

**MACHINE VISION  
FOR ENDOSCOPE CONTROL AND NAVIGATION**

by

Gul Nawaz Khan, B.Sc. (Eng), M.Sc.

April, 1989

A thesis submitted for the degree of Doctor of Philosophy of the  
University of London and for the Diploma of Membership of the  
Imperial College.

Department of Computing  
Imperial College of Science, Technology  
and Medicine  
University of London  
London SW7 2BZ

## ABSTRACT

This thesis presents two new methods of image analysis, which have been applied to the problem of guiding an endoscope inside the human colon. In the human colon, there are two types of useful information: curved contours and darker regions.

The first method concerns contour extraction. Traditionally contours are extracted by detecting edge points, thresholding, and then linking them sequentially. While retaining the traditional approach, this new method departs from these techniques in a number of ways. A simple edge detector is used to prepare an edge map. In contrast to normal detection methods, the weak edges are not removed because they may be a significant part of the contours. Instead an attempt is made to group them into short line segments by filtering on the basis of perceptual criteria. The grouping process is local, highly data directed, and it is implemented in parallel. Next the line segments are linked, perceptually and hierarchically, into contours. The method employs perceptual grouping in a unified way, relying on the bottom-up organisation of edge data. It has been tested experimentally on a number of endoscopic images and the results are very encouraging.

The second method is based on region extraction and introduces the use of variance in a pyramid structure for detecting coherent regions. The method has been extended for general purpose segmentation and tested on a variety of medical images including endoscopic colon images. The novel feature of both methods is their parallel implementation on a pyramid based computer architecture.

The regions and contours are represented in a new world and search space representation (QL-Tree) for navigational purposes. The QL-Tree representation consists of a series of planes represented by quadtrees and it can be incrementally constructed by integrating information from a sequence of images. Ease of updating, access, and efficient search make this representation ideal for navigation.

This research is dedicated to my family:  
my mother, son, and wife.

## PREFACE

Machine vision is a field of research which has many goals and objectives. A primary goal has been to build a computer vision system that can provide information to general purpose robots about their surroundings in the same way as we receive information from our own visual system. The research described in this thesis takes an important step towards a similar but restricted goal of providing sensing capabilities for an automatic endoscope. The machine vision techniques that have been developed in the course of this research detect navigational landmarks for guiding the endoscope inside human colon. These landmarks are detected from monochrome colon images in the form of occluding contours formed by the inner colon muscles and darker regions which correspond to the deeper and obstacle free areas in the colon.

A second major goal of machine vision research is to provide a computational understanding of human vision. This research has many implications in understanding the human vision particularly in the area of perceptual organisation and the grouping phenomena of human vision studied in depth by the Gestalt psychologists. An attempt has also been made to relate this computational work to the relevant areas in the psychology of vision and neurophysiology. In Chapter 2, there is a review of the early visual data organisation in animal vision from the point of view of neurophysiology. The psychology of human vision is also explored, particularly in the area of perceptual organisation and grouping.

One of the most important conclusions arising from this research is that partial image segmentation, in terms of contour and region extraction, can commonly be achieved without having a knowledge of the scene and there is no need to assume a certain type or level of noise in the images. In Chapter 3 and 4, one of the

segmentation methods developed during three years of this research is described. This method partitions images by detecting the boundary contours. Perceptual grouping plays a vital role by providing direct relations among two-dimensional features (edges, line segments etc.) of an image. These grouping processes are employed hierarchically to filter out image features due to noise and for extracting relevant contour structure. Their performance is compared with the existing techniques quantitatively and qualitatively. In Chapter 5, a second method to detect darker regions is presented. The dark region extraction is very closely related to region based image partitioning and the method has been shown to work well for general purpose image partitioning. It employs a variance-average pyramid representation. The grouping based on proximity and similarity in grey level is utilised for image partitioning. The hierarchical and parallel nature of these two visual processing techniques makes them applicable to real-time image analysis.

In Chapter 6, the endoscope navigation system is described in an effort to integrate our machine vision research with the navigation of endoscope. A new world and search space representation is proposed which can be constructed from the information provided by the vision system. In Chapter 7, some final conclusions are made and the overall research plan for endoscope navigation is presented.

The thesis can be easily divided into two main sections: contour extraction and region based image partitioning. After reading the first two chapters, the reader can approach these sections individually. For example, Chapter 5 can be read before Chapters 3 and 4.

## ACKNOWLEDGEMENTS

I wish to thank Dr. Duncan F. Gillies for supervising this research. His advice and encouragement kept me on the right track during the three years of this research.

I am indebted to Dr. Peter Burger, Mr. B. Kim, and other members of the Computer Vision and Graphics Group for providing an active and cheerful environment, in which this research was conducted.

I am also thankful to Dr. Christopher B. Williams of Endoscopy Section, St. Mark's Hospital London, for his enthusiasm for this project. He has also provided us different types of colonoscopy video tapes.

My thanks goes to my parent department, Ministry of Education, and the Committee of Vice-Chancellors and Principals of the Universities in the United Kingdom for their financial support.

The generous grant from the Olympus Optical Co. Tokyo for the endoscope project is also highly appreciated.

## TABLE OF CONTENTS

	Page
ABSTRACT . . . . .	1
PREFACE . . . . .	3
ACKNOWLEDGEMENTS . . . . .	5
TABLE OF CONTENTS . . . . .	6
LIST OF FIGURES . . . . .	9
LIST OF TABLES . . . . .	12
<b>1. INTRODUCTION</b>	
1.1 Machine Vision and Endoscope Navigation . . . . .	13
1.2 The Endoscope . . . . .	14
1.3 Endoscope Image Analysis . . . . .	16
1.4 Perceptual Organisation . . . . .	19
1.5 Image Segmentation and Perceptual Grouping . . . . .	23
1.6 Research Objectives and Motivation . . . . .	26
<b>2. VISUAL DATA ORGANISATION AND GROUPING</b>	
2.1 Introduction . . . . .	29
2.2 Early Data Organisation in Biological Vision . . . . .	30
2.2.1 Low Level Feature Detection . . . . .	31
2.2.2 Visual Signals, Pathways and Organisation . . . . .	34
2.3 Pyramidal Architecture and Organic Vision . . . . .	39
2.4 Psychology of Vision . . . . .	43
2.4.1 Background . . . . .	43
2.4.2 Form Perception . . . . .	44
2.5 Classification of Grouping Processes . . . . .	48
2.5.1 Perceptual Grouping . . . . .	49
2.5.2 Hough Techniques . . . . .	50

	Page
<b>3. EXTRACTION OF CURVED LINE SEGMENTS</b>	
3.1 Introduction . . . . .	54
3.2 Motivation and Problem Definition . . . . .	55
3.3 Previous Work on Extracting Line Segments . . . . .	57
3.4 Perceptual Grouping of Edges into Line Segments . . . . .	62
3.5 Line Segment Extraction: Implementation Details . . . . .	67
3.5.1 Edge Point Detection . . . . .	67
3.5.2 The Grouping Process . . . . .	71
3.5.3 Selection of Thresholds . . . . .	75
3.6 Experimental Results and Conclusions . . . . .	77
3.6.1 Artificially Generated Images with Added Noise . . . . .	77
3.6.2 Endoscopic Colon Images . . . . .	87
3.6.3 Conclusions . . . . .	98
 <b>4. CONTOUR DETECTION FROM LINE SEGMENTS</b>	
4.1 Introduction . . . . .	101
4.2 Shape and Contours . . . . .	102
4.2.1 Contour Types . . . . .	103
4.2.2 Three-Dimensional Structure from Image Contours . . . . .	105
4.3 Different Approaches to Contour Extraction . . . . .	108
4.3.1 Grouping Edge Points into Contours . . . . .	108
4.3.2 Aggregation of Straight Line Segments . . . . .	109
4.3.3 Multi-Resolution Based Contour Extraction . . . . .	111
4.4 An Hierarchical Method to Detect Contours . . . . .	113
4.4.1 From Line Segments to Contours . . . . .	113
4.4.2 The Algorithm . . . . .	116
4.4.3 Implementation Details of the Algorithm . . . . .	120
4.5 Experimental Results and Discussion . . . . .	121
 <b>5. ESTIMATING RELATIVE DEPTH BY REGION EXTRACTION</b>	
5.1 Introduction . . . . .	129
5.2 Depth from Intensity in Colon Images . . . . .	131
5.3 Region Extraction Techniques . . . . .	133
5.3.1 Region Merging . . . . .	133
5.3.2 Region Splitting . . . . .	135
5.3.3 Pyramid Based Techniques . . . . .	136

	Page
5.4 The Pyramid Structure for Region Extraction . . . . .	137
5.5 Detection of Dark Regions . . . . .	140
5.5.1 The Algorithm . . . . .	140
5.5.2 Implementation Details and Experimental Results	146
5.6 Extension to Image Partitioning . . . . .	152
5.6.1 Image Partitioning into Uniform Regions . . . . .	152
5.6.2 Results and Comments . . . . .	154
5.7 Parallel-Serial Region Extraction: Discussion . . . . .	158
5.8 Concluding Remarks . . . . .	161
6. NAVIGATION OF THE ENDOSCOPE	
6.1 Introduction . . . . .	163
6.2 Navigation Techniques: A Review . . . . .	164
6.2.1 Machine Perception for Autonomous Vehicles . . . . .	165
6.2.2 Find-Path Problem . . . . .	167
6.3 Navigation System for the Endoscope . . . . .	171
6.3.1 The Find-Path Problem and Guiding the Endoscope	171
6.3.2 An Hierarchical Navigation Control System . . . . .	172
6.3.3 Environment Representation for Navigation . . . . .	175
6.3.4 Colon Model for Endoscopy . . . . .	177
6.4 World and Search Space Representation . . . . .	181
6.5 Concluding Remarks . . . . .	185
7. CONCLUSIONS AND FUTURE WORK	
7.1 Introduction . . . . .	187
7.2 Contour Extraction . . . . .	189
7.3 Region Extraction . . . . .	191
7.4 Discussion . . . . .	192
7.5 Future Work . . . . .	194
REFERENCES . . . . .	197
BIBLIOGRAPHY . . . . .	213

## LIST OF FIGURES

	Page
Figure 1.1: A typical endoscope. . . . .	16
Figure 2.1: Schematic diagram of the path from retina to visual cortex. . . . .	34
Figure 2.2: Block diagram of the layers in monkey's LGN and visual cortex. . . . .	36
Figure 2.3: An elementary volumetric unit of the visual cortex.	38
Figure 2.4: Data representation in tabular format for the elementary unit of visual cortex. . . . .	40
Figure 2.5: The pyramid architecture. . . . .	43
Figure 2.6: The laws of organisation. . . . .	46
Figure 2.7: The $(\rho, \theta)$ parameterisation of the line in Hough transform. . . . .	51
Figure 3.1: Line parameterisation in terms of intersection points between the line and the bounding box. . . . .	60
Figure 3.2: Grouping edges into curved line segments. . . . .	64
Figure 3.3: Relation between edge orientation and line direction. . . . .	65
Figure 3.4: Stimuli showing theta-aggregation and combination of theta-aggregation, curvilinearity and orientation drift. . . . .	66
Figure 3.5: A $3 \times 3$ image window centred at $(x, y)$ . . . . .	70
Figure 3.6: An artificial image having $\pm 10\%$ added noise and the output of Sobel edge detector. . . . .	79
Figure 3.7: An artificial image having $\pm 22\%$ added noise and the output of Sobel edge detector. . . . .	79
Figure 3.8: Line segments extracted by employing connectivity grouping. . . . .	80
Figure 3.9: Line segments extracted by employing grouping based on similarity in edge pixel intensity. . . . .	81
Figure 3.10: Line segments extracted by employing grouping based on similarity in edge contrast. . . . .	82
Figure 3.11: Line segments extracted by employing grouping based on orientation drift and theta-aggregation. . . . .	83

	Page
Figure 3.12: Line segments extracted by employing O'Gorman and Clowes method. . . . .	84
Figure 3.13: Line segments extracted by employing perceptual grouping criteria. . . . .	85
Figure 3.14: Line segments extracted without applying perceptual grouping. . . . .	86
Figure 3.15: First colon image and the output of Sobel edge detector. . . . .	89
Figure 3.16: Second colon image and the output of Sobel edge detector. . . . .	89
Figure 3.17: Third colon image and the output of Sobel edge detector. . . . .	90
Figure 3.18: Line segments extracted using connectivity grouping for colon image of Figure 3.15. . . . .	90
Figure 3.19: Line segments extracted using connectivity grouping for colon image of Figure 3.16. . . . .	91
Figure 3.20: Line segments extracted using connectivity grouping for colon image of Figure 3.17. . . . .	91
Figure 3.21: Output line segments for the colon image of Figure 3.15 on the basis of $8 \times 8$ window. . . . .	92
Figure 3.22: Output line segments for the colon image of Figure 3.15 on the basis of $4 \times 4$ window. . . . .	93
Figure 3.23: Output line segments for the colon image of Figure 3.16 on the basis of $8 \times 8$ window. . . . .	94
Figure 3.24: Output line segments for the colon image of Figure 3.16 on the basis of $4 \times 4$ window. . . . .	95
Figure 3.25: Output line segments for the colon image of Figure 3.17 on the basis of $8 \times 8$ window. . . . .	96
Figure 3.26: Output line segments for the colon image of Figure 3.17 on the basis of $4 \times 4$ window. . . . .	97
Figure 4.1: Shape from Contour: Constraints. . . . .	106
Figure 4.2: A $4 \times 4$ (50%) overlapped pyramid scheme. . . . .	114
Figure 4.3: Typical three-neighbours used to group line segments. . . . .	117
Figure 4.4: Line segments extracted for the first colon image. . . . .	122

	Page
Figure 4.5: Line segments extracted for the second colon image. . . . .	122
Figure 4.6: Line segments extracted for the third colon image.	123
Figure 4.7: Line segments extracted using connectivity grouping for the artificial test image. . . . .	123
Figure 4.8: Groups of line segments formed. . . . .	124
Figure 4.9: The output contours overlaid on the images. . . . .	126
Figure 5.1: The farthest part of the colon receives the smaller amount of incident light. . . . .	132
Figure 5.2: The endoscope approaching a bend. . . . .	134
Figure 5.3: The quadtree based pyramid structure. . . . .	139
Figure 5.4: A part of an intensity variance-average pyramid.	145
Figure 5.5: Dark region extraction in the first colon image.	148
Figure 5.6: Dark region extraction in the second colon image.	149
Figure 5.7: Dark region extraction in the third colon image.	150
Figure 5.8: Dark region extraction in the fourth colon image.	151
Figure 5.9: Partitions of the first heart ventricular image. . . . .	155
Figure 5.10: Partitions of the second heart ventricular image.	156
Figure 5.11: Partitions of the computer generated image. . . . .	157
Figure 5.12: Partitions of the artificial image with added noise.	159
Figure 6.1: A simple model of the colon. . . . .	178
Figure 6.2: A homogeneous generalised cylinder. . . . .	179
Figure 6.3: Moving co-ordinate system from point O to Q in space. . . . .	183
Figure 6.4: A section of Quad-List tree, (QL-tree) environment and search space representation. . . . .	185

## LIST OF TABLES

	Page
Table 3.1: Output line segments for the artificial image with varying degree of noise and for different grouping processes. . . . .	87
Table 3.2: Output line segments for colon images using different perceptual and other grouping processes. . . . .	99
Table 4.1: Image data reduction at the different levels of visual representation. . . . .	127

# CHAPTER 1

## INTRODUCTION

### 1.1 Machine Vision and Endoscope Navigation

Computer vision deals with enabling the computer to understand the environment from visual information. How the information is processed and which intermediate representation is used to achieve ultimate understanding, are very significant factors in the overall structure of a vision system. The key ideas behind the development of a high performance computer vision system are its competence and structure. In contrast to many computing tasks, the performance of a vision system on an unseen set of images cannot be guaranteed. The competence with which it will deal with the new information will depend in part on the representations it uses to describe the world. The structure of a vision system is generally taken to be a sequence of levels of representation. One of the well known structures for computer vision was proposed by Marr [1976, 1982] in the form of a raw primal sketch for early visual processing, and a  $2^{1/2}$ -D sketch at the intermediate level of visual processing.

There is a lot of controversy over, whether the processing in a vision system should be data-driven or goal-driven. Normally early visual processing is data-driven while higher levels are controlled by goals and expectations. Intermediate levels are often a combination of goal-driven (top-down) and data-driven (bottom-up) operations, both to compensate errors and to avoid computational overload. Intermediate processes are divided between low and high level processing and they have received little attention from computer vision researchers until recently. Due to this lack of research, plenty of the work done on early

processing is independent and difficult to integrate with the higher levels of the system.

Most of the vision systems can be categorised on the basis of their use which may be assembly (manufacturing), inspection, navigation, or recognition. The computer vision techniques described here have been developed as part of an autonomous guidance system for an endoscope which will enable the instrument to navigate inside the human colon. Navigation of the endoscope inside the human colon is a complex task. The colon is analogous to an unlit tunnel, closed at the far end, and with many bends, twists and pockets. The endoscope is like an articulated chain being pushed at the rear end. When it bears on the colon wall it will in some cases distort the shape of the colon, and produce paradoxical behaviour at the tip. The cross section of the colon is not uniform and can be completely collapsed in certain places, making the centre line difficult to see. The endoscope navigation is in no way similar to that of mobile robots or autonomous vehicles.

This is the first attempt at applying computer vision to the automation of endoscopy. The work is mainly directed towards colonoscopy, but the techniques are general enough to be used in most applications of endoscopy. Although the majority of endoscopes provide coloured images, all of the work presented in this thesis is based on grey level image analysis. Colour may well prove to be a useful property for guidance, however, it can increase the computing time required to process the image by a large factor. There is, as we shall see, plenty of useful information available in the grey scale images. As a first stage of the research, a detailed study on the application of computer vision in endoscopy has been carried out [Khan and Gillies 1987].

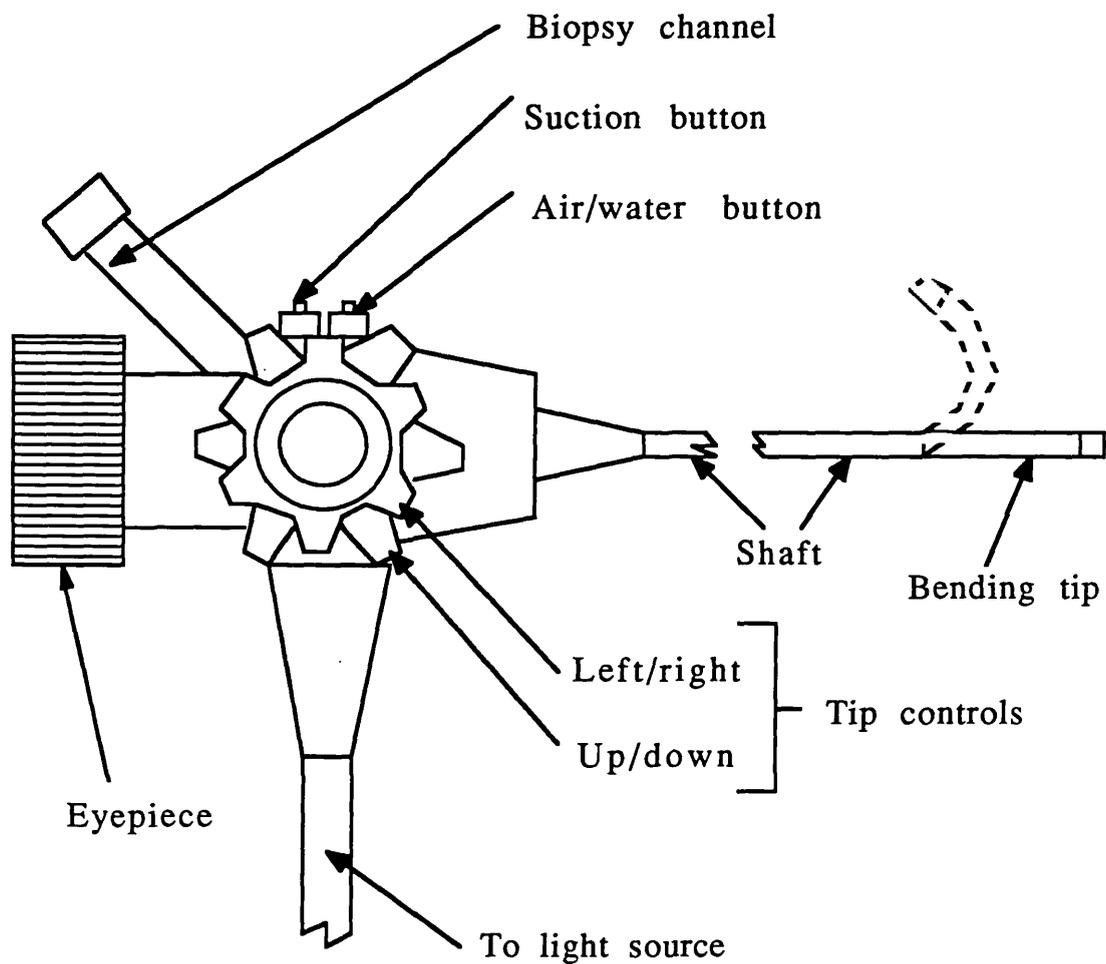
## **1.2 The Endoscope**

The endoscope is a medical instrument used for observing the inner surfaces of the human body. It is typically used for

diagnosing different colon and upper gastrointestinal diseases including cancer. More specialised instruments are also used for observing and diagnosing conditions in different divisions of the bronchus and biliary ducts. Fibrescopes are the most commonly used instruments but next generation of endoscopes, based on the utilisation of a CCD video chip at the tip, have recently been introduced.

A fibrescope consists of a head with eye piece and control, and a flexible shaft with a manoeuvrable tip (see Figure 1.1). The heart of the scope accommodates an optical system for viewing, control wires for tip movement, and two or more operating channels allowing passage of flexible instruments such as biopsy forceps. The head of the scope is connected to a unit containing a cold light source and air and water supplies. The video endoscopes which have recently been developed by most of the manufacturers are based on CCD video cameras. These cameras place an image sensor at the focal plane of the image formation lens in the front portion of the camera, and the detected image data is transmitted over wires to the output portion of the system. In this way the endoscope can be controlled by looking at the images on a monitor, rather than looking through the eye piece continuously during the whole diagnostic process. The tip of the instrument is controllable by control wheels providing up/down and left/right tip movement in some endoscopes or with a joy stick in others, while the shaft transmits the rotary movement to the tip.

Using a conventional endoscope for colonoscopy, the consultant inserts the endoscope by estimating the position of the colon centre line, called the lumen, from the shape of the colon muscular curves and from the apparent deepest region of the image, which he sees either through the eye piece or on a TV monitor. He then steers the tip appropriately. This task requires simultaneous movement of two control wheels in addition to push pull and rotational operations on the endoscope itself. Consequently a high degree of skill is required in utilising the instrument.



*Figure 1.1: A typical endoscope.*

### 1.3 Endoscope Image Analysis

In the current generation of endoscopes, there is only a single camera, and thus there is no direct measurement of depth in the image. Therefore depth must be estimated from the two-dimensional information. In the future, depth measurement may be incorporated into the endoscope, either by means of stereo

vision or by the introduction of a dedicated depth sensor on the endoscope tip. These possibilities are currently being investigated, and may or may not prove feasible. Since, there is only a remote chance of utilising stereoscopy or direct depth measurement for obtaining three-dimensional shape, it has been necessary to investigate other techniques.

Traditional methods based on the use of a single camera, for example shape from shading, are still insufficiently well developed to be used for noisy colon images. Moreover, they are far too expensive in processing time to be useful in a real-time navigation application and they also require special illumination arrangements and other shape constraints which are not feasible at the current level of development of endoscopes. Shape from camera motion is another possibility which at first sight appears to be worth investigating. However, for the time being, the planned automatic endoscope insertion will only control the direction of the tip. The push pull and rotational operations will still be carried out by the medical practitioner and the navigation system will therefore have no control on the forward, backward, or rotational movements of the instrument tube which are transmitted directly to the tip. It follows that the exact three-dimensional movement of the tip, where the camera is effectively sited, is not known. In addition to that, the non-rigid structure of colon presents further difficulties for depth estimation using the camera motion.

A higher confidence for image analysis can be achieved by integrating information from a sequence of images, providing the analysis is carried out sufficiently quick to ensure that no substantial changes occur in the image. This means that the processing of the images must be achieved in a fraction of a second, and ideally within the video frame rate.

The colon is illuminated by a point like, light source at the tip of the endoscope during diagnosing. Some endoscopes have multiple light sources but they are so close to each other that for all practical purposes a single point light source can be assumed.

Moreover the distance between the object and light source is very small. Under these illumination conditions the colon surfaces which are nearer to the light source are more brightly illuminated than the farther surfaces. Therefore in most of the endoscopic images the darkest area corresponds to the deepest and obstacle free region.

Secondly, the inner walls of a human colon contain circular rings of muscle. These rings are clearly distinguishable in the endoscope images since they form occluding edges. When the endoscope is directed along the centre line of a straight section of the colon, the muscle rings appear as closed curves in the image. More commonly, only part of the muscle curves are visible, the remainder being hidden either by bends, other muscles closer to the endoscope tip, or by irregularities in the colon wall. In the case of closed curves, the centre coincides with the correct direction of insertion, namely the centre line of the colon. In other cases an estimate of the insertion direction can be made from the average centre of curvature of the visible curves. The nearer rings of muscles are not difficult to see because of the pronounced occlusion caused by their contours.

It is clear from the above discussion that two types of image features, the darker regions and the image contours, require extraction from the endoscopic images. Although later analysis of contours is used for extracting some clues about the third dimension, these features are generally two-dimensional. The endoscopic colon images are processed to extract primitives in the form of edge segments and coherent regions, obtaining a map-like representation which is similar to the Marr's primal sketch. To extract the contours, the edge point representation is then refined by applying a perceptual grouping processes. Different principles from perceptual organisation are used first during the construction of the initial representation and then in the refining process at higher levels.

## 1.4 Perceptual Organisation

Human perception allows us to recover relevant regularities or structure from an image without the prior knowledge of the corresponding scene. The fact that this is possible suggests that there are identifiable properties of the scene which can be used by the visual system to connect primitive features. These are termed *perceptual criteria*, they are used to form *perceptual groupings*, and the structure they impose is called *perceptual organisation*.

Understanding and defining the description of the perceptual criteria which the human visual system uses to extract groups from images is one possible approach to the development of an effective and general purpose computer vision system. After identifying the set of perceptual primitives, the main task is to provide the computational techniques for their recovery from the raw image data. The issues concerning the use of perceptual organisation in machine vision are far from fully resolved and currently are being pursued actively by researchers in the computer vision field. Recent work on the application of perceptual organisation has concentrated on two types of approaches which differ in the way they depend on the nature of the perceptual primitives.

In the first of these approaches the primitives describe the scene specific features and attributes. Therefore the early vision processes are concerned with the recovery of environmental regularities (e.g. rigidity, axes of symmetry etc.) which are used by the cognitive processes at later stages. There have been a number of suggestions about the possible perceptual primitives since the late 1970s. Barrow and Tenenbaum [1978] proposed a computational framework for the recovery of point properties of the visible scene surfaces in terms of their orientation, reflectance, incident illumination, and range. The basis of their argument is in the fact that humans normally recover these characteristics regardless of their familiarity with the scene. These intrinsic image properties are more meaningful than the image intensity

and they also describe the image formation process. Kass and Witkin [1985] have gone further by using the decomposition of two-dimensional image intensities into more or less independent primitives in the form of *flow fields*. The flow fields are used to describe the oriented pattern in an image which are produced by propagation, accretion, or deformation. Zucker [1985] has proposed the description of perceptual features which also depend on oriented patterns. Two types of features, one-dimensional contours and two-dimensional flows, have been mentioned. The one-dimensional contours underlie the perception of occluding edges and shadow boundaries while the two-dimensional flows are related to the perception of surfaces similar to furs, hairs, wheat fields, grass, water falls, and snow. Separate computational approaches for identifying these descriptions have also been provided. There are other processes which shape the world and a lot of effort has been put into decomposing some of the basic patterns in natural scenes into parts. Witkin and Tenenbaum [1983] have acknowledged the role of scene structure in machine vision. Pentland [1986a, 1986b] proposed a theory based on intermediate types of models which are in-between the point-wise primitives of Barrow and specific object models, and are known as parts. These primitive models can be thought of analogous to a lump of clay which can be deformed or shaped without changing the main perceptual notion of the primitive model. The three-dimensional geometric primitives are another possibility which have been employed in higher level processes. Marr and Nishihara [1978] have suggested similar primitives based on generalised cylinders which they identify from the primal and  $2^{1/2}$ -D sketch.

The second approach assumes that partitioning of the image into coherent regions is the main goal of the perceptual organisation. The perceptual primitives in this case depend on the image content rather than the corresponding scene. These techniques are based on the organising principles provided by Gestalt psychologists [Wertheimer 1923, Koffka 1935], which ignore the concepts of scene geometry, illumination etc. The Gestaltists have provided a highly convincing set of organising laws which indicate

how humans perceive an object from a scene, and these are largely accepted as valid part of the perception process. The Gestalt laws explain how humans group stimulus elements together during perception, and they include:

*Proximity:* The elements that are close together tend to be perceived as a group.

*Similarity:* Similar elements tend to be grouped.

*Continuity:* The group which minimises change or discontinuity, and thus maximises good continuity is preferred.

*Closure:* Stimulus elements tend to be grouped into complete figures.

*Simplicity:* When more than one grouping exists and there is competition between groups, then the ambiguity tends to be resolved in favour of the simplest alternative.

*Symmetry:* Line drawings and regions bounded by symmetrical borders tend to be perceived as coherent figures.

*Common Fate:* The elements which move together with a uniform velocity through a field of similar stationary elements, are perceived as a coherent group.

Some presentations of the Gestalt principles distinguish between the laws of figure ground segregation and the laws of grouping. In texture discrimination the main emphasis is in segregating different features and then inserting partitioning boundaries where the texture is different. Psychophysical studies on human texture perception has led to the discovery of some conspicuous features (textons) which are detected by the pre-attentive vision

system instantaneously and without any effort [Julesz 1983]. This supports the existence of segregation phenomenon which identify these features at the very early stage. It has also been suggested that human visual analysis is functionally divided into an early pre-attentive level of processing where simple features are detected in parallel, and a later stage in which the focus of attention is applied to join these features into coherent objects [Treisman 1985].

Although it is very important to identify what to measure by looking into the human visual system, the more important step is how to combine these measurements into meaningful image entities. This process of combining can sometimes also be employed in identifying the significant early primitives from different alternatives. Stevens [1978] has attempted to solve this problem in order to identify locally parallel structures. He proposed an algorithm for selecting significant virtual lines for a particular neighbourhood. In it, an orientation peak is determined by histogramming the orientations of the virtual lines. Then the virtual line whose orientation is closest to the peak is selected as the significant virtual line. In this way the grouping process itself is used to choose the significant perceptual primitives. More generally, researchers have used an initial process to label the image pixels according to their local image properties (intensity, edge type, colour, or local texture). This produces a representation similar to the raw primal sketch described by Marr. The labelled pixels are known as *place markers* [Attneave 1974] or *place tokens* [Marr 1976, 1982]. Marr has suggested a more organised labelling process in which edges are identified at different resolutions (using different size edge detector masks) and those edges which exist at most of the resolutions are selected. This selection procedure is inadequate in the extraction of all the significant edge structures as explained in Section 3.5. The place tokens are defined as the significant places in the image, and they can be chosen in a variety of ways such as short line segments, the end points of lines (if the lines are not too short), blob positions, or a higher order group of place tokens. The local geometrical relations (orientation, position, and separation of similar adjacent

elements) between these tokens are made explicit by inserting virtual lines which join nearby place tokens. These place tokens are then grouped into contiguous regions, or coherent contours. This process is more readily known as image segmentation and has been recognised as a central problem of computer vision.

### **1.5 Image Segmentation and Perceptual Grouping**

The segmentation of images is defined as the process of isolating and identifying the regions of interest or partitioning the image into meaningful shapes. Discontinuities in the scene properties (e.g. distance, material composition, or motion) are the main clues for possible places where to insert a partition. The critical issue in image segmentation is that of relating the intensity variations with the corresponding physical discontinuities in the scene. Many segmentation techniques have been proposed and all of them are based on either detecting similarities or discontinuities in some pixel value (e.g. intensity, colour, or range). One of these techniques utilises the concept of similarity by extracting uniform regions and then obtaining the boundaries of those regions. This method is known as *region based segmentation* and it discovers uniform regions and in consequence their boundaries. A second method uses the concept of discontinuity, and works by detecting edge points which are those where the pixel values change abruptly. Then the edge points are grouped into boundaries between homogeneous regions of some property. These image partitioning techniques based on edge detection are carried out in a way which is thought to be analogous to the biological visual systems and they assume that most of the useful information is embedded in the boundaries between different regions.

In image segmentation, whether we adopt the edge detection or the region based method, grouping of pixels is at the heart of the process in isolating uniform regions and their boundaries. In region based segmentation, the adjacent pixels are merged into uniform regions, while in edge detection techniques, edge points are grouped into linear or curved segments to form the

boundaries of the objects. Perceptual grouping is a process which belongs to the early and intermediate levels of vision. It can be employed in grouping any of a large number of possible primitives including edges, bars, short segments, and corners or at the symbolic level, place tokens and virtual lines, to build a line drawing of the scene. Grouping can also be employed for three-dimensional reconstruction from surface patches with depth information. Almost all of the grouping processes try to bring together those elements in the image, which belong to the same part of the same object in the scene. The choice of grouping method will determine the primitives which must be generated by earlier processes and what information they should carry to aid the grouping. This will apply both at the early and intermediate levels.

The psychology community has carried out many investigations of human performance on specific grouping problems but most of the research is focused on performance rather than the mechanism of grouping.

Perceptual grouping based on proximity was the first criterion used to cluster dots. In following the Julesz's [1962] view that perceptual grouping can be achieved by clustering based on the geometrical properties of proximal dots, Zahn [1971] devised an algorithm for detecting Gestalt clusters based on proximity by using minimum spanning trees. Afterwards the same method was applied in identifying space curves by grouping the given points or short line segments [Zahn 1973]. Lester [1975] has gone a step further by assigning an additional attribute, *edge strength* to the links between proximal dots, based on the location of neighbours and the distance between dots. The decisions about the grouping of edges into segments and boundaries can be postponed until additional information becomes available. This type of delaying in grouping decisions was formalised later on by Marr as his *principle of least commitment*. O'Callaghan [1974, 1976] has surveyed different techniques of dot grouping and developed a local operator for extracting boundaries of different dot patterns.

Some of his work is also concentrated on edge and line organisation.

In the vision theory of Marr [1976, 1982], grouping has been used extensively in early visual processing for interpreting the primal sketch by partitioning it into *unit forms*. It was emphasised that grouping should be carried out on the basis of length, orientation, size, contrast, and spatial density. Two types of orientation based groupings were advocated. Using *curvilinear aggregation*, place tokens are merged into a group which preserve their orientation while in *theta-aggregation* similar oriented items (e.g. virtual lines) are grouped into a unit whose orientation differs from the items. The place tokens are also grouped into regions directly or the output contours from curvilinear aggregation are used to define the boundaries of the regions. Most of the grouping work put forward by Marr is rather speculative, and neither fully specified nor implemented in computer programs.

According to Lowe and Binford [1982], the main task of early vision is to find meaningful groupings in the image. The *meaningfulness* is defined as the likelihood that a given grouping truly reflects an inter-dependence of its elements and has not arisen from some accidental alignment of independent elements. In this way meaningfulness of grouping is not only domain independent but also independent of the world knowledge. The groupings which they have considered included collinearity, curvilinearity, predominant orientation, repetition and symmetry. The authors have suggested that instead of examining all possible groupings in an image, a search should be carried out on those classes of patterns known to be easily handled by the human visual system. These ideas have been implemented using a computer program and some results were obtained on the detection of meaningful linear groups among random dots.

None of these implementations of perceptual grouping provide a unified approach to image segmentation. Their inputs are either dot patterns or some very simple images. Therefore it is not possible, on the basis of their results, to make any claims as to the

methods applicability to general purpose image segmentation. In this thesis we will discuss the application of perceptual grouping in a unified way to achieve the segmentation of endoscopic colon images.

## 1.6 Research Objectives and Motivation

The main aim of this research is to develop machine vision techniques for a computer vision system, which provides sensing capabilities for navigating the endoscope. The endoscope vision system identifies at least two types of information: curved contours and darker regions. This information needs to be extracted in real-time, which in turn almost certainly means that the algorithms must be implementable in parallel.

The image contours are one of the main features used for guiding the endoscope and a considerable effort is concentrated on their extraction. Traditionally contours are extracted by detecting edge points which are then linked sequentially to build contours. A lot of effort has been spent on detecting edges starting from the development of edge detectors by Roberts [1965], Prewitt [1970] and Sobel [Duda and Hart 1973] and continuing with more recent work by Marr and Hildreth [1980], Canny [1983], Haralick [1984], Nalwa [Nalwa and Binford 1986], and Noble [1988]. Most of the edge detection techniques have been idealised for step edges. However, colon images have a variety of edges. Therefore any approach based on detecting step edges will not extract all the useful information available to construct the contours from colon images.

The motivation behind applying perceptual organisation in contour detection arose when simple edge detectors (e.g. Sobel, Prewitt, or Isotropic) produced unsatisfactory output, despite the fact that when the same image was presented to humans, they perceived the image contours without any difficulty. This led to a belief that some more complex perceptual organisation was being applied to the images by humans, and if it could be formalised, a

better segmentation method would result. Accordingly, some of the perceptual grouping criteria have been implemented to extract image contours from the output of any reasonable edge detector. No assumptions have been made about the level of noise in the images, since, as we shall see, perceptual grouping provides a highly effective noise filtering. Similarly, no assumptions have been made about the level at which edge points have been detected.

The key feature of this new algorithm for contour extraction from endoscopic images is the use of a simple edge detector from which an edge map is prepared without any significant thresholding. Thus, most of the intensity change information, including useful weak edges and edges due to noise, is retained. Then perceptual grouping is applied in building contours from the edge map. Different perceptual grouping techniques are applied at a number of hierarchical levels of refinement. As far as we know this is the first algorithm in which the perceptual grouping has been used for filtering edges and line segments from the processed image data. Previous implementations of perceptual grouping have, as mentioned earlier, been confined to dot patterns and other artificial data.

Normally in existing contour detection techniques, the edge point detection is assumed to be a local and parallel process. While the grouping is assumed as a global and sequential process. We will see that most of the early and intermediate level grouping techniques are implementable in parallel by following the psychophysical findings which suggest a purely local relationship between proximity and similarity in orientation and brightness [Zucker et al. 1982, Zucker 1983]. The contour extraction method described here is implementable in parallel using pyramid computer architecture [Khan and Gillies 1989a].

Region based segmentation is the more appropriate method for extracting the dark lumen from colon images in cases where it is directly visible. We have developed an algorithm for dark region extraction which uses a pyramid structure and is also

implementable on pyramid architecture based parallel computers [Khan and Gillies 1989b]. A variance-average pyramid is constructed starting from the bottom level and moving to the top and during this process the coherent darkest square region is also identified. The recursive variance calculation is formulated in such a way that for calculating the variance of the parent block, only the mean and variance of its children are employed. The whole process of dark seed region extraction is worked out in a single pass at the completion of the pyramid. To obtain an accurate region, the identified dark region can be used as a seed and similar neighbouring regions are merged with it. The method has been extended to general purpose region based segmentation and tested experimentally on a variety of images.

The next objective is to devise a suitable world and search space representation which can be incrementally constructed by integrating information from a sequence of images. Generally in most of the navigation systems two different representations for world and search space are used. But a mapping between the world and search space representation is required, which may make the updating of representations expensive. A single representation for world and search space called the *QL-Tree*, based on a linked-list of quadtrees, is proposed which has the inherent features of easy updating, access, and search for navigation.

## CHAPTER 2

### VISUAL DATA ORGANISATION AND GROUPING

#### 2.1 Introduction

There is now a widely-held belief among the computer vision community that it is impossible to proceed from the pixel level image data to image understanding in a single step. However, this was not always the case. In the 1970s a cognitive approach was adapted to computer vision, which avoided a large amount of computation at image level and employed symbolic manipulation. This turned the machine vision research towards representing and manipulating facts about a particular domain and exploiting the domain specific knowledge. But the available techniques proved inadequate to bridge the gap between the pixel level image data and the desired symbolic description. Therefore in 1974 Marr's work at MIT directed attention towards the search for a collection of intermediate representations known as the raw primal sketch and  $2^{1/2}$ -D sketch, which would ultimately bridge the gap. Later on, Barrow and Tenenbaum [1978] termed these types of representations *intrinsic images*.

Almost all of the recently proposed vision systems are based on the *signal-to-symbols* paradigm, in which we start from the pixel level signals and describe successively more organised attributes of the data. The representation gap between the pixel level data and symbolic descriptions is filled by a set of visual data representations which are arranged in an hierarchy of increasing abstraction. At each level of this hierarchy, we need to define a vocabulary of primitives which makes the information explicit for recognition or utilisation at the next hierarchical level. One of the most important factors in defining these intermediate visual representations are the transformational processes which

translate one representation to the other. These representations depend on the way in which the data is organised at each level and the processes that are used at different levels of hierarchy. It is clear that grouping and organisational processes are essential parts of these intermediate representations and therefore they must be studied in detail to obtain a better set of visual representations. Psychology and neurophysiology provide models of organic vision and they have influenced research in machine vision. From the point of view of organisation and grouping of pictorial data, most of the existing computer vision models do not follow those advocated in psychology or neurophysiology. Neural networks or connectionist models are the only systems constructed to make use of some of the organisational principles that are thought to be used in the brain. The psychology models usually deal with the overall input and output of perceptual behaviour in a much broader sense. The research in this area does not provide a general solution to the problems in vision. For example it is not clear how features in an arbitrary image are mapped to an interpretation. On the other hand there is little known about the models of neurophysiology. For example, Hubel and Wiesel's [1977] pioneering experimental work was aimed at determining how the low level image data is organised and aggregated into a tabular format in the visual cortex of monkey. Although, it cannot be stated definitely that these principles constitute the information used by humans for shape analysis, we can explore the organisation and architecture of visual cortex by employing the different aggregation principles discovered so far.

## **2.2 Early Data Organisation in Biological Vision**

Human vision seems so simple and effortless that we rarely realise the complexity and difficulty of the problem. The biological vision system may be considered as a process which transforms the input image intensities into perception. The world is created from a series of images projected onto the retina. The high-level retinal processing is related inversely to the intelligence and evolutionary complexities of animals. In the case of frogs, most of

the feature detection is performed at retinal level while in mammals and other higher vertebrates, geometric feature detection takes place in the visual cortex. The eyes of various mammals are generally similar and for experimental purposes cats and monkeys have been studied rather than humans. From the experimental data we only have a basic knowledge of the early stages of image data organisation and processing, which itself is not completely understood.

### 2.2.1 Low Level Feature Detection

Some very interesting experiments have been performed with frogs, cats, and monkeys which explain the nature of groups of neurones capable of extracting various edge like features from the input image [Lettvin et al. 1959, Hubel and Wiesel 1962, 1963, 1968, 1977]. It appears that these features are organised in an hierarchical manner. The visual systems of animals are divided into two broader classes. The first of which are the visual system of frogs, rabbits, squirrels, and other lower vertebrates. The ganglion cells in their retina perform low level feature detection. The visual detectors of these animals are sensitive to edges, orientation, and directional movements.

The second category of visual systems are associated with higher vertebrates including cats, monkeys, and humans and their ganglion cells are only responsible for measuring contrast and colour. The eye of higher vertebrates serves as a sensor while the visual cortex is the main place in brain where the actual vision mechanism takes place. The eyes have two distinct types of visual systems based on two types of photo-receptors, rod cells and cone cells. In a general sense, the rod and cone photo-receptors are the transducing elements which transform the focused image on the retina into electrical energy signals. The cone cells extract colour information and are used for detailed vision. They are smaller in size and densely populate the centre of retina. In the foveal region they communicate with the brain (visual cortex) through bipolar and ganglion cells (see Figure 2.1). The rod cells are more sensitive to light and their density is greater in the periphery of retina.

They work in groups and feed the visual cortex through a much smaller number of bipolar and ganglion cells. It is observed that rod cells also detect movement and other anomalies in the visual field and then cone cells are used for detailed analysis by slewing and focusing the eye. In the dark, rod cells perform the task of night vision and take over the additional responsibilities of shape perception. The On-Centre and Off-Centre type contrast sensitive concentric receptors exist at the retinal and lateral geniculate level of cats and monkeys. A line stimulus produces a significant response if it covers a large part of the centre and only a small part of the surround. These cells respond well to lines of any orientation due to their circular symmetry. Their outputs are connected to the primary visual cortex which is also known as area 17.

There is plenty of evidence from neurophysiology that at the cortical level of cats and monkeys, geometrical edge features are computed. Ideally we are interested in knowing what happens in the human's visual cortex but it is not possible with the existing techniques to ethically perform experiments on humans. Most of the experiments on the visual cortex of cats and monkeys are due to Hubel and Wiesel and it is usually assumed that a similar image processing takes place in the human brain. The visual cortex seems to be structured in an hierarchy of computation complexities. In the cortex of the monkey a large number of concentric cells are found which behave like geniculate cells. These cells also have circular symmetric fields and it appears that these less sophisticated cells are immediately connected to the inputs from lateral geniculate nucleus (LGN) and retina. All of these cells are located in the lower part of one layer in the cortex, known as layer IV. In addition to these concentric cells three distinct types of receptive fields have been observed in the visual cortex named as simple, complex, and hypercomplex cells.

### *Simple Cell*

The simple cells, which are at the lower level of hierarchy, are located in the so called area 17 of cortex. It seems that these cells

receive their input directly from groups of cells with centre-surround symmetrical receptors: the type of cells found in layer IV of cortex. They are sensitive to bars, slits, and edges of specific orientations.

### *Complex Cell*

The complex cells can be understood by supposing that they receive inputs from many simple cells, all with the same orientation preference. These feature detectors are at the next hierarchical level, also found in the area 17 of cortex and they respond to spatially oriented bars and edges. A complex cell is probably just as precise in its orientation specificity as the simple cell but it is less particular about the positioning of the bar. In other words the complex cells are invariant to translation but not to rotation. It produces a strong response if a line is kept in the optimal direction and is moved across the receptive fields.

### *Hypercomplex Cell*

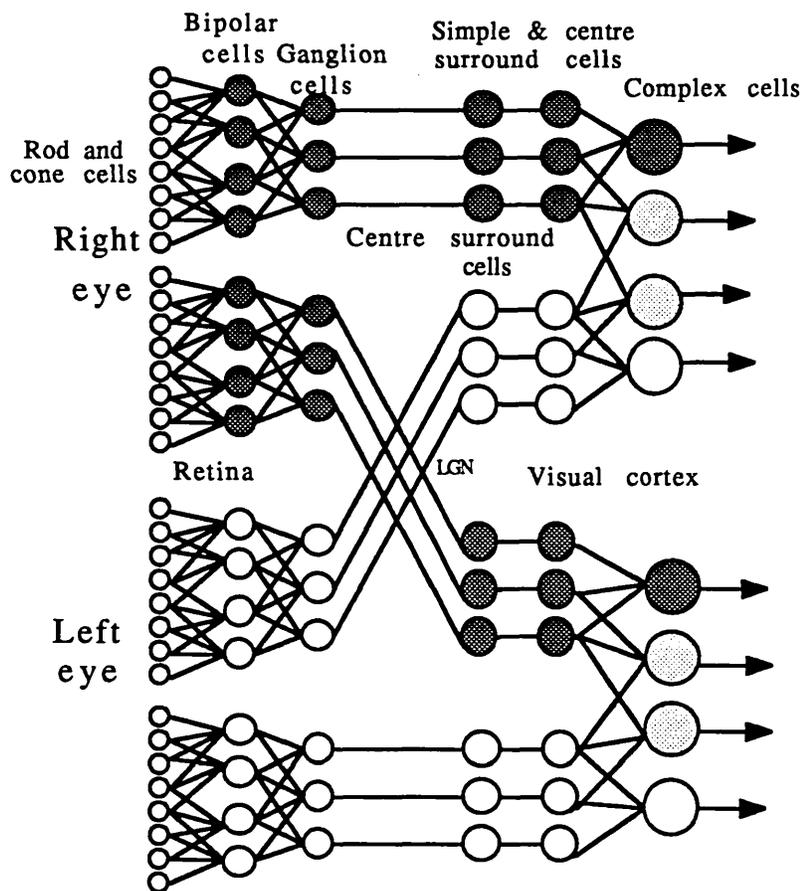
These are the receptive fields at the highest point of hierarchy in the cortex. Hypercomplex cells resemble the complex cells but if the line extends beyond its region of response, its response is reduced or completely abolished. There are two categories of hypercomplex cells: type I and II. The type I cells respond to moving, oriented, and directionally selective lines. They specifically respond to ends of lines. The hypercomplex II cells respond to corners. Actually both of these cells underlie the importance of the physical terminations of input patterns.

We have discussed the functioning of the low level feature detectors cautiously due to the experimental paradigm in neurophysiology. It is not clear whether the functions of these biological receptive fields are programmed at birth or learned in the early stages of contact with the environment. It seems clear that low level vision in animals is characterised in terms of line segments and edges moving in specified directions and orientations. How this data is organised so that we are able to

perceive different shapes, is not yet clear. A possible answer, using the neurophysiological view, is presented next.

### 2.2.2 Visual Signals, Pathways and Organisation

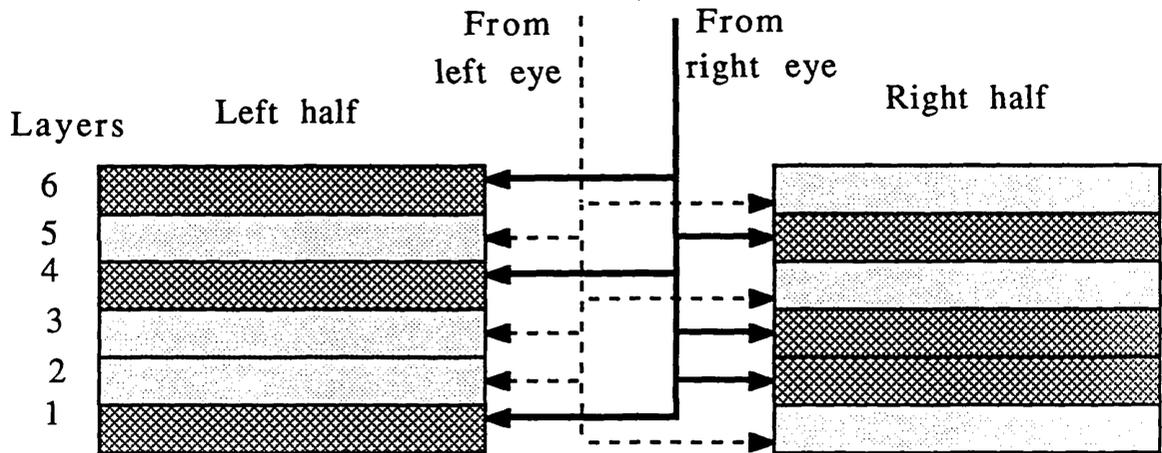
It appears that the organisation of the different features described in the previous section takes place at the initial stages of image analysis. This point can be argued mainly from the results of neurophysiological experimental data. The research by Hubel, Wiesel, and others has provided some clues about the organisation of data in the visual cortex. It is still an open question whether or not the aggregating principles which have been discovered so far underlie the initial process in shape perception.



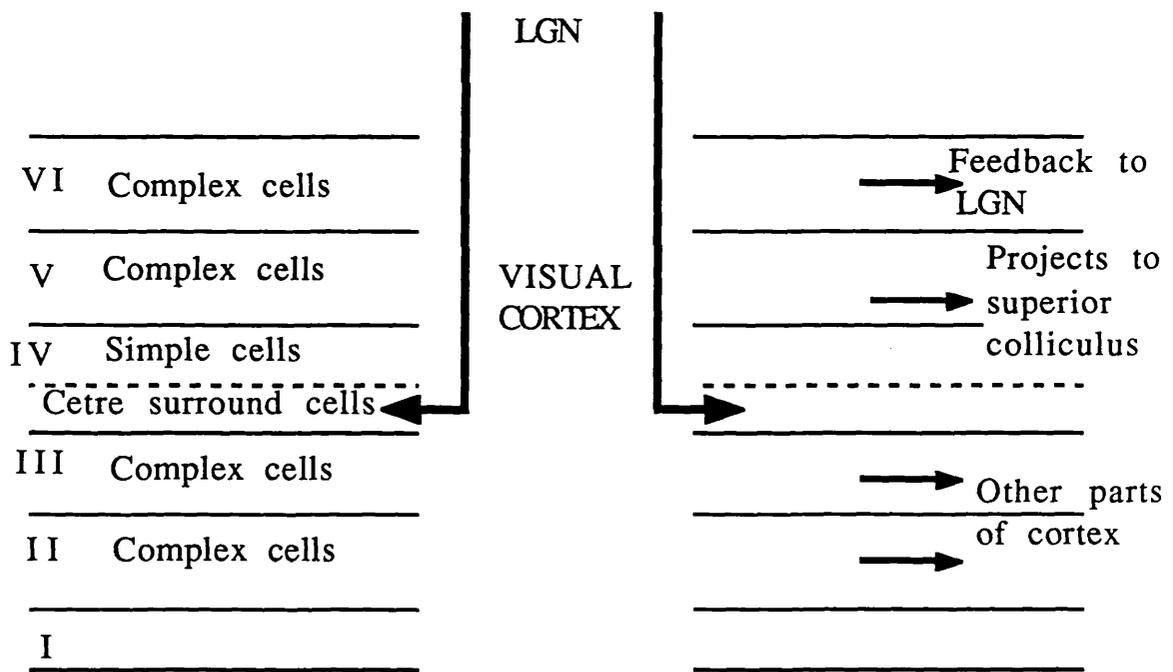
*Figure 2.1: Schematic diagram of the path from retina to visual cortex.*

We start by tracing visual signals from the ganglion cells, which provide inputs to the LGN cells from both the left and right eyes. The path from the retinal cells through the ganglion cells is shown schematically in Figure 2.1. The individual LGN cells are dedicated to process inputs from only one eye. An organisation of image data also takes place at LGN level in terms of the field of view. The input signals to the left of the LGN originate from the right side of the visual field and vice versa. There are six distinct monocular layers of the cells in both the left and right side of the LGN. Figure 2.2a explains these layers of cells in terms of their physical locations and origin of their input signals. The cells in layer 1, 4, and 6 are fed from the eye on the same side while layer 2, 3, and 5 are connected to the signals from the eye of opposite side. A high degree of order in the spatial relationship between cells has been found. Along the vertical section of these layers of cells, it is observed that the receptive fields originate in the same spatial neighbourhood of the field of view.

