac{P(u | SC_k) P(SC_k)}{P(u)} P(u | R_t^j) \\
 &= P(SC_k) \sum_{R_t^j \in I_t} P(R_t^j | I_t) \sum_{u \in CB_t^j} \frac{P(u | SC_k) P(u | R_t^j)}{P(u)}.
 \end{aligned} \tag{13}$$

The probability $P(u)$ is the prior probability of code u and is addressed in Section 5.2. In remote sensing database, large volumes of data exist. It is difficult to know the prior probability $P(SC_k)$. Therefore, the first factor, the prior probability $P(SC_k)$, is assumed to follow a uniform distribution and thus can be omitted. The probability $P(u | R_t^j)$ is calculated in a deterministic way:

$$P(u | R_t^j) = \begin{cases} 1 & C_t^j(R_t^j) = u \\ 0 & \text{otherwise} \end{cases}. \tag{14}$$

The probability $P(R_t^j|I_t)$ reflects the importance of the region in interpreting the image I_t . Given the assumption that every region is equally important, the probability is calculated as

$$P(R_t^j | I_t) = 1/|I_t|, \quad (15)$$

with $|I_t|$ representing the number of regions belonging to the image I_t . If the assumption that the regions occupying larger areas provide more evidences with respect to the semantic concepts of the image, the probability is expressed as

$$P(R_t^j | I_t) = \frac{Area(R_k^j)}{Area(I_t)}, \quad (16)$$

where $Area(\cdot)$ is the number of pixels belonging to the corresponding region or image. Therefore, $P(R_t^j|I_t)$ is the percentage of the image I_t covered by the region R_t^j .

7.2 Interactive Learning via Simple Bayesian Network

In order to calculate the posterior probabilistic $P(SC_k|I_t)$, the probability $P(u|SC_k)$ has to be learned based on a user-supplied training set T_k consisting of regions R_k^j that reflect SC_k . Then the code frequency is calculated as

$$CF_k(u) = \sum \varepsilon(R_k^j, u) \quad \forall R_k^j \in T_k, \quad (17)$$

whereby the selection function ε is defined as

$$\varepsilon(R_k^j, u) = \begin{cases} 1 & C_k^l(R_k^j) = u \\ 0 & otherwise \end{cases}. \quad (18)$$

Having obtained the code frequencies from the training set T_k , the probability $P(u|SC_k)$ is approximated by

$$P(u|SC_k, T_k) = \frac{CF_k(u)}{\sum_v CF_k(v)}. \quad (19)$$

If the user supplies another training set T'_k that has no intersection with the training set T_k , that is, $T'_k \cap T_k = \emptyset$, then the probability $P(u|SC_k)$ is updated by $P(u|SC_k, T'_k, T_k)$. Let CF'_k be the code pair frequencies obtained from the training set T'_k , then the new probability $P(u|SC_k, T'_k, T_k)$ is calculated as

$$P(u | SC_k, T'_k, T_k) = \frac{CF_k(u) + CF'_k(u)}{\sum_v (CF_k(v) + CF'_k(v))}. \quad (20)$$

In Equation (13), the prior probability $P(u)$ is another factor that affects the calculation of $P(SC_k|I_l)$ and hereafter is approximated by the probability $P(u|D)$, whereby D is the set of all extracted regions in the database. Then the actual calculation of $P(u|D)$ is according to Equations (17) – (19).

However, the simple Bayesian network takes no account of the spatial information in the inference process. This drawback of the simple Bayesian network does not affect the performance of the interactive learning approach proposed by Schröder, et al., (2000) since the image features are extracted based on Gibbs-Markov Random Field models [29], [30], which incorporate sufficient spatial information during the feature extraction process. However, if the features without or with little spatial information are applied, the colour histogram, the performance of the simple Bayesian network is limited. Therefore, this research designs the context-sensitive Bayesian network, which takes the spatial information into account during the semantic inference process and thus increases the flexibility of utilising features with no or little spatial information.

7.3 Semantic Concept Inference via Context-Sensitive Bayesian

Network

Human beings have the ability to recognise and interpret objects or images, which exhibit complex spectral and textural patterns. These objects or images are very likely to be decomposed into a set of separate regions at the segmentation stage. Although each of these separated regions contains pixels that are homogenous with respect to spectral or textural characteristics, different regions may exhibit very different spectral or textural patterns. Human beings can easily recognise that these regions form the same semantic object since the co-occurrence of multiple spectral or textural patterns indicates the existence of certain semantic object. From the regional perspective, Human beings infer the semantic concept of a region based on not only the region itself but also its contexts, that is, the adjacent regions. For example, a bright / saturated image region may be a part of a technical installation on the rooftop of a skyscraper, a cloud, or a rocky area in the mountains. Without sufficient context

information, even a human being has difficulty in precisely interpreting the semantic concept of that region. Therefore, the adjacent regions of a region should play an important role in inferring the semantic concept of the region. Inspired by the context-sensitive inference ability of human beings, this research suggests a context-sensitive Bayesian network to automatically infer the semantic concept of a region R_t^j in the image I_t . The inference is based on the individual code $C_t^l(R_t^j)$ of the region R_t^j and the codes of its adjacent regions. Let $adj(R_t^j)$ be the set of regions that are adjacent to the region R_t^j . The region R_t^j and each of its adjacent regions $R_t^k \in adj(R_t^j)$ form a region pair (R_t^j, R_t^k) , where R_t^k is the contextual region of the region R_t^j . The corresponding code pair $(C_t^l(R_t^j), C_t^l(R_t^k))$ contains the spectral / textural information of the region R_t^j and its spatial context R_t^k . Due to the introduction of region pair and code pair, the region R_t^j is more precisely represented by $|adj(R_t^j)|$ code pairs than the single regional code $C_t^l(R_t^j)$, where $|adj(R_t^j)|$ is the number of adjacent regions. For a code pair (u, v) , the first code u is called regional code while the second code v is called context code. Thus, the code pair (u, v) is order-sensitive. If the codes u and v belong to the l level of a tree-structured codebook CB_t^l , then there are $L \times L$ possible combinations of code pairs. This greatly increases the number of possible regional representations from L to L^2 , which is another factor that improves the semantic inference performance. In order to perform the semantic inference, the extent to which the image I_t is related to the semantic concept SC_k is measured by the posterior probabilistic $P(SC_k|I_t)$ through the context-sensitive Bayesian network.

Figure 20: Context-sensitive Bayesian network

Figure 20 shows the context-sensitive Bayesian network representing the probabilistic relations among remotely sensed imagery, region pairs, code pairs, and semantic concepts. According to the context-sensitive Bayesian network, the posterior probability is calculated as

$$\begin{aligned}
 P(SC_k | I_t) &= \sum_{\substack{R_t^j \in I_t \\ R_t^k \in adj(R_t^j)}} P(SC_k | (R_t^j, R_t^k)) P((R_t^j, R_t^k) | I_t) \\
 &= \sum_{\substack{R_t^j \in I_t \\ R_t^k \in adj(R_t^j)}} P((R_t^j, R_t^k) | I_t) \sum_{u,v \in CB_t^l} P(SC_k | (u,v)) P((u,v) | (R_t^j, R_t^k)).
 \end{aligned} \tag{21}$$

According to Bayes' formula

$$P(SC_k | (u,v)) = \frac{P((u,v) | SC_k) P(SC_k)}{P((u,v))}. \tag{22}$$

Then, Equation (21) can be expressed as

$$\begin{aligned}
 P(SC_k | I_t) &= \sum_{\substack{R_t^j \in I_t \\ R_t^k \in adj(R_t^j)}} P((R_t^j, R_t^k) | I_t) \sum_{u,v \in CB_t^l} \frac{P((u,v) | SC_k) P(SC_k)}{P((u,v))} P((u,v) | (R_t^j, R_t^k)) \\
 &= P(SC_k) \sum_{\substack{R_t^j \in I_t \\ R_t^k \in adj(R_t^j)}} P((R_t^j, R_t^k) | I_t) \sum_{u,v \in CB_t^l} \frac{P((u,v) | SC_k) P((u,v) | (R_t^j, R_t^k))}{P(u,v)}.
 \end{aligned} \tag{23}$$

From Equation (23) and Figure 20, the conditional probability $P((u,v)|SC_k)$ models the stochastic links between code pairs (u,v) ($1 \leq u \leq L, 1 \leq v \leq L$) and semantic concepts SC_k ($1 \leq k \leq Z$). Note that the probability $P((u,v)|SC_k)$ can be obtained through learning from the user's feedback or pre-provided training samples and is addressed in Section 8.1. The probability $P((u,v))$ is the prior probability of code pair (u,v) and is also addressed in Section 8.1. The first factor, the prior probability $P(SC_k)$, is assumed to follow a uniform distribution and thus can be omitted. The probability $P((u,v)|(R_t^j, R_t^k))$ is closely related to the applied vector quantisation algorithm. If fuzzy vector quantisation [49] is applied, regions R_t^j and R_t^k may be encoded as multiple codes with different membership values. The membership values of regions R_t^j and R_t^k can be expressed as probabilities $P(u|R_t^j)$ and $P(v|R_t^k)$, which reflect the probabilities of R_t^j and R_t^k being encoded as u and v . However, in this research the applied vector quantisation algorithm works in a non-fuzzy way, thus the probability $P((u,v)|(R_t^j, R_t^k))$ is calculated in a deterministic way

$$P((u,v)|(R_t^j, R_t^k)) = \begin{cases} 1 & C_t^j(R_t^j) = u \wedge C_t^k(R_t^k) = v \\ 0 & \text{otherwise} \end{cases} \quad (24)$$

The probability $P((R_t^j, R_t^k)|I_t)$ reflects the importance of the region pair (R_t^j, R_t^k) in interpreting the image I_t . The probability $P((R_t^j, R_t^k)|I_t)$ can be decomposed into regional importance function $RI(\cdot)$ and context regional importance function $CRI(\cdot, \cdot)$, whereby $RI(R_t^j)$ measures the importance of region R_t^j in interpreting the image I_t and $CRI(R_t^j, R_t^k)$ measures the importance of context region R_t^k in interpreting the region R_t^j . If every region is equally important in interpreting the image I_t , $RI(R_t^j) = 1/|I_t|$. If every context region is equally important in interpreting the region R_t^j , $CRI(R_t^j, R_t^k) = 1/|adj(R_t^j)|$.

Hence, the probability $P((R_t^j, R_t^k)|I_t)$ equals product of the regional importance function and the context regional importance function

$$P((R_t^j, R_t^k)|I_t) = 1 / \left(|I_t| |adj(R_t^j)| \right) \quad (25)$$

If the regions covering larger areas provide more evidences with respect to the semantic concepts of the image, $RI(R_t^j) = \text{Area}(R_t^j) / \text{Area}(I_t)$. If the context regions covering larger areas provide more evidences with respect to the semantic concepts of the region R_t^j , $CRI(R_t^j, R_t^k) = \text{Area}(R_t^k) / \text{Area}(adj(R_t^j))$. Then, the probability $P((R_t^j, R_t^k)|I_t)$ is expressed as

$$P((R_t^j, R_t^k) | I_t) = \frac{Area(R_t^j)}{Area(I_t)} \times \frac{Area(R_t^k)}{Area(adj(R_t^j))}. \quad (26)$$

For Equation (25), it can be easily proven that

$$\sum_{R_t^k \in adj(R_t^j)} P((R_t^j, R_t^k) | I_t) = \frac{1}{|I_t|} \text{ and } \sum_{R_t^j \in I_t} \sum_{R_t^k \in adj(R_t^j)} P((R_t^j, R_t^k) | I_t) = 1. \quad (27)$$

For Equation (26), it can be proven that

$$\sum_{R_t^k \in adj(R_t^j)} P((R_t^j, R_t^k) | I_t) = \frac{Area(R_t^j)}{Area(I_t)} \text{ and } \sum_{R_t^j \in I_t} \sum_{R_t^k \in adj(R_t^j)} P((R_t^j, R_t^k) | I_t) = 1. \quad (28)$$

Equations (27) and (28) prove that Equations (25) and (26) prevent probability $P(SC_k | I_t)$ from being biased towards regions with more adjacent regions and images with more regions.

If Equations (24) and (25) are used to estimate the probabilities $P((u, v) | (R_t^j, R_t^k))$ and $P((R_t^j, R_t^k) | I_t)$, and $P(SC_k)$ is omitted, the posterior probability $P(SC_k | I_t)$ can be simplified as

$$P(SC_k | I_t) \propto \frac{1}{|I_t|} \sum_{\substack{R_t^j \in I_t \\ R_t^k \in adj(R_t^j)}} \frac{1}{|adj(R_t^j)|} \sum_{u, v \in CB_t^j} \frac{P((u, v) | SC_k) P((u, v) | (R_t^j, R_t^k))}{P(u, v)}. \quad (29)$$

7.4 Code Co-occurrence Matrix

In order to simplify the calculation of $P(SC_k | I_t)$ and clarify the meaning of Equation (29), the concept of code pair importance (CPI) is introduced. The extent to which the code pair (u, v) is related to the semantic concept SC_k is measured by the code pair importance (CPI) function $CPI_k(\cdot, \cdot)$, which is the posterior probability of the semantic concept SC_k given the code pair (u, v) :

$$CPI_k(u, v) = P(SC_k | (u, v)) = \frac{P((u, v) | SC_k) \cdot P(SC_k)}{P((u, v))}. \quad (30)$$

where the prior probability $P(SC_k)$ is uniform and thus can be omitted. Hence, the simplified equation for the $CPI_k(\cdot, \cdot)$ is

$$CPI_k(u, v) = \frac{P((u, v) | SC_k)}{P((u, v))}. \quad (31)$$

According to Equation (31), Equation (29) can be expressed as

$$P(SC_k | I_t) \propto \frac{1}{|I_t|} \sum_{\substack{R_t^j \in I_t \\ R_t^k \in adj(R_t^j)}} \frac{1}{|adj(R_t^k)|} \sum_{u,v \in CB_t^l} CPI_k(u,v) P((u,v) | (R_t^j, R_t^k)) \quad (32)$$

As indicated before, the codes u and v belong to the l level of a tree-structured codebook CB_t^l , and there are $L \times L$ possible combinations of code pairs. These forms the so-called code co-occurrence matrix $\mathbf{M}_k$, which describes the code pair importance (CPI) with respect to the semantic concept SC_k . Based on $\mathbf{M}_k$, Equation (32) can be further expressed as

$$P(SC_k | I_t) \propto \frac{1}{|I_t|} \sum_{\substack{R_t^j \in I_t \\ R_t^k \in adj(R_t^j)}} \frac{1}{|adj(R_t^k)|} \sum_{u,v \in CB_t^l} M_k(u,v) P((u,v) | (R_t^j, R_t^k)) \quad (33)$$

According to Equation (24), the following equation holds:

$$\sum_{u,v \in CB_t^l} M_k(u,v) P((u,v) | (R_t^j, R_t^k)) = M_k(C_i^l(R_t^j), C_i^l(R_t^k)) \quad (34)$$

Therefore, Equation (33) can be expressed as

$$P(SC_k | I_t) \propto \frac{1}{|I_t|} \sum_{R_t^j \in I_t} \frac{1}{|adj(R_t^j)|} \sum_{R_t^k \in adj(R_t^j)} M_k(C_i^l(R_t^j), C_i^l(R_t^k)) \quad (35)$$

7.5 Semantic Score

In this section, the concept of semantic score is introduced, which measures the extent to which a region or an image is related to a semantic concept. The relationship between semantic score and posterior probability $P(SC_k | I_t)$ is also clarified. The semantic score can be classified into two categories: regional semantic score and image semantic score. The regional semantic score function $SR_k(\cdot)$ that measures the extent to which a region R_t^j is related to the semantic concept SC_k is defined as

$$SR_k(R_t^j) = \frac{1}{|adj(R_t^j)|} \sum_{R_t^k \in adj(R_t^j)} M_k(C_i^l(R_t^j), C_i^l(R_t^k)) \quad (36)$$

whereby $|adj(R_t^j)|$ is used to prevent the semantic score function $SR_k(\cdot)$ from being biased towards regions with more adjacent regions. Based on $\mathbf{M}_k$ and $SR_k(\cdot)$, the image semantic score function $SI_k(\cdot)$ for an image I_t is defined as

$$SI_k(I_t) = \frac{1}{|I_t|} \sum_{R_t^j \in I_t} SR_k(R_t^j), \quad (37)$$

whereby $|I_t|$ is used to prevent the semantic score function $SI_k(\cdot)$ from being biased towards images with more regions. The semantic score function $SI_k(I_t)$ is used to measure the extent to which I_t is related to the semantic concept SC_k . If $SI_k(I_t)$ is larger than a predefined threshold, then the image I_t is assumed to possess at least the semantic concept SC_k . According to Equations (35)–(37), the following conclusion can be drawn:

$$P(SC_k | I_t) \propto SI_k(I_t). \quad (38)$$

Note that this conclusion is a special case of Equation (23), whereby Equations (24) and (25) are considered and the prior probability $P(SC_k)$ is uniform. From this conclusion, the image semantic score is consistent with the context-sensitive Bayesian network.

