THE CITY UNIVERSITY

DEPARTMENT OF SYSTEMS SCIENCE

VEHICLE GUIDANCE BY AUTOMATED SCENE ANALYSIS

DJAMEL GUENTRI

A thesis submited for the award of the degree of Doctor of Philosophy in Systems Engineering.

August 1981

Table of Content

ACKNOWLEDGEMENT

FIRST,I WOULD LIKE TO THANK MY PROJECT SUPERVISOR,PROFESSOR L. FINKELSTEIN,HEAD OF THE PHISICS DEPARTMENT OF THE CITY UNIVERSITY, WHO DIRECTED THIS WORK GENERALLY,FOR HIS COMMENTS AND CONSTRUCTIVE CRITICISM IN THE FINAL WRITING OF THE THESIS.

I AM MOST GREATFUL TO MY SECOND PROJECT SUPERVISOR,MR L. NORTON-WAYNE FOR BRINGING MUCH INFORMATION TO MY ATTENTION,AND IN HELPING ME GENERALLY IN MY WORK.

I AM ALSO GREATLY IN DEBT TO [redacted] WHO HELPED ME TYPE THIS THESIS.

LAST,BUT NOT LEAST,I WOULD LIKE TO ACKNOWLEDGE THE HELP OF MY FAMILY,AND PARTICULARLY [redacted] WHOSE SUPPORT DURING THE LAST THREE YEARS HAS BEEN INVALUABLE.

ABSTRACT

This thesis describes research directed towards extracting information sufficient to guide a vehicle through a street,using individual monochromatic street scene images.This description emphasizes more the procedures and the techiques,used during the analysis,than the structure of the FORTRAN programs in which they were implemented.

The research consists of a study in three principal parts. A sequence of processes capable of distinguishing the street from the remainder of the scene is presented and illustrated to describe the first part. The second part involved the development of a technique,based on photogrammetric principles,to measure distances inside the street. The third and final part involved locating the obstacles inside the street.

The first chapter attempts to place the work in the wider field of artificial intelligence,analyses the task of driving,and defines the principal aims of the research.The second chapter identifies the different facets of image analysis,and reviews in depth the state of the art in this field. The third chapter reviews the different techniques which are available for image processing.Because of the importance of the human eye-brain system for future developments in machine vision,chapter outlines the physiological structure and the

psychological bevaviour of the human visual system. The chapter five discusses the many computational facilities used during the research.

The chapters six,seven and eight,describe the three principal parts of the research.The chapter six describes the extraction of the street from the remainder of the street scene.The chapter seven describes the principles and the technique used for calculating distances.The chapter eight describes a technique for locating obstacles inside the street.

The conclusion and general discussions are presented in chapter nine. In the conclusion,the results are discussed,the need for futher research is emphasized,and suggestions for future work are outlined.

1 INTRODUCTION

1.1 Summary Of The Research

The present thesis describes work in a quite recent field of science refered to usually as artificial intelligence.Artificial intelligence uses as its tools computers,and does things ,that would require 'intelligence' if done by people.The concept and definition of intelligence are discussed below.Although the work described is not directly concerned with the general theory of artificial intelligence,the description,in this introductory chapter,of this science will help to place the work in wider context.

With the generalisation of the utilisation of computers and under the influence of cybernetics,there has emerged a new scientific discipline whose goals are to make computers more useful and to understand the principles which make intelligence possible.This new discipline,under the influence of Wiener(1948), emerged in 1950,with the publication of Alan Turing's book,'Computing Machinery and intelligence'.In its beginning it was primarily concerned with the development of working computer systems and their potential for imitating human thought processes,and its approach was philosophical and speculative.But in the last two decades a large amount of research has been undertaken to develop practical systems for machine vision, representation of knowledge,speech recognition and synthesis,learning and robotics.Generally ,research carried out in recent years has tended to be in specific fields of artificial intelligence,for example

probleme solving (chess),machine vision and speech,but very little research combining all the aspects of the science has been undertaken to develop a system whose intelligence compares well with human beings.

The phrase artificial intelligence has been used because of the close relation between this kind of intelligence exhibited by machines,and the 'natural' intelligence exhibited by human beings and generally by animals.

The term intelligence has not a very precise definition,although many diverse definitions have been put forward.It has been defined by Terman (1921) as'the ability to carry on abstract thinking',by Woodrow (1921) as 'the capacity to acquire capacity',by Thorndike (1921) as 'the power of good responses from the point of view of truth or fact',by Burt (1955) as'innate general cognitive ability', by Wechsler (1958) as 'the aggregate or global capacity of the individual to act purposefully,to think rationally,and to deal effectivelly with his environment', by Boring as 'the capacity to do well in an intelligence test'. A general definition is that intelligence has something to do with the ability to adapt to the environment,the capacity for learning and the ability for abstract thinking.The previous definitions were,mainly,given by psychologists,and were on the whole concerned with human intelligence.When the concept of intelligence was extended to machines,by Wiener (1948) and Turing (1950),its definition became wider and was centered on the ability of

processing information.

The idea evolved from the concept of a simple feedback mechanism.With the increasing distinction between machines as tools and machine as prime movers, which encroach on the preserve of human intelligent qualities,such as handling non numerical data of all kinds,learning from past experience to abstract features common to different problems,and extracting general truths by inference from examples (induction), the concept of simple feedback has been extended to systems capable of storing and retrieving information(computers).The storing and retrieving are done in such a way as to optimise the extraction of information towards some predetermined goal.The processing consists generally in discarding partially,temporarily or totally information not required for a particular goal,and enhancing whatever is required.

This new understanding of intelligence is based on the concept of information ,and the realisation that the outstanding characteristic of man is not simply his ability to use tools(monkeys use sticks),but its ability to formulate complex ideas(abstract thoughts),and to communicate them.The concept of information has recently been formalised and information defined in such a way that quantities of information can be measured and treated as physical quantities such as force or energy (Shannon (1950)).

The modern theory of information is based on the two basic postulates that the purpose of information is to reduce

uncertainty,and that the information measure of two statistically independent messages should be equal to the sum of the information measures of the two messages. Any message which has high probability of occurence ,conveys little information,thus most information is conveyed by the least probable message. It has two general orientations,based on a common probabilistic base:one developed by Wiener and the other by Shannon.Wiener's work centered on developing techniques,which permit recovering from a signal,corrupted by noise or non linearity,an original uncorrupted message.Shannon carried this a step further by providing an information measure which was defined as the logarithm of the inverse of the probability of occurence of the message.He then proceeded by providing rules which govern the translation of information from one form of encoding to another (1st law),showed that information can be transmitted without any error over a noisy channel at a rate dependent only on its bandwidth and signal/noise ratio (2nd law),and also showed how the extraction of information from enciphered messages depends on redundant information provided in the message,leading to the concept of unicity distance. Hamming(1952) devised ways for checking and correcting errors during communication,by introducing redundancies.Huffman(1962) devised a method for the construction of minimum redundancy codes.

The theory has been successfully applied to the transmission of electrical signals.Because light is a very important source of information,especially for biological systems,and with the advances of electronics,the application of the theory has been extended to

the optical information processing and communication (Gabor(1961)).

Intelligence may be regarded as an ability to process information. Based on the concept of information we will define an intelligent system as a system endowed with a representation of the world,which it constantly updates,and having the ability of processing the data, gathered by its sensors,for communicating and resolving problems caused by its environment,with the purpose of optimising its operational time (its life) or the operational time of the system which comprises it.The intelligece test for a machine might be,as suggested by Turing (1950),the test of rational conversation.

A major discipline of artificial intelligence is machine vision ,which is trying to develop a machine,based on computers,which would imitate the human eye-brain system,and with appropriate simplifying constraints,is proving to be a possible tool for a wide range of applications.Current industrial applications range from simple systems that measure or compare to sophisticated systems such as character recognition,label reading and registration,metal strip inspection,silhouettes recognition and printed circuit board inspection. Although all industrial robots today work in very constrained environments, with known objects carefully presented, there is an emergence of programmable automation in which the robot,using visual and tactile sensors,can react to its environment.

