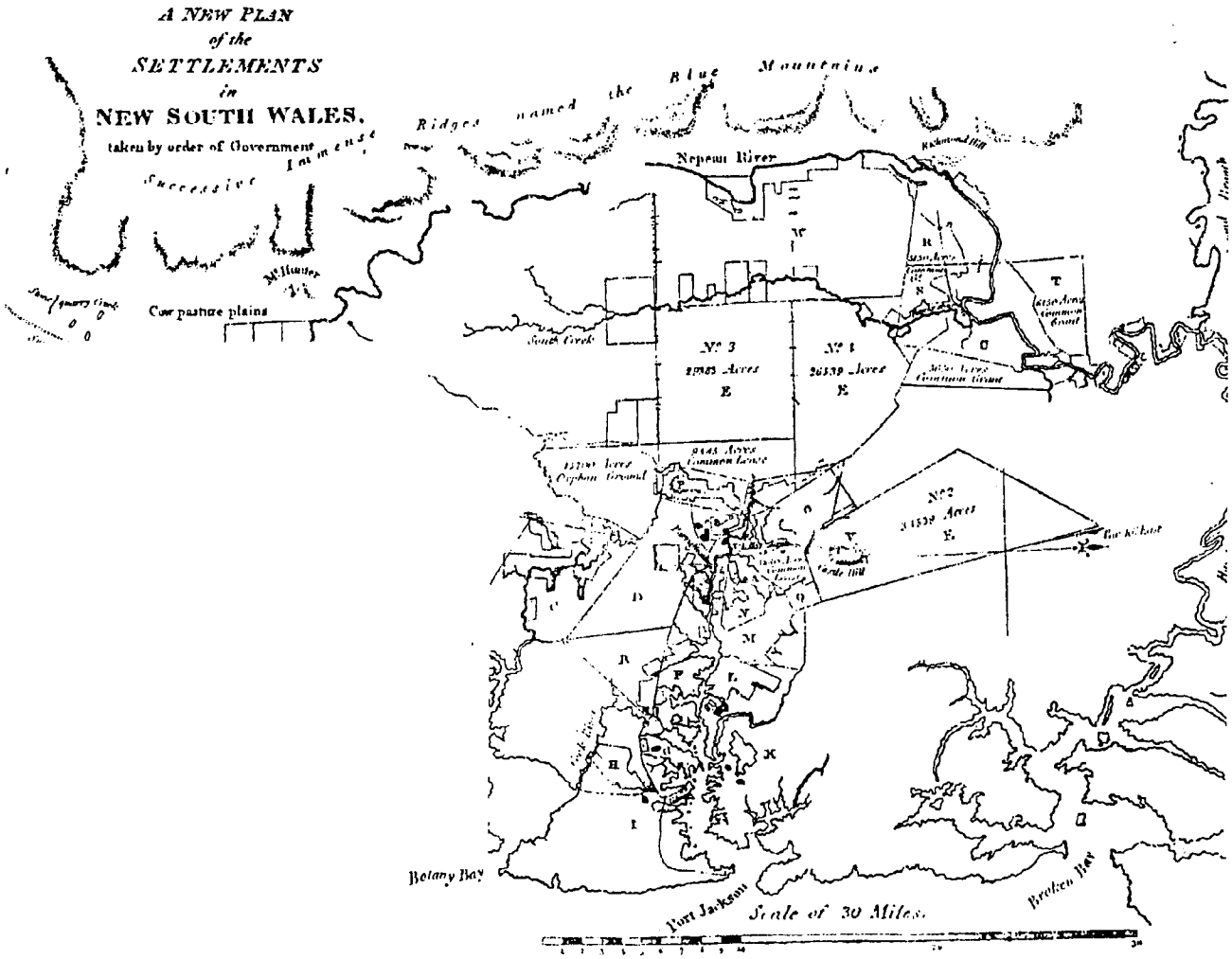
KNOWLEDGE-BASED ROAD EXTRACTION FROM AERIAL IMAGES

YANDONG WANG



UNISURV S-62, 2001

Reports from

SCHOOL OF GEOMATIC ENGINEERING

THE UNIVERSITY OF NEW SOUTH WALES UNSW SYDNEY NSW 2052 AUSTRALIA

UNISURV REPORT S-62, 2001

KNOWLEDGE-BASED ROAD EXTRACTION FORM AERIAL IMAGES

YANDONG WANG

Received: February 2001
Accepted: February 2001

SCHOOL OF GEOMATIC ENGINEERING
UNIVERSITY OF NEW SOUTH WALES
UNSW SYDNEY NSW 2052
AUSTRALIA

UNISURV REPORTS

Series Editor: Dr. J. M. Rüeger

National Library of Australia

Card No. and ISBN 0 - 7334 - 1767 - 1

Foreword

Digital image processing is currently an active area of research in photogrammetry. Topics include digital elevation model determination, and the extraction of features required for digital map updating and GIS database compilation. Development of automatic procedures for these operations is required so that the full potential of digital imaging in photogrammetry can be achieved. The School of Geomatic Engineering at UNSW has undertaken research on semi-automatic or automatic methods of feature extraction, in particular roads and buildings, since the early 1990s. Recently this work has been undertaken in collaboration with Associate Professor Arcot Sowmya of the School of Computer Science and Engineering. The research described in this thesis on automatic road extraction from low and high resolution images, has been supported by an ARC Small Grant and a grant from the Strategic Partnerships with Industry – Research and Training scheme (SPIRT) in 1996-1998, in collaboration with Associate Professor Sowmya. The research work is also supported by a Large ARC grant to investigate, jointly between the Schools of Geomatic and Computer Science and Engineering, the extraction of features by machine learning.

This thesis presents a knowledge-based method, using Prolog programming language, for automatic road extraction from aerial images based on a semantic road model in which roads are defined in both high- and low-resolution images and relationships between roads are defined. The method uses image analysis methods for feature extraction, and generation of symbolic representation of road structures and artificial intelligence techniques for manipulation of knowledge and recognition of roads. The recognition of roads includes hypothesis generation of roads from high-resolution images, verification of hypotheses, prediction of missing road segments and detection of occlusions occurring in images. Hypotheses of roads are generated from high-resolution images in a bottom-up process, which includes three different levels, i.e. low-, intermediate- and high-level.

The methods have been tested on a series of aerial photographs in Australia and on SPOT satellite images. The tests demonstrate examples of the procedures that can be taken to

detect and recognise roads on remote sensing images. While the methods achieve satisfactory results on many of the samples tested, there are examples where the algorithms do not perform satisfactorily. Further developments are required to overcome some of these deficiencies, as well as test the scalability of the algorithms over a range of image scales.

Emeritus Professor John C. Trinder
Visiting Professor
March 2001

Abstract

This thesis presents a knowledge-based method for automatic road extraction from aerial images based on a semantic road model in which roads are defined in both high- and low-resolution images and relationships between roads are defined. The method uses image analysis methods for feature extraction and generation of symbolic representation of road structures and artificial intelligence techniques for manipulation of knowledge and recognition of roads. The recognition of roads includes hypothesis generation of roads from high-resolution images, verification of hypotheses, prediction of missing road segments and detection of occlusions occurring in images. Hypotheses of roads are generated from high-resolution images in a bottom-up process which includes three different levels, i.e. low-, intermediate- and high-level. In low-level processing, features, such as edges, lines points are extracted from the original images. Intermediate-level processing concentrates on the generation of symbolic representation of road structures. It consists of generation of road structures, computation of their attributes and grouping of generalised antiparallel pairs. A generalised antiparallel pair is proposed to represent the structure of a road in high-resolution images, and its position in object space is determined by the Snake technique. To remove the effects of occlusions, such as trees, shadows, cars on road surface, etc., generalised antiparallel pairs are grouped based on their similarities in geometric and photometric properties. The generated road-like objects are represented symbolically in terms of a number of geometric and photometric attributes. In high-level processing, the knowledge of roads is applied to the generated road-like objects to yield recognition of road segments. The knowledge of roads includes their geometric and photometric properties and characteristics of road networks, which is expressed as rules in Prolog.

As hypotheses of road segments are generated in a local context, non-road objects may also be hypothesised if they have similar shapes and properties as roads. Therefore, a verification process is required to remove spurious hypotheses. In this study, a novel approach for verification of hypotheses has been developed, which is based on the

relationships between hypothesised road segments from high-resolution images and road networks derived from low-resolution images.

Occlusions are a common phenomenon in images. They usually break a road into several short segments which may not be hypothesised in the process of hypothesis generation. A process has been developed for finding missing road segments, based on the spatial relationships between verified road segments. It first uses the knowledge about context of roads and the established relationships between verified road segments to infer missing road segments. A top-down procedure is then invoked to find the road-like objects between them.

Occlusions not only break a road into segments, but also cover some parts of roads which cannot be recognised in the recognition process and found in detection of missing road segments. Therefore, detection of occlusions on road surface is very important as relationships between detected occlusion and verified road segments can be used to infer the existence of road segments in the occluded areas. A method based on supervised classification is used to detect occlusions in this study. It uses the maximum likelihood classification method to classify the pixels in the occluded area. If they are classified as trees, shadows or road surface, a road segment is then reasoned, and the disjoined road segments are linked using interpolation.

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my supervisor, Professor Trinder who brought me to the field of digital photogrammetry. I am very grateful for his consistent guidance, strong support and encouragement during my four year study in the School of Geomatic Engineering. I am also grateful of his careful reading and correction of my thesis. His invaluable advice, suggestions and financial support are gratefully acknowledged.

I would also like to thank Mr Brian Donnelly in the Photogrammetry Laboratory for his continuous support. He provided me with test image data and other technical support for this research. My thanks also go to all the staff members in the School of Geomatic Engineering for their assistance. In particular, I wish to thank Associate Professor Bill Kearsley, Coordinator of Post-graduate Studies in the school, Ms Helve Frangoulis and Ms Maria Ponce for their kind help during my study at UNSW.

I would also like to thank Dr Olof Henricsson at Swiss Federal Institute of Technology, Zurich for his providing me with the source code of SE operator.

Finally, I would like to express my deep gratitude to my family, especially my wife, Jing, for their support. Without their support, this thesis would not be possible.

Author
Sydney, Australia
February, 1999

Table of Contents

List of Tables

List of Figures

Chapter One

Introduction

1.1 Background

With the rapid development of computer technology, a transition from analytical to digital is under way in photogrammetry. Digital photogrammetry uses digital or digitized images and methods of image processing and image understanding to automatically derive geometric, radiometric, and semantic information about objects in the real world. With the revolutionary progress in computer vision and image understanding, digital photogrammetry has the potential of automating most processes in photogrammetric production and thus, increasing the efficiency of photogrammetric production. During the last two decades, digital photogrammetry has received considerable attention and remarkable progress has been made in automatic determination of positions of terrain points. This includes automatic interior orientation (Lü, 1997), automatic relative orientation (Tang and Heipke, 1996; Heipke, 1997), high-precision image matching (Ackermann, 1984; Grün, 1985; Rosenholm, 1987; Li, 1989; Baltsavias, 1991), automatic selection and transfer of tie points, automatic aerial triangulation (Schenk, 1997; Tang et al, 1997), automatic exterior orientation (Gülch, 1995; Drewniok and Rohr, 1997), and automatic generation of digital terrain models (DTM) and digital orthophotos (Greve, 1996). There are a number of commercial digital photogrammetric systems on the market (Ebner et al, 1991; Heipke, 1995) which can automatically generate DTMs and digital orthophotos, although human intervention is still required to handle difficult situations.

The procedure that is limiting the full implementation of automation in digital photogrammetry is the automatic location and recognition of objects, i.e. extraction of semantic information about objects, their description and characteristics. Objects in images have traditionally been recognized by an operator. This is a time-consuming

process. Therefore, there is a high demand for the automation of object recognition in order to increase the efficiency of photogrammetric production. In recent years, research on the automatic recognition of objects in photogrammetry has attracted considerable attention, but progress has been relatively slow. This is because the recognition of objects from images is an inherently more difficult task than the determination of their positions. With the successful solution of automatic determination of positions of terrain points, future research of digital photogrammetry will be focused on the extraction of semantic information of objects, and this trend has been reflected in a series of recent workshops on this subject (Grün et al, 1995; Grün et al, 1997; Förstner and Plümer, 1997).

Object recognition from images is the central task of image understanding, which concentrates on the establishment of relationships between objects in the real world and features in images with respect to specific knowledge about the objects. In contrast to conventional image analysis, it emphasizes the extraction of semantic information of objects from an image, rather than transforming an image from one form to another. It uses image analysis methods and artificial intelligence techniques for knowledge representation and manipulation. Up to now, a number of systems for object recognition using artificial intelligence techniques have been developed. The methods used in these systems can be adopted in the field of photogrammetry together with rigorous photogrammetric models for accurate measurement and interpretation. This study will explore methods of image understanding for automatic extraction of roads from aerial images.

1.2 Objectives of the Research

This thesis aims at developing a knowledge-based method for automatic road extraction from aerial images. Its main objectives include:

- The development of a general procedure for automatic extraction of roads from aerial images. According to current image understanding methods, object recognition is a complex process which consists of a number of interrelated stages.

Object recognition is usually performed in a certain order defined by a control strategy. Different control strategies may produce different results. Therefore, it is very important to select a suitable strategy for automatic extraction of roads.

- The definition of a suitable semantic road model for the recognition of roads. In image understanding, objects are recognized using a semantic model in which object classes and their relationships are defined. It is recognized that the scale of the image affects the appearance of an object in the image. Different aspects of the object are reflected in different resolutions of images. A combination of different resolutions of images can provide information of objects from different aspects, and therefore, better recognition results can be expected. This study will explore the use of different resolutions of images in automatic road extraction.

- The exploration of the use of knowledge of roads in automatic road extraction. A road is a man-made object which has distinct properties that distinguish it from other objects. Photogrammetrists have accumulated rich experience in interpreting remotely sensed imagery which can be used for automatic road recognition.

- The investigation of a suitable method for knowledge representation for road extraction from aerial images based on the first-order logic.

- The analysis of the effects of occlusions in the process of road recognition, and the development of a method for their detection. Occlusions are a difficult problem in road extraction which has not been studied thoroughly in previous research. They cause a loss of information about roads in images which will cause the recognition to be more difficult. On the other hand, they can provide contextual information about the environment in which roads occur, which will be useful in the recognition process. In this study, the characteristics of occlusions in aerial images and their detection will be analyzed. The use of contextual information relating areas of occlusions and roads will be studied in the process of road recognition.

1.3 Organization of the Thesis

This thesis consists of eight chapters. Following the introduction, Chapter 2 reviews the existing methods of road extraction from digital images, based on their classification into semiautomatic and automatic, and their suitability is analysed. The problems with these existing methods are also discussed.

Chapter 3 introduces the basic concepts and processes of image understanding, techniques for knowledge representation in artificial intelligence and control strategies used in image understanding systems. Problems in these processes are presented. Some existing knowledge-based systems for object recognition from aerial and satellite images are reviewed.

Chapter 4 deals with the methods of feature extraction in image understanding. Some commonly used algorithms and methods for extraction of edges, lines and points in computer vision and photogrammetry are introduced and their performances are analysed. This chapter also presents a number of techniques for preprocessing involved in low-level image processing, and methods for edge/line tracking, split-and-merge and linking developed in this study are described.

Chapter 5 describes the generation of symbolic representation of roads. It includes the generation of road structures in high-resolution images based on generalised antiparallel pairs, determination of their positions in 3-D space, computation of their attributes and grouping of generalised antiparallel pairs. To determine the positions of generalised antiparallel pairs in 3-D space, a modified active contour model is developed, in which the geometric constraints on the smoothness and parallelism of road boundaries are incorporated into the geometric energy model. Generalised antiparallel pairs are grouped based on their similarities in geometric and photometric properties, and two spatial constraints, i.e. collinearity and proximity.

Chapter 6 describes the method developed for road recognition from aerial images which includes bottom-up hypothesis generation of roads from high-resolution images, verification of hypotheses and detection of missing road parts and occlusions. Feature extraction and generation of symbolic representation of roads in the hypothesis generation are described in Chapters 4 and 5, thus only high-level processing will be presented in this chapter. Before presenting the details of the recognition methods, a semantic road model in which roads are defined in both high- and low-resolution images together with their relationships and associated knowledge is described. Hypotheses of roads are generated by applying the knowledge of geometric and photometric properties of roads expressed as rules in Prolog to the road-like objects. These hypotheses are verified using the relationships between them and the road network extracted from low-resolution images. Once the hypothesised roads are verified, their spatial relationships are determined for inferring the missing road parts caused by occlusions.

Test results of the developed methods for aerial images are presented in Chapter 7 in terms of four aspects: generation of hypotheses of roads from high-resolution images, verification of hypotheses, locating missing road parts, and detection of occlusions. Chapter 8 presents the conclusions and some recommendations for future research in automatic road extraction.

Chapter Two

Review of Road Extraction from Digital Images

Automatic road extraction from remotely sensed imagery has been an active research area in computer vision and digital photogrammetry for more than two decades. A number of methods and algorithms have been developed for road extraction from aerial photographs and satellite images. These methods and algorithms can be classified into semi-automatic and automatic according to the way and degree of human intervention. Semi-automatic methods usually delineate a road in the image, based on the edge information of the image or the analysis of road surface profiles. The starting position and direction, or the approximate position of a road, are provided by an operator. Automatic approaches aim to locate a road in the image without *a priori* knowledge of its initial position. They have two tasks, i.e. automatic recognition of a road and determination of its position. A road is recognised based on its specific properties, including its geometric and radiometric properties, while its position is determined using edge or surface information of the image. This chapter reviews existing methods of automatic road extraction and comments on their suitability and efficiency.

2.1 Semi-automatic Road Extraction

Semi-automatic methods delineate a road in the image based on the information on its initial position provided by an operator. The conventional methods consist of three major steps, i.e. road finding, road tracking and road linking. In road finding, the local properties of the image are tested. At this stage, an algorithm based on edge or surface information, or an algorithm combining edge and surface information is used to detect seed points. The detected seed points are then tracked to form road segments based on the contextual information of the image. In the case where occlusions exist due to shadows cast by trees and overpasses, vehicles on the road, surface anomaly, etc., the tracked road segments are often separated. Therefore, a linking process follows to

connect the separated road segments based on the geometric constraints. Table 2.1 lists some major semi-automatic methods and algorithms for road extraction developed during the last two decades. The following are the abbreviations used in the table:

LR-AP	Low Resolution Aerial Photograph
HR-AP	High Resolution Aerial Photograph
MSS	Multi-Spectral Scanner images
TM	Thermatic Mapper images
SPOT-P	SPOT Panchromatic images
SPOT-XS	SPOT Multi-Spectral images

Quam (1978) developed an algorithm for road tracking in high-resolution aerial photographs based on a road surface model and a path model. The path model is a list of the detected road points used for extrapolation of the road path. The surface model is an array of intensity values selected from the image in the direction perpendicular to the road path. With the initial position, direction and width of the road provided by an operator, the position of the next road point is predicted using the path model. The cross-section is then extracted from the image at the predicted point, and a cross-correlation is performed based on the surface model to determine the error in the predicted position. If the correlation peak is poor, the algorithm uses the path model to predict ahead for another point and an another profile is selected for testing the cross-correlation. This process will continue until a good match is found or the length skipped is larger than the longest expected anomaly. If there is a large number of anomalous points, a surface change is hypothesized and a new surface model is extracted. The system assumes that a road has a consistent surface pattern and a constant width, and that there are possible anomalous points and change in the surface material. However, the problems which may be caused by road width changes due to gradual changes in surface patterns that occur where a new lane is added to the road or a lane is deleted, are not considered.

Table 2.1 Overview of Semi-automatic Road Extraction

Author(s)	Year	Input Image	Spatial	Local Properties	Geometric Constraint	Radiometric Constraint
Quam	1978	HR-AP	0.3 -- 0.9	intensity profile	local parabola	consistent profile
Kestner and Kazmierczak	1978	LR-AP	4.5 -- 9.0	intensity profile	constant road width	constant intensity
Fischler et al	1981	LR-AP	?	ridge	constant road width smoothness	uniform intensity high contrast
McKeown and Delinger	1988	HR-AP	1.0 – 3.5	edge intensity profile	constant road width local parabola	consistent profile high contrast
Fua and Leclerc	1990	HR-AP	?	edge	smoothness	high contrast
Grün and Li	1994	SPOT-P	10	generic model	smooth, low curvature	homogeneity high contrast
Trinder and Li	1995	SPOT-P	10	edge or ridge	smooth curve	high contrast
Vosselman and Knecht	1995	HR-AP	1.6	intensity profile	constant curvature	consistent profile
Neuenschwander et al	1995	LR-AP	?	edge	smooth curve	homogeneity
Geman and Jedynak	1996	SPOT-P	10	ridge	low curvature	homogeneity high contrast
Grün and Li	1996	HR-AP	0.6	edge template	smooth curve	homogeneity, high contrast

Kestner and Kazmierczak (1978) used a correlation follower and a region-based follower for road extraction from medium resolution images. The correlation follower compares the road intensity profile with the expected road intensity profile against the background in the direction of the predicted road path, when the initial position and direction of the road are given by an operator. When the correlation follower fails to find an acceptable path, a region-based method is utilized to re-acquire the road. The region-based method extracts a 2-D area from the image and searches the whole area for points with the expected intensity profile. Each point in the area is marked with a score which indicates how well it matches the expected profile. All points except the best points are removed and the remaining points are then linked together using geometric constraints to find the most likely road path. It is reported that both correlation follower and region-based tracker complement each other perfectly, but it is not clear how they work together. A binary method is also described. The method first generates a binary image of the section of the image containing the road by thresholding all points within the expected road intensity range to 1 and other points to 0. It then eliminates regions which are too wide to be part of the road and links the remaining thin regions.

A road tracing algorithm based on least squares profile matching and Kalman filtering is presented in Vosselman and Knecht (1995). With the initial road segment selected by an operator, an estimation of parameters that describe the position and shape of the road is made using Kalman filtering. This estimate allows the prediction of the position of the next road profile. The profile at the predicted position is chosen and matched with the model profile which is the average of the cross-section profiles of the initial segment. The shift between the selected profile and the model profile is determined and used by the Kalman filter to update the parameters describing the position and shape of the road. The above process is applied again to estimate the road path until some stop criterion is fulfilled. The advantage of least squares matching is that it can estimate the precision of the profile shift which can be used to evaluate the success of the matching algorithm. Moreover, in least squares profile matching, not only the road position, but also the road width can be estimated. When the road width changes, least squares matching can obtain good results, while cross correlation will fail. Thus, the algorithm can cope with intersections, fly-overs and cars.

McKeown and Denlinger (1988) argue that a single correlation road tracking algorithm or a single edge-based road tracker cannot generate satisfactory results and presented a cooperative method for road tracking in aerial imagery. The proposed method consists of three different levels, i.e. low-level road tracking, intermediate-level feature detection and high-level symbolic description. In low-level road tracking, a Quam's road tracker and an edge-based tracker incorporating the road tracking method of Nevatia and Babu (1980) are used. These two trackers work independently to generate the estimate of the centre line of the road, its width and other local properties. Intermediate-level processes monitor the state of the low-level road tracking and evaluate the results of each tracker. Six detectors are used to detect road intersections, width changes, overpasses, surface material changes, vehicles and occlusions. The high-level provides overall control and user interface. When one tracker fails, it will be restarted by the model generated by other trackers. Finally a symbolic description of the road will be generated in terms of a number of attributes of the road, such as positions of the centre line, road width, surface material, overpasses, and an indication of potential vehicles on the road.

A method for detecting roads in low-resolution aerial imagery is introduced in Fischler et al (1981), which combines local information from multiple sources, including various line and edge detection operators, maps which can be used to derive information about the likely path of roads in the image, and the generic knowledge about roads, to facilitate road extraction. The image operators are classified into two types based on their error characteristics. Type I operators almost never incorrectly classify artifacts as instances of the structure they are searching for, but may often miss correct instances of the structure. Type II operators can accurately determine the parameters of all true instances, but may falsely classify and incorrectly parameterize non-instances. Type I operators generate a score for each pixel which indicates the likelihood of local feature presence and Type II operators produce a cost array. The results produced by both Type I and Type II operators are combined in such a way that the cost array is modified so that zero cost corresponds to a very high likelihood of local road presence. The *a priori* knowledge can be introduced through the cost transformation. The best path of the road is found by minimising the sum of the costs along the road.

A model-driven method for road extraction from aerial and satellite images based on dynamic programming is presented by Grün and Li (1994). This method uses a generic model to describe a road in the image with geometric and photometric properties. The model is formulated by some constraints and a merit function. The optimization of the model is achieved by a time-delayed dynamic programming algorithm. Once the approximate position of the road is provided by an operator, the algorithm can automatically find the optimal road path. For road sharpening, a wavelet function is applied to enhance the image and generate a multi-resolution representation of the image. Because the geometric and photometric constraints are enforced globally, small gaps such as those caused by sparse trees can automatically be bridged. It can work in a monoplotting mode and in 3-D where two or more overlapping images are used to derive 3-D coordinates with an existing digital terrain model (DTM). The knowledge used includes photometric and geometric properties of a road.

Geman and Jedynak (1996) presented a model-based approach for road extraction from SPOT images. In their method, a road is represented as a set of segments. The difference in direction between two neighbouring segments is represented by a number in the set {0,1,-1} in which 0 means no change in direction while 1 and -1 stand for small deviations to the left and right. Thus, a road in the image is described by a sequence of numbers in {0,1,-1}, which is modelled as a stochastic process. Based on the assumption that the curvature of a road has an upper bound and that a road surface is homogeneous and has distinct contrast against its background, road tracking becomes a process of computing the Maximum *A Posterior* (MAP). The method has a good foundation in mathematics, however, it does not consider the problems caused by disturbances, such as trees, shadows, etc. which may cause tracking to fail.

In recent years, Snakes technique has attracted considerable attention from computer vision and digital photogrammetry. Snakes are active contour models in which a linear feature is represented as an energy model (Kass et al, 1988). The energy model contains internal and external energies which are defined by the geometric constraints and photometric properties of linear features respectively. The extraction of a linear feature

is achieved through the minimisation of total energies. A model-driven method for road extraction from aerial images based on the snake technique is presented by Fua and Leclerc (1990). In their method, the photometric energy of a linear feature is defined by the sum of image gradients along the feature, and the geometric energy is the sum of the squares of the curvatures of the feature. In their implementation, a road is represented by a polygonal curve forming the centre of the road and the associated road width. The sum of the photometric energies of the two boundaries is taken as its photometric energy, and its geometric energy is the sum of geometric energies of the two boundaries plus an additional term which is the sum of the squares of the differences in road width between two neighbouring vertices along the polygonal curve. This additional term enforces the parallel constraint. Therefore, problems such as occlusions or ill-defined edges on one side of the road can be overcome.

Trinder and Li (1995) developed a 3-D snake for road extraction from satellite images in which two or more overlapping images are used. In 3-D space, a linear feature on the ground is described by a cubic B-spline curve. Its geometric energy is defined by the first and second derivatives of the curve, and photometric energy is defined by the sum of the photometric energies in each image. The photometric energy in the image is computed from a Chamfer image which is derived from the edge image and in which the value of a point is determined by its distance to the closest edge. The total energy contains a third term which is used to define the boundary conditions. Once a few seed points are given in one image, the approximate 3-D coordinates of the feature are calculated using an existing DTM. During the iterative computation, the feature points are interpolated using the defined cubic B-spline. This method has been extended to road extraction from aerial images (Trinder and Wang, 1997a). In aerial images, a road is described by a ribbon. When the seed points are provided along one side of the road by an operator, the snake can automatically determine the positions of both sides of the road using the extracted edge information.

Grün and Li (1996) combined the least squares template matching with Snakes technique and presented a least squares B-spline snake (LSB-Snake) for road extraction from satellite and aerial images. In LSB-Snakes, a linear feature is approximated by a

B-spline curve and three types of observations are used to describe its properties. The first is the photometric observations in which a template describing an edge point in the image is used. The relationship between the template and the image patch is given by six parameters that model the geometric transformation (Grün, 1985). The seed points given by an operator are considered as the second type of observations. The smoothness constraints of the feature are introduced as the third type of observations. The final position of the feature is determined by minimising the L_2-norm of the residuals in all three types of observations which is equivalent to the total energy of Snakes. To obtain 3-D coordinates of the feature, two or more images and an existing DTM are used. Because the solution of the feature position is achieved by least squares method, a theoretical accuracy of the feature location can be determined. Another advantage of LSB-Snakes is that it has the ability to detect occlusions. In LSB-Snakes, occlusions are detected as blunders, and the photometric energy in the portion of the occlusion is adjusted as zero. In this case, the solution is totally determined by the geometric constraint.

2.2 Automatic Road Extraction

In contrast to semi-automatic approaches, automatic road extraction aims at locating a road in the image without input of its initial position from an operator. It has two tasks, i.e. recognition of a road and determination of its position. There are two types of automatic methods. The first follows the paradigm of the conventional methods, i.e. road finding, road tracking and road linking. They concentrate on the automatic determination of the starting position of the road. Road tracking and linking are then performed to determine the position of the whole road. The second type of approaches use artificial intelligence techniques to detect and recognise a road automatically. In these methods, a knowledge base containing the knowledge of a road and *a priori* knowledge about the image and the real world is used to recognise the road. Some major automatic road extraction methods are given in the Table 2.2.

An early attempt in automatic road detection was made by Bajcsy and Tavakoli (1976) when the first generation of earth resource satellite Landsat-1 images became available. The IFOV of Landsat-1 MSS images is 57×79 m^2, thus only major roads with three or four lanes could be found. A model for recognition of roads in Landsat-1 MSS band 2 images was presented based on its physical and geometric properties in the image. The road extraction starts with a threshold technique to detect road points. Road segments are found using 52 defined templates and are linked based on the criteria of proximity and collinearity. After linking, the short isolated segments are treated as noise and removed. The remaining segments are then labelled based on the positions and directions of their end points. In this method, it is assumed that the grey values of the road surface in the image are within a given threshold. In reality, they are often beyond the defined threshold because the grey value of the road surface in the image depends on not only the spectral properties of the surface material, but also on other factors such as the time and season of imaging, weather conditions, the spectral signatures of the surrounding objects, etc. Therefore, it is unavoidable that some non-road points may be detected or some road segments missed.

Nevatia and Babu (1980) presented a method for road extraction based on antiparallel lines. In their method, six masks with different edge directions are used to detect edge points. Each mask has a size of 5×5 pixels and an edge direction from 0° to 150° in intervals of 30°. The image is first convolved with the defined masks. The results of the convolution are the magnitudes and directions of edge points. Meaningful edge points are extracted by thresholding. The boundaries are traced by checking the directions of edges in a 3×3 window and are approximated to straight lines by linear segments

Table 2.2 Overview of Automatic Road Extraction

Author(s)	Year	Input Image	Spatial Re-solution (m)	Local Properties	Geometric Properties	Radiometric Properties	Topological Information	Contextual Informatio
Bajcsy and Tavakoli	1976	MSS	79	brightness	no	high intensity	no	no
Nagao and Matsuyama	1980	HR-AP	0.5	edge	elongated area	homogeneity	no	no
Nevatia and Babu	1980	LR-AP	?	edge	local anti-parallel lines	high contrast	no	no
Groch	1982	LR-AP	1.2 – 9.0	intensity profile	straight line	high contrast uniform grey value	no	no
Hwang et al	1986	HR-AP	0.75	semantic model	rectangular area	no	yes	yes
Zhu and Yeh	1986	LR-AP	3.0 – 4.0	edge	long, local antiparallel	intensity	yes	no
Wang and Howarth	1987	TM	30	intensity profile	local smoothness	high gradient	no	no
Lemmens et al	1988	HR-AP	0.28	edge	road width, length	average intensity	no	no
Vasudevan et al	1988	TM	30	ridge	?	?	yes	no
Ton et al	1989	TM	30	ridge	no	high gradient	no	no
Maillard and Cavayas	1989	SPOT-P	10	edge	local smoothness	grey value, edge magnitude	yes	no

Table 2.2 Overview of Automatic Road Extraction (continued)

Author(s)	Year	Input Image	Spatial Re-solution (m)	Local Properties	Geometric Properties	Radiometric Properties	Topological Information	Contextual Information
Cleynenbreugel et al	1990	SPOT-XS	20	ridge	terrain-specific	high gradient	yes	no
Wang et al	1992	SPOT-P	10	gradient profile	no	high contrast	no	no
Barzohar and Cooper	1993	HR-AP	?	geometric and stochastic model	consistent road width, smoothness	consistent grey value, high contrast	no	no
Zlotnick and Carnine	1993	HR-AP	?	edge	local anti-parallel lines	high contrast	no	no
Ruskoné et al	1994	HR-AP	?	edge	local anti-parallel lines	long segment high contrast	yes	no
Heipke et al	1995	HR-AP	0.25	edge	local parallel lines	homogeneity grey value	no	no
Gunst	1996	HR-AP	0.4	semantic model	width, area	grey value	yes	yes
Gong and Wang	1997	HR-AP	1.6	edge	no	grey value	no	no
Baumgartner et al	1997	HR-AP	0.25	edge	antiparallel edges	grey value	yes	yes

piecewise. It is assumed that a road has two parallel boundaries with opposite contrasts, called antiparallel lines. Therefore, road segments are found by searching antiparallel lines in the image. One obvious limitation of the method is that no information about the specific properties of a road is applied to the generated antiparallel lines. This means that the generated antiparallel lines do not necessarily correspond to road segments.

Due to the effects of image noise, poor contrast, occlusions, etc, the extracted segments are highly fragmented. They must be linked to form roadlike features before high-level processing is performed. An algorithm based on the criteria of proximity and collinearity was developed by Vasudevan et al (1988). The algorithm works in two steps. It first finds all segments within a given distance around one end of the given segment. The difference in spatial directions between the given segment and the found segments is then computed. The segment which has a difference within the given threshold is selected as the neighbour of the given segment. The algorithm can work well when the image has a simple structure. However, it may lead to wrong linking in the case where other features exist around the segments to be linked.

A low level expert system for road extraction from aerial photographs is described in Zhu and Yeh (1986). In this system, a road is defined as a homogeneous area bounded by two parallel edges. The road extraction starts with the generation of antiparallel pairs based on the extracted edges. Antiparallel pairs with grey values which are close to the mean of grey values of all antiparallel pairs are selected as seeds for road growing. In road growing, the focus of attention is defined on an end of the selected road segment, based on its width and end direction, and edge information is extracted. A road segment is extended by virtue of the goodness of the detected edge being considered. To judge the goodness of an edge, four measures, i.e., continuity, sharpness, straightness and divergence are used. A number of production rules are formed based on these four measures for the acceptance of the detected edge. One problem with the system is that the knowledge used for the selection of road seeds is not always true. Therefore, a verification process is required to testify the detected road segments. As well, the situation in which more than one edge occurs in the region of attention is not considered in the system.

An automatic road network interpretation system was developed at the French Institut Geographique National (Ruskoné et al, 1994). The system consists of a low-level process invoking starting point detection and road following, and high level hypothesis checking. The starting points are detected based on the heuristic knowledge of a road segment. Parallel edges with a considerable length and distinct contrast against the background are selected as road seeds. Road growing is based on the homogeneity of the texture of the road surface. On both sides of a road seed, regions with similar texture to the road seeds are selected as the possible extension. All possible road segments being tracked form a number of road paths. Each path has a cost, the minimum corresponding to the optimal path. In high level hypotheses checking, topological information is used to link disconnected road segments. Finally, the positions of road points are adjusted by the Snakes technique. The method suffers from the same problem as the method of Zhu and Yeh (1987) in the selection of road seeds because only limited knowledge is used and the texture of the road surface is easily affected by disturbances.

Barzohar and Cooper (1993) use a geometric-stochastic model for the automatic extraction of main roads from aerial images. This model is based on the assumption of the photometry and geometry of a road. Auto-regressive processes are designed to model the road centre line, road width, grey level within the road, edge strength at the road boundary, and regions outside the roads and adjacent to road boundaries. Roads are found by MAP (Maximum *A Posteriori* probability) estimation. To obtain road candidates, the image is divided into square windows and a dynamic programming technique is used to find the segments that fit the road model with high probability. The detected road segments are then extended by moving a test window in which dynamic programming is used again to produce the best estimation of road extension. Finally short segments which do not connect to the recognised roads are removed from the hypotheses.

Wang and Howarth (1987) developed a method for automatic road network extraction from Landsat TM imagery. Due to the low spatial resolution of Landsat TM imagery (30 m), a road in the image appears as a line feature with a width of one to two pixels. To detect lines, an operator based on the second derivative of a road in its normal

direction is presented, and 14 masks with eight different directions for the calculation of second derivatives are designed. The convolution of the image with the defined masks gives the magnitudes and directions of lines, i.e. a magnitude image and a direction image. A point with a magnitude larger than the given threshold and with two among its 8-neighbours having similar magnitudes and directions is selected as the starting point. Line following is then performed as graph searching, in which the pixels in the direction image are nodes and connections between two neighboring nodes that have similar directions are arcs. Each arc is associated with a cost which is defined by the magnitude of the first node. The path with minimum cost is taken as the final result. Because only local magnitude is taken into account in the selection of the starting points, the selected road seeds may not be true road points.

To overcome the problem of Wang and Howarth (1987) method, a two-phase method was developed for the automatic road detection from Landsat 4 TM images (Ton et al, 1989). The method consists of low-level road detection and high-level road labelling. The low-level road detection is similar to the method of Wang and Howarth, except that the definition of the masks is slightly different. The magnitude of the road operator is not only used for the selection of road seeds, but also to classify pixels in the image into different levels. In high-level road labelling, the detected road segments in low-level processing are linked using geometric constraints, which are represented in rules and are classified into different levels according to their lengths and curvatures. One advantage of the two-phase method is that it can avoid errors in the selection of road seeds, using the geometric measures of road segments. However, simple geometric measures cannot solve the problem completely.

Gong and Wang (1997) applied classification techniques to road network extraction from multi-spectral aerial images. They use a clustering algorithm, a supervised maximum classifier and a cover-frequency based contextual classifier to recognize roads from images. Four different types of road surfaces, i.e. new asphalt, old asphalt, concrete and railroad, and four different land covers, i.e. residential area, industrial area, well irrigated grass and dry grass, are selected for training. The classification results show that the highest accuracy, 74.5 percent, is achieved by the clustering method. To

compare them with other automatic road extraction methods, a gradient direction profile analysis (GDPA) method is selected. It was found that GDPA is superior to all three classification methods and it reached an accuracy of 78.7 percent.

Existing maps exhibit abundant information about objects in an image, including roads. Such information is very useful for road extraction from images. Maillard and Cavayas (1989) developed an automatic method for road extraction from SPOT imagery under the guidance of topographic maps with a scale of 1:50,000. The process starts with the extraction of the roads in the image that exist in the map after the map and image are registered correctly. It is assumed that there may be a deviation between the positions of a road in the map and in the image. Therefore, a search area for the road point in the image is defined along the existing road segment. A point in the search area is taken as a road point if its grey value and edge magnitude, and the ratio between the edge magnitude and absolute difference of grey values with its preceding point, are within the predefined thresholds. Small gaps between the detected road points are bridged by linear segments. Intersections between new roads and the existing roads are found by searching both sides of the extracted roads in a similar way. After the intersections are detected, the same procedure as above is applied to look for the new road segments. With the guidance of information about the existing roads, road extraction becomes easier and its reliability is improved.

Cleynenbreugel et al (1990) argue that the road model used in the existing methods is not complete because only the photometric and geometric properties of a road are utilized. The information about the terrain type and land cover is actually very important since different types of land cover correspond to different types of road network topologies, and different terrain types usually correspond to different road patterns. To incorporate the knowledge of terrain type and land cover into road extraction, a knowledge-based method for road extraction from multispectral SPOT imagery was developed. Based on the expert's knowledge, two different road models for forest path networks and mountain roads are given. An existing GIS is used to derive the information on the terrain type, land cover and existing roads. For each land cover subclass or terrain type, there is a specific knowledge base for implementing a typical

road network model, which consist of two types of knowledge: procedural and declarative. The procedural knowledge activates the segmentation algorithm to extract line features and the declarative knowledge is applied to the generated line features to derive the final road network. The use of knowledge about the terrain type and land cover can improve the quality of road extraction. However, the road models only contain the general properties of road structures and no specific properties of a road are used. This may generate spurious recognition of roads.

Heipke et al (1995) presented a hierarchical approach for automatic road extraction from aerial imagery. In their method, two different scales of images are used. In low resolution images, pixels exhibiting significant contrast against the background and grey values within the given threshold are detected and grouped into contours. To eliminate non-road points, the local properties of the detected points are examined. Only pixels which are the local maximum in the direction perpendicular to the contour are accepted as road points, and contours with a larger percentage of accepted pixels than the given value are retained. In high resolution images, edges with approximate directions are detected. The mean and variance of the grey values in the areas bounded by the parallel edges are computed and used to test the homogeneity of the road surface. Parallel pairs are accepted as road segments only if their mean and variance are within the predefined thresholds. Parallel pairs are extended in their neighbouring regions by checking the homogeneity in these areas. To increase the reliability of road extraction, the results from these two different scales of images are combined using a number of heuristic rules. The use of two different scales of images can improve the quality of road extraction because their results are used to verify each other. But this still cannot guarantee that the extracted roads are true.

Baumgartner et al (1997) explored the use of context in automatic road extraction. They analysed various situations at which roads are possibly related to other objects and classify context into context sketches and content regions. Context sketches describe relationships between roads and related objects while content regions define the type of an area. Road segments are extracted using the method described in Heipke et al (1995), starting from the easiest area using the knowledge of content regions derived from

texture analysis of the image. The gaps between the extracted road segments are then analysed using contextual information of roads, i.e. context sketches which are generated by using a Digital Surface Model (DSM). If the classified object corresponding to the gap between two road segments is consistent with the knowledge of context of roads, a road part is then hypothesised and the gap is bridged.

A knowledge-based system for automatic road extraction from aerial images was developed by Gunst (1996). The system uses a multi-scale semantic road model in which a road is represented in three different scales, i.e., small, medium and large scales. The model includes the definition of object classes that correspond to the meaningful parts of a road network, object-specific properties and the relationships between the object classes. Objects in three different scales are related by part-whole relationships and the generalised objects in each level correspond to specialised objects. The knowledge base is developed based on the semantic model, in which the properties of the object classes and relationships are represented in the form of object definitions and object relations. Object definitions contain the properties of geometric and radiometric properties of the each object class which are used for recognition, and object relations define the spatial, part-whole and specialisation-generalisation relationships between object classes. Priorities are attached in the object relations which determine the order of the subsequent reasoning. The contextual information is utilized to infer other objects and invoke image segmentation for searching new objects based on the recognised objects. The knowledge base also includes alternative object definitions and relations to generate alternative hypotheses and trigger re-segmentation with upgraded parameters, if inconsistencies occur. Outdated maps are used to generate hypotheses. When a hypothesis is generated, a chain of actions are taken using the relationships in object relations.

2.3 Summary of Previous Work

A large number of existing methods and algorithms for automatic road extraction from aerial images and satellite imagery have been reviewed. These methods and algorithms

are classified into semi-automatic and automatic, according to the degree of human intervention. In semi-automatic methods, a road is defined as a linear feature or a homogeneous surface. Thus, the task of road detection becomes the tracking of an edge or a homogeneous surface after the initial position and orientation of a road are given by an operator. Road points are detected using an edge operator or a surface profile analyser, or their combination. An edge operator works in a local area and computes the magnitudes of gradients of an image. A thresholding process is applied to classify pixels into road points and non-road points. As only local radiometric properties of pixels are used, it is impossible to determine road points accurately. Surface profile analysers detect road points by comparing road surface profiles with the surface model. When surface anomalies occur, no or wrong points will be found. The combination of an edge operator and a surface profile analyser can improve the quality of road detection as they can supplement each other. However, they may fail when poor contrast and a surface anomaly occur at the same time. In road tracking, detected road points are traced using simple geometric constraints. Since these constraints are enforced locally, non-road segments may also be tracked. Due to the effect of occlusions, such as trees, shadows, vehicles, overpasses, etc., the tracked roads are usually separated. Therefore, a linking process is usually included to connect separated road segments. In road linking, geometric constraints are used locally. This may lead to wrong connections. To solve these problems completely, knowledge of context and the overall structure of the road network should be used.

In active contour models (Snakes), a road is modelled by energy which is represented by the geometric and radiometric constraints. The extraction of a road is achieved through the optimization of the energy. The Snakes technique has two advantages: (i) the geometric constraints are directly used to guide the search for the feature, and (ii) the photometric information is integrated along the entire length of the curve, providing large support without including the irrelevant information of points not on the feature (Fua and Leclerc, 1990). As the initial positions of the feature are provided by an operator, the overall structure of the road is implicitly defined. This can ensure that the extracted results are reliable and accurate. Even when there are some points which cannot be detected by an edge operator or small occlusions such as trees, shadows,

overpasses, change of surface materials, etc. Snakes can determine the positions of the feature correctly, as geometric constraints are enforced during the optimisation process. Therefore, the Snakes technique is more robust and efficient than the traditional methods. However, it cannot completely solve the problem caused by large occlusions, e.g. a road going through a forest. To solve these problems, knowledge of the context of objects is required.

Automatic road extraction includes automatic determination of positions and recognition of a road. In most existing methods, much effort has been made on automatic detection of the starting point or segment of a road. The radiometric properties of the point or segment, such as intensity values and contrast against the background, are usually used to classify pixels as the starting point or segment of a road. Since they vary with many factors, such as time, season and weather conditions of photography, the same object in different images may have different grey values, while different objects in one image may have similar intensity. Therefore, the radiometric properties alone cannot classify an image point (segment) as a road point (segment) or a non-road point (segment) reliably.