According to Equations (35) – (38), the following lemmas hold:

Lemma 1: The non-zero elements in the code co-occurrence matrix $\mathbf{M}_k$ determine the number of the stored remote sensing images that have non-zero semantic scores with respect to the semantic concept SC_k .

Lemma 2: if the different choices of $P((R_t^j, R_t^k) | I_t)$ always produce values greater than zero, they will not affect the number of the stored remote sensing images that have non-zero semantic scores with respect to the semantic concept SC_k .

Chapter 8

This chapter introduces three different RF approaches, which are designed to improve the retrieval performance from different perspectives.

Relevance Feedback

The concept of RF was firstly invented in the domain of text information retrieval [92]. Then, it was applied into CBIR during early and mid 1990's [113]. RF approaches are very popular in some CBIR systems and some remote sensing image retrieval systems since they are able to interactively capture the semantic concepts of the user's query. The approaches free the user from the direct modification of internal parameters in order to optimise the retrieval precision. Instead, the users only evaluate the retrieval result iteratively. Then, the approaches gradually adjust the query parameters to improve the retrieval accuracy. The success of the RF approaches can be attributed to their real-time response and their sensibility to the end user's subjectivity. However, four problems of the RF approaches [42] limit their applications in the remote sensing image retrieval. Firstly, most RF approaches cannot completely solve the inconsistency between low-level features and high-level semantic concepts, that is, most RF approaches cannot efficiently capture the user's query intention if the positive images are scatteredly distributed in the feature space. From the remote sensing perspective, this gap is caused by the distinction between signal classes and semantic concepts (information classes) [65], [54], whereby semantic concepts are those categories at semantic level and may refer to different ground cover types such as forest, farm, rock, etc. Signal classes are clusters of pixels, regions or images that are distributed homogenously in the feature spaces. Secondly, the performances of RF approaches are often limited by the asymmetry and sparsity of the feedback examples. The sparsity is caused by the fact that users often have limited patient with providing sufficient feedback examples, while the asymmetry refers to the fact that the negative images often outnumber the positive images, where the negative images are marked by the users and irrelevant to the users' query intention.

Thirdly, the real-time response of most RF approaches cannot be guaranteed for large image databases, for example, remote sensing databases. Fourthly, the RF approaches lack memory. The knowledge obtained in a query session is totally discarded at the end of the query session.

This research proposes three types of RF approaches to solve the fore three problems while the last problem is addressed in the next chapter. The first RF approach interactively learns a semantic score function to measure the extent to which a stored remote sensing image is related to the user's query intention. Since the semantic score function is a special case of the context-sensitive Bayesian network, the first approach is able to link different code pairs to the user's query intention and is called Bayesian network relevance feedback (BNRF) approach. According to the fact that each code refers to a signal class in the feature space, the BNRF approach is able to overcome the first problem at least in theory. Moreover, the codes of the regions are cross-image representations [32] and each code is related to multiple images in the database. Since the BNRF approach is based on the code representations, the knowledge obtained from a small number of training images is able to cover a large fraction of stored images. Thus, facing the sparse feedback examples, the learning efficiency is significantly increased. To further solve the problem of asymmetry, the second RF approach focuses on learning from the negative images. In addition to the semantic score function, the second RF approach also interactively learns a pre-filtering function, which is able to synthesis the knowledge from both the positive images and the negative images. For convenience, the second RF approaches is so-called PBNRF (pre-filtering Bayesian network relevance feedback) approach. Experimental results have shown that the pre-filtering function can continuously improve the precision even when only negative images are eventually provided. Therefore, the PBNRF approach is able to improve the retrieval performance based on the asymmetric feedback examples. The real-time response of RF approaches is essential since the users have limited patience with waiting for the retrieval results. However, this is not guaranteed for large remote sensing databases. Therefore, the BNRF approach is used to pre-select a set of candidate images, which is much smaller than the original database. Then, a more expensive RBRF approach is performed on the candidate images to reduce the response time.

8.1 Relevance Feedback via the Context-Sensitive Bayesian Network

The BNRF approach utilises the semantic score to measure the extent to which a remote sensing image I_t is related to the user's query intention. In order to calculate the image-specific semantic score $SI_k(I_t)$, the probability $P((u,v)|SC_k)$ has to be learned based on a user-supplied positive training set T_k consisting of regions R_k^i that reflect SC_k . Then the code pair frequency (CPF) is calculated as

$$CPF_k(u,v) = \sum \varepsilon((R_k^i, R_k^{j'}), (u,v)) \quad \forall R_k^i \in T_k \quad \wedge \quad R_k^{j'} \in adj(R_k^i), \quad (39)$$

where $R_k^{j'}$ may not belong to the training set T_k . The selection function ε is defined as:

$$\varepsilon((R_k^i, R_k^{j'}), (u,v)) = \begin{cases} 1 & \text{if } C_i^l(R_k^i) = u \quad \wedge \quad C_i^l(R_k^{j'}) = v \\ 0 & \text{otherwise} \end{cases}. \quad (40)$$

Having obtained the code pair frequencies from the training set T_k , the probability $P((u,v)|SC_k)$ is approximated by

$$P((u,v)|SC_k, T_k) = \frac{CPF_k(u,v)}{\sum_{v,\omega} CPF_k(v,\omega)}. \quad (41)$$

If the user supplies another positive training set T'_k that has no intersection with the positive training set T_k , that is, $T'_k \cap T_k = \emptyset$, then the probability $P((u,v)|SC_k)$ is approximated by $P((u,v)|SC_k, T'_k, T_k)$. Let CPF_k' be the code pair frequencies obtained from the training set T'_k , then the new probability $P((u,v)|SC_k, T'_k, T_k)$ is calculated as

$$P((u,v)|SC_k, T'_k, T_k) = \frac{CPF_k(u,v) + CPF_k'(u,v)}{\sum_{v,\omega} (CPF_k(v,\omega) + CPF_k'(v,\omega))}. \quad (42)$$

If $T'_k \cap T_k \neq \emptyset$, then the probability $P((u,v)|SC_k)$ is estimated by $P((u,v)|SC_k, T'_k \cup T_k)$, the calculation of which follows Equation (41). In Equation (31), the prior probability $P((u,v))$ is another factor that affects the calculation of the code pair importance and hereafter is approximated by $P((u,v)|D)$, whereby D is the set of all extracted regions in the database. Then the actual calculation of $P((u,v)|D)$ is according to Equations (39) – (42). According to Equation (31), $P((u,v)|D)$ is inversely proportional to the code pair importance $CPI_k(u,v)$ and measures the probability of code pair (u,v) in the context of the entire database. Therefore, if a code pair (u,v) happens frequently in the database, it plays a less important role in representing the user's query intention or the

desired semantic concept. In fact, the reciprocal of $P((u,v)|D)$ is analogous to the inverse document frequency (IDF) and inverse image frequency (IIF) [57].

8.2 Confusion Code

If the semantic inference is merely based on positive code relations, that is, those code pairs that represent the desired semantic concept, then the inference performance is significantly restricted by the performance of the vector quantisation algorithm that generates the codebook. Therefore, the definition of a confusion code pair is introduced. A code pair (u,v) is identified as a confusion code pair for the semantic concepts SC_k if it satisfies

$$P((u,v)|SC_k) \neq 0 \quad \wedge \quad P((u,v)|\neg SC_k) \neq 0, \quad (43)$$

whereby u and v are the confusion regional code and confusion context code, respectively. From the perspective of the code co-occurrence matrices, this translates into

$$\mathbf{M}_k(u,v) \neq 0 \quad \wedge \quad \mathbf{M}_k^-(u,v) \neq 0, \quad (44)$$

where the code co-occurrence matrix $\mathbf{M}_k$ is generated based on positive region set T_k , while the code co-occurrence matrix $\mathbf{M}_k^-$ is produced from negative region set NT_k provided by the user. The negative region set NT_k contains the regions that are interpreted by the user as the semantic concepts other than SC_k . The confusion code pair (u,v) describes incorrectness in the inference process of the semantic concept SC_k by highlighting code combinations that are related to other semantic concepts as well. Therefore, it is necessary to exclude the corresponding images. In order to simplify the filtering progress, a pre-filtering approach was designed to analyse the image regions, which have the confusion regional code u but are interpreted as semantic concepts other than SC_k . Although this approach does not take the confusion context code into consideration, it nevertheless improves considerably the precision of the semantic inference.

Based on Equation (43), a regional code u is identified as a confusion regional code for the semantic concept SC_k if it satisfies the condition

$$\sum_v P((u,v)|SC_k) \neq 0 \quad \wedge \quad \sum_v P((u,v)|\neg SC_k) \neq 0 \quad (45)$$

and in terms of the code co-occurrence matrices according to Equation (44)

$$\sum_v M_k(u, v) \neq 0 \quad \wedge \quad \sum_v M_k^-(u, v) \neq 0. \quad (46)$$

Once a confusion regional code u is detected, the corresponding region sets $rs_u(T_k)$ and $rs_u(NT_k)$ that contain code u are selected from the positive and negative training sets $\{T_k, NT_k\}$ and filtered.

8.3 Pre-filtering Function

The pre-filtering function is used to improve the precision of the BNRF relevance approach. For this purpose the user is required to provide a set of negative region sets NT_k in addition to the positive region set T_k . Thus, the confusion code pair can be detected based on the positive region set and negative region set.

Given that the code u is part of the codebook which was generated in the feature space represented by feature function F_i , let μ_p^u be the centre of positive region sets $rs_u(T_k)$ and μ_n^u be the centre of the negative region set $rs_u(NT_k)$ in the feature space. The distance between a region R_i^j ($C_i^l(R_i^j)=u$) and a positive centre μ_p^u is calculated as

$$d_p^u(R_i^j) = \sqrt{(F_i(R_i^j) - \mu_p^u)^T (F_i(R_i^j) - \mu_p^u)}, \quad (47)$$

and equivalently for a negative centre μ_n^u

$$d_n^u(R_i^j) = \sqrt{(F_i(R_i^j) - \mu_n^u)^T (F_i(R_i^j) - \mu_n^u)}. \quad (48)$$

If $d_p^u(R_i^j)$ is larger than $d_n^u(R_i^j)$, then the region R_i^j is closer to the negative centre and thus weaker related to the semantic concept SC_k and vice versa. Hence, given $C_i^l(R_i^j)=u$, the $P(SC_k | C_i^l(R_i^j), F_i(R_i^j), T_k, NT_k)$ is estimated by the pre-filtering function h

$$h = \begin{cases} 1 & C_i^l(R_i^j) \notin CC \\ 1 & d_p^u(R_i^j)/d_n^u(R_i^j) \leq \theta \quad \wedge \quad C_i^l(R_i^j) \in CC \\ 0 & d_p^u(R_i^j)/d_n^u(R_i^j) > \theta \quad \wedge \quad C_i^l(R_i^j) \in CC \end{cases} \quad (49)$$

where CC identifies the set of confusion codes and θ representing an experimentally determined threshold.

Based on this probability, Equation (37) for the semantic score is redefined as

$$SI_k(I_t) = \frac{1}{|I_t|} \sum_{R_i^j \in I_t} P(SC_k | C_i^l(R_i^j), F_i(R_i^j), T_k, NT_k) \cdot SR_k(R_i^j), \quad (50)$$

whereby the semantic score of image I_t takes no account of the regions that are predicted to be unrelated to the respective semantic concept. Then the PBNRF approach is conducted based on Equation (50).

8.4 Composite Relevance Feedback Approach

RF approaches have shown their effectiveness and efficiency in learning the user's query intentions and gradually improve the retrieval performance. Most of them are developed for the image retrieval systems that use global descriptors. One major problem is observed in applying these approaches to remote sensing imagery retrieval systems, that is, the database has to utilise local image descriptors instead of the commonly used global information vectors due to the complexity of the stored satellite images. Recently, RBRF approaches have been successfully developed to capture the user's query intentions in the RBIR systems. However, these approaches use the computational expensive similarity measurements that measure image-to-image similarity based on all segmented regions within the images. Thus, these region-based approaches have difficulty in achieving real-time response especially for large image databases, for example, remote sensing databases. Therefore, this research proposes a CRF approach to obtain real-time response in remote sensing databases. The basic idea of the CRF approach is that the BNRF approach is used to pre-select a set of candidate images, which is much smaller than the original database and highly related to the user's query intention. Then, a more expensive RBRF approach is conducted on the candidate images to reduce the response time and increase retrieval precision. In this research, the RBRF approach [56] [57] proposed by Jing, et al., is combined with the BNRF approach. Based on the assumption that every region could be helpful in retrieval, Jing, et al. combine all of the regions of the initial query image and the positive examples into a region set. The region set forms a pseudo image that is used as the optimal query image at the next iteration of retrieval and feedback process. Since more positive examples accumulate after a number of feedback iterations, the number of the regions in the pseudo image increases rapidly and the retrieval speed gradually slows down. In order to solve this problem, the regions in the pseudo image are compressed into fewer larger regions using the k -means algorithm. However, the k -means does not work very well if the number of the regions in the pseudo image is not very large. Moreover, the important information may be lost during the

compression, since the k -means algorithm causes a precision drop [57]. Therefore, instead of using the k -means algorithm, the CRF approach only uses the BNRF approach to reduce the response time. Jing, et al., uses the earth mover's distance (EMD) to measure the similarities between the pseudo query image and the stored images. To further reduce the computational cost, the CRF approach uses the IRM algorithm to measure the image-to-image similarities. The proposed CRF approach is flexible since it can be assembled with any other RBRF approaches.

8.5 Results and Assessment

In the experiments, 26 multispectral satellite scenes, acquired by IKONOS in 2003, were utilised. Each scene was rotated by 16 different angles and the interim results were cut into image tiles. After removing unsuitable image tiles, that is, subimages that incorporated undefined areas resulting from the resampling process, a total set of 25'827 sub images with a size of 300×300 pixel were available for the database. Note that the word “image” in the experiments refers to the “sub image”. The dataset covers more than 37'190 km² with a variety of different landscape types, while the rotated versions of the original data ensure that a sufficiently large amount of similar images exists and can also be used to test the proposed approaches are invariant to the rotations.

Prior to the retrieval experiments, the stored images were manually classified into seven semantic concepts, that is, cloud (981), water (7'913), forest (4'398), urban areas (7'781), farmland (1'873), bare soil (1'750) and rock (1'328). The numbers in brackets indicate how many related images were associated with the corresponding semantic, whereby a given image may belong to multiple semantics. This manual classification provides the foundation against which individual retrieval results can be benchmarked. The evaluation was carried out in terms of $P=|A \cap Z|/|A|$, the number of correct images $N=|A \cap Z|$, and recall rate $R=|A \cap Z|/|Z|$, where A and Z are the sets of automatically and manually classified images, respectively, belonging to a certain semantic concept.

The experimental prototype is implemented in MATLAB 6.5 and runs on a NEC Powermate with an Intel Pentium IV 2.66GHz CPU, 1 GB memory. The platform is Windows TH 2000 server. All the experiments are performed by the author.