Through the 1960 very few industrial robots(Unimation Inc

provided almost the sole exception) were available, but since then there have been numerous attempts to develop robots of one kind or another(Nottingham University System SIRCH,Heginbothan(1973)).In 1978 10,000 industrial robots were installed over the world,and 110 firms offered 250 different types .Examples of the products available are 'Unimate' which is a digitally controlled machine,'Trallfa' which is an analog controlled robot that was developed specifically for spray painting applications,and 'Auto-Place' which is a 'pick and place' robot with mechanically adjustable travel and limited motion sequences, used as an automation system component.Industrial robots ,which embrace three categories(remotely manned,automated and self contained types),have a broad range of applications including die casting,forging,stamping,welding,molding,machine tool operations, spray coating,material transfer and assembly.

The two immediate objectives in robotics are the multifunctionalisation of hand mechanisms ,and the development of walking or driving machines. The possibility of realising a versatile walking robot,operating under computer control,has been greatly enhanced in this decade.They could be octoped,hexaped,quadruped or biped,and have the ability of being all- terrain vehicles.This research could be seen as part of the drive to improve the capability and flexibility of robots(with the provision of real time visual feedback,adaptive corrections of the trajectory of the robot can be made so as to automate its guidance)

The intense interest and significant research in machine vision,has been motivated by the importance of human vision in sensing and interpreting the environment.It is estimated that 75 per cent of the information received by a human is visual.Despite possible improvements after laborious training of tactile,auditory and olfactory sensory capabilities,it is hardly necessary to emphasise the severe handicaps and limitations imposed on the mental and physical activities of the blind. For the past decade,many researchers, among whom Rosenfeld (1980) is noteworthy,have been slowly developing understanding in this field,and has began to implement simple but increasingly sophisticated machine vision techniques in medical diagnosis,manufacturinng processes, photo-interpretation and vehicle guidance.

Machine vision has as its purpose to derive pragmatic information ,useful for the execution of a given task,from the symbolic information describing an image. The image will have been formed by an optical system(camera) from an arrangement of bodies forming a scene which may be diverse and three-dimensional. The objects in the scene have to be recognised and their spatial relationships determined,to provide a description appropriate to a particular application,from the raw image.Irrelevant visual data is discarded,while needed relationships between parts of objects is deduced from their optical projections.Machine vision is concerned with the generation of pragmatic descriptions and uses these descriptions to permit a controlled system take appropriate actions.Much still has to be learned about what kind of pragmatic

descriptions of the three dimensional reality can be reasonably derived ,by a vision system,from an image of intensity readings.

Machine vision has been largely derived and built on subjects such as image processing,pattern recognition and scene analysis.Image processing concerns itself with the production of new images,by application of techniques from linear systems,from existing images so as to improve their appearance to a human viewer.It comprises mainly restoration and enhancement,both of which deal with symbolic information only. Although all the techniques of image processing are not relevant to machine vision (e.g. restoration),some of them ,for processing grey level images,are still of importance.Scene analysis is largely concerned with the transformation of descriptions into more abstract descriptions,such as transforming a line drawing description into a description of three-dimensional solids and how they relate spatially.The above example requires that the system can extract a good line drawing from a grey level image.Scene analysis could be seen as comprising a set of transformations which start with symbolic information,which is composed of the basic building block symbols(intensity readings),and finishes with pragmatic information for practical use. Intermediate stages in the processing involve syntactic information (rules limiting allowable combinations of symbols),and semantic information involving abstract descriptions of what is perceived in the scene.This latter information is required to obey constraints of semantic consistancy.

Parallel to the development of techniques for machine vision ,from image processing and scene analysis,there has been another approach (Nitzan(1977),Horn(1977))based on understanding of the physics of image formation.Understanding how the image proprieties change with the lighting,shape and surface material of the objects being imaged,will help to provide methods for inverting the imaging process by exploiting the physical constraints and thus building a three-dimensional description of the scene.

Machine vision research,in common with most of the work of artificial intelligence is still conducted primarily as an experimental science.Although a complete scientific theory of machine vision is not yet available ,there has been a large accumulation of methods for processing images.It is therefore ,sensible to use what is known,to situations in which simplifying physical constraints can be applied to yield viable solutions to simpler ,more constrained but general classes of tasks.This approach might suggest new strategies and directions for the elaboration of a general theory.

The object of the research was chosen to fulfil two requirements.The first was that the research should investigate present image analysis techniques,adapt them or develop new ones if required.The second was that the chosen problem should be original and make a contribution to the engineering science.Another aim was that the work could easily be adapted for practical applications.Given the restrictions imposed by the definition of the

problem,the choice which was arrived at was to try to lay the foundations for the automatic guidance of vehicles.To try automating the guidance of all kinds of vehicles in all kinds of situations is a formidable task and could not be undertaken with success ,in the short time allocated to the research.Because automatic vehicle guidance in completely unconstrained environments,is not feasible ,given the current state of the art ,a compromise,which still leaves a very complex and general task, was adopted.This comprised guidance in a semi constrained environmemt,such as a network of street and roads.

Having defined the general context,and the overall aims of the research, a precise definition of the problem is necessary.For two reasons the work does not aim at designing a machine which will be operational at the end of the research.First,the amount of time allocated for the work is too short,and secondly the resources,ideally required, for the task are not available ,at present ,in the university. The work described in this thesis does not concern itself with the design of hardware.Instead,its aim is to provide and implement in software a methodology to extract information ,from street scene images,which will be used to automate the guidance of vehicles. However the eventual implementation of the methodology in hardware is not ignored.

Four principal methodologies have been investigated for scene analysis for vehicle automated guidance .Levine(1979) has used colour in images to delineate different objects with high

reliability, but his program takes eight hours per scene on a P.D.P.10 computer, and may thus be considered too cumbersome.Since human beings work satisfactorily with monochrome images anyway,the complication of colour seems unnecessary.

The second method consists of working with a time sequence of images. By analysing pairs of consecutive images it is possible to extract objects moving against a constant background,even when the camera is moving. Several investigators such as C.L.Fennema and W.B.Thompson(1979) and notably H.H.Nagel(1976) have used monochrome images considered as a time sequence; Nagel extracts nodal points in the scene,which are associated by observing their behavior as the time sequence proceeds.Points fixed in the background seem to move radially outwards,for example.

A third possibility is to use images in stereo pairs,with analysis of the stereoscopic displacement used to determine distances .The analysis consists first of getting a three dimensional image by convolving the two stereo images , then ,using the three dimensional image to isolate the street from the rest of the image and locate the obstacles inside it.The approach is well described by D.B.Genery (1979) whilst Hans.P. Moravec (1977) has actually produced a working system which drives a vehicle through cluttered environments (though certainly not through streets) under computer control guided solely by images perceived through an onboard TV camera.

Our approach notes that a one eyed human being can drive vehicles perfectly satisfactorily,using as input single monochrome images only.This method has the advantage of using far less input data than the alternatives and the advantage of the availability of a rig for digitising black-and-white images,based on a 1732 element CCD array line scan camera.

To recapitulate let us define precisely the object of the research.It is to develop and implement in software,a methodology which will extract from a single monochrome image,information which could be used to automate the guidance of a vehicle driving along a street.

1.2 driving

Driving involves the guidance of a vehicle along a street.Visual data describing the view in front of the car is processed and information is extracted so that the position ,the speed and the direction of the vehicle can be adjusted so as to approach a particular destination in the shortest time.At the same time,the driver must avoid collisions and maintain an appropriate position in the road(there are rules such as keeping to a particular side of the street and interpreting road signals,which must be obeyed).

An important simplifying constraint in driving is that it is usually done on bounded two dimensional surfaces,hence the system,used, will concern itself only with objects which could reasonably be encountered in a street(usually other vehicles and signals). Because of this,although there are more constraints in driving and the speed of driving is much greater than that of walking,on the whole the information necessary for driving is less than that required for walking.Thus automating driving could be considered as a first step in the development of an all-terrain walking robot.

The concept of driving emerged from the concept of moving from one place to another using wheels and outside agents(horses,mechanical motors...) to do the work.The use of this outside agent was necessary not only because the human being was reluctent to use his own biological motor system but because its power was not sufficient to move fast enough.To solve this problem of speed humans used animals which were faster,stronger and more resilient than they were.So they mounted horses,donkeys,lamas and elephants,but the real breakthrough came with the invention of the wheel.