Classification is a common technique for interpretation of an image, which uses spectral information of objects in different spectral bands. However, in aerial images or high-resolution satellite imagery, the spectral values of an object may change greatly due to the image noise, its texture, different weather conditions, sun angle and slope, etc. This may lead the existing classification methods to fail.

To recognise an object from an image correctly, not only the geometric and radiometric properties of the object should be used, but also the relationships between the object and its surrounding objects should be considered. Knowledge-based systems provide a good mechanism for representing and manipulating various types of knowledge. They use the knowledge of geometric and radiometric properties of objects, their structures, relationships with other objects, and *a priori* knowledge of the imaging system and the related world to interpret the objects. As various types of knowledge are used, more

reliable results can be achieved by knowledge-based methods. Most existing methods are based on a specific road model in which specific properties of a road are included. To generate reliable results for most cases, more general semantic road model is needed and more complete knowledge should be used.

Chapter Three

Concepts in Image Understanding

Image understanding has been an important research area in computer vision for decades. Its purpose is to construct a corresponding relationship between image features and objects in the real world, i.e. to derive recognition of objects from images with respect to domain knowledge. During the last two decades, a number of theories and methods have been developed for object recognition, including Marr's theory of vision. These theories and methods can be adopted for photogrammetry for recognition of objects from aerial and satellite images. This chapter will introduce some general concepts of image understanding and processing involved in image understanding. Some existing image understanding systems will also be reviewed.

3.1 Process of Image Understanding

Image understanding is a very complex process. This is mainly because there is a significant difference in the representation of an image and objects in the real world. An image records the radiometric properties of objects in the real world by projection. It is an array of pixels which only possess information of intensities and positions while objects in the real world are usually linked by relations (Ballard and Brown, 1982). It is impossible to apply object matching process to the image directly unless the domain is extremely simple and heavily constrained (Hanson and Riseman, 1988). To bridge this gap, intermediate representations such as points, edges/lines, regions, object structures and their relations are required. With these features, object structures and relations, a link between the image and objects in the real world may be established with respect to domain knowledge. Therefore, an important task of an image understanding system is to design the appropriate representations and algorithms to relate them to each other. Representations are generally classified into three levels, i.e. *low-*, *intermediate-* and

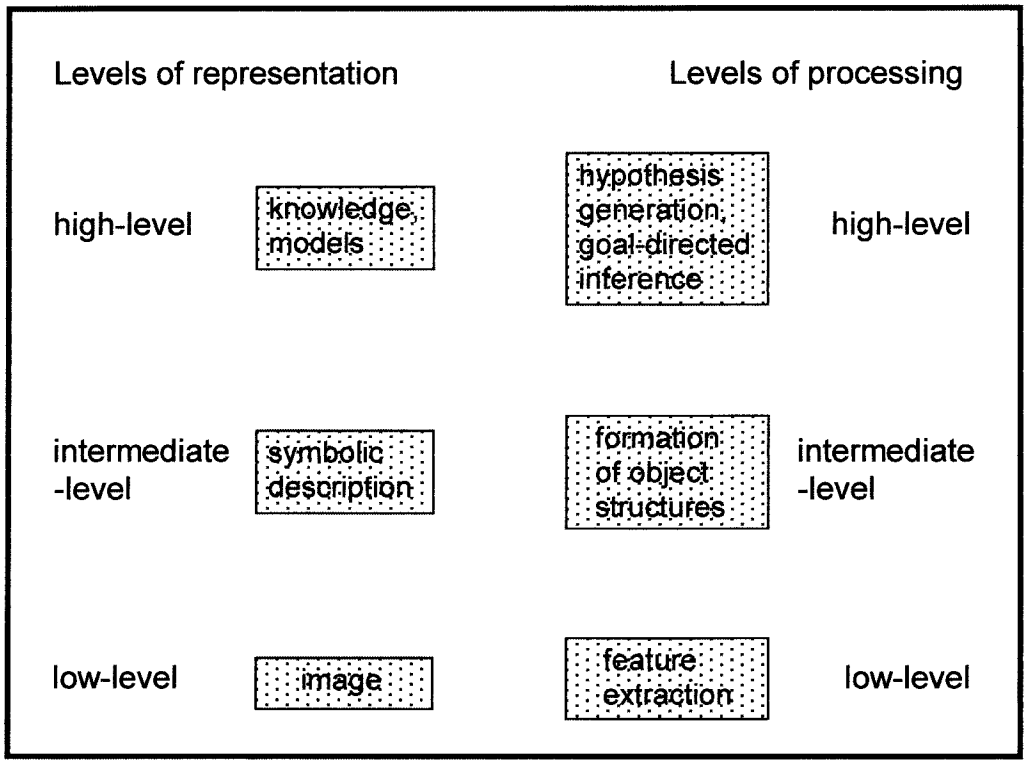


Figure 3.1 Levels of Processing and Representation

high-levels (Marr, 1982; Hanson and Riseman, 1988; Förstner, 1993) although they might be categorized into four levels (Ballard and Brown, 1982). *Low-, intermediate- and high-level representations* correspond to iconic, symbolic descriptions of objects, and knowledge and models of objects respectively (Figure 3-1).

At *low-level representation*, input data, i.e. the image is represented. An image is the projection of objects in the real world. It consists of an array of pixels which have intensity or grey values and positions in the image plane. With intensity values of images, it is only possible to classify large areas in satellite images or objects with very simple structures and distinguishable spectral properties. However, it is impossible to generate recognition of objects in large scale images if only intensity values are used because different objects may have similar or even same spectral properties. The image must be processed to derive meaningful information about the objects, such as points, edges/lines and regions.

Intermediate-level representation refers to symbolic representation. At this level, symbolic tokens for points, edges/lines and regions with attributes which can be extracted from the input data, are represented. At the same time, object structures which are formed by grouping image features are also represented symbolically in terms of their associated attributes and relations. They are stored in what is called the "Short Term Memory (STM)" of the knowledge base.

At *high-level representation* are the knowledge and models of the objects, which are the highest abstraction of objects. It describes the structures of objects to be interpreted, their properties and relationships in the real world. They are expressed in explicit forms such as rules and frames, and stored in the so called "Long Term Memory (LTM)" of the knowledge base. They are applied to the derived symbols from images to generate interpretation of objects, or to invoke appropriate processes to find instances of objects to be interpreted.

To derive recognition of objects from the image, a number of processes are usually required, which can be divided into three levels, i.e. *low-, intermediate- and high-level* (Hanson and Riseman, 1988) (Figure 3-1). *Low-level* processes manipulate pixel data and produce features such as points, lines /edges and regions. These features constitute the structures of objects. They can be extracted using a single algorithm or several algorithms or operators. There are numerous operators available for the extraction of these features. Most edge and point operators are based on the discontinuity of intensity, while region segmentation algorithms usually use homogeneity property of the image such as texture, chromatic values, etc., to extract homogeneous areas. For edge and area extraction, their direct results are feature points with attributes. Therefore, some processes such as edge tracking and linking, region growing, splitting and merging are included in *low-level processing* to form continuous edges and regions. Generally, no knowledge about the objects to be interpreted is used at this stage.

The results of *low-level processing* are unstructured points, lines/edges and areas. They are meaningless before the structures of objects are formed. For recognition of objects,

these features produced in *low-level* processing need to be grouped to form the structures of objects. At the same time, relationships between objects should also be established. This process is referred to as *intermediate-level processing*. Grouping is a process of aggregating relevant features to form the structures of objects using the attributes of features and their relationships. At this level, some general knowledge of objects, e.g. a road has two parallel boundaries, an industrial building has rectangular shape, etc., may be used. The use of general knowledge of objects can save computation time and overcome problems caused by image noise occlusions (Ballard and Brown, 1982).

High-level processing is based on the knowledge and models of objects stored in the knowledge base. It has the function of generation of object hypotheses and goal-directed inference. Generating object hypotheses is a process in which *a priori* knowledge and models of objects are applied to the structures and relationships of objects produced in *intermediate-level processing*. If the structures and relationships match the models or definitions of some objects in the knowledge base, the corresponding objects are hypothesised. Goal-directed inference starts with hypothesis generation of objects and invokes appropriate image processing algorithms to find instances of the hypothesised objects in the image, or triggers grouping algorithms to find complete structures of objects, based on partial hypotheses made in the previous stage. When the found features or structures meet the definitions of objects in the knowledge base, the objects are confirmed. Which processes are performed strongly depends on the selected control strategy. This will be discussed in detail in Section 3.3.

In Ballard and Brown (1982) and Sonka et al (1993), processes involved in image understanding are classified into two levels, i.e., *low- and high-levels. High-level processing* corresponds to cognitive processes, while *low-level processing* includes all processes for generating symbolic representations of objects. The number of processing levels in image understanding is not so important. However, it is generally believed that processes for generation of symbolic representation of objects are necessary, unless the

domain is extremely simple and heavily constrained so that object matching processes can be applied directly to the image to derive the interpretation of objects.

3.2 Knowledge Representation

The information which can be derived from the image consists of image features such as edges/ lines and regions and structures with their attributes and relations. They are insufficient for recognition of objects in an image. To recognize objects, the specific properties of objects, their relationships and the knowledge related to the image and objects such as *a priori* knowledge of the imaging system, etc., are required. In artificial intelligence, knowledge is defined as *the understanding of a given subject* (Durkin, 1994). It refers to all known and useful information that would lead to the recognition of objects. It includes the types of objects which may occur in the image, geometric and radiometric properties of the objects, and their mutual relationships. It also includes *a priori* knowledge about the world related to the objects and the imaging system, such as the terrain type of the area where the objects are located, and orientation parameters of the imaging system.

A knowledge base is a collection of knowledge related to solving a specific problem. Research in cognitive psychology has found that humans use different types and organisation of knowledge for solving a problem. Scientists in artificial intelligence have used the results of these studies to develop techniques to best represent these different types of knowledge in the computer. This section will introduce the types of knowledge and some major methods of knowledge representation.

3.2.1 Types of Knowledge

According to the discovery of cognitive psychologists, there are different types of knowledge humans commonly use to solve problems. These include *heuristic*, *procedural*, *declarative*, *meta-* and *structural knowledge* (Durkin, 1994). *Heuristic*

knowledge describes a rule-of-thumb that guides the reasoning process. It is empirical and is formed through the experience of solving past problems by an expert. *Heuristic knowledge* is simple and easy to implement. However, it may not lead to a correct solution in exceptional circumstances because of its experimental nature.

Procedural knowledge describes how a problem is solved. This type of knowledge is represented in the form of explicit process descriptions. Like a computational program, *procedural knowledge* not only defines the step-by-step sequences of performing a task, but also specifies the optimal algorithms to accomplish the sub-tasks. In an image understanding system, knowledge used to search for the evidence of objects or object parts is usually procedural. It contains the description of a search area and what image processing algorithms will be performed to find the instances of objects.

Declarative knowledge describes what is known about a problem. It is an assertion of the existing facts. This type of knowledge includes simple statements that are asserted as either true or false. It may also include a list of statements that more completely describe objects. For example, the following is a piece of *declarative knowledge* describing a table:

– a motorway has a smooth surface and two parallel boundaries.

In image understanding, the properties and relationships of objects are described by *declarative knowledge*. They include geometric and photometric properties of objects and object parts, semantic and spatial relationships.

Meta-knowledge describes knowledge about knowledge. It is used to guide the use of knowledge for solving a problem. It can improve the efficiency of problem solving as it can direct reasoning to the most promising areas and determine the most suitable knowledge for the given task.

Structural knowledge describes knowledge structure. The expert's mental model of concepts, sub-concepts and objects belongs to this type of knowledge.

3.2.2 Knowledge Representation

Knowledge representation is a major subject of artificial intelligence. Researchers in artificial intelligence have developed a number of methods to represent different types of knowledge. An excellent survey of these techniques is given in Barr and Feigenbaum (1981), Durkin (1994) and Crevier and Lepage (1997). Here only four of the most common techniques are introduced. They are *rules*, *semantic networks*, *frames* and *logic*.

3.2.2.1 Rules

A commonly used form of knowledge representation is *rules*. A rule is a knowledge structure that relates some known information to other information which can be inferred or concluded to be known. In a rule, known information is expressed as premises or conditions, while information to be inferred is expressed as conclusions. They are logically connected in the form of IF-THEN clauses. For example:

IF It is raining
THEN Take an umbrella

The action contained in the THEN part may be an assertion of a new fact or some procedures to be performed. In this sense, a rule can represent both *declarative* and *procedural knowledge*. In order to perform more complex operations, most rule-based systems are designed to access an external program. For example:

IF Two neighbouring road segments are not connected
THEN "find_missing_segment" between two segments

In general, a rule can have multiple premises joined by AND statements, OR statements, or a combination of both. Its conclusion part can also include several statements joined by an AND. The following is an example of a general rule:

IF	It is a road segment
AND	It is connected to a house
THEN	It is a driveway

In a rule-based system, domain knowledge is represented in a set of rules in the knowledge base. The system uses these rules together with information stored in the working memory (facts) to solve a problem. When the IF part matches the information contained in the working memory, the rule will be fired and the action defined in the THEN part of the rule is performed. The result of action may be a new fact or execution of a procedure. The produced new fact or the result of executing the procedure is added to the working memory in the knowledge base.

For solving complex problems, different sets of rules are required, which can be structured in such a way that they can exchange information with each other. Such a structure is called a *blackboard* in which each set of rules, called a module, performs a sub-problem. It shares and uses information with other modules by writing information on a blackboard and reading information from other modules to accomplish its task.

Since rules are an explicit representation of knowledge, they are easy to understand. Inference and explanations can be easily derived from them. Rules are stored independently, and hence they can be easily modified and maintained. However, solving a complex problem often requires many rules, as each rule can only represent one aspect of the problem. This may create problems in using and maintaining the system. Searching for the solution is also a problem in systems with many rules.

3.2.2.2 Semantic Networks

A semantic network is a method of knowledge representation using a graph structure. It was first introduced by Quinlian (1968). A *semantic network* consists of a number of nodes and arcs connecting the nodes (Giarratano and Riley, 1994). Nodes represent objects, object properties or property values, while arcs represent relationships between nodes. Common relationships used in *semantic networks* are "IS-A", "AKO" and "PART-OF". The IS-A means "is an instance of" and refers to a specific member of a class. For example, a canary is a bird. Thus, their relationship can be represented by IS-A. The AKO relates an individual class to a parent class to which the individual class belongs. For example, a bird is a kind of mammal. Therefore, a hierarchy among objects can be established through this relationship. The AKO is not used to relate a specific individual member to a class. "PART-OF" relate parts of an object to the object. It is used to describe part-whole relationships of objects. Figure 3.2 shows a simple example of a *semantic network.*

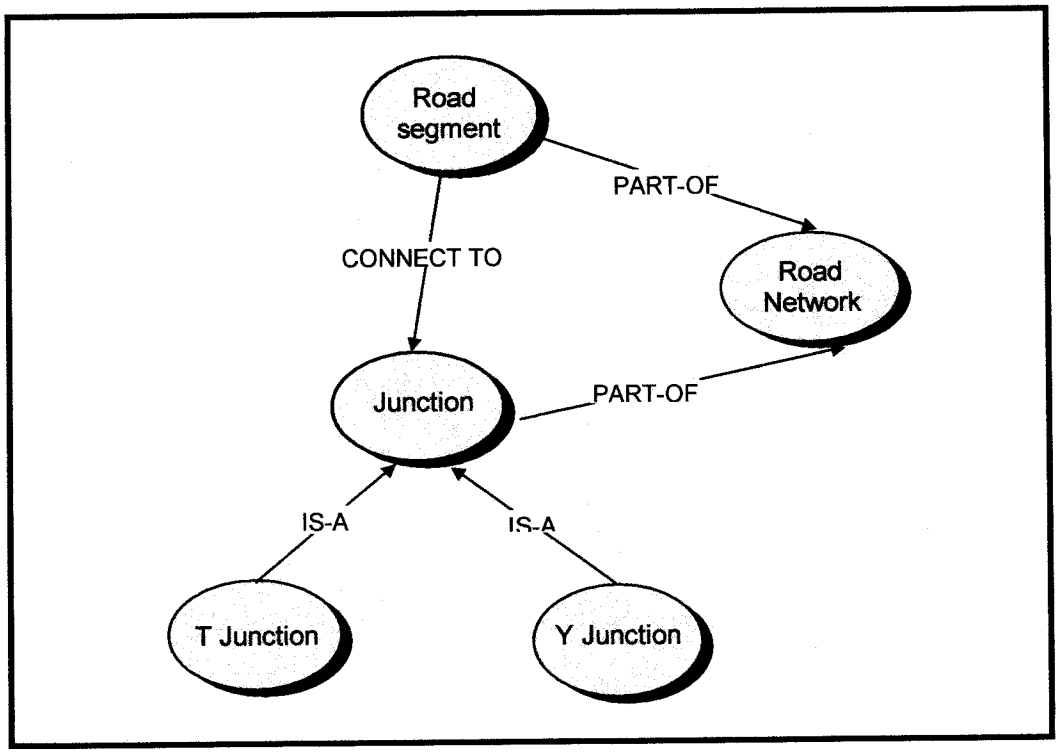


Figure 3.2 An Example of a Semantic Network

One important characteristic of *semantic networks* is inheritance. The information possessed by an object in the *semantic network* can be automatically inherited by objects related to it with the relationship "IS-A". The inheritance nature of *semantic networks* can ease the task of coding knowledge as some common properties of different objects in the *semantic network* are not required to be represented repeatedly.

The *semantic network* represents knowledge in visual form, which is easy to understand. A knowledge base built up in this form can be extended by simply adding new nodes and arcs to the defined network. However, knowledge does not appear in a computer in graphic form. Instead, various objects and their relationships are represented in the knowledge base in other forms of representations, e.g. *rules*. Searching within a *semantic network* with many nodes and arcs may be difficult. Another problem is that *procedural knowledge* is difficult to represent as the sequence and time are not explicitly represented. Therefore, the technique is used mainly for analysis purposes.

3.2.2.3 Frames

A *frame*, also called schema, is an object-oriented method of knowledge representation which was first proposed by Minsky (1975). In contrast to other representation methods, the knowledge that describes one object is grouped into a single unit, a *frame*. A *frame* contains a name, usually the name of the object to be represented, its associated attributes and relationships with other objects. In a *frame*, the knowledge is partitioned into slots. Each slot contains one property and a value or a range of values of the object. An external procedure that performs a certain action can be attached to a slot. It can be used to derive the slot values or remove slot values. Therefore, both *declarative* and *procedural knowledge* can be represented in *frames*. A basic structure of a *frame* is shown in Figure 3.3.

Frames can be classified into *class frames* and *instance frames*. *A class frame* represents the general properties of some common objects. For example, a frame "bird" which describes the general characteristics of birds is a *class frame*. The common properties of

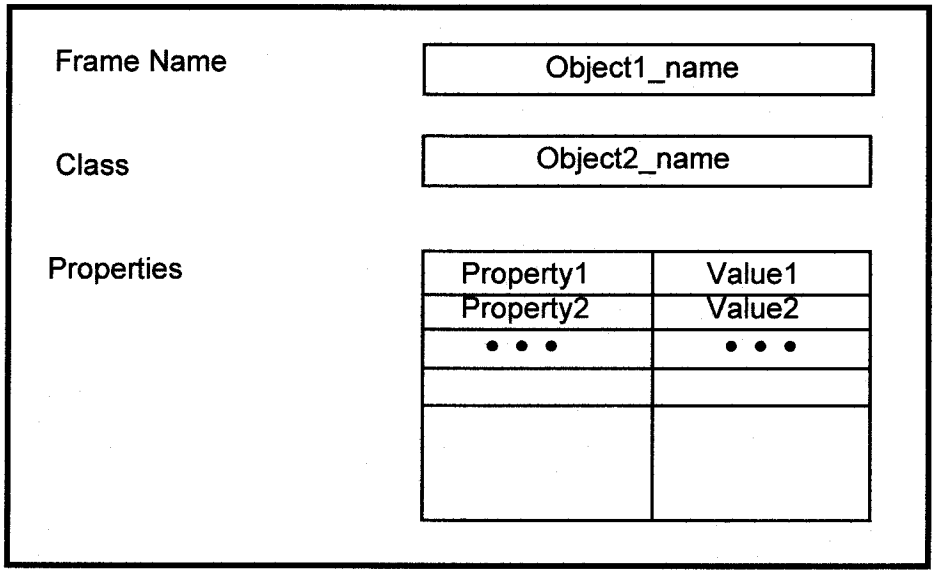


Figure 3.3 Structure of a Frame

various types of birds are included in the frame. An *instance frame* describes the properties of a specific object, e.g. a canary. It contains information more specific to its individual category. A relationship is defined in an *instance frame* which relates the frame with a *class frame*. The defined relationship usually has the form of "IS-A" or "AKO". Therefore, a hierarchy among *frames* can be established. More generic *frames* are at the top of the hierarchy, and specific *instance frames* are at the bottom. Each *instance frame* at the bottom of the hierarchy inherits the properties of objects from its top level *frame*.

There are significant advantages to object-oriented knowledge representation. First, *frames* add a third dimension by allowing objects to have structures, while *semantic networks* are basically a two-dimensional representation of knowledge. Such an important feature of frames is very useful in representing objects with complicated structures which are not easily represented by nodes and links. Secondly, a slot can have external procedures or frames attached to it. This enables spatial and contextual reasoning in *frame-based* image understanding systems.

3.2.2.4 Logic

The oldest form of knowledge representation is *logic*. *Logic* uses symbols to represent knowledge and operators applied to symbols to produce logical reasoning. Two widely used logic representation techniques in artificial intelligence are *propositional logic* and *predicate logic*. *Propositional logic* represents and manipulates knowledge with propositions. A proposition is represented by a symbolic variable, which can be of any form, e.g., A, PERSON, ROAD etc. They are connected by logical operators such as "Λ", "V", "~" and "→". For example,

If it is raining then take an umbrella

can be represented in the form:

$$A \rightarrow B$$

where

A = it is raining

B = take an umbrella

and → stands for IF...THEN.

Although the *propositional logic* is useful, it does have some limitations. The major problem is that the *propositional logic* can only manipulate complete statements. In other words, it cannot deal with the internal structures of statements. To provide a finer representation of knowledge, *predicate logic* was developed.

Like *prepositional logic*, *predicate logic* uses symbols to represent knowledge. *The first order logic* is the simplest form of *predicate logic*. Instead of representing a complete proposition with a single symbol, the *first order* logic permits symbols representing

constants, *predicates*, *variables* and *functions*. In this way, the properties of objects and their relationships can be represented by the use of *variables*, *predicates* and *functions*. Thus, knowledge processing is enhanced.

In *the first order logic*, variables are used to denote objects or object properties. The relationships between objects or properties are represented by *predicates*. For example, the connectivity relationship between two objects can be expressed as:

adjacent (X, Y)

The properties of an object are described by functions, e.g.

tall (X).

The first order logic operates with knowledge using the same operators found in *propositional logic*. At the same time, it introduces two other symbols called *quantifiers* which define the scope of variables in a logical expression. The two *quantifiers* are *universal quantifier* $\forall$ and *existential quantifier* $\exists$. The *universal quantifier* $\forall$ indicates that the expression has a true value for all values of the designated variable while the *existential quantifier* states that the expression is true for some values of the variable.

As *the first order logic* uses symbols to represent knowledge, the properties of objects and relationships between objects can be easily represented and implemented. Therefore, it is selected as the knowledge representation tool in this study. It should be pointed out that some statement may not be expressed by *the first order logic* properly, e.g.

It probably rains today!

In this statement, the "probably" means with large chance. It cannot be expressed in terms of the universal existential qualifiers. To describe "probably", a logic must provide some predicates for counting, as in *fuzzy logic* which is beyond this study.

3.3 Control Strategy

Control strategy plays an important role in image understanding. It determines the execution order of various processing involved in object recognition. This section explores different strategies for controlling information processing involved in image understanding.

3.3.1 Hierarchical Control Strategy

Under hierarchical control, different levels of processing are organized according to their order in the process of image understanding. Hierarchical control has three different forms, i.e. bottom-up, top-down and hybrid control strategies.

3.3.1.1 Bottom-up Control Strategy

Bottom-up control, also called data-driven control, proceeds from image data to interpretation of objects through the formation of object structures and relationships (Figure 3.4). Bottom-up image understanding starts with feature extraction in which raw image data are converted into feature points, lines/edges and regions which are supposed to correspond to the meaningful parts of objects. As no knowledge is used at this stage, the processing is domain-independent. The extracted points, lines/edges and regions are then grouped to form the object structures and relationships. Symbolic descriptions of objects are generated for the high-level recognition process. Finally, the knowledge and models of the objects are applied to the derived object structures and relationships to generate recognition of the objects.

Obviously, the results of bottom-up image understanding are heavily dependent on the quality of low- and intermediate-level processing. Good results can be expected if unambiguous data are processed and the lower levels of processing give reliable and precise representations for later processing steps. This is true when the objects have

simple structures and the image is of high quality, e.g. recognition of industrial parts in robot vision. However, when the image data contains complex structures and disturbances, e.g. aerial and satellite images, the derived structures of the objects by low- and intermediate-level processing may not be complete and correct, as errors may be introduced and useful information could be lost in this process. Thus, unreliable interpretation of objects may occur.

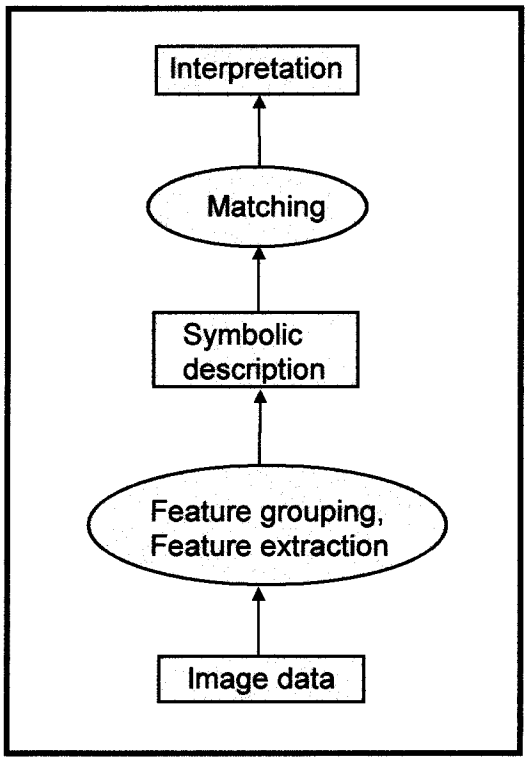


Figure 3.4 Bottom-up Control

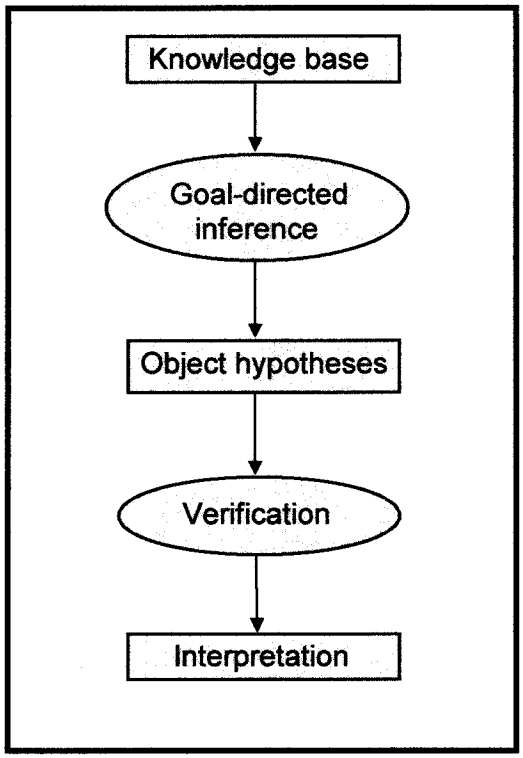


Figure 3.5 Top-down Control

3.3.1.2 Top-down Control Strategy

Top-down control, also called model-driven control, is a goal-directed process (Figure 3.5). Goals at higher processing levels are divided into sub-goals at lower processing levels, which are divided again into sub-goals until they are simple enough to find instances in the image. A common top-down procedure is "hypothesise-and-verify". The hypotheses of objects or object parts are generated based on the knowledge in the knowledge base, in which the structures of the objects or object parts and their positions

in the image are predicted. The appropriate image processing algorithms are then invoked to find the instances of the predicted objects or object parts in the predicted area. If the evidences of the objects or object parts are found in *low-level processing*, the hypotheses are confirmed. Otherwise, the alternative processing algorithms will be selected and tested again.

Model-driven image understanding relies on the knowledge and models of objects stored in the knowledge base. To achieve reliable results, the properties of objects and their relationships should be included in the knowledge base. This is especially true for natural objects. Goal-directed processing can avoid a large amount of *low-level processing* and potential errors introduced at this stage, as algorithms and parameters suitable to the tasks to be performed are selected. However, goal-directed inference needs information about the initial positions of the objects. Otherwise, its advantages will be lost.

3.3.1.3 Hybrid Control Strategy

Hybrid or combined control strategy uses a combination of both bottom-up and top-down control strategies (Figure 3.6). It usually gives better results than a single control strategy and is widely used in today's image understanding systems. The top-down process uses knowledge in the knowledge base to guide the search for objects at lower levels of processing, so it makes low-level processing easier. However, when the knowledge representing the objects is not complete and accurate, it might be difficult to predict object appearances and positions precisely and reliably. In this case, the incorporation of bottom-up control is very useful since some properties of the objects can be derived through this process.

In hybrid control, the way of combining bottom-up and top-down control strategies depends on the problem to be solved. One common hybrid control strategy uses bottom-up control to generate initial interpretation of objects and top-down control to find the missing object structures and instances in the image, based on the derived partial interpretation in the bottom-up process. Both bottom-up and top-down control strategies

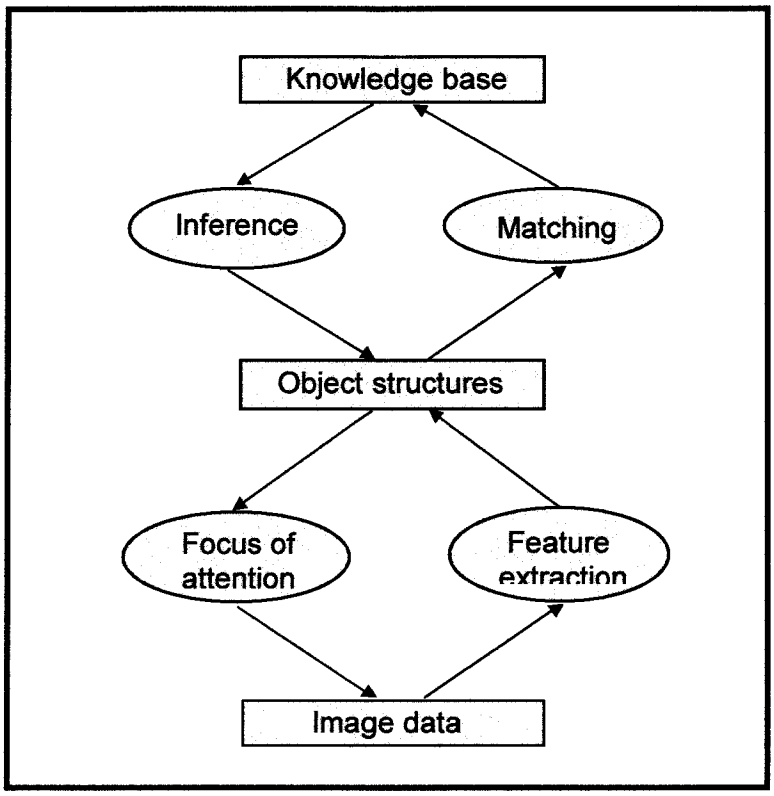


Figure 3.6 Hybrid Control Strategy

complement each other and therefore, better results can be produced.

3.3.2 Heterarchical Control Strategy

In heterarchical control, the processes are viewed as cooperating and competing experts at the same level, and the expert who can help most in finding the solution of the subtask is selected. Knowledge is shared among the experts. While new pieces of evidence are found and new hypotheses are generated, an appropriate expert is chosen to analyse them and create new hypotheses. Each expert relies on the information the other experts supply.

Blackboard is an application of the heterarchical control strategy. In a blackboard system, the results derived from independent knowledge sources are directly written on the blackboard or short term memory (STM). Different knowledge sources exchange information by writing on the blackboard. Potential actions awaiting execution are

recorded in the agenda of the blackboard. A scheduler is used to decide which knowledge source should be performed next, based on those stored in the agenda. The consistency among the representations of the emerging solutions is maintained by a consistency enforcer.

3.4 Overview of Knowledge-based Image Understanding Systems

Over the last two decades, a large number of knowledge-based systems for interpretation of objects from remotely sensed imagery have been developed. These systems use different knowledge representation techniques and control strategies and, therefore, exhibit different structures. A good survey of the existing knowledge-based image understanding systems can be found in Argialas and Harlow (1990), Crevier and Lapage(1997) and Trinder and Sowmya (1997). This section will review some of the existing systems based on the knowledge representation approach and control strategy used. Some examples are listed in Table 3.1

VISIONS is an early knowledge-based system for interpreting natural scenes developed at the University of Massachusetts (Hanson and Riseman, 1978, 1987). VISIONS is based on a number of assumptions. The first assumption is that the initial segmentation proceeds in a bottom-up fashion by extracting information from the image without knowledge of its contents. The second is that every stage of processing is inherently unreliable. Finally it assumes that local ambiguity and uncertainty in object hypotheses can be removed only by satisfying the relations between objects and object parts. The system consists of three different levels of representation and processing, i.e. low-, intermediate- and high-level. It uses a hybrid control strategy in which a bottom-up process is performed to generate partial interpretation of objects through building intermediate symbolic tokens of objects. A top-down control is then applied to direct further grouping, splitting and labelling processes at the intermediate level, to construct complete object structures by activating the knowledge related to the hypothesised objects. The knowledge of objects is represented by schemas (frames) each of which has

Table 3.1 Overview of Knowledge-Based Image Understanding Systems

Author(s)	Time	System Task	Input Image	Knowledge Representation	Control Strategy
Hanson & Riseman	1978	interpretation of natural scenes	close-range image	frames	hybrid control
Nagao & Matsuyama	1980	interpretation of suburban scenes	multi-spectral aerial image	production rules	heterarchical control
Levine & Shaheen	1981	interpretation of natural scenes	close-range color image	production rules	bottom-up control
McKeown et al	1985	interpretation of airports	aerial photographs	production rules	bottom-up control
Matsuyama & Hwang	1985	interpretation of houses and roads	aerial photographs	frames	hybrid control
Nicolin & Gabler	1987	interpretation of suburban scenes	aerial photographs	semantic networks	hybrid control
Strat & Fischler	1991	interpretation of natural scenes	intensity, range, color image	rules (context sets)	top-down control
Füger et al	1994	interpretation of man-made objects	aerial photographs	blackboard	bottom-up control
Stilla	1995	interpretation of urban scenes	aerial photographs	blackboard	bottom-up control
Gunst	1996	interpretation of highways	aerial photographs	frames	top-down control
Schilling & Vögtle	1996	interpretation of areal objects	Landsat TM-2	semantic networks	bottom-up control
Koch et al	1997	interpretation of landscape	aerial photographs	semantic networks	top-down control

a declarative and procedural component. The class and instance schemas are related by PART-OF and IS-A relations. Low-level processing includes two distinct parallel segmentation algorithms. The first aggregates edges into boundaries while the second uses a global histogram and a local spatial analysis procedure to form regions. The extracted regions, lines and surfaces with attributes are aggregated to form structures by grouping, splitting, and/or modifying and represented symbolically in intermediate-level processing. The initial hypotheses of objects are created by applying the object hypothesis rules to the generated tokens. Once an object is hypothesised, the knowledge about its properties will be used to guide further intermediate-level processing to find a complete object structure and instance in the image.

Levin and Shaheen (1981) described an image understanding system for segmentation and interpretation of natural scenes. The system has a modular structure and interprets scenes in a bottom-up mode. The modules incorporated in the system include low-level processors, a measure analyser, hypothesis initialiser, hypothesis verifier, focus of attention and scheduler. They communicate through the long-term memory and short-term memory similar to blackboard architecture. The low-level processors determine basic properties of regions such as the mean value and standard deviation of three colour features. The feature analyzer computes measures over regions and other structures. The hypothesis initialiser utilizes the region descriptions and model information stored in the long-term memory to create interpretations. The hypothesis verifier uses measures of regions, relations between regions and the generated hypotheses to confirm and update interpretations. The focus of attention module recognises situations of interest and generates actions. The scheduler controls execution of modules. The knowledge of objects are represented as rules and stored in the long-term memory.

Strat and Fischler (1991) argue that it is possible to describe natural objects using precise geometric models and devised a context-based vision system - Condor for interpretation of natural scenes. Condor interprets objects using relationships between objects and context, instead of using specific object models. The central component of the system is a special-purpose knowledge base, in which the knowledge is represented as context sets

(rules). Each context set (rule) contains a set of conditions and an action. The action is fired only when all the conditions in the context set are satisfied. The interpretation is performed in a top-down process and contains four major steps: candidate generation (hypothesis generation), candidate comparison (hypothesis evaluation), clique formation and clique selection. In candidate generation, all candidate generation context sets are evaluated, and their associated operators produce candidates for each class. They are evaluated using candidate comparison context sets and a preference ordering is set up among them. In clique formation, candidates which are mutually coherent are grouped together. Finally, one clique is chosen as the best interpretation of the image, based on the portion of the image that is explained and the reliability of the operators that contributed to the clique.

Nagao and Matsuyama (1980) presented a blackboard system for the interpretation of suburban scenes from multispectral aerial images. The system consists of a number of object-detection subsystems. Each subsystem concentrates on detection of a specific object and contains the knowledge related to that object. The knowledge is represented as production rules, and each specified method for locating a specific object is expressed as one rule. There is no direct link between two different subsystems. They communicate via a blackboard, i.e. by writing information to the blackboard or reading information from the blackboard. The blackboard contains the information derived by low-level processing and properties of detected objects. The system itself does not set up a schedule for activation of subsystems. Which subsystem is invoked depends on the current information stored in the blackboard. When the conditions for activating a specific sub-system are satisfied, the subsystem will be fired and a goal-oriented analysis will be performed to locate a specific object. Once the subsystem locates the object, its properties are written onto the blackboard and can be shared with other subsystems. This enables other subsystems to find "ambiguous" objects which were not recognized because of poor reliability. The updated information in the blackboard will fire other subsystems and another goal-oriented object locating cycle commences.

A knowledge-based system for interpretation of suburban scenes from aerial images is described by Nicolin and Gabler (1987). Their system is composed of four main modules: method-base, long-term memory, short-term memory and control module. The method-base module comprises a set of domain-independent processing methods, such as image preprocessing, segmentation and structural analysis. The long-term memory stores the generic knowledge about suburban scenes, which is represented by semantic networks and meta-knowledge used to control knowledge during the course of interpretation, while the short-term memory contains the results obtained in the image analysis process. The control module controls the whole process of image interpretation and decides which process is executed in a certain situation. The system uses a hybrid control and starts with building initial hypotheses of objects in a bottom-up process. The top-down process is performed to generate hypotheses about possible instances of the objects and to find instances of the objects in the image, based on the initial interpretation of the objects and object models stored in the long-term memory. It is believed that reliable interpretation of objects may not be obtained through a single pass of bottom-up and top-down. Therefore, an iterative process is applied in this system in order to achieve more reliable results.

McKeown et al (1985) developed a rule-based system - SPAM for interpretation of airports from aerial images. The system is composed of three major components: a map database, image processing tools and a knowledge base. The map database collects the facts about the existence and locations of man-made or natural objects. It provides facilities to compute geometric properties and relationships of objects which serve as constraints during hypothesis verification. Image processing tools include a region-growing program, a road/road-like feature follower, a stereo analysis program and an interactive segmentation system. These tools perform low- and intermediate-level tasks and generate primitives of objects for high-level processing. The knowledge about the spectral and geometric properties of objects, spatial and semantic relationships of objects, etc., is represented as production rules and loosely organized into six classes in the knowledge base. The interpretation consists of build, local evaluation, consistency check, and a functional area model evaluation processes. It starts with the interpretation

of individual objects based on their geometric and spectral properties derived in low- and intermediate-levels of processing. The hypothesised objects are then combined to form functional areas such as runways, taxiways and tarmac if they are compatible with each other, after a consistency check is performed. Finally, a model evaluation is done to resolve the inconsistency among the generated functional areas.

A knowledge-based system - SIGMA for interpretation of houses and roads in urban areas is described by Matsuyama and Hwang (1985) and by Hwang et al (1986). The system consists of three major parts: geometric reasoning expert, model selection expert and low level vision expert. The geometric reasoning expert is the central reasoning module of the system which uses the accumulated evidence obtained during interpretation to produce new hypotheses of objects. The model selection expert reasons about the most promising models of object appearances for searching for the objects in the image, based on the contextual information provided by the geometric reasoning expert. Once the appearance model of the object is selected, the low-level vision expert performs image processing to extract image features corresponding to the specified appearance model. The knowledge of objects is represented by frames. Each frame includes the attributes of the objects, constraints among object attributes, and relations between the objects. The system integrates bottom-up and top-down processes during interpretation. It starts with hypothesis generation of objects with simple appearances. A goal-oriented low-level image processing is then performed to extract object features. The features which match the selected appearance model are returned to the model selection expert, which instantiates the corresponding object instances. The geometric reasoning expert uses the generated object instances and evidences to infer the missing parts of the objects and generate hypotheses of new objects based on the knowledge of object relations.

Existing maps and databases in geographical information systems (GIS) exhibit abundant information about terrain and various types of objects on the terrain. They offer information on terrain types, locations, shapes and sizes of objects and their relationships. Such information is very useful in image understanding, as it can be used to

predict the initial positions and appearances of objects to be interpreted, and improve the reliability of interpretation. Stilla (1995) presented a method for interpretation of aerial images based on structural analysis of existing digital maps. In his method, digital maps in GIS are analysed to derive contextual information of objects. Using a production net, objects are represented as image description graphs in which they are decomposed into primitives. The objects and associated image description graphs are stored as rules in a blackboard. The interpretation of the image starts with image processing to detect image primitives such as straight lines. The extracted image primitives are assessed and assigned an expectation based on their deviation from the attributes produced from maps. The control unit selects the primitives with best assessment. This process stops when the objects to be examined have been found in images with sufficient instances.

A similar method is described by Koch et al (1997). In their method, two types of knowledge, i.e. fact knowledge and procedural knowledge are defined. Fact knowledge describes the generic models of objects, their attributes and relationships. It is represented by semantic networks in which nodes represent objects or object features. Nodes are implemented by frames and related by a set of defined relations such as instance-of, part-of, con-of, etc. Procedural knowledge defines the order of the image interpretation process and is represented as rules. The image is interpreted in a top-down procedure. Firstly, objects and their relationships in maps are generated and stored in the knowledge base. With the generated object types, their positions and relationships from maps, and hypotheses of objects in the image are generated. They are verified in the image by selecting specific algorithms.

Gunst (1996) developed a knowledge-based system for updating road networks. The system uses a multi-scale semantic road model in which road networks are represented in three different scales, and are related by spatial relations and specialisation-generalisation relationships. The knowledge of road properties and the internal relationships of road parts is derived from road construction manuals and is represented as frames. The system works in a top-down scheme. The information of the existing road networks is generated from GIS and used to guide the search for roads in the image. When any change is

detected in the image, specific algorithms incorporated in *frames* are executed to locate the road. Spatial relationships between road parts are used to generate hypotheses of corresponding road parts and specially designed algorithms are selected to find their instances in the image.

3.5 Summary of Chapter

In this chapter, the elements of an image understanding system, some techniques for knowledge representation and control strategies were introduced. Image understanding is a complex process which generally consists of low-, intermediate- and high-level processing. Each processing step performs specific tasks and is related to other processing steps. Knowledge representation is of primary importance in artificial intelligence and image understanding systems. Knowledge can be classified as heuristic, procedural, declarative, meta- and structural. It is represented in a knowledge base in a certain form. Production rules, semantic networks, frames and logic are common approaches for knowledge representation in image understanding systems. Each of them has advantages and disadvantages. Which method is selected depends on the problem to be solved. Control strategy is also an important issue in image understanding. It has four different forms, i.e. bottom-up, top-down, hybrid and heterarchical. Each control strategy has its own characteristics. Previous research shows that reliable results may not be achieved by pure bottom-up or top-down control, especially when the image to be interpreted contains complex structures. What control strategy is used is determined by the problem to be solved and the knowledge representation scheme selected.

Chapter Four

Feature Extraction

Feature extraction is an important step in image understanding. It aims at detecting features of objects such as points, edges, lines and regions from an image without using specific knowledge about the structures of the objects. These features are aggregated in the subsequent processing to form the symbolic representations of the objects which are used in high-level processing to generate interpretation of the image. As no specific knowledge about object structures are used in feature extraction, this process is often called *low-level processing*.

Feature extraction has been an active research area in computer vision over the last three decades. A large number of methods and algorithms for feature extraction have been developed, which are generally based on two basic properties of an image: discontinuity and similarity. In the first category, the approaches are to partition an image based on abrupt changes in grey value to detect isolated points, edges and lines in the image. The approaches in the second category are based on thresholding, region growing, region splitting and merging to generate distinct regions. This chapter will concentrate on the first category, i.e. detection of point, edge and line features. Some preprocessing involved in feature extraction such as noise removal, edge tracking, edge linking, etc. will also be discussed.