8.5.1 BNRF and PBNRF Approaches

In this experiment the spectral codebook was chosen to contain 300 codes and the provided query included five remote sensing images for each of the seven selected semantics. The retrieval performance was tested based on five query sessions, each of which was started with one image in the query set and spanned three iterations of user-system interaction. Table IV shows the precision of the BNRF approach for each semantic concept, whereby the depicted values were averaged over the five individual query sessions. It can be seen that the precision of the different semantic concepts changed irregularly depending on the iteration. On average, a 3% decrease is observable throughout the iterative process. However, the number of correct images N increased steadily, which means more and more qualified images were retrieved, that is, a 619% increase from 42 images at the first iteration to 260 images at the last iteration were achieved. This phenomenon is caused by the iteratively increasing number of non-zero elements in the code co-occurrence matrix. The new non-zero elements are introduced by the newly added images with semantic scores greater than zero. However, these additional images result in instable precision results, since the number of qualified images in the additionally incorporated images is affected by the overall image distribution in the database. In general, the minor drop in precision with simultaneous increase of correctly retrieved images was observed consistently. Therefore, the BNRF approach without pre-filtering is suitable during the first stage of a retrieval process. Once the user is satisfied with the number of correct images, tougher constraints help to improve the precision.

Table IV: Precision and recall rates using BNRF

Semantic concept	Initially		1 st iteration		2 nd iteration		3 rd iteration	
	P	N	P	N	P	N	P	N
Cloud	0.64	26	0.55	54	0.49	84	0.46	111
Water	0.94	12	0.84	75	0.94	221	0.95	273
Forest	0.63	20	0.77	114	0.78	204	0.74	244
Buildings	0.86	46	0.87	122	0.83	211	0.85	346
Farmland	0.84	89	0.76	174	0.75	231	0.72	276
Soil	0.84	46	0.86	186	0.83	248	0.83	339
Rock	0.68	53	0.72	105	0.71	169	0.67	234
Average	0.78	42	0.78	119	0.76	195	0.75	260

For the assessment of the PBNRF approach with pre-filtering, the experimental query set was increased to 350 images, i.e. each semantic concept corresponded to 50 examples. The pre-filtering threshold θ was set to 1.0, 0.85 and 0.70 for the respective iterations, whereby the monotonic decrease was chosen in order to improve the rate of convergence. Again five query sessions were conducted with ten images as starting point in each query set. During each of the following iterations additional four user-identified negative images from the retrieval result were incorporated, while no new positive images were included. Table V shows the average precision values and numbers of correctly retrieved images for different semantic concepts for the five consecutive query sessions. Table V shows that the precision for different semantic concepts increases steadily with increased iteration number. On the average, a 15% increase is observable for P , and the precision exhibits the greatest increase in the first iteration. In the following iterations only minor improvements are observable. This characteristic is of advantage for the user who receives an instantaneous feedback and hence can easily assess whether the query image is suitable. Moreover, in no case could a drop in precision with increased iteration number be observed. Thus, the approach depicts a stable performance. Comparing the increase of P , N averagely decreases amounts to 20% over the three iterations.

Table V: Precision and recall rates using PBNRF

Semantic concept	Initially		1 st iteration		2 nd iteration		3 rd iteration	
	P	N	P	N	P	N	P	N
Cloud	0.35	36	0.49	33	0.57	27	0.65	24
Water	0.95	160	0.99	140	0.99	132	0.99	121
Forest	0.61	197	0.70	186	0.74	172	0.77	155
Buildings	0.82	261	0.88	246	0.90	227	0.93	191
Farmland	0.68	331	0.74	325	0.81	308	0.86	283
Soil	0.85	409	0.88	407	0.91	391	0.95	328
Rock	0.64	296	0.68	290	0.75	280	0.81	241
Average	0.70	338	0.77	325	0.81	307	0.85	269

Figure 21: Performance of BNRF and PBNRF: (a) Average precision, (b) Average number of correct images, (c) Precision vs. Number of correct images

From Figure 21, it is found that the BNRF approach and the PBNRF approach complement each other very well. From one perspective, the BNRF approach is able to increase the number of corrected images with increased iteration number. From the other perspective, the PBNRF approach is able to improve the retrieval precision iteratively. Therefore, both approaches can be cooperatively used to improve both precision and recall rate.

Table VI and Figure 22 show that the BNRF approach is slightly faster than the PBNRF approach. Averagely, the PBNRF spends 0.32s more than the BNRF does. This is a desirable characteristic for large remote sensing archives since the response time is not affected too much by the pre-filtering function. On the other hand, the improvement of the retrieval precision is more and more significant as the number of

iterations increases. Therefore, the pre-filtering function is not only important but also necessary to be the augmentation of the BNRF approach.

Figure 22: Average retrieval time comparison of BNRF and PBNRF

Table VI: Retrieval time comparison between BNRF and PBNRF

Semantic concept	Initially		1 st iteration		2 nd iteration		3 rd iteration	
	BNRF	PBNRF	BNRF	PBNRF	BNRF	PBNRF	BNRF	PBNRF
Cloud	1.92	1.91	1.91	2.24	1.89	2.25	1.91	2.23
Water	1.83	1.83	1.81	2.14	1.84	2.09	1.83	2.11
Forest	1.83	1.81	1.83	2.16	1.81	2.20	1.81	2.22
Buildings	1.92	1.92	1.89	2.28	1.91	2.18	1.92	2.20
Farmland	1.83	1.80	1.78	2.17	1.93	2.21	1.80	2.19
Soil	1.84	1.77	1.83	2.16	1.80	2.20	1.77	2.22
Rock	1.84	1.81	1.83	2.13	1.81	2.14	1.81	2.11
Average	1.86	1.84	1.84	2.18	1.86	2.18	1.84	2.18

8.5.2 Composite Relevance Feedback Approach

In this experiment the spectral codebook was chosen to contain 300 codes and the provided query included five remote sensing images for each of the seven selected semantics. The retrieval performance was tested based on five query sessions, each of

which was started with one image in the query set and spanned three iterations of user-system interaction. In order to reveal the effect of the pre-selection stage in the CRF approach, the query images are selected from the stored images that have low top-100 precisions when the IRM algorithm is used to rank the similarities. Table V and Figure 23 show that the pre-selection stage in the CRF approach considerably improve the precision of the original RBRF approach. On average, the precision of the CRF approach is 12.5% higher than that of the RBRF approach. Moreover, Table VIII and Figure 23 show that the CRF approach is much faster than the RBRF feedback approach. On average, the CRF approach is 58s faster than the RBRF approach. In particular, the retrieval time of the CRF approach is one eleventh of that of the RBRF approach at the last iteration. Considering the precision improvement and the time reduction caused by the pre-selection stage, the CRF approach is very useful especially in large remote sensing databases.

Table VII: Precisions of top-100 images using the CRF and the RBRF.

Semantic concept	Initially		1 st iteration		2 nd iteration		3 rd iteration	
	CRF	RBRF	CRF	RBRF	CRF	RBRF	CRF	RBRF
Cloud	0.45	0.31	0.50	0.33	0.52	0.34	0.54	0.38
Water	0.86	0.38	0.88	0.61	0.91	0.73	0.93	0.74
Forest	0.61	0.38	0.66	0.55	0.72	0.65	0.78	0.73
Buildings	0.67	0.68	0.83	0.85	0.86	0.87	0.91	0.92
Farmland	0.56	0.41	0.68	0.52	0.70	0.60	0.72	0.65
Soil	0.92	0.69	0.96	0.93	0.96	0.97	0.97	0.97
Rock	0.64	0.44	0.68	0.56	0.75	0.66	0.81	0.70
Average	0.67	0.47	0.74	0.62	0.77	0.69	0.80	0.73

Figure 23: Retrieval performance of CRF and RBRF:
(a) Average precision, (b) Average retrieval time

Table VIII: Retrieval time comparison between the CRF and the RBRF feedback

Semantic concept	Initially		1 st iteration		2 nd iteration		3 rd iteration	
	CRF	RBRF	CRF	RBRF	CRF	RBRF	CRF	RBRF
Cloud	0.86	16.28	2.67	50.38	4.59	87.21	6.78	126.06
Water	1.45	12.49	4.22	36.80	7.26	63.67	10.49	92.23
Forest	1.70	15.06	5.41	47.37	8.91	80.01	12.29	111.21
Buildings	2.20	15.17	5.93	45.34	10.08	76.20	14.26	109.15
Farmland	0.98	13.82	3.52	50.60	5.90	84.97	8.31	119.61
Soil	0.95	15.07	3.00	47.84	5.27	83.90	7.48	118.82
Rock	1.24	15.71	3.21	45.87	5.59	80.95	7.91	114.43
Average	1.34	14.80	3.99	46.31	6.80	79.56	9.65	113.07

8.5.3 Heterogeneous Test

In the experimental database, the images with the same semantic concept often come from different scenes, which are captured at various locations of the Earth, that is, cloud (3 scenes), forest (10 scenes), farm (2 scenes), building (14 scenes) and water (17 scenes). This is a challenge situation for the RBRF approach. The reason is that the IRM algorithm is quite efficient in measuring the similarities between the pseudo query image and the stored images if the pseudo query image is formed by a set of visually similar and semantically relevant positive examples. However, the semantically related positive examples are not always visually similar. For instance,

Figure 24 shows the water images from 17 scenes, which include sea, river and lake and have different spectral and textural characteristics. Therefore, if the user provides a set of visually different positive examples as shown in Figure 24, the performance of the RBRF approach is limited. Inversely, the BNRF approach is designed to interactively learn the semantic score function, which is a special case of the context sensitive Bayesian network. Although the visually different regions are encoded as different codes, the context-sensitive Bayesian network is able to link these codes with the same semantic concepts. Therefore, the BNRF approach is more robust to the positive images from different scenes. In order to test this hypothesis, a heterogeneous test is performed to measure the extent to which the BNRF approach and the RBRF approach are robust to the positive examples from various scenes.

In the heterogeneous test, the tree-structure codebook was chosen and contains 300, 600, 1200 codes, resp., at the three different levels. In order to simulate the situation when positive images come from different scenes, the number of positive images equals to the number of scenes, that is, n positive images comes from n different scenes. The number of the positive images n is set to 2, 3, 4 and 5, respectively. For each choice of n , the BNRF approach and the RBRF approach were performed with five different queries for each of the seven selected semantics. The performance of BNRF depends on the codebook size, since this is a parameter that describes how many different semantic concepts can be inferred at specific granularity. In order to measure the direct dependency on a predefined size choice, the BNRF approach was carried out based on each of the three levels of the tree-structured codebook. Figure 25 and Table VII show that on the average the precisions of the BNRF approach are significantly higher than that of the RBRF approach for all three levels of the codebook. In particular, the precision of the BNRF approach is 19.5% higher than that of the RBRF approach when the codebook size is 1200. This shows that the BNRF is more robust than the RBRF in dealing with positive examples from different scenes. In order to further establish this robustness, the other heterogeneous test is performed. In this test, the number of the positive images belonging to a semantic concept equals to the number of the scenes that are related to the semantic concept. Every positive image comes from different scenes. Five query sessions are carried out for every semantic concept.

Table IX shows that on average the precision of the BNRF approach is 43% higher than that of the RBRF approach. Comparing to the previous result, this result illustrates that the advantage of the BNRF becomes increasingly significant with the number of scenes. Therefore, the BNRF can be used to perform long-term learning if the feedback examples are accumulated from different query sessions and are provided by different users. The reason is that the accumulated feedback examples often have very inhomogeneous signal characteristics.

Figure 24: Water body in different scenes

Table IX: Precisions of BNRF /RBRF approaches for different scenes

Semantic concept c	Number Of Scenes	BNRF			RBRF
		Level 1 (300 codes)	Level 2 (600 codes)	Level 3 (1200 codes)	
Cloud	3	0.56	0.64	0.88	0.29
Water	17	0.95	0.97	0.98	0.44
Forest	10	0.66	0.74	0.77	0.67
Buildings	14	0.83	0.88	0.94	0.70
Farm	2	0.78	0.91	0.89	0.41
Average	9	0.76	0.83	0.89	0.50

Figure 25: Precisions of the BNRF/.RBRF approaches with positive images from different sources

Table X: Precisions of BNRF/RBRF with positive images from different scenes

Semantic concept	Code number	2 Scenes		3 Scenes		4 Scenes		5 Scenes	
		BNRF	RBRF	BNRF	RBRF	BNRF	RBRF	BNRF	RBRF
Water	300	0.99	0.79	0.97	0.72	0.98	0.59	0.89	0.67
	600	1.00		0.98		0.98		0.92	
	1200	1.00		0.99		0.99		0.97	
Forest	300	0.67	0.80	0.72	0.75	0.77	0.77	0.75	0.73
	600	0.86		0.84		0.84		0.81	
	1200	0.94		0.87		0.87		0.88	
Buildings	300	0.91	0.83	0.84	0.61	0.81	0.73	0.76	0.69
	600	0.92		0.88		0.87		0.83	
	1200	0.94		0.82		0.90		0.86	
Average	300	0.86	0.81	0.84	0.69	0.85	0.70	0.80	0.70
	600	0.93		0.90		0.90		0.85	
	1200	0.96		0.90		0.92		0.90	

8.5.4 Verification Experiments

In order to prove that the proposed algorithm can retrieve more than the rotated versions of the query images, the experiments were also performed on the second experimental database. In this database, 28 multispectral satellite scenes acquired by IKONOS were utilised. Each scene was cut into overlapping sub-images, with a non-overlapping proportion of at least 30%. A total set of 14'874 sub-images with a size of 300×300 pixel were available for the database. Note that the word “image” in this experiment refers to “sub-image”. The overlapping sub-images ensure that a sufficiently large amount of similar images exists in the database. Prior to the actual retrieval experiment, the stored images were manually classified into seven semantic concepts, i.e. cloud (598), water (3'919), forest (3'085), urban areas (4165), farmland (764), bare soil (991) and rock (628).

In this experiment the spectral codebook was chosen to contain 300 codes and the provided query included five remote sensing images for each of the seven selected semantics. The retrieval performance was tested based on five query sessions, each of which was started with one image in the query set and spanned three iterations of user-system interaction. Table XI shows the precision of the BNRF approach for each semantic concept, whereby the depicted values were averaged over the five individual query sessions. It can be seen that the precision rates of the different semantic concepts changed irregularly depending on the actual iteration. On average, a 5%

decrease is observable throughout the iteration process. However, the number of the correctly retrieved images N increased steadily. On average, N increased 535% from 33 images in the first iteration to 166 images in the last iteration.

For the assessment of the PBNRF approach with pre-filtering, the experimental query set was increased to 280 images, where each semantic concept corresponds to 40 examples. The pre-filtering threshold θ was set to 1.0, 0.85 and 0.70 for the respective iterations, whereby the monotonic decrease was chosen in order to improve the rate of convergence. Again five query sessions were conducted with ten images as starting point in each query set. During each of the following iterations additional four user-identified negative images from the retrieval result were incorporated, while no new positive images were included. Table XII shows the average precision values and numbers of correctly retrieved images for different semantic concepts for the five individual query sessions. Table XII depicts that the precision for different semantic concepts increases steadily with increased iteration number. On average, a 7% increase is observable for P . Comparing the increase of P , N suffers from only a very moderate drop. The average decrease amounts to 39% over the three iterations.

The results are similar to the results obtained based on the first experimental database.

Table XI: Precision and recall rates using BNRF

Semantic concept	Initially		1 st iteration		2 nd iteration		3 rd iteration	
	P	N	P	N	P	N	P	N
Cloud	0.50	28	0.55	59	0.50	96	0.47	134
Water	1.00	24	1.00	80	1.00	203	0.99	331
Forest	0.72	58	0.80	140	0.81	179	0.81	196
Buildings	0.95	19	0.74	39	0.72	58	0.88	170
Farmland	0.90	36	0.83	57	0.72	85	0.73	107
Soil	0.95	21	0.83	58	0.78	98	0.73	146
Rock	0.61	28	0.59	34	0.63	60	0.65	75
Average	0.80	31	0.76	67	0.73	111	0.75	166

Table XII: Precision and recall rates using PBNRF

Semantic concept	Initially		1 st iteration		2 nd iteration		3 rd iteration	
	<i>P</i>	<i>N</i>	<i>P</i>	<i>N</i>	<i>P</i>	<i>N</i>	<i>P</i>	<i>N</i>
Cloud	0.47	134	0.52	123	0.53	111	0.55	100
Water	0.92	823	0.95	525	0.97	497	0.99	385
Forest	0.87	1017	0.89	996	0.89	988	0.92	828
Buildings	0.85	911	0.87	832	0.88	641	0.91	502
Farmland	0.70	321	0.71	306	0.75	303	0.81	244
Soil	0.77	584	0.79	349	0.81	311	0.85	210
Rock	0.58	303	0.58	266	0.59	250	0.62	222
Average	0.74	585	0.76	485	0.77	443	0.81	356

Chapter 9

Long-Term Learning

This chapter introduces a LTL approach and an associative LTL approach, both of which are able to improve the retrieval performance based on accumulated users' feedbacks.