The car is in effect a fast cart with a mechanical motor instead of a horse,and is less intelligent than the hybrid system formed by the horse and the chariot.It is notable that,although horses are not considered intelligent by human standards,they are quite capable,with only minimal human assistance, of guiding

vehicles through streets. The problem of the energy used in driving has been adequately dealt with(chemical power from oil,electric power,etc...).But the control of a mechanically propelled vehicle is still very much a human activity.Driving is still a task achieved only by humans.It can be simplified by mounting the vehicle on rails,so that it operates in a highly restricted environment. So it would be advisable to start with the analysis of the way humans perform this task and try to develop a machine based system which will do the same.

Humans use their visual system to drive and an extensive study of it has been carried out by psychologists.The system does not consist of the eyes only,but of the eye-brain system,with the brain doing ,nearly, all the processing,and the eye serving as a light sensor with some preprocessing capabilities.The discovery of how the brain recognises complex patterns,such as real objects,in real scenes is still in its infancy and much work is still to be done before all outstanding questions can be answered.

But if we concern ourselves only with simple driving problems,it should be possible to achieve some useful results.Very little has been done to relieve humans from the burden of control in driving with its requirement for constant attention.However manipulations of the environment, in which driving is done,have been carried out very early in history to make the driving much easier.These manipulations were the construction of roads,of bridges and later on of motorways,and also painting of lines on the

streets,and the utilisation of traffic lights and road signals.An important development which makes driving much easier,and which gives some hopes for the complete automation of driving,is that vehicle are usually constrained to move inside bounded regions such as roads and motorways.

The economics of developing an automatic system for driving are beyond the scope of this thesis,but we can try to give a general idea of what is involved.Most driving is done by individuals who own their own cars and enjoy driving. Automated vehicle guidance through streets is thus seen,initially,chiefly as a potential aid to drivers to reduce fatigue.So the unemployment ,which will be caused by its complete automation ,will not be very important.The jobs which will be created if the system is produced in a large scale and the indirect jobs in software development which will result from developing the system,will be sufficient to compensate for the lost jobs.The adoption of the system will also upgrade the activities of human beings,from controller of a mechanical process(driving),to jobs which hopfully will require more skills.

1.3 Strategy for automatic guidance

A strategy for automatic vehicle guidance needs to determine the kind of information to be extracted from the scene ,which is needed for controlling the vehicle.There are mainly two approaches

for automating this guidance.

Conventional methods for guidance require the provision of rails,painted lines,beams of electromagnetic radiation,buried wires or other modifications of the environment,which the vehicle is constrained to follow.One such modification could be carried out in motorways,where the control of the car would then be taken over by a system which would sense the coordinates of the car(for example its position and the carriage way it is in).To control the car we could have a simple mechanical system: in the middle of each carriage way there would be a small notch in which slide a piece of metal,comming out of the car,which will guide the car(fig 1.1). The above example has to resolve problems caused by friction,and safety such as when the metal breaks.

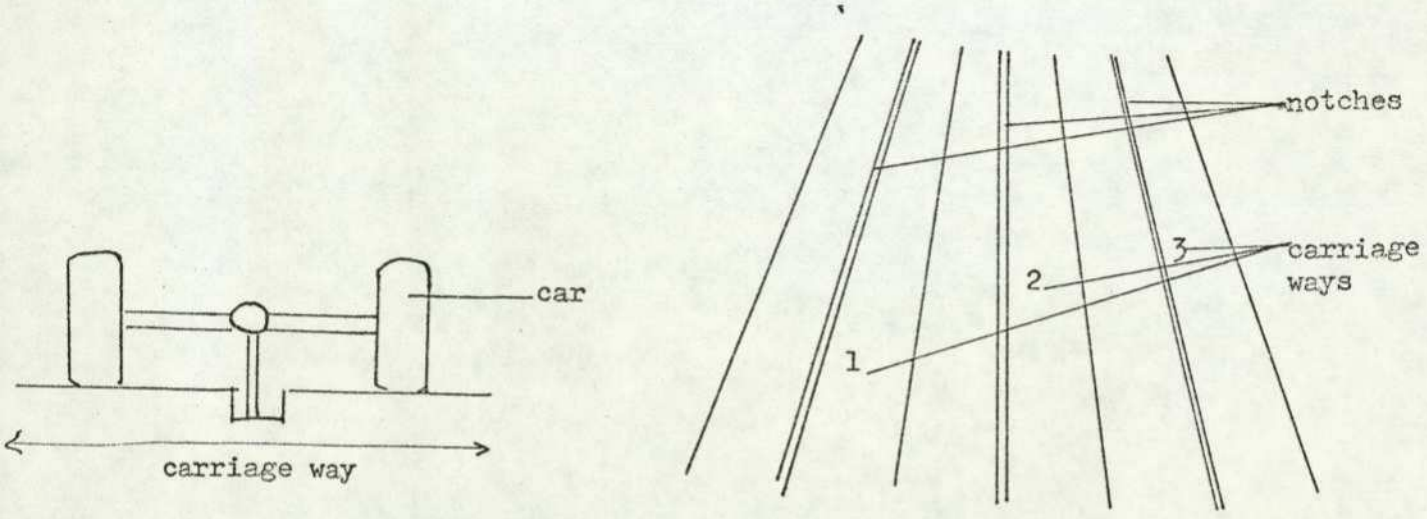


Fig 1.1: Possible system for controlling vehicles.

In the case of systems based on rails or buried wires,a simple malfunction of a small part of the system may bring to a halt,if not all the system, a large part of it.White lines must remain unobscured to provide effective guidance.In nearly all existing systems for guiding vehicles automatically,vehicles cannot generally pass one another,and evasive action to cope with unexpected alterations to the guidance track (eg,an obstacle in the rails) is generally impossible. At best the vehicle can only sense the obstacle and stop.

In many applications where vehicles are used (eg ,on the farm,the battlefield ,under the sea or in remote and hostile environments) the provision of rails ,lines or wires is expensive or impractical .Thus alternative methodologies must be provided.

The most flexible and powerful approach to vehicle guidance involves a human driver,who can examine the scene in front of him visually,extract information such as the size and location of obstacles including other vehicles (which may be moving), and take appropriate action. Thus there is reason to automate the activity of the human eye-brain system,to produce ultimately an automated vehicle guidance system, which uses visual information only.

The possible applications for a vehicle guided automatically by machine vision are numerous.They include operation in hazardous environments such as battlefields,mines, areas exposed to radiactivity,toxicity and extreme temperature,inaccessible regions

such as deep space and under water ,and applications requiring continuous operation for many days,as might occur in agriculture to take advantage of brief periods of fine weather.More immediate application might be found in driving vehicles along motorways ,where the environment is more constrained so the problem is simpler.This would enable a human driver to relax his attention.

The second approach , is much more complex,but if successful,would be much more interesting than the first one.It consists of guiding vehicles in an existing partially constrained environment.To do so,we have first to sense the environment and the the different objects contained in it. When we have catalogued the different objects which are around us,we then have to decide which of those are obstacles which must be avoided.

With the current state of scene analysis such a project will be very difficult ,but if we restrained ourselves to the less general problem of a vehicle moving in a street,some progress toward its resolution could be achieved.The environment could be sensed with radiation of any wavelength ,but it would be advantageous to use light because it is freely available during the daytime,and because it has been used for a long time by all sorts of living creatures to guide themselves in all kinds of environments(by studying such creatures ,we could develop algorithm which will copy them).During the night time infrared devices could be used.The techniques of analysing the images would remain the same for both light and infrared images.

By working in a semi-constrained environment,such as a network of streets we are simplifying our problem,but not trivialising it.Driving cars in a street is not a simple problem ,even for human drivers.Even attempting to resolve completely the problem in such a semi-constrained environment,would be a task which would require a long time to yield results.The approach is then to simplify the problem by making reasonable assumptions which will be described below.But before let us describe the general strategy which has been adopted.

The following approach has been adopted.Firstly much of the information contained in the scene is redundant so far as the required analysis is concerned,and must be discarded as early in the process as possible. This is achieved by dividing the scene into regions of two kinds, the first containing information of interest in the subsequent analysis,the second containing information of no immediate interest or no interest at all.Thus the objective is to locate as early as is possible in the analysis,using search at low resolution,the first region. The region of immediate interest ,in this project,comprises chiefly the inside of the road and particularly its boundaries.Some object lying completly outside the road,such as road signals,traffic lights and people going to cross a particular street,could be relevant in a more sophisticated analysis.