4.1 Edge, Line and Point Detection

Edges, lines and points are primitive features of objects. They play a very important role in generating object structures for interpretation of images. Numerous methods and algorithms for feature extraction from images have been developed. Surveys of early development of feature extraction can be found in Davies (1975), Peli and Malah (1982),

Ballard and Brown (1982), Rosenfeld and Kak (1982), and Venkatesh and Kitchen (1992). Recent developments are investigated by Sonka et al (1993), Pal and Pal (1993), and Lemmens (1996). This section will introduce some of the existing methods and describe the method which will be used in this study.

4.1.1 Edge Detection

4.1.1.1 Definition of Edge

an edge is a local grey value discontinuity in the image

An edge in an image corresponds to a boundary or part of a boundary of an object, which separates one part from the other. There is a distinct difference in reflectance on both sides of a boundary in object space. In the ideal case, it is a step function (Figure 4.1(a)). An edge in the image is the result of convolution of a boundary in object space with the Point Spread Function (PSF) of the imaging system (Figure 4.1(b) and (c)). Therefore, the intensity profile of an edge in the image is a smooth curve rather than a step function. Due to the effects of image noise and texture of object surfaces, the real distribution of an edge profile in the image may deviate from the curve shown in Figure 4.1(c). As an edge is defined locally, it is sometimes called a "local edge" (Ballard and Brown, 1982).

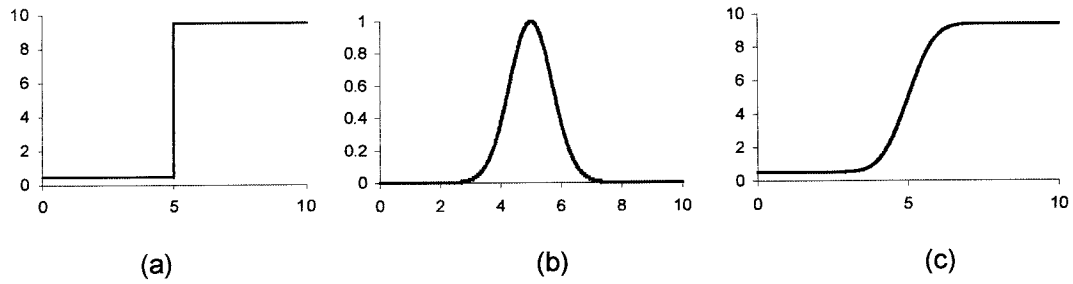


Figure 4.1 Definition of an Edge (a) Profile of a boundary in object space (b) Profile of point spread function in 1-D (c) Profile of an edge evolved from (a) and (b)

4.1.1.2 Edge Operators

Based on their definition, formulation of algorithms to detect edges is straightforward. An edge in an image can be detected by computing the intensity gradients of the image. Early edge operators such as the Roberts operator (Roberts, 1965), the Prewitt operator (Prewitt, 1970), the Sobel operator, the Laplacian operator, etc., work in this way. The Roberts operator is the earliest edge operator which computes the gradient of a pixel in a 2×2 window in two diagonal directions (Figure 4.2). The magnitude of an edge at pixel (i, j) is computed as:

$$\text{gradient} = |\Delta_1| + |\Delta_2| = |g(i, j) - g(i+1, j+1)| + |g(i, j+1) - g(i+1, j)| \quad (4.1)$$

$$\Phi = \tan^{-1}(\Delta_1/\Delta_2) \quad (4.2)$$

g is the grey value of a pixel and i, j are its image coordinates. Δ_1, Δ_2 are the difference of grey values in two diagonal directions at the pixel (i, j) and Φ is the gradient direction. As only four neighbouring pixels are used in the computation of gradients, the advantage of the Roberts operator lies on its simplicity in computation. However, it is very sensitive to image noise and scene details (texture).

1	0
0	-1

Δ_1

0	1
-1	0

Δ_2

Figure 4.2 Roberts Operator

To suppress the effects of image noise, the Prewitt operator uses local averaging of neighbouring pixels (Figure 4.3), which is actually a process of low-pass filtering. The magnitude and direction of the gradient at point (i, j) are computed as:

$$\Delta_1 = 1/3*(g(i-1, j-1) + g(i-1, j) + g(i-1, j+1) - g(i+1, j-1) - g(i+1, j) - g(i+1, j+1)) \quad (4.3)$$

$$\Delta_2 = 1/3*(g(i-1, j+1) + g(i, j+1) + g(i+1, j+1) - g(i-1, j-1) - g(i, j-1) - g(i+1, j-1)) \quad (4.4)$$

$$\text{gradient} = (\Delta_1^2 + \Delta_2^2)^{1/2} \quad (4.5)$$

$$\Phi = \tan^{-1}(\Delta_1/\Delta_2) \quad (4.6)$$

1	1	1
0	0	0
-1	-1	-1

-1	0	1
-1	0	1
-1	0	1

Figure 4.3 Prewitt Operator

There is a generalisation of the Prewitt operator which employs 8 masks representing different edge directions as shown in Figure 4.4 (Rosenfeld and Kak, 1982). The magnitude of an edge is determined by the convolution of the masks with the image. The direction of the magnitude is given by the mask showing maximum response.

1	1	1
0	0	0
-1	-1	-1

1	1	0
1	0	-1
0	-1	-1

-1	0	1
-1	0	1
-1	0	1

0	-1	-1
1	0	-1
1	1	0

-1	-1	-1
0	0	0
1	1	1

-1	-1	0
-1	0	1
0	1	1

1	0	-1
1	0	-1
1	0	-1

0	1	1
-1	0	1
-1	-1	0

Figure 4.4 Generalization of the Prewitt Operator

Instead of using the same weights for all neighbouring pixels in the Prewitt operator, the Sobel operator employs different weights, yielding better smoothing results. Larger weights are assigned to pixels lying closer to the central pixel (Figure 4.5).

1	2	1
0	0	0
-1	-2	-1

-1	0	1
-2	0	2
-1	0	1

Figure 4.5 Sobel Operator

The Laplacian is a gradient operator which uses second derivatives of the image to represent the gradient magnitude, i.e.

$$\nabla^2 = \nabla_x^2 + \nabla_y^2 \tag{4.7}$$

∇_x^2, ∇_y^2 are the second derivatives of the image in the direction x and y respectively. It has the discrete form:

$$\nabla^2 = g(i, j) - 1/4*(g(i, j+1) + g(i, j-1) + g(i+1, j) + g(i-1, j)) \tag{4.8}$$

The disadvantages of the Laplacian are that the image noise is enhanced doubly and there is no directional information available (Ballard and Brown, 1982).

Kirsch (1971) presented an operator for edge detection which uses a number of templates with different sizes and orientations, four of which are shown in Figure 4.6. The magnitudes and orientations of the gradients are computed through the convolution of the template and a local area of the image. The operator gives the magnitude and orientation associated with the maximum gradient magnitude. Some other template operators can be found in Robinson (1977), Nevatia and Babu (1980) and Grün and Agouris (1994).

An edge in an image can also be detected by fitting a designed function such as a step function to a local area in the image. The magnitude and orientation of an edge are then determined by the parameters included in the function which can be calculated through a

number of condition equations. The Hueckel operator (Hueckel, 1971), the Hummel operator (Hummel, 1979) and the operator presented by Ghosal and Mehrotra (1993) fall into this category.

-1	-1	0	1	1
-1	-1	0	1	1
-1	-1	0	1	1
-1	-1	0	1	1
-1	-1	0	1	1

1	1	0	-1	-1
1	1	0	-1	-1
1	1	0	-1	-1
1	1	0	-1	-1
1	1	0	-1	-1

0	1	1	1	1
-1	0	1	1	1
-1	-1	0	1	1
-1	-1	-1	0	1
-1	-1	-1	-1	0

1	1	1	1	0
1	1	1	0	-1
1	1	0	-1	-1
1	0	-1	-1	-1
0	-1	-1	-1	-1

Figure 4.6 Kirsch Templates

Abdou and Pratt (1979) studied the performance of some gradient edge operators based on their localisation errors and ability to detect true edge points. The study shows that the Prewitt operator and Sobel operator have similar performance, and both are superior to the Roberts operator. The performance of the Hueckel operator ranks between the Prewitt operator and the Roberts operator, but is closer to the Prewitt operator. Similar results were carried out by Venkatesh and Kitchen (1992) based on four criteria: false negatives-failure in detecting a true edge, false positives detecting non-edge points, multiple detection and localisation error. All of these studies show that the performance of the gradient edge operators decrease with the increase of image noise. This is not surprising because all gradient edge operators work in small areas in the image.

Marr and Hildreth (1980) presented an edge detection method using the zero-crossings of the second derivatives of intensity values across an edge based on the assumption

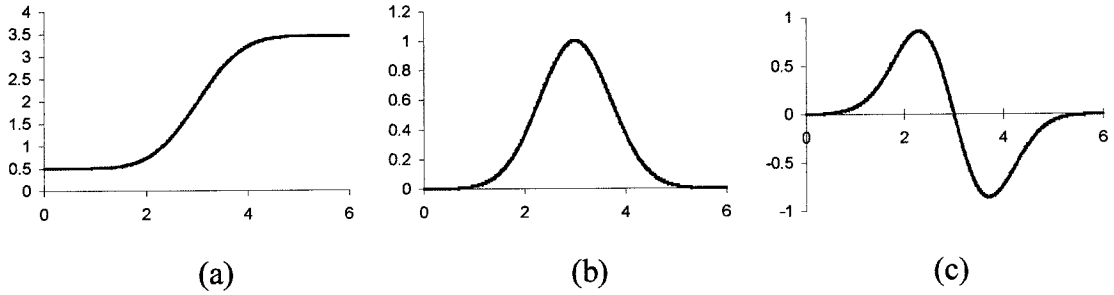


Figure 4.7 Second-zero Crossing of an Edge Profile (a) Edge profile (b) First derivative of edge profile (c) Second derivative of edge profile

that the first derivative of the image should have its maximum at the position where an edge occurs and in the direction perpendicular to the edge, and the second derivative should be zero at the same position and direction (Figure 4.7). To alleviate the effects of the image noise, the image is smoothed with a Gaussian filter which has the form :

$$G(x, y) = e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{4.9}$$

σ is the standard deviation of the Gaussian function. The second derivatives of the smoothed image can be computed using the Laplacian operator, i.e.

$$\nabla^2(G(x, y, \sigma) \otimes g(x, y)) \tag{4.10}$$

$\otimes$ stands for convolution. In the $\nabla^2 G$ image, points with value of zero correspond to edges. As the image is smoothed using a Gaussian filter, the influence of all pixels within the distance of 3σ is suppressed (Sonka et al, 1993). Therefore, the Marr-Hildreth operator yields smoother results. This is its advantage compared with other local edge operators. Another advantage of the Marr-Hildreth operator is that it does not need to define a threshold to detect edge points. However, the operator suffers from some deficiencies (Lemmens, 1996):

- Theoretically, edges defined by the zero-crossings should be continuous, but breaks occur due to the effects of image noise, texture and quantisation.
- T junctions or trihedral vertices are incorrectly detected.
- Gaussian smoothing causes merging of closely spaced edges, yielding phantom edges.
- A good method for combining the results at different scales is lacking.
- The operator is sensitive to image noise since it uses second derivatives for detection.

Canny (1986) formulated edge detection as an optimization problem based on three performance criteria:

- Good detection. True edge points should be detected and non-edge points should be avoided.
- Good localization. The detected edge points should be located as closely as possible to the centre of the true edge.
- Only one response to a single edge.

Using the above criteria, an optimal operator for a step edge in 1-D is derived, which can be approximated by the first derivative of a Gaussian function. In two dimensions, a 2-D Gussian function G given in Equation 4.9 is used. Suppose the first derivative of the Gaussian function in the direction normal to an edge is G_n, i.e.

$$G_n = \frac{\partial G}{\partial n} = \mathbf{n} \bullet \nabla G \tag{4.11}$$

The normal direction of the edge can be estimated from the smoothed gradient direction,

$$\mathbf{n} = \frac{\nabla(G \otimes I)}{|\nabla(G \otimes I)|} \tag{4.12}$$

I stands for the image. An edge point is defined to be a local maximum of the smoothed image. It is determined by the following equation:

$$\frac{\partial}{\partial n}(G_n \otimes I) = 0 \tag{4.13}$$

By substitution of Equation 4.11 into the above equation, it can be rewritten as:

$$\frac{\partial^2}{\partial n^2}(G \otimes I) = 0 \tag{4.14}$$

The magnitude of an edge point can be computed by:

$$|G_n \otimes I| = |\nabla(G \otimes I)| \tag{4.15}$$

Thus, edges are located by finding zeros of second derivative in their normal directions. To avoid streaking, a scheme with two thresholds is used, in which a high threshold is used to eliminate the edges caused by the image noise and texture while a low threshold is utilized to detect weak edge points.

The Canny operator in one dimension is similar to the Marr-Hildreth operator. In two dimensions, however, the directional properties of the Canny operator enhance its detection and localisation performance. Another important characteristic of the Canny operator is that it can provide a good estimate of the edge strength. The Canny operator has been extended for different types of edges, e.g. extraction of ramp edges (Petrou and Kittler, 1991). Some other optimal operators have been developed for extraction of edges with different shapes (Shen, 1992; van der Heijden, 1995).

Förstner (1993, 1994) presented a scheme for feature extraction. In his method, an image area is described by three types of features, i.e. points, edges and regions which

are assumed to be piecewise smooth. To distinguish these features, a homogeneity or smoothness measure is generated based on the average squared gradients of the image, i.e.

$$G_\sigma(x, y) \otimes \Gamma g(x, y) \tag{4.16}$$

with

$$\Gamma g(x, y) = \nabla g(x, y)\, \nabla(x, y)^T = \begin{pmatrix} g_x^2(x,y) & g_x(x,y) \bullet g_y(x,y) \\ g_x(x,y) \bullet g_y(x,y) & g_y^2(x,y) \end{pmatrix} \tag{4.17}$$

Regions are distinguished from points and edges using the trace of the average squared gradients. If the trace is less than a predefined threshold $T_h(\sigma)$, the image area is a region, otherwise it is an edge or point. Another quantity, the curvature, is used to separate points and edges, which is:

$$K_\sigma^2(g) = \frac{1}{\sigma^2} \frac{\lambda_2}{\lambda_1} \tag{4.18}$$

λ_1 and λ_2 are two eigenvalues of the homogeneity measure. If it is smaller than a given threshold $T_k(\sigma)$, the detected feature is an edge, otherwise a point. Both $T_h(\sigma)$ and $T_k(\sigma)$ can be estimated from the image noise, the width of the Gaussian kernel G_σ and a predefined significance level.

Heitger (1995) developed a novel approach, the SE operator, for edge and line detection. The operator is based on a complex combination of the first and second derivatives of the response modulus and the even and odd filter responses in different orientations. It detects edge and line features by suppressing filter responses which are unlikely to characterize the type and position of an image feature and enhancing the responses that correspond to well-defined image features. The SE operator has the following advantages (Heitger, 1995):

- Providing better continuity at locations of edge/line transitions,
- Dealing with periodic signal variations and feature interference, and
- Being not limited to a particular class of filter kernels (e.g. Gaussian derivatives)

The comparison study shows that the SE operator can produce better representation of line features than the Canny (Heitger, 1995). But more computational time is needed by the SE operator as it contains complex computations in different orientations.

4.1.1.3 Edge Operator in This Study

It can be seen that numerous operators and algorithms have been developed for edge detection. Simple gradient operators can easily be implemented, but they are not suitable to edge extraction of remotely sensed images because of their serious shortcoming - their sensitivity to image noise and texture. Optimal operators developed by mathematical operations are usually based on some assumptions about an edge. They may work very well for images in which edges are similar to those defined by the operators, but unsatisfactory results may be produced for other images. There are no optimal operators which can work well in all situations. From this point of view, the selection of an operator for edge detection is task dependent. The purpose of this study is not to design an edge operator for automatic road extraction. Therefore, an existing operator will be chosen according to their performances. As described in Chapter 2, automatic road extraction has two tasks, i.e. automatic location and recognition of roads from images. For the purpose of location, a high positioning accuracy of road boundaries is required while all useful features in the image should be used for recognition. This means that true image features should be extracted and non-feature information caused by the image noise and object texture should be avoided. Thus, the criteria for selecting an edge operator can be summarized as:

- Good localisation of road boundaries,
- Good detection of road boundaries.

The above study has show that both the Canny and SE operator can provide accurate localisation and good detection of road boundaries, and they are insensitive to image noise. Therefore, they are used for *low-level* feature extraction in this study.

4.1.2 Line Detection

A line in an image corresponds to an abrupt change in intensity against its background. Figure 4.8 shows some examples of line profiles that are commonly encountered in a digital image.

A line in the image can be detected by computing the intensity gradients, usually by the convolution of the image with designed masks (Rosenfeld and Kak, 1982; Fischler et al, 1981). One possibility is to use 14 different masks, each of which has a size of 5×5 and corresponds to a defined line orientation (Sonka et al, 1993). Four of them are listed in Figure 4.9 and the others can be obtained by rotation. The magnitude of the gradient is determined by the convolution of the image with the masks and the direction is the orientation of the mask giving the maximum gradient. One serious shortcoming of this type of line operator is that it may detect other patterns which are not line-like, e.g. an edge. To avoid this problem, some conditions should be used. For example, the following conditions should be satisfied for a vertical dark line:

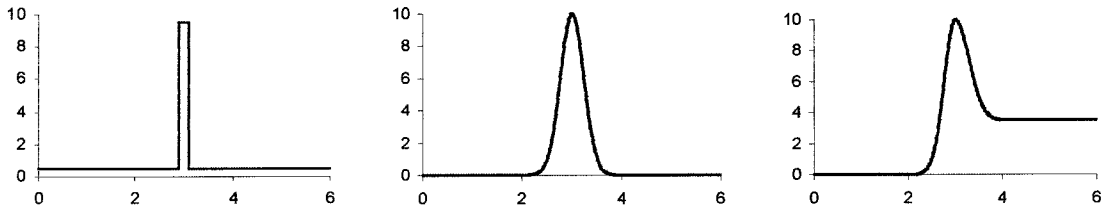


Figure 4.8 Common Line Profiles in Image

0	0	0	0	0
0	-1	2	-1	0
0	-1	2	-1	0
0	-1	2	-1	0
0	0	0	0	0

h_1

0	0	0	0	0
0	-1	-1	-1	0
0	2	2	2	0
0	-1	-1	-1	0
0	0	0	0	0

h_2

0	0	0	0	0
0	0	-1	-1	-1
0	2	2	2	0
-1	-1	-1	0	0
0	0	0	0	0

h_3

0	0	0	0	0
-1	2	-1	0	0
0	-1	2	-1	0
0	0	-1	2	-1
0	0	0	0	0

h_4

Figure 4.9 Masks for Line Detection

$$g(i-1, j-1) < g(i-1, j), \quad g(i-1, j) > g(i-1, j+1)$$
$$g(i, j-1) < g(i, j), \quad g(i, j) > g(i, j+1)$$
$$g(i+1, j-1) < g(i+1, j), \quad g(i+1, j) > g(i+1, j+1)$$

As these types of operators work in small areas, they are sensitive to image noise. Also, when the width of a line is different from that defined by the masks, the line will not be detected.

Mathematical morphology provides a flexible mechanism for image processing. It uses four basic operations, i.e. dilation, erosion, opening and closing, for computing intensity gradients of an image. One typical edge operator, referred to as *the morphological gradient*, uses the difference of the dilated image and eroded image for computation of image gradients (Dougherty, 1992). It is defined as:

$$GRAD(g) = (g \oplus s) - (g \varphi s) \tag{4.19}$$

$\oplus$ and φ stand for dilation and erosion operations respectively, and s is a flat structuring element which is constant on its domain.

A dark line in an image can be detected by the *top-hat transform* which has the following form:

$$HAT(g) = g - (g \circ s) \tag{4.20}$$

o stands for the opening operation, and s is a flat structuring element. As opening is an antiextensive operation, the image after the opening operation lies below the original image. So HAT(g) is always non-negative. This is illustrated in Figure 4.10.

A bright line in an image can be detected in a similar way. The corresponding operator is defined by the operation $(g \bullet s) - g$. $\bullet$ means closing operation. It is equivalent to:

$$-g - [(-g) \circ s] = HAT(-g) \tag{4.21}$$

It is also called a *top-hat transform.*

To detect both dark and bright lines, the above operators can be applied separately, or both the opening and closing are performed simultaneously and then subtracted to obtain $(g \bullet s) - (g \circ s)$.

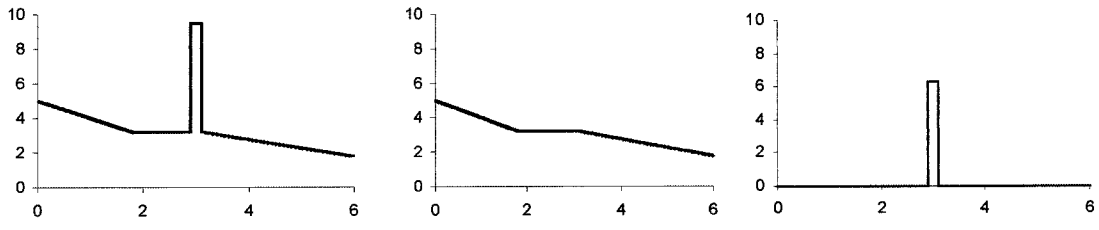


Figure 4.10 Top-hat Transform (a) Image section containing a line (b) Image section after opening operation (c) Image profile after top-hat transform

A line profile in the direction perpendicular to the line direction can be approximated by a second or third order Taylor polynomial. Line points are then detected by selecting points with high directional second derivatives, i.e. high curvature. There are two ways to determine the coefficients of the polynomial. One uses a least squares method to fit the polynomial to the image data (Haralick et al, 1983; Busch, 1994). The advantage of this approach is that lines can be detected with sub-pixel precision, but it usually leads to multiple responses to a single line (Steger, 1996). The other method determines the coefficients of the second order polynomial by convolving the first derivative of a Gaussian kernel with the image (Steger, 1996). One obvious advantage of this approach is that it can be scaled to lines with arbitrary width and detects lines with sub-pixel precision. It can also avoid the problem of multiple responses. There are some other techniques which can be used for detection of lines, e.g. wavelet transforms (Grün and Li, 1994).

As has been shown, morphological operators can easily handle different widths of line by changing their structural elements. This is very advantageous to road extraction as a road in low-resolution images may change. At the same time, morphological operators can easily be implemented. Thus, a morphological operator will be used in this study,.

4.1.3 Point Detection

A point feature in an image has a distinct difference in intensity against its background. It can be an isolated image point, corner point, or intersection of two edges/lines. Like detection of edges and lines, a point feature in the image can be located using a local image gradient. The simplest corner detector is the Moravec operator (Moravec, 1977) which is based on the image gradients in a local area. The Moravec operator computes the interest value (IV) for every pixel in the image by:

$$IV = \min\{v_1, v_2, v_3, v_4\}$$

$$v_1 = \sum (g(i, j) - g(i+1,j))^2$$

$$v_2 = \Sigma (g(i, j) - g(i,j+1))^2 \tag{4.22}$$
$$v_3 = \Sigma (g(i, j) - g(i+1,j+1))^2$$
$$v_4 = \Sigma (g(i, j) - g(i+1,j-1))^2$$

with $i = m - k, \ldots, m + k - 1$, $j = n - k, \ldots, n + k - 1$, $k = w/2 - 1$.

v_1, v_2 are gradients in vertical and horizontal directions, and v_3 and v_4 are gradients in two diagonal directions. w is the width of the window. When a point has an IV larger than the given threshold and IVs of its neighbours in the window, it is selected as the corner point. By definition, it can be found that the IV value of an intersection would be zero if one of two intersecting edges or lines lies on one of the four directions. Therefore, intersections of edges or lines cannot be detected properly by the Moravec operator.

The Förstner operator (Förstner and Gülch, 1987) is a well-known operator for point detection in the photogrammetric community. The operator was developed based on five criteria: distinctness, invariance, stability, seldomness and interpretability. It detects interest points in a two-step procedure: selection of the optimal windows in which point features are likely to appear and location of the point in the selected window. Optimal windows are chosen using two parameters w and q:

$$w = \frac{trN}{\det N} \tag{4.23}$$

$$q = \frac{(trN)^2}{4 \det N} \tag{4.24}$$

$$N = \begin{bmatrix} \Sigma g_u^2 & \Sigma g_u g_v \\ \Sigma g_u g_v & \Sigma g_v^2 \end{bmatrix} \tag{4.25}$$

where g_u and g_v are the intensity gradients of points in two diagonal directions, and N is the covariance matrix of gradients in a window with certain size such as 5×5, which is similar to Equation 4.17. trN and detN are the trace and determinant of the matrix N respectively. q is the roundness of the error ellipse of the estimated point and w is its weight.

The optimal windows are determined using the following conditions:

$$q > T_q = 0.5 - 0.75 \text{ and}$$
$$w > T_w = f \cdot w_{mean},\ f = 0.5 - 1.5.$$

W_{mean} is the average weight of all selected windows. When an optimal window is selected, the position of the point feature is determined based on all edge points in the window by a least squares method. One advantage of the Förstner operator is that it can detect points where more than two edges intersect. However, the computation cost of the Förstner operator is very high as it needs to compute weight matrix for every pixel.

Based on an investigation of the Moravec and Förstner operators, Lü (1988) presented a new algorithm for point detection which introduces a two-step procedure. It first computes the gradients of pixels in both the horizontal and vertical directions. Points with gradient magnitude above the given threshold are selected as point candidates. As the gradient is computed in only two directions, the computation time will be greatly reduced compared with the Moravec and Förstner operators, but edge points may also be detected in this step. To determine the interest points, the interest values, which can be the summation of the absolute gradient magnitudes in their 8-neighbours or q and w in Equations 4.23 and 4.24, are computed. Finally, points with interest values within the defined ranges are selected.

Heitger et al (1992) presented a scheme for point detection based on the first and second derivatives of the oriented energy filters. The first derivatives in the direction of the modulus reflect the oriented structures such as corners and junctions, while the second

derivatives correspond to blobs and large curvature. The operator generates a scalar feature called a "key-point map" in which points with significant variations in intensity in two directions are represented. In this study, this operator is used to detect intersections of roads in low-resolution images. As road intersections are distinct points in low-resolution images, a high threshold is employed for point selection. Henricsson (1996) compared the results of point extraction by using the SE operator and Förstner operator and found that they are quite similar. In this study, the SE operator will be used for point extraction from low-resolution images.

4.2 Noise Estimation for Feature Extraction

After a feature extraction operator is applied to an image, an image comprising the intensity gradients is generated, which is called a *feature image* or *feature map*. Due to the effects of image noise and texture of object surfaces, the generated feature maps include two components: true image features and non-feature points, i.e. noise. To facilitate subsequent processing, noise must be removed from the generated feature image. This is usually done by a thresholding operation in which a threshold is used to separate the image noise from the true image features. The threshold may be determined manually through the visual check of the feature maps. However, this method suffers from serious problems. Firstly, the selection of the threshold is subjective. Secondly, it is inefficient as the selection of an "optimal" threshold needs a large amount of testing and comparison. Voorhees and Poggio (1987) and Busch (1996) have shown that the threshold can automatically be derived by the analysis of the distribution of gradient magnitudes of the image. In Voorhees and Poggio, the threshold for edge detection is automatically determined based on the estimation of image noise. They assume that the image contains white Gaussian noise, and the gradient magnitudes of noise are Rayleigh distributed. Noise contributes to the low gradient magnitudes while the true edge points only affect the tail of the distribution as their gradient magnitudes are usually much larger than those of the noise. The threshold T is determined by:

$$T = \xi \sqrt{-2 \ln s} \tag{4.26}$$

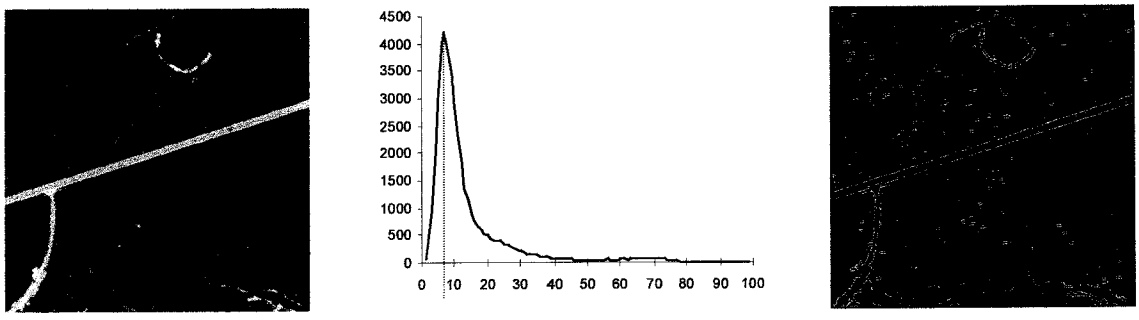
where ξ is the peak value of Rayleigh distribution and s is the risk probability.

Busch (1996) developed an approach for automatic determination of the threshold for edge and line detection. It is assumed that true features have a certain length and isolated feature points correspond to image noise. Therefore, the method starts with the selection of isolated points in the edge or line maps and generation of the distribution of their gradient magnitudes. Given a significance level τ, the corresponding threshold T can be calculated as:

$$T: \int_{-\infty}^{T} p(c_s) d c_s = \tau \tag{4.27}$$

where $p(c_s)$ is the probability function of the distribution.

It should be pointed out that not only noisy edge information is removed in the thresholding operation, but also weak edge elements will be deleted in this process. Therefore, one should be very careful in selecting a threshold. For images with simple object structures and high contrast, a high threshold may yield good results. For remotely sensed imagery, information of objects derived is usually incomplete due to occlusions,



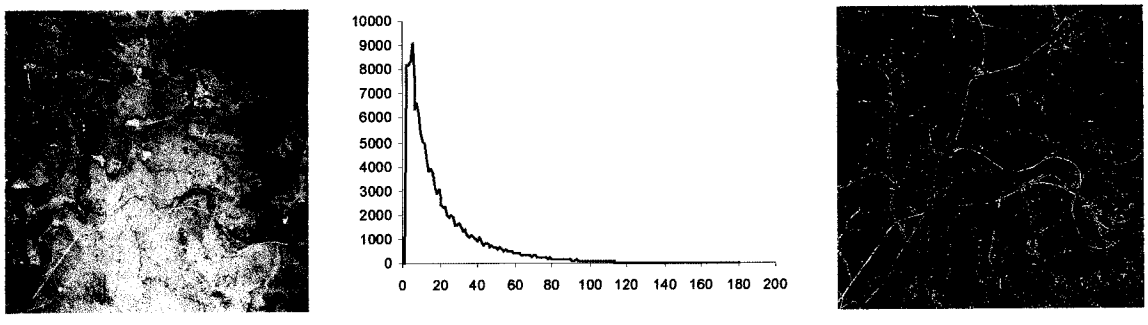
(a) (b) (c)

Figure 4.11 Noise Estimation in Edge Extraction (a) Original image (b) Derived distribution of gradient magnitude (c) Edge image after noise removal

complex structures, etc. If a high threshold is chosen, edges with poor contrast will be lost. This will make the subsequent processing very difficult, sometimes producing unreliable results. Therefore, for object recognition from remotely sensed images a conservative threshold is usually selected. In this study, a method similar to Voorhees and Poggio's is used to determine the threshold for edge detection, and in order to preserve weak parts of road boundaries, a risk probability of 1% is chosen, which yields the threshold equal to $3.0 \cdot \xi$.

Figure 4.11 shows an example of the thresholding operation in which Figure 4.11(a) is an aerial image with 500×500 pixels. The edges are detected by the Canny operator and the distribution of their gradient magnitudes is shown in Figure 4.11(b). The peak value of gradient magnitude corresponding to the of edge points is 6.9 in this example, and the corresponding threshold is 20.7. The result after thresholding is given in Figure 4.11(c). It can be seen that most noise components are removed from the edge image.

It has been found that the threshold determined by Voorhees and Poggio's method is too low when a morphological operator is used to extract line features from aerial images. To determine the threshold for line detection, the method described in Busch (1996) is used. Figure 4.12 gives an example of noise removal for line detection. The threshold with the significance level of 75 percent is 17 in this example. The result after noise removal is shown in Figure 4.12(c).



(a) (b) (c)

Figure 4.12 Noise Estimation for Line Extraction (a) Original image (b) Derived distribution of gradient magnitude (c) Line image after noise removal

4.3 Edge/Line Tracking and Linking

4.3.1 Edge/Line Tracking

In a thresholded feature image, most noise components are removed, and hence most of the remaining edge elements correspond to true edges. Edge elements are called *edgels* in computer vision. The unstructured *edgels* have no meaning until they are aggregated into edges which are often called *contours* in the computer vision community. There are a number of methods for edge tracking which can be classified into two categories - global and local. The global approaches include those formulating edge tracking as a graph search and dynamic programming. In a graph search, the directions of image gradients are taken as nodes and their magnitudes are treated as weights of the nodes. Arcs between nodes are generated based on the connectivity and requirements of both magnitudes and directions of the gradients of nodes to be connected. To find an optimal path between two nodes in the graph, the technique of heuristic search can be used (Nilsson, 1980). In heuristic search, a cost function describing the cost from the start node to the goal node is expressed as the sum of the costs from the start node to the current node and from the current node to the goal node. The optimal path is then traced by finding nodes which can make these two costs a minimum. The cost can be the function of the gradient magnitudes of edge points, curvature, etc.

Dynamic programming is a technique for finding an optimal solution to a problem which consists of a number of interrelated stages. Instead of testing all possible solutions at the same time, dynamic programming only considers the possible situations related to the current stage and tries to find the optimal solutions of the current stage, based on an evaluation function. The optimal solutions are found by minimising this evaluation function. In edge tracking, it can be defined as a function of magnitude and direction of gradients of edgels. Edgels which can optimize the evaluation function are taken as the optimal successors.

Edges in an edge image can also be tracked by using the Hough transform technique. The original Hough transform was devised to detect straight lines and curves which can be represented by an equation with a number of parameters, e.g. two parameters for a straight line. In parameter space, a straight line or curve corresponds to a point with certain coordinates. Therefore, edge tracking in the image becomes the task of detection of points with corresponding parameters in parameter space. To detect curves with arbitrary shapes, a generalized Hough transform (Ballard, 1981) and fuzzy Hough transform (Philip, 1991) were developed. However, they still need some prior information about shapes of regions or objects to be tracked.

Local approaches only consider the surrounding pixels of an edgel, usually its 8-neighbours. The algorithm presented in Nevatia and Babu (1980) falls into this category. They examined the 8-neighbours of each edgel to determine its predecessors and successors. Two conditions are used in their algorithm. The first is that the change of spatial direction of an edge should be less than 90°. This limits the search to three pixels in front of the edgel. The second condition is that the difference of gradient direction between the edgel and pixel to be tracked should be less than 30°. The problem with their algorithm is that it cannot track edges correctly when an intersection occurs.

A simple tracking algorithm based on gradient magnitude is described in Rosenfeld and Kak (1982). The algorithm uses two thresholds - a high threshold and a low threshold. Edgels with gradient magnitudes above the high threshold are selected as the starting points. The algorithm then checks the neighbouring edgels of these points at their next rows. If the gradient magnitudes of the neighbouring edgels are above the low threshold, they are labeled as successors. This process continues until no new points are found. Obviously this algorithm only uses connectivity of edgels without considering the gradient orientation information. The method presented in Nalwa and Pauchon (1987) is based on connectivity information of edgels. To derive the connectivity information among edgels, a finer grid and masks with different orientations are used. Edges are thinned before the connectivity relationship is generated in order to limit the search to 8-neighbours. In the thinned image, an edgel normally has two connected neighbours.

Therefore, an edge can be tracked by extending an edgel in both directions. Orientation information is only considered at junctions.

In contrast to using gradient magnitude, Burns et al (1986) utilises the gradient orientation as the criterion for extracting straight lines. Their algorithm first groups edgels with similar gradient orientations to form "line-support regions". A planar surface is then fitted to each line-support region and straight lines are extracted by intersecting the fitted plane with a horizontal plane representing the average intensity of the region weighted by local gradient magnitudes.

Most existing local approaches use information on magnitude and orientation of edgels to track edges. Good connectivity between edgels is usually required by these methods. One problem with local approaches is that they cannot track edges correctly when they pass through a corner or an intersection. To solve this problem, Henricsson (1996) presented a novel approach for generation of a contour graph. The method first selects the starting points for tracking, based on the homogeneity of gradient magnitude, orientation and types of points in a given neighbourhood. The tracking proceeds from the selected starting points in both directions. To determine the successor of a tracked point, 8 points in front of the tracked point in a 5×5 window are examined and a point with highest priority is selected. The reason for using a 5×5 window instead of 3×3 is that the tracked points will be more consistent and small gaps can be bridged. The tracking will stop when an end point or a key-point is found, or it meets a tracked edge or line.

A tracking algorithm based on the connectivity between points, gradient magnitude and orientation, and spatial direction of edges is proposed in this study. The proposed algorithm is a local operator which is similar to the algorithm presented in Nevatia and Babu (1980). Corners and junctions are not considered in this operator as road boundaries are smooth edges, and all edgels in the thresholded image are treated equally in order to avoid missing weak edges. Therefore, any edgel in the image can be the starting point for edge tracking. To find the successor of the starting point, its 8

neighbours are examined and points with similar gradient orientation and magnitude are selected. Tracking is then performed in two opposite directions (Figure 4.13 (a)). To determine the successor of a tracked point, two conditions are used. They are that the change of spatial direction of an edge should not be larger than 90°, and points belonging to the same edge should have similar gradient orientations and magnitudes. The first condition limits the search area within five points of which three are in front of the tracked point and two are in the normal directions of the edge. The reason that the search area is expanded to five points instead of three points used in Nevatia and Babu (1980) is that in real situations an edge may have an abrupt change in its direction because of the effects of image noise and disturbances from surrounding objects. Tracking may be lost in this case if only three points are considered. However, higher priority will be given to the three points in front of the tracked point during tracking to ensure that the tracked edges are smooth. Two points in the normal directions of the edge are only considered if no points are found in front. In the search area, the point with maximum gradient magnitude and similar orientation to the tracked point is chosen. When a point is tracked, it is flagged in order to avoid re-tracking of the point. When there is no point found in the search area or it reaches a tracked edge, tracking stops.

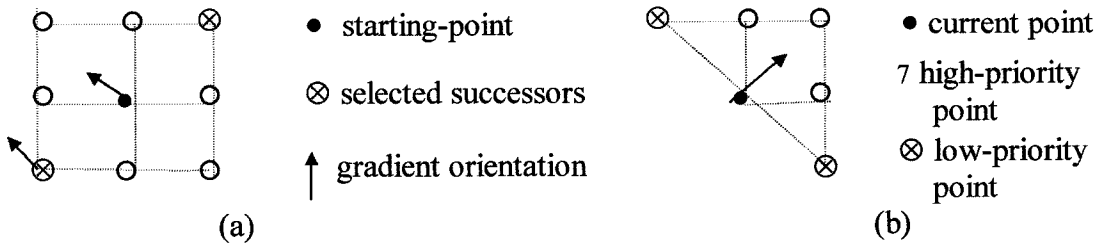


Figure 4.13 Edge Tracking

A tracked edge is represented by the image coordinates of its starting point, end point and two point chains. Each point has two pointers, one pointing to its successor and the other pointing to its predecessor. In this way, points on the edge are arranged sequentially in forward and backward directions. The advantage of using two point chains is that knowing any point on an edge, the positions of other points on the edge

can be tracked from them. This is very useful in determining generalized antiparallel pairs from the tracked edges (see Section 5.2).

4.3.2 Edge/Line Linking

Due to the effects of image noise, poor contrast, occlusions, limitation of operators, etc., extracted edges/lines are usually highly fragmented. The fragmented edge/line segments need to be linked to form elongated segments before they are used to generate structures of objects. It is generally assumed that two edge/line segments belong to the same edge/ line if they are close to each other and have similar spatial orientations. These two criteria: proximity and collinearity, are used in most existing linking algorithms (Broder and Rosenfeld, 1981; Zhu and Yeh, 1986; Vasudevan et al, 1988;). In most cases, edges/lines can be linked correctly using these two criteria. However, in areas with rich features, an edge/line segment may have more than one possible connection. In this case, use of these criteria alone may lead to wrong connection. To connect edges correctly, more information about edge segments to be linked should be used. In this study, an algorithm has been developed based on the gradient magnitude, orientation of edge segments together with the spatial constraints, i.e. proximity and collinearity. As the disturbance of image effects is usually limited to a few pixels, only small gaps will be considered. For example, only gaps less than two pixels are bridged in the Henricsson's linking algorithm. In this study, gaps less than four pixels will be bridged in order to generate smooth edge/line segments. It is known that the change in spatial direction within a distance of a few pixels along a smooth edge/line is usually small. Thus, it is mainly determined by the accuracy of calculation of spatial orientation and is set as 30° in this study, as the accuracy of spatial orientation is about 15°. Therefore, two edges will be linked if they meet the following conditions:

-- Distance between two edges is less than four pixels.

-- Difference between the spatial orientations of two edges and their connection is less than 30°

-- Two edges have similar gradient orientations and magnitudes.

The algorithm works as follows:

1. Select an edge as the first segment.
2. Determine a search area. For the first segment, a search area will be established around its starting or end points to find other starting or end points. To determine the search area a 5×5 window in the direction of the first segment at its starting or end point is selected. Using the criterion of collinearity, the search area will be limited to a small area of the window, e.g. 12 points for a 5×5 window (shaded area in Figure 4.14(b)).
3. Find candidates for connection. Edges with their starting or end points falling within the search area are selected. If their gradient orientations and magnitudes meet the requirements for connection, they are chosen as candidates for connection.
4. Determine connection for first segment. If there is only one candidate and the difference between its spatial direction and its connection to the first segment is not larger than 30°, it is connected with the first segment. If there is more than one candidate, the one with the smallest difference between its spatial direction and its connection to the first segment, provided it is not larger than 30°, is connected.
5. Go to step 1. Choose another edge and repeat steps 2 to 4.

Lines are connected in the same way as that used in edge connection.

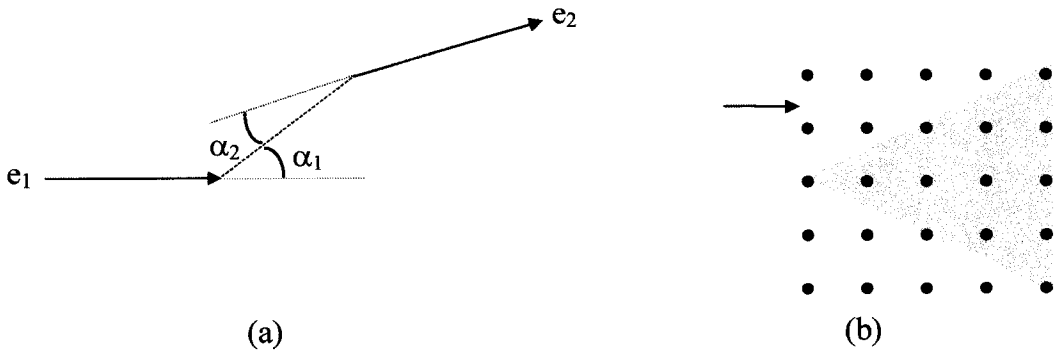


Figure 4.14 Edge/Line Linking

4.4 Edge/Line Splitting and Merging

After edge/line tracking and linking, numerous edge/line segments are generated. They can be classified into two groups: one is the collection of boundaries of objects to be recognised and related objects and the other correspond to boundaries of objects which are not related to the objects to be recognised, e.g. leaves of trees. Obviously, the former is the necessary information for the recognition process and should be kept while the latter will increase the complexity of subsequent processing, sometimes even make the results of recognition unreliable and therefore, should be removed as much as possible. It is well known that roads generally have continuous and smooth boundaries and natural objects usually have irregular shapes. Thus, a split-and-merge process can be designed to split extracted edge/line segments into short segments and to merge split segments if they have a smooth change in spatial direction. Most edge/line segments of natural objects can be removed by a thresholding operation as they are split into short segments after the split-and-merge operation.

There are various algorithms for operation of edge/line splitting (Ramer, 1972; Grimson, 1990; Medioni and Yasumoto, 1987) using different criteria. A simple algorithm for splitting of edge/line segments is the one described in Grimson (1990) which is very similar to the algorithm presented in Ramer (1972). It splits an edge/line into straight segments in four steps:

1. Given a contour, build the connection between its two endpoints and measure the maximum deviation of the contour from the connection.
2. If the deviation exceeds a given threshold, the curve is split into two segments at the point with maximum deviation and the original connection line is replaced by two straight lines which connect the endpoints and the new splitting point.
3. Repeat the above steps until the deviation is below a threshold.
4. Finally, check consecutive segments to see if they can be merged into a single straight segment without exceeding the threshold.

A modified version of the above algorithm is used in this study in order to generate smooth segments instead of straight segments. The modified algorithm works in the same way as the above method in the splitting stage, but the merging process includes approximation of split segments and merging of consecutive segments into smooth segments. The threshold used in this study is set as three pixels as a road is usually very smooth. The reason that a constant is chosen instead of a function of the length of the connection between two endpoints is that the point with the maximum deviation does not always appear around the middle of the endpoints. This can ensure that the split segments are smooth.

In the splitting process, not only are those segments with high curvatures separated, but also smooth edge/line segments are divided into several smaller segments, as only deviation of an edge/line from the connection of its two endpoints is used. Segments belonging to smooth segments should be merged together to form longer segments. Therefore, two consecutive segments will be merged if they have the same spatial direction at their connection.