The visual cortex cells are at the next level of hierarchy. There are six layers in the visual cortex which are shown in Figure 2.2b. The outputs of the centre-surround cells from LGN are connected to the centre-surround or simple cells in the bottom part of the layer IV of visual cortex. Crick et al. [1980] have hypothesised that these signals are a high resolution filtered version of the image. From the bottom of the layer IV, the outputs of the cortical and LGN centre-surround cells form groups to feed simple cells which are also found in layer IV. The processing in the layer IV is still monocular. The complex cells have been found in four of the other layers II, III, V, and VI. This is the place in visual cortex where the data from both eyes converge as input signals to single binocular complex cells. Almost half of the complex cells are fed from binocular data while others get monocular data. In the case of binocular cells two inputs generally produce the same output signal with respect to bar pattern, orientation, and directional movement but the strength of the output varies when the same stimulus is presented to the left and right eye individually. From one cell to the other, all degrees of ocular dominance has been



(a)



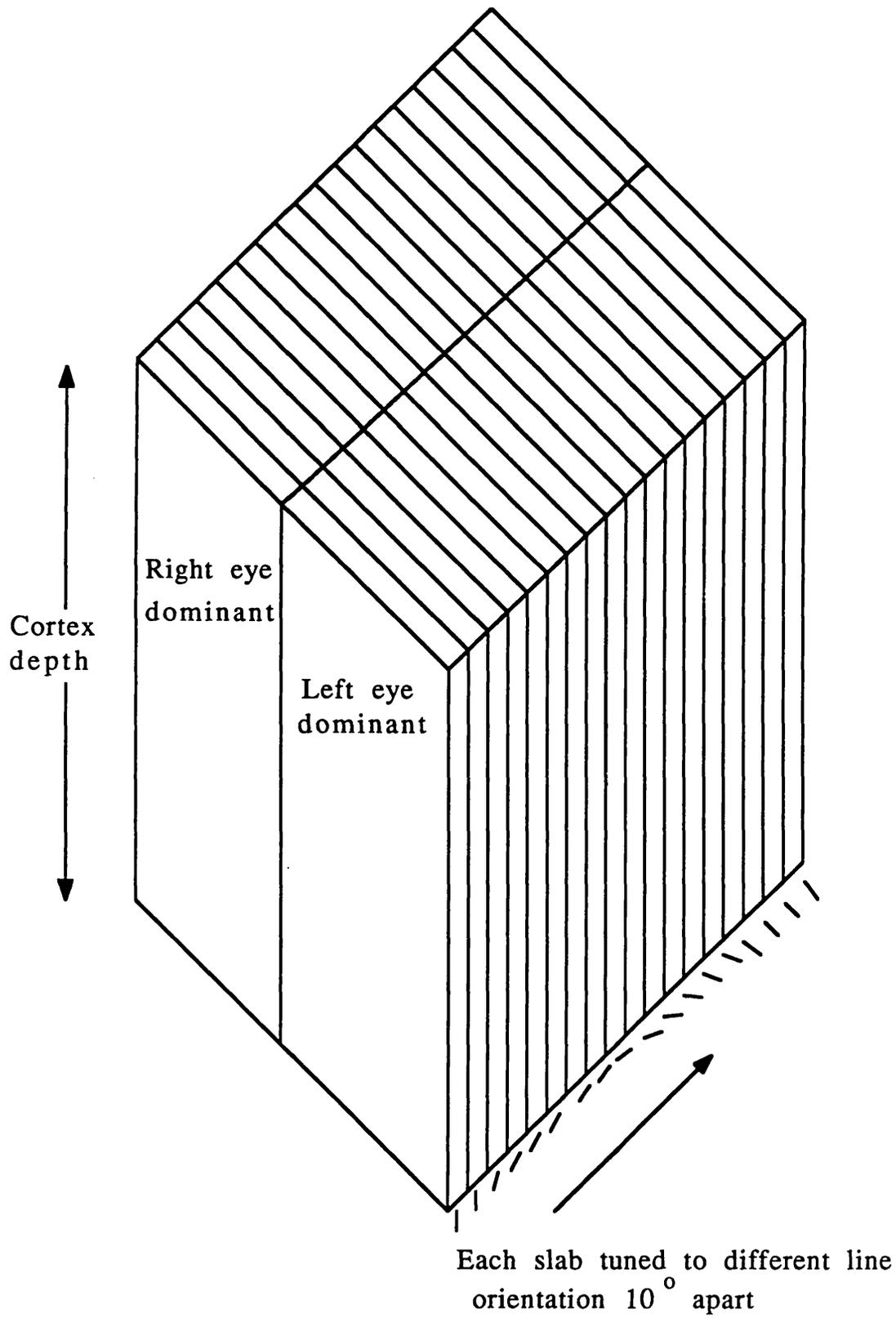
(b)

Figure 2.2: Block diagram of the layers in monkey's LGN and visual cortex.

reported, from complete dominance of one eye through equality to complete dominance by the other eye. The way the grouping of cortical cells has taken place is clear in the sense that the cells of similar complexity are grouped, with the centre-surround cells in the bottom of layer IV, the simple cells above them and the complex cells in layers II, III, V, and VI. These different layers have been organised by keeping in view the destination of their signals. For example, the deepest layer VI projects mainly back to the lateral geniculate body, layer V conveys its signals to the superior colliculus (a visual station in brain), and layer II and III send their signals to other parts of the cortex.

The next thing in the study of visual cortex is the position of its receptive field in the visual field. The spatial ordering in LGN described earlier is maintained in the visual cortex producing a cortical map of the visual field. This means that a structural relationship is maintained from the photo-receptors in the retina to the cells in the visual cortex. The visual cortex is made up of a number of layers and to study its properties it is necessary to investigate the cells in two directions. First is the perpendicular direction to the surface of the cortex while the second is in the horizontal or oblique plane. If one travels in the perpendicular direction passing cell after cell into the deeper layers, the receptive fields mostly overlap with each new field heaped on all the others. There is some variation in the size of these fields but each bar detector is tuned to the same angular orientation. Hubel and Wiesel [1977] referred these columns of receptive fields as *aggregate fields*. The variation in the size of the aggregate field depends on the distance of its cell's receptive field from the centre of the field of view. The investigation in the oblique direction of cortex surface has revealed slightly displaced aggregate fields and after every one to two millimetres, there is always a new aggregate field.

The visual cortex can be conceptualised as being subdivided into roughly parallel columns of tissue which may be swirled rather than planar as shown in Figure 2.3, approximately 1mm × 1mm in



*Figure 2.3: An elementary volumetric unit of the visual cortex.*

cross section and two millimetres deep. These volumetric units correspond to one aggregate field and originate from only one area of the visual field. Each of these units (except layer IV) contains a complete set of orientation columns and is partitioned into 50  $\mu\text{m}$  thick slabs with similar receptive field orientation. The adjacent slabs have a  $10^\circ$  shift in their line orientation. Slabs are arranged into coherent blocks with each block containing a right eye dominant and a left eye dominant column. Blocks near to the centre of gaze have tiny receptive fields while the peripheral blocks have larger receptive fields. It appears that the visual field is sampled to get the edge orientation and then the data is collected in a cortical table for further processing.

The organisation of the primary visual cortex (area 17) is explained in Figure 2.4 in the tabular form. Three variables: orientation, relative position of the aggregate field, and the ocular dominance are filled in the table which may be utilised by some higher level process in the brain. Apart from the vertical column structure, there is no experimental evidence about the use of this tabular organisation for high level perception. Additionally there is no answer to the question why the data is organised in this way, but it has been observed that this is a compact and efficient way of storing information. Perhaps the table is used to obtain histograms to use in some type of transformation (like a Hough transform) [Levine 1985] for recognising shape but this is pure speculation.

### **2.3 Pyramidal Architecture and Organic Vision**

Fast detection of global structures from digital images is an essential component of real-time machine vision. The real-time performance of human perception on complex images indicates that our visual system does not use conventional parallel processing. Reaction time experiments on human beings show that the recognition of complex objects is completed within roughly 400 to 800 milliseconds. There is also some evidence that the

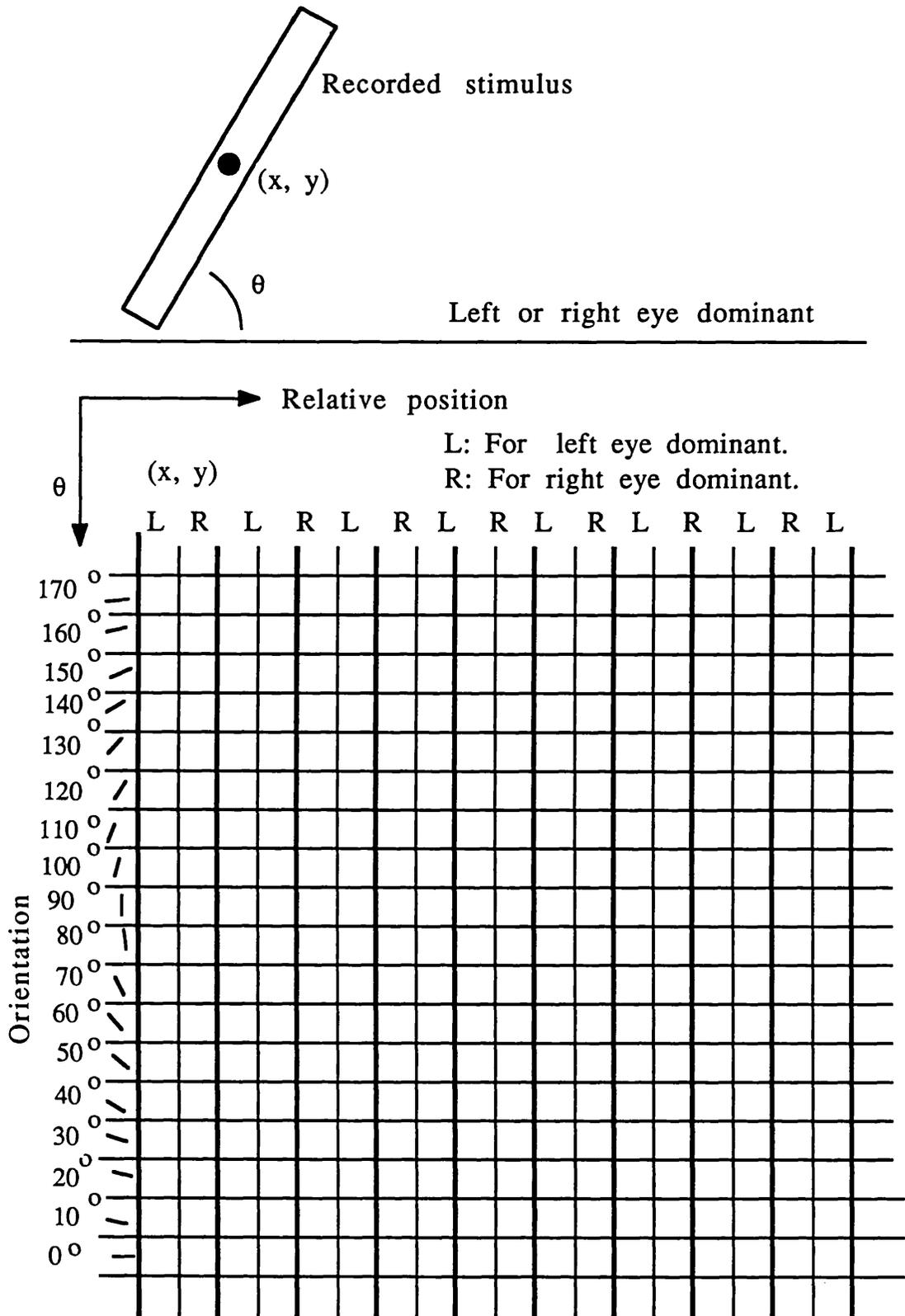


Figure 2.4: Data representation in tabular format for the elementary unit of visual cortex.

more complex the object, the longer it takes to be recognised. The experiments on the monkey's temporal cortex also reveal that they respond to a face or a particular face in only 70 to 200 milliseconds. This high speed for recognition is achieved despite the fact that the brain's basic processing elements, the neurones take about 1.5 millisecond to fire and send their response to other neurones. The contrast with machine vision is striking, vision programs take minutes or even hours to process a simple static scene. This is despite the fact that logic gates in today's computers are one to several hundred times faster than a neurone.

The brain is massively parallel but crucial to its success is the fact that it has a parallel-serial and hierarchical structure. It consists of millions (in the order of  $10^{10}$  to  $10^{11}$ ) of basic processing elements, the neurones which are organised in sets of columns side-by-side through all six layers of cortex. About  $\frac{2}{3}$  of the columns of neurones are pyramidal cells whose processes rise vertically through different layers of cortex and whose axons link to other cortical and sub-cortical areas. The actual links found between cortical areas and the other results of physiological experiments suggest a parallel-serial structure of cortex rather than a strict hierarchy. The information flows from visual cortex through at least twenty other visual and non-visual brain areas which are involved in perception. The overall structure of these areas also appears to be a parallel-serial hierarchy with each area richly linked to other areas. There is also evidence that by moving up through these areas, more and more abstract and complex features in larger regions of the retinal field are detected.

Perception is obviously a massively parallel process. The organic retina has million of rod/cone cells. For machine vision, the TV camera will also provide thousands, or millions of individual pixels. Only massively parallel architectures are likely to be fast enough to process the large resulting image arrays and successive resulting structure of image information. The parallel-hierarchical pyramid structure is among the most attractive candidate. Although a multi-computer designed in the form of pyramid is far simpler than the brain's perceptual system, it still keeps, in

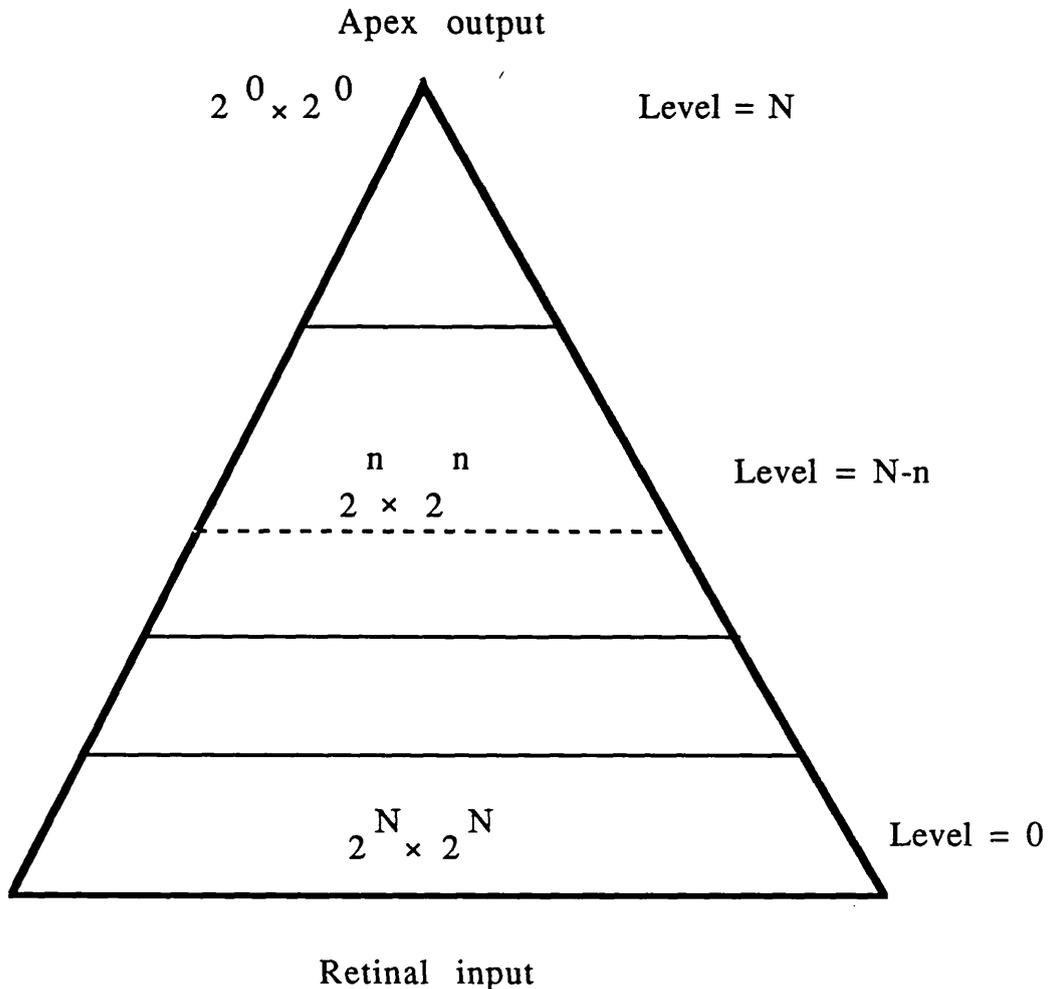
abstract and simplified form, many of the brain's essential features and more importantly its spirit.

A pyramid in its most basic form consists of successively smaller arrays of computers stacked over one another and linked via a tree as shown in Figure 2.5. The largest array is at the bottom of the stack and is known as retinal input and the smallest array (generally consisting of one processor) is at the top and called, the apex output. The links inside an array provide an efficient implementation of parallel local operations upon a small window, while the logarithmic based links between arrays of consecutive levels reduce the distance of order  $N$  (for an  $N \times N$  image or retinal size) between nodes in each of the array to the order of  $\log_2(N)$ .

The pyramid-style parallelism provides a much faster method of parallel computation to achieve fast recognition of global patterns. When implemented in parallel on suitable cellular pyramid hardware, most of the image analysis techniques require processing times of the order of the logarithm of image diameter. The contours are one of the most important global features in an image. Similar global patterns cannot be reliably recognised using the conjunction of local features. Pyramid or multi-resolution techniques provide different means of explicitly extracting the global structure in the image. When a fragment of contour is found at the lower level, the information can be passed up along with whatever more detailed information may be needed to specify exactly where that fragment started and ended. In this way the parents can stitch it properly with other fragments and pass that information up.

In chapter four the techniques based on pyramid processing of contour data (line segments and curves) will be described. Pyramids are often used to generate coarse to fine methods in edge and contour detection. The other image feature detection method is based on pyramid processing of intensity data (grey level or colour) or a local property map derived by applying feature detection operators to intensity arrays. A similar approach

for region based image segmentation using the pyramid will be described in chapter five.



*Figure 2.5: The pyramid architecture.*

## 2.4 Psychology of Vision

### 2.4.1 Background

The main aim in discussing the psychology of vision here is to explore some of the algorithmic techniques used by human visual system. These techniques can only be judged by their failures and successes in interpreting both natural and artificial images.

There are two main sources of information which can be employed by humans for visual perception. The first is the available sensory information and the second is relevant past experience and knowledge. There is a lot of controversy over the relative importance of these two factors. Gibson [1966] emphasises the role played by the stimulus information while Gregory [1970] and others argue for the constructive and hypothesis testing processes. The main focus of the Gibson theory is to provide an explanation of how humans perceive the environment in a veridical way. A radically different approach is argued by several theorists including Gregory and others. They regard perception as an active and constructive process. According to Gregory, the perceptual experiences are constructions from the data provided by sensors and drawn from the brain memory. This approach can readily account for perceptual errors and many visual illusions. But it seems that visual perception largely follows the bottom-up approach adapted by Gibson when the viewing conditions are good. It may involve top-down processes argued by Gregory and others, increasingly as the viewing conditions deteriorate.

#### 2.4.2 Form Perception

One of the most obvious and interesting facts of visual perception is that it is almost always organised. The important part of this organisation is the partitioning of the visual field into two parts *figure and ground*. The figure usually appears to be nearer than the ground which is extended uniformly behind the figure. This figure-ground organisation is one of the names given to perceptual organisation and it comes about fairly automatically. The parts of a scene may correspond to objects already seen but normally it is impossible to see the same object in the same configuration, illumination, and from the same perspective in space. This means that humans must be able to partition a scene into coherent, organised, and independently recognisable entities without prior knowledge.

The next question is how humans partition the scene into independent parts without knowing what might be present. The Gestalt psychologists argue that it reflects the basic and innate functioning of the human visual system. The Gestaltists were interested in some of the ways in which visual perception is organised. Their fundamental principles of perceptual organisation are a set of generic criteria which underlie the procedures discovered by nature for partitioning the visual field. One of the earliest and intuitively most acceptable collection of such laws are proposed by Wertheimer in 1923 and then elaborated by Koffka [1935]. These laws are based on a single fundamental principle (the law of Pragnanz) which is described by Koffka as follows:

"Psychological organisation will always be as good as the prevailing conditions allow. In practice a good form is the simplest or most uniform and organised of the available alternatives."

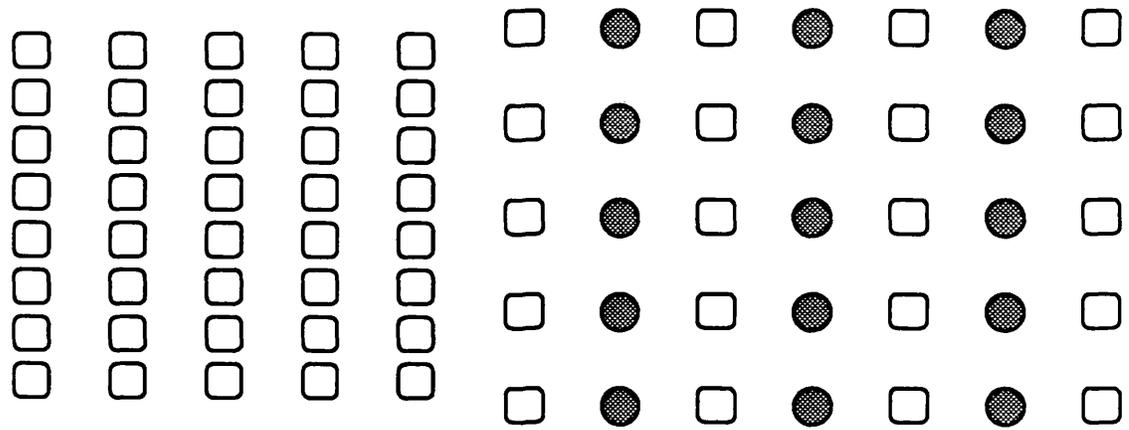
The laws of organisation have been formulated on the basis of their use in identifying ambiguous patterns similar to those shown in Figure 2.6. They will be seen to underlie the rules for perceptual grouping. These Gestalt laws include:

#### *The Law of Proximity*

The stimulus elements which are closer tend to be perceived as one entity. It will be observed that the closer elements in Figure 2.6a can be perceived as groups forming vertical columns.

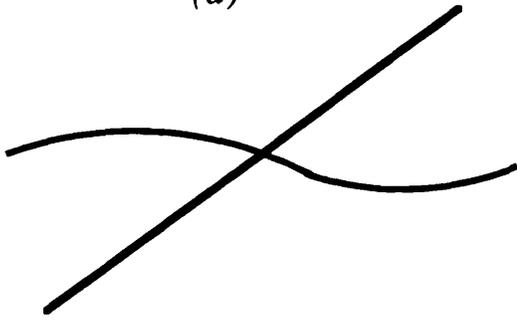
#### *The Law of Similarity*

Similar elements of a stimulus tend to be part of a unit. This similarity may be in grey level, colour, orientation, or shape as shown in Figure 2.6b.

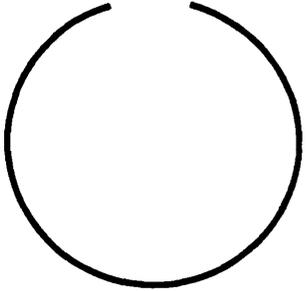


(a)

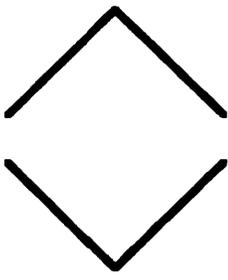
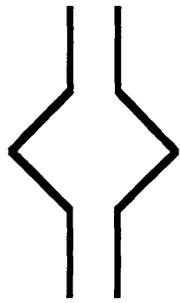
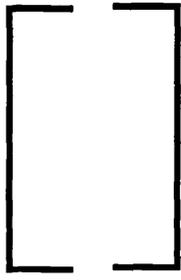
(b)



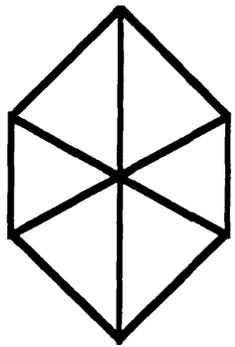
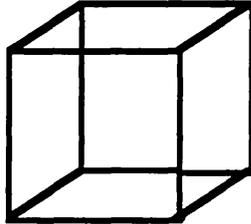
(c)



(d)



(e)



(f)

Figure 2.6: The laws of organisation.  
 (a) Proximity. (b) Similarity. (c) Continuity.  
 (d) Closure. (e) Symmetry.  
 (f) Simplicity; From two different projections of cube, the simpler (3-D) interpretation is perceived.

### *The Law of Good Continuity*

Stimuli tend to form a group which minimises a change or discontinuity as demonstrated in Figure 2.6c, which is perceived as two lines with first order continuity.

### *The Law of Closure*

The stimulus elements tend to be grouped into complete figures which are most commonly known. The stimulus in Figure 2.6d will be generally perceived as a circle despite the fact that some part of it is missing.

### *The Law of Symmetry*

The regions which are surrounded by symmetrical borders are perceived as coherent figures in the scenes shown in Figure 2.6e.

### *The Law of Simplicity*

In the stimulus where more than one figure can be perceived, the ambiguity is resolved in favour of the simplest alternative. For example if a smaller number of different angles or lines are required to interpret a figure as three-dimensional instead of two-dimensional, the observer will normally choose the three-dimensional alternative. This effect is shown in Figure 2.6f.

### *The Law of Common Fate*

If a group of elements are moving with a uniform velocity through a field of similar elements, the moving elements are perceived as part of a coherent group. In the study of obtaining structure from motion, Ullman [1979] has used this law of common fate.

The major problem with the above set of laws is their lack of explanatory power. It is possible to argue that all perceptual tendencies are implied explanations of how sensed data relates to the scene content. One of the explanations is that any partitioning

decision must satisfy the criteria of completeness, stability, and limited complexity. This viewpoint provides a broader basis for understanding Gestalt laws. Some additional ideas about the nature of perceptual organisation, such as the existence of a pre-attentive visual system and a vocabulary of perceptual primitives has been discussed in the previous chapter.

## 2.5 Classification of Grouping Processes

In machine vision, organisation of the intermediate levels receive image level features from low level image data and produce different intermediate representations. The main process in building intermediate representations from different types of features is to group them. The perceptual organisation can also be defined as the basic capability in human vision to derive relevant groupings from an image where groups of features lead to structures. Many areas of research in computer vision, e.g. structure from motion or stereo matching, are basically grouping problems where the pixels are grouped into sets of related features. The grouping processes establish relations between different elements of the image which hopefully will survive at the higher levels of vision. The best rules for grouping lead to those groups which are retained intact during the higher levels of the machine vision process. This principle is similar to the Marr's [1976] *principle of least commitment*.

There are two distinct classes of grouping. One is based on expectations and a-priori knowledge. Hough techniques are a good example of this category. An assumption is made as to the probable shape of the group, which is usually a first or second order curve. Then a search is made for the instances of that shape.

The second class of grouping is based on perceptual organisation and uses the different laws of organisation described in the previous section. This type of grouping is currently supported by Witkin, Tenenbaum [1983] and Lowe [1985] and is based on the argument of *non-accidental*. This is the degree to which an image relation is not arisen by accident. It is normally assumed by

psychophysical researchers that there is little, or no chance of any regular relation existing in the image by accident. The non-accidental argument in the first instance, seems to eliminate expectation but actually it only reduces the importance of the role of prior knowledge of the scene content at early and intermediate stages of vision.

### 2.5.1 Perceptual Grouping

The perceptual grouping is carried out on the basis of image content and does not require any scene specific knowledge. This means that this type of grouping is very useful and effective for the development of a general purpose and domain independent vision system. As these grouping processes do not utilise any domain specific knowledge, they can only be used as an early or intermediate level processes. However, they can be used to play a very important role in reducing the amount of image level data effectively without losing any useful information. Following perceptual grouping some knowledge of the scene can be employed for recognition.

The aim of this research is to use perceptual grouping for partial segmentation of colon images and thus avoid exhaustive search procedures. In this approach, the perceptual grouping is a part of early and intermediate processing which produces different intrinsic characteristic images.

In the case of segmentation through edge detection, perceptual grouping has not been employed significantly in the previous work. The different laws of perceptual organisation have only been used for dot grouping or for segmenting a very simple class of images. In the next two chapters, we will see that these laws can be applied to extract early features similar to those found in animal vision (e.g. linear edge segments) and then to group these features into contours. The edge level linear segments can be extracted at different resolutions by grouping on the basis of similarity in edge orientation, magnitude and edge pixel value in addition to proximity, continuity, and connectivity. At higher

levels, in contour extraction, linear segments can be grouped at different resolutions by using the principles of collinearity, curvilinearity, and theta-aggregation. Most of the laws of grouping define those image relations which are independent of any changes in viewpoint and therefore lead to a predictable structure in the scene.

### 2.5.2 Hough Techniques

As mentioned earlier, the Hough techniques belong to that class of grouping which is based on expectations and a-priori knowledge. In contrast to the perceptual groupings which are mostly data-driven, grouping based on Hough techniques utilise the knowledge of important scene structures. These techniques are well developed and more than 150 research, development, and application projects on them have been reported since the time Hough [1962] introduced the transform. The Hough techniques include a vast variety of clustering, histogram analysis and estimation strategies. They transform data in such a way that the shape of interest will form into clusters. The Hough transform was first developed for grouping features into simple geometric lines [Duda and Hart 1972] and curves but more recently it has been generalised and can be implemented to group two or three-dimensional features [Ballard 1981]. A comprehensive survey of Hough techniques has been completed recently [Illingworth and Kittler 1988], which can be consulted for detailed applications of Hough transform.

Basically the Hough transform is a mapping from one representation (e.g. primal sketch which includes edge information) into a new space in which elementary shapes or shape features are easy to extract. The simplest form of the Hough transform is described for line detection. Consider the line equation:

$$x \cos\theta + y \sin\theta = \rho \quad (2.1)$$

Suppose we have detected features with a local edge orientation and a measure of edge contrast.

For a line we can define a  $(\rho, \theta)$  parameter space. Restricting  $\theta$  in the interval  $[0, \pi]$  and  $\rho$  in the interval  $[-\sqrt{(x^2+y^2)} < \rho < \sqrt{(x^2+y^2)}]$ .

A line in x-y plane will map to a point in  $(\rho, \theta)$  plane as shown in Figure 2.7 with different points  $(x_i, y_i)$  on the line which map to the same point in  $(\rho, \theta)$  space.

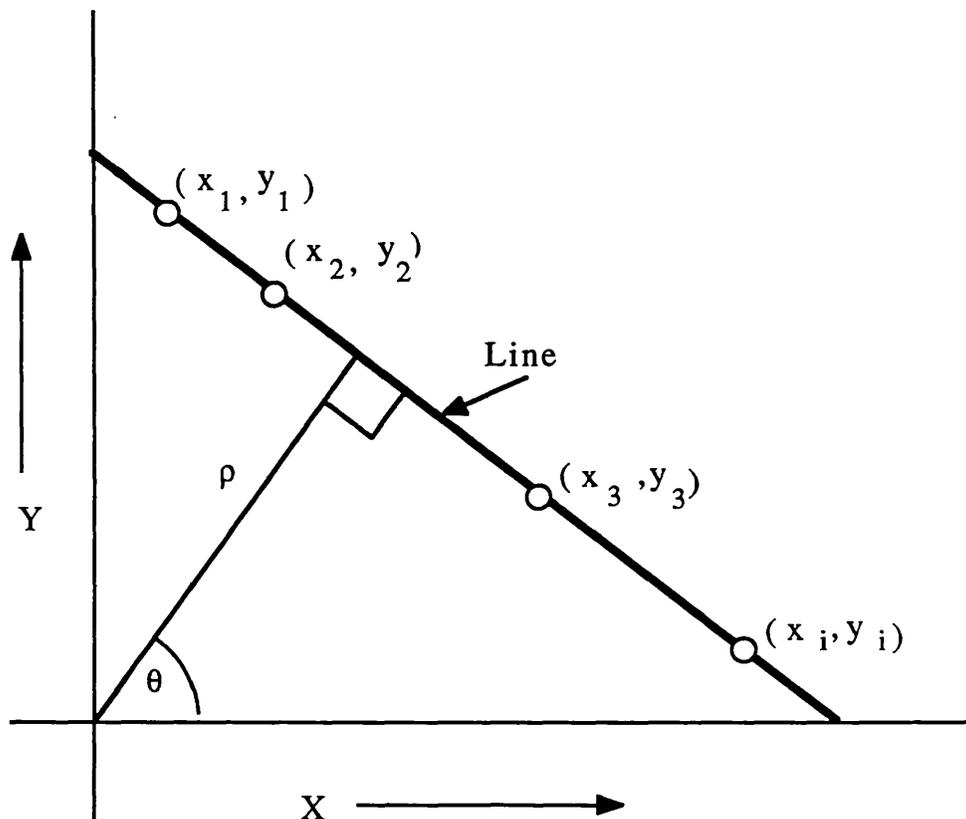


Figure 2.7: The  $(\rho, \theta)$  parameterisation of the line in Hough transform.

This relation between image space and parameter space can be formally extended to the following algorithm for line detection:

-Quantise parameter space for maximum and minimum values of  $(\rho, \theta)$ .

-Form an accumulator array  $A_c(\rho, \theta)$  with its initial value as zero.

-If the edge contrast exceeds some threshold, for each such edge point  $(x_i, y_i)$ , increment each array element along the appropriate line.

$$\text{i.e. } A_c(\rho, \theta) = A_c(\rho, \theta) + 1$$

for  $\rho$  and  $\theta$  satisfying the line equation.

$$\rho = x \cos\theta + y \sin\theta$$

-Local maxima in the accumulator array will correspond to collinear points in the image array. The values of the accumulator array give the number of points on the line.

Generally this technique can be extended to other curves and shapes represented by a function  $F(\mathbf{X}, \mathbf{P})$  where  $\mathbf{P}$  is the parameter vector and  $\mathbf{X}$  represents the initial representation. For example to detect a circle parameterised by the equation.

$$(x - a)^2 + (y - b)^2 = r^2 \quad (2.2)$$

The parameter vector,  $\mathbf{P}$  will be:

$$\mathbf{P} = (a, b, r)$$

and the Cartesian vector  $\mathbf{X}$  is:

$$\mathbf{X} = (x, y)$$

The major draw back of the Hough techniques is that their computation cost and the size of the accumulator array increases exponentially with the increase in number of parameters for complex shapes. There are, however, a number of extensions which can be employed to reduce the computation, for example we can utilise gradient direction in detecting conics like circles and ellipses. Similarly magnitude of edge contrast can be utilised as a heuristic in incrementing the accumulator.

Brady [1983] has criticised Hough techniques as means for obtaining representations at intermediate level vision due to the following factors.

- Widely spread weak evidence can become strong evidence after transformation.

- In some cases a small portion of contour can guarantee the presence of an object. However, Hough's voting system may reject the detection of that object due to its small number of votes.

- It does not provide means for detecting localised imperfection, in the objects.

Most of this criticism of Hough transform is of the representation that it provides, rather than its use as a grouping process. The biggest criticism against the Hough transform, in the context of perceptual organisation, is that it entirely ignores the proximity criterion for grouping. This drawback has been overcome in the implementation which will be described in the next chapter.

## CHAPTER 3

### EXTRACTION OF CURVED LINE SEGMENTS

#### 3.1 Introduction

In machine vision, the boundary of a uniform region is usually a very important image feature. Boundaries provide useful information for segmenting images into meaningful regions. There is also a widely held belief that humans isolate object boundaries in a scene before recognising them. In endoscopic colon images, the contours due to the occluding edges of the inner muscles of human colon, can be used for guiding the endoscope. These contours also provide clues for building a three-dimensional representation of the inner colon. There are a number of approaches to form a representation of object boundaries or contours in an image. One approach is to extract uniform regions and then process them for their boundaries. These region based techniques always form closed boundaries. An alternate approach assumes that a series of edge points define the boundary. The edge points are detected by sensing where the pixel value (intensity, colour, texture, or range) changes abruptly. These edge points can be linked into lines or curves. Computation of the gradient in pixel intensity or any other image property, can be achieved by some differentiation operator. The spatial and other relationships among the edges are utilised to infer more global entities in the form of boundaries.

The approach to extract curved contours and boundaries, described in this thesis, utilises a bottom-up organisation of edge point data. Our method of organising the image data uses the information presented in the data itself and perceptual organisation rules which are domain independent. Most of the boundary extraction techniques link edge points into boundaries directly and sequentially without an intermediate representation.

An alternative strategy is to form an intermediate representation of the desired contours before obtaining a final representation. This will not only limit the use of domain specific knowledge at the early stages of visual data organisation but will also reduce the amount of data to be processed at higher levels. The straight line segment representation is a good candidate to be formed from edge point data. An hierarchical line segment representation has been used at two resolutions by grouping edge points in different sized image windows. The grouping is directed by employing the different laws of perceptual organisation described in the previous chapter. It is easy to perform these processes on a local basis, and they are highly effective in filtering out noisy edges. Moreover, these grouping operations are amenable to parallel computation. When the line segments have been obtained, they are grouped hierarchically using a pyramid structure to form curved contours. This last part also allows parallel implementation on a pyramid based computer architecture. The step of extracting short line segments plays the central role in our method. Its ability to filter out noisy edges and to produce significant line segments has a considerable influence on the later, higher level, process of boundary formation. The perceptual grouping of edge points into straight line segments is treated in detail in this chapter before describing the overall contour extraction algorithm.

### **3.2 Motivation and Problem Definition**

The motivation behind the introduction of the straight line segment representation for curved boundary formation comes mainly from the neurophysiological studies of early data organisation in animal vision. This subject has been discussed extensively in the previous chapter. The experimental data about the different processes in the visual cortex of cats and monkeys indicates that a line segment representation is formed by simple and complex cells in the visual cortex. The straight line segments are detected by grouping the point data supplied by contrast sensitive centre-surround cells in lateral geniculate nucleus (LGN) and the visual cortex. The well known *signal to symbol paradigm*

also supports the idea that an intermediate representation should be formed as the first step in forming a boundary representation of an image.

Short line segments are good candidates for forming an intermediate representation. Clearly, they can be used to represent the boundaries of polyhedral objects. Moreover, curved object boundaries can also be approximated by piece-wise straight line segments. They thus have the ability to represent any type of contours and boundaries.

From the computational point of view, any representation in-between the edge point data and boundary contours will reduce the image data to be processed at higher levels. Generally, those intermediate representations which can be formed by local, independent and parallel processes, are preferred for early visual processing. Short line segments can be extracted independently and in parallel. Moreover, their extraction does not require any domain specific knowledge. Marr [1976] has also included the short line segments and bar like features in his well known raw primal sketch. We believe that short line segments extracted at different resolutions provide enough information for forming complete contours and boundaries.

The contours in endoscopic images are formed by the occluding edges of rings of muscle and do not come from the finer texture details and other artefacts of the human colon. The main problem with the extraction of these contours is the inherent noise in this type of medical images. Other problems appear due to specular reflections, uneven surface texture, and the presence of other matter in the environment of the colon. In a single image frame, visual inspection can only locate one or two rings of muscle usually, and at the most four to five rings. These are only partially visible due to bends, twists, and other irregularities in the colon. One useful fact about occluding contours is that they cannot cross each other. At the nearest point one contour will end on another. Therefore, the contours are well apart, and so, in order to find straight line segments which approximate the curved muscle

rings, we have assumed that, in a small square window of the image, there exists only a single line segment. The whole image is broken up into small windows which are processed independently of each other. Many of the rings of colon muscle have very low contrast due to uneven illumination. Therefore, in each window we are looking for a meaningful line which may be weak but perceptually significant. The corners and joints of the actual contours can be filled in afterwards by using edge point data. The size of the window is determined by the amount of detail which one wants to extract in a given image. We have found that the grouping of edge point data at two selected sizes of windows is adequate for extracting reasonable boundaries.

### 3.3 Previous Work on Extracting Line Segments

As argued in the previous section, the detection of short line segments by grouping edge point data is an important and initial step towards contour extraction. We will now review the prior line extraction techniques before describing a novel approach, developed for detecting short line segments. There are two main categories of edge point grouping techniques: global and local edge linking. The techniques using global criteria for linking are difficult to implement in parallel. They also become computationally expensive when the amount of edge data is large. Some of the global edge linking techniques are based on graph theoretical search methods where the edges are viewed as nodes of a graph and a cost factor is associated with each link between nodes [Martelli 1976]. The cost can be a function of the proximity and direction of the edge elements. The minimum cost paths in the graph are taken to correspond to the desired boundaries. Ramer [1975] has used an heuristic search technique to find paths. Although this was basically a global method, he needed to use an intermediate organisation based on streaks which were detected by searching the edge point data bi-directionally. Zahn [1971] built a minimum spanning tree in the graph to detect clusters. Recently the minimum spanning tree was used for extracting curvilinear features at global and local window levels [Suk and

Song 1984]. The results indicate a deterioration of boundaries in the local window mode. The global edge linking methods cope with noise by using domain specific knowledge and heuristics which make them less applicable to early and intermediate visual processing for general purpose vision. Due to their global nature, the computational cost is also not low enough for their use in real-time applications like the automation of an endoscope.

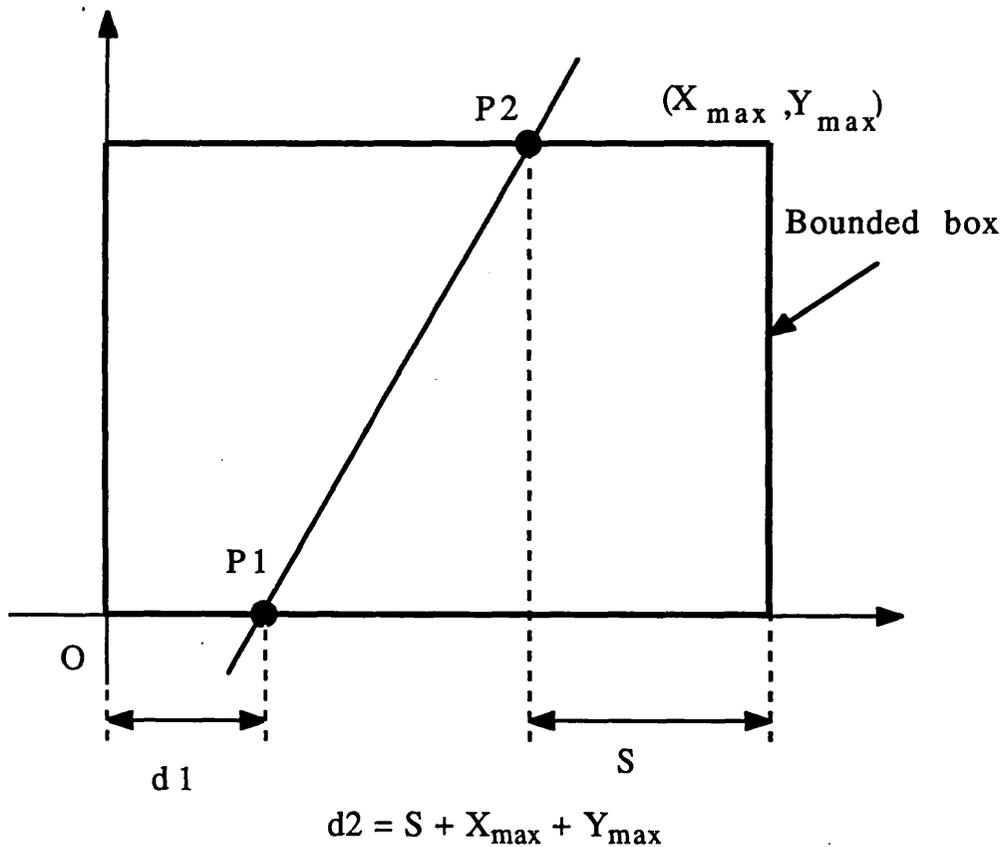
In the case of local edge linking, we will restrict ourselves to the extraction of only short straight line segments as an intermediate step for contour extraction. The problem of line extraction by grouping edge points has been worked on, since the early days of scene analysis. Roberts [1965] linked edges into straight segments locally in a  $4 \times 4$  window and edges due to noise were eliminated by ignoring those which do not have at least one neighbour with a direction within a fixed tolerance ( $23^\circ$  of tolerance was mentioned). Griffith [1973] used a more complicated technique for linking edges into straight lines. Two sets of feature points, vertical and horizontal edges, were used as input to the line extraction algorithm. The vertical and horizontal edges were projected on the y-axis and the x-axis respectively by rotating the axes to about 200 angles. The direction of the lines was determined by histogramming the occurrence of edges for each different angle of the axes.

The Hough transform, which has been introduced in the previous chapter as a grouping process, provides an interesting means of straight line extraction. This method is based on the transformation of points into straight lines using a parameter space. There are a number of parameterisation schemes developed for implementing the Hough transform and we will now discuss some of them here. Initially the line was parameterised in terms of its slope and intercept [Hough 1962]. The slope-intercept parameter space is unbounded which complicates its implementation. The angle-radius  $(\rho, \theta)$  parameterisation, discussed in chapter two, was proposed to overcome this problem where the radius,  $\rho$  is the distance of line from the origin and  $\theta$  is the angle of its normal [Duda and Hart 1972]. Generally, for a

given feature point the  $\rho$  value is computed for each of the quantised values of  $\theta$  using equation (2.1). The values of  $\rho$  are then quantised into a fixed number of intervals of width  $\Delta\rho$ . In this way, the image is divided into bar shaped windows. The selection of optimal quantisation of the angle-radius parameter space is a difficult problem in itself. The resolution of lines for a given quantisation changes with  $\theta$  and  $\rho$ . An extension of a line peak not only depends on its length and width, but also on the quantisation of the parameter space. The finer quantisation reduces the extent of the peak and it also enhances the effect of image quantisation. Van Veen and Groen [1981] have discussed quantisation errors in the Hough transform.

Wallace [1985] has proposed a new parameterisation in which the lines are represented by two points on the opposite ends of image boundary. The image is assumed to be bounded by a rectangular box whose sides are parallel to the x-axis and y-axis, extending from the origin to the vertex  $(X_{\max}, Y_{\max})$ . A line passing through the image is parameterised by the two crossing points where the line intersects the bounded box. The intersection point parameter space is explained in Figure 3.1, where P1 and P2 are the two intersection points for a given line. The points can be represented simply by their distances  $d_1$  and  $d_2$  from the origin along the perimeter of the bounding box such that  $d_1$  is less than  $d_2$ . Forman [1986] has argued for a hybrid parameterisation in-between the angle-radius and Wallace parameters. His angle-point parameterisation is based on the line direction and its intersection point with the image boundary. The *foot of normal* parameterisation was proposed by Davies [1986], in which the line is parameterised by its point of intersection with a normal vector from the image origin. It is also suggested that this parameterisation is suitable for small image windows.

We have adapted the Wallace parameter space for grouping edge points perceptually to extract line segments. The main advantages of this parameter space are its suitability for application to rasterised images, the constant line resolution throughout the



*Figure 3.1: Line parameterisation in terms of the two intersection points P1 and P2 between the line and the bounding box.*

image space, and its low computational cost because no trigonometric functions are involved. In addition to that, it is highly suitable for extracting line segments in small image windows due to two main reasons:

- It is easy to decide and compute which feature point votes for which line within the parameter space. Therefore sidelobes [Brown 1983] and bias [Cohen and Toussaint 1977] inherent in the original Hough transform can be reduced.

- For a small image window, where the accuracy of line end points is not important, the intersection

points can be used as the end points for the line extracted. This eliminates the need for computation of end points required in the traditional Hough space.

As discussed in the previous chapter, Hough techniques have been criticised on the violation of proximity criterion which is extensively used by humans in collinearity grouping. O'Gorman and Clowes [1976] tried to rectify the Hough transform line detection process by using gradient direction to determine relevant edges for a line in a given direction. Their grouping process is based on collinearity and rejects completely the principle of theta-aggregation advocated by Marr [1976]. Moreover, they have used a thresholded edge magnitude in the voting process which may suppress the contribution from weak but meaningful edges. Van Veen and Groen [1981] followed the O'Gorman and Clowes approach and suggested weighting the Hough transform with the probability density function of the gradient direction.

Almost all of the above boundary line extraction algorithms have been developed to detect boundaries of simple polyhedral objects. Their performance is also demonstrated on very simple noise free indoor scenes with ideal illumination conditions, containing only block and toy like objects. So we see that these methods require considerable improvements to cope with outdoor scenes with curved objects.

Canny [1983] has also linked the edge points into short contour segments (edgels) in his well accepted work on edge detection. His method is based on connecting edges in such a way that some portion of the contour is above a high threshold while the rest must be above a low threshold. The method is not good enough to detect a weak but meaningful long edge. Blicher [1984] has also pointed out a similar problem with Canny's edge linking and suggested the use of some Hough like method for linking edges into straight lines.

The use of edge orientation instead of magnitude in the initial organisation of image data has recently been advocated by Burns et al. [1986]. Previously edge magnitude was used in one form or another as a dominant measure. Their line extraction algorithm involves grouping of pixels into line support regions based on the similarity of edge orientation. The lines are extracted from each region by fitting a plane to the intensity surface of the region. In addition to other problems mentioned by the authors, the method does not handle the approximation of curved lines by straight line segments.

### **3.4 Perceptual Grouping of Edges into Line Segments**

The existing techniques for line segment extraction can not cope with different types of noise effectively. They are also unable to identify those line segments which are part of a curved contour or a curved surface. Therefore a new method, for grouping edge points into short line segments, is required. The method of grouping edges into short line segments, that has been developed, can be carried out in a local window mode and the process is implementable in parallel. Aggregating edge points is an early process in machine vision and therefore it is highly data directed. The information which is used to group edge points, is mainly carried by the edge points themselves in the form of:

- Edge location.
- Edge orientation and magnitude.
- Edge pixel value (grey level, colour, or range).

In the previous chapter, we have discussed different laws of organisation which can be utilised to group image data. Among them, the proximity, connectivity, and similarity in edge orientation, magnitude, and edge pixel intensity have been used for aggregating edge points into significant line segments. These grouping criteria seem very simple and ordinary but to date, no one has exploited their strength effectively and completely for segmenting images in a unified way. The grouping principles

follow the *non-accidental* argument, which is based on the fact that there is a very small probability of a regular relationship occurring by chance. The likelihood of a line structure, due to noise or some other accidental phenomenon decreases with its consistency under the different laws of perceptual grouping. The problems in using these aggregation methods is the lack of an effective implementation. There is also the unresolved problem of combining the results when different criteria give different outputs. It is difficult to deduce which criteria are the more effective. The individual aggregation methods are discussed first, and afterwards we shall consider how to combine their results.

### *Proximity*

The grouping based on spatial proximity has been used extensively in many clustering problems. The performance of proximity grouping depends on the accurate localisation of edges. Those edge points are grouped which are closer and lie on the straight line. Providing that the number of votes required to qualify as a line is sufficiently large, the use of a small window establishes the proximity criterion. In our case, the window sizes are  $12 \times 12$  and  $6 \times 6$  pixels and the minimum number of votes are eight and four respectively.