Hence the first step in the strategy is to extract the road from the rest of the scene as viewed by a camera,mounted in front of

the vehicle, simulating normal driver's view.

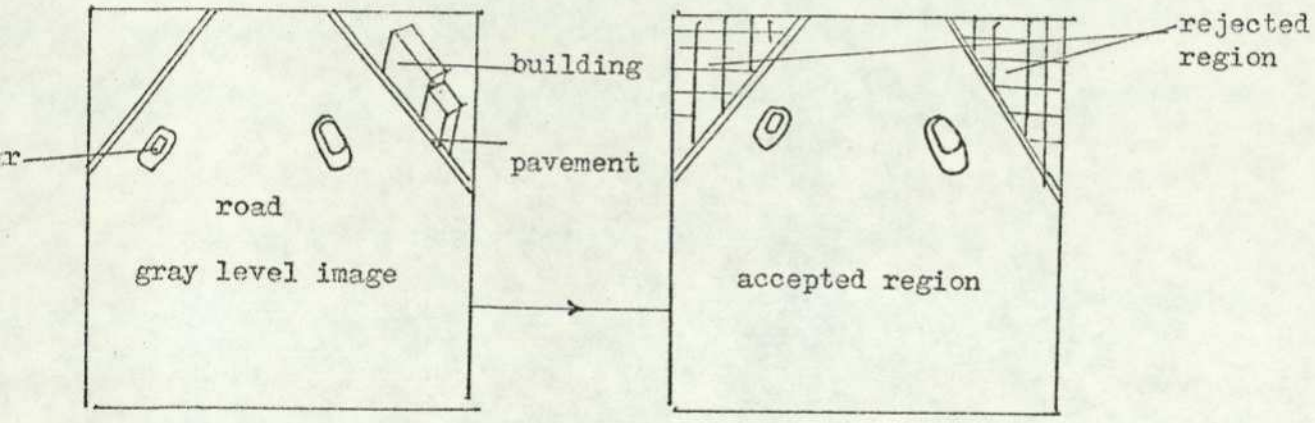


Fig 1.2: First step of the strategy.

The second step in the strategy would be to develop a method for evaluating distances from a single image when given the height and the tilt angle of the camera. This could be achieved by techniques of photogrammetry,which is the science of obtaining reliable measurement by means of photography in order to determine geometric characteristics such as size and form and position of the photographed objects in a scene.

The third and final step in the strategy will be concerned with locating and identifying obstacles inside the danger zone which is the zone of the road which if occupied by an obstacle would result in a collision because the dynamics of the vehicle are such that its path cannot be modified in time. The determination of the exact

boundaries of such a zone will depend on the dynamics of the car and of the safety factor used.Having located an obstacle we will use the method developed in step 2 to calculate the distance of the obstacle from the controlled vehicle.

It would be very complicated to apply this strategy to every kind of street scene.If we want to obtain results ,we first have to start by resolving simpler problems,which can then be used later to resolve the general problem. Therefore ,instead of analysing street scenes in general we will concentrate on a particular set of street scenes whose general form is illustrated by the following diagram:

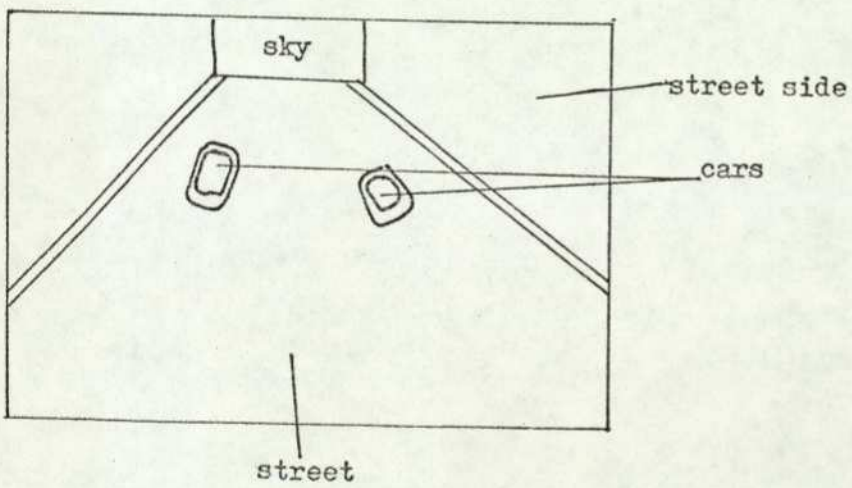


Fig 1.3: Model for street scene to be analysed.

Such picture could be taken by a camera mounted on the top of a car and tilted in such a way as to limit the view at a given distance from the car as illustrated by the following diagram:

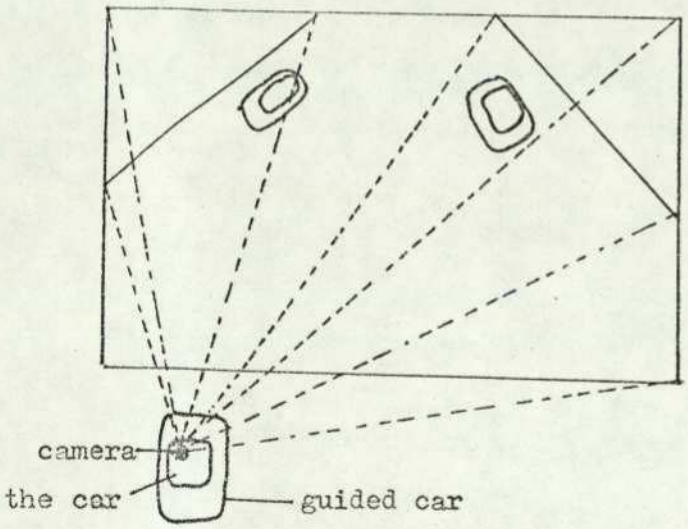


Fig 1.4: View from above of a vehicle with a mounted camera on the top in a street scene.

So we will deal with streets with approximately straight borders,and with cars located randomly in it.Each vehicle must be isolated from the others(no occlusion). This assumption could easily be ignored in future work.

The whole strategy for analysing the images could be summarised by diagram 1.From diagram 1 we can see that the driving system could be considered as a feedback system illustrated by diagram 2.

The camera would be the transducer which converts light into electrical signals. The control system ,instead of being a system with a simple transfer function is a complex one which is supplied with symbolic information,and which gives as output pragmatic