After the split-and-merge operation, short segments (<five pixels) are treated as noise and removed. The purpose of noise removal is to remove most irrelevant information. As not only road boundaries are extracted, but also the edges of other objects, such as trees, tracks, etc., are detected. The existence of non-road edges will greatly increase the computation time of the subsequent processing and even make the recognition process very difficult if they are brought into the *intermediate-level processing*. However, some true edge segments may be deleted in this process. This will be compensated in the grouping process which will be discussed in the next chapter.

Figure 4.15 shows an example of the split-and-merge operation. Figure 4.15(a) is a portion of an aerial image in rural area, which contains a curved road segment. The Canny and SE operators are used to extract edge and point features which are shown as red lines and blue dots respectively in Figure 4.15(b). As can be seen, most of the splitting points lie on non-road object boundaries, such as trees, various tracks, etc., as

they usually have irregular shapes. After merging, most split road boundaries are merged together as shown in Figure 4.15(c) while most splitting points on edges of non-road objects (green dots) still exist. Finally, edge segments with a length less than ten pixels are removed in order to delete edges of non-road objects further and the results are shown in Figure 4.15(d). Obviously, most irrelevant information has been removed while road boundaries are well retained.

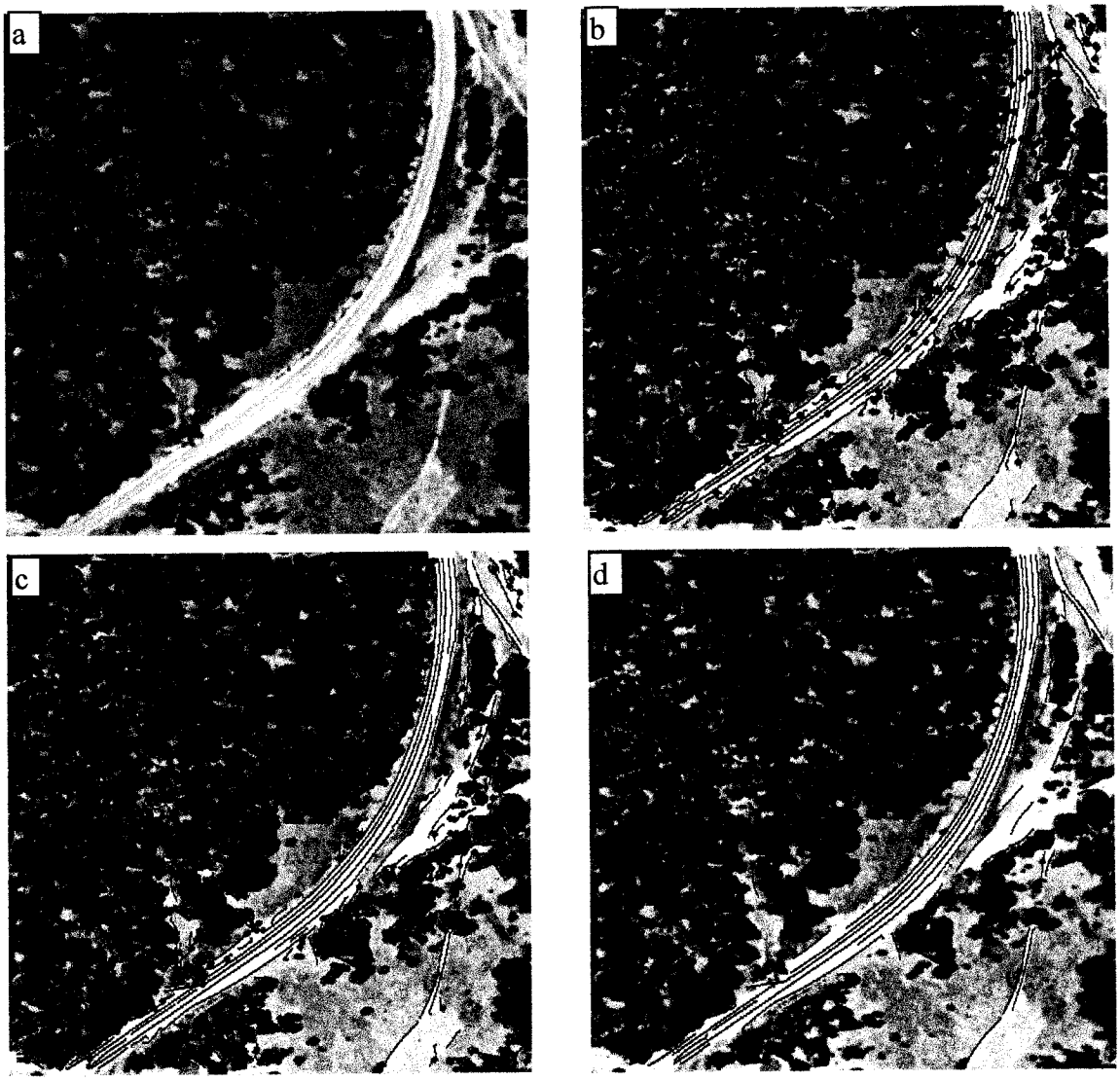


Figure 4.15 Split-and-Merge Operation (a) Portion of aerial image (b) Extracted edges with splitting points (c) Edge image with splitting points after merging (d) Final results after elimination of short edge segments

4.5 Summary of Chapter

This chapter has discussed some commonly used methods and algorithms for feature extraction in image understanding and some preprocessing techniques involved in low-level processing. The various methods or algorithms for feature extraction have different characteristics depending on the assumptions on which they are based. Previous study has shown that no single method or algorithm can work well in all situations. Therefore, the selection of a feature extraction operator is task-dependent. Preprocessing, such as edge tracking and linking, edge splitting and merging, etc., is also important in low-level processing. The results of low-level processing affect the result of image understanding directly. An edge/line tracking and linking algorithm has been developed based on the connectivity of edge points and their gradient magnitudes and orientations. To generate smooth edge/line segments, a split-and-merge procedure based on the deviation of edge/line points and their spatial orientations has also been proposed in this study.

Chapter Five

Generation of Structures of Road Segments

One important process in image understanding is the generation of object structures from image features. In bottom-up systems, objects are recognized by applying knowledge or models of objects to the generated object structures from the image data. The use of object structures can reduce the complexity of processing as knowledge of the objects can be introduced during the generation of object structures. One popular technique for formation of object structures in computer vision is aggregating subsets of features together which are likely to come from the same object. The process of organising coherent features into object structures is called *grouping* in computer vision.

There are two practical reasons for aggregating coherent features into object structures. Firstly, the number of features derived from the image is usually large, especially when the image has a complex structure. This will cause the search space to be extremely large when individual features are directly used in recognition. In grouping, image features which are likely to belong to the same object are aggregated to form structures which are used to generate hypotheses of objects, or to verify the generated object hypotheses. At the same time, spurious and irrelevant features are rejected in this process by introducing knowledge of objects. This will reduce the search space dramatically in the recognition phase (Lowe, 1985; Grimson, 1990). Secondly, when objects are recognized through establishing correspondence between image features and model features of objects, unreliable results are unavoidable because of the limited information delivered by individual features. In addition, due to the effects of image noise, occlusions, etc., some parts of objects cannot be recognized correctly. In grouping, not only are the properties of features but also their relations used (Lowe, 1985; Mohan and Nevatia, 1992; Fuchs and Förstner, 1995; Henricsson, 1996). Therefore, missing features may be recovered in this process. This will lead to more reliable object recognition.

This chapter will propose a generalised antiparallel pair to describe road structures in images and introduce the method for determining road structure in both 2-D and 3-D space. A method for grouping generalised antiparallel pairs into road-like features will then be described. Finally, road-like features will be represented in terms of a number of attributes which serve as facts in the recognition of roads in high-level processing.

5.1 Definition of Road Structure

A road is a man-made object. One important characteristic of man-made objects is that they usually have regular shapes and distinct contrast against their background. A road has a homogeneous surface which is bounded by two parallel boundaries and has distinct contrast against its background. Its appearance in the image depends on the image scale. In high-resolution images, a road is an elongated areal object bounded by two parallel boundaries while it appears as a line in low-resolution images. As the properties of road surfaces are not measurable in low-resolution images, the structure of a road segment will only be defined in high-resolution images. In Nevatia and Babu (1980), a road is described as an elongated area bounded by two parallel straight lines which have opposite gradient orientations (Figure 5.1(a)). This is only true when a road is straight. Such a definition is insufficient as the boundaries of a road are usually smooth curves rather than straight lines, which causes two problems. Firstly, only straight lines will be considered in generating road structures, causing some road parts with curved boundaries to be missed. Secondly, it cannot define the positions of curved road boundaries accurately. Therefore, to precisely describe road structure, a generalised antiparallel pair, briefly called *antipair*, is introduced instead of antiparallel lines. A generalised antiparallel pair is defined as (Figure 5.1(b)):

Two smooth curves which are pointwise parallel to each other and have opposite gradient orientations.

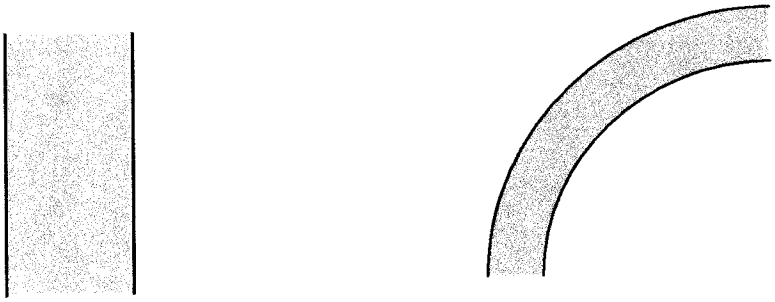


Figure 5.1 Structure of Road Segment (a). Straight road segment (b). Curved road segment

This definition reflects the situation in the real world and can give accurate location of road boundaries.

5.2 Generation of Generalised Antiparallel Pairs

5.2.1 Generation of Generalised Antiparallel Pairs in 2-D image

From the definition of generalised antiparallel pairs, their generation is straightforward. Two conditions embedded in the definition can be expressed as:

$$|\alpha_a - \alpha_b| \leq T_\alpha,$$
$$g_{\alpha 1} \cdot g_{\alpha 2} = 1.$$

α_a, α_b and $g_{\alpha 1}$, $g_{\alpha 2}$ are the spatial directions and gradient orientations of two edges a and b respectively. T_α defines the tolerance for difference between spatial directions of two edges, which is determined by the accuracy of computing spatial directions of edges. The first condition alone cannot guarantee the parallelism of edges due to errors in determination of spatial directions. To ensure the parallelism of edges, the condition that distance between two edges is a constant should be used. This condition can be represented by the standard deviation of the distance between two edges which can be described as:

$$\sigma_d < T_d$$

E_a E_b

The spatial direction of an edge at point i can be determined by the positions of its neighbouring edge points. To reduce the effects of position errors of edge points, the edge points which are two pixels away from the point i in forward and backward directions are used. The spatial direction is defined by:

$$\alpha_i = \tan^{-1}((y_{i+2} - y_{i-2})/(x_{i+2} - x_{i-2})) \tag{5.1}$$

where x_{i-2}, y_{i-2}, x_{i+2} and y_{i+2} are the image coordinates of point i-2 and i+2.

The gradient orientation of an edge in this case is not the absolute direction of the gradient, but the sign of the change in grey values across the edge relative to another edge. It is determined by the intensity values on both sides of the edge and its relative position to another edge. The grey values on both sides of an edge are computed by taking the average of grey values within a region defined by two parameters - offset and width (Henricsson, 1996). The offset defines how far away the region is from the edge and is used to reduce the blurring effects of the edge. The width determines the range of the region (Figure 5.2(a)). It should be chosen such that there are enough sample points and no other edges in the region. Suppose the grey values on both sides of an edge E_a are g_{int} and g_{ext} and there is a neighbouring edge E_b as shown in Figure 5.2(a), then the gradient orientation of edge E_a is defined as:

$$g_\alpha = \begin{cases} 1 & \text{if}(g_{int} - g_{ext}) > 0 \\ -1 & \text{if}(g_{int} - g_{ext}) < 0 \end{cases} \tag{5.2}$$

The RMS of the computed distance between two edges is determined by

$$\sigma_d = \sqrt{\Sigma(d_i - d_0)^2 / n} \tag{5.3}$$

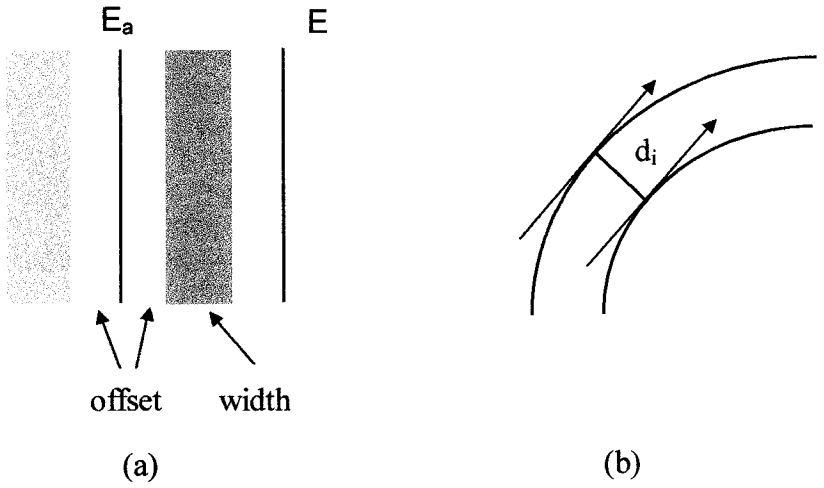


Figure 5.2 Definition of Region for Computing Attributes of Edge and Width of Road
(a) Offset and width defining the region for computation of attributes of an edge
(b) Distance between two edges

with $d_0 = \sum d_i/n$, n is the number of edge points being used in the computation. d_i is the distance between two edges at the point i (Figure 5.2(b)).

Based on the conditions defined above, the generation of generalised antiparallel pairs starts with the search for edges with the same gradient orientation. For a given edge, edges with the same gradient orientations on both its sides are selected. For each selected edge, its spatial directions are checked against the given edge point by point. Starting with the selected point of the given edge, the difference of spatial directions at the corresponding points on two edges and their distance are computed (Figure 5.2(b)). When the difference is within the defined threshold T_α, the point is marked, and the checking proceeds along the given edge. This process stops when the difference is larger than T_α for three consecutive points. Finally, the standard deviation of the distance between the given edge and the selected edge is calculated. When their standard deviation of distance is less than the given threshold T_d, they are taken as a generalised antiparallel pair. To reduce computation time, the search is limited to a range which is determined from the knowledge of road width in object space and the scale of the image. The threshold T_α is determined by the accuracy of determining spatial direction which is about 15°, and therefore it is set as 30° in this study. T_d is

determined by the accuracy of computation of distance which is about 1 pixel in image space. Therefore, it is set as one pixel in this study.

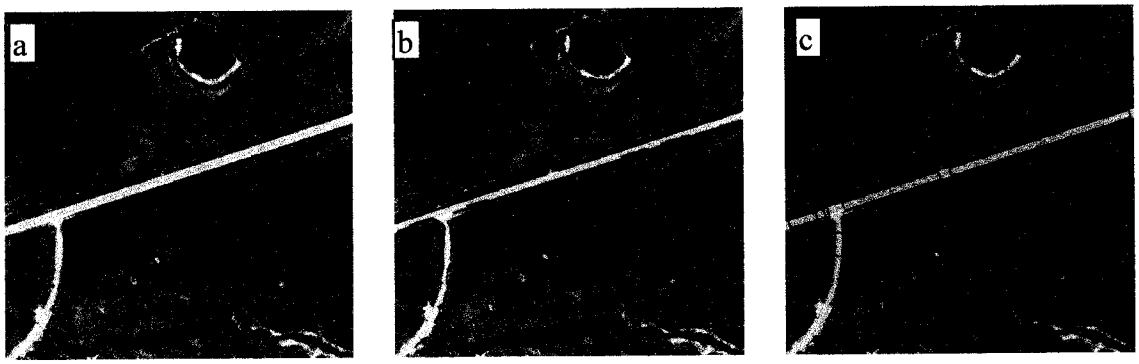


Figure 5.3 Generation of Generalised Antiparallel Pairs (a) Test image (b) Extracted edges (c) Generated antipairs

Figure 5.3 shows one example of generation of antipairs in the image. The image in this example is a portion of an aerial image with 500×500 pixels as shown in Figure 5.3(a). The result of edge extraction after edge tracking, linking and noise removal are shown in Figure 5.3(b). By applying the above method to the edge image, five antipairs are generated, four of which correspond to the motorway (Figure 5.3(c)).

5.2.2 Generation of Generalised Antiparallel Pairs in Object Space

The generation of generalised antiparallel pairs in object space is based on active contour models. In active contour models, a linear feature in an image is expressed by the sum of photometric and geometric energy, and external energy (Kass et al, 1988; Trinder and Li, 1995):

$$E = E_{geo} + E_{photo} + E_c \tag{5.4}$$

The geometric energy E_{geo} is determined by the geometric properties of the feature to be extracted, usually the smoothness of the feature. The photometric energy E_{photo} can be

provided by an operator. In Trinder and Li (1995), a linear feature in 3-D space is approximated by a cubic spline, and the geometric energy is defined by the first and second derivatives of the spline function in object space. The photometric energy is computed from the gradients of intensities along the feature on both left and right images while the external energy E_c is determined in the optimisation process using boundaries conditions. The total energy has the following form:

$$\begin{aligned} E = & \alpha X^T N_s^T N_s X + \beta X^T N_{ss}^T N_{ss} X \\ & + \alpha Y^T N_s^T N_s Y + \beta Y^T N_{ss}^T N_{ss} Y \\ & + \alpha Z^T N_s^T N_s Z + \beta Z^T N_{ss}^T N_{ss} Z \\ & + (f(x, y))^T (f(x, y)) + E_c \end{aligned} \tag{5.5}$$

where X, Y and Z are coordinate vectors of the feature, N_s and N_{ss} are the vectors of the first and second derivatives of the spline function. α and β are constants which control the influence of geometric energy against photometric energy. $f(x, y)$ are the gradients of the feature.

For a generalised antiparallel pair, its total energy should be the sum of the energies of both its sides (Trinder and Wang, 1998a), i.e.

$$E_{tot} = E_l + E_r \tag{5.6}$$

E_l and E_r are the energy of the left and right sides of a generalised antiparallel pair.

The two boundaries of a generalised antiparallel pair are parallel pointwise. This condition can be expressed as:

$$\Sigma \left(\left(\frac{\partial X_l(s)}{\partial s} - \frac{\partial X_r(s)}{\partial s} \right)^2 + \left(\frac{\partial Y_l(s)}{\partial s} - \frac{\partial Y_r(s)}{\partial s} \right)^2 + \left(\frac{\partial Z_l(s)}{\partial s} - \frac{\partial Z_r(s)}{\partial s} \right)^2 \right) = \min \tag{5.7}$$

where $\frac{\partial X_l(s)}{\partial s}, \frac{\partial Y_l(s)}{\partial s}, \frac{\partial Z_l(s)}{\partial s}$ and $\frac{\partial X_r(s)}{\partial s}, \frac{\partial Y_r(s)}{\partial s}, \frac{\partial Z_r(s)}{\partial s}$ are the directional derivatives of the corresponding points on the left and right side of the antiparallel pair with respect to the length s. It can be expressed in matrix form:

$$X^T P_s^T P_s X + Y^T P_s^T P_s Y + Z^T P_s^T P_s Z = \min \tag{5.8}$$

where $P_s = (N_s - M_s)$. N_s and M_s are the matrix of the first directional derivatives of points on the left and right side of the generalised antiparallel pair. This should be treated as an additional geometric energy of the generalised antiparallel pair. Therefore, the total energy of a generalised antiparallel pair is (Trinder and Wang 1998a):

$$E_{tot} = E_l + E_r + X^T P_s^T P_s X + Y^T P_s^T P_s Y + Z^T P_s^T P_s Z \tag{5.9}$$

Substituting equations (5.5) into the above equations, they become:

$$\begin{aligned} E_{tot} = {} & \alpha X_l^T N_s^T N_s X_l + \beta X_l^T N_{ss}^T N_{ss} X_l \\ & + \alpha Y_l^T N_s^T N_s Y_l + \beta Y_l^T N_{ss}^T N_{ss} Y_l \\ & + \alpha Z_l^T N_s^T N_s Z_l + \beta Z_l^T N_{ss}^T N_{ss} Z_l \\ & + (f_l(x, y))^T (f_l(x, y)) \\ & + \alpha X_r^T M_s^T M_s X_r + \beta X_r^T M_{ss}^T M_{ss} X_r \\ & + \alpha Y_r^T M_s^T M_s Y_r + \beta Y_r^T M_{ss}^T M_{ss} Y_r \\ & + \alpha Z_r^T M_s^T M_s Z_r + \beta Z_r^T M_{ss}^T M_{ss} Z_r \\ & + (f_r(x, y))^T (f_r(x, y)) + E_c \\ & + X^T P_s^T P_s X + Y^T P_s^T P_s Y + Z^T P_s^T P_s Z \end{aligned} \tag{5.10}$$

where l and r stand for the left and right side of the antiparallel pair respectively. X_l, Y_l, Z_l and X_r, Y_r, Z_r are the coordinates of the left and right side of the antiparallel pair. N_{ss} *and* M_{ss} are the matrix of the second directional derivatives of points on the left and right side of the antiparallel pair.

Suppose

$$X = \begin{pmatrix} X_l \\ X_r \end{pmatrix}, Y = \begin{pmatrix} Y_l \\ Y_r \end{pmatrix}, Z = \begin{pmatrix} Z_l \\ Z_r \end{pmatrix}, O = \begin{pmatrix} N & 0 \\ 0 & M \end{pmatrix}, O_s = \begin{pmatrix} N_s & 0 \\ 0 & M_s \end{pmatrix},$$

$$O_{ss} = \begin{pmatrix} N_{ss} & 0 \\ 0 & M_{ss} \end{pmatrix}, \text{ and } G^T = ((f_l(x, y))^T \ (f_r(x, y))^T),$$

The Equations (5.10) can be rewritten in a simplified form:

$$\begin{aligned} E_{tot} = {} & \alpha X^T O_s^T O_s X + \beta X^T O_{ss}^T O_{ss} X + X^T P_s^T P_s X \\ & + \alpha Y^T O_s^T O_s Y + \beta Y^T O_{ss}^T O_{ss} Y + Y^T P_s^T P_s Y \\ & + \alpha Z^T O_s^T O_s Z + \beta Z^T O_{ss}^T O_{ss} Z + Z^T P_s^T P_s Z \\ & + G^T G + E_c \end{aligned} \tag{5.11}$$

To minimise the total energy, the following conditions should be satisfied:

$$\frac{\partial E_{tot}}{\partial X} = \frac{\partial E_{tot}}{\partial Y} = \frac{\partial E_{tot}}{\partial Z} = 0 \tag{5.12}$$

A further development of the above equations will lead to the equations for estimating the coordinates X, Y and Z, which have the following forms:

$$(B + O^T F_X^T F_X O + \lambda I)\Delta X + O^T F_X^T F = 0 \tag{5.13}$$

$$(B + O^T F_Y^T F_Y O + \lambda I)\Delta Y + O^T F_Y^T F = 0 \tag{5.14}$$

$$(B + O^T F_Z^T F_Z O + \lambda I)\Delta Z + O^T F_Z^T F = 0 \tag{5.15}$$

where F_X, F_Y and F_Z are the summation of the first derivatives of all images to space coordinates X, Y and Z respectively and $B = \alpha O_s^T O_s + \beta O_{ss}^T O_{ss}$. ΔX, ΔY and ΔZ are the corrections to approximations X_0, Y_0 and Z_0.

With the generated generalised antiparallel pairs on the left or right images as approximations and an underlying DTM, their positions in 3-D space can be determined. The generated antipairs in object space are called 3-D antipairs.

5.2.3 Merging of Generalised Antiparallel Pairs in Object Space

As described in the last section, 3-D antipairs can be determined based on the generated antipairs on the left or right images and an underlying DTM of the region. Due to the effects of image noise, occlusions, such as trees, shadows, vehicles on the road surface, etc., the generated antipairs on the left and right images corresponding to a road may not be same. Thus, 3-D antipairs generated from them may not coincide. Sometimes, they overlap each other. In this case, they should be merged to form a complete 3-D antipair.

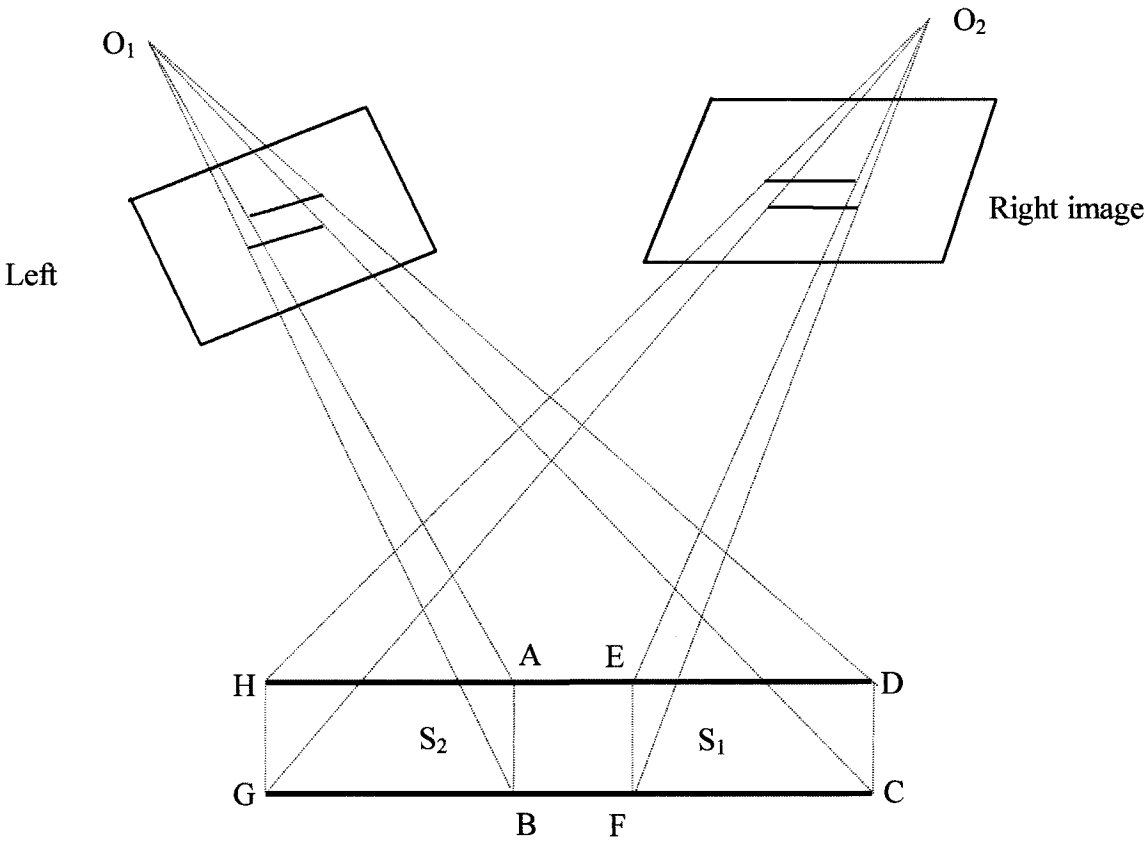


Figure 5.4 Merging of 3-D Antipairs

Figure 5.4 shows two overlapping 3-D antipairs generated from the left and right images. Suppose the antipairs on the left and right images are s_1 with end points a, b, c, d and s_2 with end points e, f, g, h. Based on s_1 and s_2, 3-D antipairs S_1 with endpoints A, B, C, D and S_2 with endpoints E, F, G, H are generated. They overlap in the section ABEF. They are merged and replaced by a new 3-D antipair with end points C, D, H and G. The process of merging 3-D antipairs can partially overcome the problem caused by image noise and occlusions by bridging gaps between the antipairs on a single image.

5.2.4 Attributes of 3-D Antipairs

After 3-D antipairs are generated, their attributes are determined in order to generate their symbolic representations for the subsequent processes of grouping of antipairs, recognition of road segments, etc. The attributes of a 3-D antipair include the properties of its two edges and the area bounded by the edges. To describe the homogeneity of a road surface, the average intensity value and standard deviation of intensity values within the area between two edges are computed. As road surfaces are usually made of concrete and asphalt, only monochrome values are computed rather than chromatic values.

In addition to the properties of the region bounded by the edges of a 3-D antipair, attributes of the two edges are also computed. These include the gradient orientation and positions of the edges, width of the antipair and spatial directions of the antipair at both ends. The gradient orientation has a value of -1 when the change in intensity across the edge has an inward direction and 1 for outward direction. The positions of two boundaries are given by two point chains which contain the positions of all points. The end directions are taken as the normal to the connection of two end points at the same end. With these computed attributes, a generalised antiparallel pair can be expressed by a *structure* in Prolog which will be introduced in the next chapter. It has the following form:

antipair(antipair_no,
left_side(L1), right_side(L2),
attribute(end_directions, length, width,
gradient, average_intensity)).

5.3 Grouping of 3-D Antipairs

Due to the effects of image noise, occlusions, such as trees and shadows, poor contrast between road surface and its background, etc., the generated 3-D antipairs are often separated. To derive road structures for high-level processing, they need to be grouped to form road-like features. This section will introduce some approaches for grouping of object features into structures and describe the method for grouping of 3-D antipairs.

5.3.1 Perceptual Organisation in Computer Vision

Perceptual organisation is defined as "a basic capability of the human visual system to derive relevant groupings and structures from an image without prior knowledge of its contents." (Lowe, 1985). The study of perceptual organisation reached its heyday during the 1920s and 30s with the foundation of the Gestalt School of Psychology. The major contribution of the Gestalt psychology to computer vision was the demonstration of the role of grouping phenomena in perception of object structure. These phenomena, found by the Gestalt psychologists, are called the *Gestalt laws* and can be categorized roughly into six groups (Lowe, 1985):

(1) Proximity – elements that are closer together tend to be grouped together,

(2) Similarity – elements that are similar in physical attributes, such as colours, size or orientation, are grouped together,

(3) Continuation – elements that lie along a common line or smooth curve are grouped together,

(4) Closure – there is a tendency for curves to be completed so that they form closed regions,

(5) Symmetry – any elements that are bilaterally symmetric about some axis are grouped together, and

(6) Familiarity – elements that are usually seen together are grouped together.

Unfortunately, the Gestalt psychologists did not explain *why* and *how* perceptual organizsation is computed. To answer the question *why*, several theories were developed, one of which is the principle of non-accidentalness proposed by Witkin and Tenenbaum (1983) and Lowe (1985). The non-accidentalness implies that perceived structures are unlikely to have arisen by accident, instead they are formed by image features related by causal relationships.

The use of perceptual organisation in computer vision was initially proposed by Marr (1976) in his initial work on the primal sketch. Marr developed the idea that the primal sketch should include not only edge information, but also groupings of edges into larger structures. He suggested groupings on the basis of curvilinearity, and lines should be grouped based on parallelism and collinearity. Unfortunately, this was never developed in detail in his later work. Witkin and Tenenbaum (1983) analysed the role of structure in vision and found that humans perform organisation in perception processes without knowing what the structures to be organised are. They pointed out " the naive observer often sees essentially the same things an expert does, the difference between naive and informed perception amounting to little more than labelling the perceptual primitives. It is as if the visual system has same basis for grouping *what* is important without knowing *why*." Witkin and Tenebaum discussed the implication of structure for computational vision. However, they did not explain how perceptual organisation should be computed.

Based on the principle of non-accidentalness, Lowe (1985) classifies image relations into accidental and causal. Accidental relations may be caused by an accident of viewpoint or other factors while causal ones result from the meaningful structures of objects. He uses the prior expectation of a relation, i.e. probability to evaluate its significance of non-accidentalness and to find instances of certain relations, such as

collinearity, proximity of end points and parallelism. Following Lowe's work, many methods for computing perceptual organisation have been proposed (Fischler and Bolles, 1986; Boldt et al, 1989; Dolan and Weiss, 1989; Mohan and Nevatia, 1989, 1992). In computing perceptual organisation, primitive structures are usually determined first by using simple geometric relations such as collinearity, proximity, parallelism, etc. Larger structures are then derived by grouping the generated primitive structures. In Mohan and Nevatia (1989), to create structures of buildings with rectangular shapes, the relation parallel is generated first. U structures and rectangles are then derived based on the generated parallels. In their later work (Mohan and Nevatia, 1992), more relations such as continuity, symmetry, closure, etc. are used to compute more complex structures.

Fuchs and Förstner (1995) presented a polymorphic grouping technique for generating object structures based on extracted point, edge, line and region features and their topological relationships. Relations between neighbouring features are generated by a process of exoskeletonisation and represented by a Feature Adjacency Graph (FAG). Spurious features, missing features and merged features can be detected through the analysis of FAG. Relations between various features can also be derived from the FAG.

Most methods for computing organisation are based on one or more properties of the Gestalt laws. They largely rely on the spatial arrangement of primitive features. They can build object structures effectively when objects have simple geometric shapes, such as rectangles, U shapes, etc. However, they may create unsatisfactory results when objects have complex structures (Henricsson, 1996). Henricsson argues that features belonging to the same object are likely to have similar region properties and presented a grouping method based on similarity on position, orientation, photometric and chromatic properties of features. This method can generate reliable results as not only geometric properties of features, but also their photometric and chromatic properties are used. A road is a areal feature in high-resolution images and the generated antipairs possess both geometric and radiometric properties. Therefore, this method can be used for grouping of antipairs.

5.3.2 Grouping of 3-D Antipairs

The purpose of grouping in this study is to aggregate antipairs to form road-like structures for high-level processing. Antipairs are primitive structures of roads which not only have spatial positions and orientations, but also possess geometric and photometric properties. They have similar properties if they belong to the same road. At the same time, two antipairs should be close to each other and collinear if they are neighbours in space. These conditions can be formulated into one *rule* in Prolog (Trinder et al, 1997, 1998c), i.e.

connect(X, Y) :- *relation*(X, Y,
attribute(distance,
difference_in_direction, difference_in_height, difference_in_gradient,
difference_in_intensity, difference_in_width)),
distance < $\mathbf{T}_d$,
difference_in_direction < $\mathbf{T}_\alpha$,
difference_in_height < $\mathbf{T}_h$,
difference_in_gradient is 0,
difference_in_intensity < $\mathbf{T}_g$,
difference_in_width < $\mathbf{T}_w$.

Sometimes, several possible connections may be found for a selected antipair when the above *rule* is applied. In this case, a similarity value is calculated based on their similarity in geometric and photometric properties, positions and orientations. The antipair with the maximum similarity value is then selected as the true connection of the selected antipair. Each similarity component is determined by one or more of properties of antipairs and represented by a score between 0 and 1. A high score means that two antipairs are very similar, while 0 stands for no similarity between two antipairs. To assess photometric similarity of two contours, Henricsson (1996) used a Gaussian function which has the following form:

$$S_L(\Delta L) = e^{-\frac{\Delta L^2}{2\sigma_L^2}} \tag{5.16}$$

where ΔL is the Euclidean distance between the means of photometric attributes of edges. σ_L is the control parameter defined by the user.

For an antipair, its photometric attributes include average intensity value and standard deviation of intensity values in the area defined by two edges of the antipairs. Thus, the photometric similarity of two antipairs is defined in a similar way to Henricsson's:

$$S_P = \begin{cases} e^{-\frac{\Delta G^2}{2T_G^2}+\frac{\Delta \sigma^2}{2T_\sigma^2}} & \text{if } |\Delta G| < T_G \text{ and } |\Delta \sigma| < T_D, \\ 0 & \text{otherwise} \end{cases} \tag{5.17}$$

where ΔG and $\Delta \sigma$ are the difference of intensity values and standard deviations of intensity values of two antipairs respectively. T_G and T_σ are two control parameters which define the tolerance of intensity value and standard deviation of intensity values. They can be determined through a large number of measurements on the image.

Two antipairs should have same width if they correspond to the same road. Therefore, geometric similarity can be defined as:

$$S_G = \begin{cases} 1 & \text{if } |\Delta W| < T_W, \\ 0 & \text{otherwise} \end{cases} \tag{5.18}$$

where ΔW is the difference of widths of two antipairs and T_W is the corresponding threshold which can be determined by the accuracy of computed width and is set as one pixel in image space in this study.

A road should not have a large change in height in a limited distance. This can be expressed as:

$$S_h = \begin{cases} 1 & \text{if } |\Delta h| < T_h, \\ 0 & \text{otherwise} \end{cases} \tag{5.19}$$

where Δh is the difference of heights of two antipairs and T_h is the corresponding threshold which can be obtained from the manuals of road construction.

The similarity in position of two antipairs is solely determined by the distance between their ends. The closer they are in space, the more similar they look. The similarity in position is given by:

$$S_D = \begin{cases} e^{-\frac{D^2}{2T_D^2}} & \text{if } D < T_D, \\ 0 & \text{otherwise} \end{cases} \tag{5.20}$$

where d is the distance between two antipairs and T_D is the threshold. T_D defines the maximum distance between two antipairs which could be linked by grouping. In an image, occlusions could be a few hundred pixels long. Thus, to remove the effects of occlusions, a large T_D should be chosen. However, a large T_D may correspond to a large difference of spatial orientations of antipairs. This means that a large threshold for spatial orientation should also be selected in grouping. This will decrease the reliability of grouping. Therefore, to ensure the reliability of grouping, the threshold T_D should be considered together with another threshold T_α.

The similarity in orientation of antipairs is determined by the difference of their spatial orientations. It is defined as:

$$S_O = \begin{cases} e^{-\frac{\Delta\alpha_1^2+\Delta\alpha_2^2}{2T_\alpha^2}} & \text{if} |\Delta\alpha_1| < T_\alpha \text{ and } |\Delta\alpha_2| < T_\alpha, \\ 0 & \text{otherwise} \end{cases} \tag{5.21}$$

$\Delta\alpha_1$ and $\Delta\alpha_2$ are the differences between the spatial orientation of two antipairs and their connection, and T_α is their threshold. A road is a curved linear object with smooth change in spatial orientation. To ensure the reliability of grouping, a small T_α should be chosen. However, this will limit the length of occlusions which could be recovered in grouping. Therefore, a trade-off between the possible length of occlusions and the threshold T_α should be made. In this study, a moderate value, 30°, is chosen for T_α since antipairs belonging to different objects may be grouped if it is too large while antipairs belonging to the same road will be rejected if it is too small. According to the manual of road construction (Underwood, 1991), this value will yield a road length of about 58 m for a road with a speed limit of 60 km/h. This correspond to about 80 pixels in the image with a scale of 1:25,000 and a pixel size of 25 μm.

The similarity of two antipairs is the sum of their similarity in position, orientation, geometric and photometric properties, i.e.

$$S_{tot} = S_P + S_G + S_h + S_D + S_O \tag{5.22}$$

For each antipair, its similarities with all other antipairs will be calculated and the antipair with maximum similarity is selected as its neighbour.

Two examples of grouping of antipairs are shown in Figure 5.5. Both images shown in these examples are parts of an aerial photograph with a scale of 1:25,000 and have a size of 500×500 pixels. Figure 5.5(a) contains a straight road segment and five antipairs are generated using the method described in Section 5.2 while a curved road segment is shown in Figure 5.5(c) and 35 antipairs in this image. In the first example, the generated antipairs belonging to the same road are separated due to disturbance near the road

boundary and an intersection. The gaps between antipairs are bridged during grouping based on the above similarity grouping method (Figure 5.5(b)). In the second example, the generated antipairs belonging to the road segment are separated at the upper left and lower left, where the contrast between road surface and its background is very low and there is a tree near the road boundary. Again they are successfully grouped to form a longer antipair which correspond to the road segment (Figure 5.5(d)).

5.4 Summary of Chapter

In this chapter, a generalised antiparallel pair is presented to describe road structures in high-resolution images. It not only describes the spatial structures of roads, but also presents the geometric and photometric properties of road surfaces. Its generation in image space is solely based on the geometric and radiometric properties of edges. To generate 3-D antipairs, a modified Snakes method has been developed, in which the geometric energy is determined by both the smoothness of road boundaries and the parallelism of two boundaries. The generated antipairs are represented as a structure *antipair* in Prolog, with a number of geometric and photometric attributes. To form road-like objects, antipairs are grouped, based on their geometric and photometric properties, spatial positions and orientations. A rule *connect* is formulated for the grouping of antipairs. The formed road-like objects are described by the structure *feature* which is similar to *antipair* and stored as facts in the knowledge base for high-level recognition of roads.

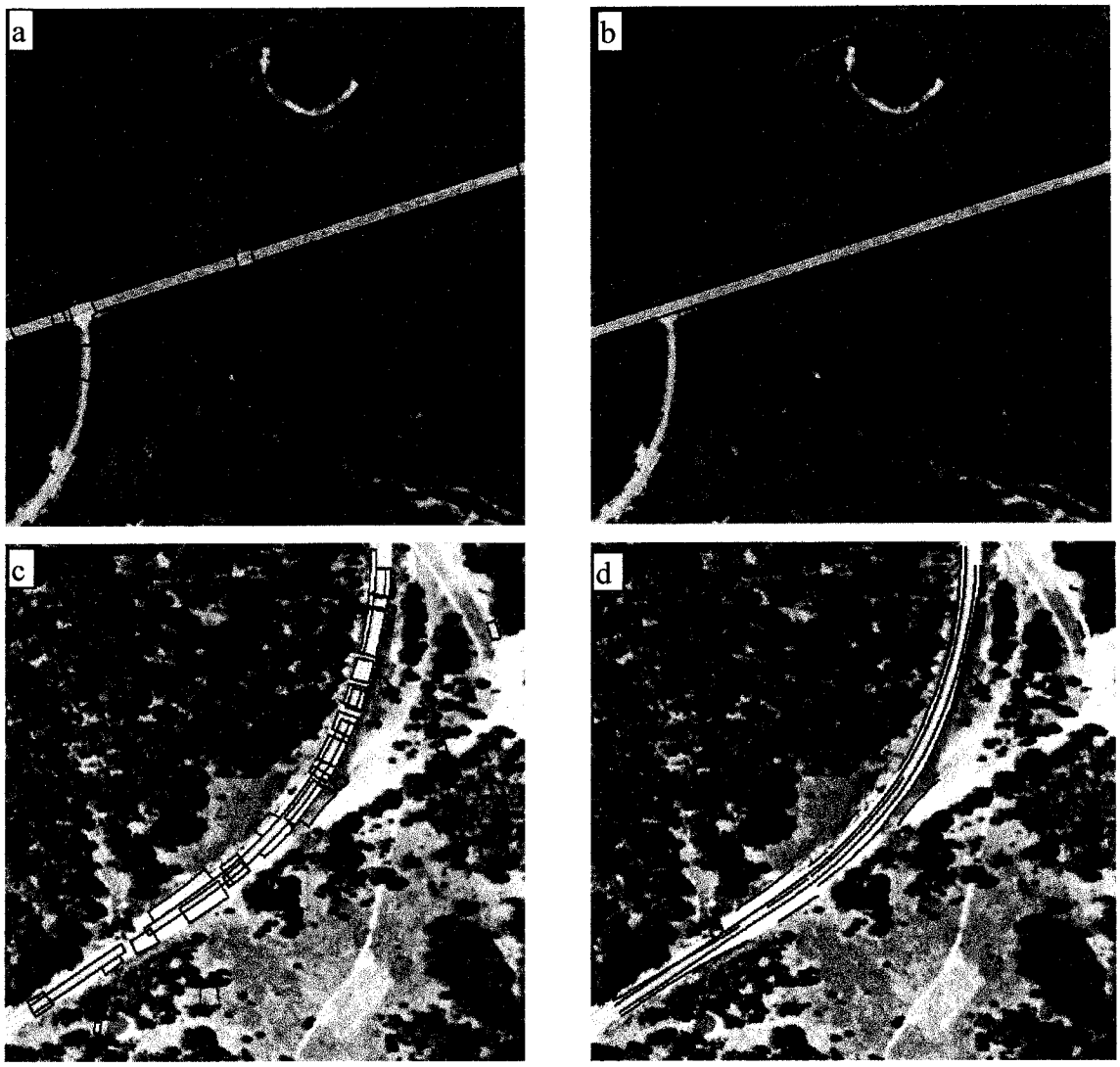


Figure 5.5 Grouping of Antipairs (a) Generated antipairs of straight road segment (b) Formed road-like features after grouping (c) Generated antipairs of curved road segment (d) Formed road-like features after grouping

Chapter Six

Knowledge-Based Road Recognition

Knowledge-based road recognition refers to automatic recognition of roads from images using artificial intelligence techniques. As discussed in Chapter 3, a knowledge-based object recognition system usually consists of three basic components, i.e. image analysis for feature extraction and formation of object structures, a knowledge base which contains domain knowledge of the objects to be recognised and intermediate results derived from image analysis and reasoning, and a control mechanism which manages the execution of the various processing involved in recognition. Since image analysis techniques have been introduced in the previous chapters, this chapter will concentrate on road recognition from aerial images using domain knowledge of roads. A semantic road model for recognition, including definition of roads in different resolutions of images and their relations, is defined in Section 6.1. Section 6.2 will discuss what knowledge can be used for recognition of roads and how it is represented in the knowledge base. A complete process for road recognition will be described in Section 6.3 while Section 6.4 gives the details of the recognition process.