### *Connectivity*

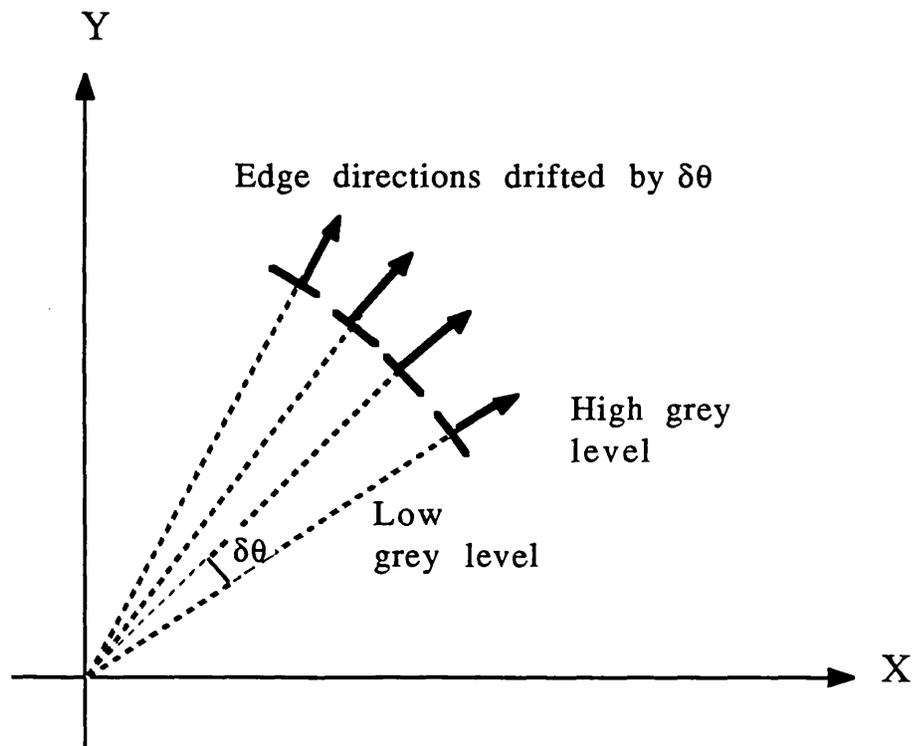
This is an important image relation since it is preserved over all possible viewpoints. Those edge points are grouped which are connected to each other on a straight line. For the size of window that we are using, no distance tolerance between two edge points is allowed for proximity grouping. Therefore, connectivity is automatically observed.

### *Similarity in Edge Orientation*

This is the most effective and widely used criterion for straight line extraction. Almost all of the previous line extraction methods use edge orientation in one way or another. Our method is

influenced by the work of O'Gorman and Clowes [1976] but we depart from their method in many respects. The problem we are facing is the extraction of curved contours of curved surfaces, which are approximated by line segments. Therefore, instead of selecting those edge points whose orientation is perpendicular to the line segment, a slow drift in the orientation is allowed from one edge point to the neighbouring edge point on the line. This drift helps in forming pseudo straight segments representing curved segments in the image. The slow change in edge orientation as one moves along a curved boundary is illustrated in Figure 3.2. A line segment can be represented by using an angle-radius ( $\rho, \theta$ ) parameterisation described earlier by the following equation:

$$x \cos \theta + y \sin \theta = \rho \tag{3.1}$$



*Figure 3.2: Grouping edges into curved line segments.*

By following the O'Gorman and Clowes approach, the angle  $\theta$  also represents the edge orientation and only those edges whose orientation is within a tolerance  $\delta\theta$  of the angle  $\theta$  (Figure 3.3) are allowed to vote for the line. In our method of orientation grouping a slow drift is allowed in the orientation of consecutive edges against a large angular tolerance. In this way, the grouping relationship formed is more perceptually stable and regular for extracting contours due to curved surfaces. The results on endoscopic images which are presented in this chapter also support this claim.

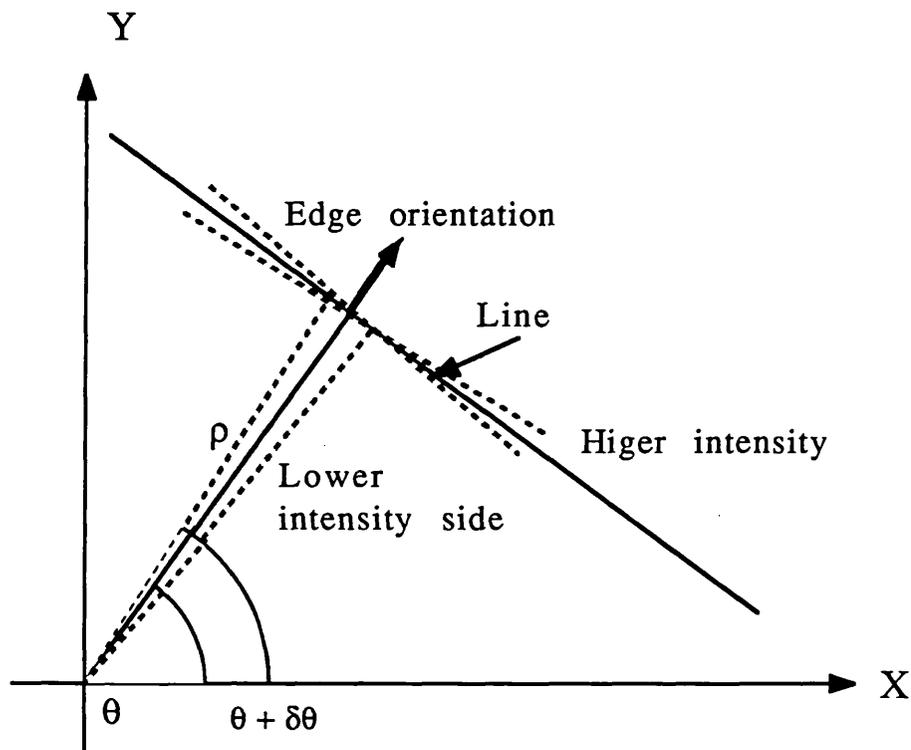
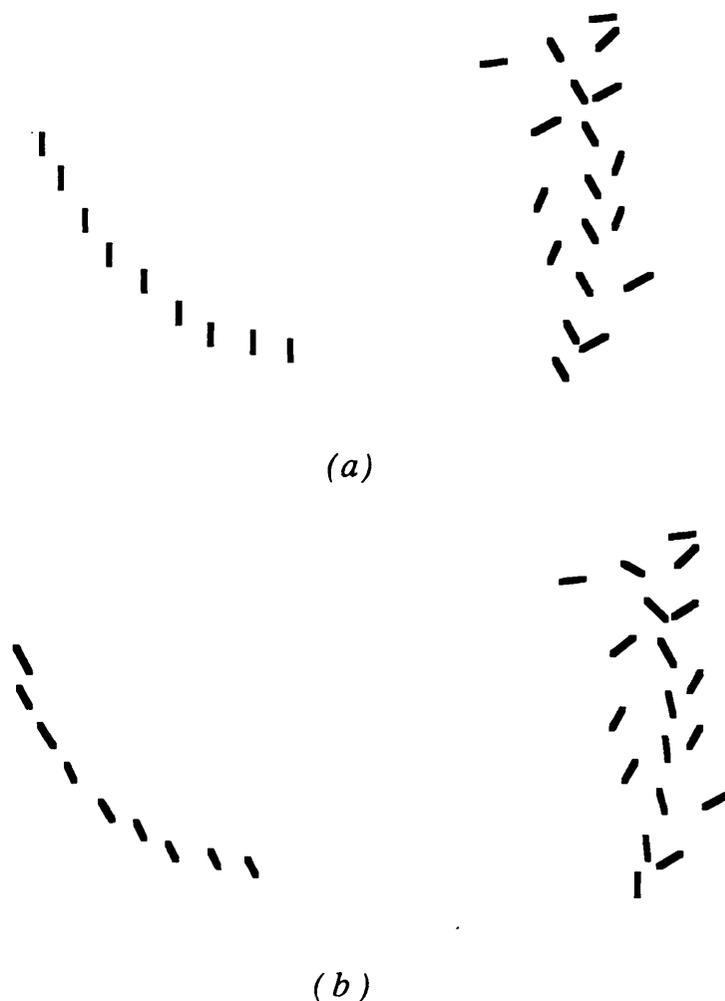


Figure 3.3: Relation between edge orientation and line direction.

The aggregation on the basis of similarity in edge orientation is also allowed whether the change in intensity is normal to the line segment or not. This is equivalent to Marr's grouping *principle of theta-aggregation* for line segments. The theta-aggregation

principle groups bar like features on the basis of parallelism and collinear displacements. Assuming that the edges are very short segments having directions, theta-aggregation is used to recover significant lines. It has been observed that theta-aggregation recovers those significant line segments which are generally missed. This principle is also combined with curvilinearity and orientation drift to extract useful line structures. In Figure 3.4 some stimuli are shown to demonstrate that humans use these grouping criteria to recover curves and lines.



*Figure 3.4: (a) Theta-Aggregation principle based stimulus. (b) Combining theta-aggregation with curvilinearity and orientation drift.*

### *Similarity in Edge Magnitude*

This is another principle for grouping edge points which can be used to filter out noisy edges irrespective of their strength and recover lines which consist of edge points having the same edge magnitude. The probability of recovering a significant line increases with the similarity in edges within a small neighbourhood on that line.

### *Similarity in Edge Pixel Intensity*

The grouping of edge points based on similarity of intensity uses consistency in intensity rather than contrast. Theoretically, it should prevent the line segment from crossing over between the background and foreground sides of the boundaries. Edge orientation can also be used to do the same thing. Generally, the groups of edge points based on edge intensity are not very accurate in terms of localisation.

The grouping based on similarity in edge orientation, magnitude and edge pixel intensity are all supported by the *non-accidental* argument. There is a very little chance that the edge points due to noise and other artefacts would have similar orientation, magnitude, and intensity values.

## **3.5 Line Segment Extraction: Implementation Details**

The two main steps in extracting line segments are edge point detection and local perceptual grouping of edge points into straight lines in a small image window.

### **3.5.1 Edge Point Detection**

The published work on edge detection is so extensive that it is difficult to discuss all the well known techniques here. Abdou [1978] and Blicher [1984] have surveyed and evaluated in detail the past and current edge detection methods. Most of them are

based on detecting ideal step edges, while in the real world scenes, step edges are a small percentage. Three main requirements of the edge detector have been pointed out and these are:

- Good detection; which needs a good sensitivity criterion. In other words, good detection means, maximising signal to noise ratio.

- Good localisation of edges.

- Single response of a single edge.

An additional requirement is the computation of accurate edge orientation for weak (low contrast) edges. Since similar orientation is an important perceptual criterion used in our method, accurate edge direction computation is very important for the success of the perceptual grouping process.

The simplest way of extracting edge points is by using masks to calculate gradient magnitude and direction. Basically edge detection is an ill-posed problem. Large edge masks, which are larger than the image features, smooth the image and sometimes remove the relevant image features completely. The smaller size masks identify multiple edges for larger features. To overcome these difficulties, Marr and Hildreth [1980] used multiple size operators. They also tried to regularise edge detection by blurring the image with a gaussian filter in their DOG operator. In addition to that they used the zero crossings of the second directional derivative to detect edges, rather than relying on gradient magnitude. They have settled for the orientation independent Laplacian operator ( $\nabla^2 G$ ) by assuming a linear variation. The main problem with the multiple size operators is that their size should match with the image events and generally before processing the image one may not know the image events. Moreover, Marr and Hildreth do not provide any clue as to how to combine and group the outputs of different size operators.

Ridge-valley and step edge detectors reported by Haralick et al. [1983, 1984] have been tried on the endoscopic images in this study. These operators detect events in the image by fitting a cubic surface over a pre-defined window size. They were tried because of the existence of ridge like muscles in the colon images. The performance of these detectors on endoscopic images was poor due to excessive noise. The output from ridge-valley detector was sparse and missed useful and significant structures.

Although we have the choice of using any one of the well accepted edge detectors (e.g. the Canny, Marr and Hildreth, or Haralick Operators), we have chosen a simple edge detector to demonstrate the capabilities of grouping edge points perceptually. The simple edge operators have the advantage that they are easy to implement and can perform edge detection at video rate on most of the existing image processing hardware systems (e.g. Imaging Technology Series 151 Image Processor). The basic 3×3 edge operator is defined by two masks which compute the edge values in the horizontal and vertical directions. Supposing that the operator is centred at a pixel location (x, y) in the image and the pixel values in the surrounding 3×3 window are denoted by:

$$\begin{array}{ccc} I_{(x-1,y-1)}, & I_{(x,y-1)}, & I_{(x+1,y-1)} \\ I_{(x-1,y)}, & I_{(x,y)}, & I_{(x+1,y)} \\ I_{(x-1,y+1)}, & I_{(x,y+1)}, & I_{(x+1,y+1)} \end{array}$$

as shown in Figure 3.5.

The edge magnitude in the horizontal and vertical directions (termed as  $E_h$  and  $E_v$ ) at the pixel (x, y) is defined as:

$$\begin{aligned} E_h = & \\ & I_{(x+1,y-1)} - I_{(x-1,y-1)} + \\ & K \times I_{(x+1,y)} - K \times I_{(x-1,y)} + \\ & I_{(x+1,y+1)} - I_{(x-1,y+1)}. \end{aligned} \tag{3.2}$$

$$E_v = \begin{aligned} & [I_{(x-1,y-1)} + K \times I_{(x,y-1)} + I_{(x+1,y-1)}] - \\ & [I_{(x-1,y+1)} + K \times I_{(x,y+1)} + I_{(x+1,y+1)}]. \end{aligned} \quad (3.3)$$

$I_{(x-1,y-1)}$	$I_{(x,y-1)}$	$I_{(x+1,y-1)}$
$I_{(x-1,y)}$	$I_{(x,y)}$	$I_{(x+1,y)}$
$I_{(x-1,y+1)}$	$I_{(x,y+1)}$	$I_{(x+1,y+1)}$

*Figure 3.5: A 3×3 image window centred at (x, y).*

The coefficient  $K$  is different for different operators. In the case of Sobel operator [Duda and Hart 1973] its value is two, while for Prewitt [1970] and Isotropic operator the values are one and  $\sqrt{2}$  respectively.

The directional accuracy of the isotropic operator is the best, but due to the floating point computation, the Sobel operator is often preferred. Moreover extensive studies for correcting the edge magnitude and orientation have been carried out and their results can be used to achieve accurate edge orientation [Abdou 1978, Kittler 1983]. For the Sobel edge operator, the edge magnitude,  $E$  in terms of horizontal and vertical edge values described in equations (3.2) and (3.3) is:

$$E = \sqrt{(E_h^2 + E_v^2)} \quad (3.4)$$

and the edge orientation,  $\alpha$  is:

$$\alpha = \tan^{-1}\{E_v/E_h\}$$

when  $E_v/E_h$  value is within the interval  $[0 \leq E_v/E_h \leq 1/3]$

and

$$\alpha = \tan^{-1}\{(3E_h - 11E_v - \sqrt{(112E_v^2 + 16E_h^2 - 64E_hE_v)})/(-7E_h - 9E_v)\}$$

when  $E_v/E_h$  value is within the interval  $[1/3 < E_v/E_h \leq 1]$

(3.5)

A major difference in our approach from earlier work is that we retain the low contrast edges, which may be a part of the perceptually significant line, rather than removing them with an arbitrary threshold. Only a few edge points, below a very low threshold, are removed since their orientation would not be sufficiently accurate.

### 3.5.2 The Grouping Process

After forming an edge point representation of the image, the next step is to aggregate the individual points into short line segments. The most straight forward method for line extraction is least-square fitting. This method is effective and feasible when firstly, the edge point data is free of noise and secondly, the edge points which belong to a particular line have been identified. However, in many cases, including ours, these criteria are not met. Moreover our interest is not just in fitting straight line segments but in performing perceptual grouping along with the noise filtering process.

The edge point data is processed to extract line segments through two different size window channels. The contour extraction algorithm, which will be described in the next chapter uses a multi-level line representation in a pyramid to link these line segments. The image is divided into pre-selected size square windows ( $6 \times 6$  and  $12 \times 12$  are used in this particular implementation) and a local grouping process for straight line extraction is applied to each square. The windows are overlapped by one third of their size. For example the  $12 \times 12$  square window is overlapped by four pixels in the horizontal and vertical directions. In this way it is assumed that the line extracted in  $12 \times 12$  window belongs to an  $8 \times 8$  square image centred at the middle of the  $12 \times 12$  window. Each local grouping process is independent of the others, and therefore it can be implemented in parallel. The main assumption is that in any square window there can only be one significant line segment. The following are the different steps involved, in the grouping process.

*Step 1:*

Apply the modified Hough transform for straight line extraction introduced by Wallace [1985], to each square window of the image. For each candidate line, the co-ordinates of its voting edge points are stored in addition to the total number of votes for that line.

*Step 2:*

From the candidate lines found, select at the most  $L_{\max}$  lines. Those lines are selected which are best (in the sense of number of votes) from all the candidates and whose votes are larger than a minimum voting threshold  $V_{\min}$ .

*Step 3:*

For each candidate line selected in Step 2, repeat the following grouping processes.

*Proximity P-Grouping:* Apply proximity grouping on the basis of euclidean distance between each voting edge and identify the largest cluster of edge points.

The edge points which are not part of the largest cluster are dropped from the voting strength of the particular line. In this way the initial filtering of edge points is performed. If the number of edge points in the largest cluster drops below the threshold for minimum votes  $V_{\min}$ , it is assumed that the selected line segment is not significant and further grouping is aborted for that particular line. It is not necessary to have a separate test for proximity which is established by the connectivity grouping described next.

*Connectivity C-Grouping:* In this particular implementation, the connectivity of edge points is ensured by not allowing any gap between adjacent points on the line. In the case of a small window the value of the minimum voting threshold  $V_{\min}$ , proximity grouping, and connectivity are tied to each other. If  $V_{\min}$  is approximately taken as the same as the window size then proximity and connectivity grouping are established by the same criterion.

*Orientation O-Grouping:* Grouping based on similarity in edge orientation is only applied on the largest cluster of connected edge points. The slow drift in orientation is allowed by computing the difference of orientation for each pair of adjacent edge points. This is equivalent to differentiating orientation with respect to distance or calculating curvature along the curved line segment. The edges are partitioned in such a way that each partition consists of those edge points whose orientation change is within a tolerance  $\delta\theta$ . The edge count for the largest partition is now taken as the voting strength of that line and is compared with  $V_{\min}$  to test for the significance of line. If the count is less

than  $V_{\min}$ , then the selected line does not qualify as a significant line based on orientation grouping.

*Contrast and Intensity CI-Grouping:* Group connected edge points for a line on the bases of similar edge magnitude and edge pixel intensity. The edge points whose edge magnitude and intensity are within their respective tolerances are clustered together. The edge magnitude tolerance  $\delta M$ , compensates for errors in edge strength computation while intensity tolerance  $\delta I$ , helps to overcome errors due to quantisation in grey levels. Similar to the orientation grouping process, edge points in the largest cluster are qualified to vote for a particular line and their count determines the significance of line.

If the selected line fails to qualify in both orientation and CI-Grouping, it is assumed that the line is not significant enough to participate in contour extraction

*Step 4:*

If none of the selected lines qualify as a significant line in the grouping processes of Step 3 then no useful line segment is present in that window. When more than one line qualifies in one or both grouping processes, decide as following:

-Preference is given to orientation grouping and if more than one line qualifies in O-Grouping, the line with maximum number of votes is selected. The preference can also be given to the line whose direction is normal to the edge orientation.

-If orientation grouping fails to identify a line in a given window then the lines are tested solely on the basis of CI-Grouping. We select the line with the largest number of votes.

The above steps only describe the method in principle. The algorithm has been implemented in a more efficient way. The CI-Grouping is only required when the O-Grouping fails. Alternatively, the above procedures could be implemented using parallel processing for the individual grouping criteria. In our implementation on endoscopic images only 10% of the useful lines are selected on the basis of CI-Grouping because O-Grouping is more stable and it identifies most of the significant line segments.

### 3.5.3 Selection of Thresholds

The grouping algorithm described above uses a number of thresholds at various steps and the reader may wonder about the selection of these thresholds and their sensitivity. However, the algorithm is robust to these choices and works well on widely different endoscope images without changing thresholds. These thresholds have been optimised by experimental results on a large number of endoscopic images. The thresholds may require some fine tuning to enhance the performance of the algorithm for a particular class of images. The effects and the bases for these choices are discussed below.

Thresholding of edge points on the basis of their strength is also known as amplitude thresholding. The amplitude of response from an edge operator is a function of the magnitude of the edge, its orientation, and its distance from the centre of the edge operator. Edge removal on the basis of edge strength also removes the low amplitude, perceptually significant edges. It is very difficult to rely solely on this threshold for filtering noisy edges. The noise amplitude will not be constant throughout the image due to changes in illumination condition and contrast. In our algorithm the edge threshold is intended only to remove low accuracy edge points. It can therefore be set for the whole image. The values that were used were six to ten for computer generated images (depending on the amount of noise) and around eighteen for endoscopic images. These proved low enough to keep all the significant edges but removed those edges whose orientation may not be accurate.

The choice of maximum number of lines,  $L_{\max}$  retained from the original Hough transform, and the minimum number of votes,  $V_{\min}$  required for each line, both depend on the selection of image window size and the amount of maximum background noise which may be present in the image. In this algorithm, the maximum number of lines selected is taken to be equal to the image window side. The minimum number of votes,  $V_{\min}$  is also affected by the amount of overlapping of windows and the tight control on grouping principles. The minimum votes for qualifying as a significant line is fixed at the window side minus the overlap size (eight and four for  $12 \times 12$  and  $6 \times 6$  windows respectively).

The tolerance threshold for orientation  $\delta\theta$  determines the limits on the curvature of contours to be extracted. In the case of theta-aggregation it also depends on the psychology of vision. The orientation tolerance is easy to estimate from the psychological studies which demonstrate how much tolerance humans allow in orientation when they group similar oriented patterns. The data organisation in visual cortex can also be used as a clue in the selection of orientation tolerance. The cortical table, described in the previous chapter, shows that the maximum orientation resolution for line segments is  $10^\circ$ . Our choice of  $5^\circ$  tolerance for endoscope images and  $15^\circ$  for computer generated images is influenced by all of these factors and has achieved good results.

The grouping thresholds  $\delta M$  and  $\delta I$  for edge magnitude and edge pixel intensity produce stable results over a wide range of scene illumination. The stability improves when both grouping criteria are employed together. When employed individually, the pixel intensity threshold is affected considerably by the errors in edge localisation while for edge contrast a small variation changes the results considerably. For both  $\delta M$  and  $\delta I$  the tolerance is fixed at three. These thresholds are also affected by the errors due to sampling and the quantisation processes.

### 3.6 Experimental Results and Conclusions

In many medical applications of image analysis, noise is present in the digitised images. Typically, pictures of internal organs obtained from radiological images or otherwise, have a great degree of noise. This means that to obtain meaningful results, segmentation algorithms must be tailored for the application. Although the line extraction algorithm was developed primarily as part of a contour detector for endoscopic colon images, it is expected that it will prove to be of much wider applicability. The method can be applied to the general problem of line segment extraction in images containing a variety of unknown noise.

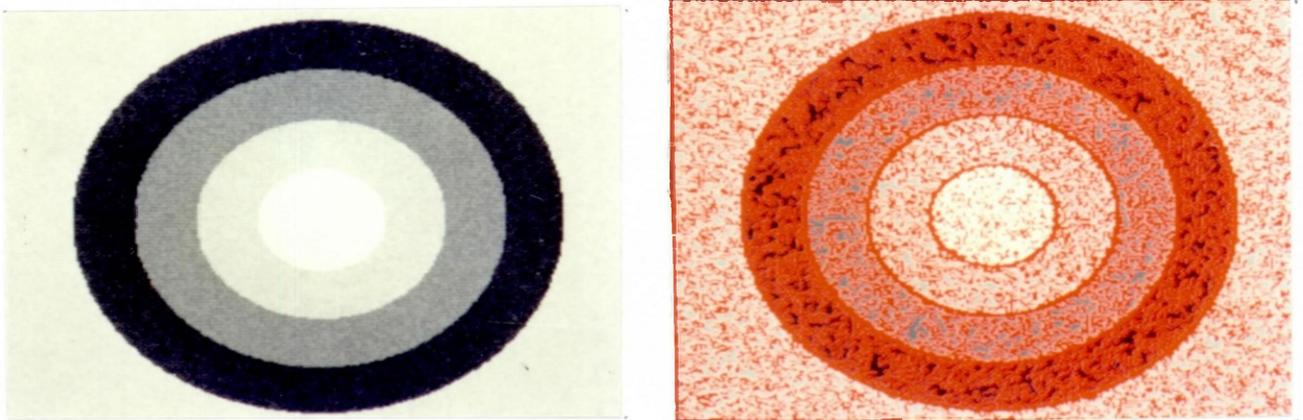
#### 3.6.1 Artificially Generated Images with Added Noise

The performance of the technique was analysed by testing it on some computer generated images with known amount of random noise. The results on one of the images with varying amount of noise are presented here. The image consists of curved elliptical contours which are well apart. The random noise was added at each pixel of the image amounting to  $\pm 10\%$  for one test image and  $\pm 22\%$  (of the average signal level) for the other. Both of these images along with their Sobel edge detector outputs are shown in Figure 3.6 and 3.7. The image in Figure 3.6 contains  $\pm 10\%$  random noise while the image in Figure 3.7 contains  $\pm 22\%$  random noise. The outputs for the individual grouping operations are presented for two resolutions (based on  $12 \times 12$  and  $6 \times 6$  images windows) in Figures 3.8 to 3.11 for demonstrating the capabilities of perceptual grouping based on connectivity, orientation drift and theta-aggregation, similarity in edge pixel intensity and edge contrast.

It can be concluded from these results that for random noise of up to  $\pm 10\%$ , the grouping based on connectivity filters out most of the noise but for  $\pm 22\%$  noise, the output also contains a considerable number of line segments due to noise. The orientation grouping performs comfortably well for a large range of noise, although its performance deteriorates with the increase in noise. This problem has been overcome in the results presented here by increasing the

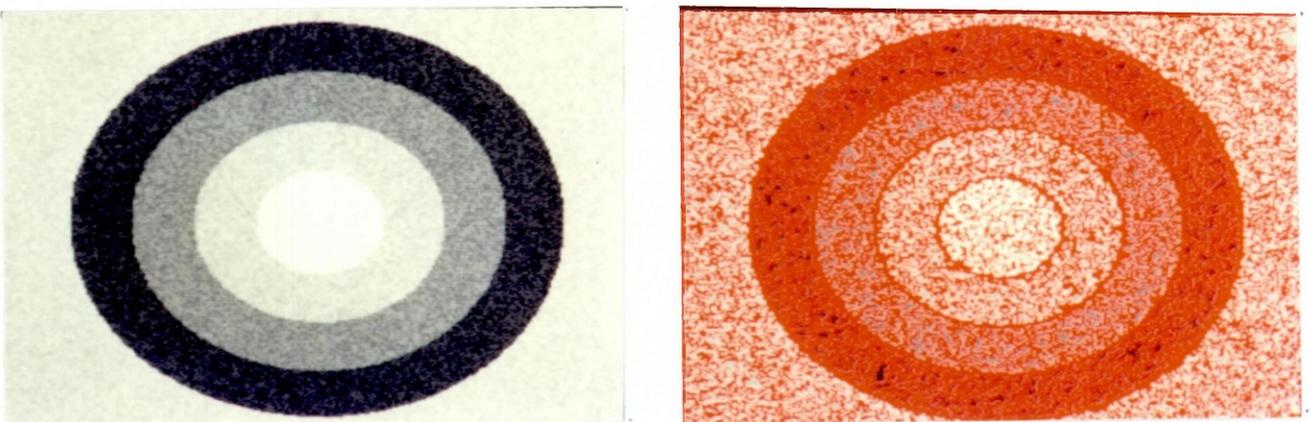
edge magnitude threshold from six to ten for  $\pm 22\%$  noise. The edge point grouping based on edge pixel intensity and edge contrast provides better results for  $\pm 10\%$  random noise but the overall results are poor for larger noise and their performance is not comparable to the orientation or connectivity grouping.

Line extraction based on the O'Gorman and Clowes method has also been implemented and applied to the same two images containing  $\pm 10\%$  and  $\pm 22\%$  random noise. Their method only groups those edge points whose orientation is perpendicular to the line segment direction (within a tolerance). The line segments of curved contours are difficult to extract in this way and a high tolerance between the line normal and edge orientation is required for extracting all the useful line segments, which may in turn produce line segments due to noise. The method is optimised by increasing the tolerance between the edge orientation and the line normal direction from  $\pm 5^\circ$  to  $\pm 25^\circ$ . The maximum tolerance is used beyond which the line segments due to noise start appearing in the output. Figure 3.12 shows the optimum best results for line segments extracted by O'Gorman and Clowes method at two resolutions. Their method misses more than 50% of the useful contour segments compared to our orientation grouping as given in Table 3.1. The amount of noise filtering achieved by our perceptual grouping technique can be estimated by comparing its output line segments with the total number of segments extracted without applying perceptual grouping. In Figure 3.13 the line segments extracted by employing perceptual grouping for two resolutions are shown while Figure 3.14 contains the line segments detected without applying perceptual grouping. The reduction in the line segments extracted by using perceptual grouping is considerable from 40 to 75% depending on the amount of noise and without losing too many useful line structures. Table 3.1 summarises the total number of line segments extracted by different grouping processes.

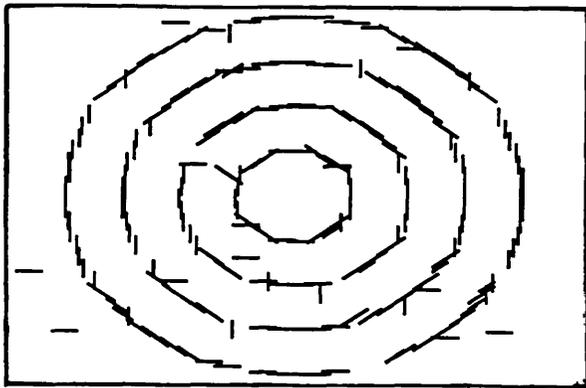


*Figure 3.6: An artificial image having  $\pm 10\%$  added noise and the output of Sobel edge detector with thresholding.*

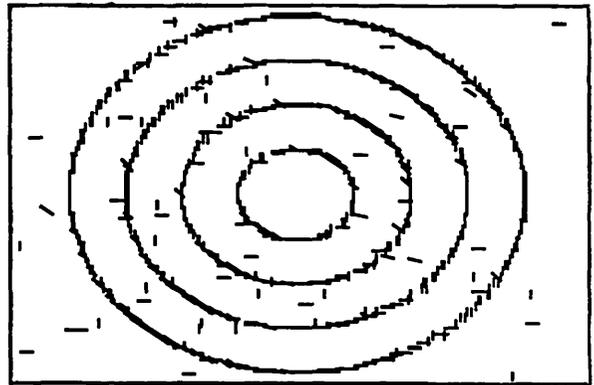
*In this and subsequent figures the edge points above a threshold are shown in red.*



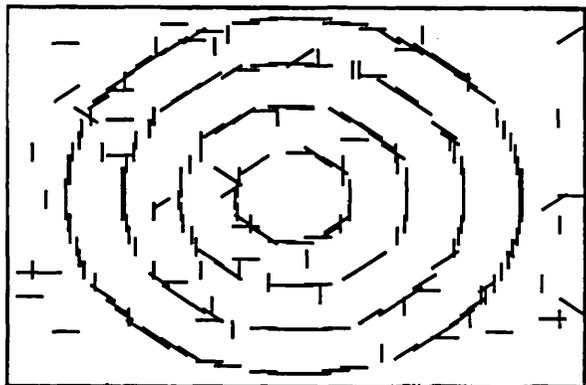
*Figure 3.7: An artificial image having  $\pm 22\%$  added noise and the output of Sobel edge detector with thresholding.*



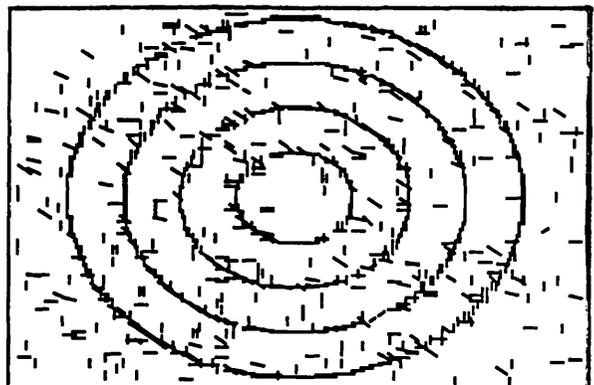
(a)



(b)



(c)



(d)

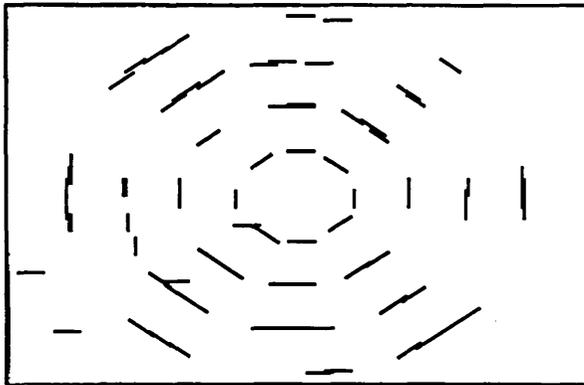
*Figure 3.8: Line segments extracted by employing connectivity grouping.*

*(a) For image containing  $\pm 10\%$  noise and based on  $8 \times 8$  window.*

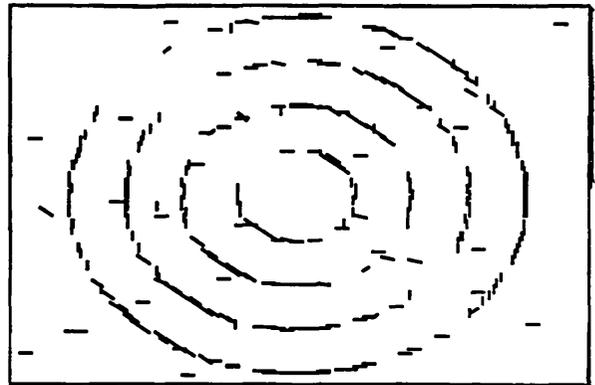
*(b) For image containing  $\pm 10\%$  noise and based on  $4 \times 4$  window.*

*(c) For image containing  $\pm 22\%$  noise and based on  $8 \times 8$  window.*

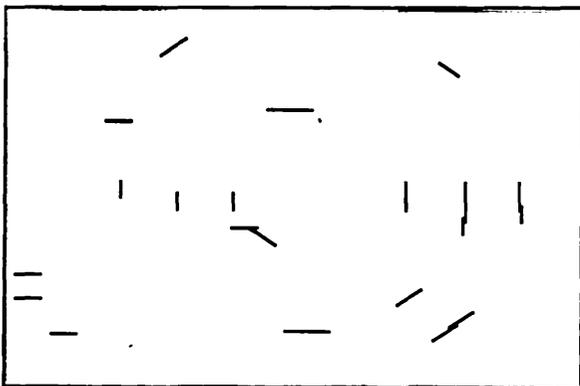
*(d) For image containing  $\pm 22\%$  noise and based on  $4 \times 4$  window.*



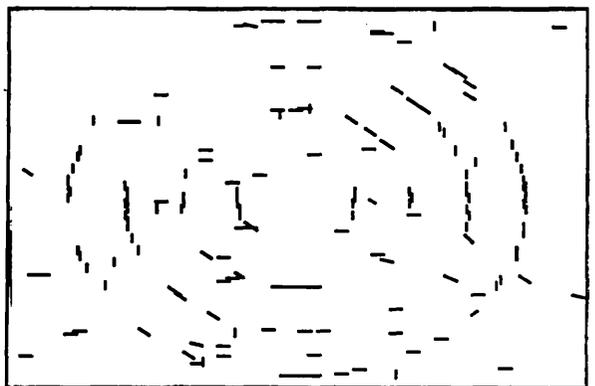
(a)



(b)



(c)



(d)

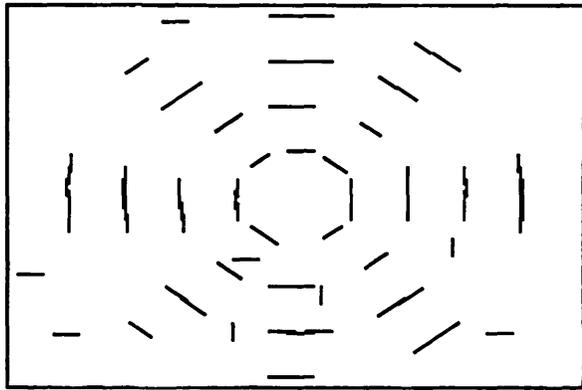
*Figure 3.9: Line segments extracted by employing grouping based on similarity in edge pixel intensity.*

*(a) For image containing  $\pm 10\%$  noise and based on  $8 \times 8$  window.*

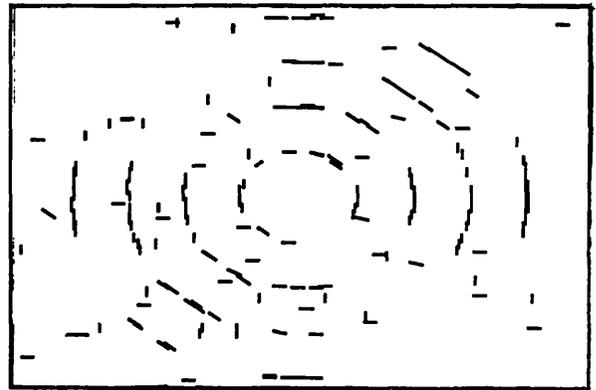
*(b) For image containing  $\pm 10\%$  noise and based on  $4 \times 4$  window.*

*(c) For image containing  $\pm 22\%$  noise and based on  $8 \times 8$  window.*

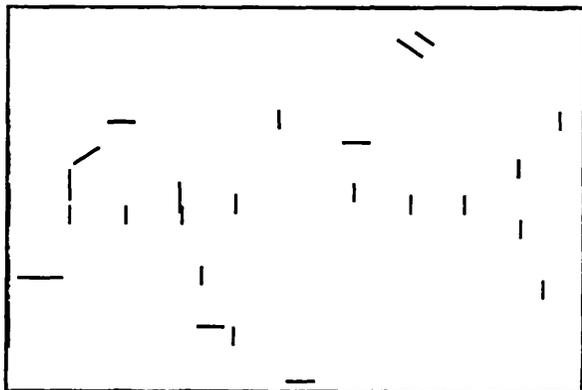
*(d) For image containing  $\pm 22\%$  noise and based on  $4 \times 4$  window.*



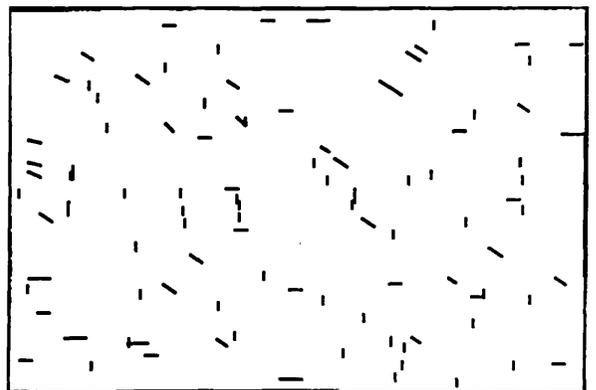
(a)



(b)



(c)



(d)

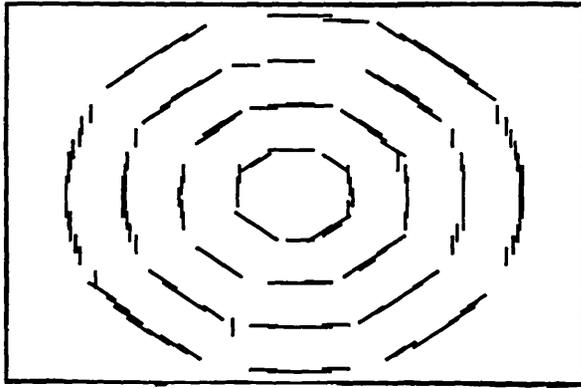
Figure 3.10: Line segments extracted by employing grouping based on similarity in edge contrast.

(a) For image containing  $\pm 10\%$  noise and based on  $8 \times 8$  window.

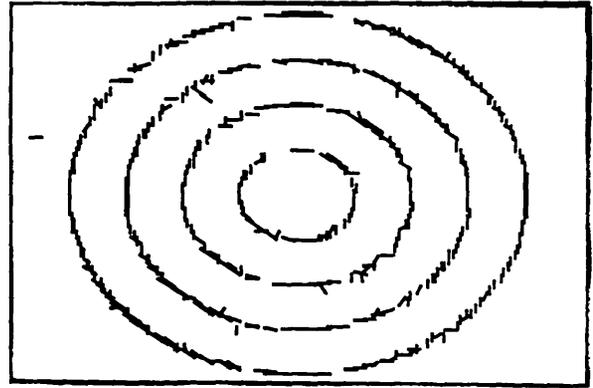
(b) For image containing  $\pm 10\%$  noise and based on  $4 \times 4$  window.

(c) For image containing  $\pm 22\%$  noise and based on  $8 \times 8$  window.

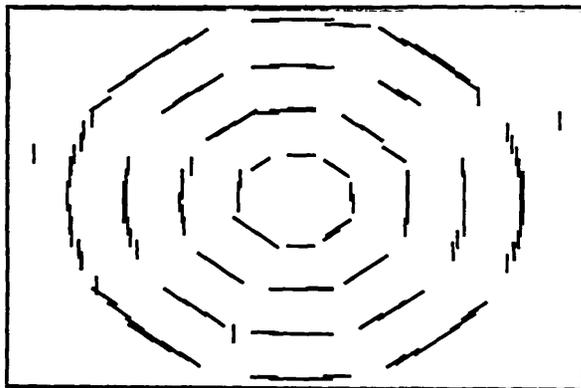
(d) For image containing  $\pm 22\%$  noise and based on  $4 \times 4$  window.



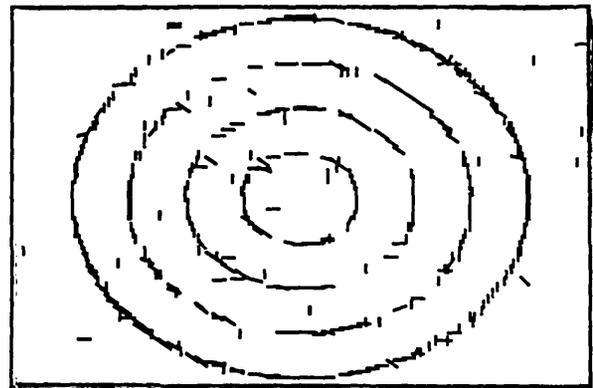
(a)



(b)

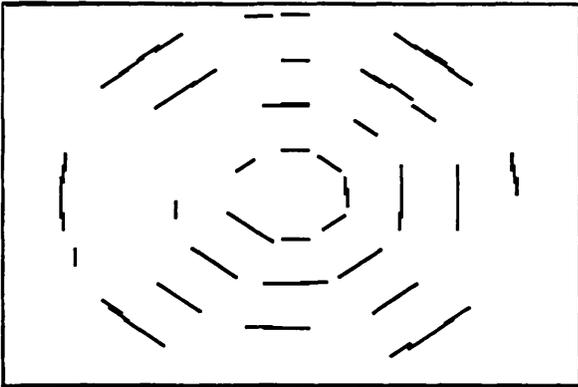


(c)

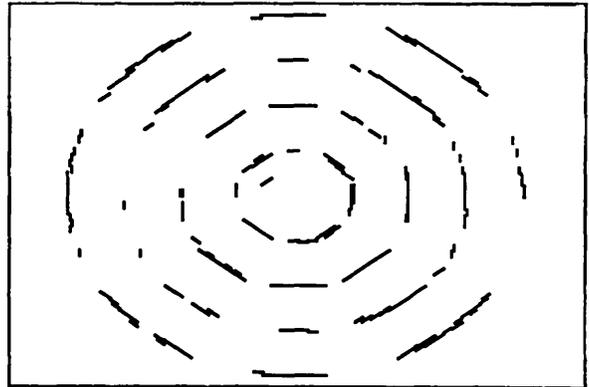


(d)

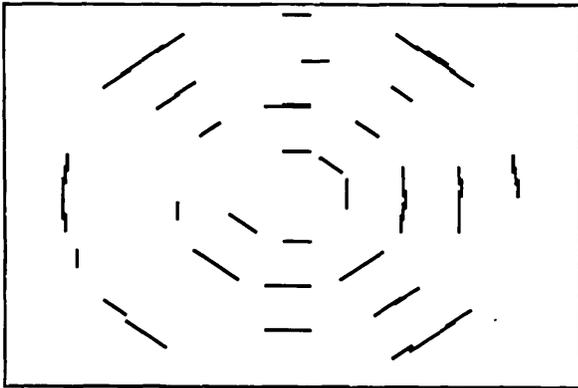
*Figure 3.11: Line segments extracted by employing grouping based on orientation drift and theta-aggregation.  
(a) For image containing  $\pm 10\%$  noise and based on  $8 \times 8$  window.  
(b) For image containing  $\pm 10\%$  noise and based on  $4 \times 4$  window.  
(c) For image containing  $\pm 22\%$  noise and based on  $8 \times 8$  window.  
(d) For image containing  $\pm 22\%$  noise and based on  $4 \times 4$  window.*



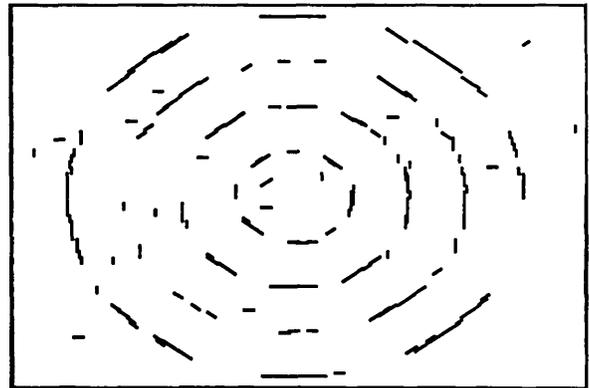
(a)



(b)



(c)



(d)

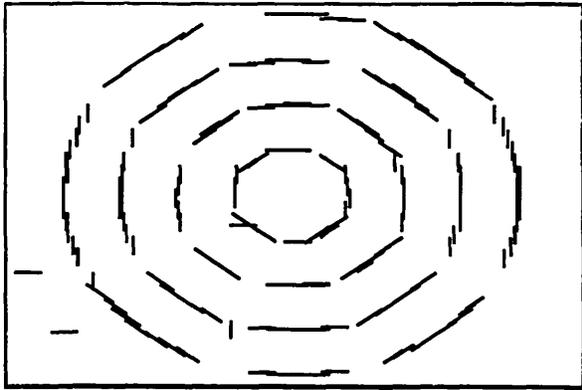
*Figure 3.12: Line segments extracted by employing O'Gorman and Clowes method.*

*(a) For image containing  $\pm 10\%$  noise and based on  $8 \times 8$  window.*

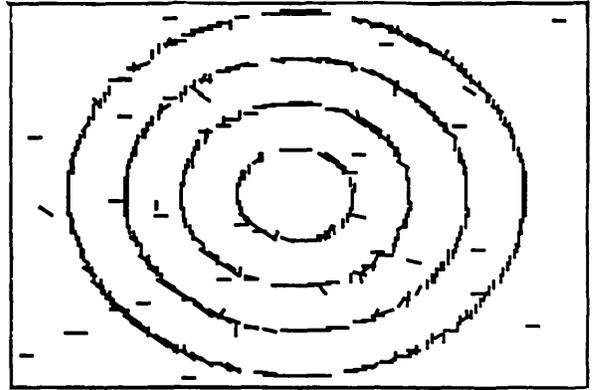
*(b) For image containing  $\pm 10\%$  noise and based on  $4 \times 4$  window.*

*(c) For image containing  $\pm 22\%$  noise and based on  $8 \times 8$  window.*

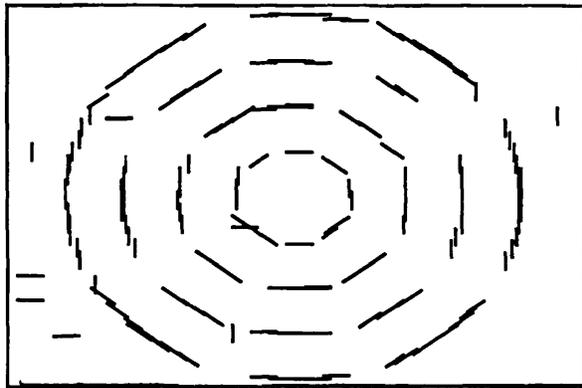
*(d) For image containing  $\pm 22\%$  noise and based on  $4 \times 4$  window.*



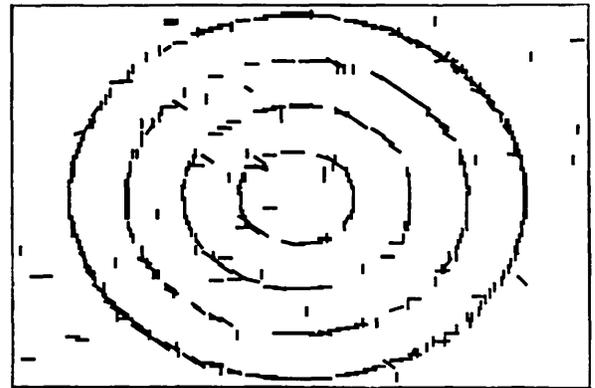
(a)



(b)



(c)



(d)

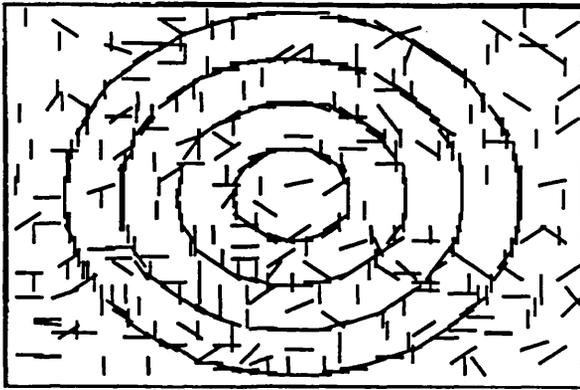
*Figure 3.13: Line segments extracted by employing perceptual grouping criteria.*

*(a) For image containing  $\pm 10\%$  noise and based on  $8 \times 8$  window.*

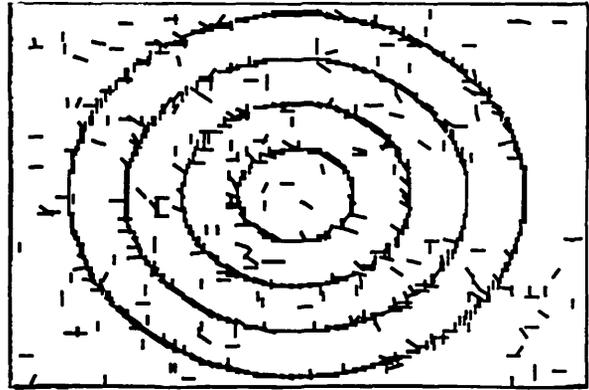
*(b) For image containing  $\pm 10\%$  noise and based on  $4 \times 4$  window.*

*(c) For image containing  $\pm 22\%$  noise and based on  $8 \times 8$  window.*

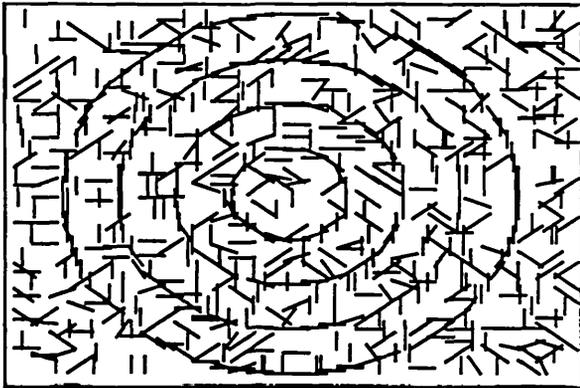
*(d) For image containing  $\pm 22\%$  noise and based on  $4 \times 4$  window.*



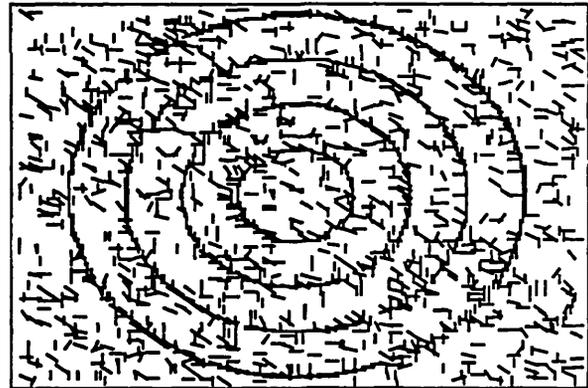
(a)



(b)



(c)



(d)

*Figure 3.14: Line segments extracted without applying perceptual grouping.*

*(a) For image containing  $\pm 10\%$  noise and based on  $8 \times 8$  window.*

*(b) For image containing  $\pm 10\%$  noise and based on  $4 \times 4$  window.*

*(c) For image containing  $\pm 22\%$  noise and based on  $8 \times 8$  window.*

*(d) For image containing  $\pm 22\%$  noise and based on  $4 \times 4$  window.*

Test Image Grouping Process	Figure 3.6		Figure 3.7	
	$\pm 10\%$ Noise 8 $\times$ 8	4 $\times$ 4	$\pm 22\%$ Noise 8 $\times$ 8	4 $\times$ 4
Line segments extracted without applying perceptual grouping principles.	481	898	721	1728
Line segments detected by using perceptual grouping.	183	535	164	468
Line Segments found by O'Gorman and Clowes method of collinearity grouping.	75	231	63	225
Orientation grouping with slow drift in orientation and theta-aggregation.	177	507	159	461
Connectivity grouping.	238	697	272	961
Grouping based on similarity in edge pixel intensity.	84	330	27	168
Similarity in edge contrast.	88	207	27	111

*Table 3.1: Output line segments for the artificial image with varying degree of noise and for different grouping processes.*

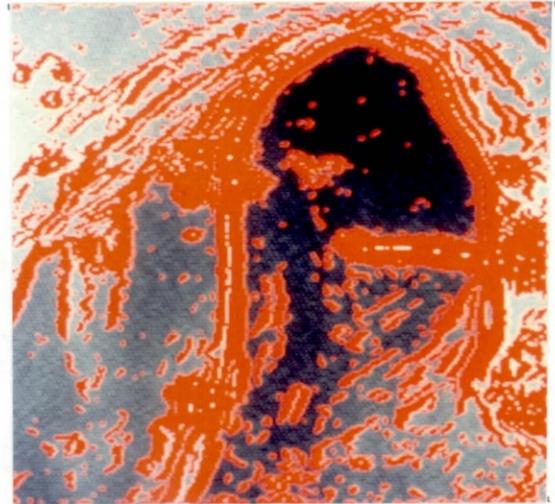
### 3.6.2 Endoscopic Colon Images

The line extraction method has also been tested for a large number of representative colon images. The images have been digitised at 256 grey levels from a video tape of colonoscopy procedures. There are a number of sources of different types of

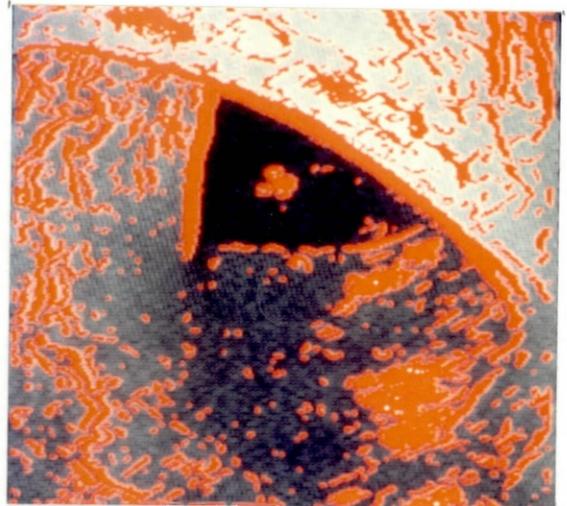
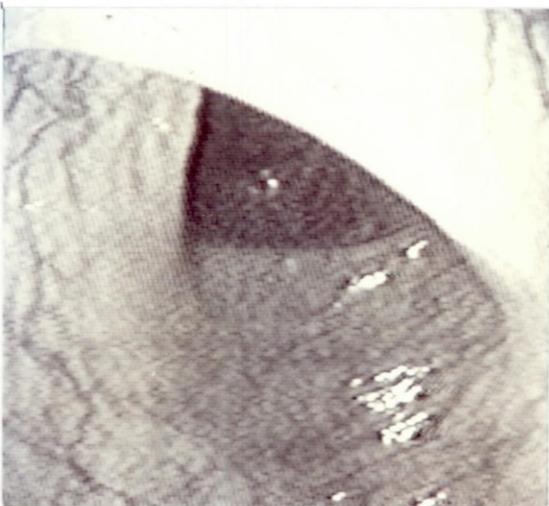
noise in endoscope images which include but are not limited to variable specular reflection, texture, and different types of matter in the colon. The magnetic media has also left some additional noise in the digitised images due to the process of recording. The results for three of these images are presented here for demonstration purposes.

