Location of roads in high-resolution, panchromatic satellite images of urban areas

Kalyan Takru

2005

Kalyan, T. (2005). Location of roads in high-resolution, panchromatic satellite images of urban areas. Master's thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/2669>

<https://doi.org/10.32657/10356/2669>

Nanyang Technological University

Downloaded on 26 Apr 2023 22:55:26 SGT

Location of Roads in High-Resolution, Panchromatic Satellite Images of Urban Areas

Kalyan Takru

School of Computer Engineering

A thesis submitted to the Nanyang Technological University in fulfillment of the
requirement for the degree of Master of Engineering

2005

ACKNOWLEDEMENTS

Firstly, I would like to express my sincere gratitude to my Supervisor, Associate Professor Charles Graham Leedham for his support. He allowed me the freedom to find my own way and at the same time, was always there to assist and advise me when I faced problems.

I would very much like to thank Assistant Professor Timo Rolf Bretschneider for the guidance he provided me in the fascinating field of satellite image analysis.

I must also thank Mr. Tan, the Laboratory Technician at the Intelligent Systems Laboratory for his help.

Finally, my appreciation for all those people who helped me along the way in one way or another.

Table of Contents

CHAPTER 1: INTRODUCTION	1
1.1 RESEARCH FOCUS	2
1.2 SATELLITE IMAGING	2
1.2.1 History of Satellite Imaging	2
1.2.2 Types of Sensors	3
1.3 OBJECTIVE OF THIS RESEARCH	4
1.4 CONTRIBUTIONS AND PUBLICATIONS	4
1.5 STRUCTURE OF THESIS	6
CHAPTER 2: REVIEW OF LITERATURE ON ROAD DETECTION	8
2.1 REVIEW OF PUBLICATIONS	10
2.1.1 Road detection in panchromatic SPOT satellite images	10
2.1.2 Efficient algorithm for detection of road-like structures in satellite images	12
2.1.3 Robust detection of road segments in noisy aerial images	13
2.1.4 Vehicle detection on aerial images: a structural approach	13
2.1.5 Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation	14
2.1.6 An image analysis system, application for aerial imagery interpretation	15
2.1.7 Automatic road extraction using fuzzy mask concepts	16
2.1.8 Automated detection of road intersections from ERS-1 SAR imagery	17
2.1.9 Road detection in spaceborne SAR images using a genetic algorithm	18
2.1.10 Application of spatial reasoning methods to the extraction of roads from high resolution satellite imagery.	19
2.1.11 The line segment match method for extracting road network from high-resolution satellite images	21
2.1.12 A wavelet transform based method for road extraction from high-resolution remotely sensed data	23
2.1.13 Road network extraction from airborne digital camera images: a multi-resolution comparison	24
2.1.14 The research of road extraction for high resolution satellite image	25
2.1.15 Towards knowledge-based extraction of roads from 1m resolution satellite images	26
2.1.16 Detection and extraction of road networks from high resolution satellite images	27
2.2 OVERVIEW AND CONCLUSIONS	27
2.2.1 Overview	27
2.2.2 Conclusions	31
CHAPTER 3: STUDY ON DETECTION OF ROADS IN HIGH-RESOLUTION SATELLITE IMAGES	34
3.1 ROAD MAP DETECTION ALGORITHM (STAGE ONE)	34
3.1.1 Thresholding	35
3.1.2 Parse and Merge	38
3.1.3 Parse and Remove	41
3.1.4 Connected Component Labeling/Analysis	43
3.1.5 Removal of Small Components	44
3.1.6 Removal of Dense Components	45
3.1.7 Removal of Irregular Components	49
CHAPTER 4: RESULTS OF ROAD DETECTION (STAGE ONE)	54
4.1 TYPICAL FAILURES	58
4.1.1 Loss Due to Thresholding	58
4.1.2 Loss Due to Concealment	59

CHAPTER 5: STUDY ON EXTRAPOLATION OF INCOMPLETE ROAD NETWORKS	63
5.1 ROAD NETWORK EXTRAPOLATION ALGORITHM (STAGE TWO)	63
5.1.1 Extraction of Initial End-point Candidates	64
5.1.2 Properties of the Existing Road	66
5.1.3 Extrapolation of End-points	67
5.1.4 Detection of Intersection	69
5.1.5 Subsequent Iterations	70
5.2 RESULTS AND DISCUSSION (STAGE TWO)	72
CHAPTER 6: STUDY ON CONVERSION OF SKELETON TO ROAD MAP: DETERMINING THE SEALED SURFACE	74
6.1 ALGORITHM FOR CONVERSION OF SKELETON TO ROAD MAP (STAGE THREE)	75
6.1.1 Extraction of Road Orientation	76
6.1.2 Estimation of Road Width	76
6.1.3 Multiple Cross-Section Extraction	77
6.1.4 Variance Extraction	78
6.1.5 Interpolation of Road Surface	80
6.2 RESULTS (STAGE THREE)	83
CHAPTER 7: CONCLUSIONS AND FUTURE WORK	85
7.1 SUMMARY OF WORK DONE	88
7.1.1 Stage One	87
7.1.2 Stage Two	89
7.1.3 Stage Three	92
7.2 CRITICAL REVIEW	93
7.3 ALTERNATIVE APPROACHES	96
7.4 FUTURE WORK	97
REFERENCES	100

APPENDIX A: ROAD EXTRACTION TEST RESULTS: ORIGINAL IMAGES AND FINAL RESULTS	111
--	------------

APPENDIX B: Document Image Analysis

CHAPTER B1: INTRODUCTION TO DOCUMENT ANALYSIS	114
B1.1 BACKGROUND	114
B1.2 DOCUMENT ANALYSIS AND UNDERSTANDING	114
B1.2.1 Practical Issues	115
B1.2.2 Document Pre-processing	116
B1.2.3 Research Focus and Structure of Report	117
CHAPTER B2: REVIEW OF LITERATURE ON DOCUMENT SEGMENTATION AND THRESHOLDING	118
B2.1 DOCUMENT SEGMENTATION TECHNIQUES	118
B2.1.1 Texture Analysis	118
B2.1.2 Mathematical Morphology	120
B2.1.3 Voronoi Diagrams	120
B2.1.4 Hough Transform	121
B2.1.5 Projection Profile	121

B2.1.6 Connected Component Analysis/Labeling (CCA)	122
B2.2 DOCUMENT THRESHOLDING TECHNIQUES	122
B2.2.1 Global Methods	123
B2.2.2 Local Methods	124
CHAPTER B3: OVERVIEW AND CONCLUSIONS FOR DOCUMENT PRE-PROCESSING	126
B3.1 DOCUMENT SEGMENTATION	126
B3.1.1 Methodologies	126
B3.1.2 Nature of the Problem	127
B3.2 THRESHOLDING	128
B3.2.1 Methodologies	128
B3.2.2 Nature of the Problem	128
B3.3 TOPIC FOR INVESTIGATION	129
CHAPTER B4: STUDY ON SKEW CORRECTION	131
B4.1 IMPLEMENTATION	131
B4.1.1 Hough Transform	131
B4.1.2 Algorithm and Illustration	134
B4.2 RESULTS	137

List of Figures and Tables

Figure 3-1: An original high-resolution greyscale image.	35
Figure 3-2: Thresholded image. Black areas represent foreground.	37
Figure 3-3: Zoomed in portion of an image showing details of the road-lane dividers.	39
Figure 3-4: The same portion as Figure 3-3 in binary form, showing the lane dividers.	40
Figure 3-5: The effect of the Parse and Merge technique. The lane dividers have been removed.	41
Figure 3-6: The effect of Parse and Remove. Compare this image to the one in Figure 3-2.	42
Figure 3-7: All the components from Figure 3-6 are marked by rectangles, demonstrating the connected component labeling. Now, analysis can be carried out based on these components.	43
Figure 3-8: The binary image after removal of small components.	45
Figure 3-9: The binary image after removal of dense components.	46
Figure 3-10: A comparison of the original to the result at this stage of the algorithm.	48
Figure 3-11: Zoomed in view showing the skeleton of a component from Figure 3-9.	51
Figure 3-12: The binary image after removal of irregular components.	52
Figure 3-13: The final image after further removal of remaining small and/or dense components.	53
Table 4-1: Table showing statistics of road length extracted and lost (using both measures).	55
Table 4-2: Table showing the statistics from Table 4-1 in terms of percentages.	57
Figure 4-1: A section of road obscured by foliage and shadows.	60
Figure 4-2: Section of thresholded image showing loss of road area due to obscuring (compare with Figure 4-1).	60
Figure 4-3: Final image showing the fragmentation caused by the obscured area.	61
Figure 5-1: The skeleton of the extracted road map after a few rounds of pruning.	64
Figure 5-2: Image after one iteration on Figure 5-1.	71
Figure 5-3: Final extrapolated road skeleton network.	71
Table 5-1: Results of the extrapolation (expressed as a percentage of total road skeleton length; The sum of the second, third and fourth columns represents the total road length)	72
Figure 6-1: Original scene with superimposed extrapolated road network	75
Figure 6-2: Cross-sections at two points of the skeleton	77
Figure 6-3: Multiple cross-sections for pixel used in Figure 6-2.	78
Figure 6-4: Plot of variance versus position	79
Figure 6-5: Result of road surface determination using the road network in Figure 6-1.	80
Figure 6-6: Result obtained after interpolation.	82

Table 6-1: Results of the extraction (expressed as a percentage of total road length).....	83
Figure 7-1: Diagrammatic representation of entire road extraction method (3 stages/algorithms)	86
Figure 7-2: Input image and final result for stage one.....	87
Figure 7-3: Input image, pruned skeleton and extrapolated final image for stage two	89
Figure 7-4: Input image and final result for stage three.....	92
Figure B4-1: Illustration of equations (13) and (14).....	132
Figure B4-2 Conversion from x-y space to Hough space.....	133
Figure B4-3: Adjustment of two points skewed with respect to each other.	135
Figure B4-4: Result of adjustment of one document image. The COGs shift from the crosses to the circles.....	137
Table B4-1: Table comparing execution times with and without Hough transform.	138

ABSTRACT

This thesis contributes to the research carried out in the field of high-resolution satellite image analysis by presenting original work on road feature extraction from panchromatic (greyscale) satellite images of urban areas with a spatial resolution of 60cm per pixel. High resolution satellite images have only recently become commercially available and they differ significantly from lower resolution images in their quality. They are extremely detailed, allowing one to see terrestrial features, such as roads, with clarity never before possible. Extraction of features from high-resolution images is a new challenge and requires a completely different approach as compared to low-resolution images.

The objective is to create an automated method to extract a high percentage of the roads from the images while minimizing noise, thus ensuring high correctness. Towards this end, three separate algorithms were created and are presented as three separate studies. Each one deals with a different portion of the extraction problem. The image resulting from the application of the first stage algorithm to the original panchromatic image is used as the input for the second stage algorithm and the result from the second stage is used as the input for the third stage. Taken together, the algorithms form a complete method for the detection and extraction of roads from high-resolution images with high precision and recall. They incorporate both low-level and high-level techniques. The images used for the tuning and testing of the algorithms have a spatial resolution of 60cm per pixel and are obtained from the QuickBird satellite. The final result is the extracted road map with an average completeness of 90.9% and a correctness of 92.1%, representing a reasonably successful attainment of the stated objective.

Chapter 1: Introduction

The research reported in this thesis revolves around the processing of digital images. The fundamental objective of any image processing is one of the following:

- To obtain an enhanced image that is more visually accessible to the human eye
- To extract information automatically from the image for image understanding and recognition
- To compress the image for transmission and later reconstruction

The processing requires the use of mathematical operations. These operations can be based on an n -by- n pixel kernel, morphology, thresholding and other more complex techniques. A wide variety of methods are available at the present moment for enhancing a digital image and obtaining useful information from it. Research into this area was initially funded by governments but now, the private sector has made leaps and bounds in image processing technology. Extremely sophisticated image processing software is now available to the novice user.

Initial image processing software and techniques were simplistic as the hardware was primitive and requirements were minimal. The images were few, of poor resolution and the information that could be obtained did not have much utility. Thus, there was no need for better algorithms. However, the progress in image processing hardware and a corresponding increase in demand for useful information that could be extracted from the images spurred a rapid improvement in techniques. For example, with the huge volume of spy satellite data being generated, automation was required to sift through thousands of images daily, clearly not a task suitable for humans. Earlier, there were comparatively few images as the images were generated by spy planes that only made infrequent trips,

Introduction

being constrained by fuel. With the advent of space travel and the realization that such imagery was potentially very useful (the Cuban missile crisis is an example), the supply of images jumped and correspondingly, so did the sophistication of methods used to analyze the images.

1.1 Research Focus

The topic of this thesis is the analysis of high-resolution panchromatic (greyscale) satellite images of urban areas for the purpose of identification and extraction of road features. Some comparatively minor work was also done on the analysis of document images for the purposes of segmentation/thresholding. That work is attached as Appendix B.

1.2 Satellite Imaging

Remote Sensing involves gathering data and information about the physical world by detecting and measuring radiation, particles, and fields associated with objects located beyond the immediate vicinity of the sensor device(s). A slightly more restrictive definition would be that Remote Sensing is a technology for sampling electromagnetic radiation to acquire and interpret non-immediate *geospatial data* from which to extract information about features, objects, and classes on the Earth's land surface, oceans, and atmosphere.

1.2.1 History of Satellite Imaging

The technology of modern remote sensing began with the invention of the camera more than 150 years ago. Although the first, rather primitive photographs were taken as "stills" on the ground, the idea and practice of looking down at the Earth's surface emerged in the 1840s when pictures were taken from cameras secured to tethered balloons for purposes of topographic mapping. By the First World War, cameras mounted on airplanes provided aerial views of fairly large surface areas that proved invaluable in

Introduction

military reconnaissance. From then until the early 1960s, the aerial photograph remained the single standard tool for depicting the surface from a vertical or oblique perspective.

Satellite remote sensing can be traced to the early days of the space age (both Russian and American programs) and actually began as a dual approach to imaging surfaces using several types of sensors from spacecraft. In 1946, V-2 rockets acquired from Germany after World War II were launched to high altitudes from White Sands, New Mexico. These rockets, while never attaining orbit, contained automated still or movie cameras that took pictures as the vehicle ascended. Then, with the emergence of the space program in the 1960s, Earth-orbiting cosmonauts and astronauts acted much like tourists by taking photos out of the window of their spacecraft.

The term "remote sensing," first used in the United States in the 1950s by Ms. Evelyn Pruitt of the U.S. Office of Naval Research, is now commonly used to describe the science—and art—of identifying, observing, and measuring an object without coming into direct contact with it. This process involves the detection and measurement of radiation of different wavelengths reflected or emitted from distant objects or materials, by which they may be identified and categorized by class/type, substance, and spatial distribution.

1.2.2 Types of Sensors

There are two types of remote sensing instruments—passive and active. Passive instruments detect natural energy that is reflected or emitted from the observed scene. Passive instruments sense only radiation emitted by the object being viewed or reflected by the object from a source other than the instrument. Reflected sunlight is the most common external source of radiation sensed by passive instruments. Active instruments provide their own energy (electromagnetic radiation) to illuminate the object or scene

they observe. They send a pulse of energy from the sensor to the object and then receive the radiation that is reflected or backscattered from that object.

1.3 Objective of this Research

The objective of the research reported in this thesis is the creation of a method to carry out detection and extraction of terrestrial features, specifically, roads, from satellite imagery of urban areas. The images chosen are high-resolution panchromatic (greyscale) images which have only recently become commercially available and differ significantly from low resolution images that have been used so far. The objective is to find road features within these high-resolution satellite images based on observations of typical properties, pattern-based operations, statistical analysis and high-level methods. Due to the high resolution, non-road features are seen in great detail and have to be carefully separated from the road features. This was not an issue when dealing with low resolution images as non-road features were obscured due to the low level of detail and road features stood out quite clearly. Another effect of the high resolution is that the roads themselves possess greater variations in intensity values, making it more difficult to identify road pixels than in the case of low resolution images where the roads generally have pretty similar intensity values. These are some of the challenges faced in the extraction of roads from high-resolution satellite images.

1.4 Contributions and Publications

Studies were carried out on the topic of satellite image analysis and form the body of this thesis. Some minor work on document image analysis was also done and is attached as Appendix B. In the case of satellite imaging, algorithms were created to extract road features from high-resolution satellite imagery. Such an algorithm has applications in cartography and can also be used onboard satellites in order to extract relevant information in real time from captured images and transmit it back to earth. This saves

Introduction

bandwidth and translates into cost savings when compared to a system that does no on-board processing and only transmits all captured data back to a terrestrial station. The images used in this study were from the QuickBird satellite.

The work on document imaging in Appendix B consists of an algorithm to correct skew in a page of handwritten text. The final objective is to utilize the skew correction in order to better segment written text into lines and words.

Three conference papers have been published/accepted based on the work on road extraction:

1. Kalyan Takru, Timo Bretschneider & Graham Leedham, *Unsupervised Detection of Roads in High Resolution Panchromatic Satellite Images*, Proceedings of the International Workshop on Advanced Image Technology (IWAIT'04), January 12-13, 2004 Singapore, pp. 15-19.
2. Kalyan Takru, Timo Bretschneider & Graham Leedham, *Automatic extraction of road maps: determining the sealed surfaces*, Accepted for presentation at the International Geoscience and Remote Sensing Symposium (IGARSS), Alaska, 20-24 September 2004.
3. Kalyan Takru, Timo Rolf Bretschneider, Charles Graham Leedham, *Extrapolation of incomplete road networks in high resolution satellite images*, Proceedings of the Asian Conference on Remote Sensing, Volume 2, pp. 1568-1573, 2004

A journal paper has been submitted based on the overall work on road extraction.

4. Kalyan Takru, Timo Bretschneider & Graham Leedham, *Unsupervised extraction of roads in high resolution panchromatic satellite images*,

Introduction

Submitted to the IEEE Transactions on Geoscience and Remote Sensing,
July 2004.

A conference paper on comparison of document thresholding techniques was published in collaboration with three other students.

5. Graham Leedham, Chen Yan, Kalyan Takru, Joie Hadi Nata Tan & Li Mian, *Comparison of some thresholding algorithms for text/background segmentation in difficult document images*, Proceedings of the 7th International Conference of Document Analysis and Recognition, 3-6 August 2003, Edinburgh, Scotland, pp. 859-863

A journal paper for submission to Pattern Recognition Letters, based on the work on text segmentation is currently being prepared.

6. Graham Leedham, Wu Xiaoying and Kalyan Takru, *Line and word segmentation in unconstrained handwriting*, In preparation for Pattern Recognition Letters.

1.5 Structure of Thesis

The literature review (Chapter 2) is sub-divided into a review of publications (Section 2.1), an overview of the research (Section 2.2.1) and the conclusions drawn from the research (Section 2.2.2). The main body of this thesis deals with three separate studies on satellite image analysis. The details of each individual study are found separately in Chapters 3, 5 and 6. Chapter 4 contains the results for the study in Chapter 3. Chapter 7 is the conclusion of the report. It is divided into a summary of the work done, a critical review of the same, possible alternative approaches and some suggestions for future work. The list of references follows Chapter 7. At the end of the thesis, two appendices

Introduction

are included. Appendix A contains the panchromatic images used in the evaluation reported in Chapter 3-6. Appendix B consists of a short report on related work in the area of Document Analysis. This research was also carried out during the M. Eng. candidature and is included for completeness.