6.1 Semantic Road Model

In order to extract and recognise roads from images automatically, it is necessary to define a model of a road. A road in Webster's dictionary is defined as "a place where one can ride; an open way or public passage for vehicles, persons and animals; a track for travel or for conveying goods, etc., forming a means of communication between one place and another." This definition is more or less a functional description of a road, and hence it would be very difficult to recognise roads in the image by it. A road is a man-made object which has some distinct properties which distinguish it from other objects. At the same time, it is related to other roads and objects because of its functions. These

specific properties and functions give a more complete definition of a road and, thus should be included in the road model for its recognition from images. This section will introduce some existing road models for road extraction from remotely sensed imagery and describe the semantic road model in this study.

6.1.1 Existing Road Models

Over the last two decades, various road models have been proposed for road extraction from remotely sensed imagery. A review of existing road models can be found in Trinder and Wang (1998b). They can be categorized into two types: generic and semantic. Generic road models define a road based on its generic properties in geometry and photometry, e.g. constant width, smooth boundaries, homogeneous surface, etc. These are directly used for recognition of roads from images (Bajcsy and Tavakoli, 1976; Nagao and Matsuyama, 1980) or formulated into a mathematical model or a process which can be executed by computer (Kestner and Kazmierczak 1978; McKeown and Delinger, 1988; Grün and Li, 1994; Trinder and Li, 1995; Vosselman and Knecht, 1995). In Bajcsy and Tavakoli (1976), the model for recognition of roads from satellite images describes a road as a linear object with bounded width and curvature, distinct grey value and a certain length. Nagao and Matsuyama (1980) define a road as an elongated linear object with constant width and smooth curvature for road recognition in rural areas. Obviously, these road models are simple and incomplete since they do not contain the specific geometric and photometric properties of roads, their spatial and contextual information. They may interpret non-road segments with similar geometric structures to roads. Grün and Li (1994) define a road as a linear object with higher intensity than its background, homogeneous intensity along its surface, smooth boundaries and consistent width. They formulated these properties into an optimisation process which is implemented by dynamic programming. Vosselman and Knecht (1995) classified road characteristics into five types: geometric, photometric, topological, functional and contextual, and used the first two in their road model for road tracing in aerial images. These two road models were developed for locating roads in images with approximate positions of roads defined by an operator, i.e. recognition is performed by an operator.

Semantic road models not only include the specific properties of roads and road parts, but also define relations between roads and with other objects (Hwang et al, 1986; Cleynenbreugel et al, 1990; Gunst, 1996; Bordes et al, 1997; Baumgartner et al, 1997). In Hwang et al (1986), a road model was developed for interpretation of roads in urban areas based on the specific properties of a road segment and its relationships with neighbouring houses. In urban areas, roads are usually straight and connect to houses which are distributed regularly. They appear as elongated objects connected to houses having rectangular shapes in large-scale images. Therefore, in their model a road segment is defined as an elongated rectangle constrained by a number of geometric and radiometric parameters, and the spatial relationships between roads and houses are also included. The geometric and radiometric properties of road segments are used to generate hypotheses of road segments, and spatial relationships are utilised to verify the generated hypotheses.

Different types of terrain usually correspond to different structures of road networks. In flat areas, roads are usually straight, but they tend to be serpentine in mountain areas. Based on these characteristics of roads, Cleynenbreugel et al (1990) presented two road models for road extraction from SPOT imagery for different types of land cover, in which roads are mainly defined by their shapes and relations. No specific properties of road surface are used because of the limitation of image scale. To obtain topographic information of terrain, such as the type of land cover and relationships between roads, a geographical information system (GIS) is used. As no specific properties of roads are used, it is possible to generate ambiguous interpretations.

The appearance of a road in the image varies with the image scale. In large-scale images, a road is an elongated homogeneous area bounded by two road boundaries, while it appears as a line in small-scale images. Based on this observation, a multi-scale road model for change detection of highways was developed by Gunst (1996). In this model, roads are defined at three different levels of image scales, i.e. small, medium and large. In small-scale images, a road is defined as a line feature while an intersection is represented as a point. At medium-scale, both roads and junctions are defined as planar features with

different shapes. Roads are classified into main carriage-ways, link roads and service roads. In large-scale images, road segments and junctions are decomposed into different road parts, such as traffic lanes, hard shoulders, edge lines, etc. They are defined by specific geometric and radiometric properties which can be obtained from the construction manuals of roads and their spectral properties. Roads in different scale images are related by part-whole and specialisation-generalisation relationships, and road parts in large-scale images are related by spatial relationships. Part-whole relationships are used to generate hypotheses of roads in larger scale images based on roads in existing small-scale maps, while spatial relationships between road parts are used for spatial reasoning, i.e. predicting the existence of road parts. The final interpretation of road parts is based on their specific properties. This model has been implemented in both medium- and large-scale images (Gunst and Vosselman, 1997). As the relationships between roads and road parts in the model are developed for a special type of road, it may not be suitable to recognition for other types of roads.

Baumgartner et al (1997) proposed a road model for road extraction from aerial images in which roads are defined in three different levels, i.e. the real world, geometry and material, and image. In the real world, a road network consists of intersections and roads which are in turn discomposed into pavement and road markings. In the geometry and material level, road parts are described by their shapes and materials of road surface, while in the image level, roads are defined in both small- and large-scale images. In small-scale images, a road network is defined as a set of straight bright lines and bright blobs. In large-scale images a road segment and an intersection are described as an elongated bright area and a compact bright region respectively. Roads and road parts at different levels are related by "concrete" relations. However, the relationships between roads at two different scales of images are not defined explicitly. The model contains contextual information, i.e. relationships between roads and other objects, such as buildings and trees, which are used to guide the connection of road segments.

Bordes et al (1997) argue that most existing road models for automatic road extraction are limited to certain types of roads, and hence are not suitable for the extraction of other

types of roads or roads in different landscapes. To solve this problem, they built an image-independent semantic road model in which geometric properties, semantic attributes and contextual information are included. The model includes geometric description, initial crossroads, final crossroads, number of lanes, type of road, position on the ground, type of land cover and neighbouring linear objects. These geometric properties, semantic attributes and contextual information of roads are derived from the existing cartographic databases and differ according to the types of roads and their positions. As the road model is defined based on an existing cartographic database, the properties and attributes included in the model need to be transferred into image space. One obvious limitation of this model is that it can only be used to determine the correct positions of the existing roads in databases based on their available information in the database, and thus it cannot automatically recognise new roads in images.

6.1.2 Semantic Road Model in This Study

It can be concluded that in order to recognise roads in images reliably, a semantic road model should include not only their specific properties, but also relationships with other objects, and the relationships between roads (contexts). The image scale should also be taken into account as it affects the appearance of a road in the image. In low-resolution images, a road appears as a line, and the geometric and radiometric properties of road surfaces are not measurable in this case. What can be measured are the gradient and curvature of the road in the image, which are insufficient to yield correct recognition of roads. However, the overall structure of the road network is well defined in low-resolution images. In high-resolution images, the structure of the road surface is well defined. At the same time, the geometric and radiometric properties of road surfaces are measurable and road markings such as lane lines may be visible, but the global structure of the road network is not exhibited because of the limitation of image size. It is clear that different resolutions of images reflect different aspects of characteristics of roads (Mayer and Steger, 1998). Low- or high-resolution of images alone cannot yield reliable results. But, the combination of low- and high-resolution images can provide more complete information of roads than a single resolution image as they complement each

other. Therefore, more reliable recognition results can be expected when both high- and low-resolutions images are used.

In this study, a multi-resolution semantic road model is proposed, in which both roads and intersections are defined in low- and high-resolution images. In low-resolution images, roads and intersections are defined as smooth lines and points respectively. At this level, the specific properties of road surfaces are not considered, but the topological relationship between roads and intersections is defined. Each road is defined by two neighbouring intersections and roads are related by spatial connectivity relationships. According to the angle at which two roads intersect, an intersection can be classified as T, Y or four-way intersection. In high-resolution images, a road segment is defined as an elongated homogeneous area bounded by two parallel boundaries, with specific properties of the road surface. These include the geometry of the road surface, such as length and width which can be obtained from manuals of road construction, and radiometric properties of road surface including gradient direction, average intensity and standard deviation of intensity. These values can be obtained by training in which a large number of roads are measured and the statistical values of their properties are calculated. Road junctions are also homogeneous areas, but have different shapes (T, Y and four-way junctions). A road can be classified into different categories by its width. In high-resolution images an image usually contains part of a road due to the limitation of image size in image processing, and hence it can be considered as part of a road in low-resolution images and is related to the road in low-resolution images by a *part-of* relationship. Road junctions are treated as specialisation of intersections in low-resolution images, and hence they are related to intersections in low-resolution images by specialisation-generalisation relationships.

In addition to the relationships between roads, relationships between roads and other objects such as buildings, trees, etc., should also be included. In rural areas, roads are not often connected to buildings, but often occluded by trees. Therefore, relationships between roads and buildings are not considered in recognition of roads in rural areas. In

this study, the spatial relationships between roads and trees and shadows are included and used to detect road parts occluded by them.

6.2 Knowledge Representation

Knowledge plays a very important role in object recognition from images. As discussed in Chapter 3, domain knowledge should be collected and represented in the knowledge base in a certain form before recognition. This section will discuss what knowledge should be used for recognition of roads from aerial images and how it is represented.

6.2.1 Knowledge for Road Recognition

Knowledge is the known and useful information for solving a specific problem. According to the definition of a road model, the essential information for recognition of roads is the structure of road surfaces and their geometric and photometric properties. They determine what a road looks like in the image and the internal relations between road parts. However, they are insufficient to generate reliable recognition of roads as objects with similar appearances may also be incorrectly recognised as roads. To avoid ambiguous recognition of roads, relationships between roads and between roads and other objects such as trees should also be used. Therefore, the knowledge for recognition of roads from images includes:

- Structure of road surface
- Geometric properties, including road width, curvature, etc.
- Photometric properties, such as average intensity, gradient orientation, etc.
- Spatial relations between roads
- Part-of relations between roads in different resolutions of images.
- Spatial relations between roads and other objects such as trees, shadows, etc.

6.2.2 Knowledge Representation Mechanism

As introduced in Chapter 3, there are different techniques for knowledge representation, such as production rules, semantic networks, frames, etc. Each method has its own advantages and disadvantages. The selection of the knowledge representation approach depends on the type of knowledge to be used for solving a specific problem. The knowledge for recognition of roads includes their specific geometric and photometric properties and relations between roads, which are used to define road segments, and hence is declarative. Relations between roads include connectivity between roads, *part-of* relationships between roads in different resolutions and specialisation-generalisation relationships. *Part-of* relations can be used to guide the search for road segments in high-resolution images or verify hypothesised road segments. They are procedural when they are used to search for road segments in high-resolution images while they are declarative if they are utilised for verification. In this study, they are mainly used for verification purpose. Specialisation-generalisation relationships are the relations between the road junctions in high-resolution images and road intersections in low-resolution images and are used to guide the search for road junctions in high-resolution images. Connectivity between roads in space are spatial relationships. They are used for the prediction of missing road segments. Both of these relationships are procedural. In this study, knowledge representation is based on rules as they can represent both declarative and procedural knowledge, and they are easy to understand and to update. The rules will be represented using Prolog.

Prolog is a commonly used programming language in artificial intelligence which is rooted in the first-order logic. It uses *structures*, *relations* and *rules* to represent knowledge (Bratko, 1990). In Prolog, an object is expressed by a variable or a *structure*. A *structure* is composed of a structure name (*functor*) and a number of components. Its components can be parts of the object or its properties. For example, a point in 3-D space can be expressed as point (X, Y, Z). X, Y and Z are variables. A component of a *structure* can be a *structure*, e.g. triangle (point (X1, Y1, Z1), point (X2, Y2, Z2), point (X3, Y3, Z3)). In this way, a representation tree of objects can be established.

One of the most common expressions in Prolog is a *relation.* A *relation* may be a connection between two objects or one property of an object. It consists of a relation name and one or more objects. If there is only one object in a *relation*, it is called unary relation, e.g. red (apple). A *relation* is referred to as binary when it includes two objects, e.g. the *relation* parallel (line1, line2). Unary *relations* are usually used to represent properties of objects while binary *relations* are employed to describe the various relations between objects.

A *rule* in Prolog is a logic expression. It contains two parts: a *condition* part and a *conclusion* (action) part. The *conclusion* and *condition* parts are sometimes called *goal* and *predicates* of the rule. The *conclusion* part of a rule is usually a *relation* while the *condition* part includes one or more than one *relations*, which are connected by AND/ OR. The following is a *rule* for recognition of a bridge in low-resolution images.

```
bridge (X) :- line (X, L),
    L > T_l,
    above (X, Y),
    river (Y).
```

The left-hand side of the above *rule* is the *conclusion* while the right-hand side is the *condition* part. This *rule* states that an object is a bridge if it is a linear feature with a length larger than a given value in low-resolution images and is above a river.

In addition to *structures*, *relations* and *rules*, another type of Prolog clause is *query* which is used by users to ask questions . For example, bridges can be found by using the following *query*:

```
?- bridge (X).
```

Once this question is asked, Prolog will find linear objects which satisfy the conditions included in the rule bridge.

6.2.3 Knowledge Representation

As introduced in Section 6.2.1, the knowledge for recognition of roads includes specific properties of roads, including their geometric and photometric properties and relationships between roads and between roads and other objects. Results generated from the *intermediate-level processing*, such as road structures derived from images should also be represented.

6.2.3.1 Relationships between Roads

The relationships between roads include their connectivity relationships in low-resolution images, neighbourhood relationships in high-resolution images and part-of relationships. These relationships are represented as *relations* in Prolog. The connectivity relationships between roads in low-resolution images are expressed by the relation *adjacent*:

adjacent (Road1, Road2)

Road1 and Road2 are two roads in low-resolution images.

A neighbourhood relationship between roads in high-resolution images is represented in the same way that the connectivity relationships of roads in low-resolution images are represented. It has the following form:

neighbour (Road_segment1, Road_segment2, D)

Road_segment1 and Road_segment2 are road segments in high-resolution images. D is the distance between them.

The part-whole relationships between roads in high-resolution images and roads in low-resolution images are defined by the relation *part-of*.

Part-of (Road_segment, Road)

There are some other relations which describe the relationships between a road and the road network. They will be introduced in Section 6.4.

6.2.3.2 Road Structure and Properties

The structure of road surface generated in *intermediate-level processing* is described by a generalised antiparallel pair with a number of geometric and photometric attributes. This can be represented by the following *structure*:

feature (feature_no, L1, L2, attribute(end_directions, length,
width, gradient, average_intensity, standard_deviation)).

Feature_no is the number of the feature being represented. L1 and L2 are two point chains which record positions of points belonging to the two boundaries of the road. *attribute* is also a *structure* which contains the geometric and photometric properties of the road, including its width, gradient orientation, average intensity value and the standard deviation of intensity values of the road surface.

A road has specific geometric and photometric properties in the image which distinguish it from other objects. These properties can be formulated into a rule based on the derived road structure *feature*. The rule has the following form (Trinder and Wang, 1997b, 1998c):

Rule one: *road_segment* (X) :-
feature (X, _, _, attribute(length, width, gradient, average_intensity,
standard_deviation),
length > $\mathbf{T}_l$,
width > $\mathbf{W}_d$ - $\mathbf{T}_d$,
width < $\mathbf{W}_d$ + $\mathbf{T}_d$,
gradient is -1,

average_intensity > $\mathbf{G_0}$ - $\mathbf{T_i}$,

average_intensity < $\mathbf{G_0}$ + $\mathbf{T_i}$,

standard_deviation > $\mathbf{R_0}$ - $\mathbf{T_R}$,

standard_deviation < $\mathbf{R_0}$ + $\mathbf{T_R}$.

$\mathbf{W_d}$, $\mathbf{G_0}$ and $\mathbf{R_0}$ are road width, average intensity value and RMS of intensity values of the road surface respectively. Road width $\mathbf{W_d}$ has different values for different types of roads and can be found in the manuals of road design. For example, it has a value of 6.8 m for a two-lane rural road. $\mathbf{G_0}$ and $\mathbf{R_0}$ are dependent on not only the material of road surface, but also many factors, such as the time and season of photography, weather conditions, condition of photographic processing, etc. It is impossible to define them for all images. Instead, they are determined by a training process in which a large number of samples of road surfaces are measured. The values obtained through a training process can be applied to images taken in similar conditions under which the training images are taken. $\mathbf{T_l}$, $\mathbf{T_d}$, $\mathbf{T_i}$ and $\mathbf{T_R}$ are the thresholds for length of road segment, its width, average intensity value and RMS of intensity values. In the ideal case, $\mathbf{T_l}$ should be equal to or larger than the dimension of the image in image space when it is assumed that a road passes through the image from its two opposite sides. However, the road segment may be occluded in the image. In this case, a value less than the image dimension is more reasonable. Assuming the occlusion occurs in the middle of the image, a value of (L - 80)/2 is chosen in image space when the image dimension is L and 80 is the maximum distance within which two separated antipairs can be grouped. $\mathbf{T_d}$ is the tolerance of variation of road width along the road. It is determined by the computation accuracy of road width and is set as one pixel in image space. The thresholds $\mathbf{T_i}$ and $\mathbf{T_R}$ are determined together with $\mathbf{G_0}$ and $\mathbf{R_0}$ in the training process. The gradient orientation of a road is defined as -1 since road surfaces usually have higher reflectance than their background.

Some other rules which represent the characteristics of a road network and relations between roads and the road network will be described in Section 6.4.

6.3 Control Strategy

Control strategy determines the execution order of processing in object recognition. As discussed in Section 3.3, pure bottom-up or top-down control cannot yield reliable recognition of objects from remotely sensed imagery in most cases. This is because remotely sensed images have complex structures and information derived from them is always incomplete due to the existence of image noise, occlusions, limitation of image processing algorithms, etc. In order to achieve more reliable object recognition, a hybrid control strategy is usually selected. This is the case for this study.

The recognition of roads from aerial images includes: hypothesis generation of road segments in high-resolution images; verification of hypothesised road segments, and prediction and finding of missing road segments. From the semantic road model defined in Section 6.1, a road segment in high-resolution images is described by an elongated area with a number of specific geometric and photometric properties. Thus, road segments can be hypothesised in a bottom-up process because of its simple structure. As only structural information of the road surface and its geometric and photometric properties are used in hypothesis generation, non-road objects with similar structure and properties may also be hypothesised. Thus a verification process is required to remove the spurious hypotheses of road segments. This is done based on the *part-of* relationships between roads in high-resolution images and roads in low-resolution images. The missing road segments are predicted by using the spatial relationships between road segments. Once a missing road segment is inferred, a top-down process is invoked to find its instance in the image and to detect occlusions between road segments. This will be discussed in the following section.

6.4 Recognition of Roads from Aerial Images

6.4.1 Hypothesis Generation of Road Segments

Hypotheses of road segments are generated in high-resolution images in a bottom-up process. It starts with edge detection, tracking and linking, and noise removal. Structure of the road surface is generated at the *intermediate-level processing* based on the extracted edge information and knowledge of the structure of the road surface, and is represented by the structure *feature*. Road segments are finally hypothesised by applying *rule one* to the generated *features.*

Two examples of hypothesis generation of road segments from high-resolution images are shown in Figures 6.1 and 6.2. In Figure 6.1(a), the image contains a straight segment of a two-way highway and part of a track which intersects the highway. Five antipairs are generated based on the extracted edge image, three of which correspond to the highway and one corresponding to the track (Figure 6.1(b)). The three antipairs belonging to the highway are separated at the junction and the middle of the road segment. They are aggregated based on the similarity of their geometric and photometric properties in grouping, and three road-like features are formed with one corresponding the road segment, one to part of the track and one to the small part of the road segment (Figure 6.1(c)). Finally, one corresponding to the road segment is recognised as a road by applying the *rule one* (Figure 6.1(d)). The other two are not hypothesised as they are too short. Figure 6.2 shows another example in which the road to be extracted is a curved segment and is occluded by a tree at the lower right, and in the upper right part the boundary line is blurred because of the poor contrast between the road surface and its background. Thus, the generated antipairs are separated at these positions (Figure 6.2(b)). They are aggregated in the grouping process and 16 road-like features are formed (Figure 6.2(c)). However, most of them correspond to non-road objects. Finally, the one corresponding to the highway with maximum length is hypothesised as a road using the *rule one* (Figure 6.2(d)).

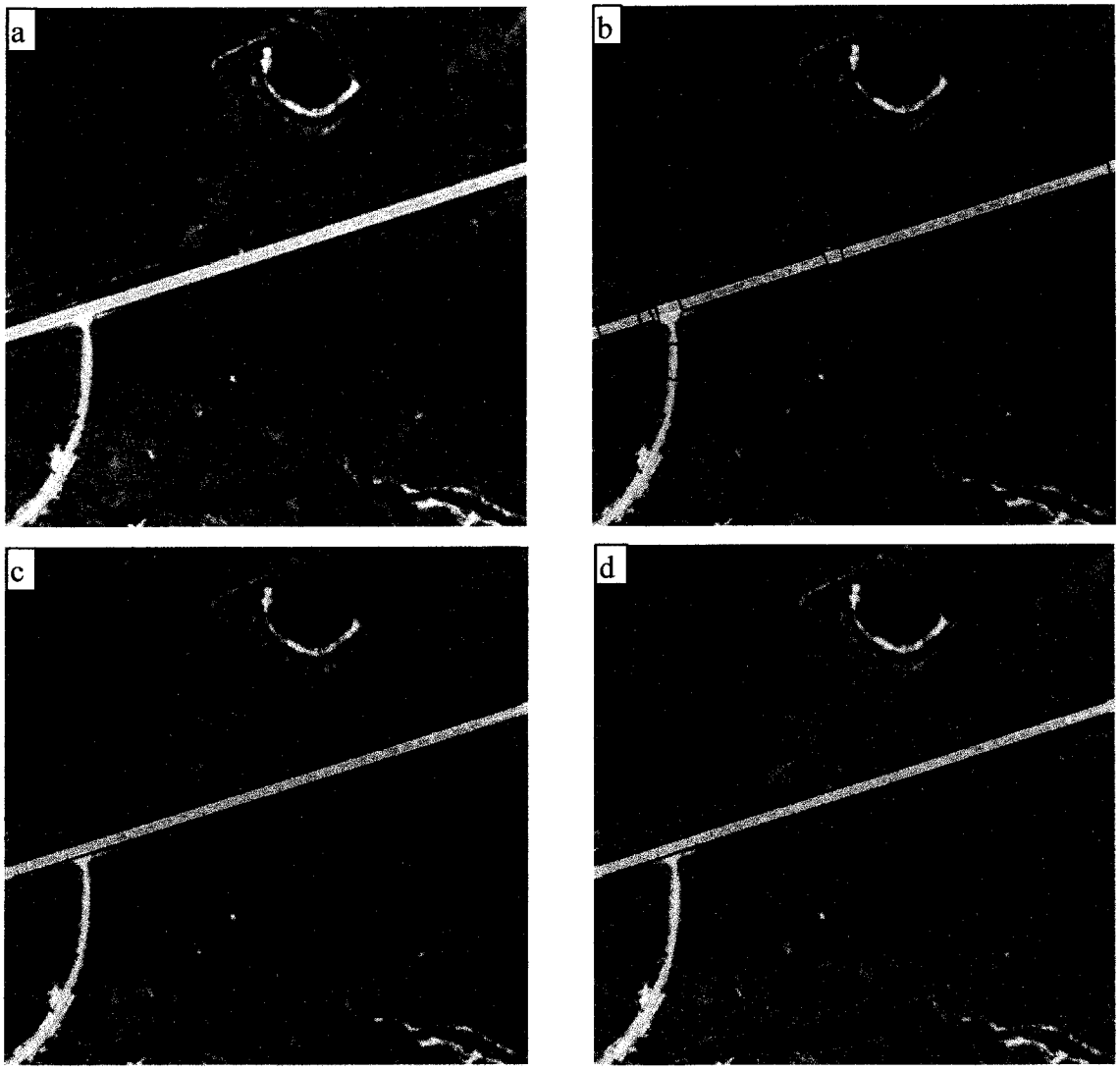


Figure 6.1 Hypothesis Generation of Road Segments from High-resolution Images
(a) Original image (b) Generated antipairs (c) Road-like features after grouping
(d) Hypothesised road segment

6.4.2 Hypothesis Verification

Once road segments are hypothesised in high-resolution images, they are verified using *part-of* relationships between roads in high-resolution images and roads in low-resolution images. The verification of hypothesised road segments includes extraction of the road network from low-resolution images, generation of *part-of* relationships and hypothesis verification.

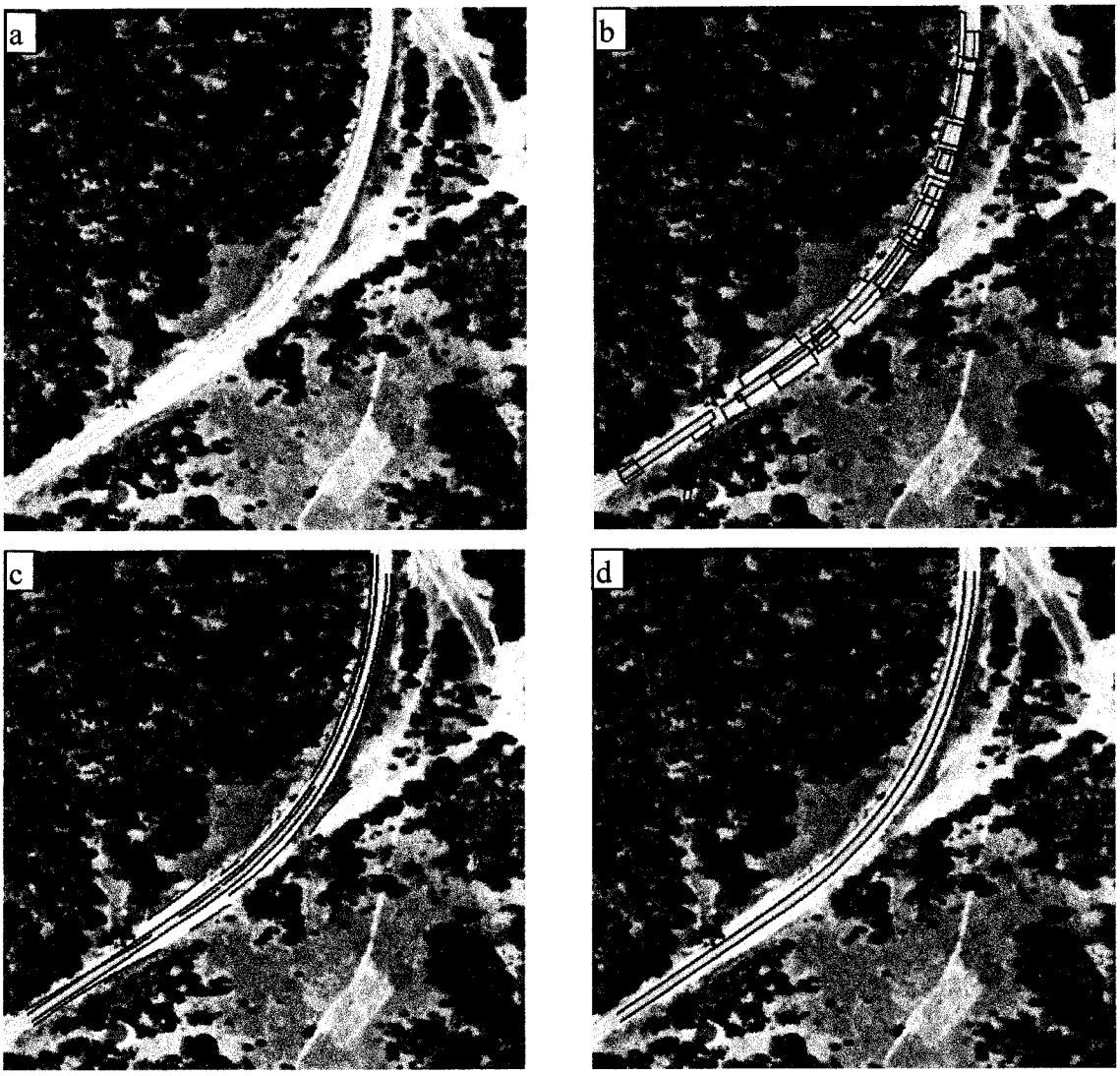


Figure 6.2 Hypothesis Generation of Road Segments from High-resolution Images
(a) Original image (b) Generated antipairs (c) Road-like features after grouping
(d) Hypothesised road segment

6.4.2.1 Generation of Topology of Road Network

Existing maps and databases in geographical information systems (GIS) possess relevant information of road networks which can be derived through digitisation or data query. However, the topology of road networks generated from these data sources cannot reflect the current status of road networks, due to the poor temporal accuracy of the existing maps and data stored in GIS. Remotely sensed images can supply users with

real-time or near real-time data of objects on the ground and therefore, they can be used to derive the most recent information of road networks. In this study, low-resolution images are used to generate the topology of road networks as they can show the global structures of the networks. Another advantage of using low-resolution aerial images is that the gaps caused by obstructions, such as trees, shadows, image noises, etc., are usually limited to a few pixels. Therefore, they can be properly bridged during line linking and grouping. The generation of topology of road networks includes detection of line and point features, grouping of line segments and representation of topology of road networks.

A road is a line feature in low-resolution aerial images. The width of the line depends on the scale of the image. A road has a width of one to three pixels in the image with a ground resolution of about 10 meters. It can be extracted using a line detector such as a morphological operator (Dougherty, 1992) which has been used in this study. The methods for preprocessing introduced in Chapter 4 are used for line tracking and linking, generation of smooth line segments and removal of non-road lines. Similarity-grouping method introduced in Chapter 5 is used to aggregate line segments to form elongated line segments. Since lines in low-resolution images do not offer surface properties, only their gradient magnitudes and orientations which can be computed in a similar way to that introduced in Section 5.3 are used as a similarity measure.

A road intersection in low-resolution images appears as a point feature which can be detected by a point operator such as the SE operator (Heitger, 1995). Because of image noise and the limitation of the point operator, not all of the detected corner points correspond to road intersections and some true intersections may not be extracted. Therefore, additional processing is required to remove spurious intersections and to add true intersections which are not extracted. An intersection is a point at which two or more line segments intersect at certain angles. Thus points will be removed if there is only one line segment around or passing through a given area. When two line segments intersect at a point with an angle within the threshold and the intersection is not detected, a new intersection is then inserted.

Due to the limitations of line operators, parts of roads at an intersection may not be extracted. In this case, the extracted line segments are extended to connect at the intersections. Line segments which are not connected to any point or line segment, i.e. isolated line segments, are removed as spurious roads. Finally, a network consisting of line segments and intersections is formed.

6.4.2.2 Representation of Topology of Road Network

A road network in low-resolution images consists of lines and points. The topology of the road network defines the connectivity relationships between these elements, which can be represented by a graph composed of edges and vertices. A graph is usually expressed as (Chachra et al, 1979):

$$\begin{aligned} &G = G(V, E), \\ &V = (v_i), \text{ with } i = 1, 2, \ldots m, \text{ and} \\ &E = (e_j), \text{ with } j = 1, 2, \ldots n. \end{aligned} \tag{6.1}$$

where G, E and V stand for the graph, vertices and edges respectively, and m and n are the numbers of the vertices and edges in the graph. Each edge in the graph is defined by two adjacent vertices, i.e. $e = (v_p, v_q)$. Such a definition is sufficient for a straight line segment. However, a road may be a curvilinear segment. It is insufficient to represent it with only the positions of its end points. Therefore, to describe a road correctly a third element is required, which represents the positions of points on the road. This can be expressed by a *structure* in Prolog (Trinder and Wang, 1998c), i.e.

edge (Edge_no, A, B, S, L)

where Edge_no is the number of the edge, A, B are two end points of the edge, S is a point chain which represents the positions of all points belonging to the edge, and L is the length of the edge. Any two intersecting edges in the network share one common point or vertex. This can be formulated as the following *relations* in Prolog:

adjacent (edge(X, A, B1, _, _), edge(Y, A, B2, _, _)), or

adjacent (edge(X, A1, B, _, _), edge(Y, A2, B, _, _)).

With the defined structure *edge* and relation *adjacent,* the topology of the road network is represented by a number of edges and connectivity relationships.

6.4.2.3 Generation of Relationships between Roads

Having generated the topology of a road network, the relationships between the hypothesized road segments in high-resolution images and the road network derived from low-resolution images can be generated. Before the relationships are established, hypothesised road segments need to be transformed to low-resolution images. When low-resolution images are generated from high-resolution images by resampling, they can be registered by a simple transformation. After transformation, a hypothesised road segment becomes a line segment which is represented by the *structure*:

h (Hypothesis_no, P1, P2, T)

where Hypothesis_no is the number of a hypothesised road segment, P1 and P2 are its two end points, and T is the length of the road segment in low-resolution images.

The relationships between transformed hypothesised road segments and edges in the road network include *part-of, relate-to and isolated* which are listed in the table 6.1.

Table 6.1 Relationships between Hypothesised Road Segments and Road Network

Relation No.	Relation Name	Relation Description
1	*part-of*	hypothesised road segment is part of an edge
2	*relate-to*	hypothesised road segment is connected to an edge
3	*isolated*	hypothesised road segment is not connected to the road network

A hypothesised road segment h is part of an edge if it lies within the buffer zone of the edge which is formed by extending the edge in the normal directions of the edge on both its sides. The area of the buffer zone is defined by the parameter *offset* (Figure 6.3). The value of *offset* is determined by the accuracy of line extraction and transformation between low- and high-resolution images. This relationship is described by the *relation* (Wang and Trinder, 1998):

pat-of (h(Hypothesis_no, P1, P2, T), edge(Edge_no, A, B, S, L))

A transformed hypothesised road segment may not belong to any edge in the road network, but connects to an edge. Such a relationship is defined by *connect-to* which is expressed as (Wang and Trinder, 1998):

relate-to (h(Hypothesis_no, P1, P2, T), edge(Edge_no, A, B, S, L), ALPHA)

ALPHA is the intersecting angle between the transformed hypothesised road segment and the edge.

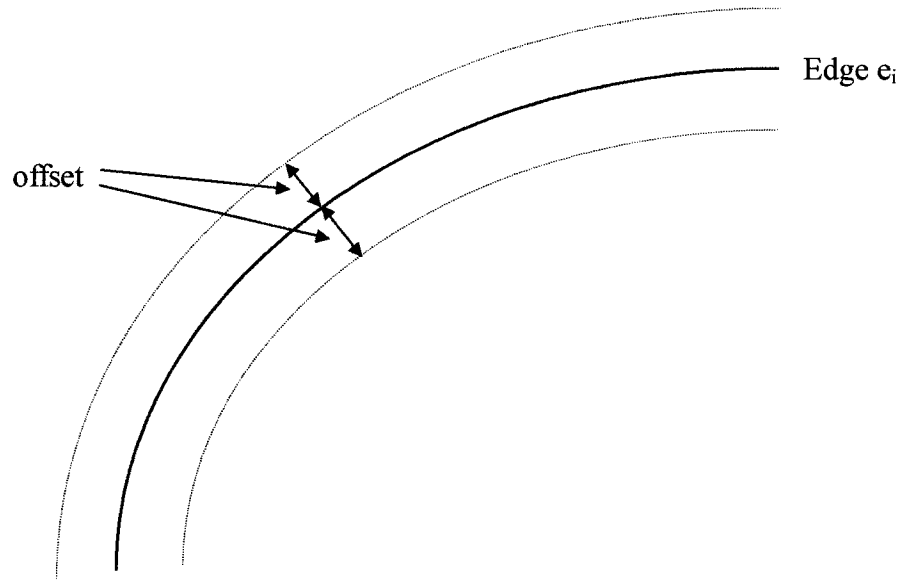


Figure 6.3 Definition of Buffer Zone of Edge

A transformed hypothesised road segment may not belong to any edge in the network, nor connect to any edge. It is called isolated road segment which is represented by the structure *isolated*:

isolated (h(Hypothesis_no, P1, P2, T))

6.4.2.4 Hypothesis Verification

With the defined relationships as above, a number of *rules* can be formed based on the characteristics of road networks, which will be used for verification of hypotheses. One important characteristic of road networks is that a road rarely exists individually, but connects to other roads. Therefore, isolated roads can be immediately treated as false hypotheses and rejected immediately. This can be formulated as the following *rule* (Wang and Trinder, 1998):

Rule two: *non-road* (X) :- *isolated* (h(X, _, _, _)).

When a hypothesised road segment is related to an edge in the road network derived from low-resolution images by the *part-of* relationship, it cannot be confirmed that the hypothesised road segment is a true road segment since the corresponding edge in the road network may not be a road. To verify whether a hypothesised road segment is a true road segment, it is necessary to check other parts of the edge. It is generally believed that the most parts of a road should be hypothesised in high-resolution images based on the fact that a road is usually partially occluded by trees, shadows, etc. For each edge in road networks derived from low-resolution images, the hypothesised road segments belonging to it can be determined by the following *query*:

?- *part-of* (h(X, _, _, _), edge(Edge_no, _, _, _, _)).

Once the hypothesised road segments belonging to an edge are determined, the ratio of their total length to the length of the edge can be computed. To relate a hypothesised

road segment with the computed ratio R, a new structure, *element* which has the form of *element* (h(Hypothesis_no, P1, P2, T), edge(Edge_no, A, B, S, L), R), is constructed. Therefore, if the ratio is larger than a predefined value, the hypotheses of road segments belong to the edge are accepted. This can be expressed as the *rule*:

Rule three: *road* (X) :- *element* (h(X, _, _, _), edge(_, _, _, _, _), R),
$R > T_r$.

T_r is the threshold for the ratio R. It should be determined based on the analysis of possible occlusions occurring on roads in different situations. However, it is generally assumed that most parts of a road are not occluded in most cases, and they can be detected in high-resolution images. Therefore, a value larger than or equal to 0.5 is usually given.

A road segment may be hypothesised in high-resolution images, but not detected in low-resolution images due to the effects of image noise, limitation of line operator, etc. Thus, for hypothesised road segments related to the road network by the relate-to relationship, they are accepted if their length is larger than a threshold and intersecting angle is within a defined range. This can be described by the *rule*:

Rule four *road* (X) :- *relate-to* (h(X, _, _, T), edge(_, _, _, _, _), ALPHA),
$T > T_0$,
$ALPHA > T_a$,
$ALPHA < T_b$.

Where T_0 is the threshold for length which is determined based on the analysis of road length in various situations while T_a and T_b define the range of intersecting angle which can be found in the manual of road construction.

After hypothesised road segments are verified, they are arranged sequentially according to their spatial positions. Every two hypothesised road segments are considered as

neighbours when they belong to the same road in the network and there is no other road segments between them. Each hypothesised road segment may have one or two neighbours, and a neighbour relationship is described as:

neighbour (h(Hypothesis_no1, _, _, _), h(Hypothesis_no2, _, _, _), D)

Where D is the distance between two neighbouring hypothesised road segments.

Ideally, hypothesised road segments should join together to form a continuous road, but they may be disconnected due to the effects of occlusions caused by trees and shadows, poor contrast between the road surface and its background, etc. Short road segments between two neighbouring road segments may not be hypothesised in this case. It is assumed that a missing road segment between two hypothesised road segments may be missed if their distance is larger than a given value. This can be formulated as a *rule*:

Rule five: *missing_part* (X, Y) :- *neighbour* (h(X, _, _, _), h(Y, _, _, _), D),
D > d_0.

D is the distance between two road segments X and Y, and d_0 is a constant. When a road segment is predicted between the road segments X and Y, a procedure will be started to find antipairs between X and Y in the high-resolution images. Once an antipair is found, a rule which is the modification of the *rule one*, i.e. *road_segment*, will be applied to it to generate the recognition of the missing road segment. The modified rule has the form:

Rule six: *road_part* (X) :-
feature(X, _, _, attribute(_, width, gradient, average_intensity,
standard_deviation),
width > $\mathbf{W_d}$ - $\mathbf{T_d}$,
width < $\mathbf{W_d}$ + $\mathbf{T_d}$,
gradient is -1,
average_intensity > $\mathbf{G_0}$ - $\mathbf{T_i}$,

$$\text{average\_intensity} < \mathbf{G_0} + \mathbf{T_i},$$
$$\text{standard\_deviation} > \mathbf{R_0} - \mathbf{T_R},$$
$$\text{standard\_deviation} < \mathbf{R_0} + \mathbf{T_R}.$$

As this rule is used to find missing short road segments, the length is set as anonymous in the rule.

Figure 6.4 gives an example of hypothesis verification. The image shown in Figure 6.4(a) is a low-resolution aerial photograph with a ground resolution of about 12 m which is generated from the original image by resampling. In the area covered by the image, there are two highways which intersect with each other. After generation of topology of the road network, three edges and one vertex are produced as shown in Figure 6.4(b). In the high-resolution image, a total of 23 road segments are hypothesized. After they are projected onto the low-resolution image, 22 out of 23 are related to the three edges by the relation *part-of*, among which six belong to edge1, ten belong to edge2 and seven correspond to edge3 (Figure 6.4(c)). These 22 road segments are accepted by applying the rule *road* to them, and one is rejected as it neither belongs to any edge, nor connects to any edge. Twenty two short road segments are found during the verification process, which are shown in Figure 6.4(d).

6.4.3 Occlusion Detection

Occlusions are a common phenomenon in images. They not only break a road into several segments, but also cover some road parts which cannot be recognised in the recognition process and found in detection of missing road segments. To extract a complete road network, road segments in the occluded areas should be detected. From the knowledge of roads, occlusions on the road surface include trees, shadows, and poor contrast, assuming that the effects of image noise, vehicles on the road, etc. are removed during the process of grouping. Trees and shadows usually change the intensity values of

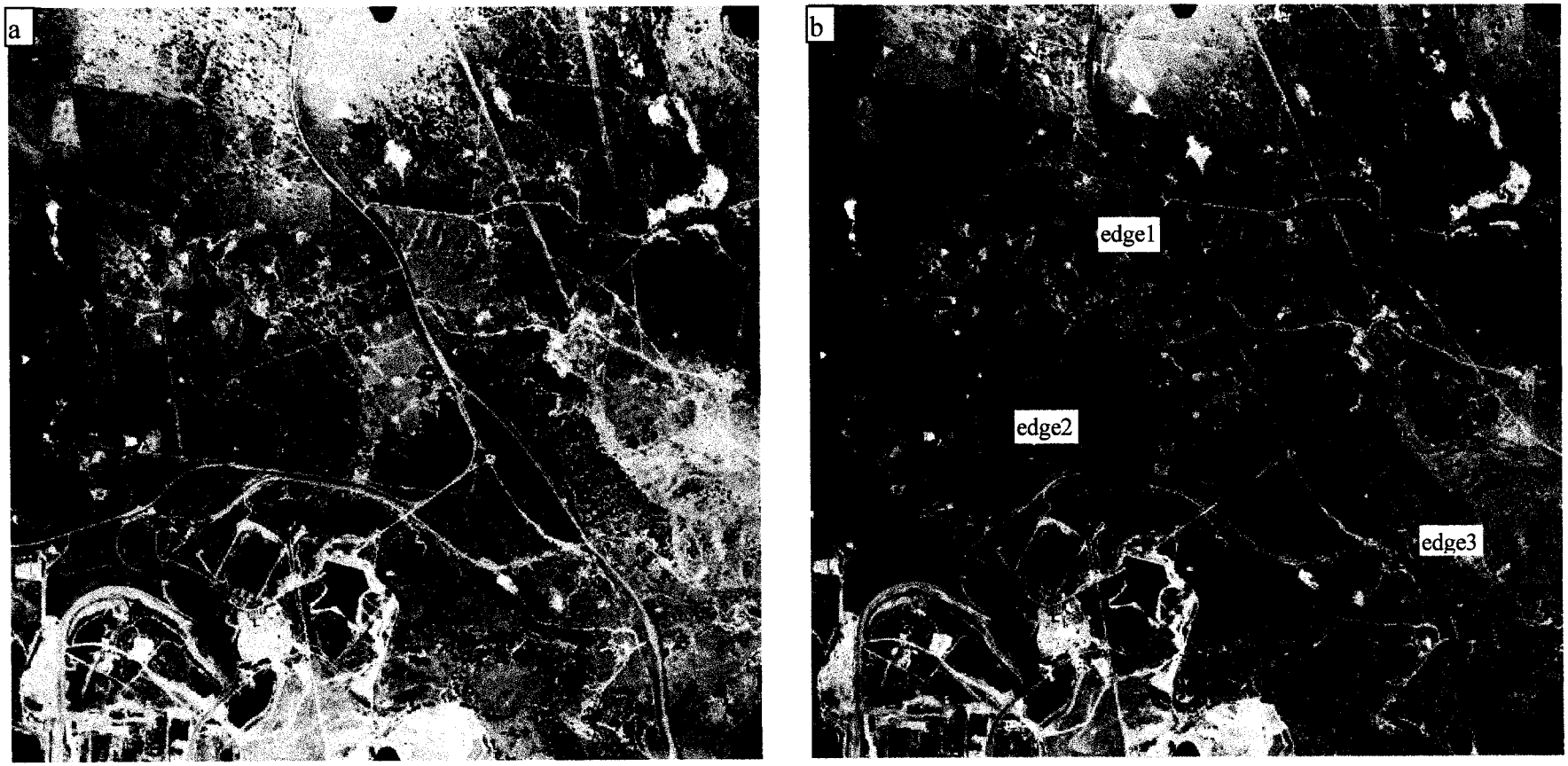


Figure 6.4 Verification of Hypotheses (a) Original low-resolution image (b) Generated road network from low-resolution image

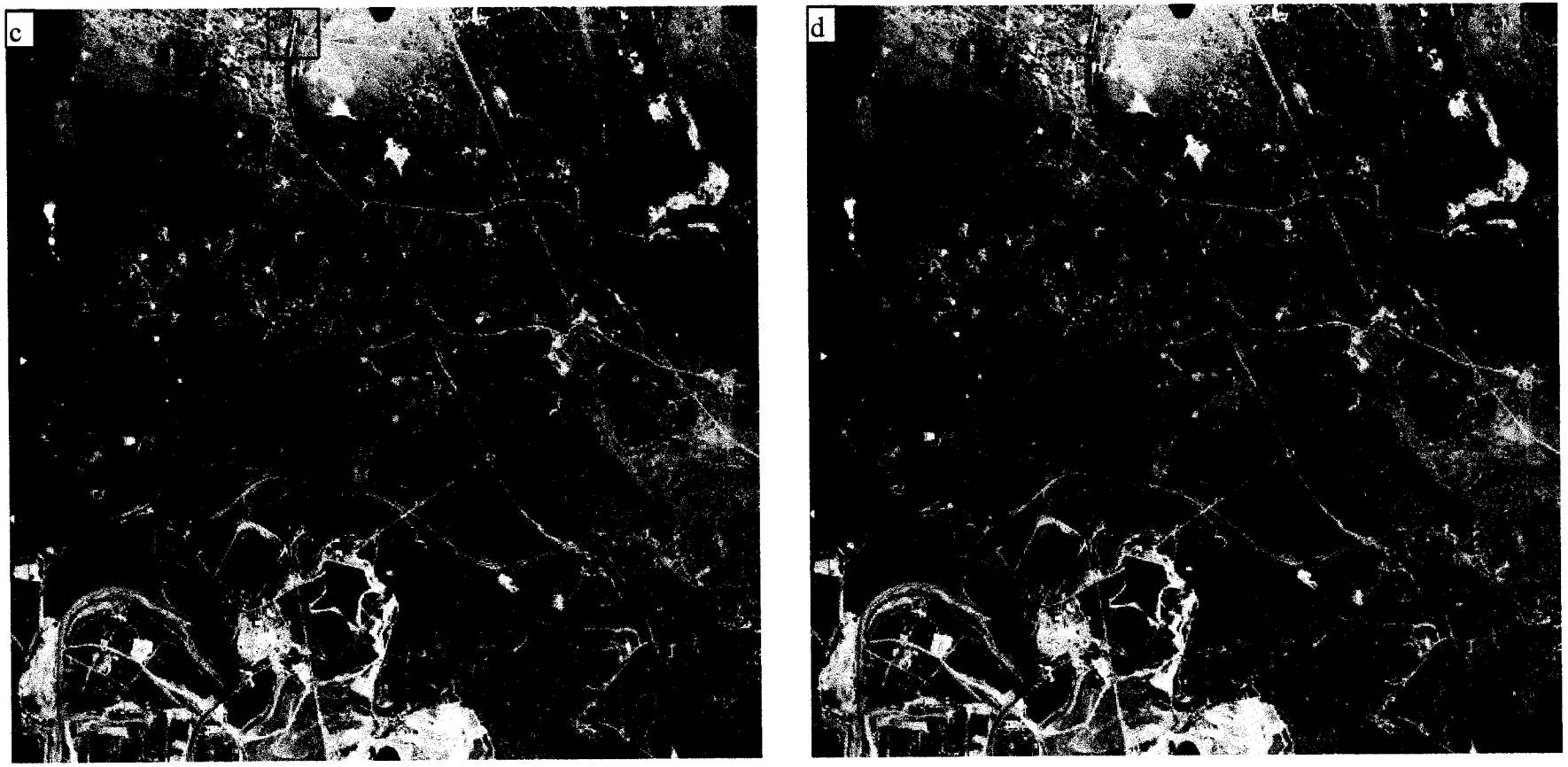


Figure 6.4 Verification of Hypotheses (continued) (c) Hypothesised road segments after transformation (d) Verified road segments

the occluded area of a road and cause road boundaries to be highly fragmented or to disappear while the intensity values of the surface do not change in case of poor contrast between road surface and its background. Therefore, occlusions can be classified using their spectral characteristics. There are a number of classification methods in remote sensing. They use the spectral properties of objects and pixels in the image for classification of pixels. Although they cannot provide accurate classification of pixels in aerial images due to the effects of image noise and object texture, they could be utilized to give a reliable classification of an area which has homogeneous properties (Abkar and Mulder, 1998). In this study, a maximum likelihood classification method is used to detect occlusions between two road segments based on the spectral values in three visible bands, i.e. blue, green and red.