The test pictures of the inner colon and the output of the Sobel edge detector for these images are shown in Figure 3.15 to 3.17. All the edge points with magnitude greater than eighteen are kept for further grouping. Our interest is only with the image contours due to inner muscles of colon while the images are littered with edges due to specularity, texture and noise. For comparison of the performance of perceptual grouping against ordinary edge linking, different results are presented in Figure 3.18 to 3.26. The line segments have been extracted at two resolutions which corresponds to an  $8 \times 8$  square and a  $4 \times 4$  square image windows. The line segments have also been detected without applying any perceptual grouping principles. The lines due to noise dominate this representation in both resolutions (see Figure 3.21b to 3.26b). The perceptual grouping principles described in our algorithm are then employed to identify relevant line structures. The results of Figures 3.21a to 3.26a demonstrate that most of the line segments due to noise are unable to qualify as significant line structures and are eliminated.

The performance of connectivity grouping, which filters out noisy edges due to random noise effectively in the artificial images, is not as effective on the endoscopic images (see Figure 3.18 to 3.20) because a large variety of noise is present in these images. The extraction of lines on the basis of connectivity and edge orientation is the most effective way of coping with noise. O'Gorman and Clowes method of collinearity grouping was also applied to the same colon images for the purposes of comparing its performance against the orientation grouping used in our algorithm. For the sake of providing same edge data to both algorithms, the connectivity grouping was also applied before using the O'Gorman and Clowes collinearity grouping. The method



*Figure 3.15: First colon image and the output of Sobel edge detector with thresholding.*



*Figure 3.16: Second colon image and the output of Sobel edge detector with thresholding.*

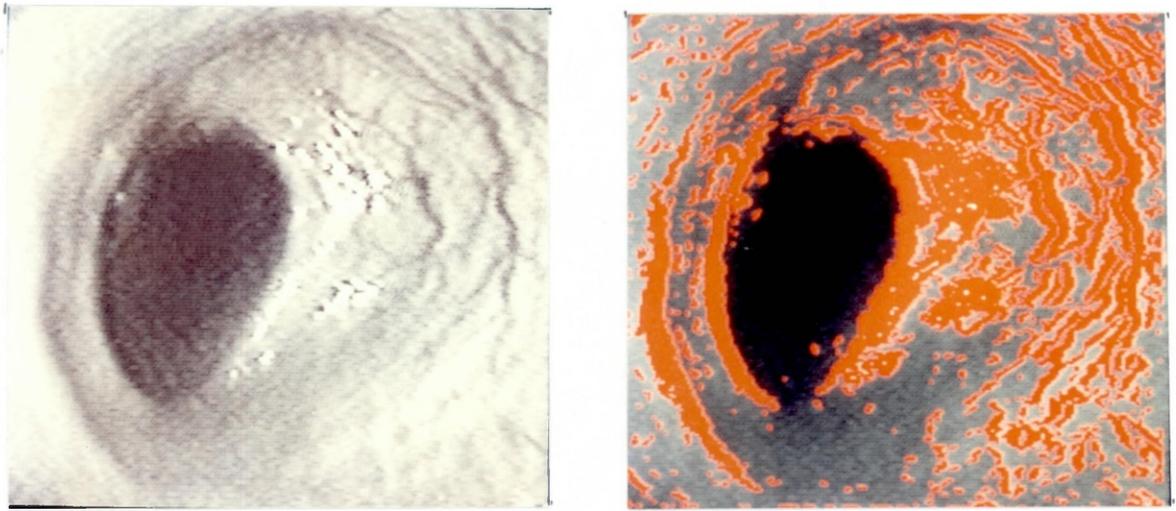


Figure 3.17: Third colon image and the output of Sobel edge detector with thresholding.

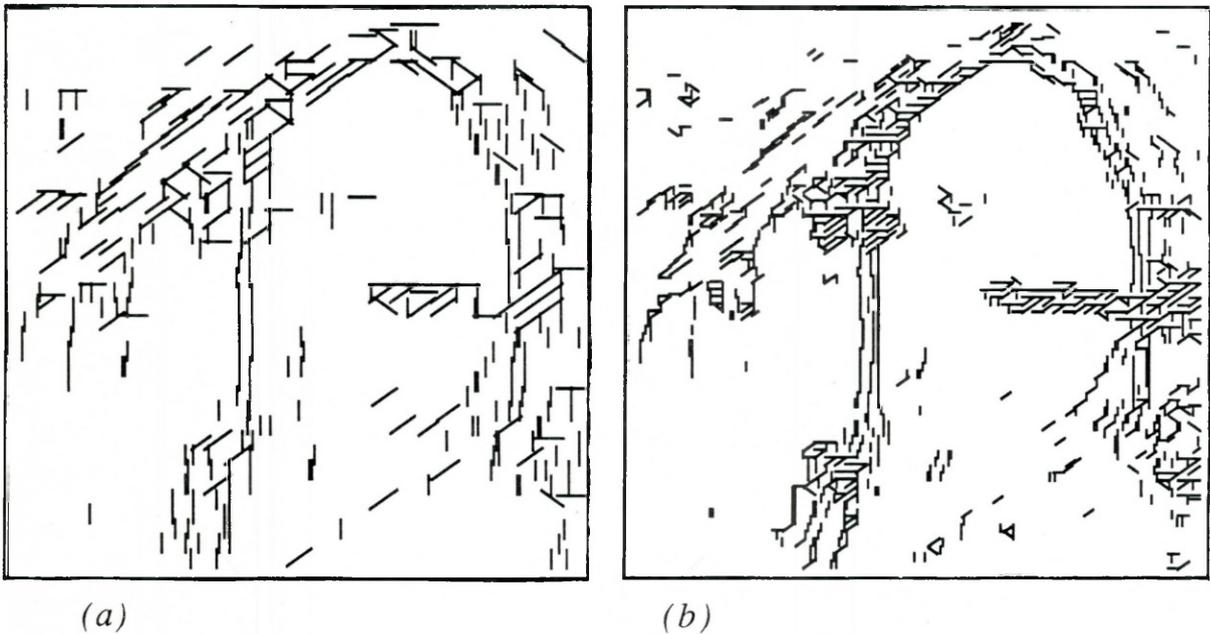
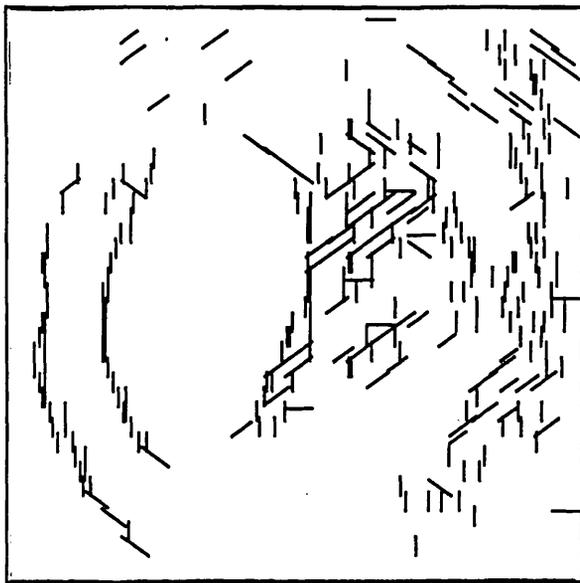


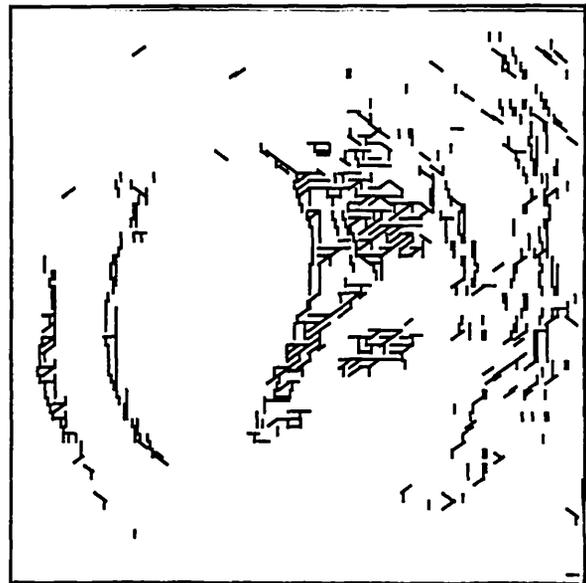
Figure 3.18: Line segments extracted using connectivity grouping for colon image of Figure 3.15.

(a) Based on  $8 \times 8$  image window.

(b) Based on  $4 \times 4$  image window.



(a)

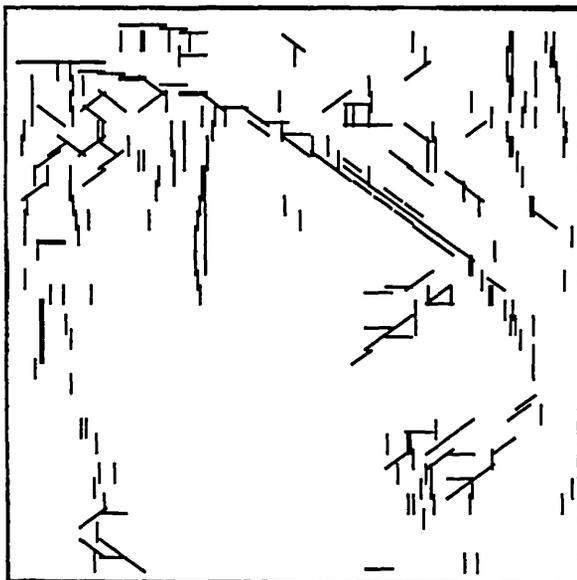


(b)

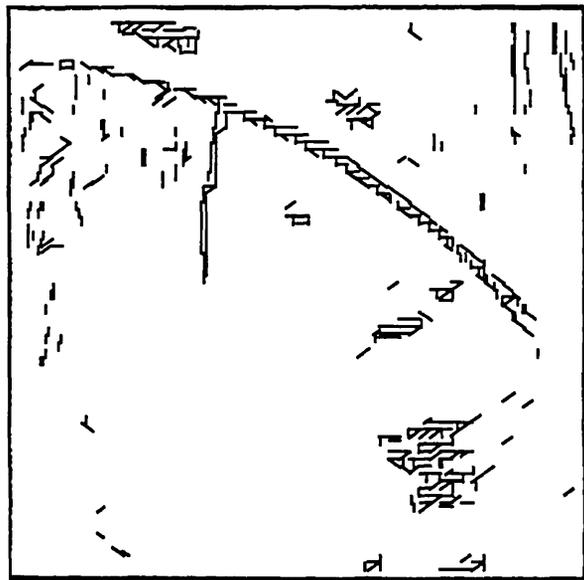
Figure 3.19: Line segments extracted using connectivity grouping for colon image of Figure 3.16.

(a) Based on  $8 \times 8$  image window.

(b) Based on  $4 \times 4$  image window.



(a)

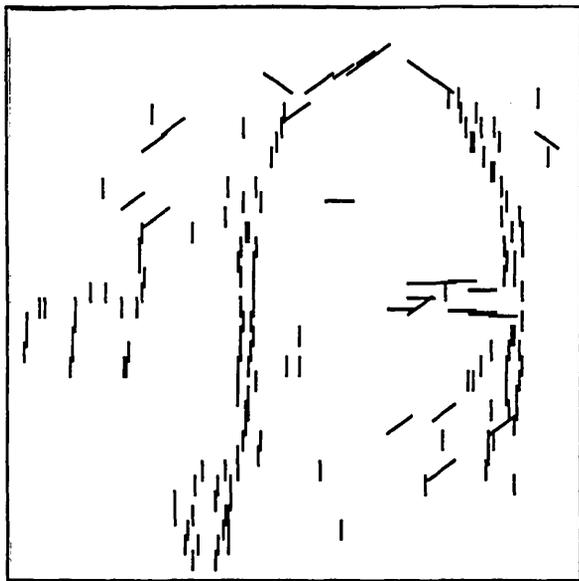


(b)

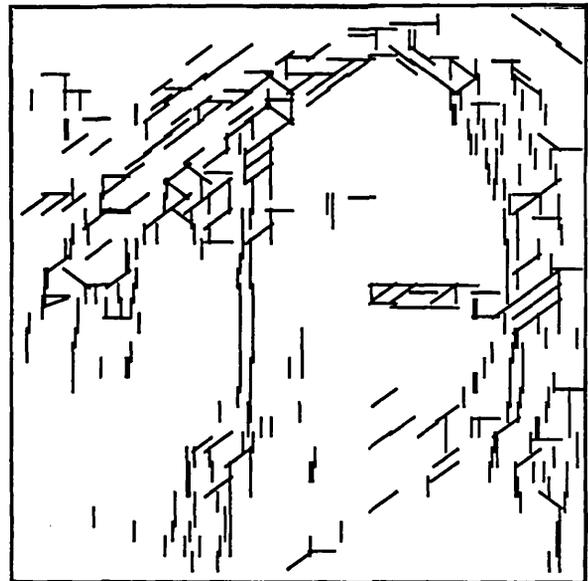
Figure 3.20: Line segments extracted using connectivity grouping for colon image of Figure 3.17.

(a) Based on  $8 \times 8$  image window.

(b) Based on  $4 \times 4$  image window.



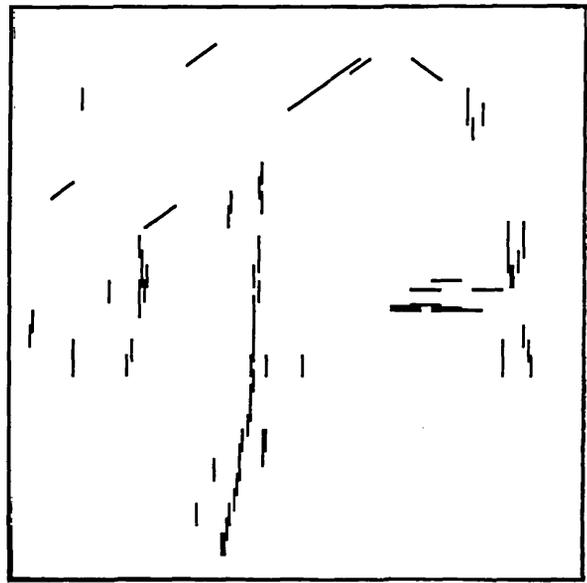
(a)



(b)



(c)



(d)

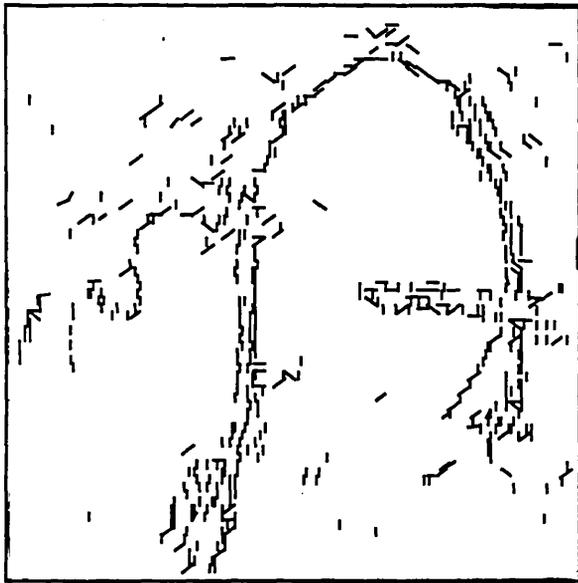
Figure 3.21: Output line segments for the colon image of Figure 3.15 on the basis of  $8 \times 8$  image window.

(a) When the perceptual grouping criteria are employed.

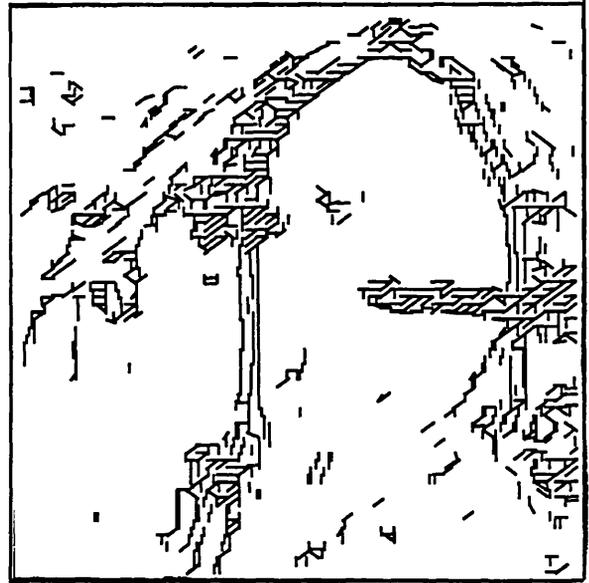
(b) Without using perceptual grouping.

(c) Grouping based on orientation drift and theta-aggregation.

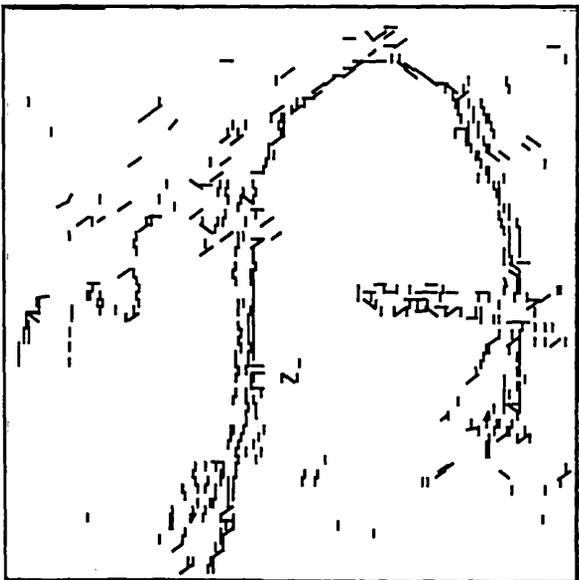
(d) O'Gorman and Clowes method.



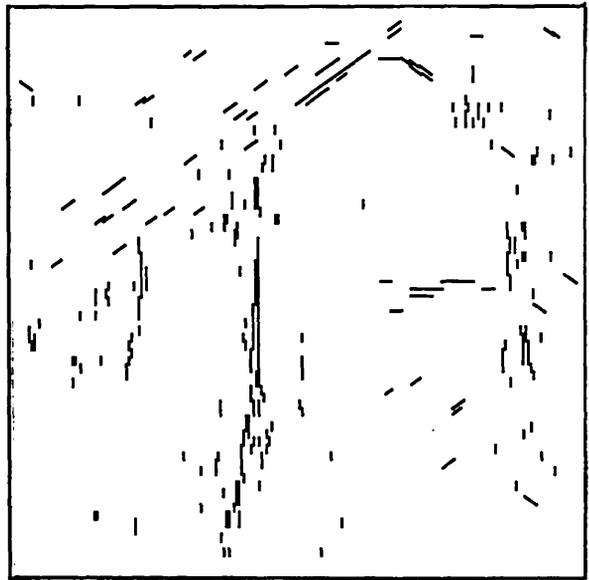
(a)



(b)



(c)



(d)

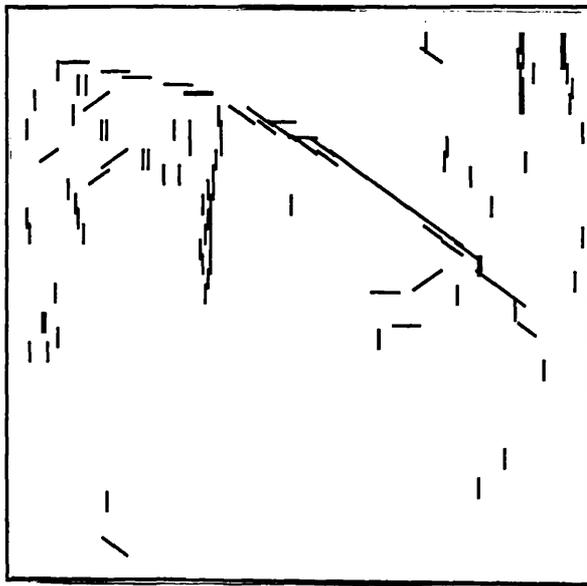
Figure 3.22: Output line segments for the colon image of Figure 3.15 on the basis of  $4 \times 4$  image window.

(a) When the perceptual grouping criteria are employed.

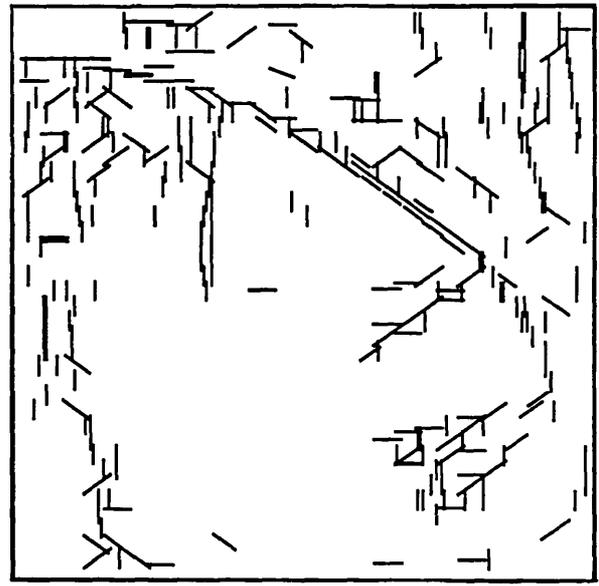
(b) Without using perceptual grouping.

(c) Grouping based on orientation drift and theta-aggregation.

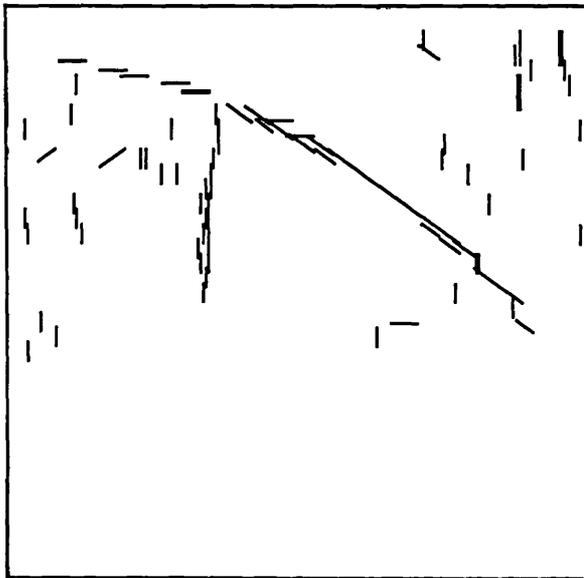
(d) O'Gorman and Clowes method.



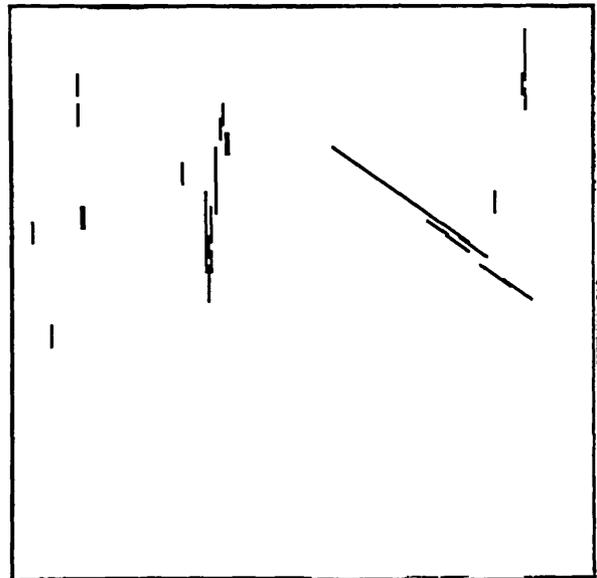
(a)



(b)



(c)



(d)

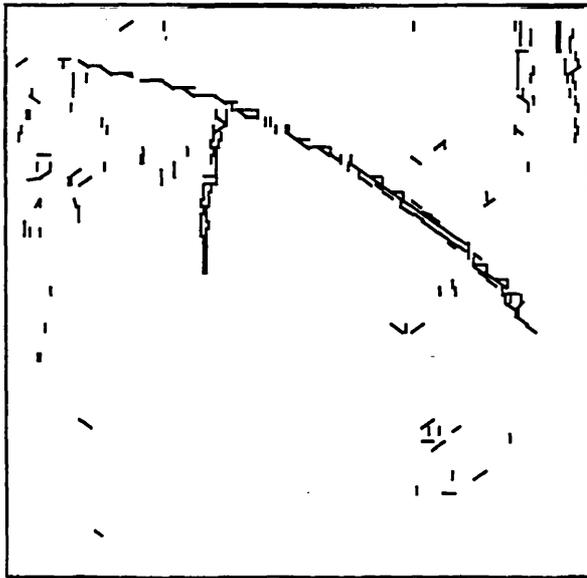
Figure 3.23: Output line segments for the colon image of Figure 3.16 on the basis of  $8 \times 8$  image window.

(a) When the perceptual grouping criteria are employed.

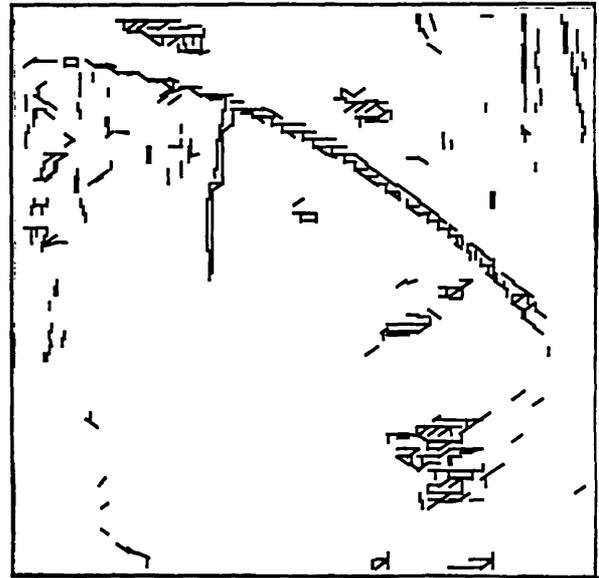
(b) Without using perceptual grouping.

(c) Grouping based on orientation drift and theta-aggregation.

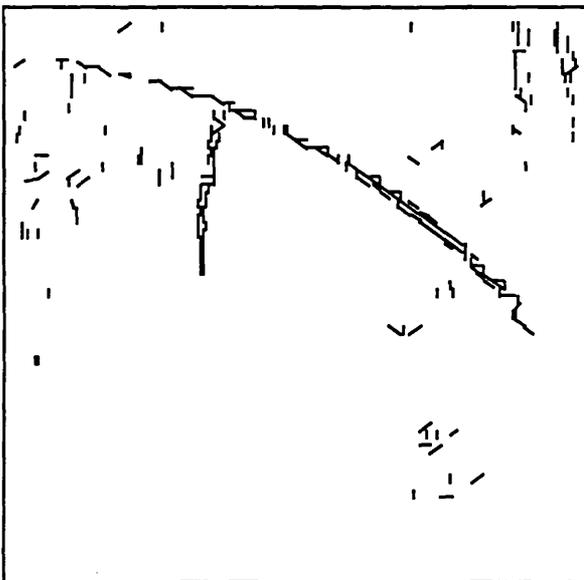
(d) O'Gorman and Clowes method.



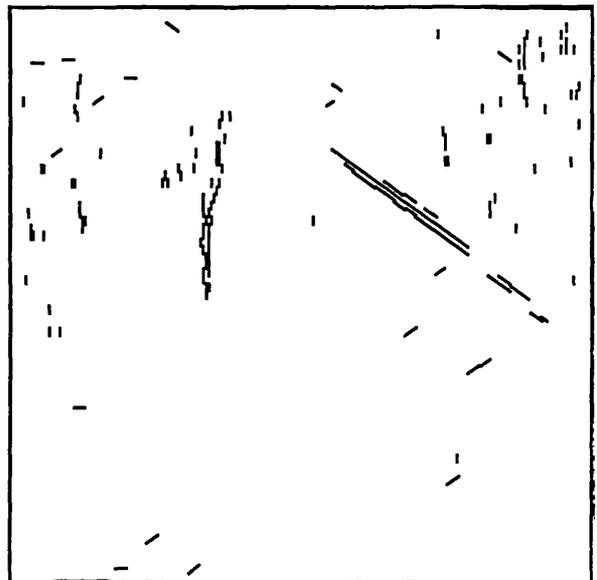
(a)



(b)



(c)



(d)

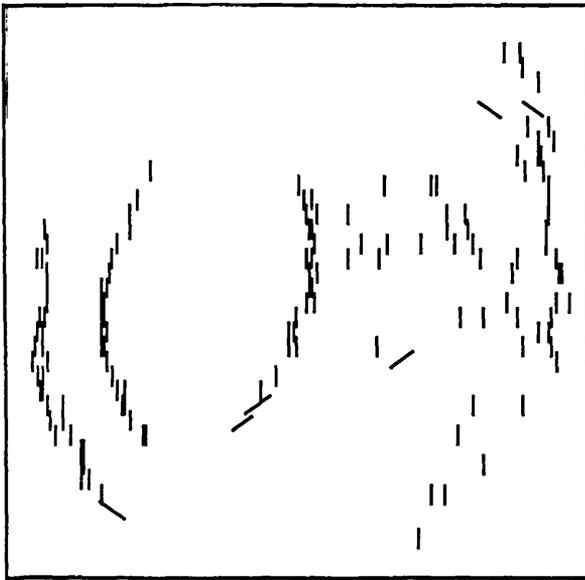
Figure 3.24: Output line segments for the colon image of Figure 3.16 on the basis of  $4 \times 4$  image window.

(a) When the perceptual grouping criteria are employed.

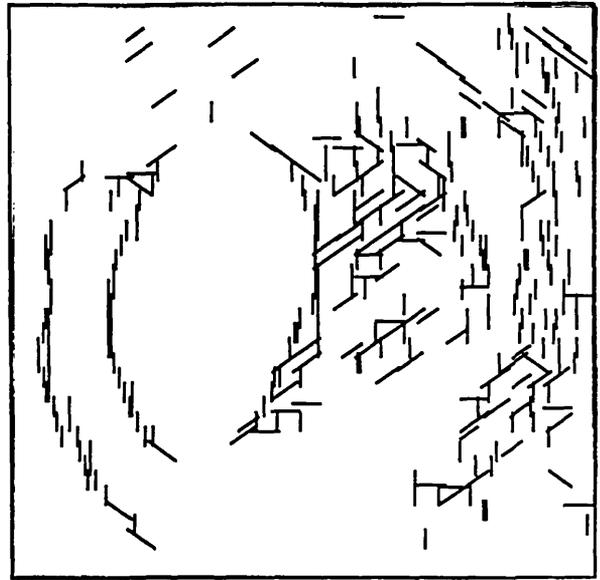
(b) Without using perceptual grouping.

(c) Grouping based on orientation drift and theta-aggregation.

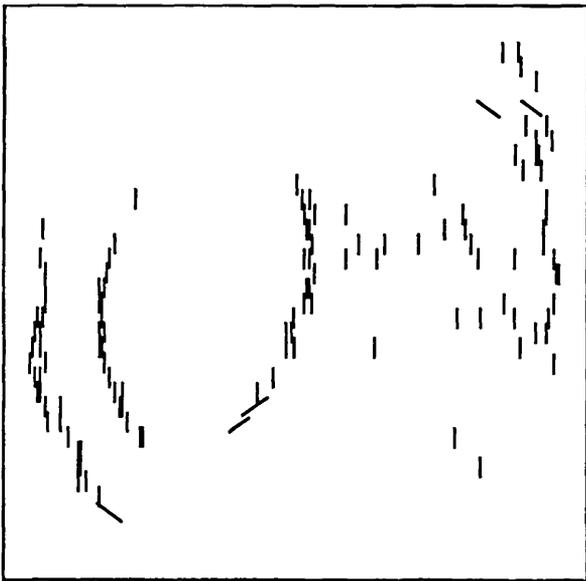
(d) O'Gorman and Clowes method.



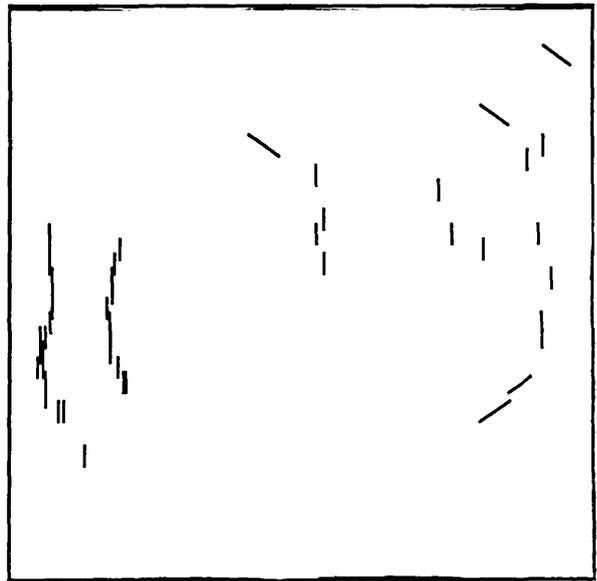
(a)



(b)



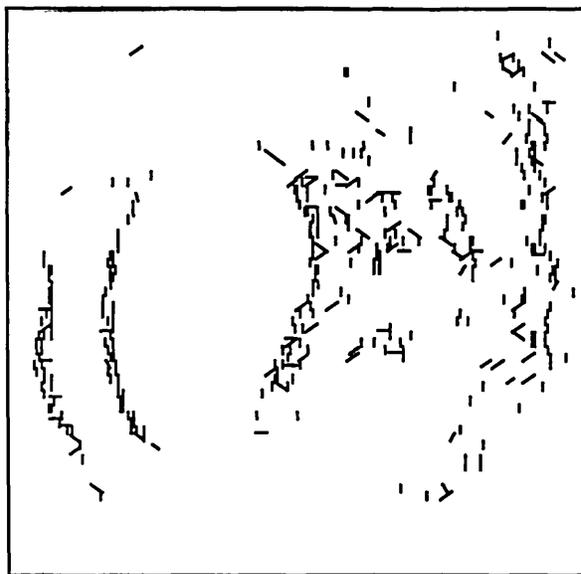
(c)



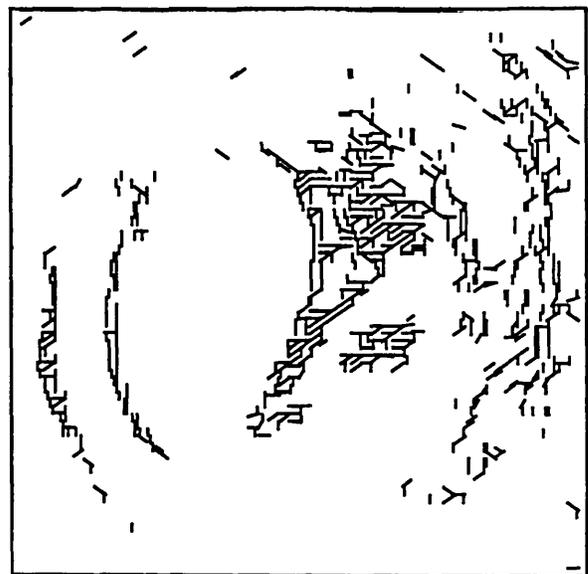
(d)

Figure 3.25: Output line segments for the colon image of Figure 3.17 on the basis of  $8 \times 8$  image window.

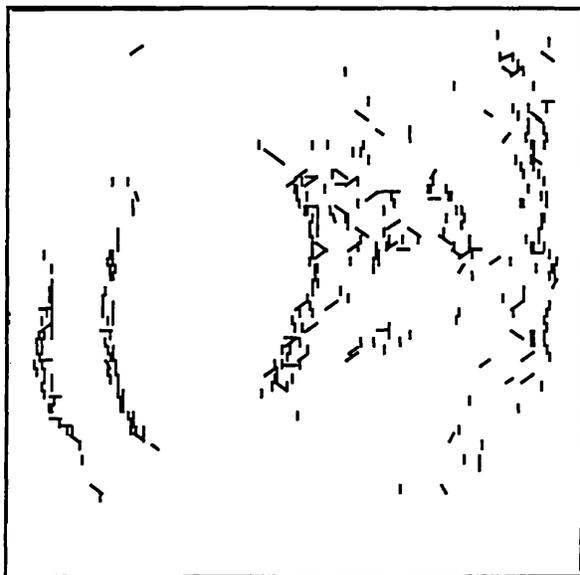
- (a) When the perceptual grouping criteria are employed.
- (b) Without using perceptual grouping.
- (c) Grouping based on orientation drift and theta-aggregation.
- (d) O'Gorman and Clowes method.



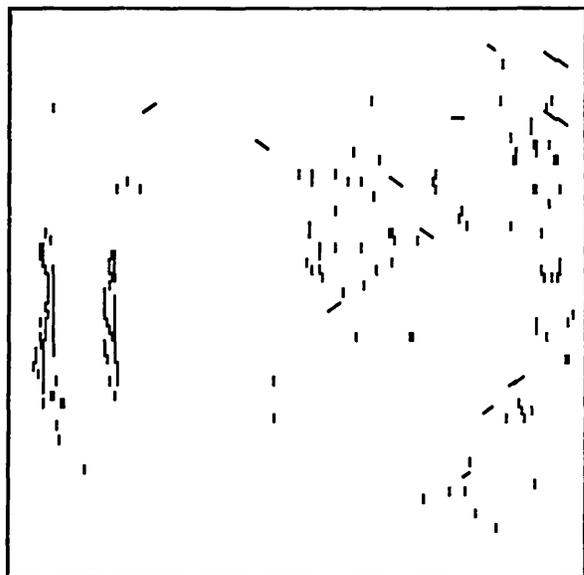
(a)



(b)



(c)



(d)

Figure 3.26: Output line segments for the colon image of Figure 3.17 on the basis of  $4 \times 4$  image window.

- (a) When the perceptual grouping criteria are employed.
- (b) Without using perceptual grouping.
- (c) Grouping based on orientation drift and theta-aggregation.
- (d) O'Gorman and Clowes method.

used by O'Gorman and Clowes filters out those edge points whose orientation is not normal to the direction of the line segment (within a tolerance), from participation in the grouping process. This type of filtering works well for polyhedral objects with strong edges and noise free images but for curved object contours the method is unable to identify all the contour structure. Different line normal and edge orientation tolerances (from  $\pm 5^\circ$  to  $\pm 25^\circ$ ) have been tried to optimise the method similar to the previous section, but the line segments which are weak yet form a useful part of the contours, were missed (see results in Figure 3.21d to 3.26d with optimised tolerance). On the other hand, our method of orientation grouping which allows a slow drift in edge orientation (typically  $5^\circ$ ) between consecutive edge points as one moves along the curved line segment, gives better results. The results of connectivity and orientation grouping based on slow drift and theta-aggregation are shown in Figure 3.21c to 3.26c. The line segments extracted by our algorithm for the colon and artificial images are also overlapped on the original images to determine the accuracy of locating these segments. These overlapped pictures are presented in the next chapter, where the line segments are grouped into curved contours.

### 3.6.3 Conclusions

It is difficult to judge the performance of different line detection methods based on the total number of extracted line segments alone. Nevertheless, with the help of the output results presented already and the statistics of the output of different grouping processes, we can gain an insight of the ability of different grouping techniques in reducing the image data for varying degree of noise and without losing useful information. The total number of line segments extracted for the artificial and colon images under different grouping methods and resolutions are provided in Table 3.1 and 3.2 respectively.

It is easy to see that our perceptual grouping method reduces the amount of line data almost to 50% without losing useful image line structures. Another interesting feature is that orientation

Test Image	Figure 3.15		Figure 3.16		Figure 3.17	
	8×8	4×4	8×8	4×4	8×8	4×4
Grouping Process						
Line segments extracted without applying perceptual grouping principles.	433	1186	346	558	348	784
Line segments detected by using perceptual grouping.	177	564	118	255	146	400
Line Segments found by O'Gorman and Clowes method of collinearity grouping.	44	320	43	171	47	191
Connectivity & orientation grouping with slow drift in orientation and theta-aggregation.	160	518	93	230	124	355
The share of CI-Grouping in total line segments found by our algorithm.	17	46	25	25	22	45

*Table 3.2: Output line segments for colon images using different perceptual and other grouping processes.*

grouping covers up to 90% of the line structures. This is all due to the introduction of slow drift in orientation and theta-aggregation. Without this, the O'Gorman and Clowes technique is unable to extract comparable line segments. The 10 to 15% share of lines found due to CI-Grouping supports our argument that the effectiveness of CI-Grouping is poor and is sensitive to tolerance thresholds.

In this chapter we have demonstrated the capabilities of perceptual organisation at a very early level of vision specifically in filtering noisy edges for image segmentation. The application of this research is for an on-line image analysis. Therefore during the whole development process of these grouping principles the possibility of parallel implementation was kept in mind. Custom VLSI hardware for the Hough transform [Sher and Tevanian 1984, Rhodes et al. 1988] encouraged its use as a part of our method. Hanahara et al. [1988] have also reported a real-time processor for the Hough transform which takes 0.79 Second for 1024 feature points. The implementation described here, in which the feature points will not be more than 144 in an image window, can be easily achieved below a millisecond. Previously, the Hough transform has been mostly used as a global method for contour detection. We have utilised it for detecting line segments locally. The bias in the Hough transform is prevalent for small image windows and precise detection of line segments cannot be guaranteed at different orientations. The image window can be rotated (e.g.  $9^\circ$ ,  $18^\circ$ ,  $27^\circ$ , and  $36^\circ$ ) to detect more accurate lines but in our particular application, line extraction accuracy is adequate and we are able to identify all the relevant contours at a reasonable accuracy as shown in the results of next chapter. The Hough transform for a small window avoids long computation time and large memory requirements and thus it is possible to implement the whole perceptual grouping process on a single VLSI chip. These individual processing elements can serve as a part of the pyramid architecture for the contour extraction process which is described in the next chapter.

## CHAPTER 4

### CONTOUR DETECTION FROM LINE SEGMENTS

#### 4.1 Introduction

The contours in a single, monocular image provide useful information about the content of a scene. Although image contours are two-dimensional, they still yield information about three-dimensional shape. In the case of endoscopic images the inner colon muscles generate occluding contours in colon images, which provide unambiguous depth information essential to navigate the endoscope. The problem, we are trying to solve is the extraction of these contours in noisy images. In the previous chapter, a new method was presented to construct a multi-resolution intermediate representation for contours by perceptually filtering noisy edges. The intermediate representation is based on the approximation of curves by short line segments. The edge points are grouped into line segments by following the different laws of perceptual organisation. There is still however a possibility that line segments may exist where no meaningful image contour does, and conversely lines may be absent where a boundary exists. The solution is to apply different perceptual grouping principles for linking only the relevant line structures and for resolving ambiguities. Lower level image data (line segments at lower resolution and edge points) can be employed to fill the gaps for fragmented parts of contours.

The pyramid based representation can easily represent the line segments at two resolutions in an organised way which supports efficient and fast searching of the image data ( $O[\log(n)]$  for an  $n \times n$  image). It also supports the parallel-pipeline implementation of the contour extraction algorithm. If we assume that the PEs (processing elements) at different levels in a pyramidal computer

can perform different operations on the image data then, for implementing contour extraction, the PEs at the lowest level of pyramid may perform edge detection and feed edge data to the PEs at level two and three where perceptual grouping can be performed using the Hough transform. The level two PEs detect line segments for a  $4 \times 4$  window while PEs at level three extract line segments for  $8 \times 8$  window. Alternatively, the line segments can be fed into the pyramid structure as inputs. The second alternative was adopted and the contour extraction method was implemented using quadtrees to simulate the pyramidal architecture. The line segments are grouped into contours by using the collinearity, curvilinearity, and theta-aggregation grouping principles in an hierarchical manner. The main grouping is performed on line segments based on an  $8 \times 8$  window and lower line segment data is only utilised when a gap is encountered, or for resolving ambiguities between equally significant competitor groups. The edge point data can also be employed to fill in short gaps between contour ends and corners.

As we have mentioned previously, the contour extraction method consists of bottom-up and data-driven processes and it is motivated by psychophysical and neurophysiological studies. The pyramid structure is employed for grouping line segment data hierarchically. The transition gap between signal to symbols (from edge points to contours) is traditionally considered a source of discontinuity for the flow of information in machine vision. Our approach which utilises a multi-resolution representation offers the potential for eliminating this discontinuity. The contours are detected irrespective of their type and source. The only requirement is that the contour should result from a reasonable intensity change in the image.

## **4.2 Shape and Contours**

There is no doubt about the power and vividness with which contours can depict shape. The main question is how do contours create three-dimensional realism? Psychophysics is unable to

provide any answer to this question. It is not clear whether there is only one distinct module in human vision which performs the job of shape from contour or there are several modules which are jointly responsible for the vividness of contour perception. The contour extraction method described in this thesis can only extract those contours which correspond to a grey level change in the image. However we are concerned with the extraction of many types of contours in the image, and it is useful to classify different types and how they can provide three-dimensional shape information specifically the third dimension (depth).

#### 4.2.1 Contour Types

There are a number of different conditions under which a contour can arise in an image, which are:

- Discontinuity in the depth.
- Discontinuity in the surface orientation.
- Change in surface reflectance.
- Different illumination effects in the scene.

The problem of shape from contour becomes more difficult when from a single, monocular image one tries to identify different sources of contours. The type of a contour in terms of its source is essential for conveying information about the shape. The main categories of contours for interpretation purposes are occluding contours, surface contours, and contours due to change in surface orientation.

##### *Occluding Contours*

These contours simply occur at discontinuities in the distance of the surface from the viewer. They are very useful in providing plenty of clues about the shape. For every occluding contour in the image, there is a particular curve on the object surface known as

the contour generator. The contour generator consists of a set of points on the surface where the surface normal is orthogonal to the direction of view. When humans interpret occluding contours there are some a priori assumptions that allow them to infer shape from an outline or silhouette. Marr [1977] has described three of these assumptions.

-Each point on the contour generator projects to a different point in the contour.

-Nearby points on the contour arise from nearby points on the contour generator.

-The contour generator lie wholly in a single plane.

When these assumptions are satisfied from a distant viewpoint (orthographic projection) and for a smooth surface, the surface can be defined by generalised cones. Conversely for a generalised cone surface, these three conditions always hold. In other words the concavities and convexities in occluding contours are important in determining the surface because they are the actual properties of the surfaces represented by generalised cones. We are proposing a three-dimensional model of the colon in terms of generalised cones, which will be described in chapter six. That model promises to be useful for the estimation of shape from occluding contours in endoscopic colon images. The only exception is when the endoscope tip is facing the colon walls (i.e. normal to the colon axis), or when the inner colon is viewed from a position where its axis is foreshortened.

### *Surface Contours*

These contours are no longer restricted to the silhouette boundaries and they can arise within the silhouette. The contour generator for the surface contours may be due to internal surface markings or different types of illumination effects. The surface contours are difficult to analyse because there is no obvious

source of their regularity which humans can use to infer shape from them.

### *Surface Orientation Contours*

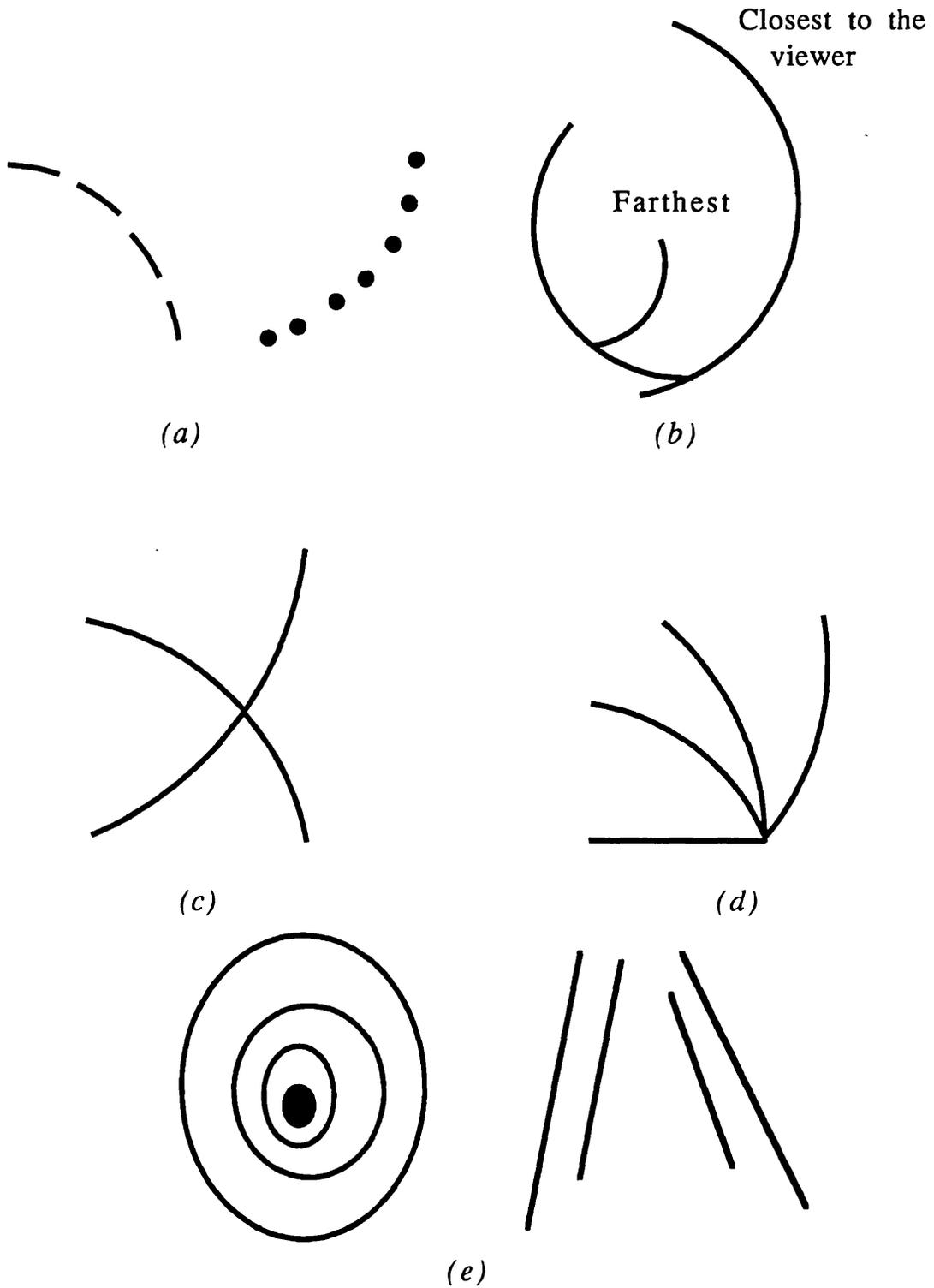
The surface orientation contours are along the loci of change in surface orientation. Generally they follow the creases on the surface. Regarding the recovery of surface geometry from these contours, it is difficult to resolve whether the contours correspond to concavity or convexity in the surface.