9.1 Long-term Learning via Context-Sensitive Bayesian Network

The proposed RF approaches learn the necessary knowledge from the user-supplied samples and use the obtained knowledge to bridge the semantic gap. They are able to effectively and efficiently improve the retrieval performance during a single query session. However, the proposed RF approaches totally discard the obtained knowledge after each query session. Thus, they cannot memorise, reuse, and communicate the obtained knowledge. Such short-term learning behaviour has two disadvantages. Firstly, a user has to repeat the same tedious feedback iterations just to get identical results he/she has obtained before [5]. Secondly, users with similar retrieval interest as other users cannot utilise these users' retrieval experiences to satisfy their query intention. Therefore, in this research a LTL approach was designed to gradually improve the retrieval performance through the knowledge accumulated from the previous users' interactions. Compared to the RF approaches, the LTL approach is performed based on the feedback examples, which are related to the same semantic concept, but more likely to have very inhomogeneous signal characteristics. This is caused by the fact that the feedback examples are supplied by different users, who may possess very different viewpoints with regard to the same semantic concept.

In order to solve the problem, the proposed LTL approach trains a set of semantic score functions based on inhomogeneous feedback examples. In the last chapter, it has been shown that the semantic score functions are robust against the feedback examples that possess heterogeneous signal characteristics. Therefore, like the BNRF approach, the LTL approach trains the semantic score function according to Equation

(41). However, unlike the BNRF approach, it trains the semantic score function of the semantic concept SC_k based on the n different sets of positive examples $\{T_k^1, \dots, T_k^j, \dots, T_k^n\}$, which are provided by n different users respectively. Therefore, the accumulated positive example T_k is the union of the n different sets of positive feedback examples

$$T_k = \bigcup_{j=1}^n T_k^j. \quad (51)$$

Based on the positive examples T_k , the code co-occurrence matrix $\mathbf{M}_k$ is generated according to Equations (39)–(42) and provides the necessary information to calculate the semantic scores with respect to the semantic concepts SC_k . From the above description, it is found that there are no differences between the mechanisms of the LTL approach and the BNRF approach. However, the BNRF approach is based on a small number of positive examples while the LTL approach has to deal with a large number of inhomogeneous positive examples. According to the experimental results, with the increased number of positive examples, the proposed LTL approach enables a significantly increased recall rate at the cost of a slight decreased precision. This indicates that the precision is consistently decreased by the accumulated quantisation errors of the additional positive examples if a large number of positive examples are supplied. In order to increase the precision, the LTL approach is extended by an associative learning strategy in the next subsection.

9.2 Associative Long-Term Learning

The basic idea of LTL is that a user can find his/her desirable images based on the feedback examples provided by other users who had the same query intention. However, the proposed LTL approach trains the semantic score functions merely based on positive code relations, that is, those code pairs that represent the desired semantic concept SC_k . Thus, the semantic inference performance is significantly restricted by the performance of the vector quantisation algorithm that generates the codebook. In order to mitigate restriction, this research proposed an associative long-term learning (ALTL) approach based on the original LTL approach.

In addition to the positive training set of the semantic concept SC_k , there exist accumulated positive training sets that represent semantic concepts other than SC_k .

These training sets can act as negative or inverse training sets for the semantic concepts SC_k if their semantic concepts are orthogonal to the semantic concept SC_k . The major advantage of the ALTL approach is that the accumulated user-supplied knowledge is fully exploited to improve the performance of the semantic inference. In this research, the pre-filtering function is used by the ALTL approach to synthesis the accumulated knowledge. Accordingly the confusion code is redefined based on Equations (43)–(44): a code pair (u,v) is identified as a confusion code pair for the semantic concepts SC_k if it satisfies

$$P((u,v)|SC_k) \neq 0 \quad \wedge \quad \exists SC_j, j \neq k : P((u,v)|SC_j) \neq 0. \quad (52)$$

From the perspective of the code co-occurrence matrices, this translates into

$$M_k(u,v) \neq 0 \quad \wedge \quad \exists M_j, j \neq k : M_j(u,v) \neq 0. \quad (53)$$

The confusion code pair (u,v) describes incorrectness in the inference process of the semantic concept SC_k by highlighting code combinations that are related to other semantic concepts as well. Therefore, it is necessary to exclude the corresponding images. In order to simplify the filtering progress, a pre-filtering function is used to detect the image regions, which have the confusion regional code u but are interpreted as semantic concepts other than SC_k . Although the pre-filtering function does not take the confusion context code into consideration, it nevertheless improves the precision of the semantic inference considerably.

Based on Equation (52), a regional code u is identified as a confusion regional code for the semantic concept SC_k if it satisfies the condition

$$\sum_v P((u,v)|SC_k) \neq 0 \quad \wedge \quad \exists SC_j, j \neq k : \sum_v P((u,v)|SC_j) \neq 0 \quad (54)$$

and in terms of the code co-occurrence matrices according to Equation (53)

$$\sum_v M_k(u,v) \neq 0 \quad \wedge \quad \exists M_j, j \neq k : \sum_v M_j(u,v) \neq 0. \quad (55)$$

Once a confusion regional code u is detected, the corresponding region sets $\{rs_u(T_1), \dots, rs_u(T_N)\}$ that contain code u are selected from the training sets $\{T_1, \dots, T_N\}$ and filtered.

Given that the code u is part of the codebook which was generated in the feature space represented by feature function F_i , let μ_p^u be the centres of positive region sets $rs_u(T_k)$ and μ_n^u be the centre of the negative region set $rs_u(NT_k)$ in the feature space, i.e.

$$NT_k = \bigcup_{j \neq k} T_j. \quad (56)$$

Then, the positive centre and negative centre are used to exclude the regions - which are predicted to be unrelated to the respective semantic concept according to the calculation of the semantic scores based on Equations (47) –(50).

9.3 Results and Assessment

The experimental data set is identical to the scenario described in Chapter 8 in order to enable cross-comparison between the approaches.

9.3.1 Comparison of SemQuery and Long-Term Learning Approach

In this performance evaluation the proposed LTL approach was compared with the well-known SemQuery system [94], [95]. The reason for this choice is that SemQuery was used for the retrieval of satellite imagery as well, employs a similar semantic concept understanding. [23], [96].

SemQuery categorises the remotely sensed imagery directly in the space of the extracted features. For a balanced performance assessment of the semantic inference in this low-level feature space and the proposed LTL approach, the SemQuery system was modified to classify images based on regional features instead of the original global feature. This enables SemQuery to classify images into more than one semantic concept. The actual feature extraction is identical for both systems.

The evaluation was carried out in terms of precision $P=|A \cap Z|/|A|$ and recall $R=|A \cap Z|/|Z|$, whereby A and Z are the sets of automatically and manually classified images, respectively, belonging to a certain semantic concept. Note that this evaluation does not attempt to assess the suitability of SemQuery versus the LTL approach but rather the benefits and disadvantages of using different approaches and parameterisations to infer the images' semantics. Moreover, the experimental results [95] show that SemQuery achieves good performance when the size of the training set is approximately 20% of the entire database. However, the size of the training set in the conducted experiment was less than 2% of the entire database, which is closer to reality – especially for large remote sensing databases.

In the first retrieval experiment only spectral features were used to assess the strategies of the two approaches. The objective behind this restriction is to investigate how much this feature (which is very dominant) for the human perception contributes towards the performance of automated retrieval. The results are presented in Table XIII, whereby the LTL approach used a codebook containing 300 codes, that is, no tree-structured vector quantisation was utilised. For SemQuery the columns show the mean μ and the standard deviation σ^2 of the precision and recall rates, which were collected over five separately conducted training and retrieval sessions to reduce a possible bias. This averaging over individual sessions is required due to SemQuery's dependency on the initial cluster seeds for the semantic inference. In contrast, this is not needed for the LTL approach since it utilises the initial feature vector space indirectly and produces identical retrieval results each time.

The results in Table XIII show that for all but one case the precision and recall rates of the proposed LTL approach are significantly higher than the mean values achieved by SemQuery. This is mainly due to the difficulty to determine the number of subclusters as well as the extent of each subcluster in the feature space. Several parameter choices were tested and the best results selected for the depicted comparison.

The semantic meaning of a satellite scene does not solely depend on the spectral characteristic of the image, but also on the textures found within homogenous regions. Therefore, in the following experiment the semantic meaning of individual regions is inferred in a separate and combined approach, respectively, for both LTL and SemQuery. To obtain the semantic meanings, codebooks with 300 codes for the spectral and texture features were generated individually, and then inference was performed. For the joint features, the inference results based on the spectral and textural codebooks were intersected. The results of the experiment are shown in Table X, where the columns describing the performance of the spectral features coincide with the corresponding columns in Table XIII and are shown for comparison. For SemQuery, only the mean values are shown, which again were computed over five separately conducted training and retrieval sessions.

The initial assumption that the spectral information is most significant can be validated from the values shown in Table X. Both the precision and the recall rates outperform the achievements by the texture features for SemQuery and the LTL

approach. However, when the inference results based on two types of codebooks were intersected then the precision is improved and superior to utilisation of each feature type separately. This is expected since a more precise description of the queried scene can be derived. As a disadvantage the recall performance decreased significantly. This is due to the well-known phenomenon that with more restricted constraints the accessible scope becomes narrower. Hence, the impact of imprecisely and / or wrongly classified images turns out to be more significant.

Table XIII: Mean and standard deviations for precision and recall rates over five separate training and retrieval sessions for semantic classification based on spectral features

Semantic concept c	No. of training regions	Precision P			Recall R		
		SemQuery		LTL	SemQuery		LTL
		μ	σ^2		μ	σ^2	
Cloud	107	0.06	0.06	0.38	0.07	0.09	0.43
Water	198	0.58	0.11	0.93	0.12	0.10	0.32
Forest	139	0.44	0.13	0.65	0.16	0.07	0.57
Urban area	202	0.63	0.12	0.77	0.25	0.14	0.30
Farmland	159	0.58	0.30	0.57	0.21	0.11	0.49
Soil	165	0.51	0.22	0.79	0.20	0.07	0.63
Rock	170	0.31	0.08	0.54	0.28	0.12	0.63
Average	163	0.44	0.15	0.66	0.18	0.18	0.48

Table XIV: Precision and recall rates over five separate training and retrieval sessions for semantic classification based on spectral and texture features

Semantic concept	Spectral features				Texture features				Spectral and texture features			
	SemQuery		LTL		SemQuery		LTL		SemQuery		LTL	
	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>
Cloud	0.06	0.07	0.38	0.43	0.15	0.13	0.22	0.35	0.42	0.01	0.50	0.19
Water	0.58	0.12	0.93	0.32	0.86	0.11	0.83	0.09	0.82	0.04	0.97	0.05
Forest	0.44	0.16	0.65	0.57	0.53	0.12	0.62	0.14	0.63	0.03	0.83	0.08
Building	0.63	0.25	0.77	0.30	0.49	0.24	0.54	0.13	0.62	0.08	0.82	0.06
Farm	0.58	0.21	0.57	0.49	0.18	0.18	0.26	0.27	0.61	0.06	0.85	0.20
Soil	0.51	0.20	0.79	0.63	0.15	0.32	0.45	0.20	0.61	0.09	0.94	0.14
Rock	0.31	0.28	0.54	0.63	0.20	0.21	0.29	0.31	0.50	0.11	0.71	0.23
Average	0.44	0.18	0.66	0.48	0.37	0.19	0.46	0.21	0.60	0.06	0.80	0.14

9.3.2 Simple Bayesian Network and Context-Sensitive Bayesian Network

In this performance evaluation the proposed LTL approach based on the context-sensitive Bayesian network was compared with a LTL approach based on a simple Bayesian network. For simplicity, the former LTL approach is denoted as CSBN (Context-Sensitive Bayesian Network) while the latter LTL approach is denoted as SBN (Simple Bayesian Network). The motivation for comparing the two techniques is that simple Bayesian network has already been used to capture the semantics in several remote sensing information miming systems, for example, the ISIMS system [26], [27], VisiMine [61] and the KIM system [32], [28]. Since these systems utilise the features extracted by Gibbs-Markov Random Field models [29], [30], the extracted features incorporate sufficient spatial information. However, if features without or with little spatial information are used, in our case, a colour histogram, we claim that the performance of the simple Bayesian network is limited. Therefore, we compare the performance of the CBSN to that of the SBN to demonstrate the need for incorporating spatial information during the inference process. The comparisons were conduct based on different codebook levels, that is, 300, 600 and 1200 codes. Moreover, all comparisons are based on the same number of feedback samples. For

each semantic concept, a remote sensing image is related to a particular semantic concept if the corresponding semantic score is larger than zero. Table XV shows that for all cases the precision of the CSBN is significantly higher than that of the SBN. We interpret this different as the effect of incorporating spatial information. However Table XII shows that for all cases the recall rates of the CSBN are significantly lower than those of the SBN. This indicates that the CSBN improves the precision at the cost of reducing recall rates. Plotting the averaged data in Figure 26 illustrates that the CSBN and the SBN have similar overall trends with regard to precision and recall rates, that is, the precision is monotonically increasing and the recall rate is monotonically decreasing with the codebook size. From this observation, it can be predicted that the precision of the SBN can be improved by increasing the number of codes as well. However, the complexity of the GLA algorithm is $o(kd)$, whereby k is the number of codes and d is the dimension of low-level features. Thus, for large remote sensing databases, an increase of k results in high computational costs especially for high dimensional feature vectors. This has only an insignificant impact on the CSBN since it requires fewer codes to achieve high precision.

Table XV: Precisions of the semantic inference based on three levels of the tree-structured codebook using simple and context-sensitive Bayesian networks

Semantic concept	Level 1 (300 codes)		Level 2 (600 codes)		Level 3 (1200 codes)	
	Context	Simple	Context	Simple	Context	Simple
Cloud	0.38	0.09	0.55	0.13	0.70	0.17
Water	0.93	0.70	0.97	0.74	0.98	0.80
Forest	0.65	0.37	0.76	0.45	0.82	0.52
Urban areas	0.77	0.42	0.86	0.48	0.92	0.57
Farmland	0.57	0.15	0.73	0.20	0.82	0.29
Soil	0.79	0.16	0.85	0.23	0.89	0.31
Rock	0.54	0.12	0.73	0.16	0.84	0.22
Average	0.66	0.29	0.77	0.34	0.85	0.41

Figure 26: Retrieval performance of CSBN and SBN at three levels of the tree-structured codebook: (a) Average precision, (b) Average recall

Table XVI: Recall rates of the semantic inference based on three levels of the tree-structured codebook using simple and context sensitive Bayesian networks

Semantic concept c	Level 1 (300 codes)		Level 2 (600 codes)		Level 3 (1200 codes)	
	Context	Simple	Context	Simple	Context	Simple
Cloud	0.43	0.76	0.32	0.71	0.21	0.64
Water	0.32	0.90	0.21	0.79	0.14	0.63
Forest	0.57	0.88	0.23	0.83	0.14	0.68
Urban areas	0.30	0.94	0.18	0.85	0.10	0.70
Farmland	0.49	0.72	0.34	0.70	0.21	0.68
Soil	0.63	0.99	0.34	0.99	0.16	0.94
Rock	0.63	0.99	0.43	0.98	0.26	0.95
Average	0.48	0.88	0.29	0.84	0.17	0.75

9.3.3 Significance of Adjacent Regions

In the utilised database, about thirty percent of the images contain urban areas, which exhibit complex spectral and textural patterns. Thus, these areas are very likely to be decomposed into a set of separate regions at the segmentation stage. Due to the different signal characteristics of these regions, the corresponding feature vectors may be distributed far away in the feature space. For the SemQuery system, this increases the difficulty of capturing the clusters of the urban areas in the feature space. Moreover, regions belonging to the same cluster may have different semantic concepts other than

urban. For example, a bright / saturated image region may be a part of a technical installation on the rooftop of a skyscraper, a cloud or snow. These two problems of separation on the one hand and combination on the other hand greatly reduce the precision and recall rates in urban areas. The proposed LTL approach solved the problem by inferring semantic concepts based on the region and its contexts. From Table X, if both the spectral and textural features are used to retrieve urban areas, the precision rates of LTL are 20% higher than those of the SemQuery system. For ground cover types that have simpler appearances in certain spectral bands, for example, water and forest, the LTL approach also achieved good results, which are generally superior to the SemQuery system. This indicates that the adjacent regions of a region play an important role in inferring the semantic concept.