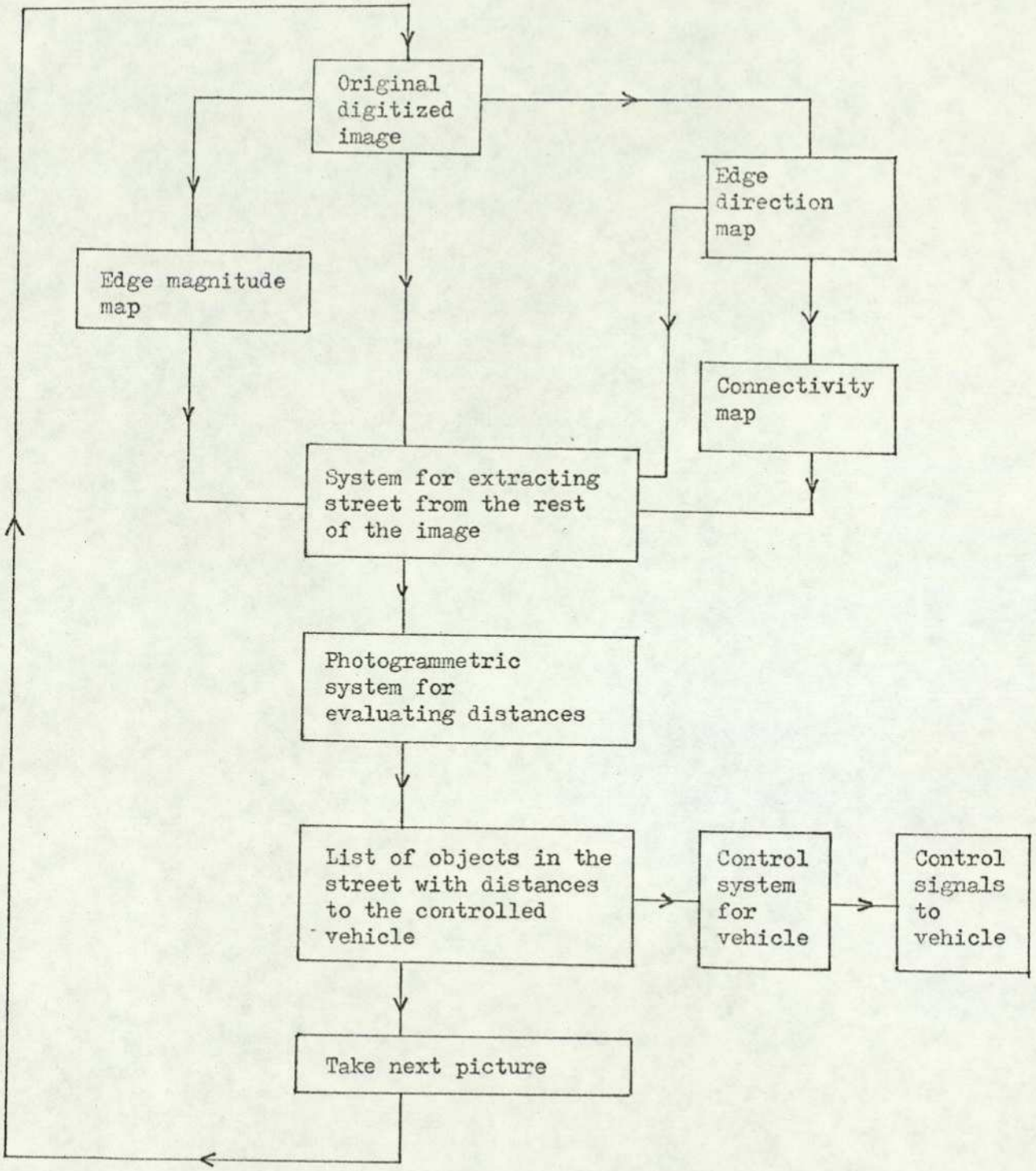


Diagram 1 :Strategy For Image Analysis.

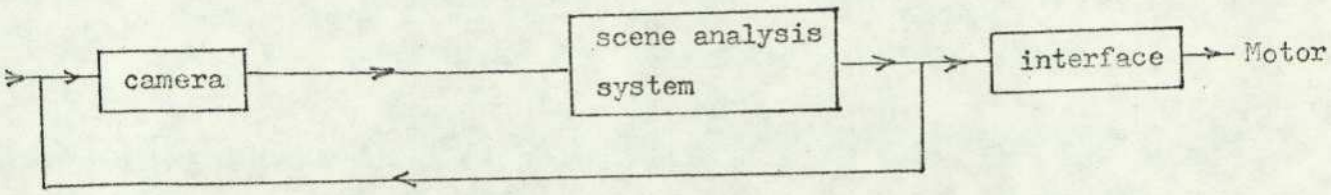
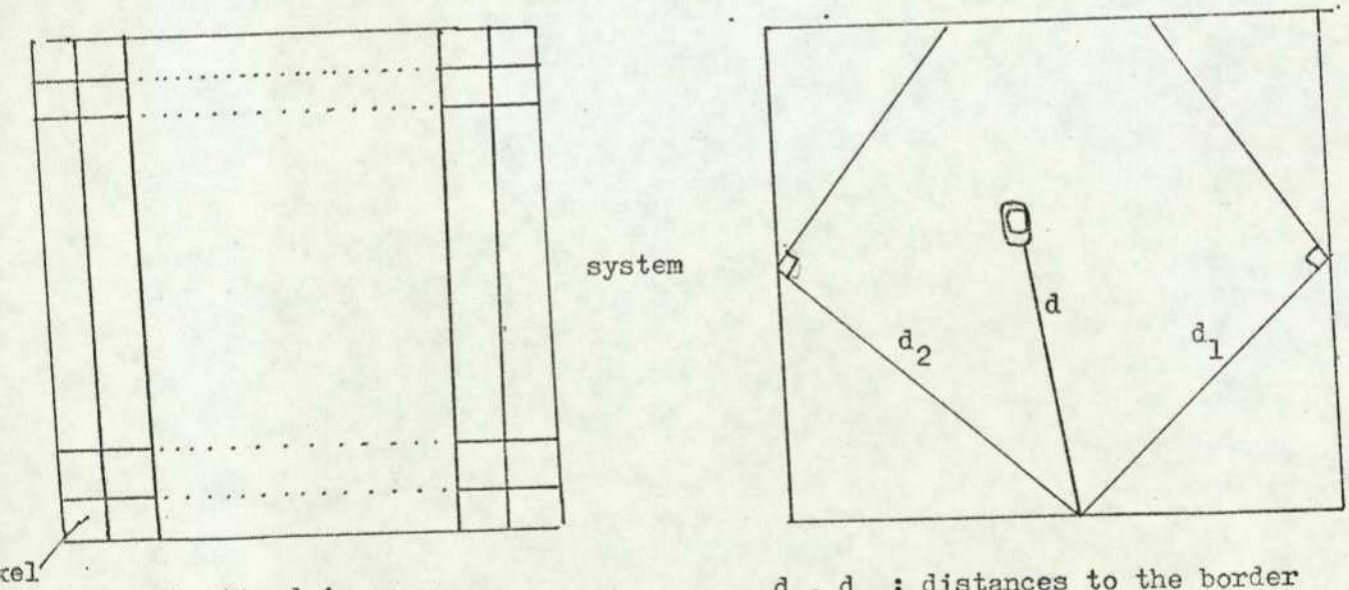


Diagram 2 : Driving Feedback System.

information. The system thus differs fundamentally from ordinary control systems, which take as input electrical analogue signals and work with symbolic information only. Input symbolic information is represented by grey levels of the pixels in the image and the output pragmatic information describes significant properties of the road in front of the vehicle, such as distances from the guided vehicle to different obstacles in the street as illustrated below.



digitised image
grey level 0 - 255

d_1, d_2 : distances to the border of street.
d : distance to obstacle

Fig 1.5: Input-output of the system.

Before finishing this introductory chapter ,I would like to stress that the strategy described above would in its final stage involve dynamic scene analysis which is the analysis of temporal features of the visual scene.Although this thesis will specially concentrate on scene analysis taking as input a single static image ,the overall strategy is based on dynamic scene analysis which takes as input a sequence of static images with a given time function relating the order and the elapsed interval between elements of the sequence. The idea is to develop a methodology to extract information from a single image and update this information by repeating the process for all the elements of the sequence,as illustrated by diagram 3.