#### 4.2.2 Three-Dimensional Structure from Image Contours

In this section we are concerned with the use of image contours in determining three-dimensional geometric structure in the image. Lowe and Binford [1981] describe some general and domain independent constraints for the interpretation of image contours. The bases of these constraints are the non-accidental argument and the coincidence assumption for the view point and light source positions. These constraints can be used to categorise image contours into their distinct classes described in the previous section. There may be a large number of constraints and sub-constraints and some of them are also helpful in the detection of contours. Here we are discussing only those constraints which interpret image contours in terms of their distance from the observer and specifically the contours in colon images. The individual situations are also explained in Figure 4.1.

### *Curvilinear Alignment*

When two contours are aligned in an image (even if they are separated by a gap), they are also aligned in space as shown in Figure 4.1a. The only exception is when contours are parallel and the viewer happens to be in the plane of the contours. This constraint is very handy in bridging the gaps in curves due to errors in the contour extraction method.



*Figure 4.1: Shape from contour: Constraints*

*(a) Curvilinear alignment. (b) Termination at a*

*continuous contour. (c) Crossing contours.*

*(d) Contours ending at a common point.*

*(e) Vanishing point.*

### *Termination at a Continuous Contour*

When an image contour terminates at a continuous curve (T Junction), the continuous curve is normally closer to the viewer than the terminating contour (Figure 4.1b). This is very useful clue for identifying the closest contour in colon images. Because during the navigation of the endoscope, the nearest contour should be avoided and the tip can be aimed towards the hypothesised centre of that contour.

### *Contour Crossing*

When two continuous contours cross one another (X Junction as shown in Figure 4.1c), it indicates either an illumination discontinuity, a transparency, or a rare combination of surface contours. Both of the crossing contours cannot belong to occluding geometric boundaries. If one of them is an occluding contour then the other must be either a wire or the edge of a partially transparent object and it must be closer to the camera. If one contour belongs to a shadow boundary then the other must be a surface marking on the same surface.

### *Termination at a Common Point*

When two or more contours terminate at a common point in the image (L, Y, K, or Higher order Junctions as shown in Figure 4.1d), then they also terminate at a common point in space. The exception is an accidental coincidence when the viewer is aligned in such a way that separate vertices in space project at a common point in the image.

### *Vanishing Point (Parallel Lines)*

Due to perspective projection in image formation, the parallel straight lines in object space converge to a common vanishing point in the image. Once a vanishing point is found for some lines, the other lines which are aligned can be assumed to be parallel. The vanishing point is very important for solving different

navigational problems and it has been used in road following [Liou and Jain 1987]. In colon images the inner muscle rings yield their own vanishing point in terms of reduction in the size of those rings as their distance from the observer increases (Figure 4.1e). When these muscle rings are partially visible, the length of their image contours also provide a distance estimate. Generally a short length contour surrounded by longer contours is further away from the viewer than the longer contours. Additionally if the colon is viewed along its axis the darkest region in the image corresponds to the vanishing point and in this way the endoscope tip can be guided towards that region.

### **4.3 Different Approaches to Contour Extraction**

The main purpose of contour or boundary extraction is to make some identifications in the image. This phase of machine vision is an intermediate step that passes its findings on for post processing by a higher level stage. There are a number of approaches to contour extraction which use edge points in one way or another to form boundaries. All the methods incorporate some sort of knowledge into the grouping operations which map edge points into contours. For example, the knowledge of where to expect a boundary, allows a considerable reduction in the search space. If a little is known about the contours, the methods rely on the general knowledge and heuristics which are true for most domains. The approaches to contour extraction we are examining in this section include the use of:

- Linking edge points directly.
- Grouping line segments.
- Multi-Resolution or pyramidal methods.

#### **4.3.1 Grouping Edge Points into Contours**

The methods which group edge points into contours directly include graph searching, the minimum spanning tree, the Hough transform, and relaxation techniques. Graph searching techniques,

the methods employing minimum spanning tree, and Hough techniques have been introduced in the previous two chapters. Relaxation is a widely used method at different levels of computer vision for refining noisy and impure information. The purpose of edge grouping is to find sets of edge points which are consistent to each other. This process can be modelled by the probabilistic relaxation method proposed by Zucker et al. [1977]. To formulate the relaxation process, each pixel is assumed to have  $n$  labels which correspond to  $n-1$  directions of the edge and an  $n$ th label to no edge. For every pixel the probability of each label is first initialised by applying an edge operator to the pixel. The general idea is to compute some probability updating contribution for the central edge as a function of the probability of neighbouring edges. By overlapping the neighbourhood and iterating the decision process, local changes propagate and affect the surrounding neighbourhoods. In every iteration the totality of changes in the local neighbourhood is used to update the existence of each edge. Several iterations may be required for the relaxation process to link edges by suppressing noisy edges and enhancing long smooth edge lines.

The number of iterations and convergence speed is affected considerably by the amount of noise in the image, which is a factor to be considered before using relaxation methods in real-time applications. The idea of refinement by relaxation is closely related to dynamic systems which has interesting implications for neurophysiology. If more edge properties are added to the probability functions for updating, a higher dimensional state space results and can cause considerable computational difficulties. Another improvement in relaxation can be sought by finding global quantities which are optimised in the solution.

#### 4.3.2 Aggregation of Straight Line Segments

There are only a few techniques of image analysis where bottom-up organisation of the image data has been adopted successfully. Marr [1976] has given a new life to the idea of bottom-up

organisation in his theory of early and intermediate levels of vision. Mainly two types of grouping process have been put forward for linking line segments into contours. The grouping principles along with the principle of least commitment play a central role in organising the raw primal and  $2^{1/2}$ -D sketch.

Theta-aggregation is one of the grouping principles in which a set of similarly oriented line segments are glued in a direction which differs from the intrinsic orientation of line segments. The principle is based on very local grouping measures to form a curve or line which has orientation associated with it rather than the individual line segments. This principle is also used for combining virtual lines and in fact it is not necessary to know individual line orientation explicitly. The second grouping principle which Marr has argued for linking line segments is known as curvilinear aggregation. In this grouping principle the line segments are grouped in such a way that the assembled contours preserve the orientation of segments. The results of applying both of these principles are given for different images but it is not clear how these methods were implemented. It has been mentioned that the theta-aggregation principle is more basic and an early process, and therefore it should be applied first on the image data. The curvilinear aggregation is more successful if applied on the resulting line segment data from theta-aggregation.

In another implementation for line extraction Weiss and Boldt [1986] have used geometric grouping to form longer straight lines by gluing together the shorter line segments. Their grouping method is based on the proximity and collinearity. A straight line is defined as a sequence of line segments in such a way that consecutive pairs of line segments are roughly collinear and have similar contrast. The main grouping process has two steps. In the first step, those pairs of straight lines are tentatively linked which obey proximity and collinearity. The linking based on proximity and collinearity is implemented by defining different conditions for grouping which are based on the distance between end points, overlap, contrast, and orientation of line segments. The main purpose of this step is to avoid a combinatorial explosion in the

search space. In the second step, the paired line segments are further merged into longer straight lines and the usage of geometric context depends on the search radius which provides an upper bound on the length of a sequence of segments tested for their straightness. The sequence of lines which passed this test is replaced by a single longer line. The authors also describe the hierarchy in grouping by defining different planes where each plane is divided into a grid whose size depends on the density of lines. The main problem with the method is that it cannot cope with texture and noisy images and it may over-merge the lines. From the contour extraction point of view, this method is restricted to detecting straight lines and therefore it is difficult to adapt it for extracting curved contours.

#### 4.3.3 Multi-Resolution Based Contour Extraction

Pyramidal or multi-resolution methods have their roots in neurophysiology. The main advantage of using these techniques is their computational efficiency. In addition to that, the edges at multiple scales can be used to analyse the underlying physical causes of the brightness changes. Although similar pyramidal techniques have been used for the analysis of multi-scale curvilinear image data, representing boundaries, we are restricting ourselves to hierarchical linking of edge data into contours. There are two alternative approaches to the problem of extracting contours using multiple resolutions:

- A grey level pyramid is constructed from the given image and boundary points are extracted at each scale. Then the grouping rules based on the proximity, continuation, and similarity in orientation and contrast are used to link the boundary points locally within each level and between adjacent levels of the pyramid.

- In the second approach the input to the pyramids can be line segments or curves at each level of the pyramid. Normally line or curve segments are fed

to the retina (bottom level) of the pyramid. Then at each level they are grouped into longer lines or curves and the result is passed on to the next higher level. In this way the contours are formed by reaching the root or apex of the pyramid.

Kelly [1971] originated the concept of multi-resolution processing for contour detection by employing two scale, coarse to fine planning. A coarser image was computed by averaging the pixels in every  $8 \times 8$  window of the original image. The boundary was detected from the coarse image which was then used to find the contours in the original image. This is a sequential approach which saves needless work in uninteresting areas of the highest resolution image and also serves to verify and localise the results obtained. A similar coarse to fine method based on two-resolution has been used for detecting lung tumours in chest radiographs [Sklansky and Petkovic 1984].

The edge pyramids are also used to extract boundaries of objects [Hong et al. 1982]. A pyramid is constructed by reducing the resolution of the image at successive levels and then edge detectors are applied at each level of the pyramid. The edges between adjacent levels and at the same level are linked based on their distance and orientation. The extraction of straight lines and smooth curves has also been carried out by using overlapped pyramid structure [Hong et al. 1983]. In this method the boundary curves can be fed into the appropriate levels of the pyramid. The contours are approximated by line segments and at each level segments from the level below are combined using local position, curvature, and direction to form longer contours. These contours are then passed on to the higher levels for further grouping. A similar approach has also been employed to link lines hierarchically for corner detection [Hartley and Rosenfeld 1985].

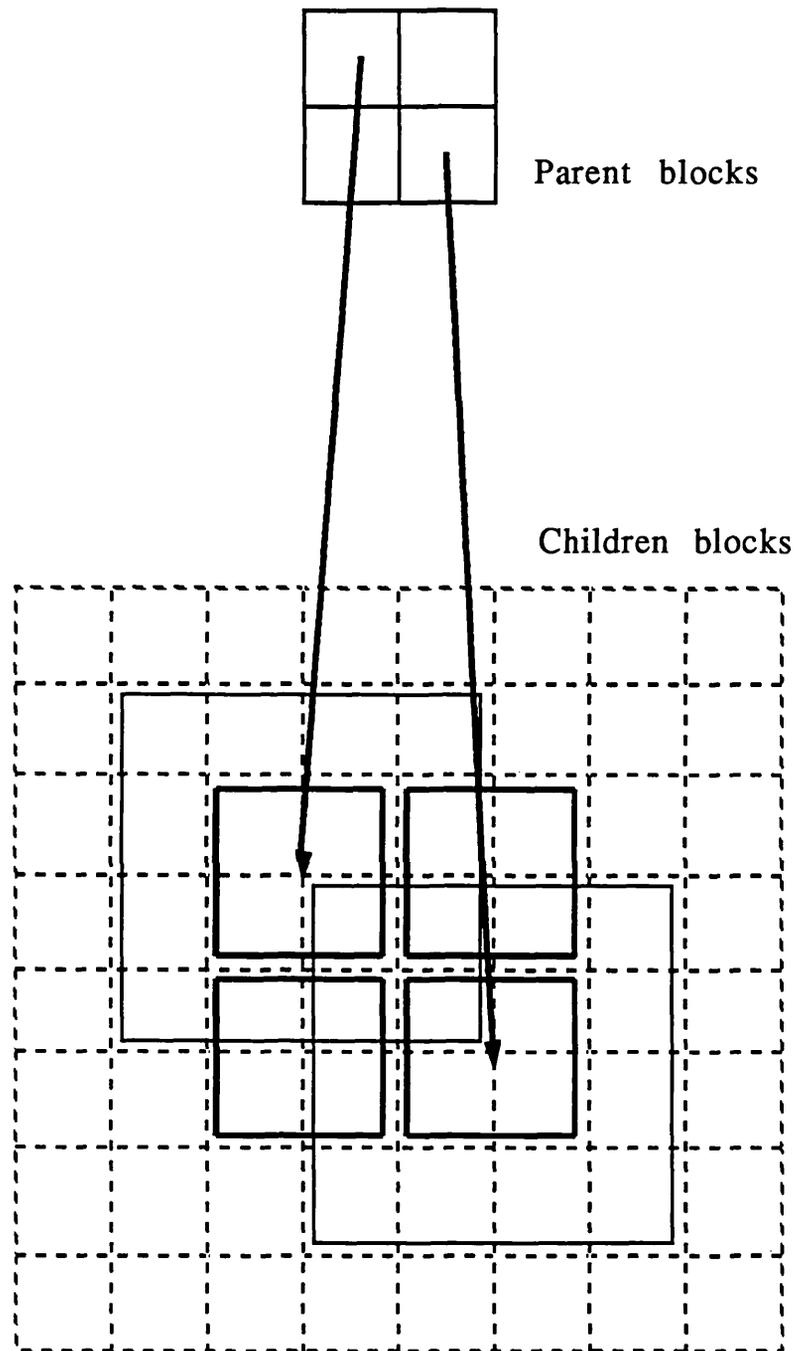
## 4.4 An Hierarchical Method to Detect Contours

Humans can perceive straight and curved contours easily even if they are broken and relatively sparse, or they are obscured by other curves. The pyramid-based techniques make the global properties of contours local and therefore global operations are easy to implement in parallel. The main difficulty in pyramidal methods is that by reducing the resolution, there is a chance of losing relevant contour information. This problem can be overcome if the information at lower levels is readily available and used in an informed way. In our contour extraction algorithm, the line segments are fed to the lower two levels of pyramid. Then higher level processing elements group the line data supplied by their children and pass on the grouped segments to higher levels for further grouping. In this way contours are constructed in a single pass when the pyramid root is reached.

### 4.4.1 From Line Segments to Contours

The approximation of image contours by line segments simplifies their representation at different resolutions by different length line segments without losing useful information. The variable length line segments are easy to detect by using line detectors at different size image windows. A sequential version of the contour extraction method which groups line segments only at one resolution has also been developed [Khan and Gillies 1988b]. The sequential contour tracing uses a number of heuristics to fill the gaps in broken parts. The method described here is extended to parallel implementation and uses line segments extracted at two resolutions.

A line pyramid is constructed in such a way that its two lower levels hold line segments extracted for  $4 \times 4$  and  $8 \times 8$  image windows. The pyramid is based on  $4 \times 4$  overlapping neighbourhoods and each node has four parents and sixteen children as shown in Figure 4.2. Each parent performs grouping on its sixteen children, but it only keeps the grouped line segments in



*Figure 4.2: A 4x4 (50%) overlapped pyramid scheme, where each node has sixteen children and four parents. Thick and smaller squares are the central 2x2 children, while the larger squares cover all the sixteen children.*

its central  $2 \times 2$  block for passing on to the parents. The bottom level line segments are only considered when gaps are encountered. Rather than using ad hoc heuristics for filling broken contours it is better to use lower level data in support of any decision.

The grouping is mainly carried out on the line segments found in the  $8 \times 8$  image window. A number of grouping criteria are used to achieve the aggregation of line segments into contours. Proximity, theta-aggregation, curvilinearity, continuity, and similarity in line contrast are the main grouping principles used in the linking process.

### *Theta-Aggregation*

For theta-aggregation line segments of similar orientations are grouped together, whether their individual orientations are preserved in the resulting contour or not. The grouping based on theta-aggregation has been explained previously. Here we have restricted this criterion by grouping only if the resulting contour direction is not normal to the individual line segments.

### *Curvilinearity*

In curvilinear aggregation only those line segments are grouped, whose orientation is roughly collinear with the contour direction. In this way this grouping process preserves the orientation of line segments.

### *Proximity*

The proximity of line segments is explicitly defined by the distance between their end points for curvilinear grouping and perpendicular distance between parallel lines for theta-aggregation. If lines are well apart they are not linked immediately and support for their grouping is sought from the line segments at the bottom level.

### *Continuity*

The continuity during linking of contour segments is preserved by grouping those contour segments whose respective end directions (tangent at contour ends to be joined) are compatible.

### *Similarity in Line Contrast*

In deciding between competitor groups of line segments, their contrast is also used and similar contrast lines are preferred.

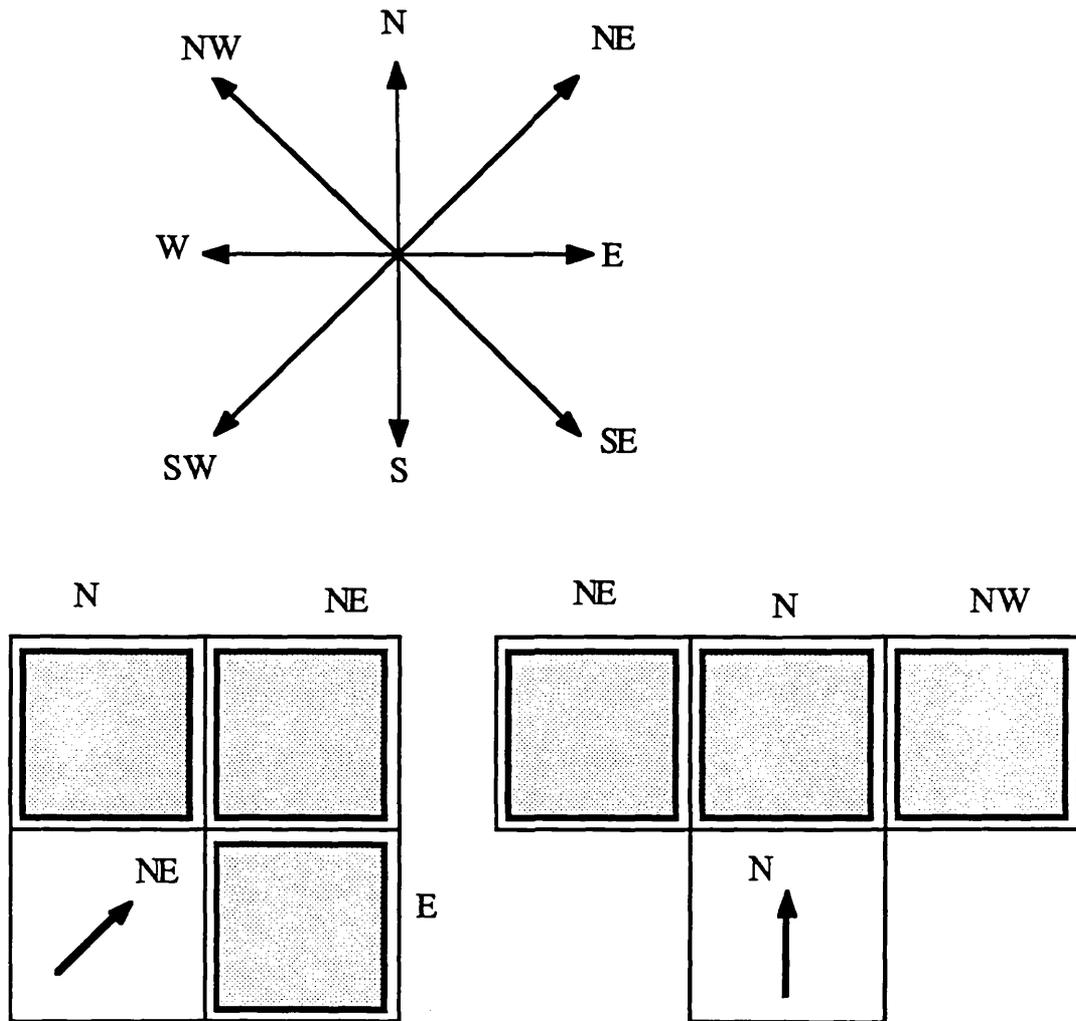
For nodes at an immediate level above the line segment nodes three neighbours are considered for linking line segments to form a contour segment. One neighbour is in the prior grouping direction and the other two are its left and right side neighbours. Initially the grouping direction is approximated from line segment orientation. For example as shown in Figure 4.3, if the grouping direction is NE, then the three neighbours considered are in NE, N and E directions. Similarly for N direction the three neighbours are NE, N, and NW. This means that contours can turn up to a maximum of 45° over an 8×8 image. For nodes at higher levels contour segments with compatible end directions are only considered and they are grouped if the neighbouring line segments from both groups follow one of the grouping principles, which are based on theta-aggregation and curvilinearity. During the linking process which is performed locally by each parent to link its children, thinning can also be performed to eliminate those line segments which are parallel to the contour segments and located within a pre-defined distance. The detail of the overall grouping process along with its parallel implementation is discussed in the next section.

#### 4.4.2 The Algorithm

The method of extracting contours in parallel using a pyramid structure is divided into following steps:

*Step 1: Construction of a 4x4 (50%) Overlapped Pyramid*

A 4x4 overlapped pyramid is constructed in such a way that its nodes at the bottom level (level zero) represent line segments extracted on the basis of 4x4 image window, while nodes at one level above (level one) carry line segments for 8x8 image window.



*Figure 4.3: Typical three-neighbours used to group line segments.*

## *Step 2: Grouping Line Segments*

The nodes at an immediate level above the line segments and the higher levels nodes perform similar grouping operations, but different approaches are followed due to different types of input data.

(i) For each node at one level above the line segments (in this particular implementation level two) contour segments are formed by grouping line segments. The three-neighbour strategy described earlier is also performed at each end of the segments. The grouping is carried out using proximity, theta-aggregation, and curvilinearity from the sixteen children as explained below.

-If the grouped line segments do not pass through the central  $2 \times 2$  block, then either no contour exists in the image area corresponding to that node or some portion of the contour which passes through the node is missing. Support for the missing portion of the contour segment is sought from the bottom level line data. If no contour segment is found then a no-contour flag is passed to its parents.

-If one or more contours pass through the central block then for every line segment, check for the possibility of grouping it with neighbouring line segments. Group two line segments: if they are parallel and located within a pre-defined distance  $D_{max}$ , or if their directions are at the most  $45^\circ$  apart and end points are within the pre-defined distance threshold,  $D_{max}$ . The central block line segments are kept as a contour segment along with its tangent directions at both ends. The tangent direction at the ends of contours formed by theta-aggregation, may be different from the individual line orientations. All contour information is passed on to the parents for further grouping. This

information also includes the identification tag for those neighbours which are in the outer block and have been identified as part of that particular contour segment.

(ii) For all the nodes at higher levels, the contour segments from the central four children are tested for grouping among themselves or with the outer twelve children. This process proceeds as follows.

-For every central contour segment, check for the possibility of grouping it with other neighbouring contour segments. Group two contour segments if their respective ends are within a pre-defined distance  $D_{max}$ , end directions (tangents) are compatible, and the neighbouring line segments of both contours follow at least one of the grouping criteria (theta-aggregation or curvilinearity). This process is repeated on the resulting contours until there are no more grouping possibilities. From the resulting groups the central children contour segments are merged into longer contours by making a single group of line segments and by modifying end tangent directions. The information about the outer children which are part of the larger group is also passed on to the parent along with the merged contour.

-The central contours, which are not merged with any other contour segment, are flagged as completed and they are passed on to their parents. The higher level nodes just pass on these completed contours to their parents until the root is reached.

### *Step 3: Filtering Short Contours*

At the root level, the length of all the contours is checked and only those with length longer than a threshold  $CL_{min}$ , are retained.

An additional step for filling the gaps can be added before step 3, where the edge point data can be used for merging close contours. At this stage of the contour extraction algorithm, the knowledge about the shape and size of contours can also be exploited to achieve better results.

#### 4.4.3 Implementation Details of the Algorithm

The contour tracing problem is basically a sequential process but we have developed this technique using a pyramid structure in such a way that it can be easily implementable on a pyramidal computer. There are two approaches which can be followed.

In the first approach the retina or bottom level processing elements can extract edge points from the image, while the processors at the next level above can be used as links to pass on edge point data to be processed at next two levels simultaneously. At these two levels, the processing elements can perform perceptual grouping to extract line segments for  $4 \times 4$  and  $8 \times 8$  image windows. In a  $4 \times 4$  overlapped pyramid structure, the perceptual grouping process can be extended to  $6 \times 6$  and  $12 \times 12$  windows (if the implementation of the previous chapter is followed strictly) or even more. The processing elements at higher levels can implement the linking algorithm by forming groups of line segments which belong to their sixteen children and pass on the resulting aggregate of segments to next higher level of processes. In this way, by the time the root of the pyramid will be reached, all the groups of line segments can be formed and replaced by different contours easily.

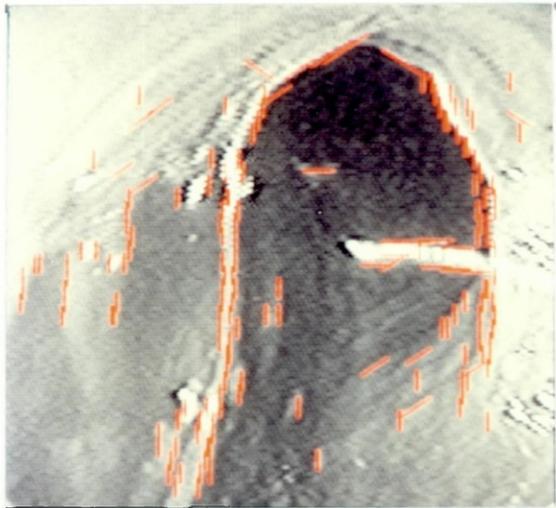
The second approach, which has been adopted in our implementation, is realised by computing edge points and then grouping them perceptually in a classical parallel way. The line data is then fed to an overlapped pyramid computer which links these line segments hierarchically similar to the first approach. The pyramid computer architecture is simulated by constructing a quadtree based on  $4 \times 4$  overlapping neighbourhood. For a  $256 \times 256$

image size, where line segments are extracted based on  $4 \times 4$  and  $8 \times 8$  image windows, the line pyramid has a total of seven levels labelled from zero to six (see Figure 2.5). The level zero which is also the bottom level of the pyramid holds  $64 \times 64$  array of line segments and level one carries  $32 \times 32$  array of longer line segments.

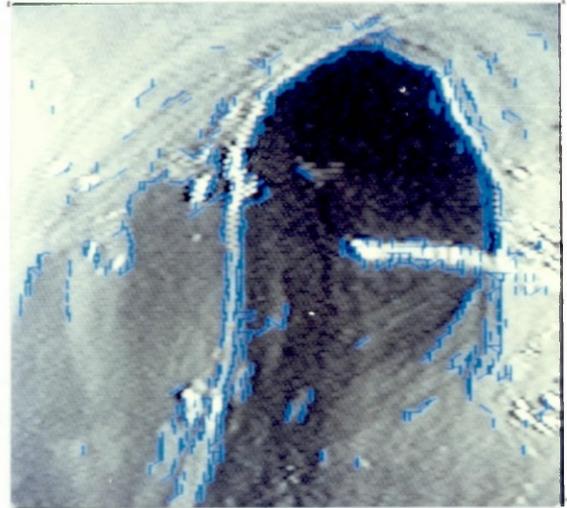
The higher level nodes require a data structure to keep multiple contour segments. A pointer field which points to the contour linked-list structure is provided in the node. For each contour segment a linked-list of line segments, ordered from one end of the contour to the other, is kept. While for multiple contours, a linked-list of contours is used and in addition to having the contour pointer, it also has fields to keep a record of the end directions of that particular contour, the number of line segment in the contour, and neighbour information about those outer block contour segments which can be grouped to form a longer contour at higher levels.

#### **4.5 Experimental Results and Discussion**

The line segments extracted by using the perceptual grouping method of previous chapter are used as input to the contour extraction algorithm. The algorithm has been tested on a large number of typical endoscopic colon images and the results from three of them with different type of contours and noise conditions are presented here. In addition to that the results of grouping line segments, extracted by employing connectivity grouping on one of the artificial images of previous chapter, is also presented. Four test images, with line segments overlaid, are shown in Figure 4.4 to 4.7. Red lines correspond to the line segments extracted on the basis of  $8 \times 8$  image window while the blue lines correspond to  $4 \times 4$  image windows. The grouped line segments for contours are shown in Figure 4.8. These groups have been formed in a single pass by traveling from the 2nd level of the pyramid to the top. The longer lines, which are based on  $8 \times 8$  image window, are dominant in these groups because the  $4 \times 4$  window based lines



(a)

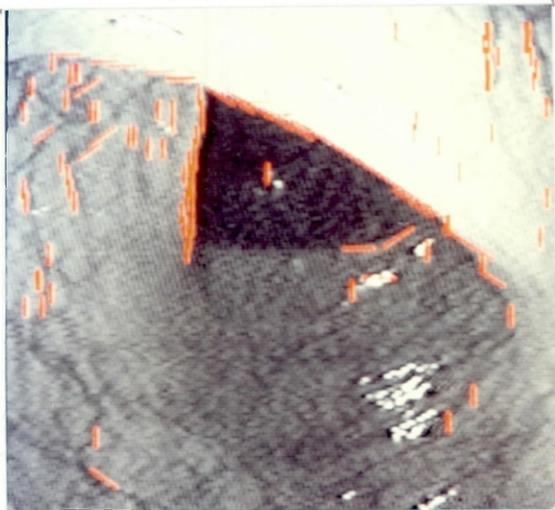


(b)

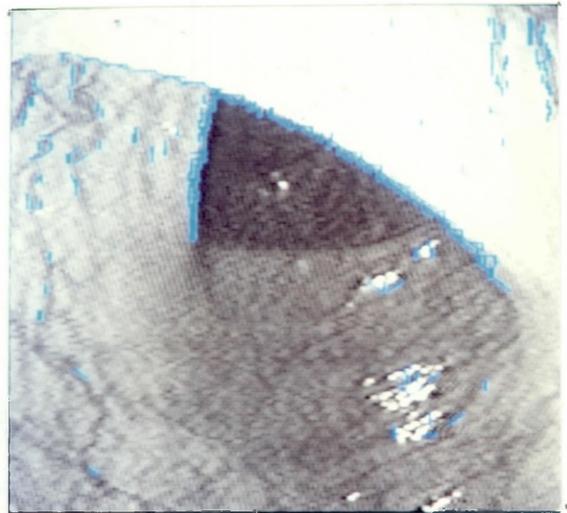
*Figure 4.4: Line segments extracted for the first colon test image.*

*(a) Based on  $8 \times 8$  image window.*

*(b) Based on  $4 \times 4$  image window.*



(a)

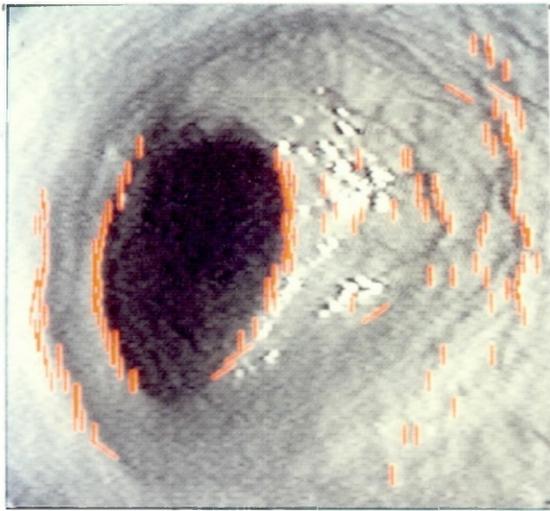


(b)

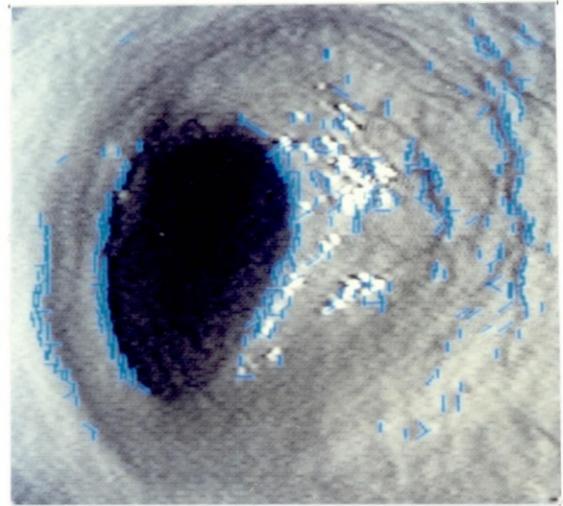
*Figure 4.5: Line segments extracted for the second colon test image.*

*(a) Based on  $8 \times 8$  image window.*

*(b) Based on  $4 \times 4$  image window.*



(a)

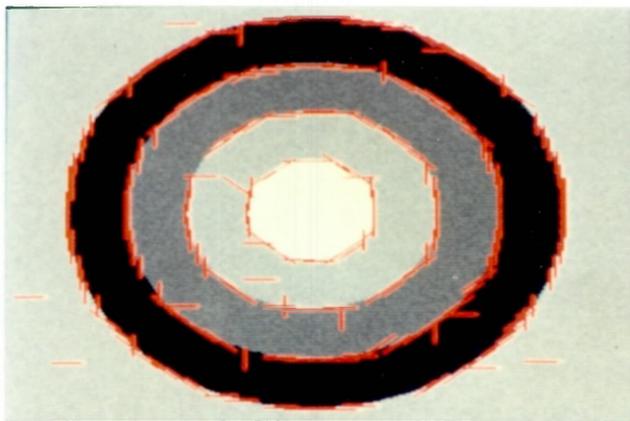


(b)

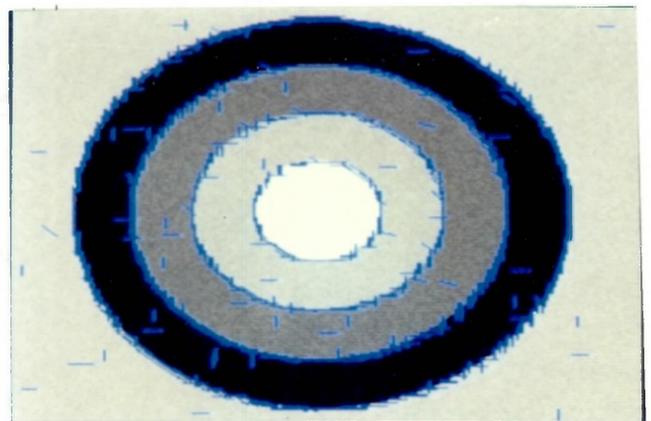
Figure 4.6: Line segments extracted for the third colon test image.

(a) Based on  $8 \times 8$  image window.

(b) Based on  $4 \times 4$  image window.



(a)

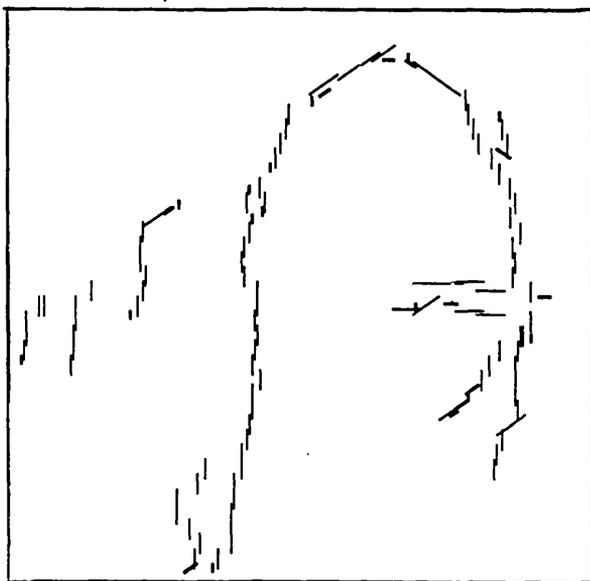


(b)

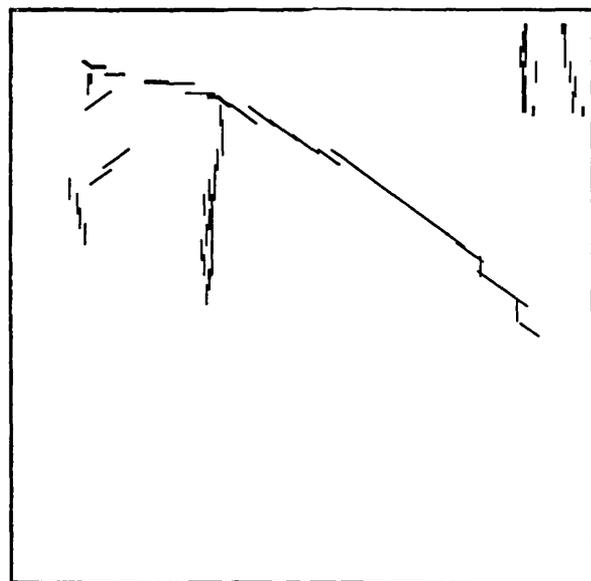
Figure 4.7: Line segments extracted using connectivity grouping for the artificial test image.

(a) Based on  $8 \times 8$  image window.

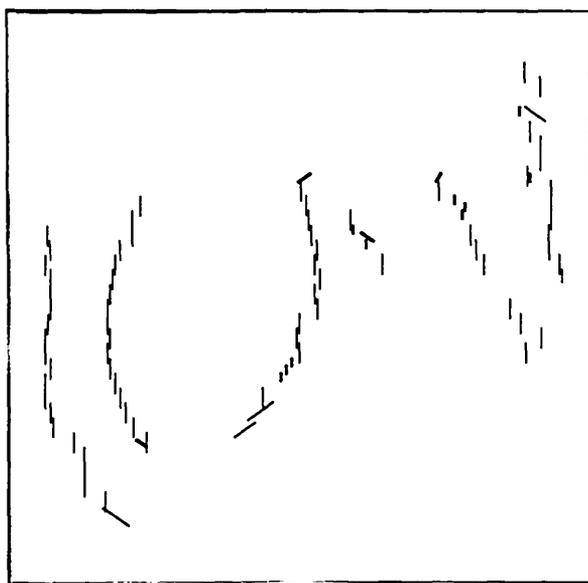
(b) Based on  $4 \times 4$  image window.



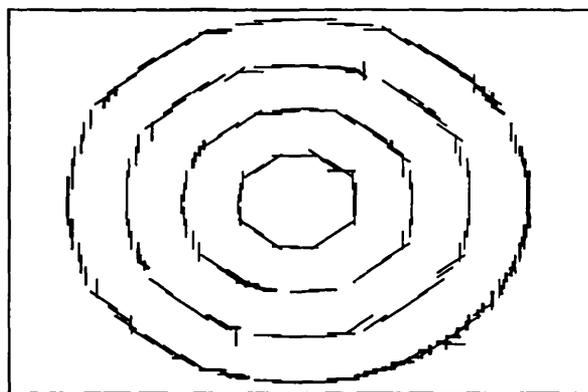
(a)



(b)



(c)



(d)

*Figure 4.8: Groups of line segments formed.*

*(a) For the first colon image line data of Figure 4.4.*

*(b) For the second colon image line data of Figure 4.5.*

*(c) For the third colon image line data of Figure 4.6.*

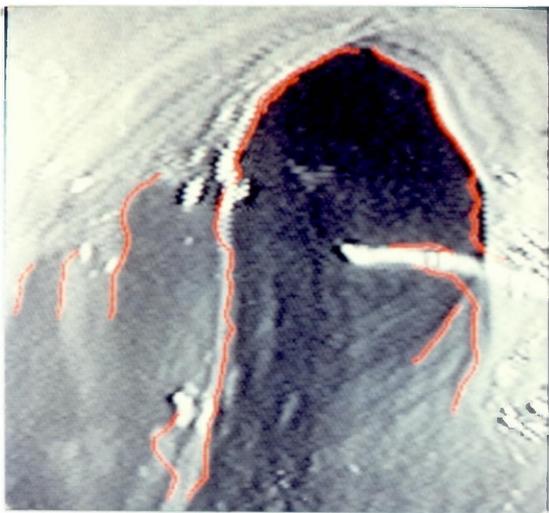
*(d) For the artificial image line data of Figure 4.7.*

(shown as thicker lines in Figure 4.8.) are only used for filling the gaps. The proximity grouping threshold ( $D_{max}$ ), is used as eight pixels for line segments which are based on  $8 \times 8$  image window and four pixels for  $4 \times 4$  lines. If the line end points (for curvilinear aggregation) or lines themselves (for theta-aggregation) are separated more than this threshold, they are not grouped.

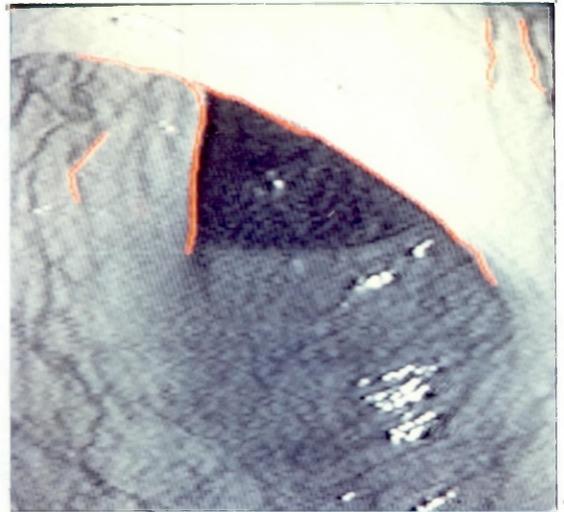
During the grouping of test line data of Figure 4.6, thinning was also performed. The group formation process, when combined with thinning, produces cleaner groups of lines as shown in Figure 4.8c. The results without applying thinning are not affected considerably because additional contours, which are of shorter length, can also be filtered out during the Step 3 of the algorithm. The groups of line segments are then replaced by contours which are shown in Figure 4.9 overlaid on the images. Only those contours are retained whose length is greater than a threshold  $CL_{min}$ , which is chosen to be at least four line segments in the results shown.

We now consider the statistics about the amount of visual data, processed at different levels of hierarchical representations, starting from the edges until contours are formed. The two visual data representations between the raw image and contours are edges and line segments. Table 4.1 provides an insight to the visual data reduction at each level of representations for the colon and artificial test images. It also illustrates the amount of filtering performed by perceptual grouping of edge points and line segments. The considerable reduction in image data from pixel to the higher level representations justifies our argument for an intermediate line segment representation in-between edge points and contours.

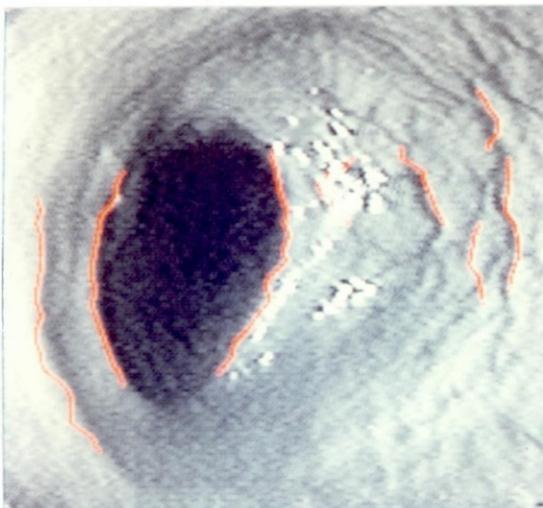
The main problem with the resulting contours are their lack of smoothness and precise location. This has happened mainly due to the use of the theta-aggregation principle. It is difficult to locate the position of contours accurately when parallel line segments are replaced by a contour which does not preserve the orientation of the individual line segments. In the results of Figure 4.9, we



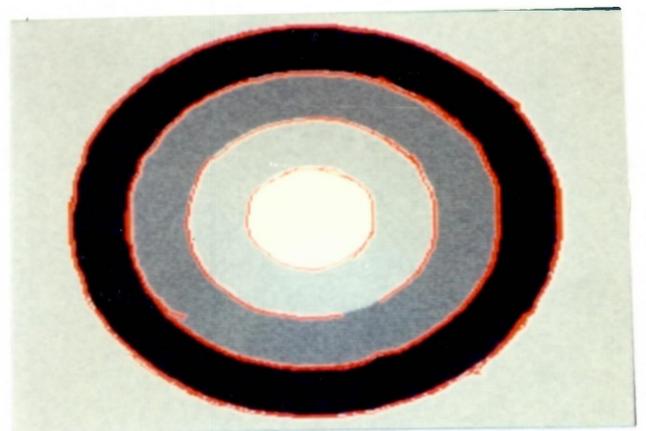
(a)



(b)



(c)



(d)

Figure 4.9: The output contours overlaid on the images. (a) First colon test image. (b) Second colon test image. (c) Third colon test image. (d) Artificial test image.

have simply placed the contours in such a way that they pass through the mid-points of each line segment. Theta-aggregation is an important principle which detects significant contours but at the cost of imprecision in locating the contours.

Visual Data Representations	Artificial Image Figure 3.6	1st Colon Image Figure 3.15	2nd Colon Image Figure 3.16	3rd Colon Image Figure 3.17
Image Size.	256×256	225×310	225×310	225×310
Number of pixels.	65536	69750	69750	69750
Number of edges retained.	28688	20925	13398	16406
Number of line segments found based on:				
8×8 window.	238	177	118	146
4×4 window.	697	564	255	400
Number of line segments emerged as part of contours.	301	137	78	103
Number of contours formed.	4	10	5	8

*Table 4.1: Image data reduction at the different levels of visual representation.*

In addition to that the method is also prone to accidentally forming curves due to texture. This problem can be overcome by using a higher threshold for the minimum length of contours. A similar problem was reported by Weiss and Boldt [1986] in their method of extracting straight lines. Nevertheless, this is the first successful attempt to use perceptual grouping in detecting arbitrary shaped contours from noisy images and the contour extraction method can be improved further for other applications.

The closest work to our method has been reported by Hong et al. [1983]. They have employed good continuity for merging contour segments based on proximity of end points and compatible end directions. Our algorithm departs from their method in a number of ways. First of all in addition to continuity, we have used two very important grouping principles based on curvilinearity and theta-aggregation. Secondly to reduce computation, children pass on the information to their parents about the prospective contour segments which can be grouped to a particular contour. In addition to that line segments are not replaced by contours until the top level is reached. This in turn helps in applying grouping criteria on neighbouring line segments, which belong to two different contour segments. The main criterion used by Hong et al. for detecting compatibility between contour ends is also not general and it depends on the radius of the circle, which they try to fit between end points. To achieve good results for straight lines the radius of the circle is set to a very high value (typically 1000) while for circular curves the radius must be a small value (typically 10).

The examples they have chosen to demonstrate their method, are very simple synthesised curves and lines. The technique has not been tried on real world images. In contrast we have shown how our algorithm performs on endoscopic and artificial images which contain a large variety of noise. The input to contour extraction algorithm contains a number of false line segments and the results demonstrate that our algorithm is robust and applicable to a more general class of real images.

## CHAPTER 5

### ESTIMATING RELATIVE DEPTH BY REGION EXTRACTION

#### 5.1 Introduction

In addition to occluding contours, the darker regions in the endoscopic colon images provide important depth cues for navigating the endoscope in colon. The task of locating these regions in colon images is closely related to region based image segmentation. The region based methods aim to extract regions of similar properties. They partition images by grouping pixels and regions on the basis of proximity and similarity in intensity, colour, range, or texture. Using these methods, the image is divided into uniform and homogeneous sub-regions in terms of some property.

The region extraction algorithms presented in this chapter, detect a uniform and coherent region of given properties. In the specific application of depth estimation in endoscopic images, the darker region is the deepest and free of obstacles. The task is to extract the lowest intensity region in noisy colon images. The endoscope application also requires that the region should be detected in real-time. The classical implementations of region extraction methods are sequential which makes them less applicable to real-time image analysis. Consequently, a pyramid structure has been used in our technique for region extraction. The use of pyramid structure has allowed us to devise a highly parallel implementation of the algorithm. The image histogram is used to estimate the average intensity of the darker region. A variance-average pyramid is constructed in which each node consists of the comprehensive properties of its corresponding region in the form of intensity mean and variance. These properties are calculated recursively by using the mean and variance of child nodes. During

this process a record is also kept for the largest dark and uniform region in each sub-tree of the pyramid under construction. In this way, starting from the pixel level, when the pyramid is completely built, the largest dark and uniform square region is identified in a single pass. This region may itself be large enough to provide insertion direction for the endoscope or, for a more general case, it can be used as a seed for the region growing process. The merging of regions can then take place to extract the complete region.

An extension of the algorithm is also implemented for image partitioning. The same technique is used to extract uniform seed regions, which can play an effective role in partitioning any given image. The improper selection of seeds can lead to inaccurate partitioning of an image. Ideally a seed region should be large and uniform. For image partitioning, a similar pyramid is constructed but only the record is kept for the uniform square regions which satisfy the uniformity criterion in terms of the region intensity variance and mean. When the pyramid is completed, the root node of each sub-tree in the pyramid contains the address of the node which corresponds to the seed region in that sub-tree. The nodes of each plane of the pyramid are treated as sub-trees and a link to the seed region for each sub-tree is established. An additional top-down pass in the pyramid is then initiated in which these seed regions are grown in parallel to identify complete uniform regions. A specific case of the algorithm is also investigated for detecting bright or dark regions without using the histogram information. The same pyramid construction steps which extract seed regions are employed with an additional restriction. For detecting bright regions, during the construction of pyramid when the seed region for a parent is selected from the seeds of its children, the brighter and larger size seeds are preferred. Similarly the darker region can be extracted by preferring darker seeds during the selection process.

In the experimental results, the endoscopic images of the human colon are presented to demonstrate the dark region extraction method. The image partitioning algorithm is applied to other

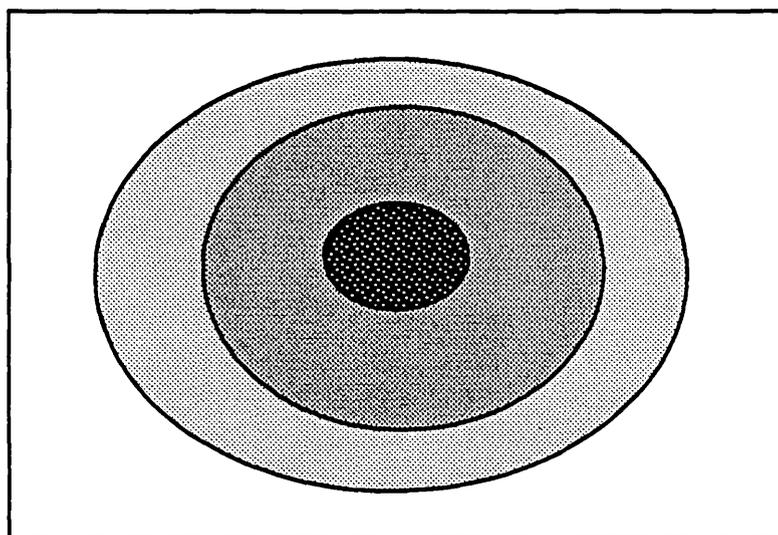
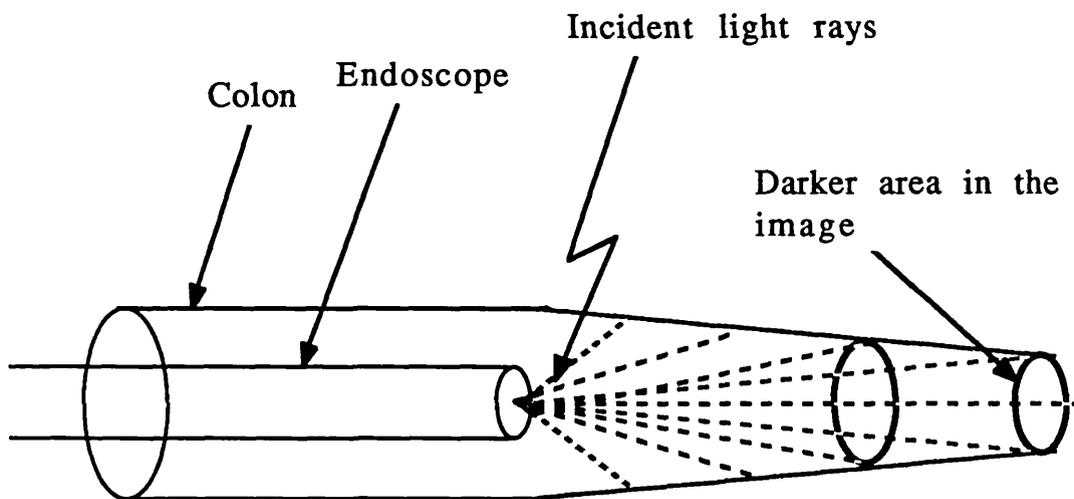
medical images and the results are also demonstrated on a variety of images including the computer generated images.