Chapter 2: Review of Literature on Road Detection

The literature on road detection can be classified in two different ways. The first classification method is based on the resolution of the image, low resolution or high resolution. In this chapter, literature on both types is reviewed but the thesis itself deals with high-resolution images from the QuickBird satellite. The analysis of the methods for extraction of roads from low-resolution images helps us understand the methodology in use before high-resolution images became available. Since low-resolution images have existed for a longer time, there is a larger body of work on it and a great deal can be learnt from it. However, this knowledge is not directly applicable to high-resolution image analysis. Thus, the analysis of techniques used for extraction of roads from low-resolution images is included in order to underline the difference between those methods and the methods used in this thesis. Not all of the work on low-resolution images is reviewed in detail, but certain publications that offer insight into the methodologies are included in the following analysis. Though the method developed in this thesis is different than that used for low-resolution images, a basic understanding was considered essential. Duta (2000), Mukherjee et al. (1996), Netanyahu et al. (1996), Barzohar et al. (1993), Solaiman et al. (1998), Geman and Jedynak (1991), Gervin and Ragan (1993), Wang and Newkirk (1988), Garnesson et al. (1990), Heutte et al. (1992), Xuan and Adali (1996) and Lin et al. (2000) comprise some of the work dealing with low-resolution images. Within the area of low resolution images, there is also related work on detection of features associated with roads, for example, vehicle detection as in Ruskone (1996). The concept is that the detection of an aggregation of vehicles can be used as a guide/indicator for the detection of roads.

It must be noted that the term ‘high-resolution satellite image’, when used in a piece of work, can only be understood in the specific context of the time period of that particular work. As an example, in Geman and Jedynak (1991) and Wang and Newkirk (1988), the authors refer to images from the SPOT satellite as high-resolution even though the resolution (about 10m per pixel) is fairly low by the standards of today’s imaging technology. But, a decade and a half back, 10m resolution images were considered high-resolution. Thus, to put the work in this thesis in the proper context, ‘high-resolution’ refers to images with a resolution of less than 1m per pixel. Any work on images with a resolution of 10m or greater is thus referred to as low-resolution in this thesis. Even 2-3m resolution images are relatively lower resolution than the images used in this thesis.

There is some existing work on road extraction from high-resolution images but most of it deals with lower resolutions than those used in this thesis. More importantly, the approaches do not produce satisfactory results, as will be clear from the detailed analysis of the relevant papers in Section 2.1. Guindon (1998), Wenzhong Shi and Changqing Zhu (2002), Gong and Wang (1997), Chen et al. (2002), Lee et al. (2000), Peteri et al. (2003) and Xuan et al. (2003) deal with road detection in images with a resolution 1-3m. Karimi et al. (1998) provides a summary of all methods in use. Also described is a scheme developed for classification of surface material in high-resolution images. The scheme comprises four levels, each level corresponding to a different level of precision in identifying the surface material.

The second method of classifying the literature on road detection is based on the type of sensor used to obtain the image. The sensor can be passive, such as those which generate optical images or it can be active, such as those that generate SAR (Synthetic Aperture Radar) images. This thesis deals with images obtained in the visible wavelength range (optical) and accordingly, most of the literature reviewed deals with optical images, but

some work on SAR imagery is included for the sake of a complete picture. Also, it is informative as there are certain useful road extraction techniques in such work. Iisaka, J. et al. (1995), Tupin et al. (1998), Dell'Acqua and Gamba (2001), Huber and Lang (2001), Byoung-Ki Jeon et al. (2002) and Tupin et al. (2002) deal specifically with road detection in SAR imagery.

From among each sub-group (based on resolution and/or sensor type) of papers on road detection, some representative papers are analyzed in detail below. The papers representative of each section are chosen based on their methodologies and the innovativeness of their proposed ideas. Wherever quantitative results are presented (quite a large percentage of the papers do not provide them), a high success rate (in terms of percentage of road detected) is a criterion for inclusion in the detailed analysis. The approach used here is to analyze each paper individually and then summarize the findings into a general overview and relevant conclusions after the analyses. The overview is presented in Section 2.2 while the individual summaries of the papers are given below in Section 2.1.

2.1 Review of Publications

2.1.1 Road detection in panchromatic SPOT satellite images

Duta (2000)

This paper describes an automatic procedure for road detection which has the following advantages: it does not require manual initialization; it detects some of the secondary roads also and is fast. Based on a study of typical satellite images of areas with roads, the author listed the defining characteristics of the image of a road. These properties are piecewise linearity, connectivity, homogeneity, separability from background and length.

The first step is to enhance the contrast by equalizing the histogram. Histogram equalization is a point process in which existing values of intensity are mapped to new values but the actual number of intensities in the resulting image is equal to or less than the original number of intensities. The end result is a redistribution of intensities leading to higher contrast. Equalization also allows for fixed values of thresholds for different images as the final histogram is more uniformly spread. The second step is to locate road segments. Road segments are composed of cross-sections placed end to end which are roughly perpendicular to the general direction of the segment. Some spatial limits are placed on the size of the cross-sections and the length of the segment itself as all images used are of a similar resolution and quality.

The initial cross-section is obtained by using a clustering algorithm, which takes into account the fact that the cluster is a 1-D structure. Subsequent scanning of the adjacent rows/columns leads to the next cluster. For clusters from adjacent rows/columns to be considered part of the same segment, the displacement should not be above a certain threshold. Once the displacement exceeds the threshold, that segment ends there and from its endpoints, a tree is constructed of possible segments that continue the road onwards. Thus, the segment is used as a road-seed. The lengths of these candidate segments are approximately equal to the segment found. The number of branches of the tree depends upon the observed degree of curvature of the roads in the image.

From the tree, the most likely candidate for a segment is chosen based on the criteria of homogeneity of intensity values and contrast with the background. A filter is used to calculate the contrast value and thus, confirm or deny the likelihood of it being a valid segment. The output of the filter is a probability value. Further extensions of the road (in tree form) carry on until the road reaches an edge of the image or merges with an extension of some other segment used as a road-seed. The final road map is an

aggregation of all extensions. The average accuracy of the method is reported as 90%. It is calculated as the ratio of the number of well labeled road segments to the number of manually tracked and classified segments. A segment is considered well labeled if it is at most 3 pixels away from a manually tracked road.

2.1.2 Efficient algorithm for detection of road-like structures in satellite images

Mukherjee et al. (1996)

This paper presents an effective enhancement technique and a segmentation technique that removes non-road pixels from the image, step by step. The segmentation process depends upon road length and contrast. Like the above paper, the authors, in this case too, made certain observations of the characteristics of the images beforehand. These properties assist in the creation and subsequent refinement of the techniques. The observations include facts about road thickness, linearity, intensity values, homogeneity and length. Thus, we observe that both papers employ similar methodologies. This paper used SPOT satellite images also along with IRS images.

The first step is enhancement in which line structures get enhanced and edges are suppressed. Two operations are used for this step. The first operation produces an enhanced image based on the above observations and the second operation reduces the contrast of pixels with high heterogeneity measures.

The second step is segmentation, which consists of three elimination processes. First, thresholds are set for the intensity values of roads. Any pixels lying outside the range are eliminated. Secondly, low contrast pixels are eliminated. This resultant image is then binarized using contrast and length of linear structures. The linear structures are obtained through analysis and subsequent morphological operations.

2.1.3 Robust detection of road segments in noisy aerial images

Netanyahu et al. (1996)

Method: The finding of roads is resolved into the simpler problem of finding straight or circular *pieces* of road. The assumption behind this approach is based on the knowledge of highway engineering practices which specify that on level ground, roads are piecewise straight, with the straight segments joined by curved pieces. The curves, except for brief transitions, are essentially circular arcs.

A local non-linear operator is used to detect pixels that are locally line-like and then robust estimation techniques are applied to find sets of pixels that lie on or near straight or circular loci. The authors explain that ordinary straight line or circular arc fitting is not adequate because of the presence of outliers, which severely perturb the fit. If it was known in advance that all segments are straight, then they can be detected using various robust estimators, which *reject outliers*. However, this approach will not handle both straight and curved segments at the same time.

Curved road segments generally curve very gently, having a large radius. A robust straight line fit will accept such an arc by rejecting certain parts which lie outside the straight line. On the other hand, a conventional robust circular arc estimator will not detect such an arc well because the arc will contain relatively few pixels and robust estimators are statistically inefficient. The authors have used a novel and statistically efficient method of robust estimation.

2.1.4 Vehicle detection on aerial images: a structural approach

Ruskone et al. (1996)

Method: Generally, the recognition of vehicles is made through the texture they engender and not through the agglomeration of isolated vehicles. In this paper, a

hierarchical model is used which has pixels as the fundamental level, then vehicles, lines and finally meta-lines. The recognition method is based on this hierarchy. The actual technique used is neural networks, which allows general classification.

The hierarchy is based on observations of general images. Vehicles are composed of pixels and the vehicles, in turn, form lines. For purposes of this paper, a line is defined as at least two identically oriented vehicles. Thus, a car line with the same orientation as its elements suggests a road and dissimilar orientation suggests a car park. Two car lines with relative orientations of 0, 45 or 90 degrees form a meta-line.

The first step is to identify and cluster pixels likely to belong to cars and then secondly, validate the presence of pixels by looking for a well-defined organization of the vehicles found. The neural network used accepts as input an original image and returns an image where each pixel value is proportional to its probability of belonging to a vehicle. Thus, the network basically defines a limit between the two sets of pixels, those which belong to cars and those that don't.

The result of the neural network is not completely correct and thus validation is required. It is carried out by forming lines and meta-lines. If a meta-line is successfully formed by pixels that have a high probability of belonging to a vehicle, it confirms the presence of vehicles. Thus, while the detection stage is equivalent to a human looking for pixel patterns that seem to belong to a car, the validation step is equivalent to a human looking for contextual clues to differentiate between the top view of a car and for example, a lift exit on a roof, both of which look similar in the image. Using this unique approach, around 87% of all vehicles are detected.

2.1.5 Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation

Barzohar et al. (1993)

Method: The approach used in this paper is to build geometric-probabilistic models for road image generation using Gibbs distribution. Given an image, roads are found by estimation.

The problem considered is the automatic extraction of main roads when road curvature, width, image intensity and edge strength can vary considerably and when a barrier along the road center may or may not be present. The modeling approach forces the designer to focus on all significant phenomena. The assumptions used for the model are that road width variance is small and width variation is gradual, direction changes are gradual, road pixels' intensity variation is gradual, there is a large intensity difference between road and background and roads (being main roads) are likely to be long.

The problem of road finding is the estimation of the road geometry by formulating the problem as map estimation. The image is divided into square windows and the system searches for road candidates that fit the road model with high probability. The road search is run four times in every window, with one search starting at each side. All the road candidates in a window are thus found. A high level processing stage extends the candidates into complete roads using shifting windows. An extension window is abutted to the side of the present window from which the road is exiting. The process is repeated until the border is reached or other criteria end the search. If the length of the final estimated road is greater than a threshold value, it is accepted.

2.1.6 An image analysis system, application for aerial imagery interpretation

Garnesson et al. (1990)

Method: This paper presents an image analysis system that reasons from generic models. The system, called MESSIE, detects salient objects and then using knowledge of those

objects and their context, it confirms or rejects object hypothesis and tries to infer new objects. The domain used to develop the system is aerial imagery interpretation.

The strategy is to begin search with low-level primitives for objects using image data. Identification of some objects and their spatial relationships allows the generation of relevant hypotheses which are validated by returning to low-level search, while making use of context. More objects are found through hypothesis testing and the process repeats until no more objects are generated.

MESSIE has three parts. The localization controller uses the context of objects to build a plan to find other objects. It analyses relations between discovered objects, reasons at a semantic level and generates goals to be further validated. The scene controller manages objects found so far in the image and accepts requests to find objects. The requests are made by the supervisor which attempts to find objects in certain limited areas.

The system works on blackboard architecture. Currently, there are four separate specialists. Each contains knowledge needed for different purposes: detection of roads, detection of buildings, detection of shadows and detection of cars.

2.1.7 Automatic road extraction using fuzzy mask concepts

Solaiman et al. (1998)

Method: In this paper, fuzzy concepts are used in order to realize road pixels extraction. The aim of the proposed algorithm is to attribute a road membership value to each pixel. A set of twelve fuzzy sets representing basic 2D road structures is first defined. The road structure, defined as a logical concatenation of these elementary predefined 2D structures, is then used to enhance the values.

A fuzzy mask aims at detecting a particular pattern of neighboring pixels. At a local spatial context level, a road is defined as a set of concatenated bright pixels comprising a particular road pattern configuration. The membership value of a given pixel to each of the twelve masks is calculated as the minimum darkness and brightness values of its constituent pixels. The membership value of a pixel to the fuzzy set is computed as the maximum membership of this pixel to the twelve fuzzy masks.

If a pixel is considered as a road pixel with a membership value to the fuzzy mask M1, then its neighboring pixels from M1 are restricted in terms of which masks they can belong to by the configuration of the pixels in M1. Thus, the possible masks can be determined and the optimal mask can be chosen for each neighboring pixel. This leads to the determination of the road segment configuration.

2.1.8 Automated detection of road intersections from ERS-1 SAR imagery

Iisaka, J. et al. (1995)

Method: Locations of road intersections were estimated automatically from single look SAR images taken by the European Remote Sensing Satellite 1 (ERS-1). A Hough Transform based approach was used. The algorithm has two parts. The first part is to apply a despeckling filter, which preserves small features in the SAR images, and the second part consists of methods to detect peaks in a Hough space. Several such methods were developed.

Ground control points in an SAR image are usually selected from point-like features such as road intersections, branch points of rivers etc. However, such points are very difficult to determine in a SAR image. The main reason is the presence of speckle noise. The other reasons are that features in a SAR image may not appear in the image because of sensitivity to the radar imaging direction and effects of the object environment. Also,

object shapes are less instantly recognizable in a SAR image and terrain feature classes do not have unique intensity ranges, thus causing overlap among different classes.

The authors chose to concentrate on road intersections as they have better characteristics as a ground control point. Although the intersection itself may not appear in the image, the location of the intersection can be obtained by extrapolating or interpolating segments of the intersecting roads that appear near the intersection point. The Hough transform can handle globally large line-like features in theory. In this paper, several practical solutions to the difficulties inherent in a Hough transform based approach were developed in order to detect road intersections in noisy SAR images.

The Hough transform is a coordinate transformation in which a straight line in the x-y plane is projected to a single point in the R- θ plane. The first step is filtering to reduce speckle, then relatively dark objects are extracted as candidates of road segments. The image is then divided into small overlapping subimages of a size such that the road segments are the largest features in the subimages. The peaks in the transform are enhanced by using the overlapping images and the line equations of road candidates are estimated. Finally, the intersections can be estimated from the line equations and checked against the road candidates obtained earlier.

2.1.9 Road detection in spaceborne SAR images using a genetic algorithm

Byoung-Ki Jeon et al. (2002)

Method: Roads in a spaceborne SAR image can be modeled as curvilinear structures that possess width. The curve segments, which represent the candidate positions for roads, can be extracted from the image using a curvilinear structure detector. Grouping these curve segments together leads to accurate road location. This paper attempts to find the

roads without any human intervention whatsoever and is thus one of the most challenging methods in literature.

As always with SAR images, speckles are reduced as they can degrade the performance of road detection. The authors used a sigma filter which suppresses speckles without blurring edges and destroying detail. As roads appear dark in SAR images, global thresholding is carried out to exclude regions of no interest before the extraction of curvilinear structures.

For the extraction of curve segments, the input image is convolved with a gaussian kernel and then each pixel is tested to see whether it is the center of a curve segment. The direction in which the curvature of the cross-section is the largest is found from the Hessian matrix computed at that pixel. Along this direction, first and second order derivatives are computed. If the 1st derivative is zero and the 2nd is high, then the pixel is at the center of a curved segment.

The next step is initial grouping, which is carried out based on perceptual factors such as proximity and cocurvilinearity. The results of the initial grouping are used as seeds for a genetic algorithm based grouping method. For each seed, the grouping is done independently around the two endpoints of the seed with the segments present in those regions. The grouped segments are used as seeds for the next stage and so on. The genetic algorithm imitates evolutionary rules of life in order to find an engineering solution. The proposed algorithm finds roads in the SAR images with an accuracy of 92.2%.

2.1.10 Application of spatial reasoning methods to the extraction of roads from high resolution satellite imagery.

Guindon (1998)

Method: As satellite imagery of higher resolution becomes available, information extraction methodologies to exploit the data must draw on both data-driven processing of images along with object-driven image understanding technologies. This paper describes a method where the feature extraction is based on spatial reasoning with a segmented rendition of the image. The recognition system recognizes residential streets using road seed identification, iterative extrapolation from the seeds to derive a consistent road network and abstraction from the segments to a road centerline rendition.

In road seed location, it is more important to minimize commission errors than to strive for network completeness. Rules to locate seeds can only rely on inherent segment attributes and not on any spatial context. As is standard in satellite imagery algorithms, some observations are made beforehand based on a typical set of the images. A visual inspection revealed that road brightness can not only vary significantly, but that contrast reversals relative to road surroundings can also occur. Thus, a brightness attribute is, by itself, a weak discriminator. Secondly, longer road segments are more easily identified. Lastly, segments are seldom linear and may even be complex in shape, but exhibit constancy in one property, their width.

Based on observations, road seeds are selected from among the segments using the following properties:

1. length to width ratio of at least 5.
2. grey level mean and variance, the tolerances of which are defined by global ground truth.
3. road width within tolerances defined by global ground truth.

The extrapolation process is based on the following characteristics:

1. the neighboring segments of each *known* road segment (referred to as parent segment) are examined to identify new road candidates.
2. road segment radiometric characteristics in a local residential neighborhood can be expected to be quite uniform compared to the broad variations seen over large urban areas. Thus, the attributes of the parent segment are used as local reference.
3. limited spatial context can be employed. For example, roads intersecting at right angles can provide supporting evidence.

The rules for road extrapolation use logical extension to search for new candidates. Logical extension is defined as: ‘ A segment X is defined to be a logical extension of segment Y if segment X is in close proximity to segment Y and at least one branch of the skeleton of segment Y, when extrapolated, leads to segment X’. The rule set consists of a series of rule groups, each of which tests one of the possible cases of logical extension, parent-to-candidate, candidate-to-parent and road extension versus intersection.

The road recognition (seed and extrapolation) rule set identifies 51% of road segments. Manual interventions are required to improve upon the performance of the extrapolation iterations.

2.1.11 The line segment match method for extracting road network from high-resolution satellite images

Wenzhong Shi and Changqing Zhu (2002)

Method: This paper proposes an approach for extracting road networks from high-resolution images of urban areas. As pointed out by the authors, methods developed before this paper rely heavily on the characteristics of the images. Therefore, in keeping with this methodology, the first step in this algorithm is analysis of the set of images. High-resolution images contain far more ground detail, thus providing a chance to

extract more detailed road features but it also has a disadvantage as in there are more distinct and visible non-road features within the image. Thus, this paper is novel as in it presents an approach according to characteristics of high-resolution images.

To simplify matters, the image is converted to binary form. The road network becomes more visible, but there is still a lot of non-road noise. From the analysis, the road network can be considered as a composition of many roughly straight and slightly curved segments. A road is several pixels wide and is in the shape of a long-narrow rectangle or a band-shaped line.

First, the straight-line segments are detected. The minimum threshold for length is 30 pixels and the maximum allowed gap between segments for them to be considered part of the same road is 3 pixels. These values are fairly stable to a certain type of network. For a given direction, the detection is done for all parallel lines from one end of the image to the other. Each time, the starting point moves by one along the edge. This complete operation is repeated for all directions, with a precision of a third of a degree. This leads to a coarse network.

The coarse network is processed by filling in the holes on the image using closing operations followed by thinning using a morphological thinning algorithm. Finally, the small gaps are connected and short splinters are deleted.