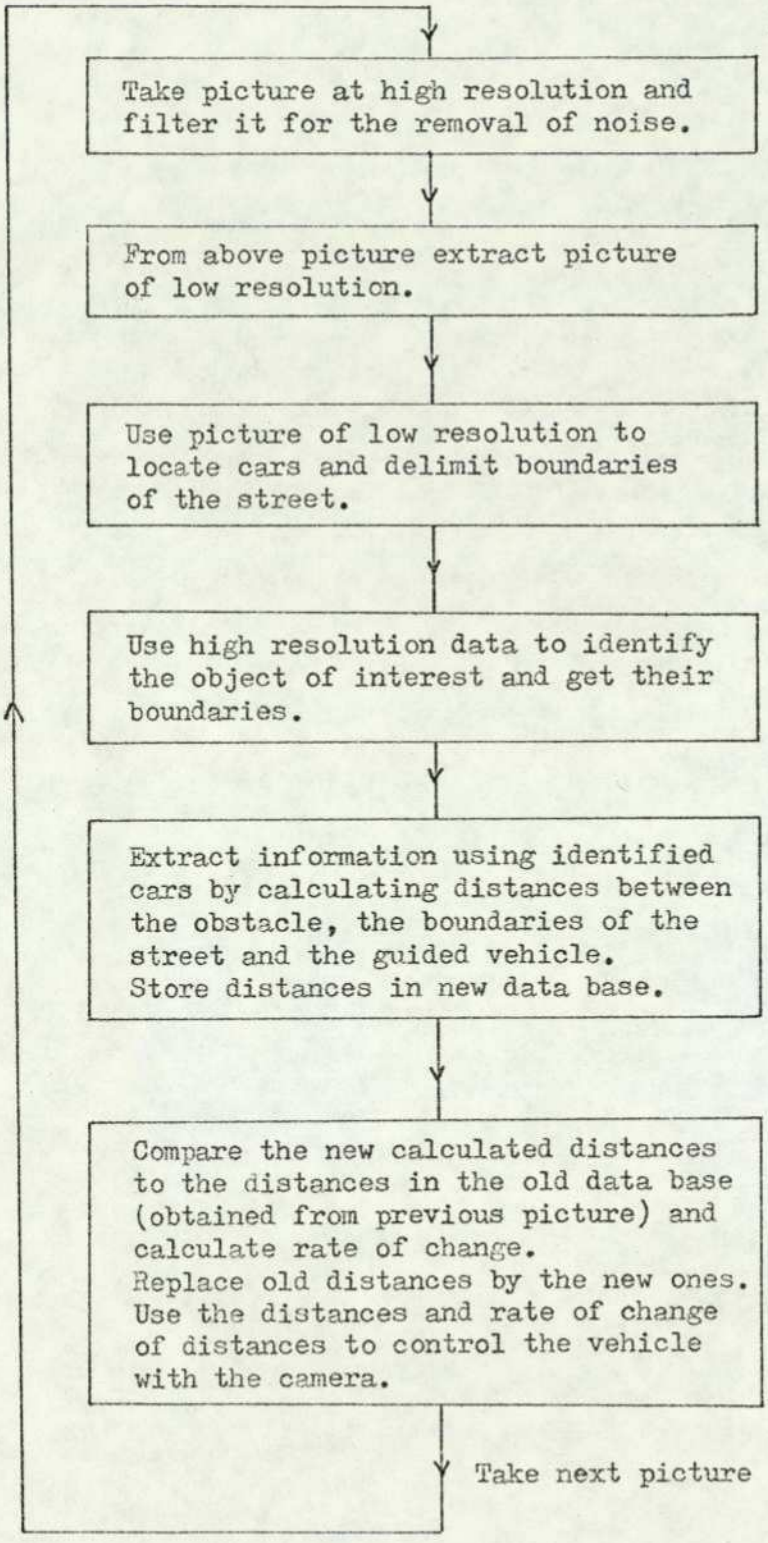


Diagram 3 .Strategy for dynamic scene analysis for automatic driving.

2 STATE OF THE ART IN MACHINE VISION

2.1 Introduction

Among the major problems,in the design of intelligent robots,such as natural language understanding,model building,planning and monitoring of performance, perception is the most important.At present computer visual abilities are admittedly primitive in comparision with the extensive human visual abilities. Visual perception,which is necessary to the developement of intelligent machines capable of loading,unloading,positioning,transfering,feeding,orienting and manipulating different kinds of objects,and moving independently in unconstrained environments,is a process where computer abilities are primitive and in need of urgent improvements.Visual abilities added to the mechanical hand motion capabilities of existing industrial robots,will permit them to interact more efficiently with complex environments,and will widen their field of applications.All actual robots have very limited visual abilities and operate in environment restricted by very strict heuristic rules.Machine vision is becoming more and more an essential part of flexible and versatile automation systems.Already limited special purpose vision systems which incorporate visual recognition,memorisation and inspection capabilities,are appearing in increasing numbers on factory floors.Standard units scan a field of view in a thousandth of a second,obtaining digitally coded signals,which are further processed digitally on a real time basis, and thus making possible the automation of visually controllable production operations whose

speed,accuracy and cost effectiveness exceed human capabilities. For advanced applications,only improvement in the machine perception of its environment ,could provide adequate flexibility.

A wide variety of approaches,to develop machine vision,have been followed. All of them rely on a common body of knowledge,which is constituted by a set of techniques for processing two-dimensional images,but use different strategies.Basically the knowledge used by a vision system,is split between knowledge imbedded in the algorithms,which are the implementation of the specific procedures designed to analyse the input data,to infer properties and to classify object in the scene,and explicit knowledge of representation models stored in the memory.

The data upon which the entire processing is based,is the digitised image array .Generally,an A/D converter is used to transform the analogue optical data from a camera into digital intensity readings,which are then stored as a two-dimensional array in a computer memory or a mass storage.The major task is then to develop techniques that can analyse the array and extract semantic and pragmatic information.The array is processed so as to find and classify the objects in the scene.The difficulty in extracting the needed information from an image,varies with the picture complexity and the kind of information to be extracted from the image.

Vision systems usually start with some preprocessing of the digitised image. This generally involves enhancement and

restoration,which will be described, in a later chapter in more detail.The preprocessing must be chosen with care ,as not to destroy needed information,and to reject as much as possible of the information which is not needed.

The next step in picture processing is the segmentation of the image.The various techniques used in segmentation have been surveyed by Riseman and Arbib(1977). The first step in image understanding is,starting with the preprocessed image,to organise the image into different disjoint regions.There are three general approaches for segmenting an image into disjoint regions of uniform or nearly uniform intensity values.The first finds the regions of fairly uniform intensity by growing regions (Brice and Fennema(1970),Pavladis (1976),Tanenbaum and Barrow(1976)).The second approach consists of thresholding at calculated levels (Ohlander(1975),Price(1976)),or clustering in multidimensional space with two or more characteristic features (Schater,Davis and Rosenfeld(1976)). The third approach starts by locating discontinuities in intensity(edge points) and connecting them to produce a closed edge or boundary by forming closed curves (McKue(1975),Ramer(1975)).A major part of the research has concentrated on segmentation using edge detection.Nagin(1979) evaluated the performance of two segmentation algorithms based respectively on symbolic labelling of pixels based upon cluster analysis of histograms,and a probabilistic relation labelling algorithm,and concluded that more research is necessary before developing an evaluation measure for their respective performance.

Having outlined the object in the image,the final step in the design of a machine vision system,is to extract semantic information from the segmented image. This step is usually refered to as scene analysis.The extraction of semantic information involves recognising and locating the objects in the scene,and giving a simple but precise description of the scene,which generally involves determining region descriptors such as perimeter,area,maximal to minimal diameter ratio and number of interior holes,and classifying the regions into classes.

2.2 Segmentation

2.2.1 Region growing

The first of the three image segmentation approaches assumes that the image of a surface is uniform or smooth in its local properties such as grey level,colour and texture.It aims at segmenting the image into a set of nearly uniform intensity regions,and hence reducing the data necessary for the description of the image.

Brice and Fennema (1970) started with digitised television images,and partitioned them into homogeneous connected

regions.They then proceeded to group adjacent atomic regions (region with elements having the same intensity),if their properties are not too different.For joining atomic regions together ,they used a criterion based on the fact that the boundaries between two regions belonging to the same surface,are generally not so strong as boundaries of regions of different surfaces.Regions sharing the weakest boundary were iteratively merged until only regions with boundaries stronger than some threshold,or a fixed number of regions,remained.After partitioning the image into its natural regions,they used a simple straight line-fitting to represent the boundaries in terms of long straight lines.As a final step they used a scene analyser to interpret the line drawing in terms of wedges,cubes,walls and floor. The segmentation based on the simple intuitive idea of surface intensity uniformity,does not deal adequately with complex images because natural regions in those images do not have approximately uniform intensity,and ,also,do not always have distinct boundaries (shadows).

Yakimovsky and Feldman(1974) combined mathematical decision theory with problem dependent information for partitioning a complex image such as a road scene image,into meaningful regions.They took a point in the image and grew around it a region consisting of all image points which can be connected to the starting point,and which have an intensity difference with the starting point below some threshold.they then sequentially merged atomic regions into larger regions,using decision theory

techniques and problem dependent information.

Tenenbaum and Barrow(1977) used knowledge from a variety of sources to merge regions in accordance with their possible interpretations. They used semantic information to segment simple outdoor and office scenes. They interactively labeled initial atomic regions with a specified size,and blocked the merging of two regions with different labels,and when regions without labels exceeded a certain size they interactively labeled them.The innovation in their procedure is that it includes a criterion for stoping bad merges, which would otherwise occur with the simple Fennema's technique.