## 5.2 Depth from Intensity in Colon Images

Depth is the most important information which any navigation system needs to know for path planning and obstacle avoidance. Humans estimate depth by means of various cues including shading, texture and stereo visual processing, while some animals such as bats utilise a *time of flight* method for distance measurement.

Although there is a possibility of redesigning the tip layout of the endoscope to allow either the use of stereo vision or addition of a dedicated range sensor, we are currently assuming that the colon scene information is only available from a single camera. In this case the properties of the illumination and physical layout of the endoscope can be used. The inner body surfaces of the colon are illuminated by a single point-like light source. Although there is a lot of reflected light due to specularity, we still assume a point light source for the purposes of estimating the deepest region. Moreover the light source and the viewer are located almost at the same position and the light source is near to the colon surfaces. This simple illumination model resulting from the physical layout of the endoscope tip has led to the development of the method for estimating depth inside the colon, described here.

Under this arrangement, the colon surfaces which are nearer to the point light source are more brightly illuminated than the farther surfaces. When the light source is close to the object, which is the case in colon images, the light rays cannot be assumed parallel and the reflected light intensity becomes a function of the distance between the light source and the surface. The above condition is illustrated in Figure 5.1. The normal assumption is that the incident light intensity  $I_{inc}$ , varies inversely with the square of the distance  $r$ , between the light source and the point of reflection on the surface.



*Figure 5.1: The farthest part of the colon receives the smaller amount of incident light.*

In other words

$$I_{inc} = I_0/r^2 \quad (5.1)$$

where  $I_0$  is the light intensity at the light source.

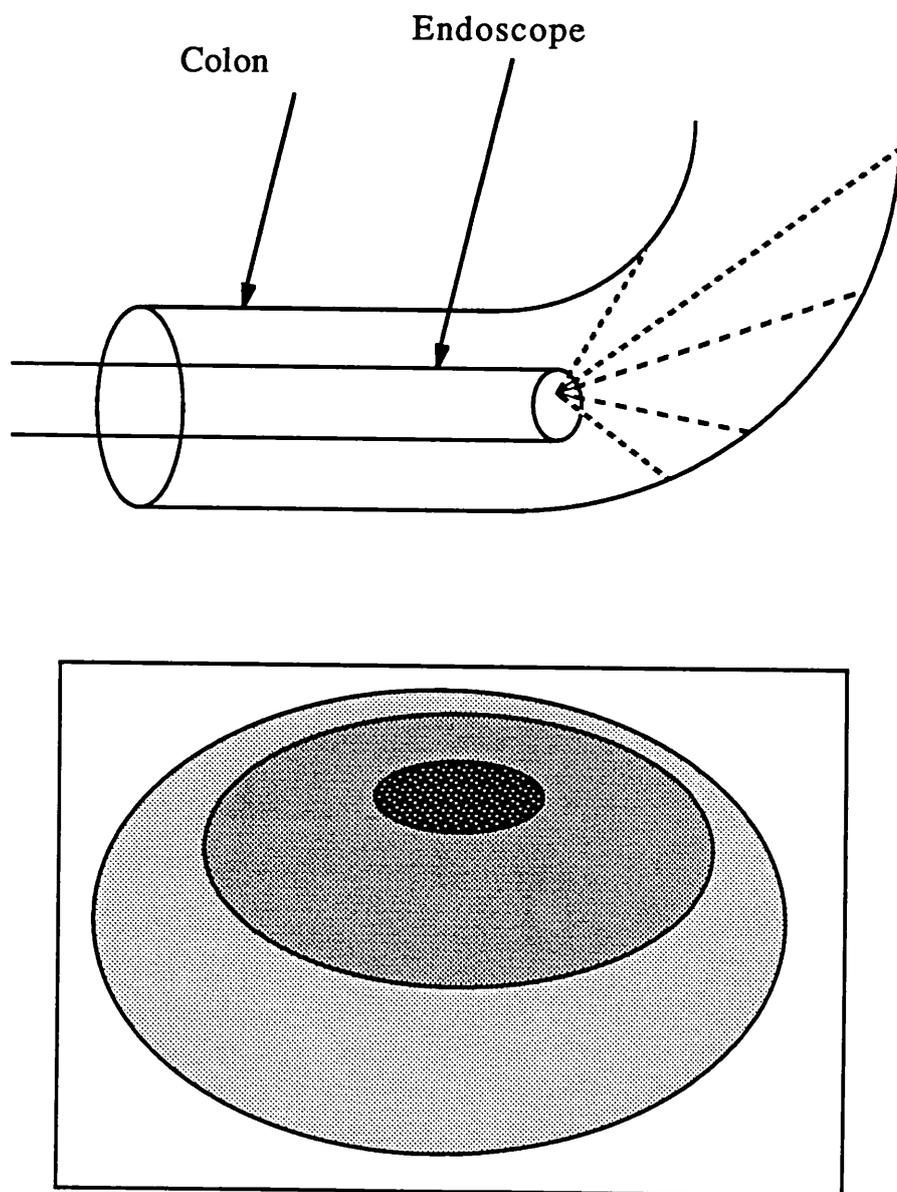
Since the light source and camera are almost at the same location, the deepest area in the colon with respect to the viewer corresponds to the darkest region in the image. This is illustrated in Figure 5.1 and 5.2. In the case of Figure 5.1, the darkest part is in the centre. Therefore the endoscope is moved straight forward while keeping its tip controlled in the previous direction. For Figure 5.2, the darkest part of the endoscopic image is in the upper part, therefore the endoscope must be advanced in the upward direction and the tip direction is controlled accordingly.

### 5.3 Region Extraction Techniques

In region extraction the image is divided into sub-regions on the basis of their properties including intensity, colour, texture, or range. Zucker [1976] has written an excellent survey on different region extraction techniques. These techniques can be categorised as merging, splitting, and split and merge.

#### 5.3.1 Region Merging

Merging starts with a uniform region of one or more properties. In the simplest case this region can be a single pixel. An attempt is made to enlarge the region by searching for the similar properties in neighbouring regions, one at a time. The whole process is sequential and the resulting region's shape may depend on the starting region or pixel, known as seed, and the direction of search.



*Figure 5.2: The endoscope approaching a bend.*

Different forms of the basic method have been successfully applied. For example, Muerle and Allen [1968] have used a three stage approach for region merging: firstly, the entire picture is segmented into square blocks (of size  $2 \times 2$ ,  $4 \times 4$  etc.); secondly, a statistical measure is determined for each of these regions, and finally the regional neighbour search method is used to merge the blocks of similar statistics. The estimate of the statistical measure for a region is updated after every merge operation, which makes it an accurate description of regions. Brice and Fennema [1970] use the so called atomic regions of constant grey level to start with, then these atomic regions are merged by applying successively *phagocyte* and *weakness* heuristics. These heuristics use the properties of the edge boundaries between regions. The adjacent regions are merged if the boundary between them is weak and the resulting region has a shorter boundary than the original regions. Pavlidis [1972] has presented an algorithm which divides the image into one-dimensional thin strips. These strips are then segmented into a small number of partitions using an approximation method. Merging is then performed utilising the partitions of every strip. Levine and Shaheen [1981] have grown regions by merging as many adjacent pixels as possible based on the colour features. The threshold for merging is adapted according to the coherence of regions for limiting the growth of less uniform regions.

### 5.3.2 Region Splitting

This is the opposite approach to merging for image segmentation. Splitting starts with the whole image which is considered as one region. It is then divided successively into smaller regions until each smaller region satisfies the uniformity criterion. Normally the histogram of a coherent or uniform region is unimodal. Therefore when a region has a multimodal histogram, an attempt is made to partition it in such a manner that the histograms of resulting regions are unimodal.

Like merging, splitting has been used successfully as the basis of several algorithms. For example, Robertson et al. [1973] have

developed an algorithm for partitioning multi-spectral images recursively. Their criterion for region uniformity, known as G-regularity, is based on the mean vector of grey levels. The algorithm continues to subdivide blocks until a sub-region is found whose mean grey level does not differ from that of its parents. The histogram information has also been used to determine a threshold for separating an object from its background. The same idea is utilised in a more general sense for region splitting. Tomita et al. [1973] have calculated histograms on the basis of texture properties, while Ohlander et al. [1978] have employed nine colour attributes. These authors have suggested a recursive region splitting technique. By using histograms of several different feature values of a given region, a threshold in one of the features can be used to split the region into sub-regions. This process is repeated on each split region until all their histograms are unimodal.

### 5.3.3 Pyramid Based Techniques

The first major algorithm to employ a pyramid was due to Horwitz and Pavlidis [1974]. They employed the principle of split and merge for segmentation. The split and merge process starts from any given partition of the image. The adjacent regions are merged if they satisfy the uniformity criterion and a single region is split, if it is not sufficiently uniform. The process continues until there are no regions to merge or split. For a given image, the pyramid is constructed whose nodes correspond to square regions and whose leaves represent single pixels. Each node has also an associated value attached to it which is the maximum and minimum brightness functions of the corresponding block. The algorithm begins with an arbitrarily chosen cut-set of the nodes (partition of the image), which is subsequently refined by splitting and merging. Merging of four nodes is performed by removing them from the cut-set and replacing them with the single node. The pyramidal data structure is abandoned by a grouping process which follows the split and merge. The segmentation is further improved by combining small regions with their neighbours and

by merging similar regions of different sizes by using an adjacency graph structure.

Other researchers have also employed different pyramidal structures for image segmentation [Burt et al. 1981, Hong and Rosenfeld 1984]. An overlapped pyramid, defined by 4×4 block averaging, has been employed successfully for image segmentation. Each block has four parents and sixteen children. In this way the blocks overlap by 50% on all four sides. The links between adjacent levels are weighted and their strength is adjusted by recalculating their values iteratively. This process is continued until the link strength shows no further changes. The links which remain intact provide the sub-trees in the pyramid and the leaves of each sub-tree represent a homogeneous region in the image [Hong and Rosenfeld 1984]. In a simpler approach, each block is linked to that of its parents whose average grey level value is closest to its own [Burt et al. 1981]. The values associated with each block are recomputed by averaging only those blocks on the level below that are linked to it. The links are changed based on these new values, if necessary. This process is repeated until there are no further changes. Pietikainen et al. [1982] have suggested different techniques for splitting and linking the overlapped pyramid. The pyramid structure has also been used for determining thresholds for blob detection [Shneier 1983], and more recently Blanford and Tanimoto [1988] have reported a variety of bright-spot detection techniques.

#### **5.4 The Pyramid Structure for Region Extraction**

Region based segmentation techniques which employ a two-dimensional pyramid based representation are the fastest currently available. They are also the most cost effective since, with recent developments in VLSI and ULSI technologies, memory is cheap and it is unnecessary to save space for the representation selected. The important features of a representation are the facility for computationally efficient operations and ease of access to arbitrary regions. The most common operation is the searching

for suitable seeds from the entire image for region growing. We have found that these operations are best supported by the pyramid structure. The root of the pyramid represents the entire image and any required region can be reached from the root in a few steps. The pyramid based image representations are based on the organisation of the interior of a region and can be categorised as a collection of maximal blocks which partition a given region of the image. They do not follow strictly the maximal block representation and their blocks are disjoint and have a standard size and location depending on the level of the tree. This provides a systematic hierarchical way of representing squared regions in the image.

A pyramid architecture computer in its most basic form consists of successively smaller planes of processing elements stacked over one another in such a way that the largest plane is at the bottom and the smallest which consists of a single processing element at the top. These planes are linked via a tree (e.g. quadtree). The internal links of each plane provide efficient implementations of parallel and local operations, while the logarithmic based connections between planes of consecutive levels reduce the distance of order  $n$  to  $\log_2(n)$ . The pyramid based techniques can be simulated using quadtrees on ordinary sequential computers. Therefore the specially tailored image partitioning methods, which utilise quadtrees, can be implemented on pyramidal computers in parallel. The methods described in this chapter are best suited for parallel implementation, specifically the step concerning the extraction of seed regions.

In quadtrees the image representation is based on the successive subdivision of image into quadrants. A quadtree is represented in the memory by a tree of outdegree four, where the root represents the whole image of  $2^N \times 2^N$  pixels. The pyramid based on a quadtree is made of  $N+1$  planes stacked one over the other, with the original image at the base and the root of the tree at the top. For a plane of size  $2^n \times 2^n$ , the level  $L$  is given by,  $L=N-n$ , and each pyramid plane is denoted by  $g_L$  as shown in Figure 5.3.

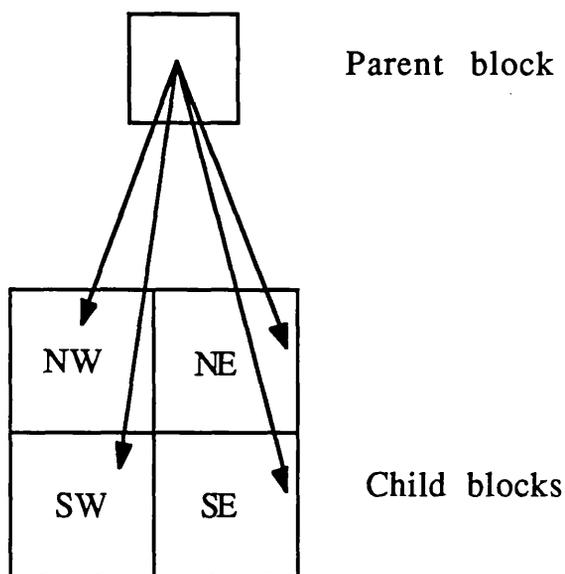
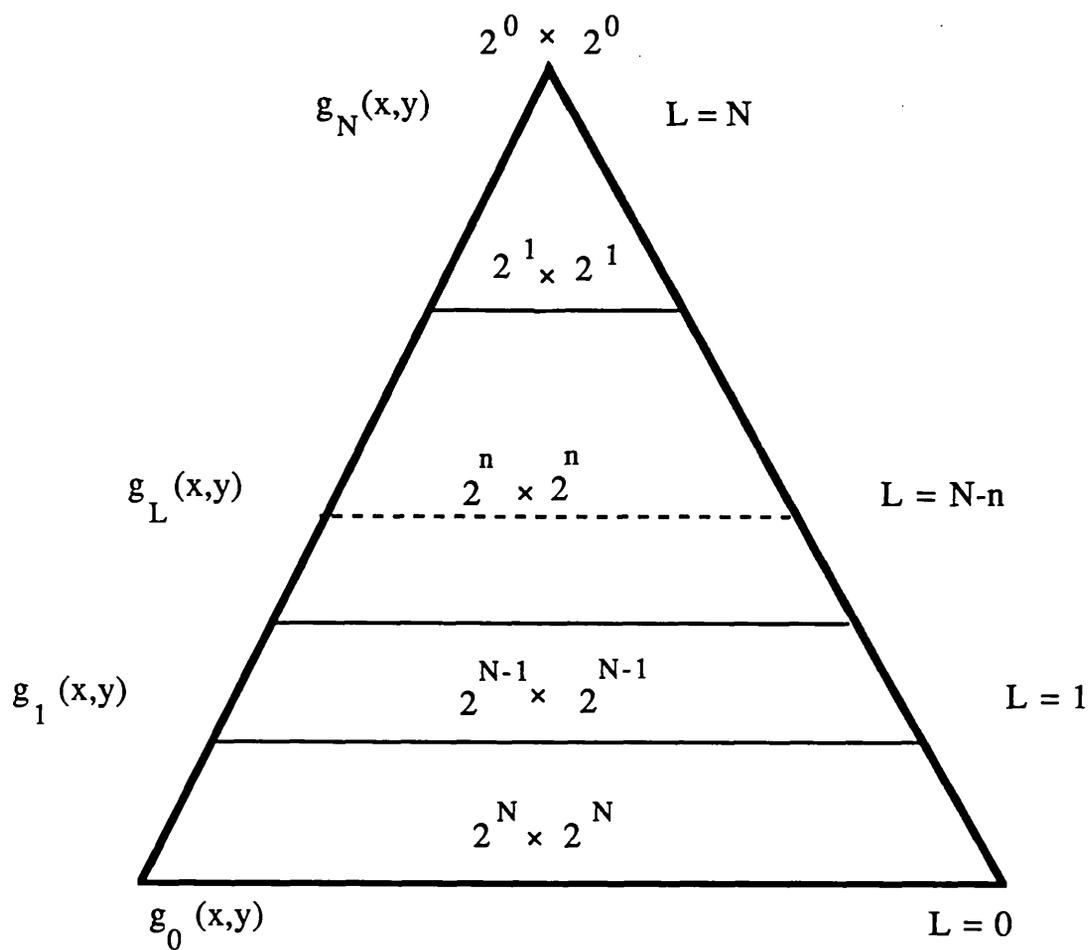


Figure 5.3: The quadtree based pyramid structure

For a given level  $k$ , the plane at level  $k+1$  is constructed by employing one of a possible set of transformations. If  $g_{k1}$ ,  $g_{k2}$ ,  $g_{k3}$ , and  $g_{k4}$  represent an intensity or other property of four quadrants on which the parent node  $g_{k+1}$  is based, Then

$$g_{k+1} = T\{g_{k1}, g_{k2}, g_{k3}, g_{k4}\} \quad (5.2)$$

Where  $T$ , is a particular transformation. One simple example of a transformation is the average transformation  $T_{ave}$ , which is:

$$T_{ave} = [g_{k1} + g_{k2} + g_{k3} + g_{k4}]/4 \quad (5.3)$$

It sets  $g_{k+1}$  to the mean of its four son nodes at level  $k$ . We have employed both the average transformation,  $T_{ave}$  and a variance based transformation. Other possible transformations are,  $T_{min}$  which selects the minimum, or  $T_{max}$  which selects the maximum of some property of the four regions. The typical properties can be one or a set of different histogram features like standard deviation, variance, skewness, energy, or entropy. The transformation based on variance,  $T_{var}$  is formulated and used, as far as we know, for the first time in our algorithm.

## 5.5 Detection of Dark Regions

### 5.5.1 The Algorithm

We now introduce the algorithm which uses a variance-average pyramid structure for detecting a dark homogeneous region in an image. The method, which has been developed from an initial version [Khan and Gillies 1988a], has three distinct steps.

#### *Step 1: Estimation of Intensity for Darker Region*

The darker areas of colon images which correspond to the deepest and obstacle free part in the colon, are often clearly visible. The first peak in the intensity histogram can be used as an estimate of the average grey level of the area desired. An intensity histogram

is constructed for the given colon image and then a range of grey levels, surrounding the first peak in the histogram, is estimated (see Figures 5.5 to 5.8). The first peak is taken to be the maximum below an intensity level of 80.

*Step 2: Detection of Seed Region*

An intensity variance-average pyramid is constructed by starting from the pixel level and averaging the pixels in each square group of four to produce an image of half the resolution. The  $T_{ave}$  transformation, described in the previous section, is used. In addition to the mean grey level, variance is also calculated recursively for every square region corresponding to a particular node in the quadtree. Using the terminology of previous section, for variances  $v_{k1}$ ,  $v_{k2}$ ,  $v_{k3}$ , and  $v_{k4}$  of the son nodes at a given level  $k$ , the variance at the parent level,  $v_{k+1}$  is calculated using:

$$v_{k+1} = T_{var}\{\mu_{k1}, \mu_{k2}, \mu_{k3}, \mu_{k4}, v_{k1}, v_{k2}, v_{k3}, v_{k4}\} \quad (5.4)$$

where  $\mu_{k1}$ ,  $\mu_{k2}$ ,  $\mu_{k3}$ , and  $\mu_{k4}$  are the mean intensity of child nodes at level  $k$ .

and

$$T_{var} = [v_{k1} + v_{k2} + v_{k3} + v_{k4} + \mu_{k1}^2 + \mu_{k2}^2 + \mu_{k3}^2 + \mu_{k4}^2]/4 - \mu_{k+1}^2 \quad (5.5)$$

where  $\mu_{k+1} = [\mu_{k1} + \mu_{k2} + \mu_{k3} + \mu_{k4}]/4$

For the one level above the bottom level of the pyramid, variance is calculated directly from the pixel intensities of the given image.

The proof that Equation (5.5) does correctly calculate the variance, can be done by induction. Starting with the definition of variance as:

$$v = \sum_{i=1}^n (x_i - \mu)^2 / n \quad (5.6)$$

expanding the squared term and rearranging we get:

$$v = \sum_{i=1}^n x_i^2 / n - \mu^2 \quad (5.7)$$

In the base case we are dealing with single pixels, and thus by definition:

$$\begin{aligned} v_{01} = v_{02} = v_{03} = v_{04} &= 0 \\ \mu_{01} = x_1; \mu_{02} = x_2; \mu_{03} = x_3; \mu_{04} &= x_4; \end{aligned}$$

where  $x_1, x_2, x_3,$  and  $x_4$  are the individual pixel intensities.

Thus, using equation (5.5) we have that:

$$\begin{aligned} v_1 &= (x_1^2 + x_2^2 + x_3^2 + x_4^2)/4 - ((x_1 + x_2 + x_3 + x_4)/4)^2 \\ &= \sum x_i^2 / n - \mu^2 \end{aligned}$$

as required.

Assuming that the result is true for  $n$ , we use equation (5.5) to perform the inductive step giving:

$$v_{n+1} = (v_{n1} + v_{n2} + v_{n3} + v_{n4} + \mu_{n1}^2 + \mu_{n2}^2 + \mu_{n3}^2 + \mu_{n4}^2)/4 - \mu_{n+1}^2 \quad (5.8)$$

Now since for any sub-tree at level  $n$  in the pyramid there are  $(2^n)^2$  pixels, we have that:

$$\begin{aligned} v_{n1} + \mu_{n1}^2 &= \sum_{[x_i \in \text{sub-tree } n1]} x_i^2 / (2^n)^2 - \mu_{n1}^2 + \mu_{n1}^2 \\ &= \sum_{[x_i \in \text{sub-tree } n1]} x_i^2 / (2^n)^2 \end{aligned} \quad (5.9)$$

And by substituting it in equation (5.8) we have:

$$v_{n+1} = \left\{ \sum_{[x_i \in \text{sub-tree } n1]} x_i^2 / (2^n)^2 + \sum_{[x_i \in \text{sub-tree } n2]} x_i^2 / (2^n)^2 + \sum_{[x_i \in \text{sub-tree } n3]} x_i^2 / (2^n)^2 + \sum_{[x_i \in \text{sub-tree } n4]} x_i^2 / (2^n)^2 \right\} / 4 - \mu_{n+1}^2$$

giving:

$$v_{n+1} = \sum_{[x_i \in \text{sub-tree } n+1]} x_i^2 / (2^{n+1})^2 - \mu_{n+1}^2$$

as required.

The calculation of intensity mean and variance is formulated in such a manner that each node uses only the intensity mean and variance of its children. During this process, it is also determined whether the area corresponding to a particular node is uniform by comparing its variance to a uniformity threshold,  $v_{thr}$ . The uniformity is a function of both intensity mean and variance of the region. A suitable heuristic law for combining both properties into one is:

$$\text{Uniformity} = [ 1 - v_k / \mu_k^2 ] \quad (5.10)$$

This region uniformity has been specifically used in the region growing step for adapting the user supplied average grey level tolerance,  $M_{usr}$  between adjacent regions to avoid over merging. A similar approach has been employed by Levine and Shaheen [1981] in their region growing method.

A special link belonging to each node, the  $u\_link$  is used to indicate the seed for the darker region in the sub-tree of that node in the pyramid. If the node mean grey level is within the grey level range estimated in step 1 and its corresponding region also satisfies the uniformity criterion, it is labelled as the seed for the whole of its sub-tree by connecting its  $u\_link$  to itself.

Otherwise, following steps are carried out to determine the largest dark and uniform region in the sub-tree.

-The four children of a given node are examined and the one which has the largest seed region in its sub-tree is selected. The seed area for that child is also identified as the seed area of the parent by assigning its `u_link` value to the parent `u_link`. An example showing several `u_links` is shown in Figure 5.4. If the seeds provided by more than one child are of equal size then that seed is selected which is more uniform and whose grey level is comparatively near to the grey level range estimated in Step 1.

-When none of the children have any seed region in their sub-trees, their `u_links` will point to a pre-defined location (e.g. NULL in C Language) and the parent `u_link` is also assigned to NULL.

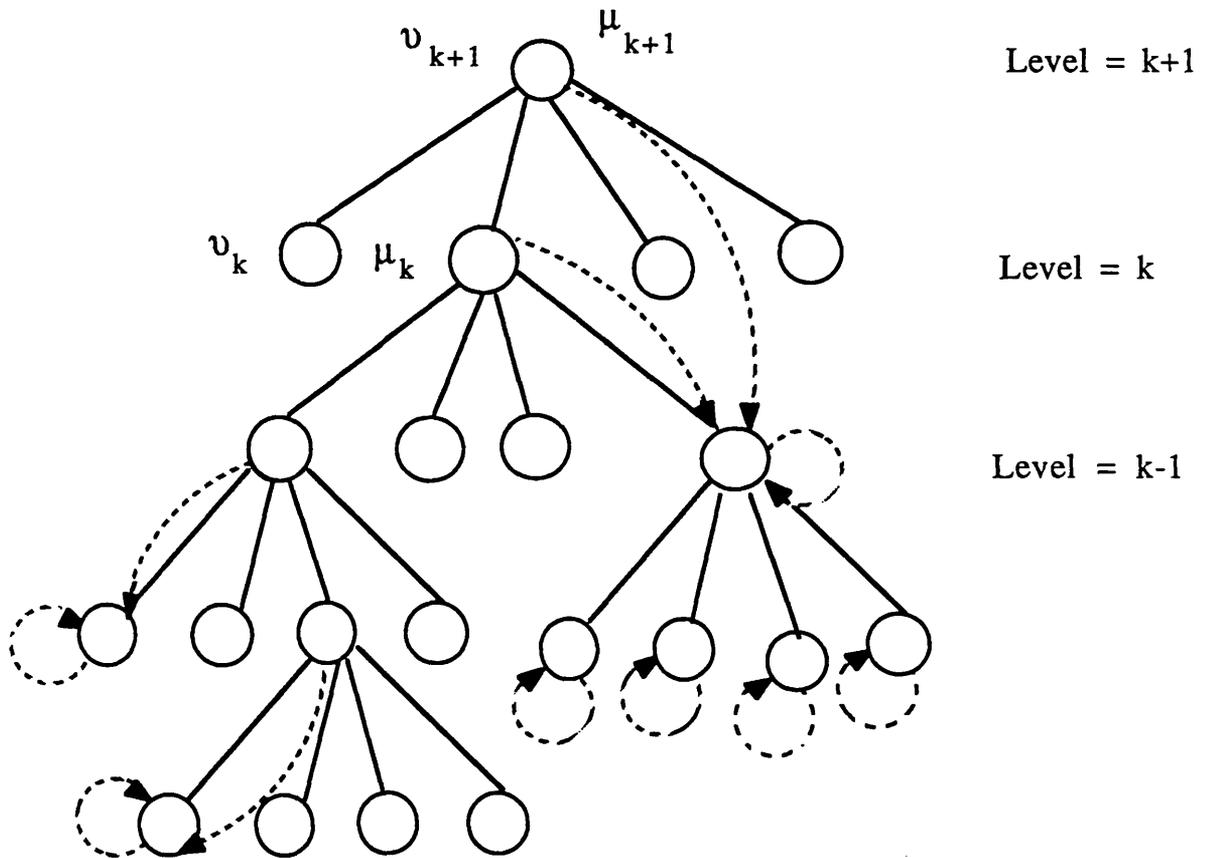
Therefore in a single pass when the pyramid is completed and the root of the pyramid is reached, the `u_links` of each node in the pyramid points to a uniform and dark square region in its sub-tree. The `u_link` for the root of the pyramid provides the address of the largest uniform and darker square region in the whole image. This region may itself be large enough to determine the insertion direction of endoscope in the colon, or for a more general case it can be used as a seed to extract a complete darker region by the region growing process of Step 3.

### *Step 3: Region Growing*

After successfully identifying the seed region, adjacent areas of the seed are considered for merging if they satisfy the following two conditions.

- (i) The intensity mean is within the grey level range estimated in step 1.

(ii) The intensity variance does not violate the uniformity criterion.



*Figure 5.4: A part of an intensity variance-average pyramid, explaining the  $u$ \_links which are shown in dotted lines.*

The neighbouring areas of the seed are examined and merged if their average intensity is equal to the intensity of seed (within some tolerance,  $M_{thr}$ ) and their variance is below the variance threshold,  $v_{thr}$ . The merging threshold in intensity mean for two adjacent regions,  $M_{thr}$  is adapted from the user supplied

threshold,  $M_{usr}$  depending on the uniformity of the merged region. The threshold value,  $M_{thr}$  decreases with the decrease in uniformity or increase in the ratio  $[v_k/\mu_k^2]$ . Therefore the growing of less uniform regions are restricted and user need to supply only a constant threshold,  $M_{usr}$ .

$$M_{thr} = [ 1 - v_k/\mu_k^2 ] M_{usr} \quad (5.11)$$

The merging process continues for all the neighbours of newly attached regions until there is no neighbour left which satisfies the uniformity and similar grey level criteria. The merging can be performed either using only those nodes of the pyramid which are at the same level, in which case the regions of equal area are connected, or by using nodes lower in the pyramid for extracting regions with more accurate boundaries.

### 5.5.2 Implementation Details and Experimental Results

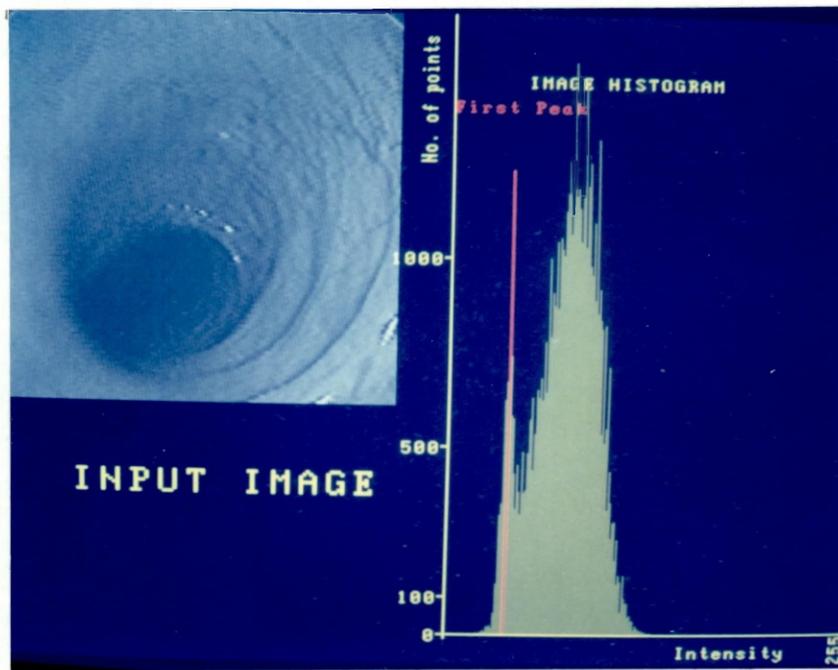
Work has already been reported for converting different two-dimensional image representations to quadtrees such as binary arrays [Samet 1980]. The method to build a quadtree from a binary image has been modified here, to deal with an image array of 256 grey levels and to compute mean and variance of intensity values for each node. Each node of the tree structure corresponds to a squared region in the image and contains four types of fields.

- The intensity mean of the corresponding region.
- The intensity variance of the region which defines its uniformity and cohesiveness.
- The relation of the node to the decomposition of its parent, which may be one of the set NW, NE, SW, or SE as shown in Figure 5.3.
- Four links (pointers) for its children and one to its parent, and an additional special link,  $u\_link$ , which points to the the largest uniform and dark region

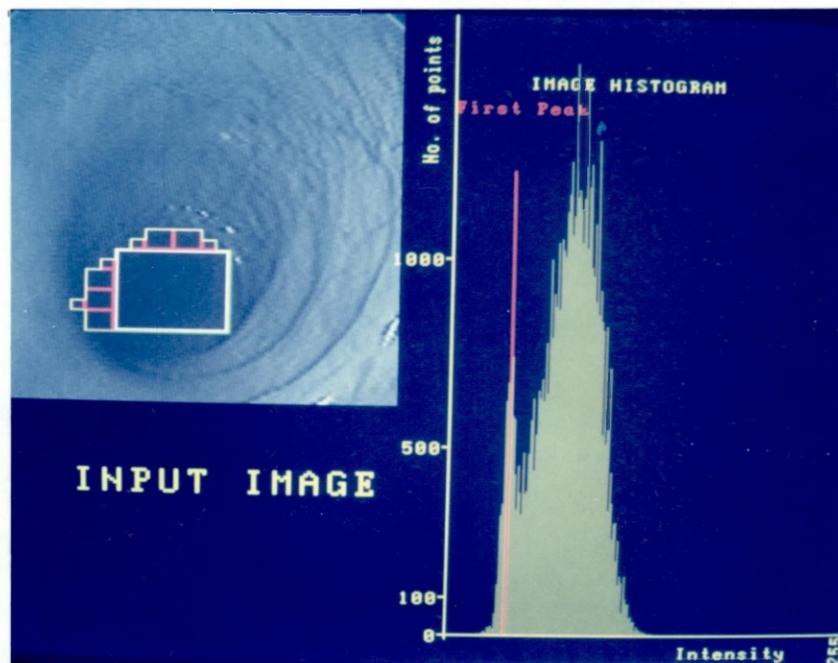
in the sub-tree of the node. If there is no uniform area, the `u_link` points to a NULL.

The algorithm has been implemented on an 80386 based host for an Imaging Technology series 151 Image Processor using XENIX operating system. The image processor grabs the image from a video recorder and constructs the intensity histogram at the video rate. The first peak in the histogram is determined using the host computer, where the pyramid structure is also constructed for the given image. The method has been employed for a number of typical endoscopic images taken inside a human colon. The chosen pictures of human colon are representative of 70 to 80% of the whole class of colon images encountered during colonoscopy. These pictures are digitised using 256 grey levels and contain a lot of noise, particularly due to specular reflections.

Four selected pictures, along with their histograms, are given in Figures 5.5a, 5.6a, 5.7a, and 5.8a. The first significant peak in the histogram is indicated by a red vertical line. The extracted regions are shown in Figures 5.5b, 5.6b, 5.7b, and 5.8b in the form of a group of squared areas. In this implementation the pyramid is only built from the second level upward but uniformity of the regions is not sacrificed in this process since the variance is calculated for every  $4 \times 4$  region and stored with the leaves of the tree. The seed regions, shown in sharp white boundaries, are detected by setting the variance threshold,  $v_{thr}$  at 100. The uniformity threshold is also varied from 80 to 250 and it is observed that for this range the detection of dark regions, for the same images, is not affected. In the merging step, the user supplied average grey level tolerance between the seed and merged regions,  $M_{usr}$  is set at seven. The merged regions have the red boundaries while the boundary of the detected region is shown in a yellowish green colour. The dark region extraction time was within eight seconds in most of the colon images we have tested. The overall boundary of the darker region is quite rough because the merging of regions is performed only down to the third level. The boundaries of the extracted regions will be smoother and more accurate when the pyramid is built from pixel



(a)

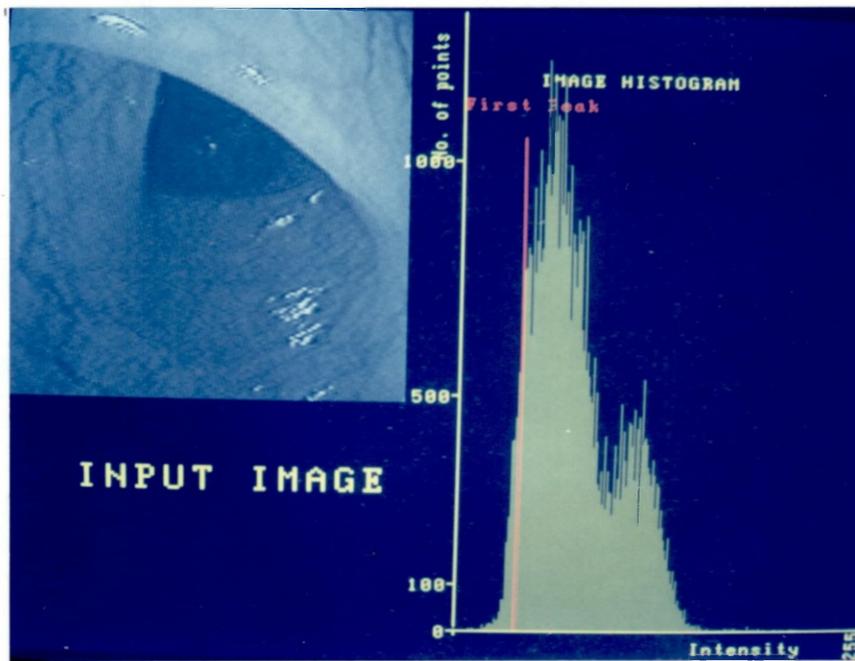


(b)

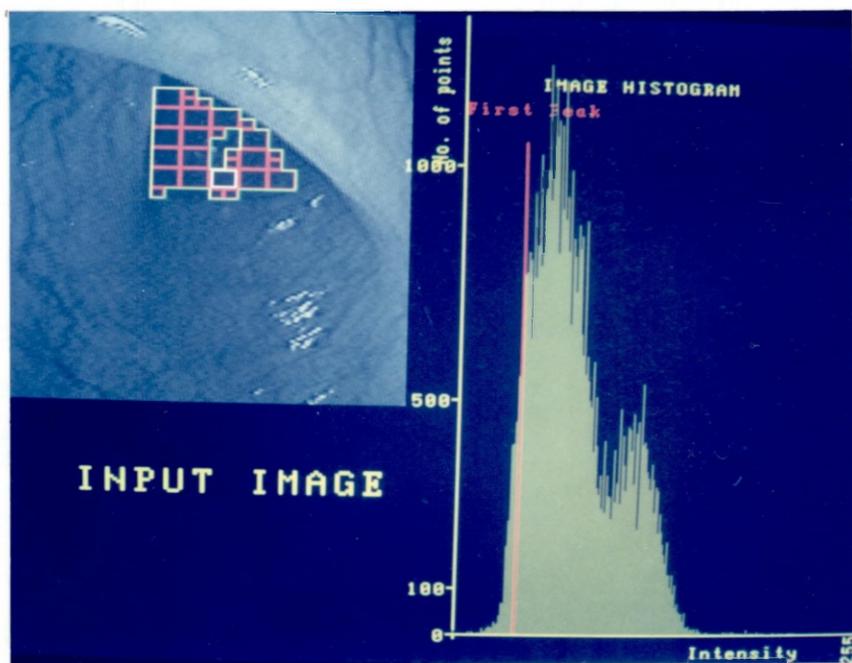
Figure 5.5: Dark region extraction in the first colon image.

(a) The image and its histogram.

(b) The seed and neighbouring merged regions.



(a)

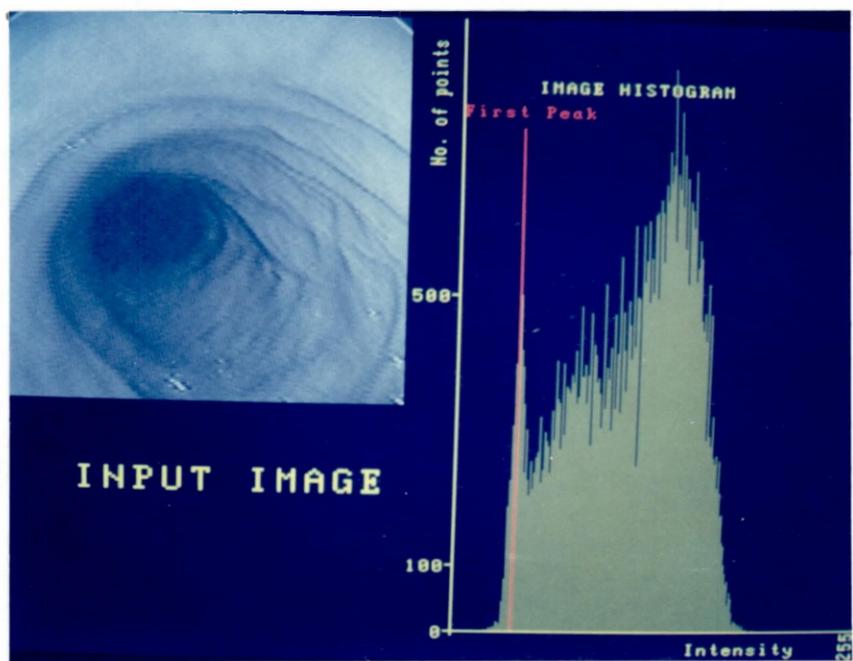


(b)

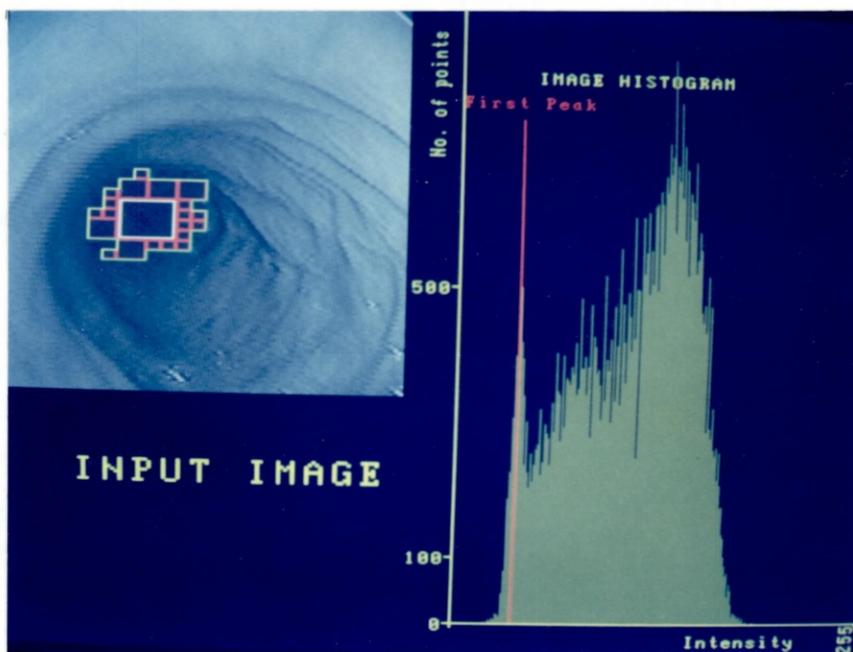
Figure 5.6: Dark region extraction in the second colon image.

(a) The image and its histogram.

(b) The seed and neighbouring merged regions.



(a)

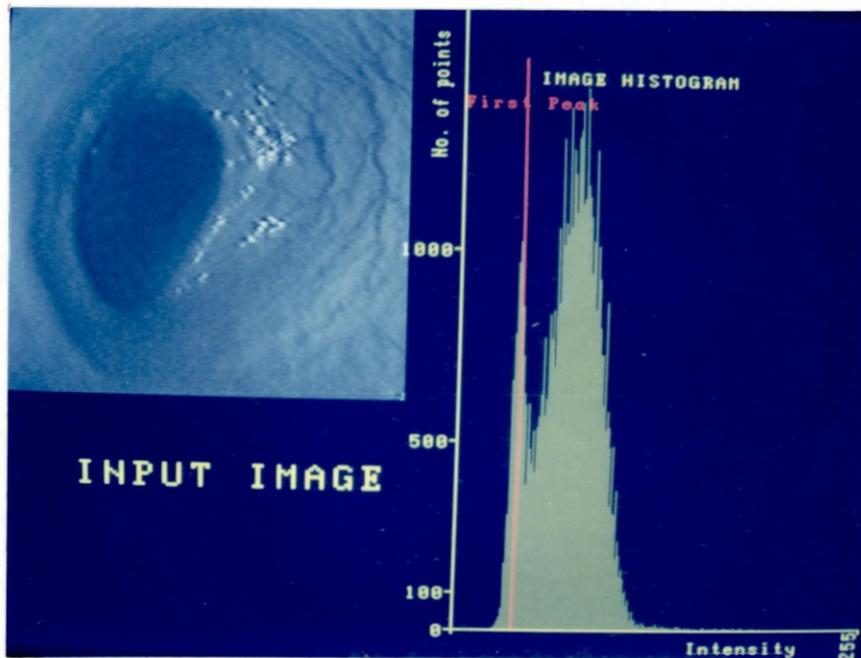


(b)

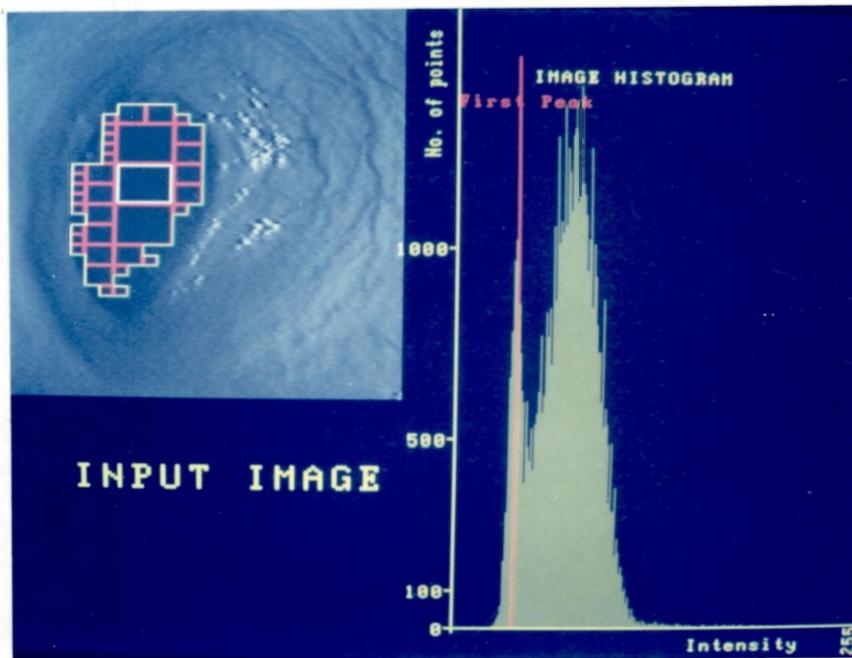
Figure 5.7: Dark region extraction in the third colon image.

(a) The image and its histogram.

(b) The seed and neighbouring merged regions.



(a)



(b)

Figure 5.8: Dark region extraction in the fourth colon image.

(a) The image and its histogram.

(b) The seed and neighbouring merged regions.

level and the merging operation is also performed down to the pixel level at the cost of a large amount of additional processing time.

## 5.6 Extension to Image Partitioning

The method to detect darker region in an image can easily be modified for extracting seed regions on the basis of uniformity and cohesiveness of image intensity or any other property. These seed regions can then be grown to uniform sub-regions in an image. It is worth pointing out that the image partitioning method described in this section gives excellent results for partitioning piece-wise homogeneous images.

### 5.6.1 Image Partitioning into Uniform Regions

For detection of seed regions to partition an image, the condition whether a seed is dark or not, is not required. Therefore no histogram is constructed. The method has two distinct steps: firstly, the intensity variance-average pyramid is constructed similar to the darker region extraction and secondly, a top-down pass in the pyramid is performed for the region growing process.

#### *Step:1 Establishing Seeds for Uniform Regions*

The intensity variance-average pyramid construction is achieved by following step 2 of the dark region extraction algorithm. Only the uniformity criterion (variance threshold,  $v_{thr}$ ) is employed in determining the seed region node for each sub-tree of the pyramid. If the node satisfies the uniformity criterion, it is labelled as the seed for the whole of its sub-tree by connecting its `u_link` to itself. Otherwise, the four children of the node are examined and the one which has the largest seed region in its sub-tree is selected. The seed area for that child is also identified as the seed area of the parent by assigning its `u_link` value to the parent `u_link` as explained in Figure 5.4.

If none of the children have any seed region in their sub-trees, their `u_link` point to a pre-defined location (e.g. NULL) and the parent `u_link` is also assigned to NULL. Therefore in a single bottom-up pass, when the pyramid is completed, the `u_link` of every node in the pyramid points to a uniform square region in its sub-tree.

### *Step 2: Parallel Region Growing*

After constructing the variance-average pyramid, the nodes of every plane in the pyramid are treated as sub-trees and each sub-tree has a distinct seed region which is grown in this step. Starting from the top of the pyramid, a top-down pass is initiated and for each plane of the pyramid following two steps are performed.

-For each node in a plane of the pyramid, a seed region is reached by testing its `u_link`. If the seed region exists, it is grown to form a complete uniform region. The neighbouring areas of the seed are examined and merged if their average intensity is equal to the intensity of the seed (within some tolerance,  $M_{thr}$ ) and their uniformity is below the variance threshold,  $v_{thr}$ . The merging threshold in average intensity for two adjacent regions,  $M_{thr}$  is adapted from the user supplied threshold,  $M_{usr}$  depending on the uniformity of the merged regions as given in equation (5.11). The merging process continues for all the neighbours of the newly attached regions until there is no neighbour left which satisfies the uniformity and similar grey level criteria. The merging can be performed by using nodes at the seed level, in which case the regions of equal area are connected, or by using nodes at lower levels in the pyramid for extracting regions with more accurate boundaries.

-All the nodes of the sub-trees of the seed region and merged regions in the merging process are

flagged as processed by assigning their `u_links` to a pre-defined node (e.g. `NULL`). Similarly all the nodes above the level of these sub-trees whose `u_links` point to the seed or merged regions are also flagged as processed.

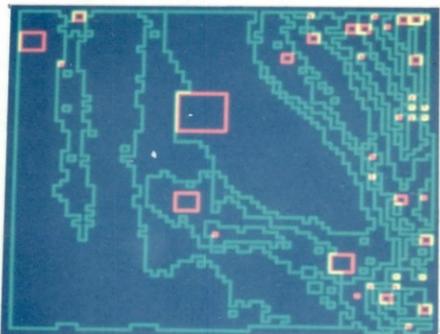
This process is continued for each plane of the pyramid until a plane in the pyramid is encountered from which no seed region is reached. If the `u_links` of all the nodes in a particular plane of the pyramid do not provide a seed region then the `u_links` of all the nodes below that plane also cannot provide any additional seed region. This also confirms that the image has been partitioned into uniform regions and there is no unpartitioned region left. The boundary of the regions can also be traced in parallel.

### 5.6.2 Results and Comments

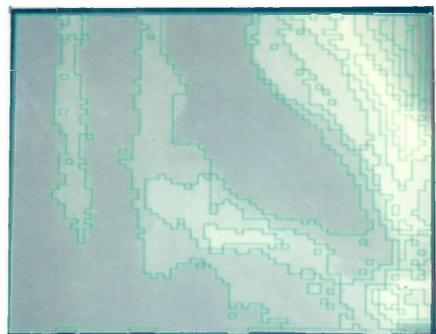
The region growing on the basis of seed regions provided by a particular plane can easily work in parallel. For the complete partitioning of an image two passes are required; the first pass is bottom-up which builds the variance-average pyramid and identifies seed regions for each node in the pyramid and the second pass, which is top-down, grows these seed regions into larger uniform regions. The method has been demonstrated on two heart ventricular images and a computer generated image which are shown in Figure 5.9a, 5.10a, and 5.11a. The pyramid in this case is also constructed from the level two and the merging is also performed down to level two. The uniformity threshold for a uniform region,  $\nu_{thr}$  is set at ninety for heart ventricular images and twenty for the artificial image. In the merging step, the grey level tolerance between the seed region and merged regions,  $M_{usr}$  is selected as seven. The extracted seed regions shown in red colour, along with the complete partitioned regions shown in green colour, for the test images are given in Figure 5.9b, 5.10b, and 5.11b. In Figure 5.9c, 5.10c, and 5.11c, the boundaries of the partitioned regions are superimposed on the test images. The same images are then partitioned by constructing the pyramid from level one upward and merging is also performed down to



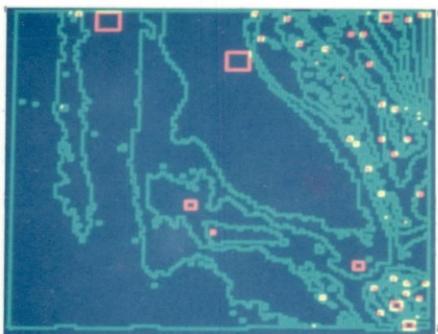
(a)



(b)



(c)

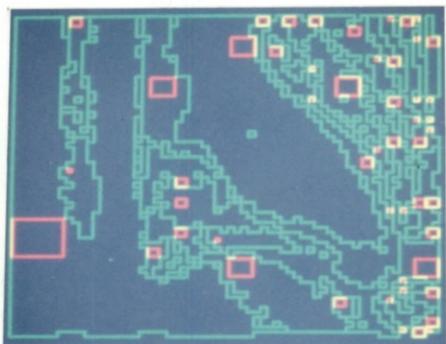


(d)

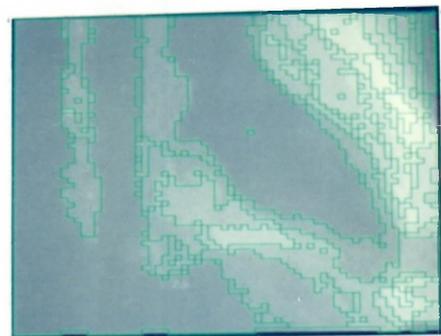
Figure 5.9: Partitions of the first heart ventricular image. (a) Heart ventricular image. (b) Seed regions and image partitions when the pyramid is built from level two. (c) Segmentation results overlaid on the image. (d) Seed regions and image partitions when the pyramid is built from level one.