Overall, this method is automatic in the sense that there is no need to select or provide seed points, but it requires intervention for purposes of setting parameters. The accuracy is calculated using the ratio of length of road correctly identified to actual length of road in the image but results are provided for only two test images. This method has been named the *line segment match method*.

2.1.12 A wavelet transform based method for road extraction from high-resolution remotely sensed data

Chen et al. (2002)

Method: For a given image, the wavelet transform of the image is proportional to the gradient of the image after smoothing. Some specific kinds of edges, such as sharp edges or nearly sharp edges can be characterized by comparing the magnitudes of their gradients under wavelet transforms at different dilation scales of the same wavelet. Very importantly, by choosing wavelets with proper supports, roads with apparent difference in width can be easily separated. Thus, extracting roads from high-resolution remote sensing images using wavelets becomes practicable.

The analysis is focused on characterizing some specific sudden changes of signals and the effects among their wavelet transforms. A cross section of a road gives a platform-like signal, and its wavelet transform has a local maximum and a local minimum. When the support of the wavelet is under a certain limitation, the magnitudes of the local extreme values are not affected. When the support of the wavelet is wide enough, the magnitudes of the local extreme values may become small. This discovery makes it possible to threshold those thin lanes and retain the main roads in an image.

The results of the one dimensional case are extended to the two dimensional case. A two dimensional wavelet transform actually consists of two one dimensional wavelet transforms followed by smoothing. Thus, the observation that the width of the support of the wavelet affects the values of the wavelet transforms still holds. Thus, a wavelet transform based technique of road extraction has been created. The new technique requires preprocessing of the image such that the information of the roads is strengthened, deleting undesirable information using wavelet transforms, and performing

some further processing to get the final result. This technique has been applied on aerial photos with resolutions of 3m and higher.

2.1.13 Road network extraction from airborne digital camera images: a multi-resolution comparison

Gong and Wang (1997)

Method: The authors use the Gradient Direction Profile Analysis (GDPA) algorithm, a clustering algorithm, a supervised maximum likelihood classifier and a contextual classifier. The resolution of the images used is 1.6m.

There are five steps to the method: data preprocessing, obtaining the initial road network, noise removal, thinning and pruning. The preprocessing enhances road network features for subsequent analysis. The technique used here is grey level dilation filtering with a 3 by 3 kernel. The initial road network is then obtained by using the GDPA algorithm. The performance of the algorithm is dependent upon three parameters that must be fixed by the user. Then, the noise removal algorithm removes pixel patches smaller than a certain size. Thinning reduces the network to one pixel in width and pruning removes short branches of dead-end roads, depending on their lengths.

Instead of the linear analysis, the other way is to use classification methods to identify road areas. Of course, this is possible only when the resolution of the image is good enough to display the roads as areas within the image instead of brightness valleys or ridges. If the resolution is too low compared to the minimum width of the road to be extracted, the road area will not appear distinctly enough to be classified accurately. The authors use eight classes in the supervised classification, including four types of road covers and four types of non road land surfaces.

78.7% of road pixels were correctly extracted using the linear extraction algorithm while the clustering method resulted in 74.5% extraction. After the analysis at maximum resolution, the images were resampled, using nearest neighbor method, to lower resolutions of 3m and 5m. For these images, the classification algorithms could not be used because of the reasons discussed above and thus, the testing was done using the linear analysis extraction techniques only. The accuracy of the extraction was 74.6% and 61.6% for the 3m and 5m resolution images respectively.

2.1.14 The research of road extraction for high resolution satellite image

Xuan et al. (2003)

Method: An automatic road extraction method is proposed for 2.5m resolution images, which converts the roads to road vectors. The steps of the method are edge detection, road following and grouping and formation of road vectors.

Roads appear as elongated regions with parallel rims. As the intensity value of a road is not constant, a Gauss function is employed as a filter to smooth images. Then, the derivatives are used to find extreme points or zero points of image intensity for detecting the image edges. The second derivative equated to zero helps find the edges within the image. In order to convert the discrete edge points into a useful linear edge, an edge-tracking technique, specifically the Hough Transform is used.

The objective of the step of road following is to eliminate road-like but non-road pixels. In this paper, region-based tracking is used with the intensity values of regions and background taken as constants. The Snake model is used to search for the edges of roads in the image. The collinearity condition helps to keep the road edges parallel and the smoothness constraint keeps the local deformation stable. At this stage, the result is incomplete due to shadows, occlusion and other objects that have the same features as

roads. Thus, Ribbon Snake is used to connect the unconnected road segments. Finally, the extracted roads are converted to road vectors using commercial remote sensing software.

2.1.15 Towards knowledge-based extraction of roads from 1m resolution satellite images

Lee et al. (2000)

Method: The approach comprises the following steps: firstly, the image is segmented using the modified hierarchical multi-scale gradient watershed transformation and the approximate road regions are identified using information about roads such as grey level values and elongation. These regions are expanded by connecting the close-by roads knowing that roads are connected to each other.

The segmentation requires that a road is not divided into too many elongated regions along the road direction and road areas should not be merged with non road areas. This is achieved using a modified version of the watershed algorithm. The next step of knowledge based extraction uses intensity and shape information to evaluate each segmented region. Each region is then either identified as a non road area and thus rejected or identified as possibly a road area and thus merged with another nearby area which has also been identified as a candidate road area. The evaluation and subsequent classification into non road or road area is based on a set of rules that depend on the knowledge gained about the typical roads in the image.

The final step is a searching technique that attempts to connect disconnected roads where road segments are missing due to imperfections in the segmentation step. The image used for testing is an airborne image downsampled to 1m resolution to simulate the quality of an Ikonos image.

2.1.16 Detection and extraction of road networks from high resolution satellite images

Peteri et al. (2003)

Method: The algorithm is divided into two sequential modules. Firstly, a topologically correct graph of the road network is extracted by a following algorithm which minimizes a cost function. Secondly, the reconstruction algorithm utilizes specific active contours combined with a Multi-Resolution Analysis (MRA) for minimizing the problem of geometric noise.

The first step aims at providing a road network with correct locations of roads and correct spatial connections. It uses a cost function that evaluates the homogeneity of the local radiometry variance for several propagation directions. From this extracted graph in which all the lines are not necessarily well registered, the lines are sampled and propagated in order to prepare for the reconstruction module. The second module uses a Snake implementation based on the *greedy* algorithm and the MRA uses the wavelet transform to perform multiresolution edge detection. This is followed by extraction of parallel road sides by using an object called the DoubleSnake and extraction of intersections using an IntersectionSnake.

2.2 Overview and Conclusions

2.2.1 Overview

After reviewing the papers on the problem of extracting roads from satellite images, certain dominant techniques and methodologies were observed. For example, the first feature of such papers is the reliance on observation of the images before committing to a method. Since the goal is rather specific, the extraction of roads, the algorithms require all the assistance that can be provided. Accordingly, the first step is to analyze the image and obtain information as to the typical *individual* and *collective* properties of road

pixels. Individual properties are basics such as the range of grey level values and contrast with background of road pixels while collective properties include the shape of segments, length of segments, degree of curvature, connectivity and homogeneity.

The properties culled from observations assist in the creation and refinement of the techniques used. In Duta (2000) and Mukherjee et al. (1996), the characteristics of the roads such as thickness, intensity, linearity, length and contrast were noted and used. The information may be used in a global manner, where it shapes the entire technique or it may be used only in a certain part of the algorithm. For example, in Guindon (1998), the properties observed were just used to select road seeds from among the various road and non-road segments found while in others such as Duta (2000), the properties are used throughout the algorithm, even in the high-level method.

It is important to note here that the properties employed by different researchers can and do vary widely. Which properties are used in a particular case depends on which properties display constancy throughout the entire set of sample images studied beforehand. In some cases, a large number of properties are reliably consistent and can thus be used effectively, whereas in some cases, there are only one or two such properties. The advantage of a large property set is that it reduces errors of commission, but increases errors of omission. A small property set leads to exactly the opposite problems.

Overall, a large property set is preferable as it is possible to retrieve the omitted sections through use of other techniques. Also, the sections that *are* detected and classified using the large property-set can be considered as belonging to the relevant class with a very high degree of confidence since they satisfy a large number of search criteria. The techniques that can be used for retrieving omitted sections include, but are not limited to,

thresholding and morphology. Even if no secondary techniques are able to retrieve omitted sections, it is not a major problem as they can be detected using the reliably detected sections and high-level methods.

The algorithms can be automatic, quasi-automatic or may require manual intervention at some stage, most often the beginning. Automatic methods only require to be provided an input and provide an output without assistance of *any* kind. Quasi-automatic methods are ones that do not require help or guidance with the actual location or detection techniques, but may require the user to set values for certain parameters, such as thresholds for length and/or intensity. Manual intervention is normally in the form of providing the algorithm with a starting seed point to detect roads or other features. Once the seed point(s) are set, the algorithm can carry on automatically. Jeon et al. (2002) presents a completely automatic method whereas Shi and Zhu (2002) is a quasi-automatic algorithm.

A common first step in road extraction algorithms is to locate segments within the image. Segments can be either straight or gently curving. Many methods have been used in the literature to detect them. In Netanyahu (1996), a local non-linear operator is used to detect pixels that are locally line-like and then robust estimation techniques are applied to find sets of pixels that lie on or near straight or circular loci. In Guindon (1998), segments are found using an edge-based approach that delineates, as entities, regions of uniform texture and brightness followed by a fusion step to create super-segments. In Jeon et al. (2002), segments are extracted using a curvilinear structure detector.

Once some of the segments have been detected, the next step is to move onto high-level methods to create an accurate and complete road network. It has been pointed out that, in most cases, the low-level technique does not detect all or even a majority of the

segments. This is, of course, obvious, since otherwise, there would be no requirement for a high-level technique. In such cases, what the low-level technique does achieve is that it locates the segments that are the easiest to detect. These segments are then used as road seeds and the remaining segments are found by high-level methods, some of which are explained below. In the rare (and trivial) case where the low-level technique detects *all* segments, the high-level technique can be used to merge them together using rules such as collinearity and homogeneity.

Some of the high-level techniques encountered during the research are described below along with their strengths and weaknesses. In Ruskone et al. (1996), an innovative method using ‘meta-lines’ is described. Once pixels belonging to vehicles are identified, an overview is obtained of the organization of the vehicles. This helps validate the presence of roads or parking areas. Since any two adjacent vehicles comprise a line, the distinction between a road and parking area is made based on orientation of the line relative to the vehicles. This is an interesting method. Even if some vehicles are not detected initially, it is still possible to form lines. Also, it is always possible to ascertain the relative orientations of vehicles and lines, thus leading to a reliable classification of roads/parking areas. However, there is one factor to be considered. If a certain road has no traffic, it will not be detected as the algorithm searches for vehicles initially. Thus, multiple images of the same area at different times or few images from peak traffic hour may be needed. The results from all the images can then be added together without the need for any calculations. This method is reliable as it provides a logical validation of the initial clustering carried out.

In Barzohar and Cooper (1993), the image has been divided into square windows within each of which road candidates are found. The high level method consists of abutting windows along the edges of a window with some candidate roads and extending the

candidates found. Abutting windows helps concentrate the search on a small area and extends the road a small segment at a time. This method is reliable in so far as the low-level technique used is reliable. Since the high-level method needs to use the low-level search, it is dependent on it. Thus, this high level method *cannot* be assessed all by itself as a stand-alone although it is logically sound *given* the low-level technique.

Two other high-level methods are the tree method and logical extension. The relevant papers are Duta (2000) and Guindon (1998) respectively. These two are described together as they are conceptually similar. Both are based on extrapolation using certain properties of the road pixels within a local neighborhood. Given a segment to be used as a road seed (from the low-level technique), the immediate area is searched for extensions. In the tree method, several possibilities are considered and all but one are rejected based on certain criteria. The tree method is based on the obvious and always true fact that roads are piecewise linear and that once the end of a linear part is reached, the next linear segment can be found by exploring the angles at the end of the first segment. Given that roads exhibit spectral characteristics different from surroundings, this is a good method. Similarly, in the extension method, neighboring segments are considered and the one that satisfies the criterion of collinearity or connection is the extension.

2.2.2 Conclusions

From this research, three facts emerged. Firstly, a variety of methods exist to extract roads from satellite images. Some methods are more effective than others, but none is complete by itself. None of the algorithms identify all the roads in an image. One of the methods was created only to identify the main roads in the image, ignoring the smaller branches for simplicity. It must be pointed out, however, that the most successful methods identify a large majority (up to 90%) of all roads present. This is quite an

achievement, considering the complexity of the problem. However, high success rates across a comprehensive set of images are confined to work on low-resolution images. Given this success ratio, any new method must aim for similar results, though the problem will have a higher difficulty level for high-resolution images. Difficulty level here refers to the complexity of the image with respect to extraction of roads.

Secondly, in each case seen, the method is built around and for a particular type of image. That is, the images all have the same resolution and are of a similar kind, panchromatic or multi-spectral, taken by the same satellite. In some cases, the images used to test a particular method are of the same terrestrial spot, for example, a particular city center or a specific agricultural area. This makes the task easier as the images share more common spectral and spatial traits, which can be exploited for the extraction procedure. There is no method that works for all images, of a given resolution, of various areas or a method that works for all images of different resolutions, of the same area. Also, a given method works only for a certain kind of image, panchromatic or multi-spectral. Thus, taking account of all the above factors, any method developed must be created such that it works for either panchromatic or multi-spectral images of a fixed resolution (depending on image set available and chosen) of a particular *kind* of terrestrial area.

Thirdly, the research led to the realization that there is relatively little work in the field of road detection from high-resolution imagery. Also, among the small body of work on high-resolution imagery, the majority of publications either do not provide quantitative results (only original and final images are provided for visual comparison and evaluation) or if they do, they only do so for one or two images. There are very few that provide quantitative evaluations of the proposed method. Guindon (1998) is one such attempt. But, the success ratio is just 51% as 99 of 196 road segments were identified.

Another is Shi and Zhu (2002), but that method is limited by its low-level nature and not applicable to the wider variety of problems or to more complicated images where the simplistic approach would not perform quite as well. Also, results, though reasonable, are provided for only two images. Gong and Wang (1997) experimented with two approaches, linear analysis and clustering, but both methods provide extraction accuracies of under 80%. More importantly, the above methods were created for images with lower resolutions than used in this thesis. Recently, panchromatic images with a resolution of 60cm have become available. Thus, there is considerable scope for original work (evaluated on a larger set of test images), not to mention improvement upon the few existing methods, in the field of high-resolution imagery. Thus, the decision was made to concentrate upon creating a method to extract roads from panchromatic images, specifically of urban areas, with a spatial resolution of 60cm per pixel.

Chapter 3: Study on Detection of Roads in High-Resolution Satellite Images

The conclusion reached from the literature review in Chapter 2 was to carry out a study on extraction of roads from high-resolution, panchromatic satellite images obtained from the QuickBird satellite with a resolution of 60cm per pixel. The images are usually a few thousand pixels along both dimensions. In such cases, sub-images were made by cutting out sections of the original image. An example sub-image is shown in Figure 3-1. This helps reduce the run-time of the algorithm. While this may not be a factor once the algorithm is complete, it is very much a factor when testing and debugging. Handling images with sizes of approximately 500 by 500 is significantly faster than the original image and speeds up the research process. Of course, the algorithm is tuned to all the sub-images, but when minor changes in code and logic need to be tested, it is better to use smaller images. It saves time without reducing the quality of the final algorithm in any way. The processing of large images follows the same process. In this study, only the processing requirements are considered. The real-time implementation of the technique is beyond the scope of this thesis.

3.1 Road Map Detection Algorithm (Stage One)

This section explains the Stage One algorithm step-by-step and the reasoning behind each step. The various steps will be illustrated by corresponding images. Firstly, we observe the image-set in order to ascertain the properties that can be used for the detection of the road features. This is important as determining the properties which are representative of road features within a set of images helps determine the approach to be used. As will be clear later, the algorithm comprises a number of steps. Each one is chosen due to the properties observed in the image resulting from the cumulative effect of all the previous steps. At each point, certain alternative techniques were considered.

Some were rejected based on conceptual analysis while others were rejected after they were implemented and found to be inferior to the other choice.

Figure 3-1: An original high-resolution greyscale image.

3.1.1 Thresholding

The initial image is greyscale. An example is shown in Figure 3-1. It is converted to binary form using thresholding. Within any image, the roads are relatively consistent in their intensity values. Assuming a size of 500 by 500 and a resolution of 60cm per pixel, that translates to a plot of land 300m by 300m in actual size on the ground. Within such a small area, it is normal for roads to be similar. Thus, thresholding is a viable option. The assumption behind the use of thresholding is that the class of pixels belonging to roads is separable from the pixels in the immediate surroundings of the roads. This is an important starting point.

However, the thresholding used is not the normal global method where any pixel greater than threshold is foreground. In this case, a *range* of pixel values is designated as foreground and pixels with intensity values on either side of that range are classified as background. The values are set adaptively, making the algorithm quasi-automatic. Given that the values are set adaptively, parity has been ensured in the performance of this approach and that of a system using calculated values. The range that is chosen for an image is a large one instead of a range that concentrates upon only the exact values of the road network. Choosing the exact values would simplify the problem more than the values chosen by an automatic thresholding method would, since such a method would not be perfect and would definitely incorporate non-road pixels into the foreground. Thus, the values are set in a realistic manner, in order to closely mirror the performance of an automatic method. Of course, it is possible that an automatic method outperforms the present one, in which case, the results would receive a positive boost from such an implementation. At the same time, it is highly unlikely that an automatic method will underperform, given the large range used currently. The point here is that the thresholding method used does not confer an undue advantage to the algorithm compared to an automatic method. In fact, care has been taken to ensure that the subsequent steps do not require high quality thresholding. It is only a minor pre-processing step.

$$m(x, y) = \begin{cases} 1, & f(x, y) > th_u \\ 0, & th_u \geq f(x, y) \geq th_l \\ 1, & f(x, y) < th_l \end{cases} \quad (1)$$

where $f(x,y)$ represents the original image with x and y as the coordinates and $m(x,y)$ is the resultant image at the end of the step. 1 represents white (background) and 0 represents black (foreground). This convention is used throughout this thesis. The lower and upper thresholds th_l and th_u are set adaptively, making the algorithm quasi-automatic.

Figure 3-2: Thresholded image. Black areas represent foreground.

Thresholding was done at the initial stage in order to remove a portion of the background. As can be seen from Figure 3-2, a portion of the non-road area has been consigned to the background and accordingly, set to the color white. It must be noted that the foreground is colored black for purposes of clarity. Thresholding has minimal negative effect as little or no road area is lost and at least some unnecessary areas are removed. The black areas in the binary image contain significant amounts of non-road pixels. This is unavoidable as there are bound to be non-road features that share the intensity values. At the beginning stages, it is a better practice to incorporate noise rather than lose road pixels.