9.3.4 Learning Ability

The most time consuming step for both the SemQuery system and LTL is the training of the semantic concepts. Thus, the impact of the size of the feedback image set on the retrieval performance was investigated. Figure 27 shows the precision and recall rates, based on 100, 200, 300, 400 and 500 arbitrarily selected feedback images that all contain forest.

In general SemQuery exhibits only a minor drop in precision with an increasing number of feedback images while the recall rate is almost unaffected. The latter aspect indicates that the approach does not exploit the additionally provided information efficiently since the overall rate is relatively low.

In comparison, when the codebook of 300 codes is utilised, the LTL approach achieves higher precision and recall rates than the SemQuery. This is due to the utilisation of codebook and integrating context information in the semantic inference process. The recall rate substantially increases with the size of the training set, at the expense of a minor reduction of the precision. This indicates that the LTL approach exploits the large variety of forest types in the training set better than SemQuery does. The reason is that various forest types have different signal characteristics, and, thus can be clustered into different signal classes in the feature space. Contrarily SemQuery uses the k -means algorithm to detect the signal classes directly based on the low-level features of the training images. However, the accuracy of the k -means algorithm is limited if the number of the training images is small or the classes have

non-spherical shapes. The k -means algorithm cannot accurately capture the semantic scopes in the feature space when the number of the training samples is small. Comparably, the LTL approach uses the GLA algorithm to detect the signal classes based on the regional features of the entire database, thus the accuracy of the detection is higher than SemQuery. Furthermore, based on the context-sensitive Bayesian network, the LTL approach is able to establish multiple-to-multiple linkage between signal classes and semantic concepts. Therefore, LTL has better learning ability than SemQuery.

Figure 27: Retrieval performance for areas partially and totally covered by forest with

respect to the size of the training image set: (a) Precision, (b) Recall, (c) Precision vs. recall

9.3.5 Codebook Size

The retrieval performance depends on the codebook size. The codebook size influences how many different semantic concepts can be inferred and to which granularity. In order to counteract the direct dependency on a predefined size choice, the LTL approach makes use of a tree-structured codebook, as described earlier. Thereby, the utilised three levels contain 300, 600 and 1200 codes, respectively.

The experimental results are shown in Table XIII and indicate that the semantic classification based on higher levels of the codebook yields higher precision and lower recall rates. Given the same set of feedback images, the semantic classification based on a higher level of the codebook classifies a smaller number of images into the same semantic concept, which causes the recall rates to drop, since a broad semantic description cannot be inferred. On the contrary, a higher level of the codebook discriminates the spectral / texture patterns in the database better than a lower level does and, hence, yields a better precision. In order to explain why recall rate goes down with more codes, let L be the codebook size, N be the number of stored images. There will be L^2 possible code pairs. On average, each code pair is related to N/L^2 stored images. Given that the training regions for semantic concept SC_k have m distinct code pairs, then $m \times N/L^2$ images are classified by CSBN into the semantic concept SC_k . The codebook size L is inversely proportional to the number of images being classified into the semantic concept SC_k . Therefore, the more codes always lead to the lower recall rate.

Table XVII: Precision and recall rates of the LTL approach based on three levels of the tree-structured codebook

Semantic concept	Level 1 (300 codes)		Level 2 (600 codes)		Level 3 (1200 codes)	
	P	R	P	R	P	R
Cloud	0.38	0.43	0.55	0.32	0.70	0.21
Water	0.93	0.32	0.97	0.21	0.98	0.14
Forest	0.65	0.57	0.76	0.23	0.82	0.14
Urban areas	0.77	0.30	0.86	0.18	0.92	0.10

Farmland	0.57	0.49	0.73	0.34	0.82	0.21
Soil	0.79	0.63	0.85	0.34	0.89	0.16
Rock	0.54	0.63	0.73	0.43	0.84	0.26
Average	0.66	0.48	0.77	0.29	0.85	0.17

(a)

(b)

(c)

Figure 28: Retrieval of forest areas based on different codebook levels in the LTL approach in comparison with SemQuery: (a) Precision, (b) Recall, (c) Precision vs. recall

In addition to the codebook size, Figure 28 shows the learning capability for the different levels of the codebook and overlays the results obtained by SemQuery given in Figure 27. It was found that an improved precision at the cost of a lower recall rate can be obtained, if a higher codebook level is used. But it is worthwhile noting that while the precision drops slowly, the recall rate increases rapidly. Considering both factors simultaneously, semantic classification based on higher codebook levels achieves better performance if sufficient training images are provided.

9.3.6 Associative Long-Term Learning Performance

The ALTL approach accumulates the positively assessed images with respect to a certain semantic provided by the different users and uses them for the retrieval. It synthesises the positive image set from the different semantic concepts to sufficiently exploit the accumulated feedback information. The numbers of positive images and positive regions of different semantic concepts are shown in Table XVIII.

Table XV and Figure 29 compare the performances of the LTL and ALTL approaches. For the ALTL approach, the implications of different threshold values θ were investigated. The experimental results indicate that the precision of the ALTL approach is consistently higher than that of the LTL approach. In most cases, a smaller threshold θ yields a higher precision and a lower recall rate. However, the ALTL approach with a lower threshold θ does not necessarily produce a higher precision. As shown in Table XV, different semantic concepts have different optimal threshold values.

Table XVIII: Numbers of positive images and positive regions

Semantic concepts	No. of accumulated positive images	No. of accumulated positive regions
Cloud	83	107
Water	144	198
Forest	80	139
Buildings	112	202
Farmland	56	159
Soil	48	165
Rock	48	170
Average	82	163

Table XIX: Comparison of precision and recall rates over semantic inference with and without pre-filtering for semantic classification based on spectral features

Semantic concept	LTL		ALTL							
	<i>P</i>	<i>R</i>	$\theta=1.0$		$\theta=0.7$		$\theta=0.5$		$\theta=0.3$	
			<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>
Cloud	0.40	0.49	0.46	0.43	0.55	0.33	0.55	0.24	0.53	0.18
Water	0.93	0.34	0.95	0.33	0.95	0.33	0.96	0.32	0.96	0.32
Forest	0.66	0.39	0.69	0.36	0.73	0.32	0.74	0.28	0.71	0.21
Buildings	0.78	0.32	0.82	0.30	0.85	0.26	0.87	0.23	0.89	0.21
Farm	0.59	0.52	0.62	0.51	0.63	0.47	0.62	0.41	0.61	0.37
Soil	0.80	0.66	0.80	0.66	0.82	0.65	0.82	0.63	0.82	0.59
Rock	0.55	0.66	0.60	0.65	0.73	0.58	0.77	0.45	0.74	0.33
Average	0.67	0.48	0.71	0.46	0.75	0.42	0.76	0.37	0.75	0.32

Figure 29: Average precision of LTL and ALTL with various threshold values

9.3.7 Thresholding of Semantic Scores

The proposed LTL approach uses a predefined threshold to decide whether an image I_t is related to the semantic concept SC_k . If the semantic score of the image is larger than the threshold v , then the image I_t is classified into the semantic concept SC_k . In this experiment, the performance of the LTL approach is evaluated for different choices of v . In addition to precision and recall rates, the F-measure [93] is employed to assess the semantic inference performance in an integrated way. The formula of the

F-measure is as follows:

$$\text{F-measure} = \frac{2PR}{P + R} \quad (57)$$

The F-measure is the harmonic mean of precision and recall rate. Only when precision and recall rate are both high, the corresponding F-measure will have a high value.

Table XX and Figure 30 show the precision and recall rates of the LTL approach for different choices of v . It can be observed that higher thresholds always yield higher precision and lower recall rates. Moreover, the average F-measure decreased monotonically with the increased threshold v , which indicates that the lower threshold has better performance considering both precision and recall rates.

Table XX: Precision and recall rates using different threshold values v

Semantic concept	$v=0.00$		$v=0.01$		$v=0.02$		$v=0.03$		$v=0.04$	
	P	R	P	R	P	R	P	R	P	R
Cloud	0.30	0.51	0.40	0.23	0.54	0.11	0.85	0.05	0.95	0.02
Water	0.89	0.29	0.96	0.12	1.00	0.07	1.00	0.06	1.00	0.05
Forest	0.67	0.31	0.84	0.13	0.91	0.04	0.96	0.03	0.99	0.01
Buildings	0.75	0.38	0.91	0.14	0.96	0.07	0.99	0.04	0.99	0.03
Farmland	0.44	0.57	0.60	0.45	0.83	0.33	0.89	0.25	0.95	0.16
Soil	0.64	0.80	0.81	0.60	0.92	0.32	0.93	0.20	0.97	0.08
Rock	0.30	0.75	0.41	0.58	0.60	0.27	0.67	0.18	0.75	0.07
Average	0.57	0.52	0.70	0.32	0.82	0.17	0.90	0.12	0.94	0.06
F-measure	0.54		0.44		0.28		0.21		0.11

Figure 30: Precision vs. recall rate for different threshold values v

9.3.8 Retrieval Example

Finally, the experimental performance assessment is concluded with query showing the actual retrieval result. The LTL approach used both codebooks based on spectral and textural features, 83 training images, and the third level of a three-level codebook, which contains 1200 codes.

Figure 31 shows the retrieval results if the user only provides the term ‘urban area’ to query the database, whereby all retrieved images match the specific semantic.

In total, the system retrieved a total of 1157 related images from 14 scenes, where there were 83 images in average for each scene. In top 100 images, 36 images are overlapped with the 83 training images.

Figure 31: Retrieval example based on the query term ‘urban area’

9.3.9 Verification Experiments

The experimental data set is identical to the scenario described in Section 8.5.4 in order to enable cross-comparison between the approaches.

In the performance evaluation the proposed LTL approach was compared with the SemQuery approach and the SBN approach. The comparison of the SemQuery and the CSBN is shown in Table XXI. For the CSBN, codebooks with 300 codes for the spectral and texture features were generated individually, and then the inference was derived. For the joint features, the inference results based on two kinds of codebooks were intersected. From Table XXI, it can be observed that the CSBN outperformed the SemQuery in both precisions and recall rates.

The comparisons of the SBN and the CSBN were shown in Table XXII and Table XXIII. It can be found that for all cases the precisions of the CSBN are significantly higher than those of the SBN. However, the results also show that for all cases the recall rates of the CSBN are significantly lower than those of the SBN.

The experimental results are similar to the results obtained based on the first experimental database.

Table XXI: Precision and recall rates over five separate training and retrieval sessions for semantic classification based on spectral and texture features

Semantic Concept	Spectral features				Texture features				Spectral and texture features			
	SemQuery		CSBN		SemQuery		CSBN		SemQuery		CSBN	
	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>
Cloud	0.05	0.36	0.34	0.35	0.28	0.08	0.21	0.36	0.37	0.04	0.52	0.21
Water	0.71	0.51	0.90	0.62	0.48	0.21	0.82	0.41	0.86	0.12	0.95	0.32
Forest	0.63	0.56	0.70	0.70	0.22	0.07	0.68	0.51	0.58	0.03	0.91	0.44
Building	0.57	0.60	0.76	0.66	0.27	0.05	0.60	0.33	0.55	0.03	0.89	0.25
Farm	0.24	0.60	0.38	0.89	0.23	0.09	0.19	0.70	0.45	0.06	0.59	0.64
Soil	0.32	0.75	0.60	0.86	0.23	0.13	0.33	0.50	0.68	0.10	0.78	0.46
Rock	0.14	0.74	0.36	0.65	0.07	0.17	0.14	0.41	0.33	0.12	0.42	0.35
Average	0.38	0.59	0.58	0.68	0.25	0.11	0.42	0.46	0.55	0.07	0.72	0.38

Table XXII: Precisions and recall rates of semantic inference based on spectral and texture features using simple and context-sensitive Bayesian networks

Semantic Concept	Spectral features				Texture features				Spectral and texture features			
	SBN		CSBN		SBN		CSBN		SBN		CSBN	
	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>
Cloud	0.10	0.79	0.34	0.35	0.08	0.80	0.21	0.36	0.15	0.66	0.52	0.21
Water	0.72	0.94	0.90	0.62	0.51	0.94	0.82	0.41	0.81	0.89	0.95	0.32
Forest	0.36	0.95	0.70	0.70	0.31	0.95	0.68	0.51	0.43	0.92	0.91	0.44
Building	0.35	0.99	0.76	0.66	0.33	0.98	0.60	0.33	0.39	0.97	0.89	0.25
Farm	0.10	0.99	0.38	0.89	0.08	0.99	0.19	0.70	0.12	0.98	0.59	0.64
Soil	0.13	1.00	0.60	0.86	0.09	1.00	0.33	0.50	0.16	1.00	0.78	0.46
Rock	0.08	0.99	0.36	0.65	0.06	0.93	0.14	0.41	0.10	0.93	0.42	0.35
Average	0.26	0.95	0.58	0.68	0.21	0.94	0.42	0.46	0.31	0.90	0.72	0.38

Table XXIII: Precisions and recall rates of semantic inference based on three levels of the tree-structured codebook using simple and context-sensitive Bayesian networks

Semantic Concept	Level 1 (300 codes)				Level 1 (600 codes)				Level 3 (1200 codes)			
	SBN		CSBN		SBN		CSBN		SBN		CSBN	
	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>	<i>P</i>	<i>R</i>
Cloud	0.10	0.79	0.34	0.35	0.12	0.63	0.50	0.21	0.16	0.47	0.67	0.15
Water	0.72	0.94	0.90	0.62	0.80	0.90	0.93	0.47	0.83	0.81	0.95	0.32
Forest	0.36	0.95	0.70	0.70	0.42	0.92	0.81	0.53	0.50	0.84	0.88	0.37
Building	0.35	0.99	0.76	0.66	0.41	0.96	0.86	0.47	0.49	0.90	0.92	0.29
Farm	0.10	0.99	0.38	0.89	0.13	0.99	0.59	0.80	0.18	0.96	0.74	0.52
Soil	0.13	1.00	0.60	0.86	0.17	1.00	0.78	0.65	0.22	0.99	0.85	0.37
Rock	0.08	0.99	0.36	0.65	0.11	0.98	0.60	0.48	0.16	0.95	0.75	0.31
Average	0.26	0.95	0.58	0.68	0.31	0.91	0.72	0.52	0.36	0.85	0.82	0.33

Chapter 10

Stepwise Remote Sensing Image Retrieval Scheme

10.1 Remote Sensing Image Retrieval

Traditional CBIR techniques retrieve remotely sensed images from the databases based on comparisons with pre-extracted low-level features. The low-level features are capable of accurately describing the characteristics of the images and are especially useful if users prefer to search the database content according to the signal characteristics. However, the retrieved images may not be related to the user's query intention due to the semantic gap between low-level features and high-level semantic concepts. Therefore, in research work several interactive learning approaches to bridge the semantic gap were developed. The proposed approaches rank the remote sensing images according to their semantic scores, which measure the extent to which an image is related to the user's query intention.

However, the semantic scores are measured based on codes, which are not as accurate as the low-level features in describing the characteristics of the images. Thus, the top-ranked images retrieved by the proposed approaches often have different signal characteristics than the query image. In order to solve the problem, a stepwise image retrieval scheme that combines the interactive learning approaches with a region-based similarity measurement IRM is developed to retrieve the images that are related to the user's query intention and have similar signal characteristics. The basic idea is that the proposed interactive learning approaches are used to pre-select a set of candidate images that are related to the user's intention. Then, the IRM algorithm [110] is used to re-rank the candidate images to ensure that the top-ranked images also have similar signal characteristics to the query.