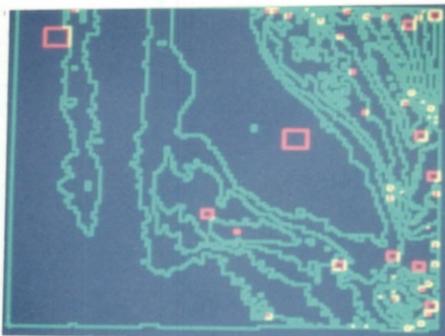
(a)



(b)

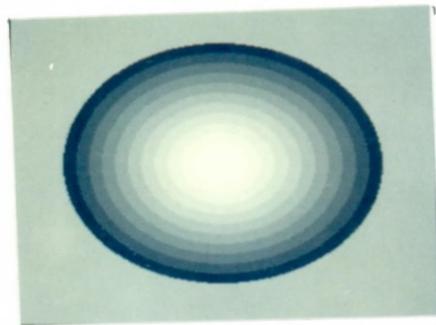


(c)

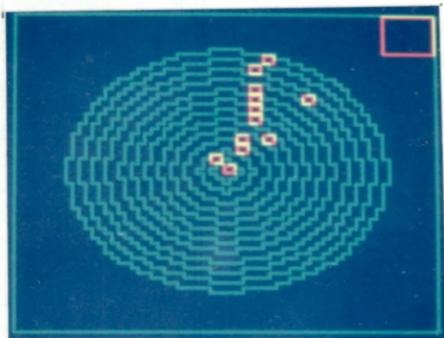


(d)

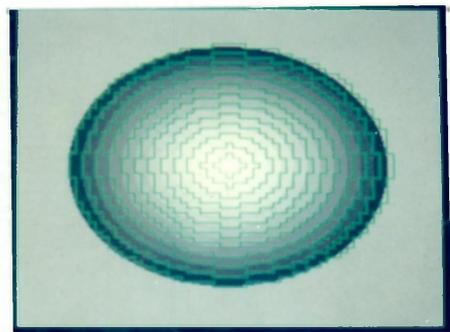
Figure 5.10: Partitions of the second heart ventricular image. (a) Heart ventricular image. (b) Seed regions and image partitions when the pyramid is built from level two. (c) Segmentation results overlaid on the image. (d) Seed regions and image partitions when the pyramid is built from level one.



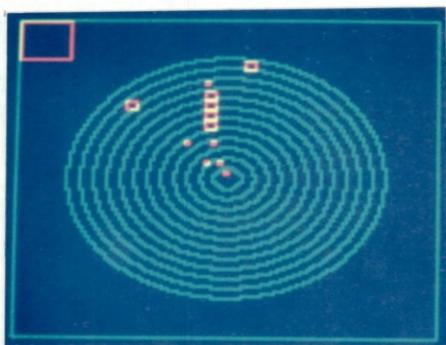
(a)



(b)



(c)



(d)

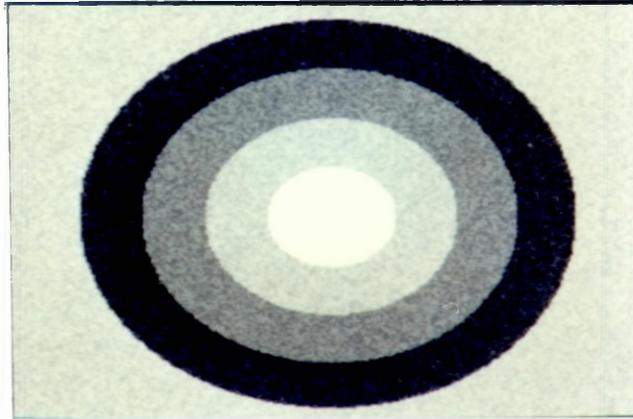
Figure 5.11: Partitions of the computer generated image. (a) The artificial image. (b) Seed regions and image partitions when the pyramid is built from level two. (c) Segmentation results overlaid on the image. (d) Seed regions and image partitions when the pyramid is built from level one.

level one. The results which are shown in Figure 5.9d, 5.10d, and 5.11d demonstrate more accurate region boundaries and image partitioning. The merging process can be extended to pixel level but it takes considerable amount of computing time. To demonstrate that the method works on images containing a large amount of noise, the artificial image of Figure 5.12a, which contains random noise amounting to  $\pm 22\%$  of the average image signal is segmented. The uniformity threshold,  $v_{thr}$  is raised to 300 for coping with noise. The partitioned image is shown in Figure 5.12b.

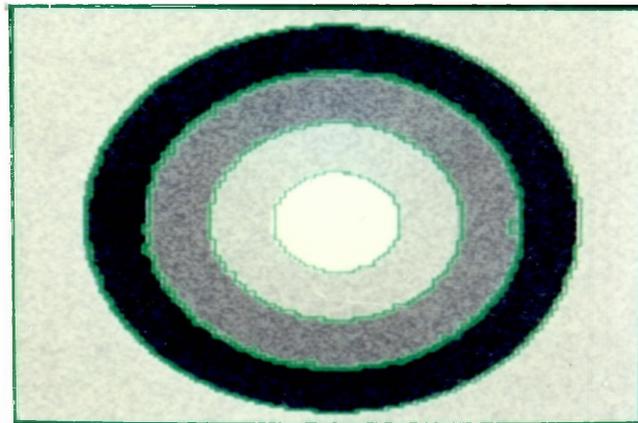
The image partitioning method is easy to modify for extracting a single region of given properties, without using the histogram information, as employed in the darker region extraction for endoscopic colon images. For bright region extraction the variance-average pyramid is constructed similarly but during the assignment of the value to the  $u\_link$ , the brightness test is also performed in addition to the uniformity. In the brightness test, the intensity mean of the seed regions of four children are compared and the seed with the higher intensity mean, which belongs to a higher level in the pyramid, is preferred. In this way the  $u\_link$  of the root provides the brighter seed region. A conflict in decision can arise between the brighter and larger size seed conditions. An additional threshold for the brightness test for adequate region size may be introduced to resolve this conflict. For darker region detection, instead of applying the brightness test the darkness condition may be used and that seed is selected, which is darker and of considerable size.

## 5.7 Parallel-Serial Region Extraction: Discussion

Fast detection of global structure from digital images is an essential component of real-time machine vision. The real-time performance of human perception on complex images indicates that our visual system uses highly evolved parallel processing. We have already discussed the close relationship between the



(a)



(b)

*Figure 5.12: Partitions of the artificial image with added noise.*

*(a) Artificial image containing  $\pm 22\%$  random noise.*

*(b) Region boundaries overlaid on the image.*

pyramidal computers and organic brain architecture in the second chapter.

On the basis of pyramid architecture, a number of specially tailored parallel computer architectures are beginning to emerge. The first version of the PAPIA chip, which has five processing elements, was developed at Rome University [Cantoni and Levialdi 1987]. The second version having three levels with twenty one processing elements is also being developed. The GAM pyramid, which is also a multi-layer pyramid structure, consists of 341 general and special purpose processing elements [Schaefer et al. 1987]. The GAM has five levels and it can implement a  $16 \times 16$  pixel array. Each of its processing elements has nine connections, which are grouped into four to the children, four to the neighbours, and one to the parent. The pyramid machine described by Tanimoto et al. [1987] uses a special VLSI chip in which each processing element can communicate to thirteen elements. They have employed eight connections to neighbours at the same level instead of the four used in GAM. The pyramid structure is implementable using  $(4^{n+1} - 1)/3$  (1, 5, 21, 85 etc.) processors. These developments in pyramid computer architecture could provide a true parallel implementation of our depth estimation and image segmentation techniques, and are capable of producing a real-time performance for endoscope navigation.

We can easily compare the performance of our algorithm with the earlier work on pyramid base segmentation. The split and merge method of Horwitz and Pavlidis [1974] does not make use of the pyramid structure for the whole process. They start from an arbitrary image partition and sometimes a large number of iterations may be required for segmentation. Moreover their method uses an adjacency graph structure for merging to achieve good segmentation results. The overlapped pyramid techniques [Burt et al. 1981, Hong and Rosenfeld 1984] are comparable but their iterative scheme for recomputation of links make these methods less applicable in real-time. The number of iterations required to stabilise the links increases with the complexity of the image. In our method, the image segmentation involves only two

passes on a pyramidal computer and does not depend on the complexity and busyness of the image. The bright-spot and blob detection methods are comparable to our dark and bright region detection techniques [Shneier 1983, Blanford and Tanimoto 1988]. In the blob detection only the thresholds are detected by using the pyramid and in bright-spot detection two types of pyramid structures have been used. One is based on the maximum transformation,  $T_{\max}$  while the other uses a mixture of average and maximum transformations. The authors are still not sure whether the averaging process will keep the bright spot in tact or not. Our method is more general and has the advantage that it can detect any type of regions including bright, dark, and other regions of given property.

## 5.8 Concluding Remarks

We started with the objective of investigating some parallel and fast way of estimating depth in colon images for automatic control of the endoscope tip. The world and illumination model for endoscopy ensures that the deepest part in the colon corresponds to the lowest intensity region in the colon images. The tracking of these darkest regions in a series of images provides an important part of the information required for guiding the endoscope. The extraction of a single region of given properties in real-time is an important goal. This led us to the development of a region extraction method based on the pyramid structure. The technique is particularly applicable to a domain where the world is illuminated only by a single light source located near the object surfaces and viewer. In this case, the darkest region in the image represents deepest point in the scene. Detection of the darkest region can therefore be used for automatic navigation in pipes and ducts for a variety of instrumentation purposes. The same method could also be applied in warehouses and buildings with appropriate lighting and for robot guidance in tunnels and automated factory environments.

The method is so effective and successful that we have extended it to partition the images into uniform regions rather than detecting single regions of given properties. The pyramidal image partitioning techniques are most promising in terms of segmenting images in real-time. None of the existing techniques provide true parallel implementation and real-time performance. Some of them do not employ the pyramid structure throughout the partitioning process, while the others use iterative techniques. The method we have introduced does not involve any iterative scheme and is implementable in parallel. Two passes in the pyramid are required to partition an image into uniform regions. For detecting a single region (dark or bright) only one bottom-up pass is needed. The main feature of our method is the ease of its implementation on a pyramid architecture based computer, specifically the bottom-up step which detects seed regions. The region growing process which uses the seed regions provided by a particular plane of the pyramid is also implementable in parallel.

Currently the algorithm is being implemented using five transputers on the same XENIX based host to process a sequence of colon images, which the system receives from a video tape made during colonoscopy. The parallel implementation of the split and merge process has already been attempted using transputers [Mansoor and Sokolowska 1988]. The main restrictions of the split and merge technique are avoided in our algorithm, as there is no change of representation from a pyramidal tree to graph structure. Hopefully with the introduction of transputers and true parallelism, the region extraction time will be reduced from eight seconds to video rate. Therefore the tracking of lumen from on-line colon images, digitised from a U-matic video recorder, will be achieved. This set up provides a simulation facility for testing the automatic control of endoscope. In the final system, the information gathered during lumen tracking, will be used by the navigation system to generate different tip control commands for the pilot sub-module. This approach provides a safe way of testing the vision and control algorithms, before using a prototype on the patients. The proposed endoscope navigation and control system is described in detail in the next chapter.

## CHAPTER 6

### NAVIGATION OF THE ENDOSCOPE

#### 6.1 Introduction

The main results of this research are the development of two new machine vision techniques for identifying contours and darker regions, which we have already presented in the previous chapters. The contours, which are formed by inner colon muscles, and dark regions, which correspond to obstacle free areas in the colon, are the most important landmarks for navigating the endoscope. In this chapter, we are investigating the ways and methods to integrate contour and region extraction techniques with the endoscope navigation system. The endoscope navigation is not very similar to that of mobile robots or autonomous vehicles. Prior navigation techniques are explored in this chapter before proposing an hierarchical navigation control system for an automatic endoscope. The endoscope navigation system consists of three sub-systems: global planner, navigator, and pilot. Perception plays an important role in navigating the endoscope and the world model of colon relies mostly on the information provided by a single CCD camera. The colon is modelled by using the generalised cylinder, which facilitates depth estimation in colon from inner muscle contours by following Marr's [1977] work on occluding contours. An efficient data structure the Quad-List tree (QL-tree), has been developed for world and search space representation. The QL-tree representation is compatible with the machine vision techniques, described earlier and the scene primitives supplied by them can be easily added to the structure directly. The QL-tree is not only suitable for world representation in navigation and it has also been employed as a more general representation for ray tracing in computer graphics [Nicholls, Khan and Gillies 1988].

A navigation system usually has three main components, the sensing system which provides information about the local and global world, the planner to plan a path to a specified goal, and the controller which executes the planned path. In general the planned path should be flexible enough to be modified for coping with any unexpected changes. The role of sensing in a navigation system depends on the domain of the world in which the autonomous vehicle is supposed to navigate. In a finite world (e.g. factory environment), the navigation system can use a global model which consists of a plan of the domain, but the model can only be utilised for planning a path to the goal. During the execution phase, the navigation system has to use the expected local model of the world. Whatever the world domain is, a navigation system is required to monitor the motion during the execution phase to verify that the movement is according to the plan. The monitoring alone requires a good sensing system. Additionally for an unknown world, sensors provide information about the environment from which the updated world model is constructed. Sensors play a vital role in adaptive navigation as well as object identification. Moreover, they are the only link to the outside world which is changing all the time. The best general purpose navigation system is considered to be that of humans and we believe that one of the reasons behind our navigational ability is our access to a variety of sensing capabilities including vision, touch, and hearing which helps directly or indirectly during navigation.

## **6.2 Navigation Techniques: A Review**

A number of research groups have reported interesting results which are relevant to the navigation of mobile robots and autonomous vehicles. A review of their work provides a current picture of the research on navigation. Instead of describing the work of individual groups separately, we are presenting the previous work in terms of different components of the navigation system such as sensing, path planning, and obstacle avoidance.

### 6.2.1 Machine Perception for Autonomous Vehicles

The Jet Propulsion Lab. (JPL) robot was originally intended for use in remote planetary exploration [Thompson 1977]. The terrain model for its navigation is constructed primarily from its vision system. The main sensing unit consists of stereo cameras and a laser range finder, in addition to proximity and touch sensors for the different manipulator processes.

The perception in the case of *Hilare* [Giralt et al. 1979] is provided by a video camera for object recognition and a laser range finder for depth measurement. Both of these units are mounted on the same scanning platform which can rotate around two axes. *Hilare* is a mobile robot developed for research and investigation purposes at the LAAS laboratory in Toulous. In addition to the main perception system, ultrasonic proximity sensors have been employed to cope with unexpected obstacles. It navigates primarily by following side walls using its sonars. Recent progress in the perception for *Hilare* is presented by de Saint Vincent [1986], describing a scene analysis module, using stereo cameras and a laser range finder, and a dynamic vision module for robot position correction and tracking world features.

The sensing system of Stanford Cart [Moravec 1979, 1983] consists of a single video camera which slides horizontally to capture a nine-way stereo vision for scanning the environment. The system moves the camera horizontally in nine precise steps to get nine images of the scene. During this whole process the cart remains stationary. Then an interest operator is applied to one image for selecting a fixed number of candidate points. The depth estimation is performed by matching these feature points in the rest of the images. This processing takes 10 to 15 minutes on PDP-10 after which the cart is only allowed to travel one meter. The same process of slider stereo is then repeated after each meter and motion stereo is also performed to confirm or add to the depth information.

The work on the Stanford Cart by Moravec was extended at Carnegie-Mellon University (CMU). Thorpe [1984] has described some additional work on the interest operators for stereo vision. Additional results on robot road following at CMU, on other mobile robots named as *Terregator* and *Neptune*, have also been reported [Wallace et al. 1985]. In these, two camera stereo was employed in addition to a ring of twenty four proximity sensors for obstacle avoidance. The main task of navigating during road-following is to keep the vehicle in the centre of the road as it moves. For road-following the road images are digitised continuously and the road edges are located. After determining the deviation of road edges from the centre line of the road, steering commands are issued for keeping the vehicle aligned to the centre of the road. At the low level of vision they have experimented with seven edge detection and three line extraction techniques. It has been observed that all of these techniques work in simple cases but give conflicting results in difficult cases. *Neptune* has achieved a continuous motion in road following at the rate of 2 cm/s. Constraining the search for road edge location to a sub-image and application of simple edge detection techniques in addition to image processing hardware have led to the reduction in processing time and therefore high speed motion as compared to Stanford Cart. The CMU *Navlab* (Navigation Laboratory) [Thorpe et al. 1988] has achieved a maximum speed of 10 cm/s during road following. *Navlab* is equipped with a TV camera and a laser range finder. Instead of tracking road boundary lines, a classification method has been used to identify clusters of on-road and off-road pixels. The pixel classification is based on colour and texture properties. The colour parameters are adapted with the changes in colour and environment. The information from a laser range finder has been used for obstacle detection and avoidance.

Recently results on the navigation of a tracked vehicle (M113A2 armoured personnel carrier) at FMC corporation has been reported. The hierarchical sensing system for the FMC vehicle can be classified at two levels of components. At the global level, machine vision is applied to gather information for building a

global model of the world [Kuan et al. 1986]. The colour images are acquired continuously and segmented into road and non-road regions by a pixel classification algorithm. Then geometric reasoning is applied for perceiving real road edges. Road side consistency, smoothness, and continuity are the most promising geometric reasoning used for perception. The local level sensing for the FMC vehicle is used by the pilot sub-system which is responsible for guiding the vehicle along a dynamically feasible route and avoiding the obstacles while maintaining the vehicle on globally planned paths [Nitao and Parodi 1985]. A sonic imaging system gives sensing information, including the information about those obstacles which may be missed by the main vision system, to the reflexive pilot. The reflexive pilot has also a fast response time to unforeseen conditions which might come up locally. This hierarchical sensing system has enabled the navigation of FMC vehicle at the speed of 8 km/hr.

Among the different sensing units, machine vision is the main source of information for constructing a world model to plan paths in addition to its use for monitoring the movement of any autonomous vehicle or robot according to the planned path. A set of algorithms for vision guidance has been proposed, which are implementable in real-time [Inigo et al. 1984]. They have used a single camera for locating road way boundaries while stereo vision is needed for obstacle detection. Most recently the use of the vanishing point concept for road-following has been put forward [Liou and Jain 1987]. Liou and Jain have argued that it is more helpful for processing low quality pictures in the search for convergent line (road sides) rather than using classical data driven line fitting and approximation techniques. After selecting the best vanishing point (where the pair of lines due to road sides meet), the vehicle can be easily guided towards that point.

### 6.2.2 Find-Path Problem

Path planning is one of the most important components of the software required for navigating an autonomous vehicle in a completely or partially mapped environment. Most of the well

developed path planning algorithms assume a completely known environment. The hierarchical path planning techniques for a cluttered and partially mapped environment may first extract path segments locally and then string together some of these segments for a complete path. The modelling of the environment is critical for path planning. It includes the representation of free space, obstacles, and the moving vehicle itself.

In the following discussion different find-path techniques are reviewed. Most of these methods, with the exception of the potential field approach, abstract the environment into a graph of possible paths. Then the graph is searched and an optimal path is determined.

### *Visibility Graph Methods*

The visibility graph, *vgraph* concept was used to navigate *Shakey* in the labs. of SRI [Nilsson 1969]. The environment of the robot is represented as a grid model which is updated continuously. A vertex graph is then constructed from the grid model. The *vgraph* consists of vertices of opposite corners of obstacles, start vertex, goal vertex, and links connecting the vertices in straight lines without overlapping any obstacle. The shortest collision free path from start to goal in *vgraph* is found by weighting each link by its euclidean distance and then searching for the lowest cost path. A similar technique is employed to plan a path for the JPL robot but the cost function associated with each link is the energy required to traverse that link [Thompson 1977]. This makes the vehicle more suitable for planetary exploration than *Shakey*. Additionally a combinatorial explosion is avoided during the derivation of links. The graph search for path planning is reduced to a tree search by properly choosing successive vertices and pruning.

All of the *vgraph* algorithms assume that the moving object is a point which is a fair approximation comparing the object size with the obstacles. It can cause problems when the size of the vehicle is larger than a narrow free space. Lozano-Perez and Wesley [1979] have introduced a solution to this problem. First a generalisation

can be made by considering the moving object to be a circular shape of radius,  $r$ . Then the obstacle vertices are moved away by the same distance,  $r$  from actual obstacle corners. The moving object is guided in such a way that its reference point moves through the new displaced vertices, producing a collision free path. This concept was introduced in *Shakey* but Lozano-Perez and Wesley have tackled the problem for more general shapes of the moving object by growing the polyhedral obstacles according to the shape and size of moving object. The utilisation of rotation is also suggested for moving objects when their shape is rectangular. Path planning in the case of the Stanford Cart is also based on the concept of *vgraph*.

### *Find-Path by Representing Empty Space*

In the free space techniques the moving object is explicitly forced to travel on those path segments which run in the middle of the free space corridors between obstacles. In the case of *Hilare*, the pre-learned floor plan is divided into empty convex regions, known, and unknown obstacles. The convex regions are formed by connecting the nearest vertices to create a representation of empty areas which are called the C-cells. A trajectory within such cells is then sought for the optimal paths. Laumond [1983] has extended the concept of free space to an hierarchical representation of C-cells in terms of topology of places such as rooms, work areas, and other parts of the known domain.

Brooks [1982] also proposes another approach to path planning which models the free space between obstacles by fitting generalised cylinders. The solution was developed for a two-dimensional plane and pathways are obtained from the fitted generalised cylinders in which the moving object can freely travel on a plane. Then the technique is extended to three dimensions by stacking these planes. Meng [1988] has modelled the free space by a spatial graph, known as Voronoi graph. The obstacles are assumed as randomly shaped. A solution to the find-path problem, in a 2-D plane, is proposed by interpreting the Voronoi diagram as

a spatially oriented graph representing the skeleton of the free space.

Lozano-Perez [1983] has provided a mathematical treatment of the find-path problem by using a configuration space approach. The path planning is treated as two problems, *find space* and *find path*. Conceptually, those parts of the free space are found in which the moving object can reside without colliding with the obstacles. Gouzenes [1984] has also addressed the problem of empty-space and find-path. Different heuristics are suggested for the construction of free-space by introducing an intrinsic tree structure to represent empty space.

### *Potential Methods*

In the potential field approach the moving object is assumed to be a rolling ball, and the floor is tilted toward the goal point. The direction to which the ball rolls for the optimal path is observed [Andrew 1983, Krogh 1984, Khatib 1986]. The obstacles are presented as hills with sloping sides so that the ball rolls away from them and seeks the path between them. Momentum can also be given to rolling ball by taking into account the energy required to accelerate, decelerate, or turn.

Another way of describing these techniques is by considering the environment as a field of forces, where the destination point attracts the moving object and obstacles or barriers produce repulsive forces for the moving vehicle. The trouble with these techniques, in common with multi-dimensional optimisation techniques, is that they can get caught in dead ends and require special procedures for backtracking and starting again. When they are combined with intelligent path planning, they can prove to be very useful and offer a quick response and thus the possibilities of real-time implementation. The use of potential methods for real-time collision avoidance has been demonstrated effectively by Khatib [1986].

## 6.3 Navigation System for the Endoscope

### 6.3.1 The Find-Path Problem and Guiding the Endoscope

Collision avoidance during the navigation of autonomous vehicles is traditionally considered as a high level path planning problem. From this point of view, different levels of control have been established. For example, collision avoidance is preferably performed at higher levels while low level control is limited to the execution of elementary operations for which the route has been precisely specified by the high level planning.

The *find-path* problem in a cluttered but completely known world is well understood and has been solved in most of the cases, as discussed in the previous section. However, all of the developed algorithms are not directly applicable in endoscope navigation because the detailed world model is not known. The development of a navigation system for the endoscope on the basis of *find-path* problem will be a disaster in terms of its real-time capabilities. It is also not feasible in any case. The interaction of endoscope with its environment will be passed through a long time cycle of planning and high level control. This actually places a limit on its navigation capabilities in the rapidly changing environment of human colon. Therefore, the role of low level control and navigation should be enhanced for real-time performance. High level navigation should not be replaced by low level functions but it is necessary to make better use of low level navigation techniques by increasing their degree of competence.

Early attempts at navigation in an unexplored world were solely based on image understanding (JPL Robot, CMU Rover). Then Crowley [1985] and Parodi [1985] provided some hierarchical approaches, where global and local models are updated from sensor information. This problem has also been researched by many scientists [Chatila 1982, Iyengar et al. 1985, Turchan and Wong 1985, Oommen et al. 1986]. Chattergy [1985] has also described some heuristic strategies to aid the navigation of a robot in an unexplored environment.

Navigation in a more general case (like endoscope navigation) requires the collision free movement of autonomous vehicles in an unexplored world. The problem of planning an optimal or near optimal path by avoiding collisions with obstacles in such an environment is a challenging task. Unlike the *find-path* problem the endoscope navigation can not be subjected to a rigorous mathematical treatment and this is because of the inherent nature of the problem. The work on endoscope navigation builds upon many of these ideas put forward for navigating robots in an unexplored world.

### 6.3.2 An Hierarchical Navigation Control System

The utilisation of sensory information is necessary for controlling autonomously the movements of the endoscope tip. An hierarchical navigation and control system is suggested which will be able to perform path planning and endoscope tip control in real-time. A multi-level production system for similar tasks has already been described [Chavez and Meystel 1984, Koch et al. 1985]. In order to simplify the problem, we are proposing the decomposition of the hierarchical navigation system into three distinct modules at different levels: global planner, navigator, and pilot sub-systems. From the previous section we argue that due to the specific nature of the problem, the navigator plays the most important role in endoscope navigation. These three sub-systems are described in an effort to define the endoscope environment, sensing capabilities, environment modelling, and path planning.

#### *Global Planner and Expert*

The global planner which is at the highest level, holds specific and abstract information about the overall environment. For example in the case of an autonomous land vehicle, it can have the map of the terrain, general weather conditions, terrain nature e.g. whether it is hazardous, passing through hills, jungles, planes, populated, or un-populated areas. Here for colonoscopy, it utilises the general model of the colon in the form of its typical shape

features like the number of bends and levels of difficulties in each portion of the colon. It can also keep some knowledge based rules from the colonoscopy expert to guide the navigator when it is stuck in some dead end and cannot continue on the basis of sensor information.

The global planner is an expert which does path planning at the highest level and advises the navigator. It also keeps itself informed about the most recent conditions of navigation by receiving current environment information from the navigator, and makes itself ready to be consulted for expert level advice. The global planner is not directly involved in issuing different navigational controls but takes the role of a specialist to be consulted. In this way, the reacting capabilities of the endoscope will be improved by excluding the global planner from the main control loop.

### *Navigator*

The navigator is at the intermediate level of the hierarchical structure of the navigation system. Its main task is to plan the path on the basis of information received from the sensing system after consulting the global planner in difficult cases. It issues different sub-goals, to achieve the planned path, and other control commands to the pilot for execution. The navigator holds a key position in the overall control system and it decides autonomously about navigation. The principle functions of the navigator for endoscopy are explained as follows.

- The navigator receives image information in the form of regions of interest and occluding contours continuously from the sensing system.

- It keeps a three-dimensional representation of the environment and updates it according to the information received from the sensing system.

-It plans the path based on the world representation already built and then generates sub-goals and other control commands for the pilot to execute.

-During path planning it consults and gets expert advice from the global planner, especially when there is an ambiguity in the world information supplied by the sensors. The global planner is also kept informed about the current environment and the whereabouts of the endoscope so that it can provide its expertise without any delay.

-The navigator instructs the pilot to execute already planned sub-goals and other control commands. It can have an option to monitor the execution of those commands. As there is no arrangement for the control of forward and backward movement of the endoscope tube, the navigator is also responsible for counting the number of rings of muscle that the tip has passed through, to make an estimate of the tip position in the colon.

These functions indicate that the navigator is the back bone of the whole navigation system and performs most of the on line tasks and control. Therefore it can be argued that the response of the automatic endoscope to any changes in the environment will be quick and in real-time due to the shorter control path. Currently we are concentrating on the development of a reliable, real-time, navigator.

### *Pilot*

The pilot has a very simple job to do in the proposed hierarchical navigation system. It executes different endoscope tip movement commands (for autonomous vehicles known as steering commands) to achieve the sub-goals issued by the navigator. It also performs other functions like spraying of water or suction of

air. Some additional tasks can also be assigned to the pilot such as avoiding unexpected obstacles which have been missed or ignored by the navigator. For these types of tasks the pilot may require its own reflexive sensors. The pilot should also be able to monitor the execution of local control commands for error correction.

The navigational control structure in the form of global planner, navigator, and pilot is a suitable basis for the development of a general purpose navigation system for an unknown or partially known environment. In the automation of endoscopy the global planner and expert keep most of the information about the world domain in the form of colon, upper gastrointestinal, or bronchus models. Therefore different types of automatic endoscopy can be performed by switching the related expert model and rules in the global planner and expert, while the same navigator and pilot modules can be employed without any changes. This is a useful feature which makes this system suitable for general purpose navigation.

As far as the navigator module is concerned, the information provided by the sensing system and the representation of the environment based on this information plays an important role in its operation. Therefore the sensing system and representation of environment requires detailed explanation in the context of navigator operations.

### 6.3.3 Environment Representation for Navigation

The selection of a data structure for the environment representation normally depends on the nature of the world whether it is pre-learned, partially known, or completely unknown. When the environment is unknown, the construction of its model for path planning is based only on the information provided by different sensors. This means that the environment model only knows what the sensors has told. Moreover, there is also some sort of uncertainty in the information provided by the sensors. The sensors (video camera) do not provide information about the environment beyond their range (field of view).

Sometimes they may track phantom objects which therefore must be removed from the representation. All of these arguments suggest that for the representation of a previously unknown environment, the data structure should support features like inaccuracy and uncertainty. In other words, it should allow the easy addition or removal of any information from the structure. Another representation which is used for performing search for the optimal paths is known as the search space representation. Configuration space, Voronoi base spaces, generalised cylinder free spaces, and medial axis free space have been used as search space representations [Lozano-Perez 1981, Canny 1985, Brooks 1982, Ruff and Ahuja 1984]. In all of these cases, mapping functions are needed to map the world space to a search space. This arrangement limits the on line reacting capabilities of the navigation system. We have developed an environment representation which provides a reliable access of the data to different types of search algorithms for finding paths. An effective representation for path planning methodologies in endoscopy should meet some of the general objectives which are given below.

-The world representation should provide different means of locating important navigational landmarks. The typical landmarks in colonoscopy are inner ring type muscles of colon which can be followed by the endoscope. There are also vanishing points supplied by the vision system in the form of different dark regions in the image. This means that the representation should be able to accommodate these regions and circular curves.

-The path planning algorithms should also be able to update the world representation to take care of inaccuracy and uncertainty in the information supplied by the sensing system. This objective is very important to solve navigational problems for an unknown or partially known world.

-The representation should be so general that it allows the movement constraints of the endoscope. This objective is helpful in the case of endoscopy because, for the time being, we do not have any control on backward, forward, or rotational movements of the endoscope.

-Where the same world representation is used as search space, the world representation should also support efficient search procedures.

A simple model of the colon is presented in the next section by keeping in view the above objectives. The environment and search space representation of the world is also proposed for navigational purposes.

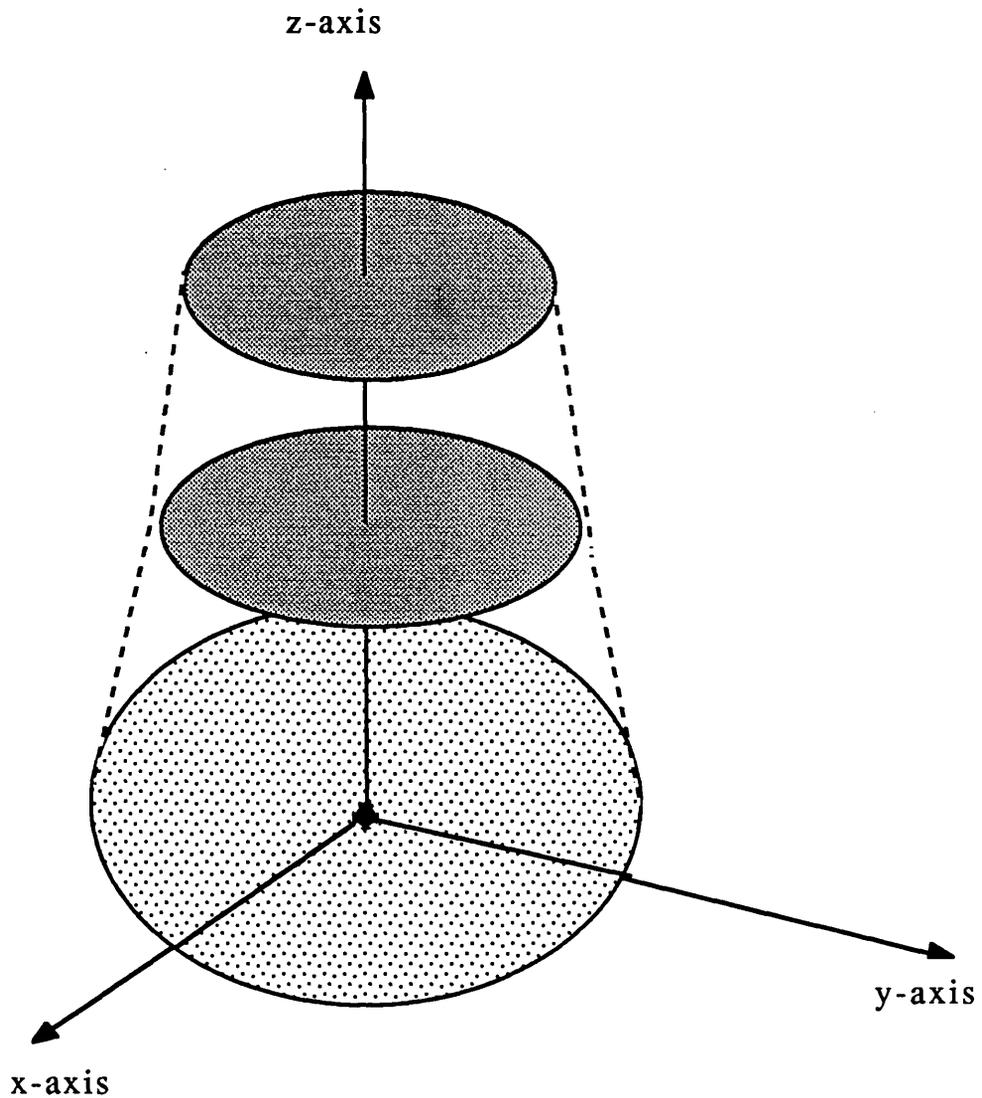
#### 6.3.4 Colon Model for Endoscopy

It is very difficult and computationally expensive to model the colon accurately. Moreover, we believe that it is also not feasible to use precise modelling for navigating the endoscope. Consequently we propose that the colon can be modelled as a series of circular or elliptical rings which are joined to give a hollow cylindrical figure. No restriction is imposed on the size of these rings so that rings of different sizes can be interconnected. When the rings are arranged in such a way that they are parallel to a plane and their centres are on a straight line normal to the planes, they model a straight portion of the colon in three-dimensional space as shown in the Figure 6.1. In this way, the inner space covered by this irregular cylinder roughly models the inside of a colon. When the condition of stacking the rings on a single plane is relaxed and it is assumed that the plane in which the subsequent rings lie can be at any orientation in the 3-D space, then the proposed scheme can model different bends in the colon.

Alternatively the colon model can be described by using generalised cylinders. Generalised cylinders were introduced by Binford [1971]. A circular homogeneous generalised cylinder can

be used to model the whole colon. It consists of four basic components which are also explained in Figure 6.2.

- (i) A space curve which acts as the axis of colon and is shown as the locus of  $L(s)$  in the vector form.



*Figure 6.1: A simple model of the colon.*

- (ii) A cross-section plane defined for every point on the axis and at some angle,  $\alpha$  to the tangent of the axis at the corresponding point. In the proposed

model the angle,  $\alpha$  can be taken as  $\pi/2$ . Moreover, instead of considering these planes at every point on the axis, the planes are taken only at those points which correspond to the inner muscular rings of the colon. For other points the two adjacent planes can be interpolated to achieve a continuous representation.

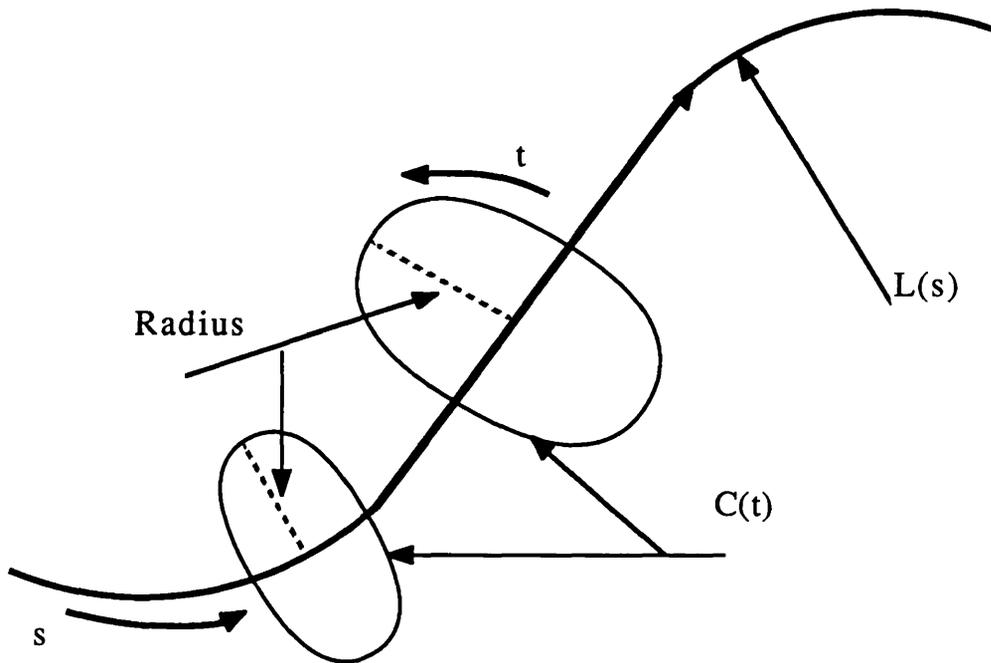


Figure 6.2: A homogeneous generalised cylinder.

(iii) The shape of the object in the plane which is described by a planar curve,  $C(t)$  on the plane. In the case of the colon a circular or elliptical curve which corresponds to the muscular rings can be assumed.

(iv) A transformation rule,  $r(s)$  which specifies a homogeneous change in the shape of the object as

the plane moves along the axis  $L(s)$ . For colon modelling if the shape can be fixed as circular, the transformation can only affect the size of the circle. If the centre of the circles are assumed to be on the axis, the transformation can be easily specified as a linear variation to the radius of the circle with respect to  $s$ .

In this way the colon is represented and modelled only by a four-tuple  $(L, C, r, \alpha)$  in terms of a generalised cylinder.

Most of the useful contours formed due to inner muscles of colon are classified as occluding contours. Therefore modelling of the colon world space by a generalised cylinder helps in extracting shape information from the contours in colon images. In chapter four, the problem of shape from contour has been discussed. In its particular implementation for endoscope navigation, the constraints which can be employed for shape extraction are:

- Each point on the contour generator projects to a distinct different point in the image contour and nearby points of the occluding contour arise from nearby points of the contour generator. It means that the occluding contours in colon images provide unambiguous information about the contour generator which is used to extract the colon shape from contours.

- As suggested by Marr [1977], when the surface is representable by a generalised cylinder, each occluding contour in the image which belongs to a contour generator must lie in a single distinct plane. These planes are distinct with respect to the viewer. In fact, this is very helpful in identifying the relative distance of image contours in colon images and then guiding the endoscope accordingly, avoiding the nearest contour first. The distance between each plane of an occluding contour can also

be approximated by the size of contours and other constraints described in chapter four.

We see that modelling of the colon world space by a generalised cylinder along with the knowledge of occluding contours in colon images, plays an important part in estimating depth and obstacle free space in the colon for endoscope navigation.

#### **6.4 World and Search Space Representation**

After modelling the inner colon space as explained in the previous section, a suitable representation is required to support the model. The best representation of the world is that which avoids excessive detail of the parts of the space and which do not affect the operation [Lozano-Perez 1981]. We have proposed the colon model as a series of planes rather than a volumetric representation by following the same heuristics. The world representation, we are introducing, not only models the world (colon), but also represents the search space. Moreover the representation structure is general enough to represent any three-dimensional space for navigational purposes, especially where the autonomous moving object has three degrees of freedom.

It is possible to represent the colon as a series of two-dimensional cross-sections with the assumption that the forward and backward movements of the endoscope at its intersection to the plane will be always in the direction of the normal vector. This assumption can only be realised, if we consider small movements of the endoscope and then shift the co-ordinates from frame to frame after each tiny movement. The distance between two planes (depth) is also provided in the data structure. This distance can vary for different plane pairs depending on the busyness and the required resolution of the environment. The better option is to fix the reference co-ordinate system in the current plane (e.g. the x-y plane) such that its origin is at the camera position. Then assume that the series of planes ahead may not be parallel to the current plane but for the first two planes they are approximately parallel

to each other and perpendicular to the line joining the camera position and the target point for navigation in the next plane. The new co-ordinate system can be established in the next plane by fixing its origin at the target point and moving the z-axis along the line joining the origins of current and new co-ordinate systems. This scheme of representing the three-dimensional space as a series of planes also takes care of all the bends in the colon. The transformation from one co-ordinate system to the next is explained below and also shown in Figure 6.3.

Suppose that the camera (endoscope tip) is at point O in the current plane and the navigator provides the target point Q(a,b,c) in the next plane with origin O. Then to move the co-ordinate system from O to Q such that the z-axis of the new co-ordinate system is along the line OQ requires one translation and two rotations one about the z-axis and the other about the x-axis. The translation and rotation matrices are:

$$\text{Translate: } \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -a & -b & -c & 1 \end{vmatrix} ;$$

$$\text{Rotate about z: } \begin{vmatrix} \text{Cos}\theta & -\text{Sin}\theta & 0 & 0 \\ \text{Sin}\theta & \text{Cos}\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} ;$$

$$\text{Rotate about x: } \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & \text{Cos}\phi & -\text{Sin}\phi & 0 \\ 0 & \text{Sin}\phi & \text{Cos}\phi & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}$$

Where  $\theta$  is the skew between the axis systems, and would normally be zero.

and

$$\phi = \tan^{-1}(c/\sqrt{a^2 + b^2 + c^2})$$

The quadtrees are best suited for the representation of each plane due to their pyramid structure. They have previously been used for representing the 3-D environment, by three orthogonal two-dimensional projections, in path planning [Wong and Fu 1985]. Octrees have also been employed for representing the three-dimensional space in path planning [Shneier et al. 1984, Ruff and Ahuja 1984, Herman 1986].

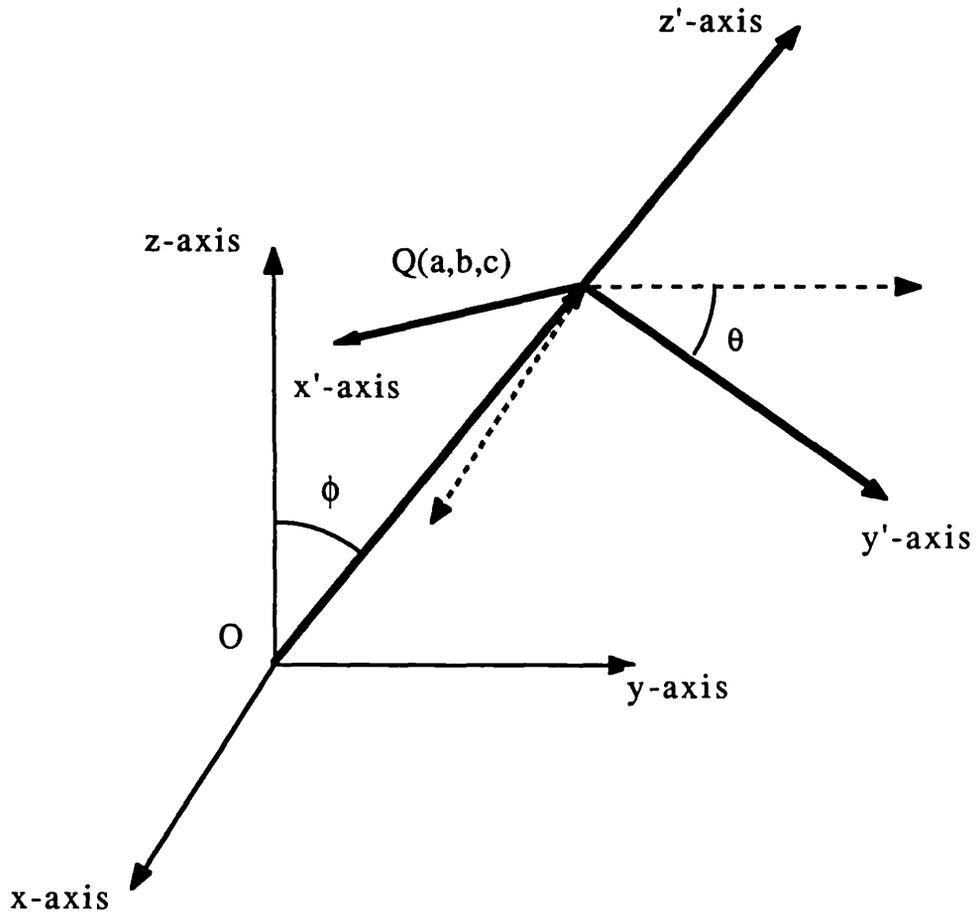


Figure 6.3: Moving co-ordinate system from point O to Q in space.

The Quad-List structure, QL-tree (quadtree based search space and world representation), we are proposing, is a tree structure in-between the quadtree and octrees. Each cross-sectional plane is represented by a separate quadtree, while the tree nodes have two additional links (previous and next) which interconnect consecutive planes to each other at each hierarchical level of the pyramid. An additional field is also provided with each node of the tree to store the distance between consecutive planes from pixel to regional levels. A small section of the data structure is also shown in Figure 6.4. The QL-tree based search space and world representation satisfies all the general objectives described earlier. It provides a spatially indexed representation of the world and can be viewed at several level of resolutions like octrees but it is less complicated in terms of memory and search operations. The representation allows an easy access to each plane represented by quadtrees and more efficient algorithms exist for searching quadtrees.

The main advantage of the QL-tree representation is that the environment information can be stored in each plane in such a way that the plane nearest to the camera contains the most updated and correct information about the objects in the space, while the further planes ahead need only to hold some rough shape of the objects. The accuracy and completeness of this information will depend on the field of view of the sensors. The scene information in the further planes is refined and corrected incrementally as the sensing system provides more data. Thus the QL-tree provides the important capability of learning as new data becomes available, and forgetting the unimportant. The entire data of a plane can be easily updated when the sensor provides more data and removed from the structure as the endoscope navigates successfully through it. The quadtree structure is also useful for other sensing tasks (e.g. image segmentation) because it provides an hierarchical image representation. It is interesting to note that in the overall endoscope control and sensing system, the pyramid structure has been used for estimating depth by region based segmentation, contour extraction, and representing the

muscle rings detected in colon images [Khan and Gillies 1988a, 1988b, 1989a, 1989b]. The effective cost of generating pyramids is much lower than if we had to generate a different representation for each task.

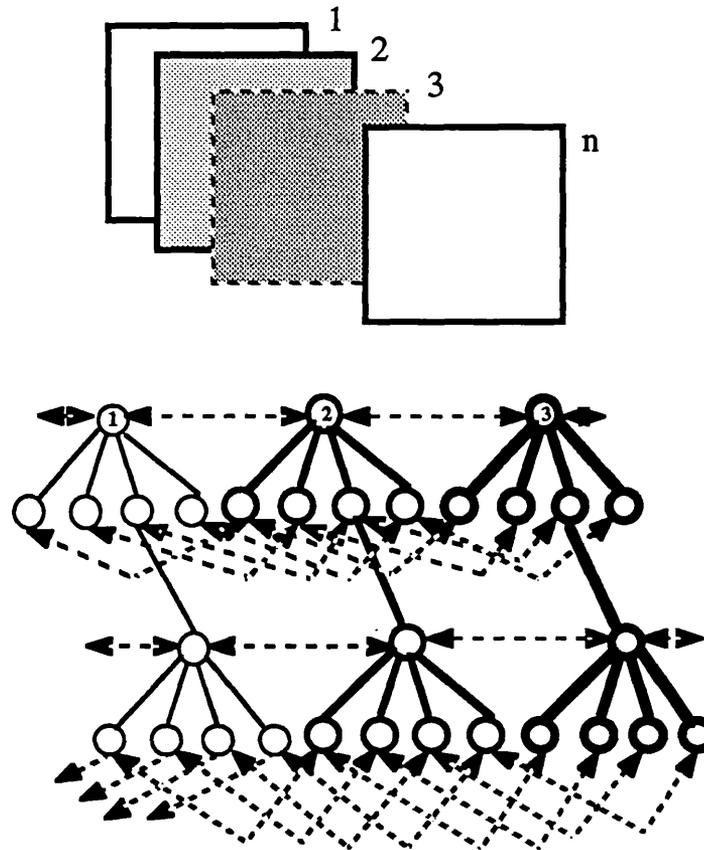


Figure 6.4: A section of Quad-List tree, (QL-tree) environment and search space representation.

## 6.5 Concluding Remarks

As part of an in-depth study for the development of a navigation system for the endoscope, different navigation techniques have been surveyed. These provide an overall picture of the research in navigation specifically in the areas like sensing units, machine perception, and the *find-path* problem. During the review it

appeared that the existing techniques for navigating autonomous vehicles are not directly applicable to endoscopy because the sensing capabilities of the endoscope are limited and its world domain is changing rapidly.

An hierarchical control structure for endoscope navigation has been presented which has three distinct levels known as the global planner, navigator, and pilot. The navigator which is at the intermediate level, receives world information from the vision system and controls the endoscope autonomously by issuing commands to the pilot. This improves the reflexive capabilities of navigation, as the global planner has only been employed as a consultant to the navigator rather than directly in the control loop. The world and search space representation is also very important for the development of a real-time navigation system. A new representation, the QL-tree has been developed and employed to model the search space. It consists of a series of planes which are organised as a list of quadtrees linked together by the distance (which is depth in terms of machine vision) between consecutive planes. This search space representation is also suitable for autonomous vehicles. The colon is modelled by using the generalised cylinder concept. The two-dimensional version of generalised cylinder has already been used for defining empty space in the find-path problem. The colon is represented by a single generalised cylinder and the model is constructed incrementally from the information provided by the vision techniques we have presented in this thesis. The model of colon also provides useful constraints to extract 3-D shape from the colon image contours. The circular or elliptical cross-sections of the generalised cylinder at regular steps directly provide the planes defined for the search space. Therefore a single representation for world and search space is established, which enhances the real-time reflexive capabilities of the endoscope navigation system.

## CHAPTER 7

### CONCLUSIONS AND FUTURE WORK

#### 7.1 Introduction

The task of navigating the endoscope inside the human colon is an ambitious one and there are a number of complications involved in achieving this objective. We have given a lead by providing a solution for the perception problem and initiating the work on navigation system for the endoscope. The navigation of the endoscope is different from that of autonomous vehicles in a number of ways. Firstly, in the current state of endoscopy there cannot be an automatic control on the forward, backward, and rotational movements of the instrument. The automation of endoscope only concerns the control of the tip movement. Secondly, the sensor system on the endoscope is limited to a single camera on its tip and there is only a remote possibility of having additional sensors for depth measurement. Motion stereo is the only other likely method for three-dimensional shape extraction of inner body surfaces. We have not considered it in this thesis, and have only treated the machine vision techniques based on single monocular images.

There are a number of constraints due to the illumination conditions and colon model which are helpful in estimating the insertion direction for controlling the endoscope tip. During the course of endoscopy procedures, the inner body surfaces are illuminated by a single light source which is in fact equivalent to a point source. In this way the surface illumination is related to its distance from the light source at the endoscope tip. The surfaces which are nearer to the tip are more brightly illuminated than the further surfaces. Therefore a uniform and dark region in an endoscope image corresponds to the deep and obstacle free area in

the colon. Another important and useful constraint arises from the contours formed by the inner colon muscles. The source of these contours is mainly occluding edges. The occluding contours are approximately circular and they provide another type of important landmark for navigating the endoscope. A line joining the centre of curvature of these contours gives a clear and obstacle free path for endoscope insertion. The representation of a colon by a generalised circular cylinder provides a useful approach to extract the shape information from these occluding contours.