3.1.2 Parse and Merge

Thresholding has one obvious negative effect on the major roads in images. Since the image is so highly resolved, it is possible to see the road lane dividers clearly. This can be seen in Figure 3-3. The problem is that the dividers have an intensity value quite distinct from the road surface. From everyday experience, it is well known that dividers are brighter compared to the darker surface of the road. The thresholding classifies the dividers as background. This causes the multi-lane roads to appear as multiple, closely spaced, parallel roads in the binary image as can be seen in Figure 3-4. This is undesirable and has to be addressed before further steps to remove non-road pixels can be used. If these roads are not corrected, it would lead to them being lost at later stages of the algorithm. The correction step has the effect of merging together adjacent, parallel road lanes which were previously separated by background pixels between them. The technique used depends upon parsing of the image along the vertical and horizontal directions. The technique, named parse and merge, detects a certain sequence of pixels along the direction of parsing. The sequence searched for is basically a short consecutive line of black pixels, followed by consecutive white pixels and so on as shown below.

$$F_5^m B_1^4 F_5^m B_1^4 F_5^m \quad (2)$$

whereby F represents foreground and B background. The subscript and superscript indices indicate the minimum number and the maximum number of pixels in the mask image that have to fulfill the pattern's constraints. Thus the investigated section is partitioned in regions corresponding to foreground and background. For each fitting pattern part, i.e. F and B , the algorithm assigns unity and inverse unity, respectively. Thus, the sum over the blocks equals unity if a correct match is found. If so, all corresponding pixels, which were used for the testing, are set to be part of the foreground. The pattern search is done over a minimum of five blocks as shown above. It

can come to an end when a consecutive line of pixels extends too far or if the edge of the image is reached. The logic of this method is that the short, consecutive black lines represent the cross-sections of the road, while the white ones represent the cross-sections of the dividers. Thus, what the method is actually searching for is a sequence of road, divider, road and so on. There is an alternative way to incorporate the lane dividers into the foreground. That is to modify the threshold range. However, the major drawback is that the range becomes too large and the benefit derived from thresholding is lost almost completely. The parse and merge solution to this problem is just as effective at recovering the dividers, but far more efficient and elegant with no negative effects on the rest of the image.

Figure 3-3: Zoomed in portion of an image showing details of the road-lane dividers.

There are parameters that define how wide a lane and a divider are expected to be and these ensure that non-road features are not merged together. The other factor that ensures detection of split roads only is the length of the sequence. Such properties are not found

Road Extraction from High-Resolution Satellite Images

in non-road features and the testing has confirmed the success of this technique at the detection and merging of split roads while not interfering much with other, non-road features. Figure 3-5 shows the result of the technique applied to the image in Figure 3-4.

Figure 3-4: The same portion as Figure 3-3 in binary form, showing the lane dividers.

Figure 3-5: The effect of the Parse and Merge technique. The lane dividers have been removed.

3.1.3 Parse and Remove

The next few steps aim at the removal of noise, which is defined as any non-road pixels. Firstly, salt and pepper noise is removed using a simple 3 by 3 mask which deletes single, isolated foreground pixels. Then, another parsing technique is employed in order to remove thin curvilinear structures. The parsing is done along the two major axes as before, but, in this case, the objective is to detect and remove narrow cross-sections, which actually represent thin, non-road structures. The parameter value for the cross-section is set at a value considerably less than the typical width of a road. If the cross-section is found to be less wide than the threshold value, that consecutive line of pixels is set as background. Thus, this technique breaks narrow curvilinear structures into smaller segments by slicing along the cross-section. Since the parsing is done along both directions, it detects any segments that are aligned with one of the axes. In some cases, the entire segment is removed as it has a narrow cross-section throughout. Some structures remain in fragmented form (Figure 3-6).

Figure 3-6: The effect of Parse and Remove. Compare this image to the one in Figure 3-2.

A diagonal parse is unnecessary because even if the structure is at 45 degrees (maximum possible value) to a parse direction, the width along the parse direction is increased only by a factor of $\sqrt{2}$, which still does not approach the actual width of a road. The technique is implemented as a state machine but for purposes of explanation, can be represented mathematically by convolution. The one-dimensional parsing of the image m for a specific row with index y can be written as

$$g_y(x) = m_y(x) * h(x), \quad (3)$$

with operator $*$ indicating convolution and h defined as [1111]. The mask h should contain ones for pixels that you are interested in. For example, if we look for the pattern ‘111’ in an image (the data is one-dimensional) that is described by ‘1011101’ then the convolution (previous padding of the image with zeros to obtain an image that matches the original in size) leads to

$$0|1011101|1 * 111 = 1223222$$

Thresholding with the sum of the filter’s coefficients, i.e., 3, results in 0001000. Hence, we detected the pattern ‘111’. Basically, Equation (3) determines whether the number of continuous foreground pixels, $g_y(x)$, starting at each pixel, is greater than or equal to a certain value. In this case, the pixel is removed if the value is less than the set limit of 4, i.e.

$$m_y(x) = \begin{cases} 0, & g_y(x) \geq 4 \\ 1, & g_y(x) < 4 \end{cases} \quad (4)$$

3.1.4 Connected Component Labeling/Analysis

These structures need to be removed along with other bigger structures that have not been addressed as yet by the algorithm. At this stage, *connected component labeling* is carried out. Connected Component Analysis (CCA) is a general image processing technique frequently used in document image analysis. It is commonly used when dealing with binary text images and is a powerful tool when used effectively. CCA is used to analyze an image once it has been divided into separate, constituent components using the labeling technique. A component is defined as pixels that are 8-connected to each other. Pixels which are not 8-connected belong to different components and pixels that are 8-connected *always* belong to the same component.

Figure 3-7: All the components from Figure 3-6 are marked by rectangles, demonstrating the connected component labeling. Now, analysis can be carried out based on these components.

Once connected component labeling is finished, the information is stored in a linked list which can be accessed for obtaining various parameters of a component. Some of the parameters stored are end-point co-ordinates and labels of the pixels within. This is an innovative step as CCA is not normally used in images of such type. That can be explained by the fact that the vast majority of previous work on extraction of roads has been done on low-resolution images. High-resolution images are more detailed and when converted to binary form, more likely to benefit from CCA. This is the property that led to the use of CCA for removal of unwanted features from such images. Figure 3-7 shows all the components in the image marked by bounding rectangles. At this stage, the number of components is extremely high and has to be reduced to those few components that comprise the road network. Accordingly, three methods were used to remove noise-based components.

3.1.5 Removal of Small Components

The first step is to remove all pixels belonging to small components as these components do not represent roads. In most cases, they belong to the fragmented segments left behind by the previous step or to other small structures. Components containing roads are quite large as the roads branch and are quite long. In fact, the largest components in any image always belong to roads. Thus, the small components can be safely removed without loss of road pixels. Equation (5) represents the logic behind this step. The pixels belonging to a component with both height and width less than a fixed value are classified as background, that is, the component is effectively deleted. The result can be seen in Figure 3-8. The fragments that were seen in Figure 3-7 are now missing, but certain components remain. Overall, the image is much less noisy. The number of components in Figure 3-7 was 433, while Figure 3-8 only has 27.

Figure 3-8: The binary image after removal of small components.

$$m(x, y) = 1 \quad \text{if} \quad \Delta C_i < 50 \wedge \Delta R_i < 50 \quad (5)$$

where ΔC_i and ΔR_i are the width and the height of the bounding box of component c_i , respectively. As before in Equation (1), $m(x, y)$ represents the image and setting a pixel value to 1 (white) represents classifying the pixel as background.

3.1.6 Removal of Dense Components

At this point, the image is at a stage where any foreground pixels belong to large components that may or may not be road-based. Thus, a property other than size has to be used in order to differentiate between them. The property used is that of component density. A component comprising only roads will have a relatively sparser distribution of pixels compared to a component that represents, for example, a patch of grassy area from the original image. Although they share similar original intensity values, the density of pixels in a grassy patch is significantly higher and using a threshold for the density, the non-road components can be detected and removed. There is the occasional component

that is non-road but less dense. Such a component will not be removed. Some components comprise road pixels, but are aligned with one of the axes, causing a high density value. This happens when the component contains an isolated linear segment of road. To avoid loss of such components, each component with a high density value also has its aspect ratio checked. If the aspect ratio is high, it points towards a high likelihood of it being a road-based component. Such components are not removed. Equation (6) represents the logic behind this step and the result can be seen in Figure 3-9. The pixels belonging to a component with both high density and low aspect ratio are classified as background, that is, the component is effectively deleted.

Figure 3-9: The binary image after removal of dense components.

$$m(x, y) = 1 \quad \text{if} \quad \begin{cases} |c_i|_F > 0.4(\Delta C_i \cdot \Delta R_i) \\ \wedge \\ \max(\Delta C_i / \Delta R_i, \Delta R_i / \Delta C_i) < 2.5 \end{cases} \quad (6)$$

where ΔC_i and ΔR_i are the width and the height of the bounding box of component c_i , respectively. $|c_i|_F$ is the total number of foreground pixels within component c_i . As in Equation (5), $m(x,y)$ represents the image and setting a pixel value to 1 (white) represents classifying the pixel as background.

Thus, the noise-removal approach is one of attrition. Once the binary thresholding has been carried out, every subsequent step aims at removing part of the noise, while minimizing loss of road pixels. None of the steps promises complete success, by itself or in conjunction with the other steps. At the same time, none causes significant loss of valuable road pixels. The parameters set and threshold values used reflect this philosophy of slowly reducing the noise step by step, while preserving the relevant pixels. Figure 3-10 shows the effect of the algorithm on the test image so far, by comparing the original image to Figure 3-9.

Figure 3-10: A comparison of the original to the result at this stage of the algorithm.

The noise having been significantly reduced, another level of parsing is employed. This time, the threshold value for the removal of the cross-section is increased. This step aims at removing small spurs which typically connect a road to a noisy, dense component, thus making it one large, sparse component. Such a component tends to not get removed by the techniques used thus far. However, if the connecting spur is detected and removed and CCA is carried out again, the noisy section can be removed easily as explained above. This step has some success as it does remove certain spurs. However, most spurs remain as they are too similar in size to actual roads and it is not possible to remove them without removing roads too. Once the spurs are removed, CCA is carried out again because the image has changed significantly due to the removal of a large number of foreground pixels and thus, the previously compiled connected component list does not represent the actual image accurately anymore.

3.1.7 Removal of Irregular Components

Given this new linked list, a method has to be found to remove noisy components that are sparse and large, thus escaping removal thus far. The property used at this stage is the irregularity of the component. A road-based component has a distinctly more regular appearance to the human eye than a noisy component, which has no alignment and no basic geometrical shape. A mathematical approach is required in order to be able to model the property of irregularity of shape. However obvious it may be visually, it cannot be translated into a useful distinguishing property unless the irregularity can be defined by a corresponding measure which is both precise and reliable. The technique decided upon was that of skeletonization. Figure 3-11 shows the skeleton of a typical, irregular component in the image. The components are thinned to their corresponding skeletons and then a measure is made of the level of branching within the skeleton. This is calculated as the ratio between the number of pixels with two or more branches $|c_i|_{>2}$

and the total number of foreground pixels $|c_i|_F$ of component c_i . If this ratio is higher than 0.3, the component is removed.

$$m''(x, y) = 1 \quad \text{if} \quad |c_i|_{>2} > 0.3 \cdot |c_i|_F. \quad (7)$$

Logically, this ratio should have a higher value for the irregular components than for road-based components. This was borne out by observation and testing of the images. This technique was able to successfully detect irregular components and remove them, leaving behind minimal noise in the images, as shown by Figure 3-12.

Figure 3-11: Zoomed in view showing the skeleton of a component from Figure 3-9.

Figure 3-12: The binary image after removal of irregular components.

Again, connected component labeling is carried out because the image has changed completely and the foreground contains fewer pixels. Thus, the old component linked list is now invalid. The step to remove small components is used once more, followed by the step to remove dense components with low aspect ratios. This time, the parameters are adjusted slightly to reflect the different nature of the image. At this stage, the images contain some noise that is essentially random, in the sense, that this noise has no consistent characteristic within the image itself or across all the images. Lacking such a characteristic, it is hard to detect such noise and remove it. Some images have no noise whatsoever, while the others do. None of the images contain a significant amount of noise and all retain a high percentage of the original road area. The actual statistics are presented in Chapter 4.

Figure 3-13: The final image after further removal of remaining small and/or dense components.

Chapter 4: Results of Road Detection (Stage One)

This chapter presents the results of the Stage One algorithm described in Chapter 3. A variety of images from the high resolution scanner QuickBird were used. The images show different types of urban areas where the roads are partly hidden from view by trees, shadows of buildings and cars. In order to estimate the performance of the detection, two measures were used. The first was the total length of road detected, ultimately expressed as a percentage of the actual total length of road and the second was a measure of road length represented by road seeds in the final image, again ultimately expressed as a percentage of total actual road length. A road seed is said to exist for a given road if that road is represented in the image by at least one component, with those components including a significant number of pixels from that road. The logic behind using road seeds as a measure is that the presence of a road seed for a given road segment provides an avenue for complete extraction of that road segment using more sophisticated techniques such as extrapolation. That is to say, a lost road segment with a road seed present in the final image is recoverable with a high degree of confidence. Thus, the statistics, concerning road length recovered in terms of road seeds, provide an estimate of the maximum possible extraction of roads by the algorithm with the aid of extrapolation. To further improve upon this projected upper limit, supplementary techniques will be required.

A straightforward length-measurement technique was used to determine the length of road detected and length of road lost in the final image. As two different measures are being used, the measurements were done twice for each image. For each image, a unit length was defined as the smallest missing linear road segment in that image only. All measurements done on that image were carried out using this unit. Thus, in the general

Road Extraction from High-Resolution Satellite Images

case, the unit length varies for different images. The values are presented below in Table 4-1 (using unit length as defined above).

Table 4-1: Table showing statistics of road length extracted and lost (using both measures).

Image No.	Length recovered	Length lost	Length recovered in terms of road-seed	Length lost in terms of road seed
1	45	13	49	9
2	74	6	80	0
3	67	13	80	0
4	71	0	71	0
5	103	4	107	0
6	79	23	90	12
7	42	6	42	6
8	60	3	61	2
9	55	11	62	4
10	43	6	43	6

From Table 4-1 above, the following observations can be made which help explain the statistics presented.

1. For a given image, the sum of values in columns 2 and 3 is always equal to sum of values in columns 4 and 5. Both sums represent the total length of road present in the image and are thus, always equal.
2. The value in column 4 is always greater than or equal to the value in column 2 for a given image. This basically means that on an image-by-image basis, the length extracted in terms of road seed is always higher than actual road length extracted. The reason for this is that if only half of a certain road has been extracted in a certain image, it significantly reduces the total length of road detected. However,

it does not affect the road seed value as that road is partially but significantly represented in the final image by at least one component (as defined above).

3. In some cases, the values in columns 2 and 4 are equal as the lost road segments are not represented in the final image by any road seeds. This normally occurs when the lost road segment belongs to a road which was isolated in the original image and has been completely lost in the final image. Thus, there is no component that represents that lost road in the final image.
4. As explained in Chapter 5, the road seeds provide an avenue for completion of the road network by using extrapolation. However, as the statistics in column 5 of Table 4-1 show, there exist road segments in most of the original images that cannot be recovered by using the road seeds in the final image. This is because those road segments have no representative road seeds. Thus, the need for recovery of obscured areas.

Table 4-2: Table showing the statistics from Table 4-1 in terms of percentages.

Image No.	% of road recovered	% of road recovered in terms of road seeds
1	77.6	84.5
2	92.5	100
3	83.8	100
4	100	100
5	96.3	100
6	77.5	88.2
7	87.5	87.5
8	95.2	96.8
9	83.3	93.9
10	87.8	87.8
Avg.	88.2	93.9

The values in the second column vary from 67.8 to 100 and those in the third column from 84.5 to 100. As expected, the values in the third column are consistently higher and the mean value is higher too by 5.7%. While the values above are encouraging, it must be remembered that some of the final images do contain random noise which is hard to categorize and thus, hard to remove in a consistent manner without compromising the recovered road map. The nature of this noise is such that any attempt to correct it using noise removal techniques will most likely lead only to a reduction of the recovery rates seen above, while not solving the actual problem. Thus, this slight noise, being hard to model, has been ignored and focus has been switched to other techniques to improve the completeness of the extraction. The techniques needed to complete the map are described in Chapters 5 and 6.

4.1 Typical Failures

The algorithm does not recover the entire road network in most of the test images. Thus, there are obviously certain weaknesses and limitations within the methods. This section explains the typical problems that lead to loss of road area in the final image.

4.1.1 Loss Due to Thresholding

As mentioned in Chapter 3, thresholding is a feasible option as the underlying assumption is that the class of road pixels is separable from the immediate surroundings based on intensity values. Without this property, the algorithm cannot work. Thresholding is vital to the success of the algorithm and the alternative is to process greyscale images instead of binary ones. However, having said that, it must be noted that the above assumption is valid only for images where the road is discernible from its surroundings by the human eye. For the road to be visually discernible, the intensity values must be distinct from that of the immediate surroundings. Thus, a simple criterion is used for selecting the test images. If a human intelligence can visually detect the boundaries of the road map, then the image is acceptable in terms of complexity and suitable for use as a test image. While it may seem that this criterion would lead to inclusion of all images, this is not true. There are images where the boundaries of the road are not clear. This is caused by adjacent, non-obscuring features whose top view has a similar intensity. These features are typically buildings and car parks. Since the images used are greyscale, there is the disadvantage of not being able to use information from multiple channels, as would be the case if color images were used.

Since the threshold values are adaptively set, inclusion of the entire road network is the ideal aimed for, but there are a few cases where the benefits of a broader threshold range are outweighed by the disadvantages of the excessive noise that would be introduced.

This is with images which have roads with widely varying intensity values. In such scenarios, the loss of some road pixels is acceptable.

4.1.2 Loss Due to Concealment

The loss of road pixels due to thresholding is not the same as the loss due to obscuring by trees and shadows. Though the end result is the same, the cause is different. Loss due to obscuring of the road is due to the fact that the features have different intensity values from the road and are thus excluded (as they should be) even though the road underneath the obscuring feature is within the threshold range. When loss is caused by thresholding, it implies that the range is not large enough, but loss due to concealment does not reflect an imperfection in the algorithm. It just demonstrates that a supplementary method is required to reverse the loss. The solution to the concealment problem is not to increase the threshold range, but to detect such problem areas and retrieve the lost road area through other means.

Road Extraction from High-Resolution Satellite Images

Figure 4-1: A section of road obscured by foliage and shadows.

Figure 4-2: Section of thresholded image showing loss of road area due to obscuring (compare with Figure 4-1).

Figure 4-3: Final image showing the fragmentation caused by the obscured area.

A technique to recover obscured areas married to a high-level extrapolation technique for joining broken roads can lead to the successful completion of the algorithm. The algorithm would be able to deal with all possible kinds of noise in typical high-resolution satellite images of urban areas and intelligently recover from mistakes made by earlier steps. As pointed out earlier, the two approaches are complementary and thus, promise higher success rates than are currently achievable.

Thus, the overall complete *road extraction algorithm* can be seen as the synthesis of low-level and high-level methodologies. As it now stands, the low-level techniques have been implemented and have produced good results, providing the basic outline of the road map. In some test images, this outline is, by itself, quite complete. It represents the roads in the original greyscale image accurately and fully. In others, it is not quite that complete and requires improvement in order to resemble the actual road map. This is where the high-level method is useful since the low-level techniques have reached a dead-end in terms of improvement. Also, the more varied and diverse that the test images used at the final stage are, the more likely that the low-level techniques will be found

lacking. And it is a fact that there is no end to the variety of roads present in high-resolution images of urban areas. Thus, a high-level technique will definitely be required to complete the map produced by the low-level methods.

Chapter 5: Study on Extrapolation of Incomplete Road Networks

Having reached a stage where a significant portion of the road area from the original image has been extracted (by the algorithm described in Chapter 3), the next step is to recover any lost road pixels. This would complete the road map. Since the remaining noise has *no defining characteristic* which can be used to distinguish it from the roads, the best way to approach this problem is to complete the road map and then analyze the final image. If the completion of the road maps has been successful at joining roads that were initially separated due to noise, concealment (obscuring) or incorrect thresholding, then the final road map would basically be one large connected component with the noise detectable and thus, separable. The main reason that the remaining noise is indistinguishable from the road area is because of fragmented road segments that can be easily confused with the noise-based components. However, such road fragments would be part of a cohesive whole once the road map is connected fully and thus, the noise would stand out due to a lack of road-based components with similar properties.