10.2 Remote Sensing Image Indexing

For large remote sensing databases, it is necessary to retrieve and index the remote sensing images in an effective and efficient way. Traditionally, most CBIR systems index the images mainly based on the corresponding low-level features. However, the detailed descriptions of the remote sensing images require high-dimensional feature vectors. Since the computational cost of indexing high-dimensional feature vectors is high, it is almost infeasible to perform a linear search within the databases with real-time response. In particular, it has been proven that the computational cost of indexing high-dimensional feature vectors is exponentially growing with increasing dimensionality [9]. In fact, the indexing problem is much more severe in RBIR systems due to the large amount of detected regions, which significantly outnumber the stored images. Moreover, the computational costs of the region-based similarity measurements are much higher than those of the global similarity measurements since the measurement is based on individual region-to-region comparisons. In order to reduce the computational costs, filtering techniques [41], [48], [50], [60], [101] are often used to balance the effectiveness and efficiency in two steps. Firstly, an inexpensive search algorithm with limited precision is used to retrieve a set of candidate images. Then, a more sophisticated algorithm is used to search within the candidate image set obtained previously. Although these stepwise techniques significantly increase the retrieval efficiency, they perform the first retrieval iteration merely based on the low-level features in an unsupervised way, which incorporates little or no semantic information. Thus, the final retrieved images are often irrelevant with respect to the users' query intention. In order to solve the problem, instead of using the unsupervised approaches, the supervised step-wise scheme is proposed and described in details in Section 10.3.

10.3 Stepwise Image Retrieval Scheme

Given a query image I_q and the semantic threshold δ_k , the set of candidate images A obtained in the first step is expressed as

$$A = \{I_p \mid SI_k(I_p) \leq \delta_k\}. \quad (58)$$

Let $SI_k^1(\cdot)$ and $SI_k^2(\cdot)$ be semantic score functions based on the code functions C_1 and C_2 corresponding to the two different types of feature functions. The thresholds δ_k^1

and δ_k^2 are set for the semantic score functions $SI_k^1(.)$ and $SI_k^2(.)$, respectively. Then the image set A can be further expressed as

$$A = \{I_p \mid SI_k^1(I_p) \leq \delta_k^1 \wedge SI_k^2(I_p) \leq \delta_k^2\} \quad (59)$$

Then the IRM algorithm is used to rank the images provided by the set A based on the query image I_q . The stepwise image retrieval scheme requires the user's interaction to accomplish the first step. The user is required to explicitly indicate his/her query intention. If no trained semantic concept matches the user's query intention, the user is asked to communicate with a RF approach in order to gather necessary knowledge with respect to the query intention.

10.4 Experimental Results

The experimental data set is the same as in Chapter 8.

10.4.1 Computational Cost of the Pre-Selection Step

In this chapter, the semantic score function is used to reduce the computational cost as well as to guarantee the semantic relevance for the retrieved images. In the last chapter, it has been shown that the semantic score function is effective and efficient in retrieving the semantically relevant images. Thus, this subsection focuses on analysing the computational costs of the semantic score function.

On average each image contains m regions. In a worst-case scenario, all regions are adjacent to all others. Then, there are $m \times (m-1)$ region pairs in each image. Thus, the semantic score function requires $m \times (m-1)$ additions and m divisions to calculate the semantic score for each image. However, the IRM algorithm needs to calculate m^2 distances among the regions of two images. If each region is described by an n_d -dimensional feature vector, it requires n_d subtractions and an equal number of multiplications to calculate the Euclidean distance between any two regions. Therefore, the IRM algorithm requires $m^2 \times n_d$ subtractions and multiplications, respectively, to measure the distances among the regions of two images. Moreover, the IRM algorithm needs to find the optimal matches between the regions. Since the computational cost of region matching is case dependent, it is omitted in the following comparison. For the sake of simplicity, additions and subtractions are regarded as first order calculation while multiplications and divisions are regarded as second order calculation. Thus, in order to compare the computational costs of the semantic score

function and the IRM, the first order calculation ratio and the second order calculation ratio are defined. The first order ratio is defined as

$$ratio^{(1)} = \frac{m^2 \times n_d}{m \times (m-1)} = \frac{m \times n_d}{m-1} \approx n_d. \quad (60)$$

From Equation (60), it is found that the number of dimensions is proportional to the first order ratio. Thus, in a high dimensional feature space, the IRM requires much more first order calculations than the semantic score function does. The second order ratio is defined as

$$ratio^{(2)} = \frac{m^2 \times n_d}{m} = m \times n_d. \quad (61)$$

From Equation (61), it is found that the IRM requires far more second order calculations than the semantic score function does. In general, the semantic score function significantly reduces the computational cost especially for the high-dimensional feature vectors.

Table XXIV shows the number of candidate images selected by the proposed LTL approach for different semantic concepts. In average, 8%, 5%, 3% of the stored images are pre-selected with respect to the different levels of the codebook. Thus, the IRM algorithm only needs to search within a small subset of the database.

Table XXIV: The number of pre-selected images for different semantic concepts in the pre-selection stage

Semantic concept	Level 1 (300 codes) N_{pre}	Level 2 (600 codes) N_{pre}	Level 3 (1200 codes) N_{pre}
Cloud	1202	654	394
Water	2894	1870	1292
Forest	2603	1389	826
Urban areas	3177	1730	946
Farmland	1645	930	519
Soil	1438	1128	366
Rock	1592	824	465
Average	2079	1218	687

10.4.2 Retrieval Performance of the Stepwise Scheme

The semantic score function is effective and efficient in measuring the extent to which an image is related to the corresponding semantic concept. However, it is not very accurate in measuring the similarities with respect to the signal characteristics. Therefore the IRM algorithm is suggested to re-order the pre-selected images according to the low-level features of the query image.

Table XVII, Figure 32 and Figure 33 show the precisions of the top-100 images retrieved by the LTL approach and the IRM algorithm, respectively, alone as well as the stepwise scheme, whereby the stepwise scheme utilises the LTL approach in the first step. The LTL approach was performed based on the three levels of the codebook corresponding to spectral features. From Figure 32 and Figure 33, it can be found that the stepwise scheme obtains higher precision and significantly reduces the retrieval time comparing to the IRM-based approach. Furthermore, it can also be noticed that the stepwise scheme obtains higher precision if the LTL approach was performed based on higher codebook levels.

Table XXV: Precision comparison (top-100 images) for IRM with and without pre-selection stages

Semantic concept	Precision			
	IRM Linear	Level 0	Level 1	Level 2
Cloud	0.27	0.45	0.60	0.75
Water	0.66	0.90	0.95	0.96
Forest	0.64	0.74	0.79	0.86
Urban areas	0.75	0.85	0.90	0.94
Farmland	0.77	0.83	0.85	0.87
Soil	0.88	0.95	0.96	0.95
Rock	0.63	0.78	0.85	0.91
Average	0.66	0.79	0.84	0.89

Figure 32: Average retrieval precision of IRM with and without pre-selection stages

Figure 33: Average retrieval time of IRM with and without pre-selection stages

10.4.3 Retrieval Example

Finally, the actual retrieval results are demonstrated. For the retrieval of images with partial urban cover, the stepwise scheme uses 83 training images, the third levels of three-level codebooks based on spectral and textural features, and the IRM algorithm for re-ordering the final result. Figure 34 shows the retrieval results obtained by using IRM alone while Figure 35 shows the retrieval results obtained by the proposed stepwise scheme. The actual query image is shown in the upper left corner. The retrieved remote sensing scenes are ordered sequentially according to their ranking. In Figure 34, only the IRM is used to retrieve images with urban areas based on the low-level features of the query image. The result shows that five images in the top-10 images have a similar appearance with respect to the query image. However, in Figure 35 the proposed stepwise scheme retrieved eight similar images in the top-10.

Figure 34: Retrieval example using IRM

Figure 35: Retrieval example using the stepwise retrieval scheme

10.4.4 Verification Experiments

Table XXVI shows that the stepwise scheme obtains higher precision. This result is similar to the result obtained for the first experimental database.

Table XXVI: Precision comparison (top-100 images) for IRM with and without pre-selection stages

Semantic concept	Precision	
	IRM Linear	LTL + IRM Level 1 (300 codes)
Cloud	0.29	0.40
Water	0.78	0.91
Forest	0.45	0.59
Urban areas	0.65	0.82
Farmland	0.74	0.81
Soil	0.60	0.80
Rock	0.40	0.52
Average	0.56	0.69

Chapter 11

Further Application of the Developed Framework

In this chapter, the developed framework is applied in the VirtSat system, which is a virtual satellite working in the World Wide Web environment. The VirtSat system is a focused web crawler that is specialised on satellite imagery. Since the set of retrieved images contains non-satellite images, the developed framework is used to differentiate the satellite imagery from the non-satellite images.

11.1 Application Background

Over the past decade satellite imagery became a substantial part of everyday life. It encompasses areas like weather forecasts and almost instantaneously provides updates on the current situation in remote regions hit by disasters like earthquakes or a tsunami. The involved technology is expensive due to the development costs for the spacecrafts. These expenses can range from a few million up to one billion US dollars, not including the daily operational and management costs. Although the resale prices dropped significantly after the commercial market was opened for high resolution imagery through Presidential Directive 23 in 1994, they are in the dollars per square kilometre range. Considering the fact that in most real world scenarios large areas have to be observed, the cost can impose significant restrictions on the effective utilisation of the image products. In particular application areas and users with a limited budget, for example, non-tertiary education and developing countries, are affected.

However, a large variety of remote sensing image data is available and free-of-charge on the Web. Although these images are partially degraded, they are useful for selected applications or at least for extensive evaluation [3], [102]. Even for certain research objectives the free image data can be useful, as long as the inherent limitations, like lossy compression, radiometric calibration, etc., are well understood. One of the main problems is to locate these satellite images on the web due to frequent updates and changed to website structures. A manual approach is highly inefficient and the utilisation of search engines like Google and AltaVista does not provide satisfactory

results, mainly due to their design requirements to cater for a huge variety of different queries. Standard search engine rely on keyword matching, and hence the retrieval for the word ‘satellite’ contains among actual satellite images also pictures showing the satellites themselves as well as receiving dishes of ground stations. What is even more severe from the user’s perspective is the fact that the images are presented ungrouped. For instance, passive imaging is based on the reflectance of the observed scene, while active approaches using microwaves transmit a signal and then measure the reflectance. Obtained images from the same scene are very different in nature and therefore should be qualified accordingly, which is almost impossible if only the keywords in the proximity of the retrieved image can be used. Correspondingly, no differentiation with respect to the image content is carried out, since it is almost infeasible to search for scenes that depict a certain ground cover characteristic. Finally, the scope of search engines is almost unlimited in terms of web pages. This means that updates, for example, on websites containing satellite image data, cannot be found as quickly as in a dedicated approach.

All the mentioned shortcomings can be solved by a focused web crawler that is specialised on satellite imagery and employ the unique properties of satellite imagery for evaluation. Since the parameterisation of such a system is not trivial, a possible solution is to make the actual crawlers perform the image evaluation process. This enables a dynamic adaptability and learning capability, which ultimately ensures a better utilisation of the available bandwidth to the Web because more valuable data is transmitted. The implementation of VirtSat on distributed computation architecture further enhances the potential of the solution, since the crawlers are not required to perform all evaluations steps themselves but can focus on the extraction of useful data. The result is an image database that comprises a structure catalogued according to the satellite images’ contents.

11.2 VirtSat System

VirtSat is written in Java to allow a straightforward portability between different operating system platforms. It utilises a heuristic search in which web pages with higher probability of containing satellite images are considered first. The actual system is divided into several threads that run on a distributed architecture. The

following paragraphs illustrate the various design and implementation issues of the system.

The VirtSat system uses multiple crawler threads to fully exploit the idle CPU time and network bandwidth. Thus, the multi-threaded crawling significantly accelerates the rate of finding and downloading satellite images. Each crawler thread is in charge of traversing web pages, extracting URL links and link text, finding and downloading satellite images. In order to manage these crawler threads, the VirtSat system utilises the thread manager to dynamically maintain URL links in the main queue, distributed these links to the crawler threads and receive links extracted by the crawler threads. The system architecture is shown in Figure 36.

Figure 36: VirtSat system architecture

The VirtSat system starts traversal by visiting URLs in the main queue initialised by seed URLs, which are manually selected from relevant web sites of satellite imagery to serve as start points. In the main queue, each URL is assigned to a domain priority, which measures the extent to which the corresponding domain contains satellite images. In order to visit the URLs with higher priorities earlier, a dynamic queue system sorts the main queue in descending order of domain priorities. The thread manager assigns different domains in the main queue to the each individual crawler threads. The maximum number of active crawler threads is defined by the user. However, the web pages of the same domain are very likely to reside on the same server. If multiple crawler threads connect to the same domain, they may significantly

increase the corresponding server hit rate and trigger an alarm. Therefore, the VirtSat systems restrict each crawler thread to a unique domain, and each crawler thread has to wait 15 seconds before starting a new connection.

In order to increase the searching effectiveness and efficiency, the crawler threads utilise heuristic search strategy to visit web pages with the higher priorities as early as possible. Before traversing the web pages, the system scanned the entire html contents of web pages to find reject keywords in order to detect undesirable web pages. If any reject keywords is found, the corresponding web page is very unlikely to contain satellite images and thus immediately rejected. If no reject keyword is found, the web page is accepted and parsed into html links, link text, image links and image description text. An example of an html link and its link text is “ Weather satellite images”, where the html link is “polar.html” and its link text is “Weather satellite images”. If the html link refers to a document file other than a URL link or an image file, it is rejected. If the accepted link is a URL link, it is placed in an appropriate thread queue or the main queue based on the following criteria: If the link is in the domain where the current crawler is working, it is inserted into the queue of current crawler. If the link is in the domain where another crawler thread is working, it is inserted into the queue of that crawler thread. If no crawler thread is working in the domain where the link resides, the link is inserted into the main queue of the thread manager. The flowchart of parsing web pages is shown in Figure 37.

Figure 37: Flowchart of a crawler thread parsing a web page

If the html link refers to an image file, it is an image link and the corresponding link text is called the image description text. In order to save network bandwidth and reduce search time, before downloading the image, the crawler thread verify the image link to make sure that the corresponding image file is a satellite image. It may reject the image links according to certain criteria to save network bandwidth and reduce overall search time. The criterion includes three aspects:

Firstly, if any image reject keyword such as “banner” is found in the description text of an image link, the image link is immediately rejected.

Secondly, an image must satisfy certain requirements, that is, minimum size and resolution, in order to eliminate icons, logos etc.

Thirdly, since the variance of satellite imagery is often higher than that of ordinary photographic imagery, the pixel density distribution of the image must have variance larger than 3000.

If these three aspects are met, the image is considered a possible satellite image and downloaded.

After visiting all the links in its own queue, a crawler thread is ended and the corresponding resource is recycled. Then, the thread manager is able to start a new crawler after it receives the information from the ended crawler thread.

11.3 Satellite Image Evaluation and Verification

The previous section showed that the crawler can provide a set of candidate satellite images based on a preliminary evaluation process. However, only general image properties were used and no actual content analysis was conducted. Thus, the set of candidate images contains non-satellite images. Even for a strict parameter choice the experimental results showed that 40% candidate images are non-satellite images, which have to be eliminated in an additional processing step. However, in order to allow a high collection rate, that is, data throughput, the crawler does not perform such action immediately. Moreover, the low level evaluation has to use relaxed filter constraints to guarantee that actual satellite images are not deselected and lost. Therefore, the developed framework is used offline to classify the candidate images into satellite images and non-satellite images. In order to perform the classification, a set of training satellite images is provided to generate the code co-occurrence matrix with respect to the semantic concept “satellite image”. Then, a semantic score function is used to measure the extent to which a candidate image is relate to “satellite image” based on the code co-occurrence matrix. Finally, the candidate images are classified into satellite images if their semantic scores are larger than a predefined threshold.