Maximum likelihood classification is a supervised classification method which is based on the statistical data of the spectral properties of training data and the spectral data of pixels to be classified. Suppose there are m classes to be distinguished which are represented by:

$$c_i,\ i = 1, 2, \ldots m$$

To determine the class to which a pixel at position $\mathbf{x}$ belongs, it is necessary to calculate its conditional probabilities

$$p(c_i \mid \mathbf{x}),\ i = 1, 2, \ldots m$$

The position vector $\mathbf{x}$ is a column vector of spectral values of the pixel. It defines the position of a pixel in multi-spectral space. In this study, it consists of the spectral values of a pixel in blue, green and red bands. The probability $p(c_i \mid \mathbf{x})$ describes the likelihood that a pixel $\mathbf{x}$ belongs to the class c_i. The maximum likelihood classification classifies a pixel by (Richards, 1993):

$$\mathbf{x} \in c_i \quad \text{if} \quad p(c_i \mid \mathbf{x}) > p(c_j \mid \mathbf{x}) \quad \text{for all} \quad j \neq i \tag{6.2}$$

According to Bayes' theorem, $p(c_i \mid \mathbf{x})$ can be calculated by:

$$p(c_i \mid \mathbf{x}) = p(\mathbf{x} \mid c_i)\, p(c_i)/p(\mathbf{x}) \tag{6.3}$$

where $p(c_i)$ is the probability that the class c_i occurs in the image, which can be estimated from the training data. $p(\mathbf{x} \mid c_i)$ is the probability of finding the pixel $\mathbf{x}$ in class c_i which can be calculated using the distribution function of the class estimated from the training data (Richards, 1993). $p(\mathbf{x})$ is the probability that a pixel appears in the image which is the sum of probabilities that the pixel belongs to every class.

After classification of occlusions, the relationships between the detected occlusions and the verified road segments can be used to infer the existence of a road segment between the verified road segments. If the occlusion is classified as road surface, shadow or trees, it can be inferred that a road segment probably is missed. The disjoined road segments are then connected by spline interpolation. If it is classified as other objects, more contextual information is needed to infer the type of the object in the occluded area, which needs further study.

An example of occlusion detection is shown in Figure 6.5 in which the hypothesised road segments are disconnected due to the poor contrast between the road surface and its background. To determine what the missing part is, a procedure for finding short road segments is first triggered. As the contrast between the road surface and its background is very low, there is no edge information extracted in this area. Therefore, no antipair is found. Then a procedure for detecting occlusions is started, in which the part between two hypothesised road segments is classified as road which corresponds to the shaded part in Figure 6.5(b) using the maximum likelihood classification. In the training stage, four images are used and five classes, i.e. trees, grass, road surface, bare soil and tree shadows are selected. The intensity values of these objects in blue, green and red bands are measured manually in the selected images, and their distributions in the spectral space

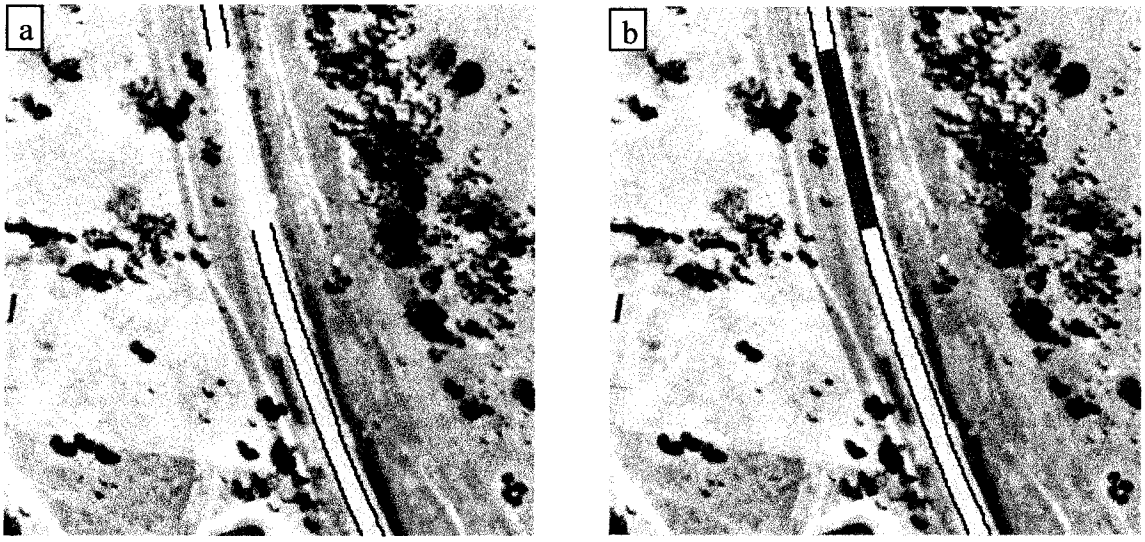


Figure 6.5 Occlusion Detection (a) Image with a road segment occluded (b) Detected occlusion between two verified road segments

are determined. With their distributions in the spectral space, the probabilities of the occluded part belonging to these objects are calculated and a classification is carried out based on the maximum probability.

6.5 Summary of Chapter

This chapter has described a knowledge-based method for road recognition from aerial images based on a multi-resolution semantic road model, in which a road network is defined in both high- and low-resolution images and a number of relations are defined to relate roads. Based on the developed road model, a bottom-up procedure has been established to recognise roads, in which road structures are generated using features and their attributes extracted from high-resolution images, and the knowledge of roads is then applied to the generated road structures to produce hypotheses of road segments. Hypothesised road segments in high-resolution images are verified using the *part-of* relationships between roads in high-resolution images and roads in low-resolution images. A number of rules have been defined based on the properties of a road and relations between roads in a road network for the recognition and verification of roads. As short road segments are not hhypothesised in the process of hypothesis generation, a

procedure for predicting and finding missing road segments is included in the verification process. Detection of occlusions is a difficult problem in road recognition. A method for locating occlusions based on the maximum-likelihood classification method has been developed. The relationships between the classified objects and the corresponding recognised road segments are used to infer what the missing part between two verified road segments could be. Some examples are included in this chapter, which show the recognition results using the developed methods. To demonstrate the applicability of the methods, more results will be given and discussed in the next chapter.

Chapter Seven

Experiments and Results

7.1 Image Data

Chapter 6 has described the process and methods of road recognition from aerial images. To test these methods, a number of experiments were made on aerial images, the results of which will be presented in this chapter. The tests in hypothesis generation, verification of hypotheses, and detection of missing road segments and occlusions will be analysed and evaluated. Also, the thresholds used in the rules for recognition of roads will be discussed. Three aerial images in the Hunter Valley and near Dungog in rural New South Wales have been used. This section will introduce some characteristics of the image data, including the size and scale of image, terrain types, road types, road structures, etc.

7.1.1 Hunter Valley Image Data

The Hunter Valley images are located in a rural area in the east of New South Wales, which have moderate undulations in elevation between 170 m and 300 m. There are two motorways passing through the area, which intersect as shown in Figure 7.1(a). The motorway on the left is a two-way two-lane rural road. The road on the right has different numbers of lanes in different parts. It is a two-lane road in the upper part of the left image, but it changes into three lanes in the middle of the image while it becomes four lanes wide in the bottom part. Roads in the images are paved with bitumen and their edges are painted with white lines which serve as the boundaries of roads during road recognition. The images also contain some other minor roads, one of which crosses the right motor way. The images have a scale of 1:25,000 and are scanned in colour and black and white with pixel sizes of 20 μm and 30 μm respectively, yielding ground resolutions of 0.5 m and 0.75 m. The black and white images have a size of 7466×7456

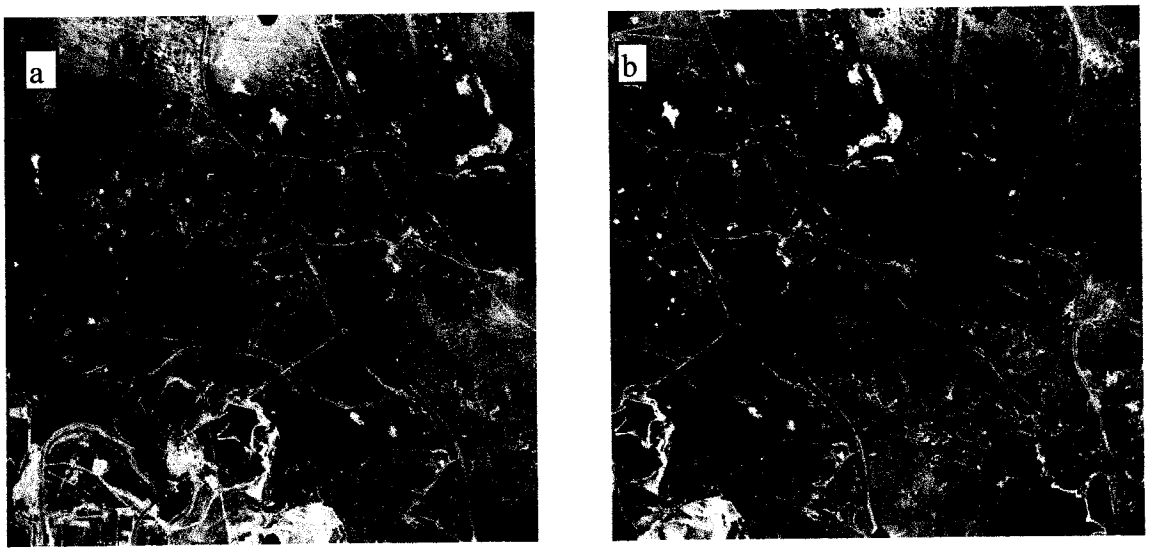


Figure 7.1 Hunter Valley Image Data (a) Left image (b) Right image

pixels. Low-resolution images have been generated by resampling the original images at a reduction factor of 15, yielding a ground resolution of about 11 m. For simplicity of processing, the original image is divided into 225 smaller sub-images, each having a size of 500×500 pixels. As can be seen in Figure 7.1, only a small portion of the original image contains linear objects. Thus, only a small part of these images is utilised in the test.

In order to test the developed methods, images containing roads have been used. Figure 7.2 shows some examples of test images in which roads have different shapes, widths, amounts of occlusion and disturbances. In Figure 7.2(a), the image contains a two-way rural road which intersects a track to its left. The road has well-defined boundary lines and distinguishable contrast against its background. Figure 7.2(b) shows a curved rural road with two lanes. The image contains extensive trees which cover the road partly on the lower left, and the boundary line on the upper right is blurred, due to the bare soil on both sides of the road. Figure 7.2(c) gives an example of roads with blurred boundary lines. Although there is a significant contrast between the road surface and its background, the boundary lines are hardly recognizable. Figure 7.2(d) shows a road with three lanes, which is partially occluded by trees and cars. In Figures 7.2(e) and (f), a three lane rural road with an overpass and a road with four-lanes partially occluded by

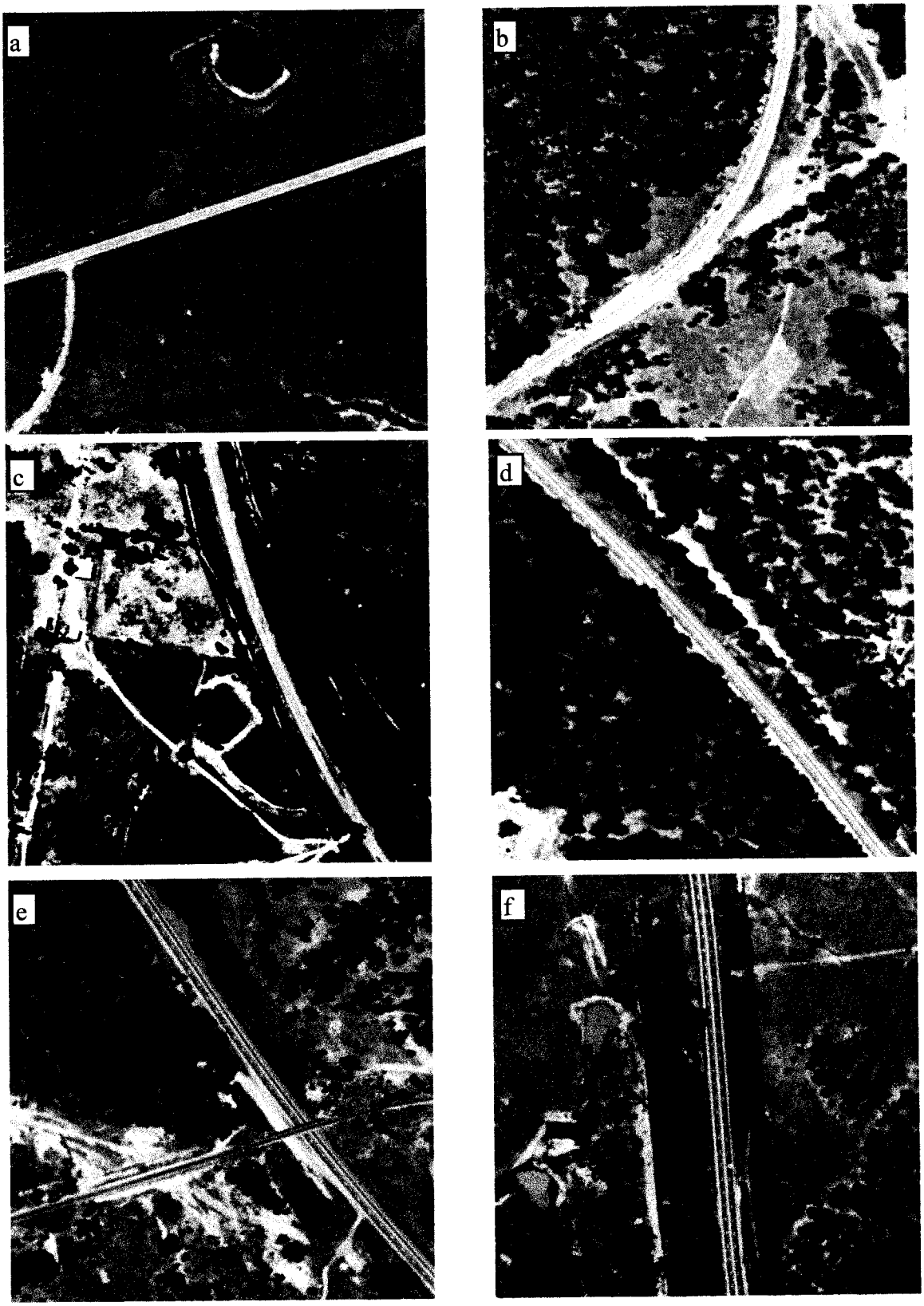


Figure 7.2 Test Images at Different Situations

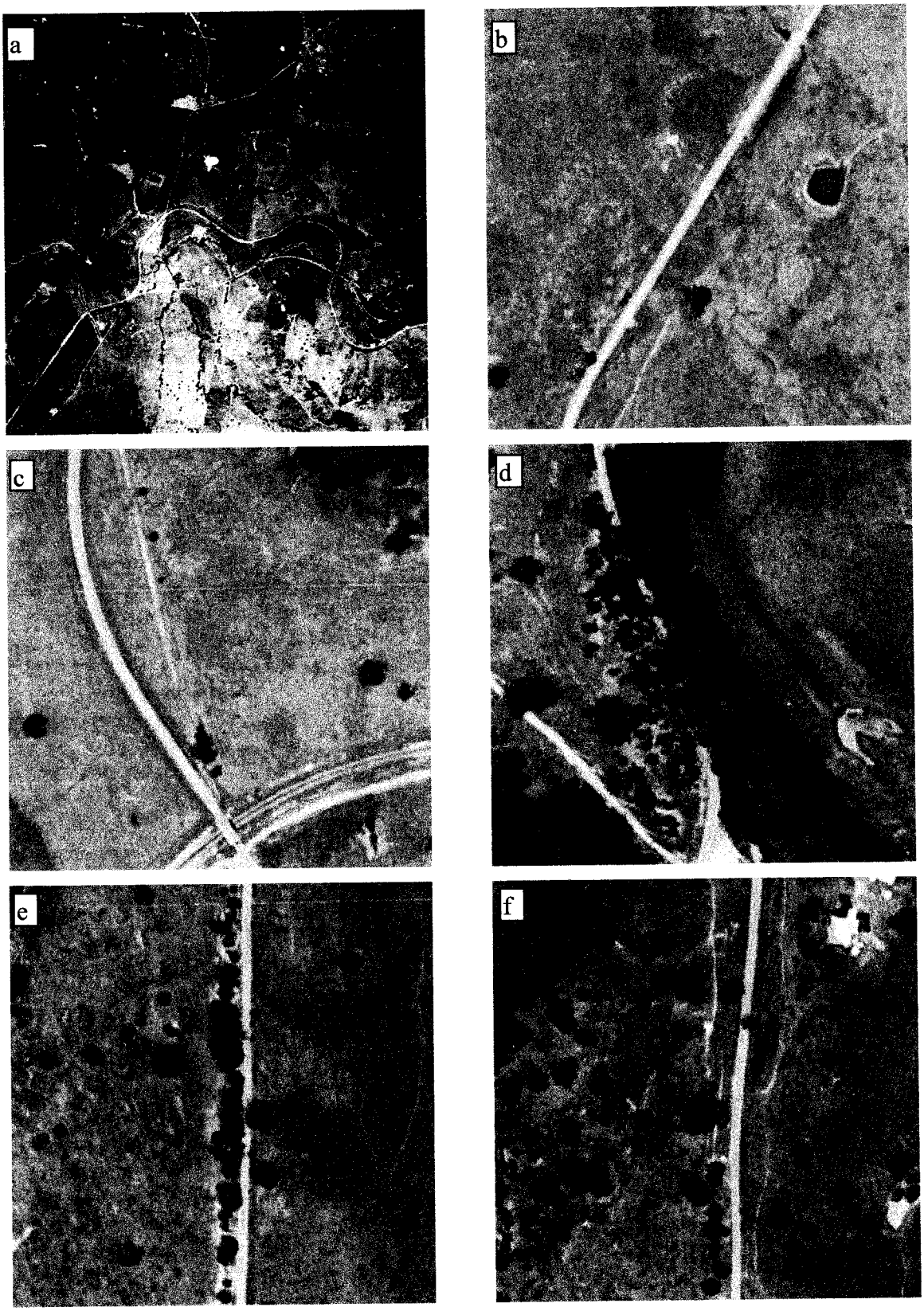


Figure 7.3 Dungog Image Data (a) Original image (b) - (f) Test images with roads

trees and cars are represented respectively. The boundary lines of roads in these two images are well-defined.

7.1.2 Dungog Image Data

The Dungog image is also located in a rural area in the east of New South Wales, and contains two rural roads and several minor roads. The rural roads are paved, but their boundaries are not painted with white lines. Thus, the natural boundaries of the roads will be used in the recognition process. The image has a scale of 1:25,000 and is scanned with a pixel size of 20 μm, producing a ground resolution of 0.5 m. The original image has a size of 10600×10600 pixels and is divided into 441 smaller sub-images with a size of 500×500 pixels. The low-resolution image is generated by resampling the original image at a reduction factor of 16, yielding a ground resolution of 8 m, and is shown in Figure 7.3(a). Figures 7.3(b) to 7.3(f) show some examples of test images used in this study. Figures 7.3(b) and 7.3(c) contain a smooth rural road which has well-defined boundaries and few occlusions and disturbances. A road passing through a forest is shown in Figure 7.3(d). In this image, most of the road is covered by trees and the lower part of the road has poor contrast. Only a small part of the road on the upper left has recognisable boundaries. Figure 7.3(e) shows a road occluded by trees on both its boundaries. Main rural roads usually have constant width, however, a simple rural road may have changes in width as shown in Figure 7.3(f).

7.2 Parameters Used in Recognition

In the recognition process of roads, a number of parameters are used in the rules for recognition, which include both geometric and photometric properties of roads. They need to be determined before recognition can be performed. The geometric parameters included in the recognition rules include length and width. The first parameter was given in Chapter 6. The road width depends on the type of the road, design speed and vehicle capacity and thus it is a function of the road type and design standards. Table 7.1 lists the widths of four different types of roads in NSW (Lay, 1990).

Table 7.1 Road Width for Different Types of Roads

Road Type	Minimum Road Width (m)	Maximum Road Width (m)
narrow two-lane road	4.6	6.4
wide two-lane road	6.5	9.1
three-lane road	9.2	11.6
four-lane road	11.6	14.8

The photometric properties of a road contain the average intensity value of the road surface G_r and a measure of the homogeneity of the road surface H_r which are represented by the mean and standard deviation of intensity values of the road surface. As they may vary along the road, the extent of their variations also need to be determined. Therefore, the standard deviation of intensity values along the road σ_G, and the standard deviation of homogeneity σ_H are calculated. These parameters are determined by measuring the grey values along the road surface interactively. For each test image, one or more areas on the road surface are selected, and the grey values of pixels in the defined areas are measured. G_r and H_r are calculated based on these measured grey values, and the average of all selected test images is taken as the average intensity of road surface G_r and the homogeneity measure of the road surface H_r. To determine the range of their variations along the road, the standard deviations of these values are computed, which are σ_G and σ_H. The parameters for the Hunter Valley image and Dungog image data are listed in the Table 7.2.

Table 7.2 Geometric and Photometric Parameters for Test Images

	G_r	H_r	σ_G	σ_H
Hunter Valley Image	120.0	12.0	21.5	4.3
Dungog Image Data	192.0	7.7	11.0	3.0

7.3 Results of Experiments

To test the developed methods for road recognition, images which contain linear objects were selected from the Hunter Valley and Dungog images, including those shown in Figures 7.2 and 7.3. The test results on these images will be presented in this section.

7.3.1 Hypothesis Generation

As explained in Chapter 6, hypotheses of road segments are generated from the high-resolution images in a bottom-up process, which includes the processes of feature extraction, generation of antipairs, grouping of antipairs and hypothesis generation of road segments. In the test, 25 test images were selected from the Hunter Valley images. A Digital Terrain Model (DTM) with a grid interval of 5 m was generated for the overlapping area on a Helava digital photogrammetric station to provide approximate 3-D information of terrain for generation of 3-D antipairs. In non-overlapping areas, only 2-D information was used. For the Dungog image, 25 images were tested, and the grouping of antipairs and generation of hypotheses of roads were performed in 2-D space as there was no ground control for this image pair. A summary of results of tests of road recognition carried out on these images is listed in the Table 7.3.

Table 7.3 Results of Hypothesis Generation

	No. of Hypotheses	True Road	Non-Road
Hunter Valley image	23	22	1
Dungog image	31	28	3

The recognition results of roads in images shown in Figures 7.2 and 7.3 are displayed in Figures 7.4 and 7.5. For each test image, its results in generation of antipairs, grouping of antipairs and hypothesis generation of road segments are demonstrated. The results of feature extraction are not shown. Both the Canny and SE operators are used for edge extraction and they generate similar results. In cases that the poor contrast occurs, the

addition, the knowledge of road widths was utilised to define the search area for the generation of generalised antiparallel pairs in order to reduce computation time. From the images shown in Figures 7.4(a), (b), (g), (h), (m), (n) and Figures 7.5(a), (b), (g), (h), (m), it can be seen that the structures of road surfaces were well established during the generation of antipairs when road boundaries are well-defined and there are no occlusions on the road surface. However, some generated antipairs do not correspond to road segments, especially when the image has a complex structure as shown in Figures 7.4(b), (h), (m), (n) and Figures 7.5(b), (g), (h). As can be seen, the generated antipairs in these images are usually separated due to the effects of occlusions, such as trees, shadows cast by trees, cars on the road surface, poor contrast between the road surface and its background, etc. They were grouped using the method described in Chapter 5. The results displayed in Figures 7.4(c), (d), (i), (j), (o) and (p), and Figures 7.5(c), (d) and (j) show that antipairs belonging to the same road were successfully grouped in most cases. However, some antipairs belonging to the same road were not linked due to their large difference in width as shown in Figures 7.5(i) and (j).

In hypothesis generation, road segments were hypothesised by applying the *rule one* given in Section 6.4 to all the generated road-like objects. Figures 7.4(e), (f), (k), (l), (q), (r) and Figures 7.5(e), (f), (k), (l) show the results of hypothesis generation, in which the hypothesised roads are displayed in black lines. In these examples, most roads were hypothesised although there are disturbances due to various occlusions on road surfaces. However, in some test images as shown in Figure 7.6(d) (red boxes), there were no hypotheses generated due to the effects of occlusions and poor contrast between the road surface and its background, and short road-like objects as shown in Figures 7.5(i), (j) and (n) were not hypothesised in this process. At the same time, some non-road objects which have similar shapes and geometric and radiometric properties to roads were hypothesised. Road-like objects with a length less than the threshold defined in the *rule one* were not hypothesised. They will be detected in the process of hypothesis verification described below.

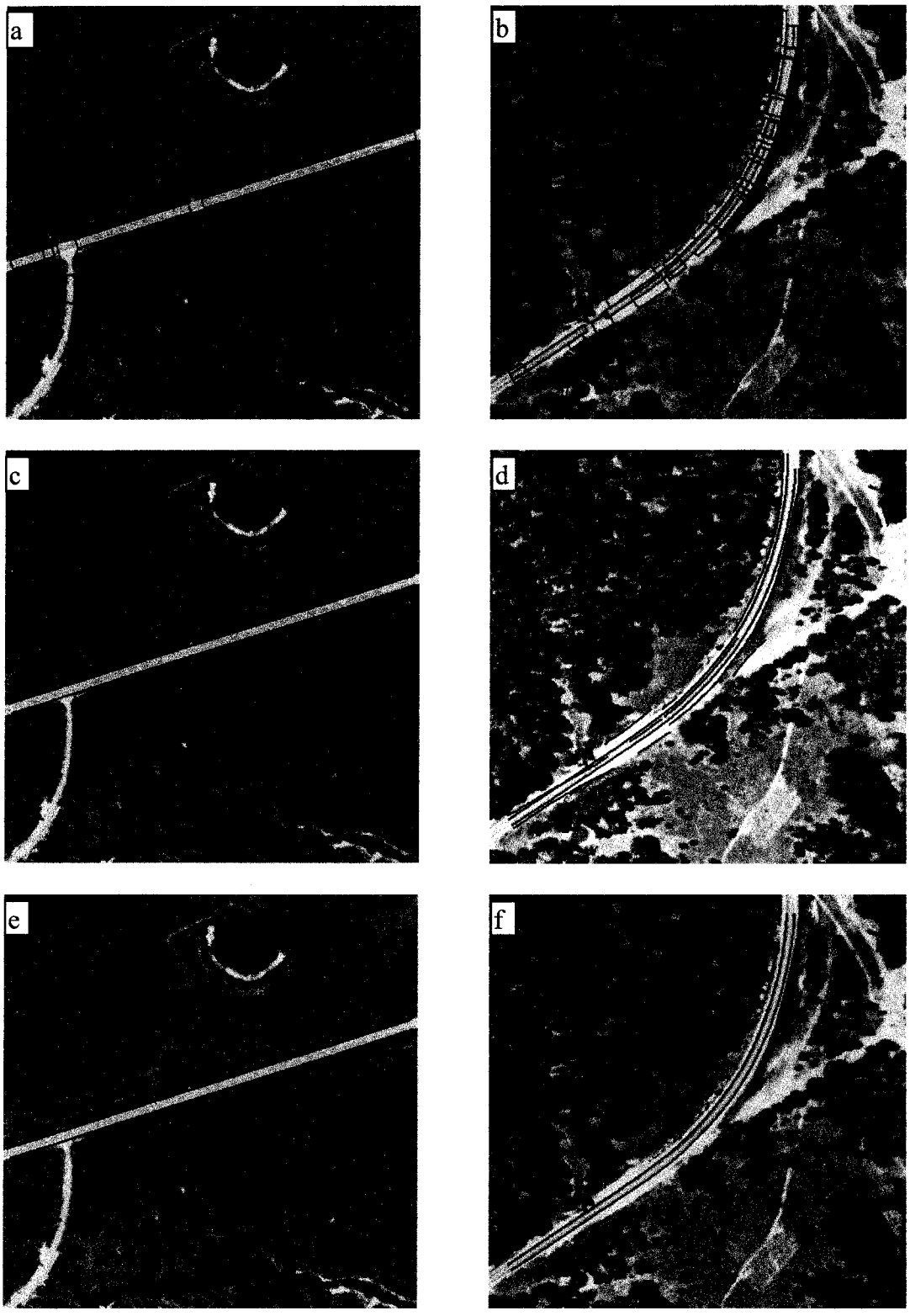


Figure 7.4 Results of Hypothesis Generation from High-resolution images – Hunter Valley Image Data (a), (b) Generated antipairs (c), (d) Formed road-like features after grouping (e), (f) Recognised road segments

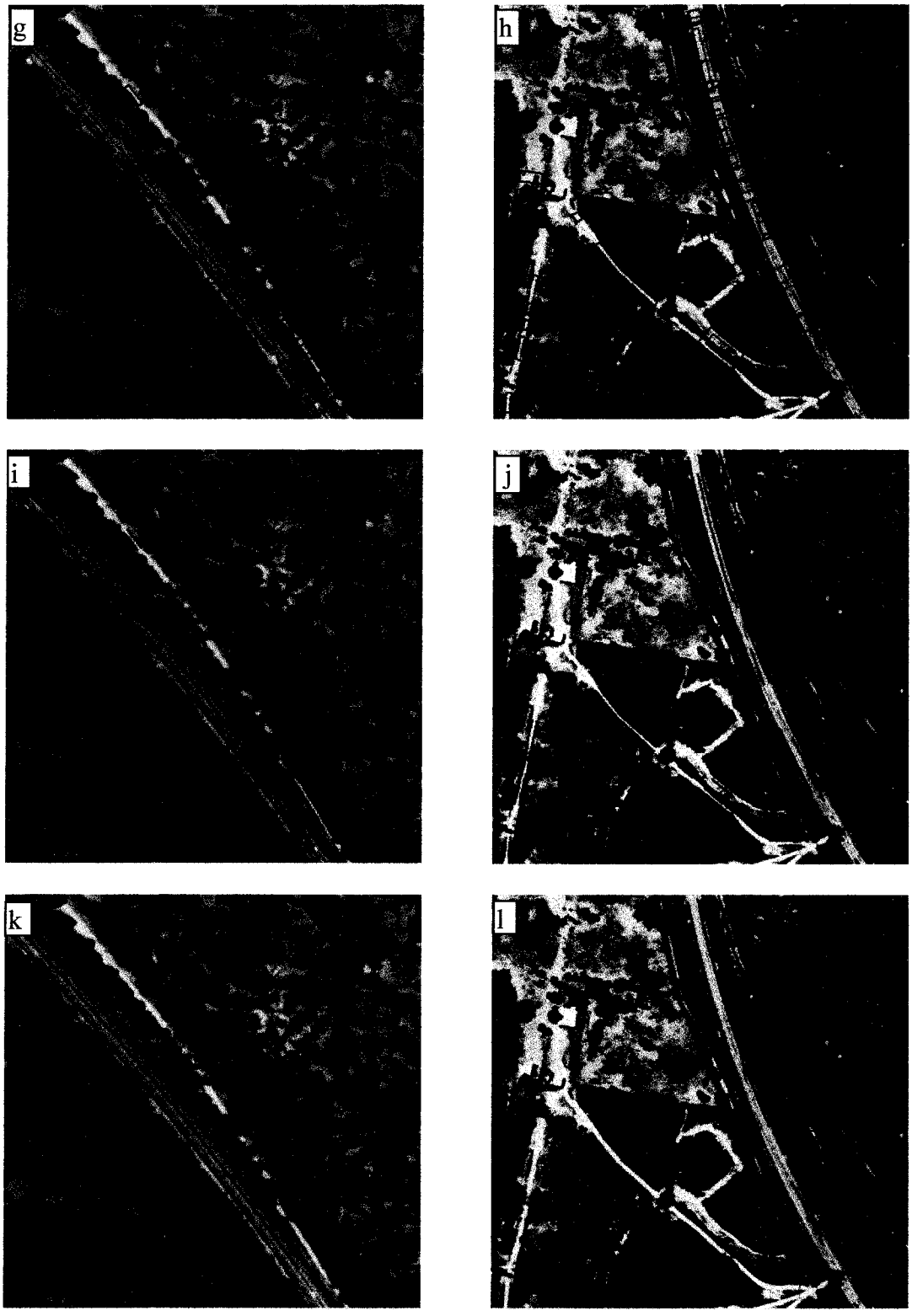


Figure 7.4 Results of Hypothesis Generation from High-resolution Images – Hunter Valley Image Data (continued) (g), (h) Generated antipairs (i), (j) Formed road-like features after grouping (k), (l) Recognised road segments

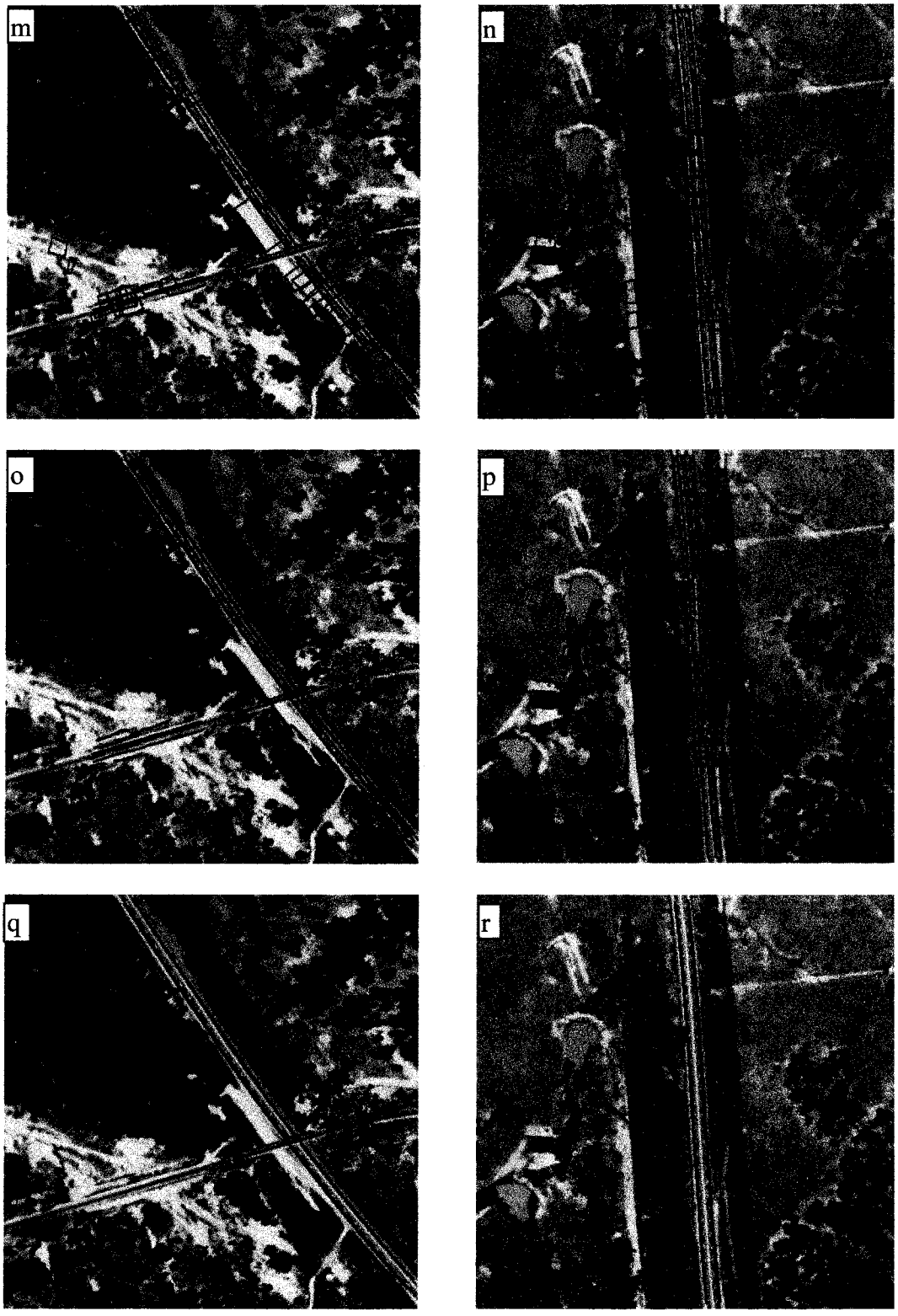


Figure 7.4 Results of Hypothesis Generation from High-resolution Images – Hunter Valley Image Data (continued) (m), (n) Generated antipairs (o), (p) Formed road-like features after grouping (q), (r) Recognised road segments

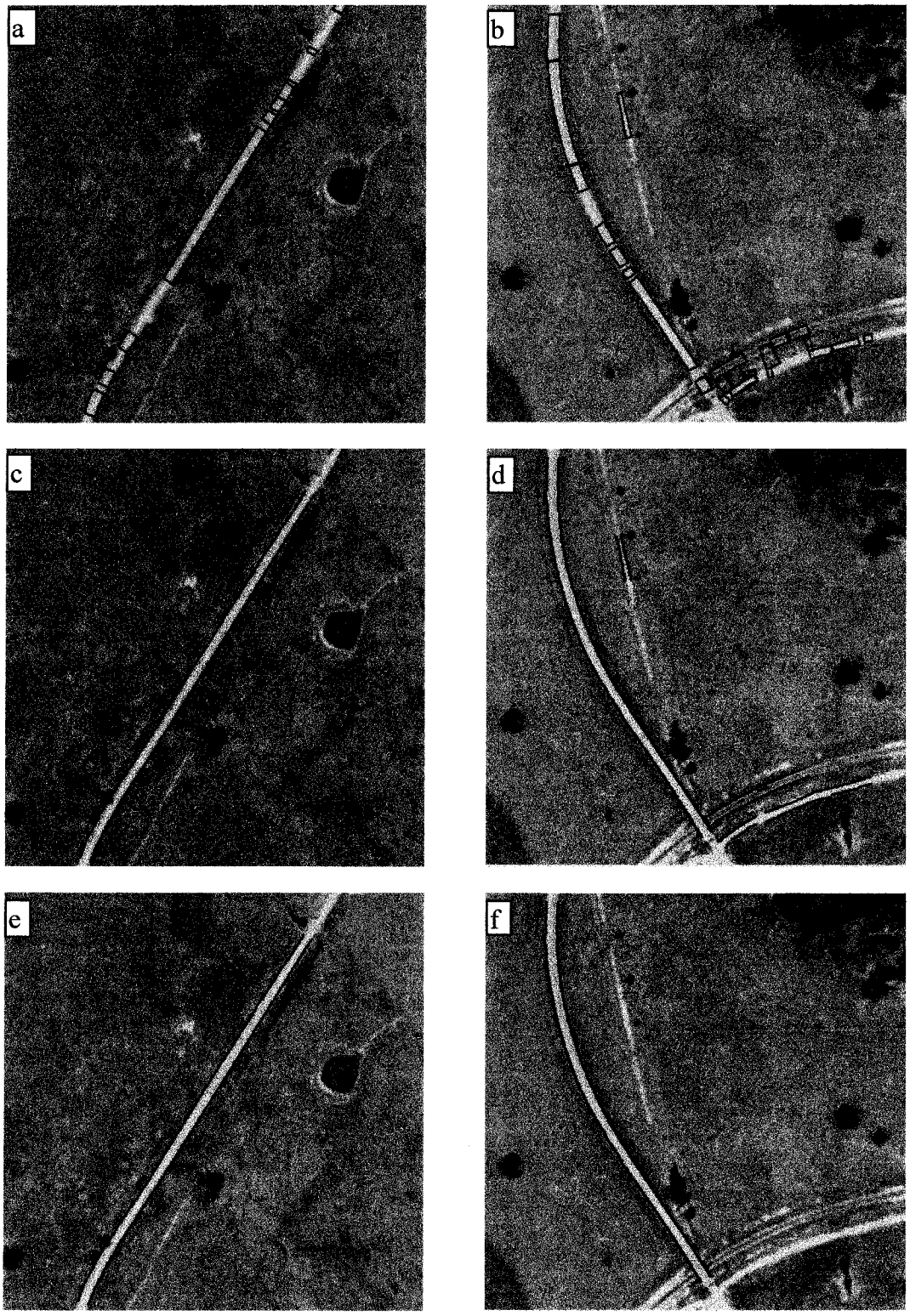


Figure 7.5 Results of Hypothesis Generation from High-resolution Images – Dungog Image Data (a), (b) Generated antipairs (c), (d) Formed road-like features after grouping (e), (f) Recognised road segments