5.1 Road Network Extrapolation Algorithm (Stage Two)

There are several complementary approaches to completing the road map by connecting incomplete road segments. One is to extrapolate the road map. Extrapolation is basically a high-level approach and requires reference to the original image. The first requirement of an extrapolation algorithm is a set of end-points of the incomplete road map, corresponding to missing road segments. These end-points can then be used as starting points for the extension. A comparison of the original image in Figure 3-1 with the skeleton in Figure 5-1 shows that several road segments are incomplete. The points at the end of those segments in the skeleton image are the candidate end-points. The concept behind the algorithm is that at each end-point candidate, a search is made to find the

direction such that the extension drawn is the most likely extension to the previously extracted road segment leading towards the concerned end-point. In order to rank the suitability of all possibilities, a probability approach is chosen to describe the model. The steps of the extrapolation algorithm (Stage Two algorithm) are described below.

Figure 5-1: The skeleton of the extracted road map after a few rounds of pruning.

5.1.1 Extraction of Initial End-point Candidates

Two complementary methods were used to locate the end-points. The first was to skeletonize the entire final image and find points with just one neighbor. This translates to points which are at the end of the skeletal lines since a thinning algorithm reduces all features to lines of single pixel width. However, it will be remembered from our earlier use of thinning that it produces several branches along the skeleton, also single pixel wide. These branches would lead to erroneous end-points. Thus, these branches are removed by pruning the skeleton. Pruning removes the branches, that is, all lines except the central one along the major axis of the skeleton. After pruning, the end-points are located and marked for later use. It must be noted that this method *will* lead to some

erroneous end-points as not all branches can be pruned away. Removing all branches would require pruning of the image until convergence, that is, until further pruning has no effect on the image. However, convergence has the effect of reducing the skeleton to closed loops, if any exist within the skeleton. Thus, it is counter-productive to prune till convergence. Instead, pruning is done a few times on the entire image to remove most branches and then the end-points are marked. Figure 5-1 shows the skeleton of the entire road map. Note that due to the pruning, the skeleton represents a significantly smaller amount of the original road network than the binary image in Figure 3-13. However, the loss is recovered by the extrapolation of the skeleton.

Possible end-points are determined in the skeleton image s whereby a candidate c at the position (x,y) has to fulfill the requirement

$$\sum_{|\Delta x|+|\Delta y| \neq 0} s(x + \Delta x, y + \Delta y) = 1 \quad \text{with } \Delta x, \Delta y \in \{-1, 0, 1\}. \quad (8)$$

However, this leads to a large number of false end-points, mostly at the end of short spurs emanating from a stretch of road. Hence, in the next step such cases are located and removed, leaving behind only a few false end-points, such as those at the end of a complete road. These, combined with the true end-points, are the initial end-points for further processing.

These end-points must be confirmed as valid via an alternative method in order to remove the incorrect ones. The second method uses a structuring element which is in the form of four intersecting lines. The lines are at 45 degrees with respect to each other, of equal length and intersect at their mid-points, thus forming an 8-point star, the eight directions being the main compass directions. This element is used as a mask. The pixel being tested as a possible end-point is placed at the central intersection point. To qualify

as an end-point, the pixel must be positioned such that two exactly opposite arms of the element coincide with only foreground and only background pixels respectively. Since the pixel is placed at the center, this implies that it is along an edge of a road and the length of the arms ensures that the position is at the end of a long stretch of road and not just any random edge. Once end-points are marked using both the above methods, final end-points can be chosen by accepting those that have been marked by both methods. The confirmation is not strictly point to point, but rather area to area. Given that one method uses thinning to find the end-point, the corresponding end-point from the structuring element method is never exactly the same pixel, but lies within the immediate vicinity. This is the test applied for confirmation of a given pixel as an acceptable initial end-point.

5.1.2 Properties of the Existing Road

This step leads to the determination of the properties of the road area immediately adjacent to a given end-point. The reason for doing so is the assumption that the characteristics of a stretch of road are spatially invariant. The properties of the immediate road that are extracted describe the existing road segment which has to be extrapolated. They can, therefore, be used for evaluation of possible extensions as the extension area is assumed to have similar properties. For each end-point, the properties of the corresponding road segment in the original image i are extracted, i.e. the mean μ and the variance σ^2 of pixels that belong to the road part that leads to the concerned end-point c with the coordinates (x,y) . Hence,

$$\mu_{(x,y)} = \frac{\sum_{r(x+\Delta x, y+\Delta y)=1} i(x+\Delta x, y+\Delta y)}{\sum r(x+\Delta x, y+\Delta y)} \quad (9)$$

and

$$\sigma_{(x,y)}^2 = \sum_{r(x+\Delta x, y+\Delta y)=1} \left(\frac{(i(x+\Delta x, y+\Delta y) - \mu_{(x,y)})^2}{\sum r(x+\Delta x, y+\Delta y)} \right) \quad (10)$$

where r indicates the road map in Figure 3-13 and Δx and Δy a local neighborhood.

Lastly, the orientation θ of the existing road segment is calculated. Assuming that the segments are piecewise straight, a linear model with slope a and offset b is fitted within a local neighborhood for each end-point in the skeleton image. Using the least square approach the y_j -coordinates of all skeleton pixels are related to their x_j -coordinate counterparts according to:

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \quad (11)$$

where n is the number of points on the skeleton.

Equation (11) can be solved by utilizing the pseudo-inverse to determine the model parameters and thus the orientation. Note that special checks must be made for the presence of vertical road segments as the actual value of the slope tends to infinity. Thus, we obtain the mean, variance and orientation of the road segment leading to the concerned end-point. These characteristics can then be utilized simultaneously, in a weighted calculation, in order to determine the optimal extension of the incomplete road segment.

5.1.3 Extrapolation of End-points

The above three properties (μ , σ , θ) form the criteria for choosing the best direction of extension. The process that leads to the optimal choice requires an analysis, which utilizes two factors. The first factor is the difference in average intensities between the

existing road and the possible extension. The second factor is the deviation of the possible extension from the original direction of the existing road. The reasons behind both requirements are the *assumption of similarity regarding the spectral properties and the assumption of linearity regarding the direction*. The assumption is based on a model of roads in high resolution satellite images that holds true to life. Firstly, at a high level of detail, a curvilinear structure which comprises discrete steps (in this case, pixels) can be resolved into short linear segments. The greater the degree of curvature, the shorter the linear segments are required to be for this to be true. This model applies well to urban roads as they generally do not curve abruptly along their main axis. Turns in urban roads are always smoothly curved, leading to the possible application of a piece by piece linear extension algorithm, using a unit of extension short enough to accommodate typical road curvatures. Secondly, a road in urban areas is sealed, thus ensuring that the surface reflectance is fairly consistent. This is the other part of the weighted probability calculation for choosing direction of extension.

The above rule (in italics) has been formalized by using absolute changes in mean reflectance and direction as evaluation factors. An artificial short road segment k is superimposed with the incomplete road skeleton for each end-point. Using different rotational representations of the artificial segment, different possible extensions are defined. From among these, the best match, in terms of the above rule, is determined. For each candidate extension, its variance σ_k^2 (defined similarly to Equation (10) but using the pixel locations of the overlaid road segment) and the difference $\Delta\mu_k$ (regarding the mean intensity μ) are calculated. Similarly, the difference $\Delta\theta_k$ of its orientation with respect to θ is recorded. For the final decision of selecting the most suitable extrapolation of the different possibilities indicated by the subscript k , the extracted parameters are fitted into the model, i.e. piecewise straightness of roads and locally invariant spectral

reflectance. Assuming zero mean Gaussian distributions for both criteria the best fit is defined by

$$\max_k \{N_{\mu}(0, \sigma^2, \sigma_k^2) + N_{\theta}(0, \sigma_t^2, \Delta\theta_k)\}, \quad (12)$$

where σ_t^2 is a user selected parameter that is selected according to the expected road straightness. However, to avoid extrapolation when it is not required, i.e. a dead-end road has to be left untouched, the σ_k^2 that matches Equation (12) is thresholded against a maximum tolerated variance.

5.1.4 Detection of Intersection

Once the optimal direction for extension has been chosen, it remains to ascertain whether the extension about to be made intersects with or touches the existing road skeleton or another extension made at a nearby end-point. In order to check for intersection, the extension of the analyzed end-point is made (conceptually) in the chosen direction. Then, if the newly determined road pixels coincide with any road skeleton pixels, the distances from the end-point to the intersection points are calculated and the minimum chosen. Any of the new road pixels that are further from the end-point than that minimum are rejected while those closer are added as part of the road network.

If an intersection exists, no new end-point is marked at the end of the newly made extension. Furthermore, if the new extension intersects or touches a previously made extension from some other end-point, the end-point marked at the end of the previously made extension is removed since the two corresponding extensions have connected and neither requires further extension.

5.1.5 Subsequent Iterations

Sub-sections 5.1.1 to 5.1.4 explain the entire first iteration. The first iteration of the algorithm is defined as ending when all the initial end-point candidates have been extrapolated by one step. Thus, the n^{th} iteration ends when the n^{th} set of end-points has been extrapolated. The n^{th} iteration produces the $(n+1)^{\text{th}}$ set of end-points. Figure 5-2 shows the result, after one iteration, on the image in Figure 5-1. The short bold road segments are the extensions added by the extrapolation algorithm.

If there was no intersection in the extension of an end-point in the $(n-1)^{\text{th}}$ iteration, the extrapolated road extension is added to the skeleton image together with a new end-point for the n^{th} iteration. At the same time, the extrapolation of an end-point provides the orientation θ of the existing road segment, with respect to the newly formed end-point. That is to say, the newly made extension is part of the road area (whose properties are to be calculated for comparison purposes) adjacent to the newly marked end-point. Thus, the orientation of the extension is subsequently used as an estimate for the direction of the existing road segment for the next iteration.

Based on that value and the length of the extension, the coordinates of the new end-point are calculated. Furthermore the spectral characteristics that describe expected road segments, i.e. μ and σ^2 , are updated. Using Equations (9) and (10), the parameters are recomputed by including the pixels of the original image i that match, with respect to their position, the added road segment in the skeleton.

Figure 5-3 shows the final result of the extrapolation of the initial skeleton in Figure 5-1. Though the skeleton Figure 5-3 is not exactly the same as a skeleton of all the roads in the original panchromatic image, it is accurate in that all the skeletal points, without being necessarily centered, definitely coincide with the roads in the original image.

Road Extraction from High-Resolution Satellite Images

Figure 5-2: Image after one iteration on Figure 5-1.

Figure 5-3: Final extrapolated road skeleton network.

5.2 Results and Discussion (Stage Two)

The algorithm was tested on the extracted road maps obtained at the end of stage one.

The original images were acquired by the QuickBird satellite. The images have a spatial resolution of 60cm per pixel and are panchromatic. The results are presented in Table 5-1.

Table 5-1: Results of the extrapolation (expressed as a percentage of total road skeleton length; The sum of the second, third and fourth columns represents the total road length)

No.	Initial skeleton	Correct Extrap.	Missing Road	Incorrect Extrap.
1	74.7	11.3	14.0	4.2
2	81.9	11.7	6.4	3.2
3	75.0	19.0	6.0	3.0
4	93.9	5.2	0.9	6.1
5	85.7	10.9	3.4	2.2
6	71.0	20.3	8.7	2.7
7	85.5	4.8	9.7	3.2
8	87.7	5.5	6.8	2.7
9	72.0	17.0	10.0	6.0
10	83.0	4.1	12.9	0.0
Avg	81.0	11.0	8.0	3.3

As mentioned in this chapter, a comparison of Figure 5-1 and Figure 3-13 shows that skeletonization causes a significant loss of road skeleton length. The completeness of the road map obtained from stage one reduces to an average of 81.0% when skeletonized (Table 5-1, second column). The projected (maximum possible) average value of road length extracted after extrapolation in Chapter 4 is 93.9%. As can be seen from Table 5-1, the actual average value of road length after extrapolation is 92.0% (sum of the second

and third column). Thus, the average correct extrapolation is 11.0% (third column). Therefore, the extrapolation algorithm not only recovers the loss of skeleton due to pruning, but also identifies more of the road network.

There are two types of shortfalls. The first shortfall is that the actual performance of the extrapolation algorithm falls short, by an average of 1.9%, of the performance projected by assessment of road seeds in the road maps at the end of stage one. The reason is that the end-point detection is not perfect. The emphasis was on preventing the incorporation of false end-points. Extrapolation of false end-points causes the extrapolation of true end-points to go awry due to intersections and makes the final road skeleton highly inaccurate. As a result of ensuring the exclusion of false end-points, some true end-points are also lost. Also, not all true end-points that are located are extrapolated completely. That is due to unpredictable variations in the intensity values of pixels adjacent to the road. These cannot be modeled and thus cannot be accounted for in the technique.

The second shortfall is that obviously not all the missing road segments are recovered by the extrapolation technique (8.0% of the road missing on average; calculated as the difference between a hundred percent and the sum of second and third columns). This is due to the reason that the extrapolation uses the pre-existing road map as a starting point. In most images, there are some lost roads, which have no remaining segment in the road map. Thus, they have no road seed that can be used to detect end-points and extrapolate from. Therefore, with respect to the proposed extrapolation technique, the first shortfall is the relevant one.

Chapter 6: Study on Conversion of Skeleton to Road Map: Determining the Sealed Surface

The studies in Chapters 3 and 5 describe two algorithms which, when applied in sequence to a test image, extract the road map, skeletonize it and then extrapolate the skeleton. The resultant image, therefore, contains a maximally extrapolated road skeleton. That image is now used as input for the algorithm described in this chapter. The algorithm described below converts the road skeleton to its corresponding road map (that is, determines the sealed surface), which is the reverse process of skeletonization. This road map is the final result of the entire body of work (on road extraction from high-resolution satellite images) in this thesis. The algorithms described in Chapters 3, 5 and 6, in sequence, comprise the entire road extraction algorithm. The division into separate studies is based on the order in which they were done and each study comprises an approximately equal amount of work, when considered in terms of both complexity and time.

The work on the detection of roads from satellite images began with the original panchromatic image from which an incomplete road map was extracted by the algorithm explained in Chapter 3. The map was then skeletonized and the skeleton was extrapolated by the algorithm explained in Chapter 5. To put the work of the previous studies in Chapters 3 and 5 into perspective, Figure 6-1 shows the original panchromatic sub-scene of Figure 3-1 with the corresponding (extracted and then extrapolated) road network superimposed. The image in Figure 3-1 was the starting point for the entire algorithm. Based on the algorithms explained in Chapters 3 and 5, the result achieved by this point is an extrapolated skeletal network of the roads in the original panchromatic image. The objective of this study is to determine the sealed surface of the roads in the image while minimizing incorporation of non-road pixels. This chapter explains in detail

how the map is obtained from the extrapolated skeleton. Using the extrapolated skeleton enables us to produce a more complete road map than was obtained at the end of the study presented in Chapter 3. Instead of fleshing out only the extrapolated parts of the skeleton - which would be faster since the road map for the original pre-extrapolation skeleton is available - this study has been approached as an attempt to find a general approach for obtaining the road map from the skeleton without any existing fragments of road map to guide the process. It is thus a greater challenge.

Figure 6-1: Original scene with superimposed extrapolated road network

6.1 Algorithm for Conversion of Skeleton to Road Map (Stage Three)

This algorithm (Stage Three) expands the road network to a full road map, which has the road edges clearly marked and road segments properly connected.

6.1.1 Extraction of Road Orientation

Calculating the orientation of the road is important as it provides information about the perpendicular direction to which the roadsides are found. Hence, before the sealed road surface can be determined, the orientation is determined at each and every point of the road network. At each point of the skeleton, a local neighborhood is extracted such that it contains a segment of the network. The size of the neighborhood is chosen based on its effect on the precision of the orientation. Smaller sizes generally lead to high inaccuracy due to possible inaccuracies in the road network, while larger sizes handicap the detection of curves. For the example shown in Figure 6-1, a neighborhood of 15×15 pixels was used. The actual determination of the orientation is equivalent to Equation (11) on page 69.

6.1.2 Estimation of Road Width

Given the instantaneous orientation θ of the network, at a particular point, the direction of analysis is fixed as $(\theta+90)$, that is, the analysis is carried out perpendicular to the direction of the road at that point. The objective is to obtain an estimate of the width of the road at that point. Towards this end, a cross-section is extracted from the image. To implement this, the sub-scene is rotated using bi-cubic interpolation to resample the data to integer pixel positions. An alternative is to use a linear structuring element that would be rotated in order to align it perpendicular to the road segment and then sampling could be done. (However, that leads to problems later in the algorithm. Basically, the range of values extracted, though continuous with respect to their position along the x-axis of the plot in Figure 6-4, is not continuous in the actual image due to the horizontal parsing done in order to sample the pixel values along the rotated structuring element. Thus, if the image itself is rotated and the structuring element remains horizontal, a range of continuous positional values denotes a range of continuous connected pixels in the image. This will be clearer after reading Section 6.1.3.). An example of a typical cross-

section is shown in Figure 6-2 by the solid line. The y-axis represents intensity and the x-axis represents pixel positions along the cross-section. The cross-section, a typical case, does not contain enough information to accurately differentiate and extract the road area from the surrounding non-road area. An example of a more complicated cross-section is depicted by the dotted line in Figure 6-2.

Figure 6-2: Cross-sections at two points of the skeleton

6.1.3 Multiple Cross-Section Extraction

The more complicated cross-section depicted by the dotted line in Figure 6-2 does not have a consistent intensity value across several positional values. Also, the pixels in between the two maxima contain two completely different levels of intensities. Thus, using a single cross-section per pixel in order to determine the corresponding road width is not a viable option. Hence, to determine the road's parameters from the cross-section,

a local neighborhood of pixels around the pixel being analyzed is extracted. As adjacent pixels have similar slope values, their respective cross-sections are typically parallel and therefore can be used to track consistent trends in intensity values. Applied to the pixel whose cross-section is depicted by the solid line in Figure 6-2, the result is as depicted in Figure 6-3.

Figure 6-3: Multiple cross-sections for pixel used in Figure 6-2

6.1.4 Variance Extraction

The positions for which the intensity values match closely across the extracted cross-sections of neighborhood pixels obviously represent road pixels. This is due to the observation that the road displays similar spectral properties, but the areas physically adjacent to the road vary in terms of their spectral properties, even when close by. This is measured by the variance σ over multiple cross-sections c_j . At positions with similar

intensities for all cross-sections the variance value is low while at others it is significantly higher. The variances for the cross-sections c_j in Figure 6-3 are shown in Figure 6-4. For successful cases, the road width is typically characterized by the valley between the two peaks and is used for the parameterization of the sealed road surface.

Figure 6-4: Plot of variance versus position

Using the described approach, the superimposed road network in Figure 6-1 is expanded into the road map shown in Figure 6-5. As can be seen, not all the pixels have an identifiable cross-section. This is because the analysis does not yield a continuous range of low variance values flanked by high variance values for all pixels in the road network. Thus, road surface determination, at this juncture, is possible only for those pixels which exhibit a variance-position plot of the kind in Figure 6-4. In some cases, even though the variance-position plot is as shown in Figure 6-4 and it is possible to extract positional

values for the road surface, the range is discarded and that pixel does not have a corresponding cross-section added to the image. This is when the range is too small or too large to be practical. In such a case, it is obviously a false range obtained by accidental spectral features and it does not actually represent a valid cross-section of road area.