11.4 Results and Assessment

In the experiment, twenty-one multispectral satellite scenes acquired by IKONOS in 2003 were utilised. Each scene was cut into image tiles. A total set of 2100 satellite image tiles with a size of 300×300 pixel were available for the experiment. These satellite images are mixed with 900 384×256 (256×384) non-satellite images from COREL. The non-satellite images mainly fall into 5 categories and 13 sub categories as shown in Table XVIII.. The satellite image set and non-satellite image set in the experimental dataset are denoted as Z_s and Z_{ns}

Table XXVII: Categories of non-satellite images

Categories	Sub categories	Number of images
Animals	Elephant	100
	Horse	100
Food		100
Human	Africa	100
Landscape	Snow mountain	100
	Beach	100
Flowers	Purple	27
	Red	43
	White	9
	Yellow	14
	Other	7
Vehicle	Double-decker bus	60
	Mono-layer bus	40

In this experiment, each of the 3000 images was segmented into 2-4 regions to produce 9543 regions in total. Each image region was described by a 64-dimensional colour histogram. The differentiation performance depends on the codebook size. In order to find out the extent to which the differentiation is dependent on a predefined size choice, a tree-structured codebook was utilised. Thereby, the utilised three levels contained 100, 200 and 300 codes, respectively. Before the differentiation, training satellite images have to be supplied to generate the code co-occurrence matrix for satellite images. In order to investigate the impact of the training set size on the differentiation performance, 100, 200,..., 1000 training satellite images were chosen. The evaluation was carried out in terms of precision $P_s = |A_s \cap Z_s| / |A_s|$, the number of

correct images $N_s = |A_s \cap Z_s|$, and recall rate $R_s = |A_s \cap Z_s| / |Z_s|$, where A_s and Z_s are the sets of automatically and manually classified satellite images, respectively.

In the experiments, the developed framework classified the satellite images based on different codebook sizes and different training set sizes. In Figure 38, the number of correctly classified satellite images N_s steadily rises with the number of training images. Moreover, the classification based on fewer codes always achieves larger N_s values.

Figure 38: Number of correctly classified satellite images based on different codebook sizes and different training set sizes

However, in Figure 39, the inverse phenomena were observed. Firstly, the precision P_s continuously drops with the number of training images. Secondly, the classification based on more codes always achieves higher precisions. For instance, if 300 codes were used, the classification achieves 97.0 % precision with 100 training images and 94% precision with 1000 training images. However, if 100 codes were used, the precision is 94.2% with 100 training images and 85.2% with 1000 training images.

Figure 39: Precision of the classified satellite images with respect to the codebook size and the size of the image training set

Figure 40: Recall rate of the classified satellite images with respect to the codebook size and the size of the image training set

In order to investigate the extent to which the classified satellite images cover the experimental dataset, the experimental results were also analysed based on recall rate R_s . In Figure 40, it has been found that the recall rate R_s rises steadily with the increased number of training images and always have lower value if the classification is based on more codes.

Figure 41: Precisions and recall rates based on different levels of the codebook: (a) 100 codes, (b) 200 codes, (c) 300 codes

According to the experimental results in Figure 39 and Figure 40, if sufficient training images can be provided, the classification achieved higher performance with more codes. For instance, if 1000 training images were provided, the classification achieves 94% precision and 90% recall rate with 300 codes. However, if the same number of training images was provided and only 100 codes were used, the precision is 85.2% and the recall rate is 91%. The advantage of large codebook is further proved by the experimental results shown in Figure 41. That is, the classification based on more codes requires more training image to achieve better optimal classification point, where precision equals recall rate.

If sufficient training images are not available, 300 - 400 training images are preferred since the classification performance improved considerably at this duration as shown in Figure 38–Figure 40.

Chapter 12

Conclusions and Future Directions

12.1 Conclusions

In this research, a novel framework for image retrieval was designed to interactively infer the semantic concepts of remotely sensed imagery. A context-sensitive Bayesian network framework was created in order to incorporate spatially adjacent information into the inference process. To simplify the calculation, semantic score functions were derived from the context-sensitive Bayesian network model, and used to measure the extent to which a scene is related to a semantic concept. The semantic specific knowledge was stored in so-called code co-occurrence matrices, which provide the necessary information for the semantic score function. In this context a novel relevance feedback (RF) approach was described to interactively construct the code co-occurrence matrices for the semantic inference and, thus, relief the inconsistency between low-level features and high-level semantic concepts. Furthermore, to improve the precision, a pre-filtering function was designed to include negative samples into the RF. Given a semantic concept, the corresponding matrix of positive samples provides the positive information, while the matrices of negative samples provide information about the irrelevant images with respect to the given concept. Therefore, the other RF approach was extended by the pre-filtering function to measure the semantic score of an image from both the positive as well as negative perspective.

The capabilities of the two developed RF approaches were tested using an extensive set of remote sensing images and assessed in terms of precision and the retrieved number of qualified remote sensing images. In summary, the RF techniques without pre-filtering cannot consistently improve the precision, but can steadily increase the number of adequate images throughout a query session. However, the trade-off between precision and retrieval extent is favourable for the depicted system, since the precision decreases significantly less than the increase of correctly retrieved images. The RF with pre-filtering can consistently and significantly improve the precision at

the cost of a moderate drop in the number of correctly retrieved images. Therefore, the two proposed RF approaches can be utilised as a cooperative approach to improve iteratively both retrieval characteristics.

Since the proposed RF approaches lack the capability of memorising information, a long-term learning (LTL) approach was developed in this research to gradually improve the retrieval performance through the knowledge accumulated from previous users' interactions. Since the proposed LTL approach conducts the semantic inference based on the context-sensitive Bayesian network, the feedback examples provided by different users can be accumulated together to train the context-sensitive Bayesian network. In this context the code co-occurrence matrices play a central role in bridging the short-term feedbacks with the LTL.

In addition, an associative learning strategy was developed to synthesis, the long-term knowledge of different semantic concepts to improve the performance of the semantic inference using the pre-filtering function. Therefore, the strategy of measuring the semantic scores of an image from both the positive as well as negative perspective was suggested. The major advantage of associative learning is that the accumulated user-supplied knowledge is fully exploited. The performance of LTL in a standard as well as in an associative learning strategy was assessed in terms of precision and recall rates. The experimental results showed that the latter approach can improve the precision at the cost of the recall rate considerably.

At the retrieval stage, a direct image matching is not adequate for retrieving the images in a large satellite database if real-time response is required. Therefore, this research adopted a stepwise image retrieval scheme to balance effectiveness and efficiency. Firstly, the developed LTL or RF approaches are used to retrieve a set of candidate images that are related to the user's semantic concepts. This intermediate result is the further evaluated by integrated region matching in order to measure the similarity between the query image and the candidate images. Since the number of candidate images was reduced significantly in comparison to the number of actually stored images, the proposed stepwise scheme reduces the retrieval time especially for large satellite databases. Furthermore, the experimental results showed that the proposed scheme is able to retrieve images that are semantically relevant as well as visually similar.

In summary, the proposed framework has three major advantages. Firstly, the developed RF approaches reduce expensive and time-consuming manual training significantly. In particular, this is due to the consideration of negative examples with a filtering of the knowledge in order to avoid the assimilation of non-meaningful information. Secondly, the proposed associative learning strategy guarantees a comprehensive exploitation of the accumulated feedback information. Thirdly, the proposed stepwise image retrieval scheme significantly reduces the retrieval time and increases the retrieval precision. However, a limitation is that a computationally expensive regeneration of the derived semantics is required if new satellite scenes are added to the database. Furthermore, like in previous work, the determination of the most suitable trade-off between precision and recall rates is still left to the user.

12.2 Contribution

A context-sensitive Bayesian network framework was proposed to integrate spatially adjacent information into the inference process. To simplify the calculation, semantic score functions were derived from the context-sensitive Bayesian network model, and used to measure the extent to which a scene is related to a semantic concept.

A novel relevance feedback (RF) approach was proposed to interactively construct the code co-occurrence matrices for the semantic inference and, thus, relieve the inconsistency between low-level features and high-level semantic concepts.

The other proposed RF approach utilises the pre-filtering function to measure the semantic score of an image from both the positive as well as negative perspective.

A long-term learning (LTL) approach was developed in this research to gradually improve the retrieval performance through the knowledge accumulated from previous users' interactions. In addition an associative learning strategy was proposed to synthesise the long-term knowledge of different semantic concepts to improve the performance of the semantic inference using the pre-filtering function.

At the retrieval stage, a stepwise image retrieval scheme is proposed to balance effectiveness and efficiency.

12.3 Future Directions

In the presented research work, the generalised Lloyd algorithm was used to partition the feature space into a set of non-overlapping signal classes. Thus, each image region is encoded as the code, which indexes to the nearest signal class. However, the mapping from the continuous feature space to quantised discrete signal classes may cause serious quantisation errors since the signal classes may not precisely reflect the original distributions of the clusters in the feature space. For example, one cluster may be separated into two signal classes and different clusters may be classified into the same signal class. As an alternative solution, the codebook can be modelled as a mixture of overlapping probability density function. Thus, each image region can be classified into more than one signal class with different membership probabilities. The major advantage of the fuzzy model is that the distributions of the regional features are considered into the image encoding process. Thus, the content of each region is more precisely represented. Therefore, in this context, the fuzzy vector quantisation algorithm has to be combined with the context-sensitive Bayesian network. The corresponding probabilities have to be estimated in a way that adapts to the fuzzy vector quantisation algorithm.

In the semantic inference process, only the adjacent relationship among regions is considered. However, other pairwise region spatial relationships also contain the high-level semantic meaning and thus need to be considered. These relationships can be grouped into three categories [2]: perimeter-class relationships, distance-class relationships and orientation-class relationships. The perimeter-class relationships include *disjoined*, *bordering*, *invaded_by* and *surrounded_by*, where *invaded_by* means about 50% perimeter of the smaller region is surrounded by the larger region, and *surrounded_by* means the smaller region is completely surrounded by the larger region. The distance-class relationships include *near* and *far*. The orientation-class relationships include *left*, *right*, *above* and *below*.

References

- [1] Boggess, A., Narcowich, F.J., 2002. A First Course in Wavelets with Fourier Analysis. Pearson Education, pp. 227–239.
- [2] Aksoy, S., Koperski, K., Tusk, C., Marchisio, G., and Tilton, J.C., 2005. Learning Bayesian classifiers for scene classification with a visual grammar. *IEEE Transaction on Geoscience and Remote Sensing*, 43(3): 581–589.
- [3] Andres, P.M., Amy, L., Steinwand, D., Lawson, J., Moffitt, R.A., 1999. Making sense of space-based Earth image data: the KidSat data system. *IEEE Transactions on Geoscience and Remote Sensing*, 37(4):1785–1800.
- [4] Barnard, K., Duygulu, P., Forsyth, D., Freitas, N.D., Blei, D.M., Jordan, M.I., 2003. Matching words and pictures. *Journal of Machine Learning Research*, 3:1107–1135.
- [5] Bartolini, I., Ciaccia, P., Waas, F., 2001. FeedbackByPass: A new approach to interactive similarity query processing. *Proceedings of the International Conference on Very Large Data Bases*, pp. 201–210.
- [6] Barclay, T., Gray, J., Slutz, D., 1997. Microsoft Terraserver: A spatial data warehouse, Technical Report.
- [7] Basset, O., Buquet, B., Abouelkaram, S., Delachartre, P., Culioli, J., 2000. Application of texture image analysis for the classification of bovine meat. *Food Chemistry*, 69:437–445.
- [8] Bechmann, N., Kriegel, H., Schneider, R., Seeger, B., 1990. The R* tree: An efficient and robust access method for points and rectangles. *Proceedings of SIGMOD*, pp. 322–332.
- [9] Berchtold, S., Keim, D.A., 1998. High-dimensional index structures: database support for next decade's applications. *Proceedings of the ACM-SIGMOD International Conference on Management of Data*, pp. 501–501.
- [10] Bergman, L.D., Castelli, V., Li, C.-S., Smith, J.R., 2000. SPIRE: a digital library for scientific information. *International Journal on Digital Library*, 3: 85–99.
- [11] Beveridge, J.R., Griffith, J., Kohler, R.R., Hanson, A.R., Riseman, E.M., 1989. Segmenting image using localized histograms and region merging, *International Journal of Computer Vision*, 2(3):311–347.
- [12] Bretschneider, T., Cavet, R., Kao, O., 2002. Retrieval of remotely sensed imagery using spectral information content. *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, 4, pp. 2253–2256.
- [13] Brice, C.R., Fennema, C.L., 1970. Scene analysis using regions. *Artificial Intelligence*, 1(3/4):205–226.
- [14] Burl, M.C., Fowlkes, C., Roden J., Stechert, J., Mukhtar, S., 1999. Diamond Eye: A distributed architecture for image data mining. *Proceedings of the SPIE*

- AeroSpace Conference on Data Mining and Knowledge Discovery: Theory, Tools and Technology, 3695, pp. 197–206.
- [15] Carson, C., Thomas, M., Belongie, S., Hellerstein, J.M., Malik, J., 1999. Blobworld: A system for region-based image indexing and retrieval. *Proceedings of the International Conference on Visual Information Systems*, pp. 509–516.
 - [16] Carstensen, J.M., 1992. Description and Simulation of Visual Texture. Technical University of Denmark: Ph.D. thesis.
 - [17] Chang, E., Goh, K., Sychay, G., Wu, G., 2003. CBSA: Content-based soft annotation for multimodal image retrieval using Bayes point machines. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(1):26–38.
 - [18] Chang, R., Chen, C., Liao, C., 2004. Region-based image retrieval using edgeflow segmentation and region adjacency graph. *Proceedings of the IEEE International Conference on Multimedia and Expo*, 3, pp. 1883–1886.
 - [19] Chang, C., Moon, B., Acharya, A., Shock, C., Sussman, A., Saltz, J., 1997. Titan: A high-performance remote sensing database. *Proceedings of the International Conference on Data Engineering*, pp. 375–384.
 - [20] Chassery, J.M., Garbay, C., 1984. An iterative segmentation method based on a contextual color and shape criterion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):794–800.
 - [21] Chen, C.Q., 2004. An enhanced generalized Lloyd algorithm. *IEEE Signal Processing Letters*, 11(2):167–170.
 - [22] Chen, Y., Wang, J.Z., 2003. Looking beyond region boundaries: a robust image similarity measure using fuzzified region features. *Proceedings of the IEEE International Conference on Fuzzy Systems*, 2, pp. 1165–1170.
 - [23] Chen, Y., Wang, Y.Z., Krovetz, R., 2005. CLUE: Cluster-based retrieval of images by unsupervised learning. *IEEE Transactions on Image Processing*, 14(8):1187–1201.
 - [24] Chica-Olmo, M., Abarca-Hernandez, F., 2000. Computing geostatistical image texture for remotely sensed data classification. *Computer and Geosciences*, 26(4): 373–383.
 - [25] Cox, I.J., Miller, M.L., Minka, T.P., Papathomas, T.V., Yianilos, P.N., 2000. The Bayesian image retrieval system. *PicHunter: Theory, implementation, psychophysical experiments*. *IEEE Transactions on Image Processing*, 9(1):20–37.
 - [26] Datcu, M., Seidel, K., 2000. Image information mining: Exploration of image content in large archives. *Proceedings of the IEEE Aerospace Conference*, 3, pp. 253–264.
 - [27] Datcu, M., Seidel, K., Pelizarri, A., Schroeder, M., Rehrauer, H., Palubinskas, G., Waleśa, M., 2000. Image information mining and remote sensing data interpretation. *Proceedings of the IEEE Geoscience and Remote Sensing Symposium*, 7, pp. 3057–3059.