Horitz and Pavlidis (1976) used a split and merge technique to perform segmentation without any kind of information other than the symbolic information contained in the image.They segmented the image by construction and manipulation of data structures.Using a segmentation tree and a split-and-merge method ,which consisted of starting with an arbitrary cutset of nodes (collections of pixels possessing some common proprety) at the same level of the tree and proceeding downward(splitting) or upward (merging) in the tree,they attempted to form the largest regions with intensities in some specifed bounds.

2.2.2 Thresholding and Clustering

The second approach in segmentation consists of thresholding and clustering. Thresholding is usually based on grey level,colour or local propreties,such as gradient and Laplacian,histograms.Ohlander (1975) developed a thresholding technique which was very effective for colour images.He started by computing histograms of various colours and hues,and then thresholded the image at the clearly separated peaks.a separate peak was defined in such a way that the ratio of maximum peak to minimum peak was greater or equal to two.He iterated the process for each segmented part of the image until no separate peaks were found in any of the histograms.He also separated textured regions ,which were only thresholded to segment the sky line,from the uniform ones by convolving the image with the Sobel edge operator,by defining the textured areas as the ones with more than 24 edges in a 9x9 window,with the thin edge boundaries not counted.All his processing was done interactively.

Price(1976) extended Ohlander's thresholding technique.He made the system almost completely automatic and greatly reduced the processing time by a planning technique which starts by segmenting initially lower resolution average images.He worked with colour and monochrome images.But he encountered difficulties when he used the intensity space on its own,and had to use texture to achieve an acceptable segmentation.

Another approach to segmentation is to look for clusters in a multidimensional space rather than using many one dimensional histograms (Schachter,Davis and Rosenfeld (1976)).Clustering consists of using two or more characteristic features,such as intensity,colour and texture,and determining the distinct clusters in the space of the characteristic features.These clusters are then assumed to represent natural regions in the image.The procedure consists of using a clustering method,usually derived from classical pattern recognition techniques ,to group points,in a given space,into clusters which are then mapped back to the two-dimensional image so as to segment it.

2.2.3 Boundary Formation

The motivating factor for this method is that most of the information within an image is contained in the boundaries between different regions.This assumes discontinuity in local properties,particularly intensity,between images of two different surfaces.The motivation is validated by the fact that biological visual systems appear to make use of edge detection and not of thresholding. The methods involves finding discontinuities in intensity (edges) and using these to segment the image into regions of approximately uniform intensity, hence transforming a colour or monochrome image into a binary line drawing image. This achieves a big reduction in the data representing the

image.The process ,generally,involves many steps.The first involves determining the magnitude and sometimes the direction of the edges,by convolving the image with edge detection masks (Davis(1975)).The second step involves some filtering and thresholding for rejecting edges which are not needed.This could involve thinning (Rosenfeld and Thurston(1971)) and skeletonisation (Fu(1975)).The third step involves linking the edges into line segments which are grouped to form the boundaries of different regions in the image (Perkins(1976)).

Early research concentrated on images containing polyhedra,and hence formed of straight lines (Roberts(1965),Horn(1973)).

The formation of segments involves eliminating false edges and merging the edge points into longer edge segments,refered to,usually,as streaks,which are then combined into boundaries.The false boundaries are then rejected.The formation of boundaries could involve heuristic searches (Montanari(1971), dynamic programming (Martelli(1976) and relaxation (Rosenfeld(1977),Lester(1978),Riseman and Arbib(1977)).Because the output of the differentiation process is usually very noisy,relaxation is used to clean it.It involves viewing edge strength as a probability which measure the confidence of the existence of an edge.The probability evaluates the compatibility of an edge with its neighbouring positions (Zucker(1976)).one example is to compare the direction of an edge to its neighbours

directions,and analyse their compatibility.

In the past a major problem with region segmentation,using edge points,has been that sometimes false edges are accepted,and some edges are missing so detected edges do not always form closed curves.Expansion-contraction techniques have been very successful for forming closed curves in binary images.Baird(1977) has used a smoothing and gap filling technique to get closed curves.Jacobus and Chien(1978) used circular operators of decreasing sizes,to fill the gaps.

Recently,Perkins(1980) developed a method based on expansion-contraction technique to resolve the problem caused by small gaps in lines.His method involved finding abrupt discontinuities in intensity(edges) ,and using them to segment an image into region of approximately uniform intensity.He first expanded the segments to close the gaps and then contracted them after the seperate regions have been labeled.His method comprised an edge finding operation,a thining operation,an expansion of active edges,a labelling of uniform intensity regions,an elimination of small uniform intensity regions,a shrinking of edge regions,a placing of the edge boundaries between uniform intensity regions,a determination of the active edge regions ,and finally an elimination of the small edge regions.

Prager(1980) has also succeeded in segmenting natural scenes.His method consisted of some preprocessing to clean up the

raw data,of the differentiation of the image,using edge masks,and then of a relaxation process to consolidate edges on the basis of local consistency requirements.The individual edges were linked to form extended line segments and low confidence lines were removed.

Another method for connecting edges together is to fit lines (Duda (1972)) or curves (Shapiro(1975)) through the edges,using Hough transform or other techniques for fitting lines.

2.2.4 Other Segmentation Methods

Although the bulk of the research in segmentation has concentrated on the segmentation of static scenes using boundary formation,region growing and thresholding and clustering,there has been attempts to combine some of the techniques together.Norton-Wayne and Guentri(1980) combined thresholding with edge detection techniques.Levine (1976) used clustering and edge detection to segment colour images.

An original method for image segmentation was developed by Will and Pennington (1971).They decomposed a scene subjected to a special form of illumination, a projector with a crossed grating,into elementary planar areas.Their method consisted in coding the various planar areas as the modulation on a spatial

frequency carrier grid so that the extraction of the planar areas became a matter of linear frequency domain filtering.

Another way of segmenting images is to use more than one image.There are mainly two approaches which respectively use a sequence of static images, with a given time or spatial function relating the order and the time intervals between the element of the sequence,viewed from the same position (dynamic scene analysis:Nagel(1980),Martin and Aggrawal(1978)),and from different positions (stereoscopic analysis and tomography).

2.2.4.1 Dynamic Scene Analysis

Dynamic scene analysis differs from the,just discussed, static analysis (analysis of single images) in that the information is extracted from several images considered in sequence,and consists of the integration into a coherent whole,of the information extracted from each element of the sequence.In the past,dynamic analysis has concentrated mainly on designing motion detection systems.Leese,Novak and Taylor(1970) determined the cloud pattern motions from geosynchronous satellite images.

Potter(1975) developed a system for segmenting a scene using motion.He used a cross shape template generated from the

first image and used a heuristic search to find a match in a second picture.He then calculated the velocity value of the match by comparing the position of its center in the first and second picture,and by grouping together points with the same velocity he was able to segment the first image.

Lillstrand (1972) and Ulstad (1973) concentrated on determining areas of change,determined by a simple subtractive process between two consecutive images of the same scene,which were rectified before the subtractive process.

Nagel (1976) used a specific form of image differencing,to extract a single moving object from a dynamic scene.The analysis was done on a sequence of images with two consecutive images chosen interactively.The first step in the analysis was the segmentation of each frame,using a modified Yakimovsky's region growing method.The second step was performed on these regions,and involved comparing consecutive images to find regions whose intensity distributions were similar.The unmatched regions were used as an estimate of the moving object. by performing cross correlations of the regions boundaries,the velocity vectors were calculated and used to normalise,superimpose and threshold the unmatched regions,so as to finally extract the moving object from the scene.

2.2.4.2 Stereoscopic Analysis

Dynamic scene analysis can only segment an image if there is something moving in the scene.When we are dealing with a static scene,the alternative to dynamic scene analysis is to use stereoscopic analysis.The analysis is carried on two stereoscopic images which are correlated to get a three-dimensional representation of the scene,and hence make its partitionning into its different components much easier(Marr,Palm and Poggis(1978),Gennery(1979),Stato(1979)).

Gennery (1979) used two stereo images to locate the street in a street scene. He correlated the two stereoscopic images to get a three-dimensional representation of the image,and by grouping the points with very low relative hight,extracted the street from the rest of the scene.