Figure 6-5: Result of road surface determination using the road network in Figure 6-1

6.1.5 Interpolation of Road Surface

For each pixel in the network that does not yield a road surface through variance analysis, the solution is to locate the nearest pixel that does. Then, the averaged positional values of the neighboring pixels are used to mark the sealed surface corresponding to the pixel being analyzed. With this method, the road surface for all pixels in the network can be determined. This approach works because the cross-sections of a road at two adjacent pixels are quite similar. The positional values of adjacent pixels

can thus be used when the analysis for a particular pixel does not yield a usable positional range. Note that obstacles like vehicles are inherently removed and therefore more accuracy is achieved.

The interpolation has a second stage. It involves addressing the gaps in the roadmap that are created due to inaccurate orientation calculations. An incorrect value for the orientation does not prevent the algorithm from finding a usable positional range. However, when these values are used at adjacent pixels (during the first interpolation stage) that have accurate orientation values, it results in uneven road surface determination. Orientation calculations result in inaccurate values when the road network branches in the neighborhood of the pixel under analysis. Thus, the recovered road map suffers from uneven contours near intersections. These are addressed by the second stage of interpolation. It uses a neighborhood from around each non-road pixel in order to correct the negative effects of incorrect orientation values. The neighborhood is extracted for each non-road pixel and if the pixel lies along a straight line path between two road pixels in the neighborhood, it is classified as a road pixel. The size of the neighborhood chosen is again critical to the success of such an interpolation. Firstly, the neighborhood is kept very small based on the criteria that a road would not curve abruptly within a short length. Secondly, the criterion used is that the two road pixels used as reference points for interpolation should belong to the same connected component. This prevents interpolation between adjacent non-connected road segments because that leads to incorrect road pixel classification. When the two interpolation techniques are applied to the image in Figure 6-5, the result is the image in Figure 6-6.

Figure 6-6: Result obtained after interpolation

6.2 Results (Stage Three)

The algorithm was tested on the extrapolated skeletons obtained at the end of stage two.

The original panchromatic images were obtained from the QuickBird satellite and have a spatial resolution of 60cm per pixel.

Table 6-1: Results of the extraction (expressed as a percentage of total road length)

No.	Extrapolated skeleton	Road map Completeness	Road map Correctness
1	86.0	86.0	88.0
2	93.6	91.6	97.8
3	94.0	94.0	92.8
4	99.1	99.1	89.0
5	96.6	96.6	88.9
6	91.3	88.3	93.3
7	90.3	90.3	92.8
8	93.2	89.2	92.0
9	89.0	87.0	93.0
10	87.1	87.1	93.7
Avg	92.0	90.9	92.1

Using the extrapolated skeleton with an average completeness of 92.0% (Column 2, Table 6-1), this step of the algorithm determines the corresponding sealed surface of the road map. As the second column in Table 6-1 shows, the completeness of the obtained road map is 90.9%, that is, it falls short by 1.1% of the total road length. But, the shortfall is not significant considering that 98.8% (90.9/92.0) of the available skeleton is converted to sealed surface by the algorithm. Secondly, as shown in Table 5-1, the average incorrect extrapolation was 3.3%. This incorrectly extrapolated skeleton is converted to false road area. When combined with the other pre-existing noise in the

images (from stage one), it produces a correctness of 92.1%. That is a total of 7.9% incorrectly incorporated image areas in the final result.

Thus, by the application of the skeleton extrapolation and skeleton to road map conversion algorithms (explained in Chapters 5 and 6), the skeleton was first increased to an average of 92.0% from 81.0% and then this extended skeleton was fleshed out into a road map with the sealed surface clearly marked. The final road map has a completeness of 90.9% and a correctness of 92.1%. Therefore, the skeleton extrapolation (stage two) algorithm and the skeleton to road map conversion (stage three) algorithm have, together, increased the completeness of the road map as compared to the road map extracted using the road detection (stage one) algorithm alone.

Chapter 7: Conclusions and Future Work

7.1 Summary of Work Done

The objective was to create a method to extract roads from high-resolution, panchromatic satellite images of urban areas. Towards that end, three algorithms were created. Taken together, they comprise a complete method for extraction of road features. Each algorithm corresponds to a stage in the extraction process and these are described in detail in Sections 7.1.1 to 7.1.3. Figure 7-1 summarizes the algorithms (stages) in a concise, diagrammatic form.

Road Extraction from High-Resolution Satellite Images

Figure 7-1: Diagrammatic representation of entire road extraction method (3 stages/algorithms)

7.1.1 Stage One

Figure 7-2: Input image and final result for stage one

The first stage utilizes the original panchromatic image as the starting point. The algorithm developed for the first stage consists of a number of steps. Firstly, the image is thresholded using a range of values. Pixels with intensity values within that range are retained as foreground while those pixels that lie outside the range are consigned to the background and lost. The thresholding utilizes a range of values as sealed urban roads typically have a certain range of intensities, using which it is possible to isolate the roads. The range also includes a large amount of non-road pixels which are consequently retained in the foreground. Therefore, the image, at this point, contains the road features along with a great deal of noise. The majority of this noise is removed in the remaining steps of the first stage. But before the image is processed to remove noise, a technique is applied that detects road dividers. Typically, urban roads contain multiple lanes and when the image is thresholded, the lane dividers are lost as they have high intensity values. This partitions the roads which contain the dividers and these roads would be lost in the subsequent steps if this effect of thresholding is not corrected before further processing is carried out. The technique created parses the image, detects the dividers and merges them together, leading to a complete road without divisions. Then, another

parsing technique is employed that detects and removes thin, curvilinear structures. Such structures do not represent roads and can be safely removed without adverse effects on the road features present in the image.

Following the parsing, connected component analysis is carried out. The image is segmented into connected components using connected component labeling and the analysis isolates components that do not represent roads based on their morphological properties. Accordingly, components that are too small, too dense and too irregular to represent roads are removed. These properties are easily calculated for each component once the components have been labeled. For determining irregularity of the component, a value is calculated that represents the level of branching in the skeleton of the component. The idea is, the higher the branching within the component, the greater the irregularity. That is to say, branching within the skeleton is a good measure of the irregularity of a component. The components with a high level of branching are removed. The removal of the components with the specified properties deletes a majority of the components present in the image. Thus, there is a significant reduction in the number of non road pixels and the result at the end of the first stage is an incomplete road map in binary form. There is still some noise remaining in the binary image. However, this minimal amount of noise cannot be removed by utilizing the properties of the components that they make up or by other techniques. Any attempt to further reduce the noise results in severe loss of road area and is counter-productive in terms of the objective. Thus, at this point in the method, the remaining noise is left as is and the focus is shifted to completing the road map (stage two). At the end of stage one, 88.2% of the roads are detected and retained in the final binary image.

7.1.2 Stage Two

Figure 7-3: Input image, pruned skeleton and extrapolated final image for stage two

The second stage uses as a starting point the incomplete road map extracted in stage one. Since the road map is incomplete, the objective is to complete it as much as possible without incorporating any more noise. The technique chosen for completing the road map is extrapolation. As the road map is, on average, 88.2% complete, there is substantial scope for improvement of the completeness with the aid of extrapolation. For extrapolation, the end-points of the incomplete road map are required. Thus, some pre-processing is done before the extrapolation algorithm is applied. The pre-processing comprises skeletonization of the extracted road map and subsequent detection of end-

point candidates. The road map is converted to skeleton as it aids the detection of the end-points. Also, the extrapolation can then be carried out piece-wise in a linear manner, with each extrapolated section being single pixel wide rather than use sections with wide cross-sections. Once the road map has been skeletonized, the end-point candidates can be found by detecting points at the end of linear stretches of skeleton. These are then confirmed by using a structuring element approach applied to the extracted road map. In order to avoid points at the end of stray branches being marked as end-points, pruning of the skeleton is carried out. This reduces the length of the skeleton to an average completeness of 81.0% from the 88.2% average completeness of the road map. Thus, the extrapolation aims not only to complete the road network, but it must also first recover the road length lost due to pruning.

The extrapolation algorithm utilizes an understanding of typical urban roads to extrapolate the incomplete roads. At each end-point candidate, several possible extensions are identified, each in a different direction. The concept is that at each end-point candidate, a search must be made to find the direction such that the extension drawn is the most likely extension to the previously extracted road segment leading towards the concerned end-point. In order to rank the suitability of all possibilities, a probability approach is chosen to describe the model based on two factors.

The first factor is the difference in average intensities between the existing road and the possible extension. The second factor is the deviation of the possible extension from the original direction of the existing road. The reasons behind both requirements are the *assumption of similarity regarding the spectral properties and the assumption of linearity regarding the direction*. The assumption is based on a model of roads in high resolution satellite images that holds true to life. Firstly, at a high level of detail, a curvilinear structure which comprises discrete steps (in this case, pixels) can be resolved

into short linear segments. The greater the degree of curvature, the shorter the linear segments are required to be for this to be true. This model applies well to urban roads as they are never designed to curve abruptly along their main axis. Turns in roads are always smoothly curved, leading to the possible application of a piece by piece linear extension algorithm, using a unit of extension short enough to accommodate typical road curvatures. Secondly, a road in urban areas is sealed, thus ensuring that the surface reflectance is fairly consistent.

When evaluating the stage one algorithm, the best possible performance of the (at that time, still to be implemented) extrapolation algorithm was projected and calculated. The calculation was based on the presence of road seeds in the extracted road map. A particular road is said to have a road seed (and is thus completely recoverable by extrapolation) if some part of that road is present in the final extracted image. The optimal performance requires that the end-points associated with each road seed are detected and that the extrapolation algorithm works perfectly. Based on road seeds, the road map can, optimally be increased to an average of 93.9% completeness, an average extrapolation of 5.7%. The projected final value is not a 100% as some roads are completely lost and have no road seeds in the final image.

The actual performance of the extrapolation algorithm is an average completeness of 92.0%. However, it must be remembered that the extrapolation algorithm starts with a skeleton of 81.0% completeness. Thus, 92.0% represents an average extrapolation of 11.0%. The extrapolation algorithm recovers the road length lost due to pruning and most of the road length recoverable by using the road seeds. The main reason that the performance of the extrapolation algorithm falls short of 93.9% (the optimal performance projected at the end of stage one, based on presence of road seeds) by 1.9% is that the detection of end-points of the road seeds is not perfect. The main emphasis was on

preventing the incorporation of false end-points. Extrapolation of false end-points causes the extrapolation of true end-points to go awry due to intersections and makes the final road skeleton highly inaccurate. Due to the exclusion of false end-points, some true end-points are also lost. Also, not all true end-points that are located are extrapolated completely. That is due to unpredictable variations in the intensity values of pixels adjacent to the road. These cannot be modeled and thus cannot be accounted for in the technique.

7.1.3 Stage Three

Figure 7-4: Input image and final result for stage three

The third stage takes the (almost complete) skeleton and converts it back to road map form, that is, the reverse process of skeletonization. It is based on the concept that the width of a road is perpendicular to the main direction of the road. Thus, the direction of the road is determined at each point on the skeletal network and analysis is done at right angles to that direction. The first step is extraction of the cross-section of the road (at each road skeleton pixel) in the perpendicular direction. From this cross-section, the pixels belonging to the road have to be identified and separated from the pixels immediately adjacent to the road. However, it was found that a single cross-section is not

sufficient to identify the road pixels as the intensity values of adjacent non road pixels are quite similar to those of the pixels on the road's edges. Therefore, multiple cross-sections are extracted from both sides of the skeletal pixel being analyzed and these cross-sections are found to be highly correlated. The concept is that adjacent cross-sections match very closely for the pixel positions that belong to the road while they vary widely for non-road areas. Thus, by calculating variance for each position of the cross-section, it is possible to identify the road pixels within the cross-section and thus, convert the skeleton to a road map.

Some cases do not yield usable cross-sections due to unpredictable variations. For these positions, the positional values are copied from the nearest skeletal pixel that does yield a usable cross-section. This interpolation technique is applied to the entire image until no further improvement is possible. This technique converts the 92.0% complete skeleton to 90.9% complete road map with an average correctness of 92.1%. The imperfect correctness is due to the noise remaining from stage one which is still present in the final image at the end of stage three. Thus, the overall method is reasonably successful and the stated objective has been achieved.

7.2 Critical Review

This sub-section takes a critical look at the work reported in this thesis. This work is applicable to high-resolution, panchromatic images of urban areas obtained at a resolution of 60cm per pixel. The algorithms are designed for such images and will thus not be as effective at extracting roads from low resolution images and/or images where the roads are not sealed (usually found only in rural areas). There are three algorithms for the three different stages. For each algorithm, there are certain particular requirements and arising from those requirements, certain corresponding limitations.

It must be noted that no algorithm is equally effective with all kinds of images. This is clear from the literature review in Chapter 2. Algorithms (in published work) are designed for and trained with a specific type of image in order to maximize their effectiveness. This is an appropriate strategy as a single satellite alone has a prodigious output of images of a given resolution. Extracting features from them requires specialized algorithms that perform optimally for those images. Following that philosophy, the algorithms in this thesis are similarly specialized for a particular type of image. The rest of this sub-section details how each algorithm performs for images that differ from the ones actually used in this thesis and how it can be adapted to those images.

In stage one, the algorithm could be adapted to different types of images with the modification of certain parameters. The parameters used depend on the resolution of the images. For example, when removing small components during stage one, the threshold used for the size of unwanted components is determined by the 60cm resolution. Such adaptations would enable the stage one algorithm to function for images of a lower resolution. But if the roads were unsealed, the algorithm would function properly (without modifications) as long as the resolution was unchanged.

For the stage two algorithm, the characteristics of urban roads are utilized, in that they have a fairly uniform surface intensity over short stretches and generally, do not curve abruptly along their main axis. The mathematical representation of these two facts is the weighted sum of probabilities used for extrapolation (Chapter 5). However, were we not dealing with sealed urban roads, the algorithm would still function if the roads had a fairly uniform intensity. However, if that does not hold true (as may happen in rural areas), the weight of the ‘spectral intensity similarity’ factor could be reduced with the factor for ‘change of direction’ given a higher weight. Similarly, if the roads display

abrupt changes of direction along their main axis, the weight of the ‘change of direction’ factor could be reduced with the factor for ‘spectral intensity similarity’ given a higher weight.

In the case that one property is not always true, the algorithm’s performance would be slightly worse, but still acceptable. However, in the case that both properties do not hold true simultaneously, the performance of the algorithm would be significantly worse. Thus, in the case of the stage two algorithm, its performance (for a different type of image) depends in a straightforward manner on how well the roads conform to the model described in Chapter 5. The parameters involved reflect the importance given to the two factors based on observations of ground truth.

In the stage three algorithm, there is no requirement for any particular type of image or any specific road model. The algorithm used in stage three utilizes a very general fact, that is, the width of the road is perpendicular to the main longitudinal axis of the road. This holds true for all types of roads in all areas. Thus, the stage three algorithm could be applied to any skeleton to obtain back the road map. The only pre-requisite is the width of the road. The road should be wide enough to produce a variance-position graph of the kind shown in Figure 6-4. Without sufficient width, it is not possible to locate the typical valley in the graph which represents the cross-section of the road. While a narrow road would produce a correspondingly narrow valley, extracting such a valley is not easy due to inclusion of a large number of false positives. Differentiating between an actual narrow valley and a random, unpredictable fluctuation in the variance-position graph is not possible and the inclusion of false positives would be too high. It does not matter whether the lack of sufficient width is due to a narrow road in a high-resolution image or a normal road in a low-resolution image. As long as the width is lacking, the algorithm

will not perform optimally. Having said that, the stage three algorithm will function for high-resolution images of all kinds as the roads display a certain minimum road width.

7.3 Alternative Approaches

There exist alternatives to the approach used in this thesis. Generally speaking, there are two broad approaches. The first is that of removing non-road areas bit by bit until only road areas are left and these can then be connected together and integrated into a network, if not already so. This is a top-down approach and is the one used in this thesis. The second broad approach is that of identifying road features from the original image and retaining only those. This is a bottom-up approach. The first approach has the advantage of extracting a high percentage of original road area but the disadvantage of incorporating a certain amount of non-road area. On the other hand, the second approach has the advantage of incorporating very little (if any) non-road area but the disadvantage of extracting a relatively low percentage of original road area. There is an example of the second approach in the literature (Gong and Wang (1997)). In this particular case, the authors report 74.5% of total road length extracted using a classification technique. However, it is not necessary that the performance of the bottom-up approach will be worse. Depending on the differentiation present amongst the various surface features in the images used, a classification technique could outperform the top-down approach used in this thesis. One possibility is to train a neural network to identify road features using a large set of sample images (comprising similar road features) and then use it on a similar set of test images. The algorithm would require careful tuning and a long time duration for execution, but it is a possible route to follow for implementing the alternative approach.

7.4 Future Work

There is definitely room for improvement and further work could be done to better the results obtained in this thesis. Firstly, there is the issue of the remaining noise (non-road areas) in the final image. Certain test images possess large amounts of noise at the end while others possess hardly any. What is needed is a sophisticated technique to separate the remaining noise from the road map. This would basically involve analyzing the final binary image in order to remove this noise and would lead to higher correctness. The problem here is that most of these non-road areas are attached to the road map. They are thus not separable by the connected component analysis used in stage one. Such noise could possibly be removed by a technique that analyses the shape of the non-road feature. Such a technique would be dependent on the size of the neighborhood chosen for analysis and would not remove all such noise, but it might likely remove non-road areas that were attached to lengthy linear segments of road. Such road segments have a distinctive shape and are easy to detect. Any attached non-road areas would be comparatively misshapen and irregular. Thus, this technique would be similar in concept to the removal of irregular components in stage one, but more complicated in terms of implementation.

A possible approach for removing non-road areas not attached to the road map is further connected component analysis. Connected component analysis was used during stage one, but still some noise remained. This is due to the fact that the characteristics of the remaining non-road areas were similar to those of some of the unconnected short road segments. Thus, those non-road areas were left as is because removing them would have caused a large scale loss of road length at the same time. The idea proposed now is that connected component analysis could be utilized after implementing an algorithm that *completes the road map into one fully connected network*, thus isolating the non-road areas from the road-based areas. In a fully connected network, all short road segments

are connected to the main road network and the noise is thus isolated when connected component labeling is done. Then, the detection and removal of the noise by CCA is easy as the short road segments (previously confused with the noise) would no longer be lost. However, the completion of the road network is easier said than done as the attempt in stage two to complete and fully connect the road map (using extrapolation) did not result in a 100% connected road map, primarily due to the lack of road seeds for certain road segments. Thus, the challenge is to recover those road segments that were completely lost in stage one and did not leave behind any corresponding road seeds. Without road seeds, they cannot be recovered by extrapolation as it requires a road seed as a starting point. Therefore, a technique is needed that can recover those road segments without use of road seeds. As a suggestion, such a technique could calculate the collective properties of all the road area pixels extracted at the end of stage three. These properties could then be utilized to identify areas within the image that possess the same collective characteristics. Such an approach might perform well when an image has an above average percentage of missing road area and thus, much room for improvement.

Another area open for future work is removing the impediments of vehicles, trees and shadows before applying the extraction algorithms. Loss of road area in stage one is caused primarily by shadows of adjacent trees etc. which have different intensity values. The idea is that if these impediments could be identified beforehand and removed without loss of adjacent or underlying road area, the results of stage one extraction would improve quite dramatically, leading to higher completeness. The major impediment is seen to be shadow area. Besides using images where the sun is directly overhead and thus casting no large shadows, a possible solution is to utilize neural networks. Neural networks could be trained on sample images to detect shadows and then applied to the test images as the very first step (while the image is still in its original panchromatic

form). Once the shadow areas have been identified, the land area hidden by them can then be replaced by useful information in terms of surface features. This requires a sophisticated technique that can estimate the spectral characteristics of the underlying surface. The estimation could use, as a guide, the spectral characteristics of all the different surface features nearby, their corresponding shapes and their projected directions of continuity.