The occluding contours and darker regions in the colon images are to be detected for seeking an obstacle free path which avoids the inner walls of colon. Most of the work, presented in this thesis deals with the detection of contours and darker regions. Due to the real-time nature of endoscope control, these machine vision techniques are implementable in parallel on a pyramid architecture based computer. The sequential algorithms do not provide real-time performance. A review on the physiology and psychology of vision has been carried out in the 2nd chapter, before describing the contour and region extraction algorithms. It was argued that early visual data organisation in organic vision follows the signal to symbols paradigm. The architectural nature of the organic visual processing is related to different parallel processing techniques and it was concluded that the pyramid based processing is a reasonable model of the brain architecture. The conventional parallel processing methods cannot be used to achieve the visual processing and recognition in real-time. The pyramid based parallel-hierarchical processing provides a useful means for explicitly extracting the global structure in images. From the point of view of psychology, different perceptual organisation and grouping principles were studied. It was argued that these principles are very useful in developing general purpose machine vision techniques, if implemented effectively and intelligently. In fact we have demonstrated, in our approach for contour extraction, that these organisational principles are very effective in isolating the relevant contour structure from noisy image data.

## 7.2 Contour Extraction

The contour extraction method is based on the bottom-up organisation of edge point data and it utilises the information presented in the data itself rather than higher level heuristics. The support for this type of early image data organisation is evident from both neurophysiology and psychological studies. The method employs different domain independent grouping rules from perceptual organisation. As far as we know, this is the first time that perceptual grouping has been utilised in a unified manner for contour extraction.

An intermediate representation based on straight line segments is formed from the edge point data rather than sequentially linking edge points into contours. There is clear evidence that straight line segments are extracted by simple and complex cells in the early stages of animal vision. Moreover, any type of contours (curved or straight) can be approximated by straight line segments. For line segment extraction, the edge points are detected by using a simple edge detector (Sobel) and all edge data, however weak, is retained providing it is sufficiently reliable. The image is divided into overlapping squares of a given size and line segments are extracted in each image square. This process has been implemented independently and in parallel for each of the image windows. The size of image window and overlapping depends on the details of the image contents and what is required for recognition. In our particular implementation, the line segments are extracted at two resolutions based on  $8 \times 8$  and  $4 \times 4$  image windows.

The grouping of edge points into short line segments is an early process and the information which is used to group edge points is carried by edges themselves in the form of their location, contrast, orientation, and intensity value. The laws of data organisation which have been employed to extract useful line segments are: proximity, connectivity, similarity in edge orientation, contrast,

and edge pixel intensity. The problem with the use of these perceptual organisation principles, in the past, has been the lack of an effective implementation. We have modified the Hough transform to implement these grouping techniques. After proximity and connectivity, similarity in edge orientation is the most effective way of extracting useful line structures. In another orientation based grouping method [O'Gorman and Clowes 1976], the edge points which are selected to support a line structure are those whose orientation is perpendicular to the line direction (within some tolerance). However, with this technique, most of the weak line segments which are part of curved contours of curved surfaces are missed as we have demonstrated by the experimental results in chapter three. To avoid this, a slow drift in the edge orientation is allowed from one edge point to the neighbouring edge on the line segment. The grouping relation formed in this way is more perceptually stable and regular for extracting weak but significant contours. The aggregation of edge points on the basis of similar edge orientation is also allowed whether the intensity change is normal to the line or not. This is equivalent to Marr's grouping principle called theta-aggregation. It has been demonstrated that these new grouping principles for edge point data recover those useful line structures which are generally undetectable. During the process of forming line segment representation by perceptual grouping, the edges due to noise are filtered out. The effectiveness of individual grouping processes is analysed by using artificial images with known added noise. In the case of random noise, connectivity grouping extracts most of the useful line segments but for endoscopic images which contain a variety of noise, orientation grouping based on slow drift in edge orientation and theta-aggregation is also needed to detect most of the useful line segments.

The next step, where we aggregate line segments into contours, was described in the fourth chapter. A multi-resolution representation, based on a pyramid, is employed to represent line segments at the two lower levels while the contour segments are represented at higher levels as groups of short line segments. The grouping of line segments into contours is implemented using a

4×4 overlapped quadtree to simulate the pyramidal architecture. The aggregation process is mainly performed on 8×8 window based line segments and lower level segments are only used for completing the fragmented parts of contours and to resolve ambiguities. The processing elements in the overlapped pyramid are linked to four parents and sixteen children. Each parent performs grouping on the line data supplied by its sixteen children independently and in parallel and passes on the grouped line segments in its inner 2×2 block children to its parents. The grouping principles used in this operation are: proximity of line end points, theta-aggregation, curvilinearity, continuity, and similarity in the line contrast. The grouping process starts from the bottom level and grouped segments are passed to the higher level processors and when the root of the pyramid is reached. Groups of line segments emerge which are then converted into contours. Different order polynomials can also be fitted on the contour data at this stage.

The bottom-up and data driven processes used for contour extraction have their roots in psychophysical and neurophysiological studies. The transition gap between edge point data and contours is a source of a discontinuity in the flow of information. Our contour extraction method offers the potential for eliminating this discontinuity. In contrast to the existing contour extraction techniques, our approach forms contours in parallel using the pyramid architecture in a single pass starting from the bottom level and moving to the top of the pyramid.

### 7.3 Region Extraction

A new method for dark region extraction to estimate the insertion direction of the endoscope was described in chapter five. A pyramid structure based on the intensity mean and variance of square blocks has been used in our method. The formulation of variance computation recursively as one moves from the bottom level to the top of the pyramid, has allowed us to devise a highly parallel implementation. The average intensity of the darker

region is estimated from the first peak of the histogram of a given colon image. The algorithm is simulated by constructing a quadtree in which each node stores the average intensity and variance of its corresponding region. A dynamic link is associated to each node of the pyramid in addition to the normal connections to its children and parent. During the pyramid construction process, a record is kept for the largest dark and uniform square region in each sub-tree. This has been achieved by connecting the dynamic link of the root of every sub-tree to its darker and uniform region node. In this way when the pyramid is completed, the node which corresponds to the largest dark and uniform region in the pyramid is identified. The region itself can be used to estimate the endoscope insertion direction or it may be used as a seed for the region growing process to extract a complete dark region.

As far as we know, this is the first time a variance pyramid has been used to extract regions of given properties. The region extraction method is very efficient and effective in terms of its parallel implementation. We have also implemented an extended version of this technique for general purpose segmentation. The first step of pyramid building, which identifies the seed regions for each node of the pyramid, is similar to the dark region extraction. An additional top-down pass in the pyramid is required for growing these seeds into complete uniform regions. In addition to the colon images tested for dark region extraction, different medical and computer generated images have also been segmented successfully.

#### **7.4 Discussion**

Both of the image segmentation techniques, we have presented in this thesis, do not now seem to require any basic modification or improvements. The tuning of the algorithms, however, may be needed when they are used in different world domains.

In contour extraction, we have assumed the presence of a single line structure in a pre-defined small window of the image. This is not a bad assumption in the context of endoscopic colon images. In other domains, where the assumption may limit the performance of algorithm, the presence of more than one line segment can be assumed depending on the threshold for the number of votes for a useful line structure. Another point which may be raised concerns the extraction of straight line segments rather than curved line segments. We believe that the choice for straight segments is an optimal one, in terms of further grouping processes (e.g. theta-aggregation), efficient implementation, and the evidence from physiological studies. There is no evidence of any loss of contours due to straight line segments, and contour location errors are within an acceptable level. The grouping techniques used in the line segment and then contour extraction steps can be strengthened by applying the well known law of common fate but this is only possible if we first calculate the optical flow from the sequence of images. The shape information from motion is discussed in the future work.

The dark region extraction method is the simplest one but most effective in determining the insertion direction. The only uncertainty involved is due to the estimation of darker region intensity from the image histogram. But the technique for region extraction itself is very sound and accurate irrespective of whether the intensity of a region is known or not. This has been demonstrated by extending the method to general purpose segmentation, where the technique does not use any information about the intensities of uniform regions. During extracting dark regions in colon images, if the first peak is un-detectable from the image histogram, it indicates that there is no dominant dark region in the image. This means that the chances of finding the deepest and obstacle free area in the colon are limited. This effect, for example, may be due to the simple fact that endoscope tip is facing a colon wall.

The region uniformity criterion for extracting the seed region is dependent only on a variance low-threshold. But when the regions

are merged there are other factors which have been included in defining the uniformity of a region. The second variance high-threshold can be defined for the merged regions which can be higher than the seed region threshold. Another tolerance threshold based on the difference in average intensities of the two regions is also employed. Ideally the average intensities of two regions to be merged should be equal. We have adapted this threshold depending on the uniformity of the merged regions to avoid over merging. The uniformity of regions can also be dynamically defined depending on the size of the region but we have found that in most of the test images this does not provide any more better results.

## 7.5 Future Work

There are a number of research areas which have to be explored for achieving automatic control of endoscope for colonoscopy. The endoscope can only be navigated successfully by merging and bridging the gaps between these diverse areas which include but are not limited to: computer vision, robotics, expert systems, 3-D modelling, sensors, control systems, and the mechanics of the endoscope itself. In this thesis we have only discussed the early and intermediate visual processing. The exploitation of full capabilities of computer vision itself requires a large amount of effort and research.

The endoscope sensing system needs a considerable amount of work for building an accurate world model of colon from the visual information available. We have already proposed a model of the colon in the form of a generalised homogeneous, circular cylinder in chapter six. The model can be extended to a more general elliptical generalised cylinder. The occluding contours can also be interpreted for extracting three-dimensional information by using the different constraints which we have described in chapter four. In addition to that, the work on the extraction of temporal information from a sequence of colon images has been started. The computation of motion information will not only

provide the colon shape information but also camera motion parameters will be determined during this process. In particular, the detection of camera motion will provide the information about the movement of the endoscope tip, which is vital for the confirmation of the execution of a tip movement command. The main work in this regard has been undertaken to investigate some effective methods for computing the optical flow from a sequence of endoscopic images. This is a difficult task, particularly considering the noise in colon images and other artefacts of the human colon.

We have introduced an hierarchical navigation system for endoscope control in chapter six. The navigation system consists of three distinct modules: global expert, navigator, and pilot. The work on the global expert will also be getting under way soon. In this regard, different expert rules on colonoscopy will be compiled by visiting colonoscopy sessions on a variety of patients. The endoscope insertion and manoeuvring techniques (for example, how to come out of a loop formed in the sigmoid colon) from different endoscope consultants will also be gathered to overcome the dead end encountered during the colonoscopy. In this way a production rule based system will be built to provide expert advice for the navigator when the visual sensing information is not adequate to guide the endoscope. Fortunately, part of the medical support for this project has been provided by the endoscopy unit at St. Mark's Hospital London.

In order to test our machine vision techniques in real-time conditions, we are implementing the dark region extraction method using a parallel-hierarchical pyramid of transputers. In the first instance, the algorithm will be implemented using five transputers on a XENIX base host for an Imaging Technology series 151 image processor. The series 151 image processor will receive the sequence of colon images from a video tape made during colonoscopy. The construction of an image histogram and the estimation of the first peak in the histogram will also be implemented on the image processor to achieve the results at video rate. The image data will be transferred to the transputer

boards for detecting the dark region. The transfer of image data is currently consuming most of the time in our proposed parallel-pipeline processing. We are investigating the use of a transputer board with the frame grabber to avoid this data transfer delay. In addition to that, the use of twenty one transputers instead of five is also being considered. Hopefully with the introduction of additional processing power and parallelism, the region extraction time will be reduced in the order of milliseconds. Therefore the system will track the lumen from on line colon images digitised from a video recorder. This set up provides a simulation for the automatic insertion of endoscope. The dark region information will be used by the navigator to generate tip control commands for execution by the pilot.

## REFERENCES

Abdou, I.E., "Quantitative Methods of Edge Detection," Ph.D. Dissertation, Image Processing Institute, University of Southern California, 1978. Also USCIP Report 830.

J. Andrew, "Impedance Control as a Framework for Implementing Obstacle Avoidance in a Manipulator", Master's Thesis, MIT 1983.

Attneave, F., "Apparent Movement and the What-Where Connection," *Psychologia*, 17, pp. 108-120, 1974.

Ballard, D.H., "Generalizing the Hough Transform to Detect Arbitrary Shapes," *Pattern Recognition*, 13(2), pp. 111-122, 1981.

Barrow, H.G., Tenenbaum, J.M., "Recovering Intrinsic Scene Characteristics from Images," in *Computer Vision Systems*, Eds. Hanson, A.R., Riseman, E.M., pp. 3-26, Academic Press, New York, 1978.

Binford, T.O., "Visual Perception by a Computer," *Proc. IEEE Conference on Systems and Control*, Miami, 1971.

Blanford, R.P., Tanimoto, S.L. "Bright-Spot Detection in Pyramids," *Computer Vision, Graphics, and Image Processing*, 43(2), pp. 133-149, 1988.

Blicher, P., "Edge Detection and Geometric Methods in Computer Vision," Ph.D. Dissertation, Computer Science, Stanford University, 1984. Also Report No. STAN-CS-85-1041, 1985.

Brady, M., "Criteria for Representation of Shape," in *Human and Machine Vision*, Eds. Beck, J., Hope, B., Rosenfeld, A., pp. 39-84, Academic Press, New York, 1983.

Brice, C.R., Fennema, C.L., "Scene Analysis Using Regions," *Artificial Intelligence*, 1(1-2), pp. 205-226, 1970.

Brooks, R., "Solving the Find-Path Problem by Representing Free Space as Generalized Cylinders," MIT Artificial Intelligence Lab. Report, AI Memo No. 674, May 1982.

Brown, C.M., "Inherent Bias and Noise in the Hough Transform," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-5(5), pp. 493-505, 1983.

Burns, J.B., Hanson, A.R., Riseman, E.M., "Extracting Straight Lines," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-8(4), pp. 425-455, 1986.

Burt, P.J., Hong, T.H., Rosenfeld, A., "Image Segmentation and Region Property Computation by Co-operative Hierarchical Computation," *IEEE Transaction on Systems, Man, and Cybernetics*, SMC-11(12), pp. 802-809, 1981.

Canny, J.F., "Finding Edges and Lines in Images," MIT Artificial Intelligence Lab. Report AI-TR-720, 1983.

Canny, J.F., "A Voronoi Method for the Piano-Movers Problem," *Proc. IEEE International Conference on Robotics and Automation*, St. Louis, Missouri, pp. 530-535, 1985.

Cantoni, V., Levialdi, S., "PAPIA: A Case History," in *Parallel Computer Vision*, Ed. Uhr, L., pp. 3-13, Academic Press, Orlando, Florida, 1987.

Chatila, R., "Path Planning and Environment Learning in a Mobile Robot System," *Proc. European Conference on Artificial Intelligence*, Orsey, France, pp. 211-215, 1982.

Chattergy, R., "Some Heuristics for the Navigation of a Robot," *International Journal of Robotics Research*, 4(1), pp. 59-66 1985.

Chavez, R., Meystel, A., "Structure of Intelligence for an Autonomous Vehicle," *Proc. IEEE International Conference on Robotics*, Atlanta, Georgia, pp. 584-591, 1984.

Cohen, M., Toussaint, G.T., "On the Detection of Structures in Noisy Pictures," *Pattern Recognition*, 9(2), pp. 95-98, 1977.

Crick, H.C., Marr, D., Poggio, T., "An Information Processing Approach to Understanding the Visual Cortex," MIT Artificial Intelligence Lab. Report AI Memo No. 557, 1980.

Crowley, J., "Navigation for an Intelligent Mobile Robot," *IEEE Journal of Robotics and Automation*, RA-1(1), pp. 31-41, 1985.

Davies, E.R., "Image Space Transforms for Detecting Straight Edges in Industrial Images," *Pattern recognition Letters*, 4, pp. 185-192, 1986.

de Saint Vincent, A.R., "A 3D Perception System for the Mobile Robot Hilare," *Proc. IEEE International Conference on Robotics and Automation*, San Fransisco, California, pp. 1105-1111, 1986.

Duda, R.O., Hart P.E., "Use of the Hough Transformation to Detect Lines and Curves in Pictures," *Communications of the ACM*, 15(1), pp. 11-15, 1972.

Duda, R.O., Hart P.E., "Pattern Classification and Scene Analysis," John Wiley, New York, 1973.

Forman Jr., A.V., "A Modified Hough Transform for Detecting Lines in Digital Imagery," *Proc. SPIE Application of Artificial Intelligence III*, Vol-635, Orlando, Florida, pp. 151-160, 1986.

Gibson, J.J., "The Senses Considered as Perceptual Systems," Houghton Mifflin, Boston, Massachusetts, 1966.

Giralt, G., Sobek, R., Chatila, R., "A Multi-Level Planning and Navigation System for a Mobile Robot: A First Approach to HILARE," *Proc. 6th International Joint Conference on Artificial Intelligence*, Tokyo, Japan, pp. 335-337, 1979.

Gouzenes, L., "Strategies for Solving Collision-Free Trajectories Problems for Mobile and Manipulator Robots," *International Journal of Robotics Research*, 3(4), pp. 51-65, 1984.

Gregory, R.L., "The Intelligent Eye," McGraw-Hill, New York, 1970.

Griffith, A.K., "Edge Detection in Simple Scenes Using a Prior Information," *IEEE Transaction on Computers*, C-22(4), pp. 371-381, 1973.

Hanahara, K., Maruyama, T., Uchiyama, T., "A Real-Time Processor for the Hough Transform," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 10(1), pp. 121-125, 1988.

Haralick, R.M. "Ridges and Valleys on Digital Images," *Computer Vision, Graphics, and Image Processing*, 22(1), pp. 28-38, 1983.

Haralick, R.M., "Digital Step Edges from Zero Crossing of Second Directional Derivatives," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-6(1), pp. 58-68, 1984.

Hartley, R., Rosenfeld, A., "Hierarchical Line Linking for Corner Detection," in *Integrated Technology for Parallel Image Processing*, Ed. Levialdi, S., pp. 101-119, Academic Press, London, 1985.

Herman, M., "Fast Path Planning in Unstructured, Dynamic, 3-D Worlds," *Proc. SPIE Application of Artificial Intelligence III*, Vol-635, Orlando, Florida, pp. 505-512, 1986.

Hong, T.H., Shneier, M., Rosenfeld, A., "Border Extraction Using Linked Edge Pyramids," *IEEE Transaction on Systems, Man, and Cybernetics*, SMC-12(5), pp. 660-668, 1982.

Hong, T.H., Shneier, M., Hartley, R.L., Rosenfeld, A., "Using Pyramids to Detect Good Continuation," *IEEE Transaction on Systems, Man, and Cybernetics*, SMC-13(4), pp. 631-635, 1983.

Hong, T.H., Rosenfeld, A., "Compact Region Extraction Using Weighted Pixel Linking in a Pyramid," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-6(2), pp. 222-229, 1984.

Horwitz, S.L., Pavlidis, T., "Picture Segmentation by a Directed Split-and-Merge Procedure," *Proc. 2nd International Joint Conference on Pattern Recognition*, Copenhagen, pp. 424-433, 1974.

Hough, P.V.C., "Method and Means for Recognizing Complex Patterns," U.S. Patent 3069654, 1962.

Hubel, D.H., Wiesel, T.N., "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex," *Journal of Physiology*, 160, pp. 106-154, 1962.

Hubel, D.H., Wiesel, T.N., "Shape and Arrangement of Columns in Cat's Visual Cortex," *Journal of Physiology*, 165, pp. 559-568, 1963.

Hubel, D.H., Wiesel, T.N., "Receptive Fields and Functional Architecture of Monkey Visual Cortex," *Journal of Physiology*, 195, pp. 215-243, 1968.

Hubel, D.H., Wiesel, T.N., "FERRIER LECTURE: Functional Architecture of the Macaque Monkey Visual Cortex," *Proc. Royal Society of London, Series B*, 198, pp. 1-59, 1977.

Illingworth, J., Kittler, J., "A Survey of the Hough Transform," *Computer Vision, Graphics, and Image Processing*, 44(1), pp. 87-116, 1988.

Inigo, R.M., McVey, E.S., Berger, B.J., Wirtz, M.J., "Machine Vision Applied to Vehicle Guidance," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-6(6), pp. 820-826, 1984.

Iyengar, S.S., Jorgensen, C.C., Rao, N.S.V., Weisbin, C.R., "Learned Navigation Paths for a Robot in Unexplored Terrain," *Proc. IEEE 2nd Conference on Artificial Intelligence Application*, Miami Beach, Florida, pp. 148-155, 1985.

Julesz, B., "Visual Pattern Discrimination," *IRE Transaction on Information Theory*, IT-8, pp. 84-92, 1962.

Julesz, B., Bergen, J.R., "Textons, the Fundamental Elements in Preattentive Vision and Perception of Textures," *Bell System Technical Journal*, 62(6), pp. 1619-1644, 1983.

Kass, M., Witkin, A.P., "Analysing Oriented Patterns," *Proc. Ninth International Joint Conference on Artificial Intelligence*, Los Angeles, California, pp. 944-952, 1985.

Kelly, M.D., "Edge Detection in Pictures by Computer Using Planning," in *Machine Intelligence*, Eds. Meltzer, B., Michie, D., Vol:6, pp. 397-409, Edinburgh University Press, Edinburgh, 1971.

Khan, G.N., Gillies, D.F., "A Review of Computer Vision with Comments on its Application to Endoscopy," Research Report DOC 87/16, Department of Computing, Imperial College, University of London, 1987.

Khan, G.N., Gillies, D.F., "A Highly Parallel Shaded Image Segmentation Method", *Proc. International Conference on Parallel Processing for Computer Vision and Display*, University of Leeds, 1988a, (Eds. Dew, P., Earnshaw, R., Heywood, T., Addison-Wesley, 1989).

Khan, G.N., Gillies, D.F., "A Perceptually-Guided Method for Curve Detection," Research Report DOC 88/12, Department of Computing, Imperial College, University of London, 1988b.

Khan, G.N., Gillies, D.F., "A Parallel-Hierarchical Method for Grouping Line Segments into Contours," *Accepted for Presentation at SPIE's 33rd International Symposium, Application of Digital Image Processing XII*, San Diego, California, August 1989a.

Khan, G.N., Gillies, D.F., "Detecting Insertion Direction for an Automatic Endoscope," *Accepted for Presentation at IFIP Sixth World Congress on Medical Informatics*, Beijing, China, October 1989b.

Khatib, O., "Real-Time Obstacle Avoidance for Manipulators and Mobile Robots," *International journal of Robotics Research*, 5(1), pp. 90-98, 1986.

Kittler, J., "On the Accuracy of the Sobel Edge Detector," *Image and Vision Computing*, 1(1), pp. 37-42, 1983.

Koch, E., Yeh, C., Hillel, G., Meystel, A., Isik, C., "Simulation of Path Planning for a System with Vision and Map Updating," *Proc. IEEE International Conference on Robotics and Automation*, St. Louis, Missouri, pp. 146-160, 1985.

Koffka, K., "Principles of Gestalt Psychology," Harcourt, Brace and World, New York, 1935.

Krogh, B., "A Generalized Potential Approach to Obstacle Avoidance Control," *Proc. SME Conference on Robotics Research*, Bethlehem, Pennsylvania, 1984.

Kuan, D., Phipps, G., Hsueh, A-Chuan, "A Real-Time Road Following and Road Junction Detection Vision System for Autonomous Vehicles," *Proc. AAAI-86*, Philadelphia, Pennsylvania, pp. 1127-1132, 1986.

Laumond, J., "Model Structuring and Concept Recognition: Two Aspects of Learning for a Mobile Robot," *Proc. 8th International Joint Conference on Artificial Intelligence*, Karlsruhe, West Germany, pp. 839-841, 1983.

Lester, J.M., "A Clustering Model of Boundary Formation," Computer Science Tech. Report No. 239, University of Wisconsin, 1975.

Lettvin, J.Y., Maturana, H.R., McCulloch, W.S., Pitts, W.H., "What the Frog's Eye Tells the Frog's Brain," *Proc. of the IRE*, **47**, pp. 1940-1951, 1959.

Levine, M.D., Shaheen, S.I., "A Modular Computer Vision System for Picture Segmentation and Interpretation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **PAMI-3(5)**, pp. 540-556, 1981.

Levine, M.D., "Vision in Man and Machine," McGraw-Hill, New York, 1985.

Liou, S., Jain, R., "Road Following Using Vanishing Points", *Computer Vision, Graphics, and Image Processing*, **39(1)**, pp. 116-130, 1987.

Lowe, D.G., "Perceptual Organization and Visual Recognition," Kluwer Academic, Boston, Massachusetts, 1985.

Lowe, D.G., T.O. Binford, "The Interpretation of Three-Dimensional Structure from Image Curves," *Proc. 7th International Joint Conference on Artificial Intelligence*, Vancouver, Canada, pp. 613-618, 1981.

Lowe, D.G., T.O. Binford, "Segmentation and Aggregation: An Approach to Figure-Ground Phenomena," *Proc. DARPA Image Understanding Workshop*, Palo Alto, California, pp. 168-178, 1982.

Lozano-Perez, T., "Automatic Planning of Manipulator Transfer Movements," *IEEE Transaction on Systems, Man, and Cybernetics*, SMC-11(10), pp. 681-698, 1981.

Lozano-Perez, T., "Spatial Planning: A Configuration Space Approach," *IEEE Transaction on Computers*, C-32(2), pp. 108-120, 1983.

Lozano-Perez, T., Wesley, M.A., "An Algorithm for Planning Collision-Free Paths Among Polyhedral Obstacles," *Communication of ACM*, 22(10), pp. 560-570, 1979.

Mansoor, W.M., Sokolowska, E., "A Hierarchical Architecture for Image Segmentation," *Proc. International Conference on Parallel Processing for Computer Vision and Display*, University of Leeds, 1988, (Eds. Dew, P., Earnshaw, R., Heywood, T., Addison-Wesley, 1989).

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

Marr, D., "Analysis of Occluding Contours," *Proc. Royal Society of London, Series B*, 197, pp. 441-475, 1977.

Marr, D., "Vision: A Computational Investigation into the Human Representation and Processing of Visual Information," W.H. Freeman, San Fransisco, 1982.

Marr, D., Nishihara, H.K., "Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes," *Proc. Royal Society of London, Series B*, 200, pp. 269-294, 1978.

Marr, D., Hildreth E.C., "Theory of Edge Detection," *Proc. Royal Society of London, Series B*, 207, pp. 187-217, 1980.

Martelli, A., "An Application of Heuristic Search Methods to Edge and Contour Detection," *Communication of the ACM*, 19(2), pp. 73-83, 1976.

Meng, A.C-C., "Free Space Modelling and Geometric Motion Planning under Unexpected Obstacles," *Proc. IEEE 4th Conference on Artificial Intelligence Applications*, San Diego, California, pp. 82-87, 1988.

Moravec, H.P., "Visual Mapping by a Robot Rover," *Proc. 6th International Joint Conference on Artificial Intelligence*, Tokyo, Japan, pp. 598-600, 1979.

Moravec, H.P., "The Stanford Cart and the CMU Rover," *IEEE Proceedings*, 71(7), pp. 872-884, 1983.

Muerle, J.L., Allen, D.C., "Experimental Evaluation of Techniques for Automatic Segmentation of Objects in a Complex Scene," in *Pictorial Pattern Recognition*, Eds. Cheng, G.C., Ledley, R.S., Pollack, D.K., Rosenfeld, A., pp. 3-13, Thompson, Washington, 1968.

Nalwa, V.S., Binford, T.O., "On Detecting Edges," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-8(6), pp. 699-714, 1986.

Nicholls, R., Khan, G.N., Gillies, D.F., "Ray-Tracing Using QL-Tree," *Proc. International Conference on Computer Graphics*, Singapore, pp. 153-162, 1988.

Nilsson, N.J., "A Mobile Automaton: An Application of Artificial Intelligence Techniques," *Proc. 1st International Joint Conference on Artificial Intelligence*, Washington D.C., pp. 509-520, 1969.

Nitao, J.J., Parodi, A.M., "An Intelligent Pilot for an Autonomous Vehicle System," *Proc. IEEE 2nd Conference on Artificial Intelligence Applications*, Miami Beach, Florida, pp. 176-183, 1985.

Noble, J.A., "Finding Corners," *Image and Vision Computing*, 6(2), pp. 121-128, 1988.

O'Callaghan, J.F., "Recovery of Perceptual Shape Organization from Simple Closed Boundaries," *Computer Graphics and Image Processing*, 3(4), pp. 300-312, 1974.

O'Callaghan, J.F., "A Model for Recovering Perceptual Organization from Dot Patterns," *Proc. Third International Joint Conference on Pattern Recognition*, Coronado, California, pp. 294-298, 1976.

O'Gorman, F., Clowes, M.B., "Finding Picture Edges Through Collinearity of Feature Points," *IEEE Transaction on Computers*, C-25(4), pp. 449-456, 1976.

Ohlander, R., Price, K., Reddy D.R., "Picture Segmentation Using A Recursive Region Splitting Method," *Computer Graphics and Image Processing*, 8(3), pp. 313-333, 1978.

Oommen, B.J., Iyengar, S.S., Rao, N.S.V., Kashyap, R.L., "Robot Navigation in Unknown Terrains of Convex Polygonal Obstacles Using Learned Visibility Graphs," *Proc. AAAI-86*, Philadelphia, Pennsylvania, pp. 1101-1106, 1986.

Parodi, A., "Multi-Goal Real-Time Global Path Planning for an Autonomous Land Vehicle Using a High-Speed Graph search Processor," *Proc. IEEE International Conference on Robotics and Automation*, St. Louis, Missouri, pp. 161-167, 1985.

Pavlidis, T., "Segmentation of Pictures and Maps Through Functional Approximation," *Computer Graphics and Image Processing*, 1(3), pp. 360-372, 1972.

Pentland, A.P., "Perceptual Organization and the Representation of Natural Form," *Artificial Intelligence*, 28(3), pp. 293-331, 1986a.

Pentland, A.P., "Parts: Structured Description of Shape," *Proc. Fifth National Conference on Artificial Intelligence*, Philadelphia, Pennsylvania, pp. 695-701, 1986b.

Pietikainen, M., Rosenfeld, A., Walter, I., "Split-and-Link Algorithms for Image Segmentation," *Pattern Recognition*, **15**(4), pp. 287-298, 1982.

Prewitt, J.M.S., "Object Enhancement and Extraction," in *Picture Processing and Psychopictorics*, Eds. Lipkin, B.S., Rosenfeld, A., pp. 75-149, Academic Press, New York, 1970.

Ramer, U., "Extraction of Line Structures from Photographs of Curved Objects," *Computer Graphics and Image Processing*, **4**(2), pp. 81-103, 1975.

Rhodes, F., Dituri, J., Chapman, G., Emerson, B., Soares, A., Raffel, J., "A Monolithic Hough Transform Processor Based on Restructurable VLSI," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **PAMI-10**(1), pp. 106-110, 1988.

Roberts, L.G., "Machine Perception of Three-Dimensional Solids," in *Optical and Electro-Optical Information Processing*, Eds. Tippet, J., et al., pp. 159-197, MIT Press, Cambridge, Massachusetts, 1965.

Robertson, T.V., Fu, K.S., Swain, P.H., "Multispectral Image Partitioning," Technical Report TR-EE73-26, School of Electrical Engineering, Purdue University, W. Lafayette, Indiana, 1973.

Ruff, R., Ahuja, N., "Path Planning in a Three Dimensional Environment," *Proc. 7th International Conference on Pattern Recognition*, Montreal, Canada, pp. 188-191, 1984.

Samet, H., "Region Representation: Quadrees from Binary Arrays," *Computer Graphics and Image Processing*, **13**(1), pp. 88-93, 1980.

Schaefer, D.H., Ho, P., Boyd, J., Vallejos, C., "The GAM Pyramid," in *Parallel Computer Vision*, Ed. Uhr, L., pp. 15-42, Academic Press, Orlando, Florida, 1987.

Sher, D., Tevanian, A., "The Vote Tallying Chip: A Custom Integrated Circuit," *Proc. Custom VLSI Conference*, Rochester, New York, 1984.

Shneier, M., "Using Pyramids to Define Local Thresholds for Blob Detection," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-5(3), pp. 345-349, 1983.

Shneier, M., Kent, E., Mansbach, P., "Representing Workspace and Model Knowledge for a Robot with Mobile Sensors," *Proc. 7th International Conference on Pattern Recognition*, Montreal, Canada, pp. 199-202, 1984.

Sklansky, J., Petkovic, D., "Two-Resolution Detection of Lung Tumours in Chest Radiographs," in *Multiresolution Image Processing and Analysis*, Ed. Rosenfeld, A., pp. 365-378, Springer-Verlog, Berlin, 1984.

Stevens, K.A., "Computation of Locally Parallel Structure," *Biological Cybernetics*, 29(1), pp. 19-28, 1978.

Suk, M., Song, O., "Curvilinear Feature Extraction Using Minimum Spanning Trees," *Computer Vision, Graphics, and Image Processing*, 26(3), pp. 400-411, 1984.

Tanimoto, S.L., Ligocki, T.J., Ling, R., "A Prototype Pyramid Machine for Hierarchical Cellular Logic," in *Parallel Computer Vision*, Ed. Uhr, L., pp. 43-83, Academic Press, Orlando, Florida, 1987.

Thompson, A.M., "The Navigation System of the JPL Robot," *Proc. 5th International Joint Conference on Artificial Intelligence*, MIT, Massachusetts, pp. 749-757, 1977.

Thorpe, C.E., "FIDO: Vision and Navigation for a Robot Rover," Ph.D. Dissertation, Department of Computer Science, Carnegie-Mellon University 1984. Also Report CMU-CS-84-168, 1984.

Thorpe, C., Hebert, M.H., Kanade, T., Shafer, S.A., "Vision and Navigation for the Carnegie-Mellon Navlab," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-10(3), pp. 362-373, 1988.

Tomita, F., Yachida, M., Tsuji, S., "Detection of Homogeneous Regions by Structural Analysis," *Proc. Third International Joint Conference on Artificial Intelligence*, Stanford, California, pp. 564-571, 1973.

Treisman, A., "Preattentive Processing in Vision," *Computer Vision, Graphics, and Image Processing*, 31(2), pp. 156-177, 1985.

Turchan, M.P., Wong, A.K.C., "Low Level Learning for a Mobile Robot: Environment Model Acquisition," *Proc. IEEE 2nd Conference on Artificial Intelligence Application*, Miami Beach, California, pp. 156-161, 1985.

Ullman, S., "The Interpretation of Structure from Motion," *Proc. Royal Society of London, Series B*, 203, pp. 405-426, 1979.

Van Veen, T.M., Groen, F.C.A., "Discretization Errors in the Hough Transform" *Pattern Recognition*, 14(1-6), pp. 137-145, 1981.

Wallace, R.S., "A Modified Hough Transform for Lines," *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Fransisco, California, pp. 665-667, 1985.

Wallace, R., Stentz, A., Thorpe, C., Moravec, H., Whittaker, W., Kanade, T., "First Results in Robot Road-Following," *Proc. 9th International Joint Conference on Artificial Intelligence*, Los Angeles, California, pp. 1089-1095, 1985.

Weiss, R., Boldt, M., "Geometric Grouping Applied to Straight Lines," *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Miami Beach, Florida, pp. 489-495, 1986.

Wertheimer, M., "Laws of Organization in Perceptual Forms," *Psychologische Forschung*, 4, pp. 301-350, 1923. (Translation in *A Sourcebook of Gestalt Psychology*, Ed. Ellis, W., Harcourt Brace, New York, 1938.)

Witkin, A.P., Tenenbaum, J.M., "On the Role of Structure in Vision," in *Human and Machine Vision*, Eds. Beck, J., Hope, B., Rosenfeld, A., pp. 481-543, Academic Press, New York, 1983.

Wong, E.K., Fu, K.S., "A Hierarchical Orthogonal-Space Approach to Collision-Free Path Planning," *Proc. IEEE International Conference on Robotics and Automation*, St. Louis, Missouri, pp. 506-511, 1985.

Zahn, C.T., "Graph-Theoretical Methods for Detecting and Describing Gestalt Clusters," *IEEE Transaction on Computers*, C-20(1), pp. 68-86, 1971.

Zahn, C.T., "Using the Minimum Spanning Tree to Recognize Dotted and Dashed Curves," *Proc. International Computing Symposium*, Davos, Switzerland, pp. 381-387, 1973.

Zucker, S.W., "Region Growing: Childhood and Adolescence," *Computer Graphics and Image Processing*, 5(3), pp. 382-399, 1976.

Zucker, S.W., "Computational and Psychophysical Experiments in Grouping: Early Orientation Selection. in *Human and Machine Vision*, Eds. Beck, J., Hope, B., Rosenfeld, A., pp. 545-567, Academic Press, New York, 1983.

Zucker, S.W., "Early Orientation Selection: Tangent Fields and the Dimensionality of Their Support," *Computer Vision, Graphics, and Image Processing*, **32**(1), pp. 74-103, 1985.

Zucker, S.W., Hummel, R.A., Rosenfeld, A., "An Application of Relaxation Labeling to Line and Curve Enhancement," *IEEE Transaction on Computers*, **C-26**(4), pp. 394-403, 1977.

Zucker, S.W., Stevens, K.A., Sander, P.T., "The Relation Between Proximity and Brightness Similarity in Dot Patterns," MIT Artificial Intelligence Lab. Report AI Memo No. 670, 1982.

## BIBLIOGRAPHY

There are a number of reputed international journals which contain material on computer vision and picture processing. The three main journals, which have been published for a long time, are "Computer Vision, Graphics, and Image Processing", "Pattern Recognition", and "IEEE Transactions on Pattern Analysis and Machine Intelligence". For the last couple of years, more international journals have been introduced due to a considerable increase in machine vision research. These journals include: "International Journal on Computer Vision", "Image and Vision Computing", "Pattern Recognition Letters", and "Machine Vision and Applications". Other journals which carry papers on machine vision and image processing from time to time are: "Artificial Intelligence", "Optical Engineering", "Computing Surveys", "Journal of Parallel and Distributed Computing", and "IEEE Transactions on System, Man, and Cybernetics, Computers, Information Theory, Biomedical Engineering, and Medical Imaging". General articles both on the psychology and the neurophysiology of vision appear in "Scientific American" and "Science". These provide an excellent summary and introduction to the various aspects of human and animal vision.

The International Conference of Pattern Recognition takes place every second year and lengthy proceedings are published. The IEEE Computer Society organises two international conferences on computer vision regularly which are CVPR and ICCV. In addition to that machine vision papers can also be found in their conference on Artificial Intelligence Applications. The yearly symposium sponsored by SPIE carries a number of conferences on machine vision, image processing, and navigation related topics. The International Joint Conference on Artificial Intelligence and the AAAI Conference also provide separate sessions on machine vision, navigation and other related topics.

The International Journal on Robotics Research and the IEEE Journal of Robotics and Automation contain material on the navigational and control aspects of robots and autonomous vehicles. The IEEE also arrange regular conferences on Robotics and Automation and publish their proceedings. The SME Conference on Robotics Research also provides a forum for presenting material on robot navigation.

During the course of this research a number of books relevant to human and machine vision, Pattern Recognition, and image processing were reviewed. A large quantity of additional articles on different topics, related to this thesis, have also been explored and some of these articles which are of considerable interest could not be referred to in the thesis. These articles and the prominent books are grouped into different areas which include machine vision, pattern recognition and image processing, organic vision, line and contour extraction, shape from contour, and pyramidal vision techniques.

### Machine Vision

Ballard, D.H., Brown, C.M., "Computer Vision," Prentice-Hall, New Jersey, 1982.

Blake, A., Zisserman, A., "Visual Reconstruction," MIT Press, Cambridge, Massachusetts, 1987.

Boyle, R.D., Thomas, R.C., "Computer Vision A First Course," Blackwell Scientific, Oxford, 1988.

Brady, M., (Ed.) "Computer Vision," North-Holland, Amsterdam, 1981.

Buxton, H., Wiejak, J., "Towards Computer Vision," in *VLSI Image Processing*, Ed. Offen, R.J., Collins, London, 1985.

Cohen, P., Figenbaum, E., (Eds.) "The Handbook of Artificial Intelligence, Vol. III," William Kaufmann, Los Gatos, California, 1982.

Fischler, M.A., Firschein, O., (Eds.) "Reading in Computer Vision, issues, problems, principles, and paradigms," Morgan Kaufmann, Los Altos, California, 1987.

Grimson, W.E.L., "From Images to Surfaces: A Computational Study of the Human Early Visual System," MIT Press, Cambridge, Massachusetts, 1981.

Hanson, A.R., Riseman, E.M., (Eds.) "Computer Vision Systems," Academic Press, New York, 1978.

Horn, B.K.P., "Robot Vision," MIT Press, Cambridge, Massachusetts, 1986.

Lowe, D.G., "Perceptual Organization and Visual Recognition," Kluwer Academic, Boston, Massachusetts, 1985.

Marr, D., "Vision," W. H. Freeman, San Francisco, California, 1982.

Nevatia, R., "Machine Perception," Prentice-Hall, New Jersey, 1982.

Pentland, A.P. (Ed.) "Pixels to Predicates," Ablex, Norwood, New Jersey, 1986.

Shafer, S.A., "Shadows and Silhouettes in Computer Vision," Kluwer Academic, Boston, Massachusetts, 1985.

Shirai, Y., "Three-Dimensional Computer Vision," Springer-Verlog, Berlin, 1987.

Ullman, S., "The Interpretation of Visual Motion," MIT Press, Cambridge, Massachusetts, 1979.

Winston, P.H., "The Psychology of Computer Vision," McGraw Hill, New York, 1975.

### **Pattern Recognition and Image Processing**

Duda, R.O., Hart, P.E., "Pattern Classification and Scene Analysis," Wiley, New York, 1973.

Gonzalez, R.C., Wintz, P.A., "Digital Image Processing," (2nd Edition) Addison-Wesley, Reading, Massachusetts, 1987.

Klinger, A., Fu, K.S., Kunni, L. (Eds.) "Data Structure, Computer Graphics and Pattern Recognition," Academic Press, New York, 1977.

Offen, R.J., (Ed.) "VLSI Image Processing," Collins, London, 1985.

Pavlidis, T., "Structural Pattern Recognition," Springer-Verlog, Berlin, 1977.

Pavlidis, T., "Algorithms for Graphics and Image Processing," Computer Science Press, Rockville, Maryland., 1982.

Pratt, W.K., "Digital Image Processing," Wiley, New York, 1978.

Rosenfeld, A., Kak, A.C., "Digital Picture Processing," (2nd Edition Vols. 1 and 2) Academic Press, New York, 1982.

### **Organic Vision: Neurophysiology and Psychological Issues**

Beck, J., "Effect of Orientation and of Shape Similarity on Perceptual Grouping," *Perception and Psychophysics*, 1, pp. 300-302, 1966.

Beck, J., (Ed.) "Organization and Representation in Perception," Lawrence Earlbaum Assoc., Hillsdale, New Jersey, 1982.

Beck, J., Hope, B., Rosenfeld, A., (Eds.) "Human and Machine Vision," Academic Press, New York, 1983.

Biederman, I., "Human Image Understanding: Recent Research and a Theory," *Computer Vision, Graphics, and Image Processing*, **32**, pp. 29-73, 1985.

Cornsweet, T., "Visual Perception," Academic Press, New York, 1971.

Crick, F., Asanuma, C., "Certain Aspects of the Anatomy and Physiology of the Cerebral Cortex," in *Parallel Distributed Processing Vol. 2: Psychological and Biological Models*, Eds: Feldman, J.A., Hayes, P.J., Rumbelhart, D.E., MIT Press, Cambridge, Massachusetts, 1986.

Fischler, M.A., Firschein, O., "Intelligence: the Eye, the Brain, and the Computer," Addison-Wesley, Reading, Massachusetts, 1987.

Gibson, J.J., "The Perception of the Visual World," Houghton-Mifflin, Boston, Massachusetts, 1950.

Gibson, J.J., "The Ecological Approach to Visual Perception," Houghton-Mifflin, Boston, Massachusetts, 1979.

Gregory, R.L., "The Intelligent Eye," McGraw-Hill, New York, 1970.

Gregory, R.L., "Eye and Brain: The Psychology of Seeing," McGraw-Hill, New York, 1978. World Universal Library (3rd Edition)

Gregory, R.L., "Mind in Science," Weidenfeld and Nicholson, London, 1981.

Grossberg, S., "Cortical Dynamics of Three-Dimensional Form, Color, and Brightness Perception, I: Monocular Theory, II: Binocular Theory," *Perception and Psychophysics*, **41(2)**, pp. 87-158, 1987.

Held, R., Richard, W., (Eds.) "Recent Progress in Perception," Readings from Scientific American, W.H. Freeman, San Fransisco, California, 1975.

Hubel, D.H., "The Brain," *Scientific American*, **241**(3), pp. 39-47, 1979.

Hubel, D.H., Wiesel, T.N., "Brain Mechanism of Vision," *Scientific American*, **241**(3), pp. 130-144, 1979.

Kohler, W., "Gestalt Psychology," Liveright Publishing, New York, 1947.

Lennie, P., "Parallel Visual Pathways: A Review," *Vision Research*, **20**, pp. 561-594, 1980.

Levine, M.D., "Vision in Man and Machine," McGraw-Hill, New York, 1985.

Olson, R.K., Attneave, F., "What Variables Produce Similarity Grouping ?," *The American Journal of Psychology*, **83**, 1970.

Prytulak, L.S., Brodie, D.A., "Effect of Length, Density, and Angle Between Arms on Gestalt Grouping," *British Journal of Psychology*, **66**(1), pp. 91-99, 1975.

Schwab, E.C., Nusbaum, H.C., (Eds.) "Pattern Recognition by Human and Machine Vol. 2, Visual Perception," Academic Press, Orlando, Florida, 1986.

Sejnowski, T.J., "Open Questions About Computation in Cerebral Cortex," in *Parallel Distributed Processing Vol. 2: Psychological and Biological Models*, Eds: Feldman, J.A., Hayes, P.J., Rumbelhart, D.E., MIT Press, Cambridge, Massachusetts, 1986.

Stevens, K.A., Brookes, A., "Detecting Structure by Symbolic Constructions on Tokens," *Computer Vision, Graphics, and Image Processing*, **37**(2), pp. 238-260, 1987.

Zobrist, A.L., Thompson, W.B., "Building a Distance Function for Gestalt Grouping," *IEEE Transaction on Computers*, C-24(7), pp. 718-728, 1975.

### **Detection of Line Segments and Contours**

Ashkar, G.P., Modestino, J.W., "The Contour Extraction Problem with Biomedical Applications," *Computer Graphics and Image Processing*, 7(3), pp. 331-355, 1978.

Casent, D., Krishnapuram, R., "Curved Object Location by Hough Transformations and Inversions," *Pattern Recognition*, 20(2), pp. 181-188, 1987.

Davis, L., "Hierarchical Generalized Hough Transforms and Line-Segment Based Generalized Hough Transform," *Pattern Recognition*, 15(4), pp. 277-285, 1982.

Dudani, S.A., Luk, A.L., "Locating Straight-Line Edge Segments on Outdoor Scenes," *Pattern Recognition*, 10(3), pp. 145-152, 1978.

Fischler, M.A., Wolf, H.C., "Machine Perception of Linear Structure," *Proc. 8th International Joint Conference on Artificial Intelligence*, Karlsruhe, West Germany, pp. 1010-1013, 1983.

Haralick, R.M., Shapiro, L.G., "Image Segmentation Techniques," *Computer Vision, Graphics, and Image Processing*, 29(1), pp. 100-132, 1985.

Illingworth, J., Kittler, J., "The Adaptive Hough Transform," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-9(5), pp. 690-698, 1987.

Nalwa, V.S., Pauchon, W., "Edgel Aggregation and Edge Description," *Computer Vision, Graphics, and Image Processing*, 40(1), pp. 79-94, 1987.

Nevatia, R., "Locating Object Boundaries in Textured Environment," *IEEE Transaction on Computers*, C-25(11), pp. 1170-1175, 1976.

Niblack, W., Petkovic, D., "On Improving the Accuracy of the Hough Transform: Theory, Simulations, and Experiments," *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Ann Arbor, Michigan, pp. 574-579, 1988.

Sankar, P.V., Sklansky, J., "A Gestalt-Guided Boundary Follower for X-Ray Images of Lung Nodules," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-4(3), pp. 326-331, 1982.

Scher, A., Shneier, M., Rosenfeld, A., "Clustering of Collinear Line Segments," *Pattern Recognition*, 15(2), pp. 85-91, 1982.

Shapiro, S.D., "Transformations for the Computer Detection of Curves in Noisy Pictures," *Computer Graphics and Image Processing*, 4(4), pp. 328-338, 1975.

Shapiro, S.D., "Feature Space Transforms for Curve Detection," *Pattern Recognition*, 10(3), pp. 129-143, 1978.

Shapiro, S.D., "Properties of Transforms for the Detection of Curves in Noisy Pictures," *Computer Graphics and Image Processing*, 8(2), pp. 219-236, 1978.

Shapiro, S.D., Iannino, I., "Geometric Construction for Predicting Hough Transform Performance," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-1(3), pp. 310-317, 1988.

Shirai, Y., "A Context Sensitive Line Finder for Recognition of Polyhedra," *Artificial Intelligence*, 4(2), pp. 95-119, 1973.

Shu, D.B., Li, C.C., Mancuso, J.F., Sun, Y.N., "A Line Extraction Method for Automated SEM Inspection of VLSI Resist," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-10(1), pp. 117-120, 1988.

Thrift, P.R., Dunn, S.M., "Approximating Point-Set Images by Line Segments Using Variation of the Hough Transform," *Computer Vision, Graphics, and Image Processing*, 21(3), pp. 383-394, 1983.

Wechsler, H., Sklansky, J., "Finding the Rib Cage in Chest Radiographs," *Pattern Recognition*, 9(1), pp. 21-30, 1977.

Zabele, G.S., Koplowitz, J., "On Improving Line Detection in Noisy Images," *Computer Graphics and Image Processing*, 15(2), pp. 130-135, 1981.

### **Shape From Contour**

Binford, T.O., "Inferring Surfaces from Images," *Artificial Intelligence*, 17(1-3), pp. 205-244, 1981.

Brady, M., Yuille, A., "An Extremum Principle of Shape from Contour," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, PAMI-6(3), pp. 288-301, 1984.

Lim, W., "Using Occluding Contours for Object Recognition," *Proc. DARPA Image Understanding Workshop, Volume II*, Los Angeles, California, pp. 909-914, 1987.

Stevens, K.A., "The Visual Interpretation of Surface Contours," *Artificial Intelligence*, 17(1-3), pp. 47-73, 1981.

Stevens, K.A., "Inferring Shape from Contours Across Surfaces," in *Pixels to Predicates*, (Ed.) Pentland, A., Ablex, Norwood, New Jersey, 1986.

Witkin, A.P., "Recovering Surface Shape and Orientation from Texture," *Artificial Intelligence*, **17**(1-3), pp. 17-45, 1981.

### **Pyramid Based (Hierarchical) Vision Techniques**

Antonisse, H.J., "Image Segmentation in Pyramids," *Computer Vision, Graphics, and Image Processing*, **19**(4), pp. 367-383, 1982.

Ballard, D.H., Sklansky, J., "A Ladder-Structured Decision Tree for Recognizing Tumours in Chest Radiographs," *IEEE Transaction on Computers*, **C-25**(5), pp. 502-513, 1976.

Cantoni, V., Levisaldi, S., "Pyramidal Systems for Computer Vision," Springer-Verlog, Berlin, 1986.

Cantoni, V., Carrioli, L., "Structural Shape Recognition in a Multiresolution Environment," *Signal Processing*, **12**(3), pp. 267-276, 1987.

Duff, M.J.B., (Ed.) "Intermediate-Level Image Processing," Academic Press, London, 1986.

Duff, M.J.B., Levisaldi, S., "Languages and Architectures for Image Processing," Academic Press, New York, 1981.

Levisaldi, S., (Ed.) "Integrated Technology for Parallel Image Processing," Academic Press, London, 1985.

Levisaldi, S., "Issues in Parallel Algorithm for Image Processing," in *The Characteristics of Parallel Algorithms*, Eds: Jamieson, L.H., Gannon, D., Douglass, R.J., MIT Press, Cambridge, Massachusetts, 1987.

Meer, P., Jiang, S., Baugher, E.S., Rosenfeld, A., "Robustness of Image Pyramids under Structural Perturbations," *Computer Vision, Graphics, and Image Processing*, **44**(3), pp. 307-331, 1988.

Rosenfeld, A., (Ed.) "Multiresolution Image Processing and Analysis," Springer-Verlog, Berlin, 1984.

Samet, H., "The Quadtree and Related Hierarchical Data Structures," *Computing Surveys*, 16(2), pp. 187-260, 1984.

Tanimoto, S.L., Klinger, A., "Structured Computer Vision: Machine Perception Through Hierarchical Computation Structures," Academic Press, New York, 1980.

Uhr, L., (Ed.) "Parallel Computer Vision," Academic Press, Orlando, Florida, 1987.

Wilson, R., Spann, M., "Image Segmentation and Uncertainty," Research studies Press, Hertfordshire, England, 1988.