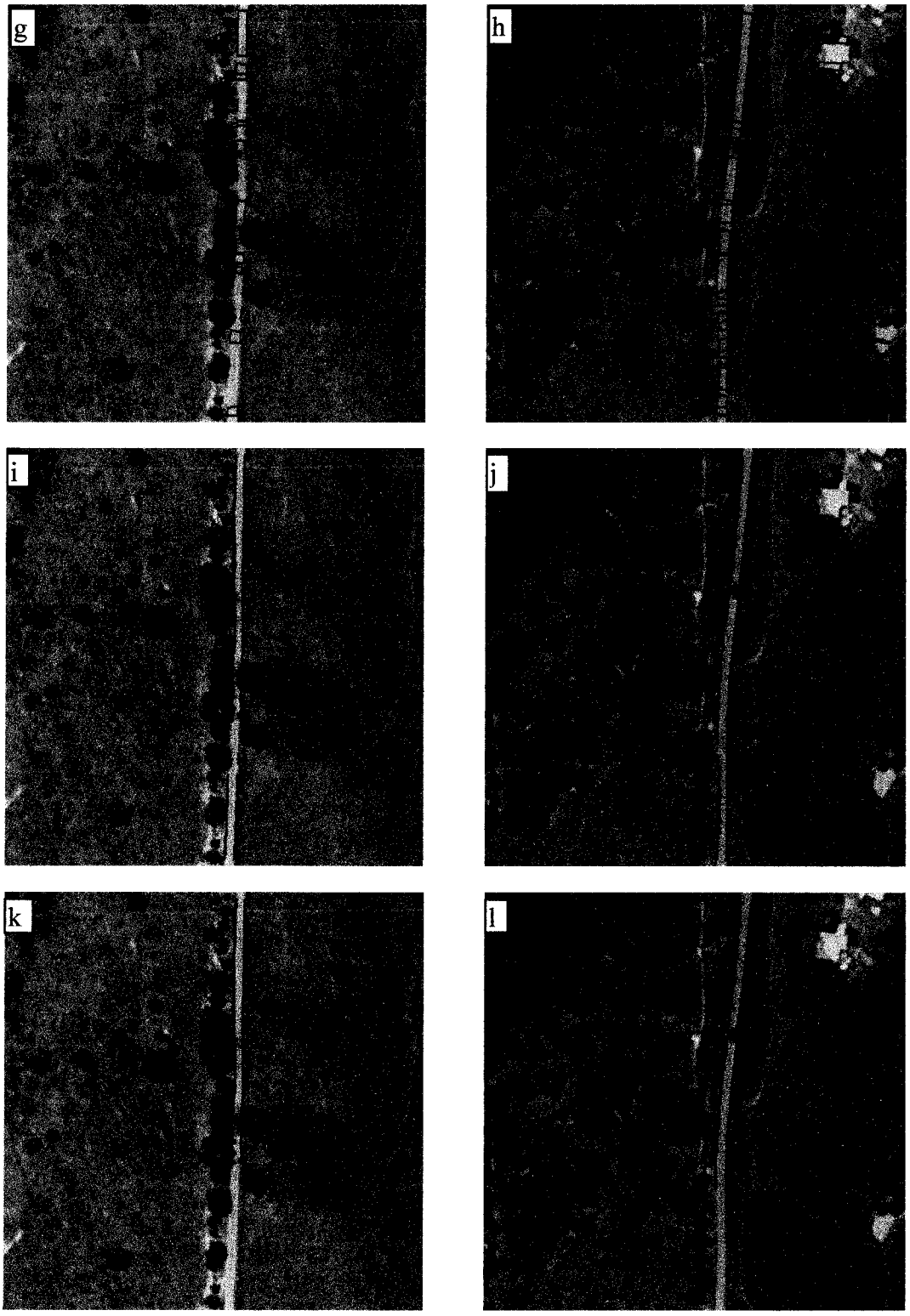


Figure 7.5 Results of Hypothesis Generation from High-resolution Images – Dungog Image Data (continued) (g), (h) Generated antipairs (i), (j) Formed road-like features after grouping (k), (l) Recognised road segments

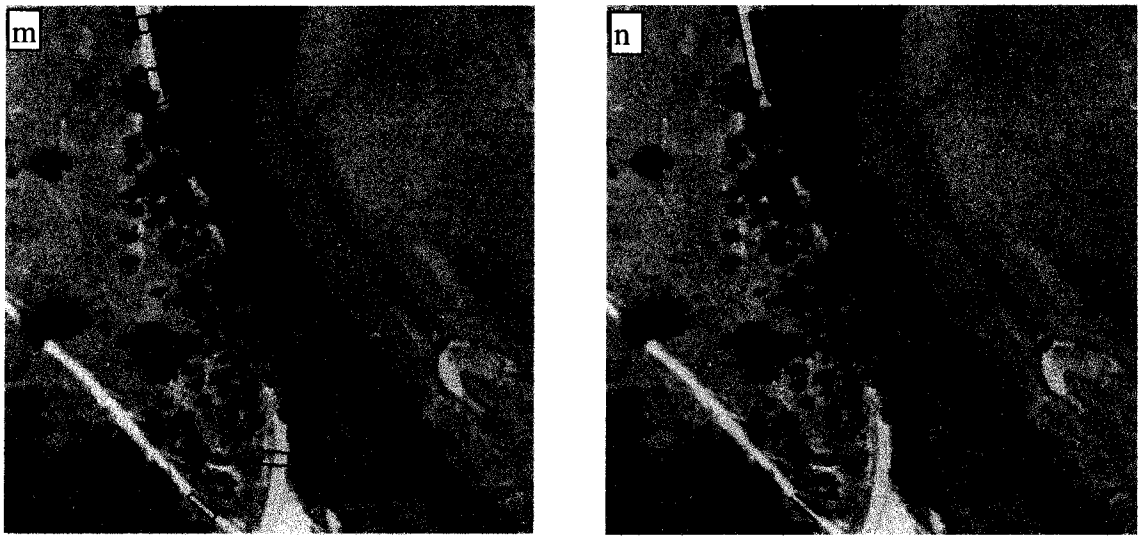


Figure 7.5 Results of Hypothesis Generation from High-resolution Images – Dungog Image Data (continued) (m) Generated antipairs (n) Formed road-like features after grouping

7.3.2 Verification of Hypotheses

As described in Chapter 6, the hypothesised road segments from high-resolution images are verified using their relationships with road networks generated from low-resolution images. To generate the overall structures of road networks, low-resolution images were derived by resampling the original images with a reduction factor of 15, which yielded ground resolutions of 11 m and 8 m for the Hunter Valley and Dungog images respectively. A morphological operator was used to extract line features from the generated low-resolution images and results are shown in Figures 7.6(a) and 7.7(a). As can be seen, not only were roads extracted, but also a large number of non-road objects. The existence of non-road objects will greatly increase the computation time of the subsequent processing and affect the reliability of the generated road networks. Thus, the irrelevant information should be deleted after line extraction. A split-and-merge process was applied in the test to split line segments corresponding to natural objects, e.g. tracks, boundaries lines, etc. A thresholding operation followed to eliminate short line segments. The results of these operations are shown in Figures 6(b) and 7(b). There is no doubt that most irrelevant line segments were removed. The remaining line

segments were then grouped based on their geometric and radiometric attributes, i.e. their spatial positions and orientations, gradient magnitudes, etc. Finally, the isolated line segments were removed to form complete networks as shown in Figures 6(c) and 7(c). Point features was extracted using the SE operator, and preprocessing was performed to delete false road intersections and to add some true intersections. The detected intersections were shown in Figures 6(c) and 7(c) (blue points). For the Hunter Valley image, the generated road network consists of three edges which correspond to two rural roads and one intersection while the road network derived from the Dungog image is composed of seven intersections and 15 edges, six of which correspond to rural roads.

Having generated the overall structures of road networks, a buffer zone was created for each edge in road networks with an offset of 3 pixels in low-resolution images. To establish the relationships between the hypothesised road segments and the road network derived from low-resolution images, the hypothesised road segments were projected onto the generated low-resolution images by a simple transformation, which are shown in Figures 7.6(d) and 7.7(d) respectively. A hypothesised road segment is related to an edge in the road network by relationship *part-of* when it lies within the buffer zone of the edge. If a transformed hypothesised road segment does not lie within any buffer zone of edges, but connect to an edge in the road network, a *connect-to* relationship is created between the hypothesised road segment and the corresponding edge. A hypothesised road segment is defined as *isolated* if it does not fall within any buffer zone of edges and does not connects to any edge in the road network. The generated relationships between the hypothesised road segments and the road networks for the Hunter Valley and Dungog images are listed in Table 7.4..

Table 7.4 Relationships between Hypothesised Road Segments and Road Network

	No. of Hypotheses	*part-of*	*isolated*	*connect-to*
Hunter Valley image	23	22	1	0
Dungog image	31	27	2	2

It can be seen that most hypothesised road segments correspond to roads (22 of 23 in the Hunter Valley image and 27 of 31 in the Dungog image). With the derived relationships between the hypothesised road segments and the road networks generated from the low-resolution images, hypothesised road segments are verified using the rules described in Chapter 6. In the Hunter Valley and Dungog images, there were one and two isolated road segments removed respectively (Figures 7.6(e) and 7.7(e)). For hypothesised road segments related to road networks by relationship *part-of*, they are confirmed using *rule two* with T_r set as 0.5. The ratios of the length of an edge to the total lengths of all transformed road segments related to it by relationship *part-of* are among 0.5 to 0.7 for edges 1, 2 and 3 in the Hunter Valley image and edges 1, 3, 4, 9, 10 and 11 in the Dungog image while they are between 0.0 and 0.1 for all other edges in the Dungog image. Hypothesised road segments with ratios larger than 0.5 were accepted as shown in Figures 7.6(e) and 7.7(e). Totally there were 22 hypothesised road segments in the Hunter Valley image and 26 hypothesised road segments in the Dungog image accepted. There are three road segments related to the road network with relationship *connect-to* in the Dungog image. However, they were rejected as they are too short (<20 pixels in the image).

7.3.3 Detection of Missing Road Segments

As short road segments between the recognised road segments were not hypothesised in the hypothesis generation, recognised road segments in both images are often disjoined as shown in Figures 7.6(e) and 7.7(e). Therefore, these unrecognised road segments should be detected after hypothesis generation, to form a complete network. The detection of missing road segments in this study is based on the spatial relationships between verified road segments. After hypothesised road segments are verified, they are related by the relationship *neighbour* based on their spatial positions. When the distance between two verified road segments is larger than a given threshold which was defined as five pixels in low-resolution images in this study, a missing road segment is predicted. Then a procedure for finding antipairs between the verified road segments is invoked and *rule six* described in Chapter 6 is applied to the extracted antipairs to recognise them.

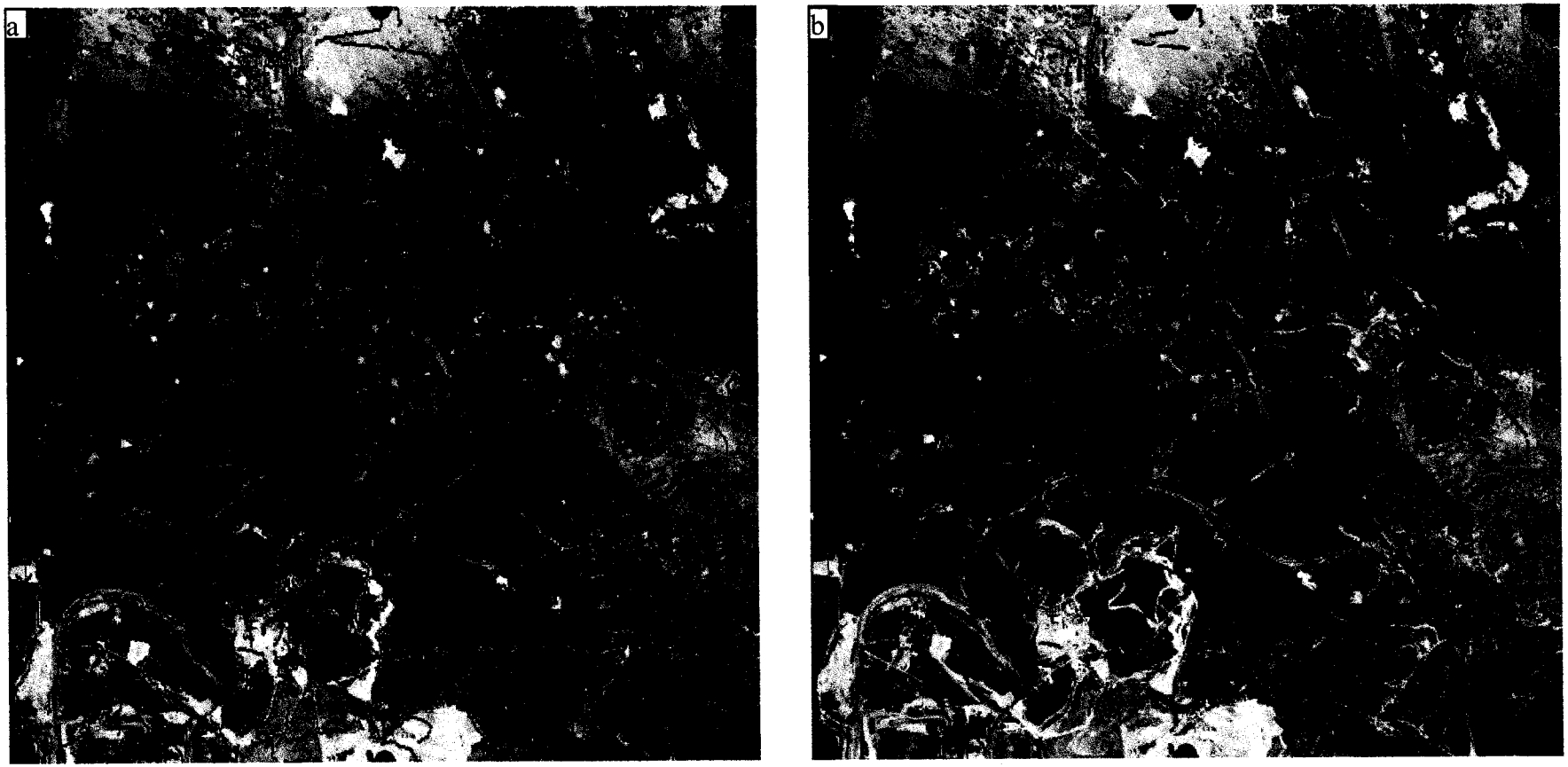


Figure 7.6 Verification of Hypotheses - Hunter Valley Image (a) Extracted lines from low-resolution image (b) Salient line segments after preprocessing

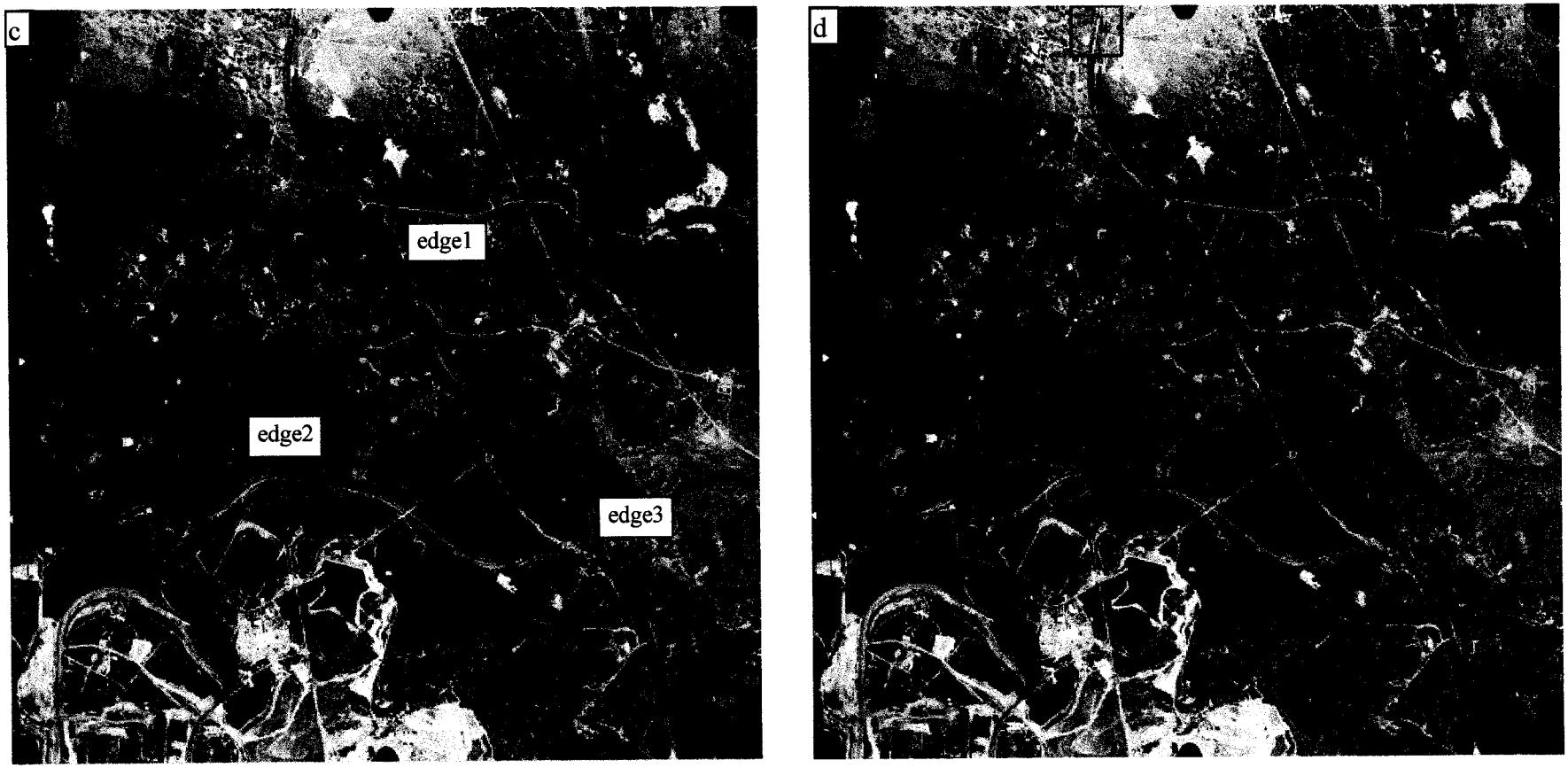


Figure 7.6 Verification of Hypotheses - Hunter Valley Image (continued) (c) Generated road network from low-resolution image (d) Hypothesised road segments after transformation

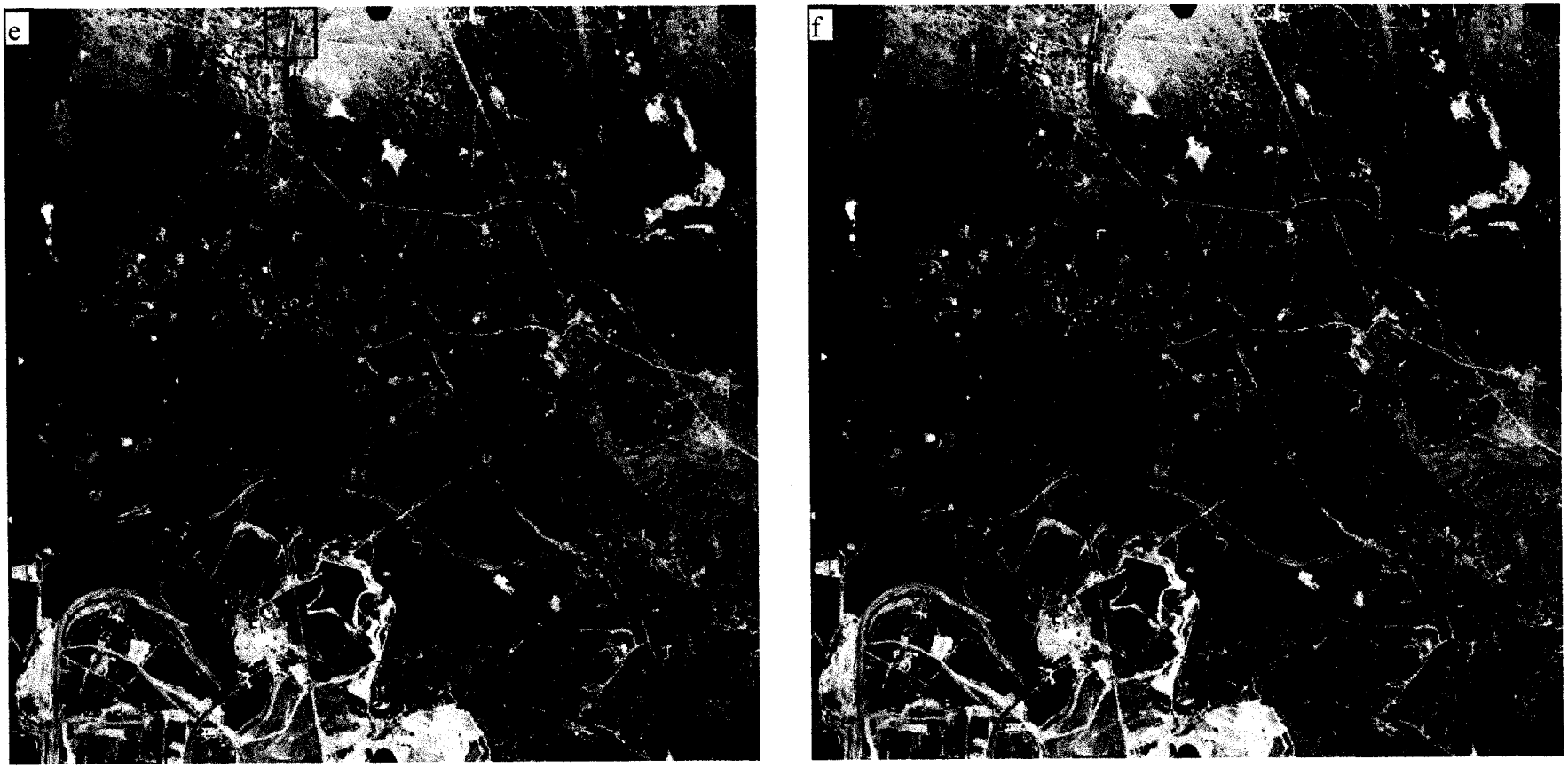


Figure 7.6 Verification of Hypotheses - Hunter Valley Image (continued) (e) Results after verification (f) Results after detection of missing road segments

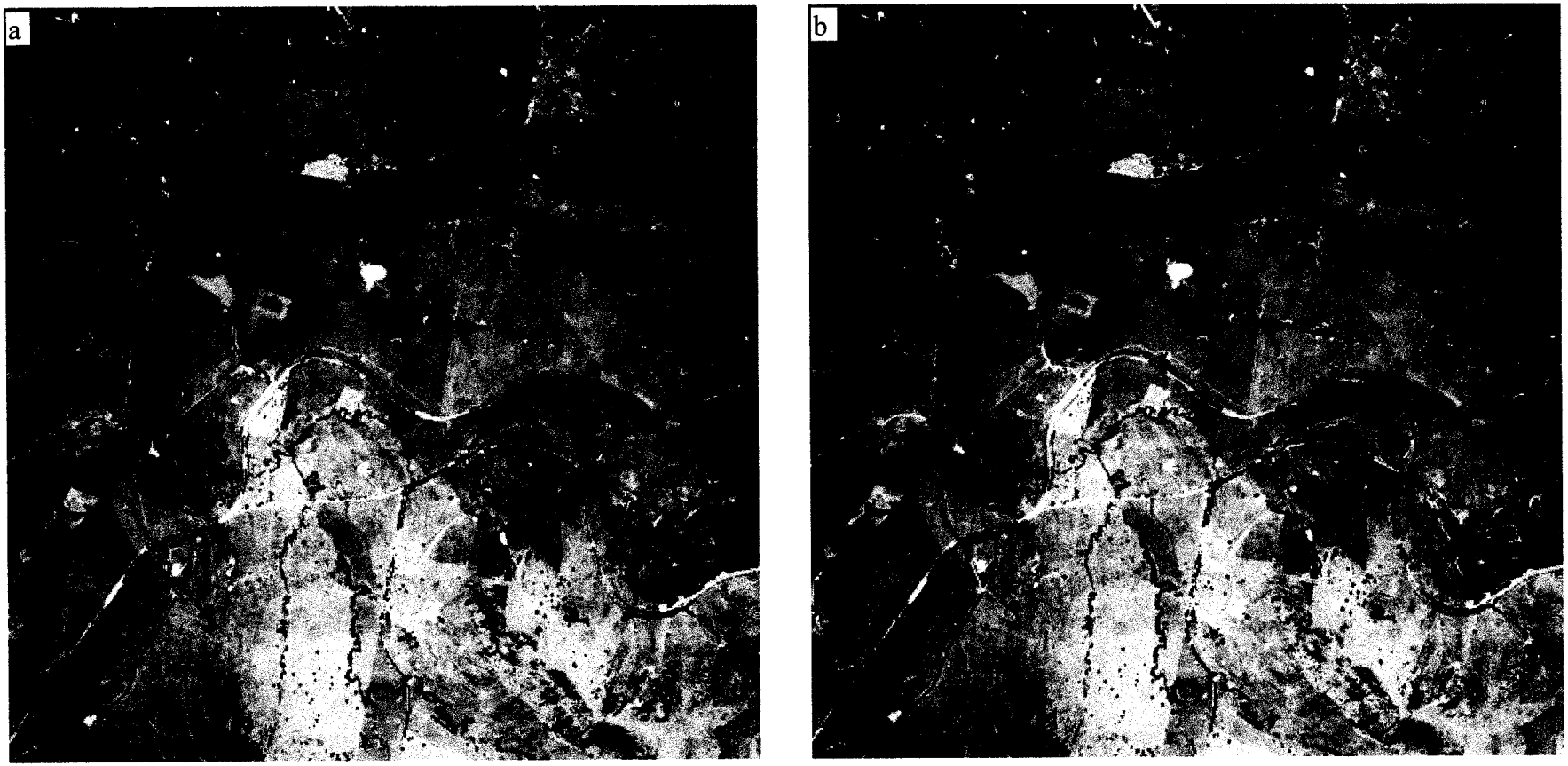


Figure 7.7 Verification of Hypotheses - Dungog Image (a) Extracted lines from low-resolution image (b) Salient line segments after preprocessing

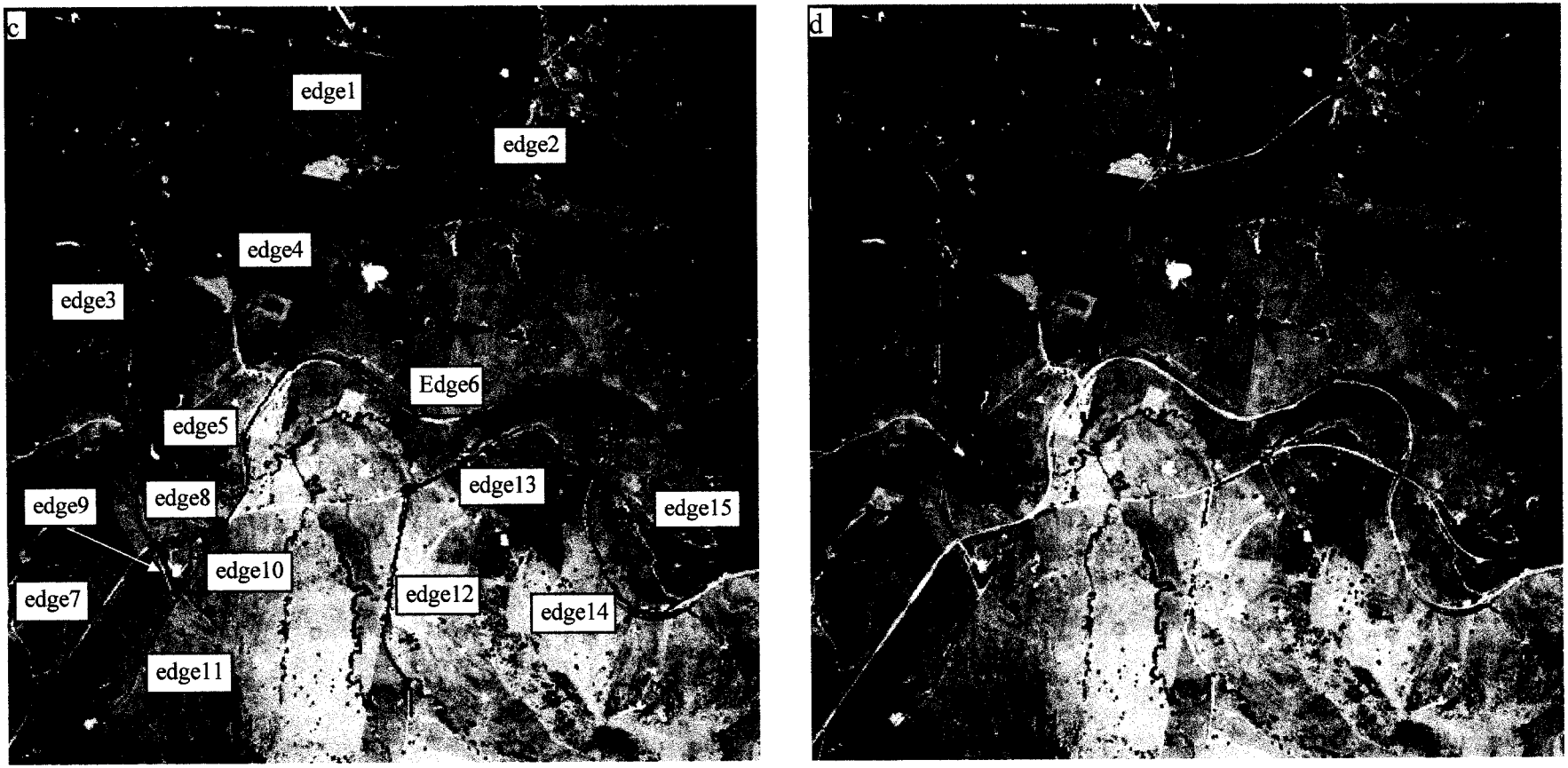


Figure 7.7 Verification of Hypotheses - Dungog Image (continued) (c) Generated road network from low-resolution image (d) Hypothesised road segments after transformation

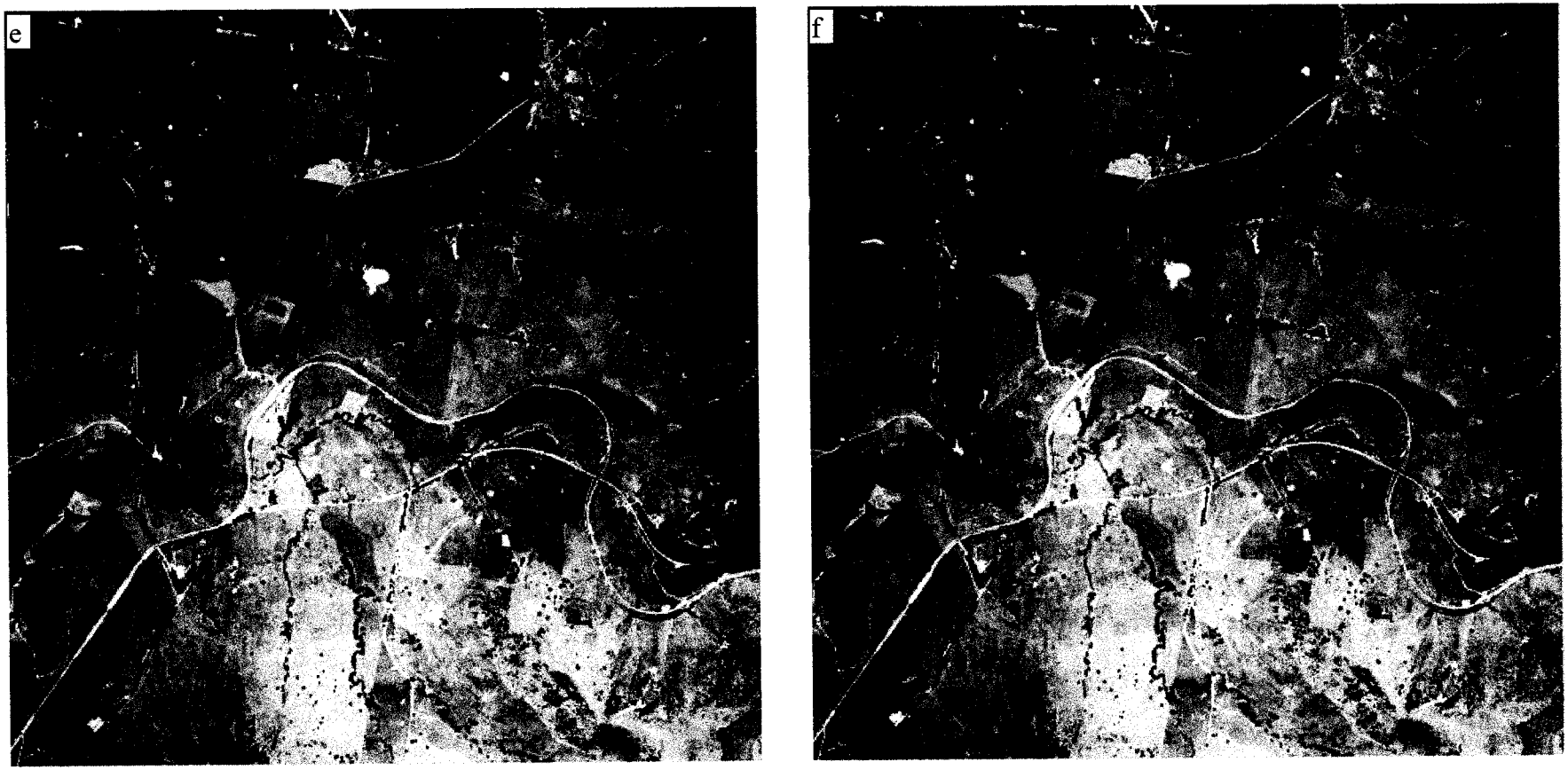


Figure 7.7 Verification of Hypotheses - Dungog Image (continued) (e) Results after verification (f) Results after detection of missing road segments

Figure 7.8 shows two examples of detection of missing road segments, in which road segments were not hypothesised in the high-resolution images because the lengths of the extracted road-like objects are less than the defined threshold for road length in *rule one* due to the poor contrast between the road surface and degraded boundary lines. Based on the positions of the recognised road segments in neighbouring test images, search areas for detecting road-like objects were defined, as shown in Figure 7.8 (white lines). In the search areas, all antipairs were generated and recognised by using *rule six*. In Figures 7.8(a) and (b), there was one road segment found in each image. However, there were still some parts of the road not detected in this process as there were no road-like objects generated. They will be detected in the process of occlusion detection. In Hunter Valley and Dungog images, there were 22 and 21 short road segments found, which are shown in Figures 7.6(f) and 7.7(f) respectively.

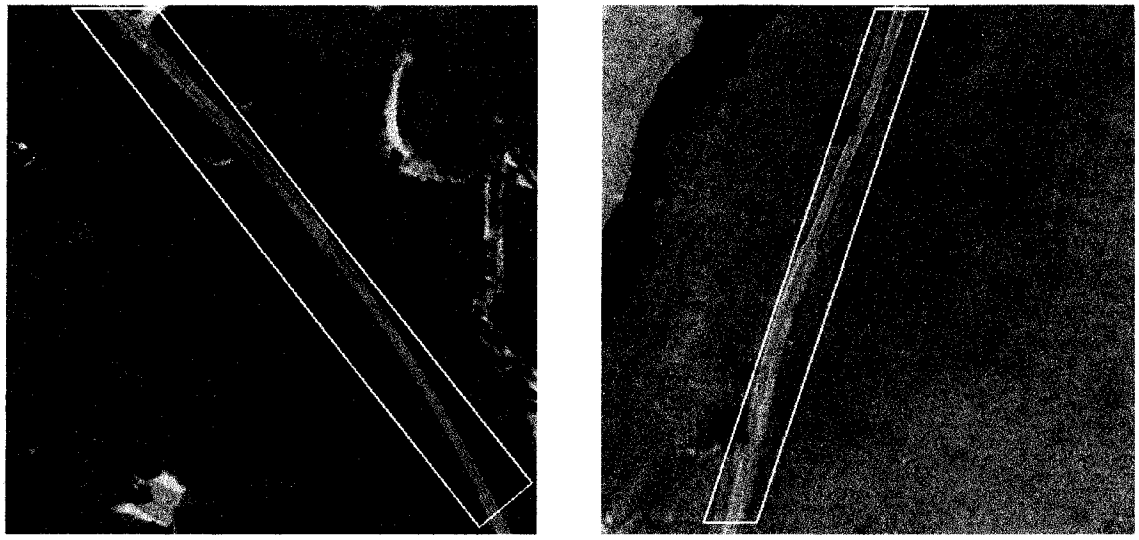


Figure 7.8 Detection of Missing Road Segments

7.3.4 Detection of Occlusions

In most cases, images contain occlusions which cause a road to be broken into several short segments that cannot be recognised in the recognition process and neither found in detection of missing road segments. On the other hand, occlusions provide contextual information between them and roads which can be used to infer the existence of roads in

the occluded areas. In this study, occlusions are detected using a supervised classification method - maximum likelihood classification. Two examples of occlusion detection are shown in Figure 7.9. In Figure 7.9(a), a road in the image is heavily occluded by trees, and hence no road segments were hypothesised in the hypothesis generation. In the detection of occlusions, a search area was defined based on the positions of two recognised road segments related by relationship *neighbour*. The pixels in the defined area were then classified using the maximum likelihood classification method described in Chapter 6.

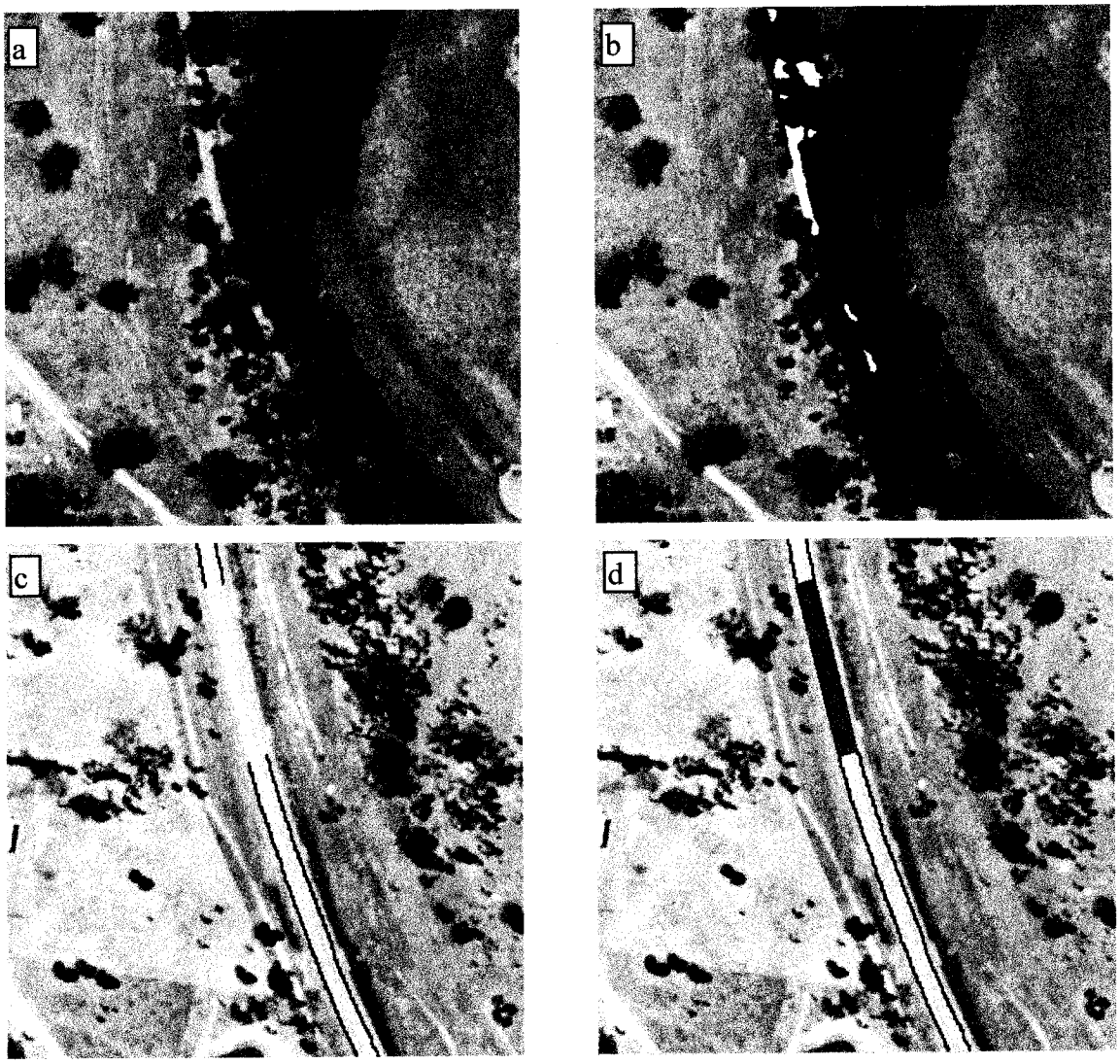


Figure 7.9 Detection of Occlusions (a) Road segment occluded by trees (b) Classified trees (shaded area) (c) Road segment with poor contrast (d) Classified road surface (shaded area)

In classification, a training process for determining the spectral distributions of typical classes of objects was performed on ten test images for five classes i.e. roads, trees, grasses, bare soil and shadows. In this example, most pixels were classified as trees which are displayed as black in Figure 7.9(b). An example of poor contrast between the road surface and its background is displayed in Figure 7.9(c) in which the black lines represent the detected road segments and the part between them was unrecognised due to the poor contrast. Using the maximum likelihood classification method, this part was classified as road surface as shown in Figure 7.9(d)(shaded area). Once an occlusion is found, its relationship with the verified road segments is used to infer the existence of road segment in the occluded area. As discussed in Chapter 6, the possible factors causing a road segment unrecognised are trees, shadows, poor contrast. Thus, a road segment was reasoned in occluded areas in these two examples. Finally, the neighbouring verified road segments were connected by polynomial interpolation when a road segment was inferred, to form a complete road network as shown in Figure 7.10.

7.4 Discussion

This chapter has presented the results of road recognition from aerial images using the methods developed in the previous chapters. The results have shown that all major rural roads in the tested images were correctly recognised. However, the minor rural roads in the images were not recognised as they have different surface characteristics and irregular shapes which are not included in the semantic road model developed in this study. The widths of roads included in the recognition rules can be obtained from the design manuals of roads. Another geometric property of a road, the length, is determined by the dimensions of the test image and the maximum gap between two road segments which can be bridged in the process of grouping. The use of this property can remove most non-road objects which have similar structures to roads since they are usually short. However, some short true road segments are rejected in the process of hypothesis generation. Therefore, a further step is required to find the missing road segments. The photometric properties of roads in the image used in the recognition rules

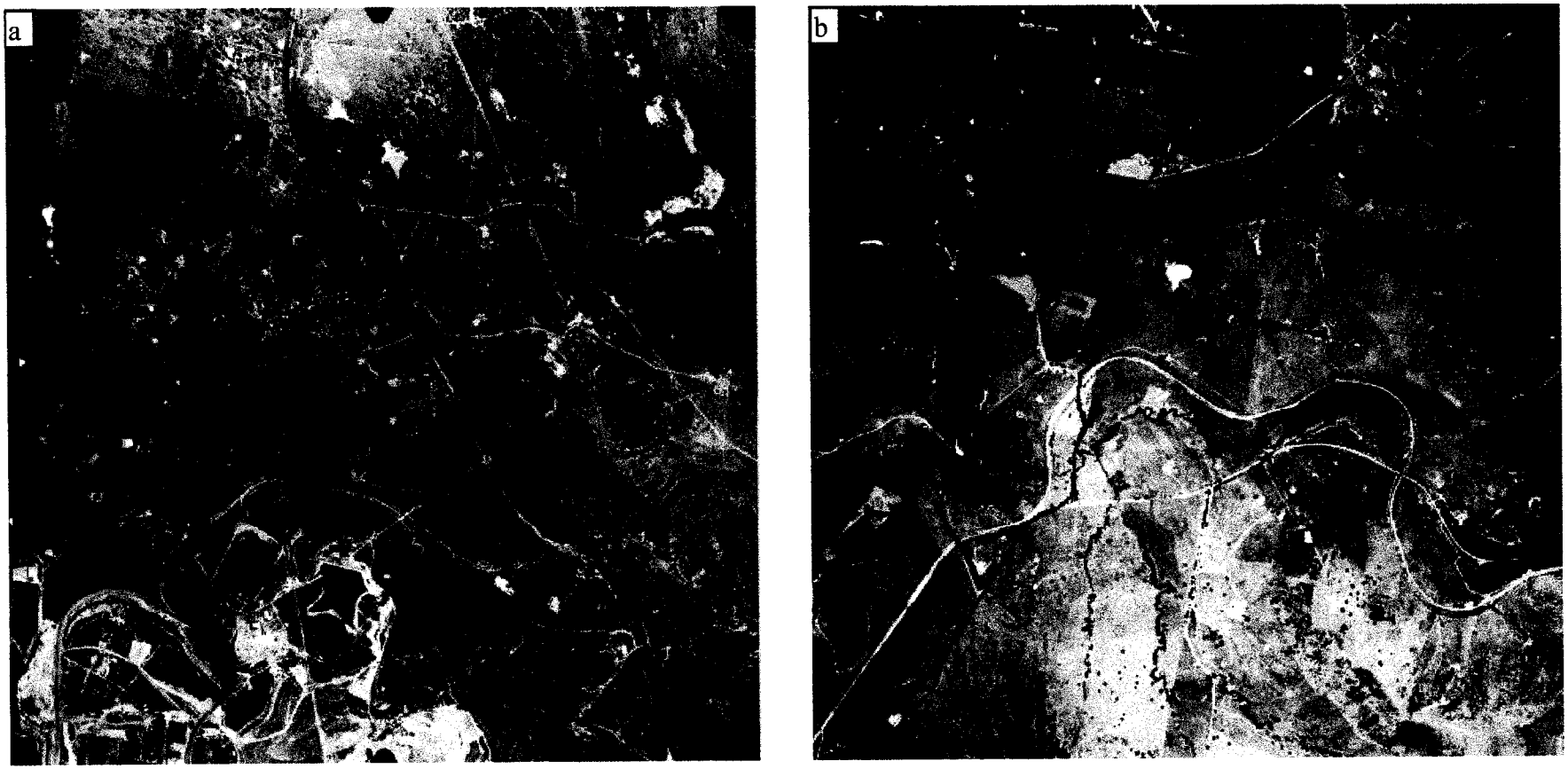


Figure 7.10 Detected Road Network after Hypothesis Verification (a) Results of Hunter Valley image (b) Results of Dungog Image

cannot be determined *a priori*, since they depend on a number of factors, such as the incidence angle of the Sun, weather conditions, slope of surface and processing conditions of the images, etc. They were determined in a training process in this study, in which a large number of test images were used for measuring the properties of road surfaces. The photometric properties of road surfaces determined by a training process can be applied to images obtained under similar conditions, but for images taken and processed in different conditions, a further training process is needed.