These are some suggestions as to the direction of the future work to improve upon the work in this thesis. Incorporation of these suggestions could lead to a road map with a higher completeness and correctness than the current values of 90.9% and 92.1% respectively.

References

Barzohar, M., Cooper, D.B., Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation, Proceedings of Computer Society Conference on Computer Vision and Pattern Recognition, pp. 459 –464, 1993.

Bernsen J., Dynamic thresholding of grey level images, Proceedings of International Conference on Pattern Recognition, pp. 1251-1255, 1986.

Burge M. and Monagan G., Using the Voronoi tessellation for grouping words and multipart symbols in documents, Technical report no. 131-95, 1995.

Byoung-Ki Jeon, Jeong-Hun Jang, Ki-Sang Hong, Road detection in spaceborne SAR images using a genetic algorithm, IEEE Transactions on Geoscience and Remote Sensing, Volume 40 Issue 1, pp. 22 –29, 2002.

Cheng H. and Bouman C., Trainable Context Model for Multiscale Segmentation, Proceedings of the IEEE International Conference on Image Processing, Volume 1, pp. 610-614, 1998.

Chinsu Lin, Chuin-Mu Wang, Chein-I Chang, Application of generalized constrained energy minimization approach to urban road detection, Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Volume 5, pp. 2080 – 2082, 2000.

Dell'Acqua, F., Gamba, P., Detection of urban structures in SAR images by robust fuzzy clustering algorithms: the example of street tracking, IEEE Transactions on Geoscience and Remote Sensing, Volume 39 Issue 10, pp. 2287 –2297, 2001.

References

Drivas D. and Amin A., Page Segmentation and Classification Utilizing Bottom-Up Approach, Proceedings of the Third International Conference on Document Analysis and Recognition, pp. 610- 614, 1995.

Duta, N., Road detection in panchromatic SPOT satellite images, Proceedings of the 15th International Conference on Pattern Recognition, Volume 4, pp. 308 -311 Volume 4, 2000.

Etemad K., Doermann D., and Chellappa R., Multiscale Document Page Segmentation Using Soft Decision Integration, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 19, pp. 92-96, 1997.

Garnesson, P., Giraudon, G., Montesinos, P., An image analysis, application for aerial imagery interpretation, Proceedings of the 10th International Conference on Pattern Recognition, Volume i, pp. 210 –212, 1990.

Geman D., Jedynak B., Detection Of Roads In Satellite Images, International Geoscience and Remote Sensing Symposium, 'Remote Sensing: Global Monitoring for Earth Management', Volume 4, pp. 2473 –2477, 1991.

Gervin, J.C., Ragan, R.M., Identifying roads and houses from scanned aerial photography and combining them with TIGER and digital terrain data of suburban areas, International Geoscience and Remote Sensing Symposium, 'Better Understanding of Earth Environment', Volume 3, pp. 1550 -1553, 1993.

Gong, P., Wang, J., Road network extraction from airborne digital camera images: a multi-resolution comparison, IEEE International Geoscience and Remote Sensing, 'Remote Sensing - A Scientific Vision for Sustainable Development', Volume 2, pp. 895 -897, 1997.

References

Govindaraju V., Srihari R.K., Srihari S.N., Handwritten text recognition, World scientific Document analysis systems, pp. 288-304, 1995.

Guindon, B., Application of spatial reasoning methods to the extraction of roads from high resolution satellite imagery, Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Volume 2, pp. 1076 -1078, 1998.

Hae Yeoun Lee, Wonkyu Park, Heung-Kyu Lee and Tak-gon Kim, Towards knowledge-based extraction of roads from 1 m-resolution satellite images, Proceedings of the 4th IEEE Southwest Symposium Image Analysis and Interpretation, pp. 171 – 176, 2000.

Heutte, L., Ogier, J.M., Lecourtier, Y., Olivier, C., Two aspects of automatic map treatment: road and texture extractions, Proceedings of the 11th IAPR International Conference on Pattern Recognition, Volume III, Conference C: Image, Speech and Signal Analysis, pp. 109 –112, 1992.

Hinds S.C., Fisher J.L. and Amato D. P. D', A document skew detection method using run-length encoding and Hough transform, Proceedings of the 10th ICPR, pp. 464-468, 1990.

Huber, R., Lang, K., Road extraction from high-resolution airborne SAR using operator fusion, IEEE International Geoscience and Remote Sensing Symposium, Volume 6, pp. 2813 -2815, 2001.

Ilg, M., Knowledge-based understanding of road maps and other line images, Proceedings of the 10th International Conference on Pattern Recognition, Volume i, pp. 282 -284, 1990.

References

- Iisaka, J., Sakurai-Amano, T., Lukowski, T.I., Automated detection of road intersections from ERS-1 SAR imagery, International Geoscience and Remote Sensing Symposium, 'Quantitative Remote Sensing for Science and Applications', Volume 1, pp. 676 -678, 1995.
- Jain A. and Zhong Y., Page Segmentation Using Texture Analysis, Pattern Recognition, volume 29, pp. 743-770, 1996.
- Jain A. and Bhattacharjee S., Text Segmentation Using Gabor Filters for Automatic Document Processing, Machine Vision and Applications, Volume 5, pp. 169-184, 1992.
- Jain A. and Yu B., Document Representation and Its Application to Page Decomposition, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 20, pp. 294-308, 1998.
- Jianhua Xuan, Adali, T., Task-specific segmentation of remote sensing images International Geoscience and Remote Sensing Symposium, 'Remote Sensing for a Sustainable Future.', Volume 1, pp. 700 -702, 1996.
- Kamel M. and Zhao A., Extraction of binary character graphics images from greyscale document images, Graph. Models and Image Processing, Volume 55, no. 3, pp. 203-217, 1993.
- Karimi, H.A., Xiaolong Dai, Khattak, A.J., Hummer, J.E., Automated roadway feature extraction from high-resolution satellite images, Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Volume 4, pp. 2065 - 2067, 1998.

References

Kaufmann G., Bunke H. and Ha T.M., Recognition of cursively handwritten words using a combined normalization/perturbation approach, Fifth international workshop on frontiers in handwriting recognition, pp. 17-22, 1996.

Kise Koichi, Sato Akinori and Iwata Motoi, Segmentation of page images using the area Voronoi diagram, Computer Vision and Image Understanding, Volume 70, no. 3, pp. 370-382, 1998.

Legourgiois F., Bublinski Z., and Emptoz H., A Fast and Efficient Method for Extracting Text Paragraphs and Graphics from Unconstrained Documents, Proceedings of the 11th International Conference on Pattern Recognition, pp. 272-276, 1992.

Lloyd D. E., Automatic target classification using moment invariants of image shapes, Technical report, RAE IDN, UK, 1985.

Maderlechner, G., Mayer, H., Automated acquisition of geographic information from scanned maps for GIS using frames and semantic networks, Proceedings of the 12th IAPR International Conference on Pattern Recognition, - Conference B: Computer Vision & Image Processing, Volume 2, pp. 361 -363, 1994.

Magerl, G., Pritzl, W., Frohling, P.W., Remote Sensing Of Road Condition, International Geoscience and Remote Sensing Symposium, 'Remote Sensing: Global Monitoring for Earth Management', Volume 4, pp. 2137 –2140, 1991.

Mukherjee, A., Parui, S.K., Chaudhuri, D., Chaudhuri, B.B., Krishnan, R., An efficient algorithm for detection of road-like structures in satellite images, Proceedings of the 13th International Conference on Pattern Recognition, Volume 3, pp. 875 -879, 1996.

References

Netanyahu, N.S., Philomin, V., Rosenfeld, A., Stromberg, A.J., Robust detection of road segments in noisy aerial images, Proceedings of the 13th International Conference on Pattern Recognition, Volume 2, pp. 151 -155, 1996.

Okutomi, M., Noguchi, S., Extraction of road region using stereo images, Proceedings of the Fourteenth International Conference on Pattern Recognition, Volume 1, pp. 853 -856, 1998.

Otsu N., A threshold selection method from grey level histogram, IEEE Transactions on Syst. Man. Cybern., Volume 9, no. 1, pp. 62-66, 1979.

Peteri, R., Celle, J. and Ranchin, T., Detection and extraction of road networks from high resolution satellite images, Proceedings of the International Conference on Image Processing, Volume 1, pp. I-301-4, 2003.

Ruskone, R., Guigues, L., Airault, S., Jamet, O., Vehicle detection on aerial images: a structural approach, Proceedings of the 13th International Conference on Pattern Recognition, Volume 3, pp. 900 -904.3, 1996.

Saitoh T. and Pavlidis T., Page segmentation without rectangle assumption, 11th ICPR, pp. 277-280, 1992.

Simon A., Pret J., and Johnson A., A Fast Algorithm for Bottom-Up Document Layout Analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 19, pp. 273-276, 1997.

Solaiman, B., Fiset, R., Cavayas, F., Automatic road extraction using fuzzy mask concepts, Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Volume 2, pp. 894 -896, 1998.

References

Tieling Chen, Jinfei Wang, Kaizhong Zhang, A wavelet transform based method for road extraction from high-resolution remotely sensed data, IEEE International Geoscience and Remote Sensing Symposium, Volume 6, pp. 3369 –3371, 2002.

Tupin, F., Maitre, H., Mangin, J.-F., Nicolas, J.-M., Pechersky, E., Detection of linear features in SAR images: application to road network extraction, IEEE Transactions on Geoscience and Remote Sensing, Volume 36 Issue 2, pp. 434 –453, 1998.

Tupin, F., Houshmand, B., Datcu, M., Road detection in dense urban areas using SAR imagery and the usefulness of multiple views, IEEE Transactions on Geoscience and Remote Sensing, Volume 40 Issue 11, pp. 2405 –2414, 2002.

Wang, F. and Newkirk, R., A knowledge-based system for highway network extraction, IEEE Transactions on Geoscience and Remote Sensing, Volume 26 Issue 5, pp. 525 –531, 1988.

Wenzhong Shi and Changqing Zhu, The line segment match method for extracting road network from high-resolution satellite images, IEEE Transactions on Geoscience and Remote Sensing, Volume 40 Issue 2, pp. 511 –514, 2002.

White J. M. and Rohrer G. D., Image thresholding for optical character recognition and other applications requiring character image extraction, IBM J. Res. Devel. Volume 27, no. 4, pp. 400-411, 1983.

Xuan Li, Yanli Qiao, Weining Yi, Zhifeng Guo, The research of road extraction for high resolution satellite image, Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Volume 6, pp. 3949 – 3951, 2003.

References

Yacoubi El, Bertille J. -M. and Gilloux M., Conjoined location and recognition of street names within a postal address delivery line, International conference on document analysis and recognition, volume 2, pp. 1024-1027, 1995.

Yeh P. S., Antoy S., Litcher A. and Rosenfeld A., Address location on envelopes, Pattern Recognition, 213-227, 1987.

Appendix A

Road Extraction Test Results: Original Images and Final Results

Appendix A: Road Extraction Test Results: Original Images and Final Results

This chapter contains the results of the remaining nine test images. The non-road areas not attached to the road map (in the final results on the right) have been removed for clarity.

Appendix A: Test results

Appendix A: Test results

Appendix B

Document Image Analysis

Chapter B1: Introduction to Document Analysis

B1.1 Background

All over the world, there are huge amounts of paper documents stored for purposes of records and retrieval. There are two basic problems with the use of paper. Firstly, accessing a particular record involves searching through massive stacks of paper and secondly, paper, being a degradable substance, tends to lose its contents with time. The information stored is thus lost irretrievably and causes corresponding monetary loss. The recent trend has been to replace all paper records with digital data. The paper is scanned and is stored as an image file. This brings about two advantages. The data can be easily located and more importantly, it is loss-proof as copies can be easily made and stored at various locations, while still reducing storage space requirements by several orders of magnitude.

Computers use the language of logic and we can thus communicate our requirements to them and they can provide us with the desired result. A whole area of research now deals with the creation of methods for intelligently scanning and understanding the contents of a document, the field of document image analysis and understanding.

B1.2 Document Analysis and Understanding

A document can be analyzed in two different ways. If the analysis is based upon human understanding, it utilizes the logical structure. However, if the analysis is based upon the layout of fundamental building blocks such as letters, words, lines, diagrams and paragraphs, the geometric structure is utilized. The geometric structure of a document is easier to determine as it is objectively defined and not subject to subjective interpretations. Document layout analysis and understanding is defined as the extraction of the geometric and logical structures of a document.

Appendix B: Document Image Analysis

It aims at conversion of document images to symbolic form for modification, storage, retrieval, reuse and transmission. It aids the transition from books to a paperless world. The document accepted as input can be of almost any kind. The processing should identify the various text areas, diagrams and the layout, segment the text into lines, words and finally identify the words based on an unlimited lexicon.

The symbolic form is a high level representation that describes the document's layout and logical structure in addition to providing its information content. At the fundamental level, research on methods of handling document images revolves around the need to automate the process of extracting information from paper documents; the challenge now facing us is to devise a complete method of handling such document images. Manpower cannot be spared or afforded for the manual perusal of such documents; thus, machine-based methods are the only solution. This has led to the immense amount of effort that has been put into coming up with innovative methods to carry out document analysis.

B1.2.1 Practical Issues

The basic idea is simple but the practical application is much more complex. The first issue is that documents vary so widely in their content and format. A human being, after years of experience and schooling, can recognize handwriting, ascertain the format of a document and the orientation of the contents with respect to the document itself. He/she can differentiate between what is spurious noise and what is actually useful. He can point out boundaries between conceptually distinct parts of the document and organize all the above information into a coherent whole to gain an understanding of what the document contains. However, the processes at work within the human brain are complex and not yet understood. Thus, they cannot be replicated in software yet, though some progress has been made in that direction using techniques such as Neural Networks. So far, most work done has been using straight-ahead logic. The image is analyzed section-by-section

and patterns are searched for. Again, the logic-based programs can only deal with simplified problems or problems that include restrictive assumptions, which make the task easier. No logic-based program can deal with all possible kinds of documents.

B1.2.2 Document Pre-processing

Once a general understanding of the area of document analysis had been obtained, the research was made more specific and aimed at pre-processing. Before a document image can be subjected to optical character recognition, it has to be prepared by using pre-processing. This can be done in the spatial domain or the frequency domain using transforms. The spatial domain methods revolve around using pixel values for neighborhood processing (as in filters) or for the image as a whole (as in histogram equalization). In the frequency domain, the same operations can be applied to the image as in the spatial domain, but with better results. This is because the spatial domain methods are approximations of the frequency domain methods in some cases. For example, to transform the image, Discrete Fourier Transform (DFT) is often used. The technique used for obtaining the DFT is the Fast Fourier Transform (FFT). After processing is complete, an Inverse DFT using FFT technique is used to obtain pixel values again.

Pre-processing comprises (general) topics such as noise removal, *document segmentation*, *document thresholding*, line segmentation (into characters), estimation of page layout, removal of underlines etc. This prepares the document for the processing stage where OCR can be carried out. OCR requires ideal conditions for the best results; thus the pre-processing should identify the text areas, remove any degrading artifacts and align the areas correctly.

B1.2.3 Research Focus and Structure of Report

The research in this case was centered on the methods of *document segmentation and thresholding* without any level of word-recognition or format-resolution involved. As pointed out above, these fall under the broader umbrella of document pre-processing. Though there are several divisions within document pre-processing, it is hard to research one without the other as the techniques are inextricably linked and it is rare to come across a piece of work which manages to devote itself to one technique exclusively without using others. This is because all the methods work towards the same goal in the end and most samples available would require more than one type of pre-processing. So, for example, before segmentation can be done on a certain image, it might be necessary to correct the skew, so skew correction is considered a part of segmentation.

The literature review in Chapter B2 provides a description of the techniques used in the two fields. That is followed, in Chapter B3, by an overview/assessment of the methodologies. Chapter B4 comprises a study on skew correction and its effect on speed of text segmentation.

Chapter B2: Review of Literature on Document Segmentation and Thresholding

This section presents a review of techniques used in the fields of document segmentation and thresholding.

B2.1 Document Segmentation Techniques

Document segmentation is defined as the separation of a document image into its constituent sections or components. The components referred to are the units of content, namely text and graphics, which are separated from the background and from each other. A segmented document has its text and graphic regions clearly marked and differentiated them from each other. Also, within the text areas themselves, the text is further segmented into main body, headings, marginal notes etc. while the graphics are divided into images, tables etc. Document segmentation lies within the overall field of document pre-processing which in turn is a sub-set of document analysis and understanding.

Tang et al. (1996) and Lee (2001) contain an excellent overview of segmentation techniques and methodologies. Some common examples of such techniques are connected component analysis, Hough transform, and texture analysis.

B2.1.1 Texture Analysis

An old texture-based segmentation method is the multichannel Gabor filtering technique. Different Gabor filters are tuned to capture desired local spatial frequency and orientation characteristics of a textured region. Jain and Bhattacharjee (1992) successfully used multichannel Gabor filters to extract text and halftone regions from document images. Though this technique is effective at texture analysis, it does not guarantee the best possible result for a given texture segmentation or classification task. A better approach is to design a small number of simpler, special purpose filters

Appendix B: Document Image Analysis

optimized for the given problem. The filters would perform the same texture extraction task as the Gabor filters. The difference is that instead of being general purpose, they are tuned for the particular problem and thus also require less computational resources.

This approach was used by Jain and Zhong (1995). They obtained a set of masks to discriminate between halftone, background, text and line-drawing regions by training a neural network on sample data from these three classes. Once the training is done, classification is carried out. During classification, each pixel in the input image, based on its neighborhood information, is classified by the network into one of the classes. In Jain and Zhong (1995), the page layout segmentation problem is basically posed as a texture segmentation problem. An image region is defined as textured if it contains some repeating gray level patterns. A neural network was employed to train the set of masks. Texture features are obtained by convolving the trained masks with the input image. These features are then used in the classification. The text in most languages consists of individual characters, which are formed of curves and single points. Thus, it represents a relatively consistent texture, which is quite different from that of the halftone and the background regions. Therefore, this texture-based page segmentation algorithm was designed to be insensitive to the language of the document. The method was directly applied to the input gray level document image, which avoided the loss of information during binarization.

The authors noted that this method shows potential in the problem of language separation, where texts of different languages have to be identified and separated. This is because of possible differences in the construction of fundamental elements of the languages and/or the placement rules of the elements. If the textural difference is consistently large enough to be captured by the texture discrimination masks, the languages can be separated.

B2.1.2 Mathematical Morphology

Mathematical morphology uses geometric operations to alter the structure (morphology) of the document. This basically translates into finding units that are relevant to each other in terms of content. The geometric operations used are opening and closing with horizontal and/or vertical elements. The optimal result is a document segmented into merged components that represent units of content. Classification of segments is done based on its height, aspect ratio etc. This is not as sophisticated a technique as texture analysis and thus, not as successful. Its success is limited to simpler documents where relevant/similar elements of a document are in physical proximity. Normally, the input image is down-sampled by a certain factor and then, the morphing is employed. Morphology was used in Fisher (1990), Lebourgeois (1992) and Pavlidis (1992).

B2.1.3 Voronoi Diagrams

Voronoi diagrams have been used in the field of image processing for the representation of input images. Burge and Monagan (1995) applied Voronoi diagrams to extract words and symbols from map images. A point Voronoi diagram is constructed from a document image by using the centroids of the connected components. This approach was used to estimate text-line orientation using the minimum spanning tree. However, the point Voronoi diagram is unsuitable for the segmentation of document images, since the approximation of each connected component as a single point is too imprecise to represent the structure of pages. Pages, generally, contain various sizes and shapes of connected components. In the case that a page has non-overlapping layout, every document component is represented as a set of Voronoi regions which are adjacent with one another. Thus, Voronoi edges represent the structure of page layout as potential boundaries of document components. The process of page segmentation is, therefore, the selection of the Voronoi edges which represent the boundaries of document components (Kise et al. (1997)). To delete superfluous edges, two characteristic features were used.