- [28] Datcu, M., Seidel, K., 2005. Human-centered concepts for exploration and understanding of earth observation images. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):601–609.
- [29] Datcu, M., Seidel, K., Walessa, M., 1998. Spatial information retrieval from remote-sensing images – Part I: Information theoretical perspective. *IEEE Transactions on Geoscience and Remote Sensing*, 36(5):1431–1445.
- [30] Datcu, M., Seidel, K., Walessa, M., 1998. Spatial information retrieval from remote-sensing images – Part II: Gibbs-Markov Random Fields. *IEEE Transactions on Geoscience and Remote Sensing*, 36(5):1446–1455.
- [31] Datcu, M., Seidel, K., Walessa, M., 2000. Interactive learning and probabilistic retrieval in remote sensing image archives. *IEEE Transactions on Geoscience and Remote Sensing*, 38(5): 2288–2298.
- [32] Datcu, M., Daschiel, H., Pelizzari, A., Quartulli, M., Galoppo, A., Colapicchioni, A., Pastori, M., Seidel, K., Marchetti, P.G., D’Elia, S., 2003. Information mining in remote sensing image archives: System concepts. *IEEE Transactions on Geoscience and Remote Sensing*, 41(12):2923–2935.
- [33] Daubechies, I., 1988. Orthogonal bases of compactly supported wavelets. *Communications on Pure & Applied Mathematics*, 41:909–996.
- [34] Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., Yanker, P., 1995. Query by image and video content: The QBIC system. *Computer*, 28(9):23–32.
- [35] Fourier, F., Card, M., 2002. Long-term similarity learning in content-based image retrieval. *Proceedings of the IEEE International Conference on Image Processing*, pp. 22–25.
- [36] Gersho, A., Gray, R.M., 2001. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, pp. 410–423.
- [37] Gevers, T., Smeulders, A.W.M., 1999. The PicToSeek WWW image search system. *Proceedings of the IEEE International Conference on Multimedia Computing and Systems*, 1, pp. 264–269.
- [38] Greenspan, H., Dvir, G., Rubner, Y., 2000. Region correspondence for image matching via EMD flow. *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries*, pp. 27–31.
- [39] Greenspan, H., Goldberger, J., Ridel, L., 2001. A continuous probabilistic framework for image matching. *Journal of Computer Vision and Image Understanding*, 84(3):384–406.
- [40] Guttman, A., 1984. R-trees: A dynamic index structure for spatial searching. *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 47–57.
- [41] Hafner, J., Sawhney, H.S., Equitz, W., Flickner, M., Niblack, W., 1995. Efficient color histogram indexing for quadratic form distance function. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7):729–736.

- [42] Han, J., Ngan K.N., Li, M., Zhang, H.J., 2005. A memory learning framework for effective image retrieval. *IEEE Transactions on Image Processing*, 12(8):924–937.
- [43] Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6):610–621.
- [44] Haralick, R.M., 1986. Statistical image texture analysis. *Handbook of Pattern Recognition and Image Processing*, The United States of America: Academic Press, Inc, pp. 247–279.
- [45] Hauta-Kasari, M., Pakkinen, J., Jaaskelainen, T., Lenz, R., 1999. Multi-spectral texture segmentation based on the spectral cooccurrence matrix. *Pattern Analysis and Applications*, 2:275–284.
- [46] He, X., King, O., Ma, W.Y., Li, M., Zhang, H., 2003. Learning a semantic space from user’s relevance feedback for image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 13(1):39–48.
- [47] Hsu, S.-Y., 1978. Texture-tone analysis for automated land-use mapping. *Photogrammetric Engineering and Remote Sensing*, 44(11):1393–1404.
- [48] Hua, K.A., Vu, K., Oh, J.H., 1999. SamMatch, a flexible and efficient sampling-based image retrieval technique for large image databases. *Proceedings of the ACM International Conference on Multimedia*, pp. 225–234.
- [49] Huang, X.D., Ariki, Y., Jack, M.A., 1990. *Hidden Markov Models for Speech Recognition*. Edinburgh University Press.
- [50] Huang, J., Kumar, S.R., Mitra, M., Zhu, W., Zabih, R., 1997. Image indexing using color correlogram. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 369–380.
- [51] Huang, T.S., Zhou, X.S., 2001. Image retrieval with relevance feedback: From heuristic weight adjustment to optimal learning methods. *Proceedings of the International Conference on Image Processing*, 3, pp. 2–5.
- [52] Iron, J.P., Pertersen, G.W., 1981. Texture transformation of remote sensing data. *Remote Sensing of Environment*, 11:359–370.
- [53] Ishikawa, Y., Subramanya, R., Faloutsos, C., 1998. MindReader: Querying databases through multiple examples. *Proceedings of the International Conference on Very Large Data Bases*, pp. 218–227.
- [54] Jensen, J.R., 1996. *Introductory Digital Image Processing: A Remote Sensing Perspective*, Prentice-Hall.
- [55] Jing, F., Li, M., Zhang, H. J., Zhang, B., 2002. Region-based relevance feedback in image retrieval. *Proceedings of the IEEE International Symposium on Circuits and Systems*, 4, pp. 145–148.
- [56] Jing, F., Li, M., Zhang, L., Zhang, H.J., Zhang, B., 2003. Learning in region-based image retrieval. *Proceedings of the International Conference on Image and Video Retrieval*, 2728, pp. 198–207.

- [57] Jing, F., Li, M., Zhang, H.J., Zhang, B., 2004. Relevance feedback in region-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 14, pp. 672–681.
- [58] Jing, F., Li, M., Zhang, H. J., Zhang, B., 2004. An efficient and effective region-based image retrieval framework. *IEEE Transactions on Image Processing*, 13(5): 699–709.
- [59] Katayama, N., Satoh, S., 1997. The SR-tree: An index structure for high-dimensional nearest neighbour queries. *Proceedings of the International Conference on Data Engineering*, pp. 360–389.
- [60] Ko, B., Byun, H., 2005. FRIP: A region-based image retrieval tool using automatic image segmentation and stepwise Boolean and matching. *IEEE Transactions on Multimedia*, 7(1):105–113.
- [61] Koperski, K., Marchisio, G., Aksoy, S., Tusk, C., 2002. VisiMine: Interactive mining in image databases. *Proceedings of the IEEE Geoscience and Remote Sensing Symposium*, 3, pp. 1810–1812.
- [62] Latif-Amet, A., Ertuzun, A., Ercil, A., 2000. An efficient method for texture defect detection: Sub-band domain co-occurrence matrices. *Image and Vision Computing*, 18(6-7): 543–553.
- [63] Lee, C., Ma, W.Y., Zhang, H.J., 1999. Information embedding based on user's relevance feedback for image retrieval. *Proceeding of the SPIE Conference on Multimedia Storage and Archiving Systems*, 5, pp. 20–22.
- [64] Li, C.Y., Shih, M.C., Hsu, C.T., 2004. Image retrieval with relevance feedback based on graph-theoretic region correspondence estimation. *Proceedings of the International Conference on Pattern Recognition*, pp. 1051–1055.
- [65] Li, J., 2003. *Information Mining in Remote Sensing Imagery*. University of Nebraska: Ph.D. thesis.
- [66] Li, J., Wang, J.Z., 2003. Automatic linguistic indexing of pictures by a statistical modeling approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1075–1088.
- [67] Li, M., Chen, Z., Zhang, H., 2002. Statistical correlation analysis in image retrieval. *Pattern Recognition*, 35(12):2687–2693.
- [68] Li, Y., Bretschneider, T., 2003. Supervised content-based satellite image retrieval using piecewise defined signature similarities. *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, 2, 734–736.
- [69] Li, Y., Bretschneider, T., 2004. Semantics-based satellite image retrieval using low-level features. *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, 7, pp. 4406–4409.
- [70] Lillesand, T.M., Kiefer, R.W., Chipman, J.W., 2004. *Remote Sensing and Image Interpretation*. John Wiley&Sons, pp. 458–460.
- [71] Liu, W., Dumais, S., Sun, Y., Zhang, H., 2001. Semi-automatic image annotation. *Proceedings of the Conference on Human-Computer Interaction*, pp. 326–333.

- [72] Ma, W.Y., Manjunath, B.S., 1997. NETRA: A toolbox for navigating large image databases. Proceedings of the IEEE International Conference on Image Processing, pp. 568–571.
- [73] MacArthur, S.D., Brodley, C.E., Shyu, C.R., 2000. Relevance feedback decision trees in content-based image retrieval. Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries, pp. 68–72.
- [74] Mallat, S.G., 1989. Multiresolution approximations and wavelet orthogonal bases of $L^2(\mathbb{R})$. Transactions of America Mathematics Society, 315(1):69–87.
- [75] Mallat, S.G., 1989. A theory of multiresolution signal decomposition: The wavelet representation. IEEE Transactions on Image Processing, 11(4):1549–1560.
- [76] Marchisio, G.B., Li, W.H., Sannella, M., Goldschneider, J.R., 1998. Geobrowse: An integrated environment for satellite image retrieval and mining. Proceedings of the International Geoscience and Remote Sensing Symposium, 2, pp. 669–673.
- [77] Mehrotra, S., Rui, Y., Ortega-Binderberger, M., Huang, T.S., 1997. Support content-based queries over images in MARS. Proceedings of the IEEE International Conference on Multimedia Computing and Systems, 2, pp. 632–633.
- [78] Minka, T.P., Picard, R.W., 1997. Interactive learning using a society of models. Special Issue of Pattern Recognition on Image Database: Classification and Retrieval, 30(4):565–581.
- [79] Mori, Y., Takahashi, H., Oka, R., 1999. Image-to-word transformation based on dividing and vector quantization with words. Proceedings of the International Workshop on Multimedia Intelligent Storage and Retrieval Management.
- [80] NASA TERRA: The EOS flagship. <http://terra.nasa.gov>.
- [81] Oldfield, R.B., Elgy, J., 1987. Geological mapping from remotely sensed data using tonal, textural, and contextual features. Proceedings of the Annual Conference of the Remote Sensing Society, pp. 533–537.
- [82] Pratt, K.P., 2001. Digital Image Processing. John Wiley & Sons, New York.
- [83] Pitas, I., 2000. Digital Image Processing Algorithms and Application. John Wiley and Sons, Inc, New York.
- [84] Picard, R.W., Minka, T.P., 1995. Vision texture for annotation. Multimedia System, 3:3–14.
- [85] Ramachandran, R., Conover, H.T., Graves, S.J., Keiser, K., 2000. Challenges and solutions to mining earth science data. Proceedings of the SPIE AeroSense Conference on Data Mining and Knowledge Discovery: Theory, Tools and Technology, 4057, pp. 259–264.
- [86] Rioul, O., Vetterli, M., 1991. Wavelets and signal processing. IEEE Signal Processing Magazine, 8(4):11–38.
- [87] Rui, Y., Huang, T.S., Mehrotra, S., 1997. Content-based image retrieval with relevance feedback in MARS. Proceedings of the IEEE International Conference on Image Processing, 1, pp. 815–818.

- [88] Rui, Y., Huang, T.S., Ortega, M., Mehrota, S., 1998. Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):644–655.
- [89] Rui, Y., Huang, T.S., 2000. Optimizing learning in image retrieval. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 1, pp. 236–243.
- [90] Saber, E., Tekalp, A.M., 1997. Region-based shape matching for automatic image annotation and query-by-example. *Journal of Visual Communication and Image Representation*, 8(1):3–20.
- [91] Salton, G., 1988. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison-Wesley.
- [92] Salton, G., 1989. *Automatic text processing*. Reading, Mass., Addison-Wesley.
- [93] Shaprio, L.G., Stockman, G.C., 2001. *Computer Vision*. Prentice Hall, New Jersey.
- [94] Sheikholeslami, G., Zhang, A., Bian, L., 1999. A multi-resolution content-based retrieval approach for geographic images. *Geoinformatica*, 3(2):109–139.
- [95] Sheikholeslami, G., Chang, W., Zhang, A., 2002. SemQuery: Semantic clustering and querying on heterogeneous features for visual data. *IEEE Transactions on Knowledge and Data Engineering*, 14(5):988–1002.
- [96] Shyu, M.-L., Chen, S.-C., Chen, M., Zhang, C., 2004. A unified framework for image database clustering and content-based retrieval. *Proceedings of the ACM International Workshop on Multimedia Databases*, pp. 19–27.
- [97] Smeulders A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R., 2000. Content-based Image Retrieval at the End of the Early Years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380.
- [98] Smith, J.R., Chang, S.F., 1996. VisualSEEK: a fully automated content-based image query system. *Proceedings of the ACM International Conference on Multimedia*, pp. 87–98.
- [99] Smith, J.R., Shih-Fu, Chang, 1996. Local color and texture extraction and spatial query. *Proceedings of the IEEE International Conference on Image Processing*, 3:1011–1014.
- [100] Su, Z., Zhang, H.J., Li, S., Ma, S., 2003. Relevance feedback in content-based image retrieval: Bayesian framework, feature, subspaces, and progressive learning. *IEEE Transactions on Image Processing*, 12(8):924–937.
- [101] Sundaram, H., Chang, S.F., 1999. Efficient video sequence retrieval in large repositories. *Proceedings of the Storage and Retrieval for Image and Video Database*, 7, pp. 108–119.
- [102] Takru, K., Bretschneider, T.R., Leedham, C.G., 2004. Automatic extraction of road maps: determining the sealed surface. *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, 3, pp. 2022–2025.
- [103] Tatem, A.J., Lewis, H.G., Atkinson, P.M., Nixon, M.S., 2001. Super-resolution mapping of urban scenes from IKONOS imagery using a Hopfield neural

- network. Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, 7, 3203–3205.
- [104] Tieu, K., Viola, P., 2000. Boosting image retrieval. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1, pp. 13–15.
 - [105] Tong, S., Chang, E., 2001. Support vector machine active learning for image retrieval. Proceedings of the ACM International Conference on Multimedia, pp. 107–118.
 - [106] Vesanto, J., Alhoniemi, E., 2000. Clustering of the self-organizing map. IEEE Transactions on Neural Networks, 11(3):586–600.
 - [107] Wood, M.E., Campbell, N.W., Thomas, B.T., 1998. Iterative refinement by relevance feedback in content based digital image retrieval. Proceedings of the ACM International Multimedia Conference, pp. 13–20.
 - [108] Walcher, W., Rehatschek, H., 1995. Design of a distributed database for the Magellan data set from planet Venus. Proceedings of the International Geoscience and Remote Sensing Symposium, 2, pp. 1010–1012.
 - [109] Wang, J., Yang, W., Acharya, R., 1997. Color clustering techniques for color content-based image retrieval. Proceedings of the IEEE International Conference on Multimedia Computing and Systems, pp. 442–449.
 - [110] Wang, J.Z., Li, J., Wiederhold G., 2001. SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(9):947–963.
 - [111] White, D. A., and Jain, R., 1996. Similarity indexing with the SS-tree. Proceedings of the International Conference on Data Engineering, pp. 512–523.
 - [112] Zhang, C., Chen, T., 2002. An active learning framework for content-based information retrieval. IEEE Transactions on Multimedia, 4(2):260–268.
 - [113] Zhou, X.S., Huang, T.S., 2003. Relevance feedback in image retrieval: A comprehensive review. ACM Multimedia Systems Journal, 8:536–544.
 - [114] Zucker, S.W., 1976. Region growing: Childhood and adolescence. Computer Graphics and Image Processing, 5:382–399.

Publications

- 1 Y. Li, T. Bretschneider, 2007. Semantic sensitive satellite image retrieval, IEEE Transactions on Geoscience and Remote Sensing, 45(4): 853–860.
- 2 Y. Li, T. Bretschneider, Comparative assessment of semantic-sensitive satellite image retrieval based on simple Bayesian and context-sensitive Bayesian networks, International Journal on Geoscience Information System, *first revision*.
- 3 Y. Li, T. Bretschneider, Remote sensing image retrieval using a context sensitive Bayesian network for relevance feedback, Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Denver, USA, *in press*
- 4 T. Bretschneider, K.K. Goh, M.E. Heng, Y. Li, Retrieval of online available satellite imagery, Proceedings of the Asian Conference on Remote Sensing, Ulaanbaatar, Mongolia, CD-ROM 2006
- 5 S. Halim, T. Bretschneider, Y. Li, P. Preiser, C. Kuss, Estimating malaria parasitaemia from blood smear images, Proceedings of the IEEE International Conference on Control, Automation, Robotics and Vision, Singapore, 2006, *in press*
- 6 Y. Li, T. Bretschneider, 2006. Learning in semantic-sensitive image retrieval. Proceedings of the Asia GIS, Johor Bahru, Malaysia, CD-ROM.
- 7 Y. Li, T. Bretschneider, 2005. Estimating multiple high-level query concepts from relevance images in content-based satellite image retrieval. Proceedings of the Asian Conference on Remote Sensing, CD-ROM.
- 8 Y. Li, T. Bretschneider, 2004. Semantics-based satellite image retrieval using low-level features. Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, 7, pp. 4406–4409.
- 9 T. Bretschneider, Y. Li, 2003. On the problems of locally defined content vectors in image databases for large images. Proceedings of the International Conference on Information, Communications & Signal Processing and IEEE Pacific-Rim Conference on Multimedia, 3, pp. 1604–1608.
- 10 Y. Li, T. Bretschneider, 2003. Supervised content-based satellite image retrieval using piecewise defined signature similarities. Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, 2, pp. 734–736.