As hypotheses of road segments are generated in a local area in high-resolution images, it is likely that a non-road object is hypothesised if it has similar shape and properties to a road. Therefore, a verification process is required to remove spurious hypotheses of roads. Low-resolution images exhibit overall structures of road networks very well, thus relationships between roads are well defined. Hypothesised road segments are related to roads in road networks by the defined relationships *part-of*, *connect-to* and *isolated*. Using these relationships, hypotheses of roads can be effectively verified.

As mentioned above, short road segments are not hypothesised in the hypothesis generation. They are inferred using spatial relationships between verified roads after verification of hypotheses and found in a top-down process. However, it should be pointed out that only short road segments which are not covered by occlusions can be found. Road segments covered by various occlusions, such as trees, shadows, etc., cannot be detected in this process.

Occlusions in the image are a difficult problem in automatic road extraction. They not only break a road into segments, but also cover some parts of the road which cannot be extracted in the processes of road recognition and detection of missing road segments. In this study, a supervised classification method, maximum-likelihood classification, is used to detect occlusions between the recognised road segments. It uses the statistical spectral properties of the classes of objects and spectral values of the pixels in the occluded area to determine the class of the pixels in the area. Once some type of occlusion is found, the contextual relationship between the detected occlusion and the verified roads can be established and used to infer the existence of a road segment in the

occluded area. The results in this study show that occlusions have been detected successfully in colour images. However, it must be pointed out that the work carried out in this study is the initial work on detection of occlusions. To test the efficiency and reliability of the method, more experiments should be done, and multi-spectral or hyperspectral images should be used in the future study in order to achieve more accurate classification. It should also be mentioned that a road segment can be inferred based on the relationship between the detected occlusion and verified roads. This does not mean that the found occlusions definitely correspond to road segments. For example, when trees are located between two verified roads, it is naturally believed that a road segment between them is covered by the trees. This is usually true, but when the found occlusion is a large forest, it cannot be confirmed that a road segment definitely exists there. To make sure of the existence of a road segment in the area of occlusion, a human check may be required.

Chapter Eight

Conclusions

8.1 Conclusions of this Study

In this thesis, a knowledge-based method for automatic extraction and recognition of roads from aerial images has been presented based on a semantic road model in which roads are defined in both high- and low-resolution images. The semantic road model was developed based on the observations that different resolutions of images reflect different aspects of roads, and reliable results cannot be achieved by using a single resolution image. In the developed road model, roads are recognised from aerial images in a hybrid process. They are hypothesised from high-resolution images in a bottom-up process, which includes low-level processing for feature extraction, intermediate-level processing for generation of symbolic representation of roads and high-level processing for recognition. Hypothesised roads are then verified using the relationships between roads in high-resolution images and roads in low-resolution images. Spatial relationships between verified roads are used to infer the missing road segments and occlusions. The main contributions of this thesis include:

(i) Preprocessing in feature extraction

Feature extraction includes their detection and preprocessing, to generate symbolic representations of the features for the subsequent processing. As there are numerous feature detection algorithms and approaches, this thesis concentrates on the development of efficient methods for preprocessing of detected features for road recognition, which includes an edge/line tracking algorithm, a split-and-merge process for generation of smooth edge/line segments and a linking algorithm. The tracking algorithm uses information of the magnitudes and orientations of edge/line points and their connectivity. The tracked edge/line is represented by its length and two point

chains recording the positions of points on the edge/line in forward and backward directions. As a least squares polynomial fitting is incorporated into the split-and-merge process, the generated edges/ lines are smooth and accurate. The small gaps which occur between the extracted edges/ lines due to the effects of image noise are bridged in the linking process. To facilitate the subsequent processing, short edge/line segments caused by image noise, object texture, etc. are eliminated.

(ii) Generation of road structure

Road boundaries are smooth linear features. In this study, generalised antiparallel pairs were proposed to define road structures in high-resolution images and an algorithm for computing them in 2-D space was developed based on the extracted edges and their attributes. A generated antipair is represented symbolically in terms of a number of geometric and photometric attributes which include position, orientation, width and the photometric properties of the area covered by the antipair. As two smooth curves are used in a generalised antiparallel pair instead of two straight lines, road boundaries are represented more accurately.

To determine the positions of generalised antiparallel pairs in 3-D space, a modified active contour model has been developed, in which the conditions of smoothness and parallelism of road boundaries are incorporated into the geometric energy model. With the developed model and an antipair computed from left and right image, the position of the antipair in 3-D can be determined efficiently and accurately. The merging process of 3-D antipairs can partially overcome the problems caused by occlusions such as vehicles on road surfaces.

Occlusions such as trees, shadows cast by trees, vehicles on road surface, poor contrast between road surface and its background, etc. are a common phenomenon in the image. They usually cover part of the road boundaries, and thus cause the generated antipairs to be broken. To remove the effects of occlusions, a grouping procedure is introduced to link the disconnected antipairs. In grouping, not only are spatial constraints used, i.e. collinearity and proximity, but also the geometric and photometric properties of

antipairs, which are formulated into a rule in Prolog. The use of the geometric and photometric properties can improve the reliability of the grouping of antipairs since they usually have similar properties if they belong to the same road.

(iii) Recognition of Roads

Roads are man-made objects with specific properties which distinguish them from other objects. These specific properties of roads are formulated as a rule in Prolog in this study, and hypotheses of roads are generated by applying the rule to the road-like objects produced in the intermediate-level processing. The geometric properties included in the rule can be obtained from the road design manuals, while the photometric properties are determined through a large number of measurements on the test images.

As only local information is used in the generation of hypotheses, non-road objects with similar structures and properties may be incorrectly recognised as roads. To remove spurious hypotheses, a novel approach for verification of hypotheses has been developed. The method uses the global information of road networks, derived from low-resolution images and relationships between hypothesised road segments in high-resolution images and the generated road network, to verify the hypotheses. The characteristics of a road network are formulated into rules and utilized to eliminate false hypotheses.

Due to the effects of occlusions, a road may be divided into several short segments which are not hypothesised in the hypothesis generation. To find the missing road segments, a procedure based on the spatial relationships between verified road segments has been established.

Occlusions in an image not only break a road into several segments, but also cover some parts of the road which cannot be recognised in the recognition process or found in the process of detection of missing road segments. To find the road segment covered by occlusions, a process based on the supervised classification method has been developed.

It uses the maximum likelihood classification technique to classify the occlusions between the verified road segments. The relationship between the detected occlusions and the verified roads are then used to infer the existence of a road segment in the occluded area.

As outlined in Chapter 1, the main objectives of this study were:

- To develop a general procedure for automatic road extraction from aerial images,
- To define a suitable semantic road model for recognition of roads,
- To explore the use of knowledge of roads in automatic road extraction and the mechanism for knowledge representation, and
- To analyse the characteristics of occlusions on road surface in the image and to develop a suitable approach for their detection.

From the previous chapters, it can be seen that all of these objectives have been successfully achieved. First, a hybrid control strategy has been developed, which includes a bottom-up procedure for generation of hypotheses of road segments, a hypothesis verification based on the spatial relationships between road segments, and a top-down process for detection of missing road segments and occlusions. It has been shown that such a control strategy is suitable to automatic road extraction from aerial images and successful results have been achieved using this control strategy. Since a road in high-resolution images has a simple structure which can be described by generalised antipairs, and specific geometric and photometric properties, it can be recognised by a bottom-up process. Occlusions are a problem encountered in road extraction, which often cause some road segments to be unrecognised in the recognition process. Using the spatial relationships between verified roads, they can be found in a top-down procedure.

Secondly, a semantic road model has been proposed, in which roads are defined in both high- and low-resolution images and relationships between roads are defined as well. It

is realised that different resolutions of images represent different aspects of characteristics of roads and reliable recognition of roads cannot be achieved by using a single resolution of images. More reliable recognition results can be achieved when different resolutions of images are used. The results carried out in this study have proved that a combination of high- and low-resolutions of images can give reliable recognition results of roads.

The properties of roads have been studied, and it has been found that not only are the appearances of roads important for recognition of roads, but also their geometric and photometric properties. The relationships between roads are of crucial importance as well. The specific properties of roads are used for their recognition, while their relationships can be used to verify the recognised results, by rules in Prolog.

A procedure for finding missing road segments has been established based on the spatial relationships between verified road segments, and a method for detection of occlusions based on supervised classification method has been proposed. In the developed methods, missing road segments are found in a top-down procedure, and occlusions are classified using maximum likelihood classification method. The results show that both missing road segments and occlusions can be successfully detected.

8.2 Further Investigations

This thesis has established a general paradigm for automatic road extraction from aerial images and investigated the formulation of knowledge of roads for road recognition. Successful results have been achieved when the developed methods are applied to aerial images for rural roads. It should be pointed out that the situations in urban areas will be more complicated, e.g. complex structures of road surfaces, more frequent occlusions, and regularly distributed houses on both sides of a road. Thus, to recognise roads in urban areas, more knowledge of roads, such as their structures and characteristics as well as their contextual information, should be used in future research.

In this thesis, the hypotheses of roads are generated in a bottom-up process. Thus, the results of recognition rely on the quality of low- and intermediate-level processing. One important issue in low- and intermediate-level processing is the selection of thresholds, which are determined based on the knowledge of roads in this study. For different images, these thresholds may need to be given different values in order to yield satisfactory results. A problem which will arise is how to determine the values of these thresholds for different images. Is there any automatic approach for determining the values of these thresholds? One possibility is to use a machine learning technique in artificial intelligence, which is a process of acquiring knowledge automatically using a certain inference strategy, e.g. induction. Such procedures should be considered for future research.

In the recognition process, only the knowledge of roads and the imaging system was used. There are a number of additional sources of knowledge about roads, e.g. existing maps and geographical information databases, which provide contextual information that would be very useful for the recognition process. Therefore, the methods of introducing this information into the recognition process is an important issue for future research.

Abdou, I. E. and Pratt, W. K. 1979. Quantitative Design and Evaluation of Enhancement /Thresholding Detectors. IEEE Proceeding **67**(5), 753-763.

Abkar, A. A. and Mulder, N. J. 1998. Likelihood-Based Classification of High Resolution Images to Generate the Initial Topology and Geometry of Land Cover Segments. International Archives of Photogrammetry and Remote Sensing, **32**, 1-7.

Ackermann, F. 1984. Digital Image Correlation: Performance and Potential Application in Photogrammetry. Photogrammetric Record, **11**(64), 429-439.

Argialas, D. P. and Harlow, C. A. 1990. Computational Image Interpretation Models: An Overview and a Perspective. Photogrammetric Engineering and Remote Sensing, **56**, 871-886.

Bajcsy, R. and Tavakoli, M. 1976. Computer Recognition of Roads from Satellite Pictures. IEEE Transactions on Systems, Man, and Cybernetics, **6**, 623-637.

Ballard, D. H. 1981. Generalizing the Hough Transform to Detect Arbitrary Shapes. Pattern Recognition, **13**, 111-122.

Ballard, D. H. and Brown, C. M. 1982. Computer Vision. Prentice Hall Inc., Englewood, New Jersey.

Baltsavias, E. P. 1991. Multiphoto Geometrically Constrained Matching. PhD Dissertation, Institute of Geodesy and Photogrammetry, ETH Zurich, Report No. **49**.

Barr, A. and Feigenbaum, E. A. (Eds). 1981. The Handbook of Artificial Intelligence, Vol. **1**, William Kaufman.

Barzohar, M. and Cooper, D. B. 1993. Automatic Finding of Main Roads in Aerial Images by Using Geometric-Stochastic Models and Estimation. In IEEE Proceedings on Computer Vision and Pattern Recognition, pp. 459-464.

Baumgartner, A., Eckstein, W., Mayer, H., Heipke, C. and Ebner, H. 1997. Context-Supported Road Extraction. In Automatic Extraction of Man-Made Objects from Aerial

and Space Images (II) (A. Grün, E. P. Baltsavias, O. Henricsson, Eds.). Birkhäuser, Verlag, Basel, pp. 299-310.

Boldt, M., Weiss, R. and Riseman, E. 1989. Token Based Extraction of Straight Lines. IEEE Transactions on Systems, Man, and Cybernetics, **19**, 1581-1594.

Bordes, G., Giraudon, G. and Jamet, O. 1997. Road Modeling Based on a Cartographic Database for Aerial Image Interpretation. In Semantic Modeling for the Acquisition of Topographic Information from Images and Maps (W. Förstner and L. Plümer, Eds.). Birhäuser, Verlag, pp. 123-139.

Bratko, I. 1990. Prolog: Programming for Artificial Intelligence. Addison-Wesley Publishing Company, Wokingham, England.

Broder, A. and Rosenfeld, A. 1981. A Note on Collinearity Merit. Pattern Recognition, **13**, 237-239.

Burns, J. B., Hanson, A. R. and Riseman, E. 1986. Extracting Straight Lines. IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-8**, 425-455.

Busch, A. 1994. Fast Recognition of Lines in Digital Images without User-Supplied Parameters. International Archives of Photogrammetry and Remote Sensing, **30**, 91-97.

Busch, A. 1996. A Common Framework for the Extraction of Lines and Edges. International Archives of Photogrammetry and Remote Sensing, **31**, 88-93.

Canny, J. A. 1986. Computational Approach to Edge Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence. **PAMI-8**, 679-698.

Chachra, V., Ghare, P. M. and Moore, J. M. 1979. Applications of Graph Theory Algorithms. Elsevier, New York.

Cleynenbreugel, J. van, Fierens, F., Suetens, P. and Oosterlinck, A. 1990. Delineating Road Structures on Satellite Images by a GIS-Guided Technique, Photogrammetric Engineering and Remote Sensing, **56**, 893-898.

Crevier, D. and Lapage, R. 1997. Knowledge-Based Image Understanding Systems: A Survey. Computer Vision and Image Understanding, **67**, 161-185.

Davies, L. S. 1975. A Survey of Edge Detection Techniques. Computer Graphics and Image Processing **4**, 248-270.

Dolan, J. and Weiss, R. 1989. Perceptual Grouping of Curved Lines. In Proceedings of DARPA Image Understanding Workshop, pp. 1135-1145.

Dougherty, E. R. 1992. An Introduction to Morphological Image Processing. A Publication of SPIE - The International Society for Optical Engineering, Bellingham, Washington.

Drewniok, C. and Rohr, K. 1997. Exterior Orientation - An Automatic Approach based on Fitting Analytic Landmark Models. ISPRS Journal of Photogrammetry and Remote Sensing, **52**, 132-145.

Durkin, J. 1994. Expert Systems: Design and Development. Prentice Hall International, London.

Ebner, H., Fritsch, D. and Heipke, C. (Eds.) 1991. Digital Photogrammetric Systems. Wichmann, Verlag, Karsruhe.

Fischler, M. A. and Bolles, R. C. 1986. Perceptual Organization and Curve Partioning. IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-8**, 100-105.

Fischler, M. A., Tenenbaum, J. M. and Wolf, H. C. 1981. Detection of Roads and Linear Structures in Low-Resolution Aerial Imagery Using a Multisource Knowledge Integration Technique. Computer Graphics and Image Processing, **15**, 201-223.

Förstner, W. 1993. Feature Extraction in Digital Photogrammetry. Photogrammetric Record, **14**(82), 595-611.

Förstner, W. 1994. A Framework for Low Level Feature Extraction. Computer Vision-ECCV 94, Vol. II, pp. 383-394.

Förstner, W. and Gülch, E. 1987. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features. Intercommission Workshop on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, pp. 281-305.

Förstner, W. and Plümer, L. (Eds.). 1997. Semantic Modeling for the Acquisition of Topographic Information from Images and Maps. Birhäuser, Verlag.

Fua, P. and Leclerc, Y. G. 1990. Model Driven Edge Detection. Machine Vision and Applications, **3**, 45-56

Fuchs, C. and Förstner, W. 1995. Polymorphic Grouping for Image Segmentation. 5th ICCV' 95, Boston, pp. 175-182.

Füger, H., Stein, G. and Stilla, U. 1994. Multi-Population Evolution Strategies for Structural Image Analysis. In IEEE Proceedings on Evolutionary Computation ICEC'94, Vol. I, pp. 229-234.

Geman, D. and Jedynak, B. 1996. An Active Testing Model for Tracking Roads in Satellite Images. IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-18**, 1-14.

Ghosal, S. and Mehrotra, R. 1993. Orthogonal Moment Operators for Subpixel Edge Detection. Pattern Recognition, **26**, 295-306.

Giarratano, J. and Riley, G. 1994. Expert Systems: Principles and Programming. PWS Publishing Company, Boston.

Gong, P. and Wang, J. 1997. Road Network Extraction from Airborne Digital Camera Images: A Multi-Resolution Comparison. In IEEE Proceedings on Geoscience and Remote Sensing, Singapore, Vol. II, pp. 895-897.

Greve, C. (Ed.) 1996. Digital Photogrammetry: An Addendum to Manual of Photogrammetry. American Society for Photogrammetry and Remote Sensing, Bethesda, Maryland.

Grimson, W. E. L. 1990. Object Recognition by Computer: The Role of Geometric Constraints. The MIT Press, Cambridge, Massachusetts, 1990.

Groch, W. 1982. Extraction of Line Shaped Objects from Aerial Images Using a Special Operator to Analyze the Profiles of Functions. Computer Graphics and Image Processing, **18**, 347-358.

Grün, A. 1985. Adaptive Least Squares Correlation: A Powerful Image Matching Technique. South African Journal of Photogrammetry, Remote Sensing and Cartography, **14**, 175-187.

Grün, A. and Agouris, P. 1994. Linear Feature Extraction by Least Squares Template Matching Constrained by Internal Shape Forces. International Archives of Photogrammetry and Remote Sensing, **30**, 316-323.

Grün, A., Baltsavias, E. P. and Henricsson, O. (Eds.) 1997. Automatic Extraction of Man-Made Objects from Aerial and Space Images (II). Birkhäuser Verlag, Basel.

Grün, A., Kübler, O. and Agouris, P. (Eds.) 1995. Automatic Extraction of Man-Made Objects from Aerial and Space Images (I). Birkhäuser, Verlag, Basel.

Grün, A. and Li, H. 1994. Semi-Automatic Road Extraction by Dynamic Programming. International Archives of Photogrammetry and Remote Sensing, **30**, 324-332.

Grün, A. and Li, H. 1996. Linear Feature Extraction with LSB-Snakes from Multiple Images. International Archives of Photogrammetry and Remote Sensing, **31**, 266-272.

Gülch, E. 1995. Automatic Control Point Measurement. In Proceedings of Photogrammetric Week'95, Wichmann-Verlag, pp. 185-196.

Gunst, M. de. 1996. Knowledge-Based Interpretation of Aerial Images for Updating of Road Maps. Publications on Geodesy, No. **44**, Netherlands Geodetic Commission.

Gunst, M. de and Vosselman, G. 1997. A Semantic Road Model for Aerial Image Interpretation. In Semantic Modeling for the Acquisition of Topographic Information

from Images and Maps (W. Förstner and L. Plümer, Eds.), Birkhäuser, Verlag, pp. 107-122.

Hanson, A. R. and Riseman, E. M. 1978. A VISIONS: Computer Vision System for Interpreting Scenes. In Computer Vision Systems (A. R. Hanson and E. M. Riseman, Eds.), Academic, New York, pp. 303-333.

Hanson, A. R. and Riseman, E. M. 1988. VISIONS Image Understanding Systems. In Advances in Computer Vision (C. Brown) (Ed.), Lawrence Erlbaum Associates, Publishers, Hillsdale, New Jersey.

Haralick, R. M., Watson, L. T. and Laffey, T. J. 1983. The Topographic Primal Sketch. International Journal of Robotics Research, **2**, 50-72.

van der Heijden, F. 1995. Edge and Line Feature Extraction Based on Covariance Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-17**, 16 -33.

Heipke, C. 1995. State-of-Art of Digital Photogrammetric Workstations for Topographic Applications. Photogrammetric Engineering and Remote Sensing, **61**, 49-56.

Heipke, C. 1997. Automation of Interior, Relative, and Absolute Orientation. ISPRS Journal of Photogrammetry and Remote Sensing, **52**, 1-19.

Heipke, C., Steger, C. and Multrhamer, R. 1995. A Hierarchical Approach to Automatic Road Extraction from Aerial Images. In Integrating Photogrammetric techniques with Scene Analysis and Machine Vision II, Proceedings of SPIE, Vol. 2486, pp. 222-231.

Heitger, F. 1995. Feature Detection Using Suppression and Enhancement. Technical Report BIWI-TR-160, Institute for Communications Technology, Image Science Laboratory, ETH, Switzerland.

Heitger, F., Rosenthaler, R., von der Heydt, Peterhans, E. and Kübler, O. 1992. Simulation of Neural Contour Mechanisms: From Simple to End-Stopped Cells. Vision Research, **32**, 963-981.

Henricsson, O. 1996. Analysis of Image Structures Using Colour Attributes and Similarity Relations. PhD Dissertation, ETH, Zurich, Switzerland.

Hueckel, M. 1971. An Operator Which Locates Edges in Digitized Pictures. J. ACM, **18**, 113-125.

Hummel, R. A. 1979. Feature Detection Using Basis Functions. Computer Graphics and Image Processing, **9**, 40-55.

Hwang, V. S. S., Davis, L. L. S. and Matsuyama, T. 1986. Hypothesis Integration in Image Understanding Systems. Computer Vision, Graphics, and Image Processing, **36**, 321-371.

Kass, M., Witkin, A. and Terzopoulos, D. 1988. Snakes: Active Contour Models. International Journal of Computer Vision, **1**, 321-331.

Kestner, W. and Kazmierczak, H. 1978. Semi-automatic Extraction of Roads from Aerial Photographs. Final Technical Report, Institute of Information Processing and Pattern Recognition, FIM, Karlsruhe, Germany.

Kirsch, R. A. 1971. Computer Determination of the Constituent Structure of Biological Images. Computers and Biomedical Research, **4**, 315-328.

Koch, H., Pakzad, K. and Tönjes, R. 1997. Knowledge Based Interpretation of Aerial Images and Maps Using a Digital Landscape Model as Partial Interpretation. In Semantic Modeling for the Acquisition of Topographic Information from Images and Maps (W. Förstner and L. Plümer, Eds.), Birkhäuser Verlag, pp. 3-19.

Lay, M. G. 1990. Handbook of Road Technology. Vol. one, Gordon and Breach Science Publishers, New York.

Lemmens, M. J. P. M. 1996. A Survey of Boundary Delineation Methods. International Archives of Photogrammetry and Remote Sensing, **31**, 435-441.

Lemmens, M. J. P. M., Bruel, E. W. and Fennis, F. 1988. Linear Feature Extraction and Road Recognition from Large Scale Digital Aerial Images. International Archives of Photogrammetry and Remote Sensing, **27**, 476-483.

Levine, M. D. and Shaheen, S. I. 1981. A Modular Computer Vision System for Picture Segmentation and Interpretation. IEEE Transactions on Pattern Recognition and Machine Intelligence, **PAMI-3**, 540-556.

Li, M. 1989. Hierarchical Multi-Point Matching with Simultaneous Detection and Location of Breaklines. PhD. Thesis, Royal Institute of Technology, Stockholm.

Lowe, D. G. 1985. Perceptual Organization and Visual Recognition. Kulwer Academic, Hingham, Massachusetts.

Lü, Y. 1988. Interest Operator and Fast Implementation. International Archives of Photogrammetry and Remote Sensing, **27**, 491-499.

Lü, Y. 1997. One Step to A Higher Level of Automation for Softcopy Photogrammetry Automatic Interior Orientation. ISPRS Journal of Photogrammetry and Remote Sensing, **52**, 103-109.

Maillard, P. and Cavayas, F. 1989. Automatic Map-Guided Extraction of Roads from SPOT Imagery for Cartographic Database Updating. International Journal of Remote Sensing, **10**, 1775-1787.

Marr, D. 1976. Early Processing of Visual Information. Philosophical Transactions of the Royal Society of London, Series B, **275**, 483-524.

Marr, D. 1982. Vision, Freeman, San Francisco, California.

Marr, D. and Hildreth, E. 1980. Theory of Edge Detection. Proceedings of the Royal Society, B **207**, 187-217.

Matsuyama, T. and Hwang, V. 1985. "SIGMA: A Framework for Image Understanding" . In Proceedings of 9th International Joint Conference on Artificial Intelligence, pp. 908-915.

Mayer, H. and Steger C. 1998. Scale-Space Events and Their Link to Abstraction for Road Extraction. ISPRS Journal of Photogrammetry & Remote Sensing, **53**, 62-75.

McKeown, D. M. and Delinger, J. L. 1988. Cooperative Methods for Road Tracking in Aerial Imagery. IEEE Proceeding on Computer and Computer Recognition, pp. 662-672.

McKeown, D. M., Harvey, W. A. and McDermott, J. 1985. Rule-Based Interpretation of Aerial Imagery. IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-7**, 570-585.

Medioni, G. and Yasumoto, Y. 1987. Cornet Detection and Curve Representation using Cubic B-Splines. Computer Vision, Graphics, and Image Processing, **39**, 267-278.

Minsky, M. L. 1975. Frame System Theory, Thinking: Readings in Cognitive Science, CUP, P.N. Johnson-Laird and P.C. Watson (Eds.), Cambridge, Massachusetts.

Mohan, R. and Nevatia, R. 1989. Using Perceptual Organization to Extract 3-D Structures. IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-11**, 1121-1139.

Mohan, R. and Nevatia, R. 1992. Perceptual Organization for Scene Segmentation and Description. IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-14**, 616-634.

Moravec, H. P. 1977. Towards Automatic Visual Obstacle Avoidance. In Proceedings of the 5th International Joint Conference on Artificial Intelligence, pp. 584.

Nagao, M. and Matsuyama, T. 1980. A Structural Analysis of Complex Aerial Photographs. Plenum Press, New York.

Nalwa, V. and Pauchon, E. 1987. Edgel Aggregation and Edge Description. Computer Vision, Graphics, and Image Processing, **40**, 79-94.

Neuenschwander, W., Fua, P., Székely, G. and Kübler, O. 1995. From Ziplock Snakes to Velcro Surfaces. In Automatic Extraction of Man-Made Objects from Aerial and Space Images (I) (A. Grün, O. Kübler, P. Agouris, Eds.). Birkhäuser Verlag, Basel, pp. 105-114.

Nevatia, R. and Babu, R. 1980. Linear Feature Extraction and Description. Computer Graphics, and Image Processing, **13**, 257-269.

Nicolin, B. and Gabler, R. 1987. A Knowledge-Based System for the Analysis of Aerial Images. IEEE Transactions on Geoscience and Remote Sensing, **25**, 317-329.

Nilsson, R. J. 1980. Principles of Artificial Intelligence. Tioga Pub. Co., Palo Alto, California.

Pal, N. R. and Pal, S. K. 1993. A Review on Image Segmentation Techniques, Pattern Recognition, **26**, 1277-1294.

Peli, T. and Malah, D. 1982. A Study of Edge Detection Algorithms. Computer Graphics and Image Processing, **20**, 1-21.

Petrou, M. and Kittler, J. 1991. Optimal Edge Detectors for Ramp Edges. IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-13**, 483-491.

Philip, K. P. 1991. Automatic Detection of Myocardial Contours in Cine Computed Tomographic Images. Ph.D. Thesis, University of Iowa.

Prewitt, J. M. S. 1970. Object Enhancement and Extraction. In Picture Processing and Psychopictorics, B. S. Lipkin and A. Rosenfeld (Eds.), Academic Press, New York.

Quam, A. 1978. Road Tracking and Anomaly Detection in Aerial Imagery. Proceedings of the DARPA Image Understanding Workshop, pp. 51-55.

Quinlian, M. R. 1968. Semantic Memory, Semantic Information Processing, M. L. Minsky (Ed.), MIT Press, Cambridge.

Ramer, U. 1972. An Iterative Procedure for the Polygonal Approximation of Planar Curves. Computer Graphics and Image Processing, **1**: 244-256.

Richards, J. A. 1993. Remote Sensing Digital Image Analysis: An Introduction. Springer -Verlag, Berlin.

Roberts, L. G. 1965. Machine Perception of Three-Dimensional Solids. In Optical and Electro-Optical Information Processing, J. P. Tippett et al. (Eds.), MIT Press, Cambridge, Massachusetts.

Robinson, G. S. 1977. Edge Detection by Compass Gradient Masks. Computer Graphics and Image Processing, **6**, 492-501.

Rosenfeld, A. and Kak, A.C. 1982. Digital Picture Processing. Academic Press, New York.

Rosenholm, D. 1987. Multi-point Matching Using the Least Squares Techniques for Evaluation of Three-Dimensional Models. Photogrammetric Engineering and Remote Sensing, **53**, 621-626.

Ruskoné, R., Airault, S. and Jamet, O. 1994. Road Network Interpretation: A Topological Hypothesis Driven System. International Archives of Photogrammetry and Remote Sensing, **30**, 711-717.

Schenk, T. 1997. Towards Automatic Aerial Triangulation. ISPRS Journal of Photogrammetry and Remote Sensing, **52**, 110-121.

Schilling, K. J. and Vögtle, T. 1996. Satellite Image Analysis Using Integrated Knowledge Processing. International Archives of Photogrammetry and Remote Sensing, **31**, 752-757.

Shen, J. 1992. An Optimal Linear Operator for Step Edge Detection. CVGIP: Graphical Models and Image Processing. **54**, 112-133.

Sonka, M., Hlavac, V. and Boyle, R. 1993. Image Processing, Analysis and Machine Vision. Chapman Hall, London.

Steger, C. 1996. Extracting Lines using Differential Geometry and Gaussian Smoothing. International Archives of Photogrammetry and Remote Sensing, **31**, pp. 821-826.

Stilla, U. 1995. Map-Aided Structural Analysis of Aerial Images. ISPRS Journal of Photogrammetry and Remote Sensing, **50**, 3-10.

Strat, T. M. and Fischler, M. A. 1991. Context-Based Vision: Recognizing Objects Using Information from Both 2-D and 3-D Imagery. IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI -13**, 1050-1065.

Tang, L., Braun, J. and Debitsch, R. 1997. Automatic Aerotriangulation - Concept, Realization, and Results. ISPRS Journal of Photogrammetry and Remote Sensing, **52**, 122-131.

Tang, L. and Heipke, C. 1996. Automatic Relative Orientation of Aerial Images. Photogrammetric Engineering and Remote Sensing, **62**, 47-55.

Ton, J., Jain, A. K., Enslin, W. R. and Hudson, W. D. 1989. Automatic Road Identification and Labelling in Landsat 4 TM Images. Photogrammetria, **43**, 257-276.

Trinder, J. C. and Li, H. 1995. Semi-Automatic Feature Extraction by Snakes. In Automatic Extraction of Man-Made Objects from Aerial and Space Images (I) (A. Grün, O. Kübler, P. Agouris, Eds.). Birkhäuser, Verlag, Basel, pp. 95-104.

Trinder J. C. and Sowmya, A. 1997. Modeling and Representation Issues for Photogrammetry, Remote Sensing and Machine Vision. In Proceeding of First International Workshop on Image Analysis and Information Fusion (H. Pan, M. Brooks, D. McMichael, and G. Newsam, Eds.), Adelaide, Australia, pp. 1-10.

Trinder, J. C. and Wang, Y. 1997a. Towards Automatic Feature Extraction for Mapping and GIS. In Geographical Information Systems and Remote Sensing Application (I. V. Muralikrishna) (Ed), Hyderabal, India, pp. 8-12.

Trinder, J. C. and Wang, Y. 1997b. Automatic Road Extraction from Aerial Images. In Proceeding of First International Workshop on Image Analysis and Information Fusion (H. Pan, M. Brooks, D. McMichael, and G. Newsam, Eds.), Adelaide, Australia, pp. 61-71.

Trinder, J. C. and Wang, Y. 1998a. Automatic 3-D Feature Extraction from Aerial Images. Paper #56 Proceeding of 9th Australian Conference of Remote Sensing and Photogrammetry (CDROM), Sydney.

Trinder, J. C. and Wang, Y. 1998b. Knowledge-Based Road Interpretation in Aerial Images. International Archives of Photogrammetry and Remote Sensing, **32**, 635-640.

Trinder, J. C. and Wang, Y. 1998c. Automatic Road Extraction from Aerial Images. Digital Signal Processing, Vol. 8, No. **4**, 215-224.

Trinder, J. C., Wang, Y., Sowmya, A. and Palhang, M. 1997. Artificial Intelligence in 3-D Feature Extraction. In Automatic Extraction of Man-Made Objects from Aerial and Space Images (II) (A. Grün, E. P. Baltsavias, O. Henricsson, Eds.), Birkhäuser Verlag, Basel, pp. 257-266.

Underwood, R. T. 1991. The Geometric Design of Roads. Macmillan Company of Australia Pty. Ltd., Melbourne.

Vasudevan, S., Cannon, R. L. and Bezdek, J. C. 1988. Heuristics for Intermediate Level Road Finding Algorithms. Computer Vision, Graphics, and Image Processing. **44**, 175-190.

Venkatesh, S. and Kitchen, L. J. 1992. Edge Evaluation Using Necessary Components. CVGIP: Graphical Models and Image Processing, **54**, 23-30.

Voorhees, H. and Poggio, T. 1987. Detecting Blobs as Textons in Natural Images. In DARPA Image Understanding Workshop, Los Angeles, California, pp. 892-899.

Vosselman, G. and Knecht, J. de. 1995. Road Tracing by Profile Matching and Kalman Filtering. In Automatic Extraction of Man-Made Objects from Aerial and Space Images (I) (A. Grün, O. Kübler, P. Agouris, Eds.), Birkhaüser Verlag, Basel, pp. 265-274.

Wang, J. F. and Howarth, P. J. 1987. Automatic Road Network Extraction from Landsat TM Imagery, In Proceeding of ASPRS-ACSM Annual Convention, Baltimore, USA, Vol. 1, pp. 429-438.

Wang, J. F., Treitz, P. and Howarth, P. J. 1992. Road Network Detection From SPOT Imagery for Updating Geographical Information Systems in the Rural-Urban Fringe. International Journal of Geographical Information Systems, **6**, 141-157.

Wang, Y. and Trinder, J. C. 1998. Use of Topology in Automatic Road Extraction. International Archives of Photogrammetry and Remote Sensing, **32**, 394-399.

Witkin, A. and Tenenbaum, J. M. 1983. On the Role of Structure in Vision. In Human and Machine Vision (J. Beck, B. Hope and A. Rosenfeld, Eds.) Academic, New York, pp. 481-453.

Zlotnick, A. and Carnine, P. D. 1993. Finding Road Seeds in Aerial Images. CVGIP Image Understanding, **57**, 243-260.

Zhu, M. L. and Yeh, P. S. 1986. Automatic Road Network Detection on Aerial Photographs. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 34-40.

Publications from the

SCHOOL OF GEOMATIC ENGINEERING

THE UNIVERSITY OF NEW SOUTH WALES
ABN 57 195 873 179

To order, write to:
Publications Officer, School of Geomatic Engineering
The University of New South Wales, UNSW SYDNEY NSW 2052, AUSTRALIA

NOTE: ALL ORDERS MUST BE PREPAID. CREDIT CARDS ARE ACCEPTED.
SEE BACKPAGE FOR OUR CREDIT CARD ORDER FORM.

MONOGRAPHS

Australian prices include postage by surface mail and GST.
Overseas prices include delivery by UNSW's air-lifted mail service (~2-4 weeks to Europe and North America).
Rates for air mail through Australia Post on application.

(Prices effective March 2001)

		Price Australia (incl. GST)	**Price** Overseas
M1.	R. S. Mather, "The Theory and Geodetic Use of some Common Projections", (2nd edition), 125 pp, 1978.	$ 16.50	$ 15.00
M2.	R. S. Mather, "The Analysis of the Earth's Gravity Field", 172 pp, 1971.	$ 8.80	$ 8.00
M3.	G. G. Bennett, "Tables for Prediction of Daylight Stars", 24 pp, 1974.	$ 5.50	$ 5.00
M4.	G. G. Bennett, J. G. Freislich & M. Maughan, "Star Prediction Tables for the Fixing of Position", 200 pp, 1974.	$ 8.80	$ 8.00
M8.	A. H. W. Kearsley, "Geodetic Surveying", 96 pp, 1988.	$ 13.20	$ 12.00
M11.	W. F. Caspary, "Concepts of Network and Deformation Analysis", 183 pp, 2000.	$ 27.50	$ 25.00
M12.	F. K. Brunner, "Atmospheric Effects on Geodetic Space Measurements", 110 pp, 1988.	$ 17.60	$ 16.00
M13.	B. R. Harvey, "Practical Least Squares and Statistics for Surveyors", (2nd edition, reprinted with corrections), 319 pp, 1998.	$ 33.00	$ 30.00
M14.	E. G. Masters and J. R. Pollard (Eds.), "Land Information Management", 269 pp, 1991. (Proceedings LIM Conference, July 1991).	$ 22.00	$ 20.00
M15/1	E. G. Masters and J. R. Pollard (Eds.), "Land Information Management - Geographic Information Systems - Advance Remote Sensing Vol. 1", 295 pp, 1993 (Proceedings of LIM & GIS Conference, July 1993).	$ 33.00	$ 30.00
M15/2	E. G. Masters and J. R. Pollard (Eds.), "Land Information Management - Geographic Information Systems - Advance Remote Sensing Vol. 2", 376 pp, 1993 (Proceedings of Advanced Remote Sensing Conference, July 1993).	$ 33.00	$ 30.00
M16.	A. Stolz, "An Introduction to Geodesy", 2nd extended edition, 148 pp, 2000.	$ 24.20	$ 22.00
M17.	C. Rizos, "Principles and Practice of GPS Surveying", 565 pp, 1997.	$ 46.20	$ 42.00

UNISURV REPORTS - S SERIES

(Prices effective March 2001)

Australian Prices *:	S8 - S20		$11.00
	S29 onwards	Individuals	$27.50
		Institutions	$33.00
Overseas Prices **:	S8 - S20		$10.00
	S29 onwards	Individuals	$25.00
		Institutions	$30.00

* Australian prices include postage by surface mail and GST.
** Overseas prices include delivery by UNSW's air-lifted mail service (~2-4 weeks to Europe and North America). Rates for air mail through Australia Post on application.

S8. A. Stolz, "Three-D Cartesian Co-ordinates of Part of the Australian Geodetic Network by the Use of Local Astronomic Vector Systems", Unisurv Rep. S8, 182 pp, 1972.

S14. E. G. Anderson, "The Effect of Topography on Solutions of Stokes` Problem", Unisurv Rep. S14, 252 pp, 1976.

S16. K. Bretreger, "Earth Tide Effects on Geodetic Observations", Unisurv S16, 173 pp, 1978.

S17. C. Rizos, "The Role of the Gravity Field in Sea Surface Topography Studies", Unisurv S17, 299 pp, 1980.

S18. B. C. Forster, "Some Measures of Urban Residential Quality from LANDSAT Multi-Spectral Data", Unisurv S18, 223 pp, 1981.

S19. R. Coleman, "A Geodetic Basis for Recovering Ocean Dynamic Information from Satellite Altimetry", Unisurv S19, 332 pp, 1981.

S29. G. S. Chisholm, "Integration of GPS into Hydrographic Survey Operations", Unisurv S29, 190 pp, 1987.

S30. G. A. Jeffress, "An Investigation of Doppler Satellite Positioning Multi-Station Software", Unisurv S30, 118 pp, 1987.

S31. J. Soetandi, "A Model for a Cadastral Land Information System for Indonesia", Unisurv S31, 168 pp, 1988.

S33. R. D. Holloway, "The Integration of GPS Heights into the Australian Height Datum", Unisurv S33, 151 pp, 1988.

S34. R. C. Mullin, "Data Update in a Land Information Network", Unisurv S34, 168 pp, 1988.

S35. B. Merminod, "The Use of Kalman Filters in GPS Navigation", Unisurv S35, 203 pp, 1989.

S36. A. R. Marshall, "Network Design and Optimisation in Close Range Photogrammetry", Unisurv S36, 249 pp, 1989.

S37. W. Jaroondhampinij, "A Model of Computerised Parcel-Based Land Information System for the Department of Lands, Thailand", Unisurv S37, 281 pp, 1989.

S38. C. Rizos (Ed.), D. B. Grant, A. Stolz, B. Merminod, C. C. Mazur, "Contributions to GPS Studies", Unisurv S38, 204 pp, 1990.

S39. C. Bosloper, "Multipath and GPS Short Periodic Components of the Time Variation of the Differential Dispersive Delay", Unisurv S39, 214 pp, 1990.

S40. J. M. Nolan, "Development of a Navigational System Utilizing the Global Positioning System in a Real Time, Differential Mode", Unisurv S40, 163 pp, 1990.

S41. R. T. Macleod, "The Resolution of Mean Sea Level Anomalies along the NSW Coastline Using the Global Positioning System", Unisurv S41, 278 pp, 1990.

S42. D. A. Kinlyside, "Densification Surveys in New South Wales - Coping with Distortions", Unisurv S42, 209 pp, 1992.

S43. A. H. W. Kearsley (ed.), Z. Ahmad, B. R. Harvey and A. Kasenda, "Contributions to Geoid Evaluations and GPS Heighting", Unisurv S43, 209 pp, 1993.

S44. P. Tregoning, "GPS Measurements in the Australian and Indonesian Regions (1989-1993)", Unisurv S44, 134 + xiii pp, 1996.

S45. W.-X. Fu, "A Study of GPS and Other Navigation Systems for High Precision Navigation and Attitude Determinations", Unisurv S45, 332 pp, 1996.

S46. P. Morgan et al, "A Zero Order GPS Network for the Australia Region", Unisurv S46, 187 + xii pp, 1996.

S47. Y. Huang, "A Digital Photogrammetry System for Industrial Monitoring", Unisurv S47, 145 + xiv pp, 1997.

S48. K. Mobbs, "Tectonic Interpretation of the Papua New Guinea Region from Repeat Satellite Measurements", Unisurv S48, 256 + xc pp, 1997.

S49. S. Han, "Carrier Phase-Based Long-Range GPS Kinematic Positioning", Unisurv S49, 185 + xi pp, 1997.

S50. M. D. Subari, "Low-cost GPS Systems for Intermediate Surveying and Mapping Accuracy Applications", Unisurv S50, 179 + xiii pp, 1997.

S51. L.-S. Lin, "Real-Time Estimation of Ionospheric Delay Using GPS Measurements", Unisurv S51, 199 + xix pp, 1997.

S53. D. B. Lemon, "The Nature and Management of Positional Relationships within a Local Government Geographic Information System", Unisurv S53, 273 + xvi pp, 1997.

S54. C. Ticehurst, "Development of Models for Monitoring the Urban Environment Using Radar Remote Sensing", Unisurv S54, 282 + xix pp, 1998.

S55. S. S. Boey, "A Model for Establishing the Legal Traceability of GPS Measurements for Cadastral Surveying in Australia", Unisurv S55, 186 + xi pp, 1999.

S56. P. Morgan and M. Pearse, "A First-Order Network for New Zealand", Unisurv S56, 134 + x pp, 1999.

S57. P. N. Tiangco, "A Multi-Parameter Radar Approach to Stand Structure and Forest Biomass Estimation", Unisurv S57, 319 + xxii pp, 2000.

S58. M. A. Syafi'i, "Object–Relational Database Management Systems (ORDBMS) for Managing Marine Spatial Data: ADCP Data Case Study", Unisurv S58, 123 + ix pp, 2000.

S59. X.-Q. Lu, "Strategies for Improving the Determination of Displacements of Sea Surface Temperature Patterns Using Consecutive AVHRR Thermal Images", Unisurv S59, 209 + xiii pp, 2000.

S60. G. Dickson, "GPS-Controlled Photography: The Design, Development and Evaluation of an Operational System Utilising Long-Range Kinematic GPS", Unisurv S60, 417 + x pp, 2000.

S61. J. Wang, "Modelling and Quality Control for Precise GPS and GLONASS Satellite Positioning", Unisurv S61, 171 + x pp, 2001.

S62. Y. Wang, "Knowledge-Based Road Extraction from Aerial Images", Unisurv S62, 178 + xi pp, 2001.

Credit Card Orders / Payment

from / to

School of Geomatic Engineering

The University of New South Wales

Name and postal address of ordering person/organisation:

_____

_____

_____

_____

_____

Date : _____

Day time phone number : _____

Credit Card debit for:

_____ $ _____

_____ $ _____

_____ $ _____

_____ $ _____

Total $ _____

Please debit my credit card: ☐ Mastercard ☐ Visa ☐ Bankcard

Card Name: _____ Signature: _____

Card No: | | | | | | | | | | | | | | | | Expiry Date : _____

Note: UNSW requires that you attach a photocopy of your credit card to this order.