They are minimum distance and the area ratio of connected components. These criteria required the use of threshold values for determining which edge should be deleted. These thresholds were either fixed or depended upon the input image and were thus determined adaptively from the input image.

B2.1.4 Hough Transform

The Hough transform maps points of Cartesian space (x, y) into sinusoidal curves in a (R, θ) space via the transformation:

$$R = x \cos\theta + y \sin\theta.$$

Each time a sinusoid curve intersects another at particular values of R and θ , the likelihood that a line corresponding to these R - θ coordinate values is present in the original image also increases. An accumulator array, consisting of A rows and B columns is used to count the number of intersections at various R and θ values. Those cells in the accumulator array with the highest number of counts will correspond to lines in the original image. As text lines are actually thick lines of sparse density, Hough Transform can be used to detect them and their orientation.

An important application of Hough Transform is skew detection. A typical method can be found in Hinds et al. (1990) where skews were correctly identified for thirteen test images of five different types of documents. Hough transform can also be used to identify text-blocks so that the document contents can be differentiated from each other.

B2.1.5 Projection Profile

An important technique used for element extraction, character segmentation and skew normalization is Projection Profile. It refers to the mapping of a 2-d region into a waveform whose values are the sums of the values of the image points along some specified directions. A projection profile is obtained by determining the number of black

pixels that fall onto a projection axis. Three-directional projection profiles (horizontal, vertical and diagonal) are used. All objects in a document are contained in rectangular blocks and blanks are present between these rectangles. Thus, the projection profile is a waveform whose deep valleys correspond to the blank areas of the documents. A deep valley with a width greater than an established threshold value can be cut as the position corresponding to the edge of an object or a block. As a document generally consists of several blocks, the process of projection should be done recursively until all the blocks have been located.

B2.1.6 Connected Component Analysis/Labeling (CCA)

A connected component is a set of connected pixels such that an 8-connected path exists between any two pixels. Different contents of a document tend to have connected components with different properties. Generally, graphics consist of large connected components while text consists of connected components with regular and relatively smaller size. By analyzing these components, graphics and text in a document can be identified, separated and grouped together into different blocks. The size and location of a component is represented by four values, the coordinates of the top-left corner and the height and width of the component since a component is represented by a bounding rectangle. The analysis of a document thus boils down to the merging of these rectangles based on their relative sizes and positions and proximity to each other. An application of CCA is envelope processing where analysis of the components is used to locate address blocks on envelopes (Yeh et al. (1987)).

B2.2 Document Thresholding Techniques

Thresholding is defined as the conversion of a grayscale image into a binary image. A grayscale image has 256 intensity levels while a binary image has two. Thus, the gray

levels are mapped to black or white depending on the value of the threshold. The threshold value depends on the algorithm used.

Thresholding is a separate area of research and requires specialized techniques. Venkateswarlu and Boyle (1995) contains an excellent summary of thresholding algorithms.

B2.2.1 Global Methods

Otsu's method (Otsu (1979)) finds a gray level threshold for which between-class variance is maximal. This is an old method and rather simple, but an important part of the literature on thresholding. Lloyd (1985) is a method that arrives at a suitable threshold using iterations. The initial threshold is taken to be the mean grey value. A mathematical formula is used to calculate the next value by employing the variance, the means and number of pixels in the two regions (separated by the present threshold value). The iterations end when two consecutive values of the threshold are equal. These are global thresholding methods as there is one threshold for all pixels irrespective of position and the neighborhood pixels.

Conventional global thresholding methods, which utilize the image grey level histogram, face the difficulty that not all features of interest form prominent peaks. Hon-Son Don (1995) solved this problem using *noise attribute* features extracted from the image. These features are based on a simple noise model and are independent of the strength of the signals (objects) in the image. The NAT method makes two assumptions: that the objects of different classes occupy a separable range in the histogram and that the noise attributes in each class are statistically stationary.

The first of the above assumptions was also made in Liu and Srihari (1997). Histogram analysis, in which candidate thresholds are chosen, is followed by texture feature

extraction. The features are extracted from the run-length histogram for each candidate threshold. The optimal threshold is then chosen based on which results in the most desirable document texture features.

B2.2.2 Local Methods

Bernsen (1986) uses a threshold value which is calculated for each pixel by using a window as reference around it. The value is the mean of the maximum and minimum values within the window. The difference between the center pixel and the threshold is equivalent to a second order derivative. If this value is greater than zero, the center pixel is foreground otherwise background. This is thus, a local thresholding method.

The non-linear dynamic algorithm (NDA) used in White and Rohrer (1983) is a thresholding method that is used for segmentation. This method is conceptually equivalent to comparing the pixel value to its neighborhood. The size of the area chosen depends upon the character size. If the pixel is significantly darker than the average of the pixels in the chosen neighborhood, it is assigned to character/graphics; otherwise it is classified as background. Another algorithm is the IFT (integrated function technique) method. It removes as much of the non-essential background information as is feasible without affecting the useful information. A gradient type operator is used to identify pixels in or very close to areas where sharp changes exist in the gray level image. The changes in intensity represent text areas or background areas with high contrast. The areas with sharp edges are then checked for evidence labeling them as either text or background. If the region contains text, it will have a certain sequence for its first and second order derivative values. Text areas are differentiated from high-contrast background regions by checking the length of the sequence. The length of the sequence is a measure of the stroke-width. Characters would have a characteristic stroke-width different from that of random background strokes.

Appendix B: Document Image Analysis

LLT (logical level technique) method used in Kamel and Zhao (1993) is conceptually equivalent to comparing the grey level of a pixel (or its smoothed value if the original image is noisy) with some local averages in the neighborhood of eight neighboring pixels. If w is the stroke-width of the characters, then the neighborhood is chosen as a square area measuring $(2w+1)$ pixels along one side. The eight neighboring pixels are defined as being situated at a distance of w pixels from the original pixel being considered and at angles of zero degrees and multiples of 90 degrees. This comparison result is similar to derivatives. Local averages being less sensitive to noise, these ‘derivatives’ are also less sensitive to noise. If the grey level of the pixel is at least T levels below four local averages, it is considered a character/graphics pixel. T is a pre-specified parameter.

Chapter B3: Overview and Conclusions for Document Pre-processing

B3.1 Document Segmentation

B3.1.1 Methodologies

The process of document segmentation has two possible approaches or methodologies.

The differences are based on the direction of the logical flow of the analysis, either top-down or bottom-up. Both approaches have their strengths and weaknesses.

The initial methods that were used for page segmentation belong to the bottom-up approach. These methods begin by extracting local information such as intensity values and then using that information to progressively determine the more complex building blocks of the geometric structure such as words, text lines and paragraphs. A good method applicable to the bottom-up approach is the connected component analysis method, where connected components are extracted from the image through a labeling technique. However, the identification, analysis, and grouping of the components is a time-consuming process as a typical document contains a large number of connected components. Another problem is that if the initial groupings are wrong, the segmentation can be completely incorrect.

Top-down approaches look for global information in the document and then segment it into smaller units. The advantage of this approach is that the time required is lower than that of the bottom-up approach. However, if the document's geometric layout is overly complex in terms of irregular shapes, various font sizes etc., the performance of the top-down approach is poor.

An alternative is to consider a uniform region such as text or graphic as a textured region. Then, page segmentation is implemented by finding textured regions in gray

scale images (based on the technique of *Texture Analysis*). One major problem associated with such texture-based approaches such as in Jain and Zhong (1995) and Jain and Bhattacharjee (1992) is that the time complexity is high. If a small mask is chosen, it is difficult to detect large-scale textures but if a large mask is chosen, the computational cost increases dramatically.

B3.1.2 Nature of the Problem

It is necessary to appreciate the nature and complexity of the problem at hand. The text areas contained within the document and the images (if any) should be identified, separated from the background and each other and stored in compressed form. The most basic image would be that of an all-text page with regular columns and clear headings with the text being machine-printed without skew. In such a scenario, the segmentation is simpler due to lack of graphics and the analysis reduces to simple OCR. However, such documents are rare and not the focus of segmentation algorithms. In the more complex cases, the problems encountered by such an algorithm are numerous.

Firstly, the original document may be dirty and smudged. This is commonly seen in very old documents such as those of historical importance. The dirt causes the text to be unclear and the information cannot be ascertained correctly. It is difficult even for humans to determine the contents of such a document. In some historical documents, there is the problem of seepage, that is, ink from the reverse side of the paper seeps into the front due to aging and the porous nature of the paper. Thus, a shadow of the reverse side of the page is imprinted on the front along with the actual text. While humans can differentiate easily between the two, a machine cannot, as the algorithm does not recognize the difference between normal and laterally inverted text. Even a method that can recognize English script and differentiate it from, for example, Devanagari script would not be able to spot the problem as the laterally inverted text still displays the

texture characteristics of the normal text (a method using neural networks exists to differentiate between scripts. It employs previous knowledge of the texture of a script. This is explained in Chapter B2 under *Texture Analysis*).

The above are the problems caused by the physical flaws in the original document itself. Another common problem is that of skew. All or part of the contents may be skewed with respect to the outline of the page. The skew has to be corrected before any later step can be taken. Otherwise, the results will be adversely affected. A final problem with the document may be that of confusion between foreground and background. Sometimes, a document may have a non-uniform, non-white background. In this case, thresholding is required to separate the two *before* anything else can be done. The foreground thus obtained is then treated as input for segmentation of text and graphics.

B3.2 Thresholding

B3.2.1 Methodologies

A number of techniques have been invented for effective thresholding. There are two kinds of thresholding methods, global and local, and some of the individual techniques of each were detailed in Chapter B2. Global methods apply one threshold to the entire image while local thresholding methods apply different threshold values to different regions of the image. The value is determined by the intensity values of pixels within the neighborhood of the pixel to which the thresholding is being applied. The neighborhood specified depends upon the particular technique being used.

B3.2.2 Nature of the Problem

The challenge when carrying out thresholding is that pixels should not be wrongly assigned, that is, foreground pixels should not be classified as background as that would cause loss of information. Also, background pixels should not be classified as foreground

as that would create objects that do not belong in the foreground and thus, should not be part of any further processing carried out on the thresholded foreground.

Global thresholding methods have been found to be lacking for complicated document images and thus, most research focuses on local methods as they perform better. The reason for this is intuitively clear. If an image has a resolution of a few hundred pixels along each dimension, it is to be expected that (in the general case) there would be wide variations in the texture of the image from one end to another. Texture here refers to unique pixel patterns. Thus, an algorithm that treats the entire image as one integrated histogram and ignores local variations would be not optimal, in fact, far from it. Ideally, a thresholding algorithm should be able to apply some level of understanding to each local portion that can then be thresholded individually. There is no relationship between one local threshold and another as they are, by definition, dependent only upon the pixels within their restricted area.

The thresholding algorithms reviewed in Chapter B2 deal with document images and not general grey scale images. This process is known as document image binarization, that is, the conversion of a document image into binary form. As mentioned above, thresholding is actually a useful step to perform before document segmentation is carried out. It simplifies the problem of segmentation, assuming that the thresholding is accurate; otherwise, it destroys information and the content of the image.

B3.3 Topic for Investigation

When segmenting a document image containing handwritten text, the aim is to separate the text into constituent lines and subsequently, the lines into their constituent words. The solution to this problem requires identification of the baseline of the lines within the text, detection of inter-line spaces and finally detection of inter-word gaps.

Appendix B: Document Image Analysis

One of the main obstacles to be overcome for accurate segmentation is the presence of line-skew. Handwritten text with several kinds of line-skew (multi-skew) is representative of real world documents. A study has been carried out to remove skew in document images and then compare the performance of a text segmentation algorithm with and without the skew correction. The algorithm created and the results of the study on skew correction are presented in Chapter B4.

Chapter B4: Study on Skew Correction

Skew is said to exist in a given document image if the text is oriented at an angle relative to the horizontal edge of the image (which represents the ideal baseline of the text).

Multi-skew exists if the skew angles of two or more skewed text lines are different.

B4.1 Implementation

B4.1.1 Hough Transform

The Hough Transform is a global method for finding straight lines hidden in large amounts of data. It is an important technique in image processing. The Hough Transform technique can transform a straight line on the normal x-y plane into a point on the Hough Transform space. This helps to solve the skewed (sloping) lines problem that is commonly encountered in document images. In the x - y space, a straight line can be expressed as:

$$Y_i = mX_i + c$$

(13)

where (X_i, Y_i) is a point on the line. The m is the slope and c is the y-axis intercept of the line. By using Hough Transform, it will be converted into:

$$\rho = X_i \cos\theta + Y_i \sin\theta$$

(14)

ρ is the perpendicular distance of the line to the origin while θ is the angle between a normal to the line and the positive horizontal axis. A simple illustration of equations (13) and (14) is shown in Figure B4-1.

Figure B4-1: Illustration of equations (13) and (14).

The derivation of equation (14) is:

$$\begin{aligned}
 \rho &= y/\cos\theta + (x - y \tan\theta) \sin\theta \\
 &= x \sin\theta + y/\cos\theta - y \sin^2\theta/\cos\theta \\
 &= x \sin\theta + y (1 - \sin^2\theta)/\cos\theta \\
 &= x \sin\theta + y \cos\theta
 \end{aligned}$$

To understand the Hough Transform, it is important to know what the Hough space is. Each point (θ, ρ) in Hough space corresponds to a line at angle θ and distance ρ from the origin in the original data space as proved in equation (14). The value of a function in Hough space gives the point density along a line in the data space. On the other hand, a point in x - y space becomes a sinusoid curve in the ρ, θ space. The most important property is that *the curves (in Hough space) of all the points on the same line (in x - y*

Appendix B: Document Image Analysis

space) will intersect in the Hough space and the intersection point represents the line which they are lying on. (Figure B4-2)

Figure B4-2 Conversion from x-y space to Hough space.

For computation purposes, the parameter space is now quantized into $P \times \Theta$, where:

- P is the quantized number of ρ ,
- Θ is the quantized number of θ [°].
- An accumulator array whose size equals to $P \times \Theta$ is created.
- Each point (X_i, Y_i) will be mapped into a sampled, quantized sinusoid in Hough space and each accumulator cell along the curve will be incremented by 1.

After processing all these points, the accumulator cell with the highest value is the parameters of the line that best explains the points. An issue with this method is that it detects just one line. But there are normally several texts line on the document image, which means the line with most pixels will be detected. Thus, the number of processing pixels must be reduced and the remaining pixels should be able to represent the layout of the original document image. For the sake of simplicity, the center of gravity (c.o.g) of each connected component is chosen to represent the image and reduce overhead.

B4.1.2 Algorithm and Illustration

- Assign the minimal and maximal value to the ρ and θ .
- For ρ , its value should not exceed the square root of image diagonal length D ($(\text{height}^2 + \text{width}^2)^{1/2}$). Thus, the ρ could be valued between $-D$ and $+D$. For θ , theoretically, it is from -90° to $+90^\circ$, but for simplicity, it is assumed that the document will only skew from -30° to $+30^\circ$.
- Initialize the Accumulator Array by size $P \times \Theta$. The value of Θ was decided above. There is no fixed rule to decide the value of P and here it is set at 200.
- For each center of gravity point, the X and Y values are substituted into the formula $\rho = X \cos \theta + Y \sin \theta$, where θ varies from the -30° to $+30^\circ$, and ρ is calculated.
- For each ρ , it is rounded off to the nearest value. The corresponding accumulator cell decided by the values of θ and ρ is incremented.
- Find a set of local maximum values in the accumulator and then get the average θ value of them. The number of values taken here is equal to the number of text lines.
- Adjust the center of gravity of each component by the θ .
- The adjusting stage is for recalculating the height of the center of gravity. An assumption made here is that all the lines that skew extract the same angle θ . The new height is equal to the original height + the original width $\times \tan \theta$. To make the last step clear, Figure B4-3 illustrates the effect after the adjustment.

Appendix B: Document Image Analysis

Assume that point A and B are the c.o.g of the two components respectively and they belong to the same text line. The original distance between A and B was D.

After adjustment, the point A is moved to point A' and B to B'. Now the distance between A' and B', which are the adjusted points, is D'. Although the two points are not in the same horizontal line as desired, D' is still much smaller than the D. This means that after the skew adjustment, a smaller region (D x width) is needed in order to include A and B. Thus, undesirable components, which cause error and/or slow calculations are avoided. Therefore, the selection of components, for merging into words and then grouping into lines, becomes more accurate.

Figure B4-3: Adjustment of two points skewed with respect to each other.

Figure B4-4 is the result of one image document which has been adjusted. The skewed angle is very large for the text lines in the original: the crosses indicate the centers of gravity before adjustment and the circles are the modified ones (the contents of components are not displayed). It shows that most of the skewed lines have been corrected to a horizontal line.

The goal behind using the Hough Transform was to enable and simplify grouping of connected components into words and then into lines. Without the Hough transform, the

Appendix B: Document Image Analysis

connected components belonging to a particular line have to be selected from a large vertical range, specifically $2.5 \times H_{avg}$, where H_{avg} is the average height of a component. $2.5 \times H_{avg}$ approximately represents the space from the top of one text line to the top of the next one, thus ensuring that all possible components are chosen. However, when the document is skewed, the typical problem is that components from one end of the first line lie at about the same vertical coordinate as the components from the opposite end of the next line. This leads to inaccurate grouping of words when forming the text lines. *The Hough Transform reduces or even completely removes skew, thus allowing more accurate grouping which leads to better segmentation results.* This is the basis of the results presented later in this chapter.

As the components have been adjusted, the region for selecting the candidate components is no longer $2.5 \times H_{avg}$. And moreover, in order to avoid some components lying higher than the rest, the vertical region is changed to be more dynamic. The system first chooses all components within the $2/3 \times H_{avg}$ region and then averages the height (H_A) of center of gravity of these components. After that, the system uses $H_A + 1/2 \times H_{avg}$ as the lowest bound to choose the candidate components.

Figure B4-4: Result of adjustment of one document image. The COGs shift from the crosses to the circles.

B4.2 Results

To demonstrate the improvement in segmentation, the algorithm was timed while carrying out segmentation with *and* without the use of the Hough Transform. Table B4-1 shows a table with the time measurement for 20 successful images. The first row contains the time required for segmentation with use of skew adjustment (including the time for performing skew adjustment). The second row contains the time for segmentation without use of the skew adjustment.

The time for the skew adjusting is normally around 20~80 milliseconds. The program was tested on a P3 Computer. Please note that the figures given here may vary when used with a different computer, However, the time required for segmenting after the skew adjustment (Hough Transform) will definitely still be lesser than the one without performing skew adjustment.

Appendix B: Document Image Analysis

Image No	a1	a2	a3	a4	a5
With Skew Adjustment	747	830	200	922	456
No Skew Adjustment	1620	1800	390	2958	1001

Image No	a6	a7	a8	a9	a10
With Skew Adjustment	198	225	120	425	2082
No Skew Adjustment	291	861	430	942	20489

Image No	a11	a12	a13	a14	a15
With Skew Adjustment	658	355	190	235	987
No Skew Adjustment	1054	765	380	126	1984

Image No	a16	a17	a18	a19	a20
With Skew Adjustment	181	164	323	905	845
No Skew Adjustment	328	453	772	376	1491

Table B4-1: Table comparing execution times with and without Hough transform.