2.3 Scene Analysis

Scene analysis is the last step in image understanding.It involves outlining the objects in the scene and labelling them with appropriate semantic interpretations. Using models (sets of heuristic rules),it recovers intrinsic scene characteristics ,for the interpretation of the segmented image.The main aim of scene

analysis is to recognise ,at least some,if not all of the objects which are contained in a scene.

This recognition could be as simple as comparing the properties of the segmented image regions with those stored in memory and describing the objects to be recognised,and determining the best match. However this is possible only for very simple scenes, usually composed namely of cubes,wedges and prisms.In the general case, because three-dimensional objects in three-dimensional scenes are not generally completely characterised by the two-dimensional propreties which could be calculated from the segmented image,this simple approach to recognition is not viable and additional semantic information has to be used with the segmented image to locate objects in complex scenes.

Scene analysis could be seen as a transformation of images containing millions of bits of symbolic information (two-dimensional array of intensity readings) into a semantic description with a much smaller number of bits. A complex street scene could hence be illustrated by a semantic description such as 'street with n cars'.This upgrading of information from symbolic to semantic is achieved by suppressing irrelevant data,integrating regions in the image into meaningful entities,characterising objects by their distinguishing properties such as size and shape,and then recognising the objects by matching them with a description hold in the memory.

Scene analysis is generally goal dependent,in that the final description will depend,very much,on the kind of information to be extracted.For a driver,road signals will be important in a street scene and therefore have to be incorporated in the final description of the scene. For someone,who is walking on the pavement,on the other hand, the road signals would be completely irrelevant and threfore would not be incorporated in the final scene description.To be efficient,the final semantic description would have to use just the minimum required information for providing pragmatic information necessary for a specific task.

Many existing scene analysis systems had been tailored for specific applications such assembly line component recognition,medical image analysis ,satellite imagery analysis,bubble chamber image analysis or specific indoor or outdoor scene analysis.Due to this specificity of the research,approaches to scene analysis have been very fragmented,and the development of a vision system,which would compare favourably in versatility or capability with humans,has not (so far as is known) been achieved.Nonetheless appreciable progress has been achieved in the last two decades.

Roberts (1965) was the first to develop a complete methodology for the interpretation of two-dimensional images as a scene of a collection of three-dimensional objects.His work was done entirely on scenes composed of simple polyhedrical objects arranged in well-specified ways.He first,through a series of operations, reduced

a digitised television image to a perfect line drawing,and then matched the description of the line drawing against a set of computer models of a defined set of geometrical objects.The geometrical objects he used were namely cubes,rectangular wedges and hexagonal prisms.For these very constrained scenes he was able to recognise the polyhedrical objects and compute their exact three-dimensional positions,with respect to the camera,by using information given by the spatial orientation of the model with the best match.Although the system worked only for a very special case of scenes,the techniques,used,remain dominant in scene analysis.

Guzman(1968) worked with scenes with solid,hole-free polyhedrical objects arranged randomly ,and assumed a perfect line drawing of the scene.His program ,SEE,partitioned the image into regions on the basis of body membership of the surfaces.He was able to partition the image into a set of regions,with each set depecting an individual object,by using two passes over the image.The first pass made local guesses about which pairs of regions depict the same body.The second pass produced sets of regions corresponding to bodies,by searching a graph with regions as nodes.The main contribution of Guzman was to resolve the problem of partially occluded objects.

A basic problem with Guzman's method was that it assumed having a perfect line drawing,which is true,only if the environment is severely restricted. Falk (1972) developed a system,INTERPRET,consisting of an integrated collection of picture

processing programs,which were able to interpret line drawings with small number of lines missing,and produced a three-dimensional representation of the scene.He worked with scenes containing a range of nine fixed prototypes with known position and orientation of the ground plane relative to the picture plane.His system involved five stages(SEGMENT,SUPPORT,COMPLETE,(RECOGNISE and PREDICT),and VERIFY),and took as input imperfect line drawings of scenes.

Using Guzman's vertex classifications to assign edges to bodies,SEGMENT partitioned the line drawing image into bodies.SUPPORT tried to determine the bodies that could conceivably support each body in the scene.RECOGNISE tried, after COMPLETE had added the missing lines,to recognise objects by matching (using a series of tests) their features against the stored properties of the prototypes. PREDICT ,using the positions of the prototypes which had been determined by RECOGNISE,involved predicting the possible image of the scene.VERIFY compared the predicted and real images ,and went back to RECOGNISE if any input line had not been predicted or more than three lines had been falsely predicted.

Falk's main innovation was the attempt at dealing with imperfect line drawings. However because he adopted Guzman's vertex classifications,his program still had the same problems as SEE,by almost totally relying on local image based heuristics.

A better inclusive alternative to Guzman and Falk's method,which deduced body membership of two surfaces from the

appearance of the corners they shared ,is the Huffman-Clowes labelling algorithm.Observing that Guzman's function category must have one of a small number of corner interpretations,and restricting themselves to two-line and three-line functions and three-surface corners,Huffman and Clowes used a small set of corners described by predicates convex,concave and occluding.Huffman labeled non-hidden lines with a '+' for convex edges,a '-' for concave edges and arrowhead label for occluding edges with their non-hidden surfaces on the right ,when moving in the direction of the arrow.Both Huffman and Clowes's procedures label edge in the image and recover some of the hidden structure by attaching to the occluding edges,surfaces that are attached to them and are turned away from the viewing direction.This approach has been a substantial improvement on Guzman's strategy.

Waltz(1972) extended Huffman-Clowes labelling algorithm to corners with many surfaces ,and to scenes with shadows.He designed a new algorithm which extended the set of lines labels used by huffman and Clowes ,and improved the mechanism of search for coherent interpretations. His achievement was to extend the use of labelling techniques to less restricted scenes.

Although Guzman(1971) described a system which could deal with complex line drawing images,all the above algorithms have been mainly concerned with scenes containing polyhedra ,and used binary images only.Another approach ,based on finding suitable descriptions for images,is the linguistic approach which was inspired by the

closeness between images and natural language. Kirsh(1964),Ledley(1964),Narashnhan(1966),and Anderson(1968) wrote grammars for restricted classes of pictures,whilst Clowes(1972),Evans(1969),Shaw(1969) and Stanton(1972) attempted the development of description languages for more general images.Although there is some doubt about the adequacy of this approach,research is still going on and is contributing further insight into image understanding.

Describing a difficult scene is much more complex than was thought when machine vision research was first started.Although there has been progress in processing binary images and images representing specific,usually simple scenes,there has not been great success in the development of a universal analyser which would compare in versatility to ,at least animal visual systems.

Current vision software is an amalgam of edge detection,thresholding and clustering techniques,heuristic techniques for finding lines and regions, fixed and probabilistic decision trees and context aided scene analysis, with the help of projective geometry which contribute to modelling the floor plane of particular environments.Futur progress will depend on growths of understanding of how the human vision system works,on the development of special hardware,and on the need of the society at large for such a system.In the coming decade a large number of machine vision systems for particular tasks,are going to be developed,with the evolution of better hardware and better visual

processing software proceeding together and reinforcing each other,and could give insights in the ways of developing a more general vision system.

An interesting approach ,which should be mentioned,involves understanding how the light is reflected by different surfaces with different texture,and tries to reverse the process so as extract a three-dimensional description from a two-dimensional image (Mackworth(1973) and Horn(1975)).

2.4 Existing Machine Vision Systems

Existing industrial machine vision systems are primarily concerned with visual inspection,material handling and restricted automatic motion (Ejiri(1973),Yoda(1973),Agin(1975),Marlow(1975) and Olsztyn(1975)). Most of these systems work almost exclusively with binary images,and are therefore restricted to operation in environments with good contrast between the background and the components,and with simple components whose geometrical outline is sufficient for their recognition. Almost all the systems concentrate on feature extraction and automatic classification procedures.They isolate the components from the background,usually by simple thresholding,and attempt to compute various features,which generally charactarise the position,orientation and shape of the component as a silhouette.

After obtaining a meaningful set of features,the problem of recognition becomes a simple classification problem,using classical pattern recognition techniques,with the machine trained to recognise a set of particular classes.Nonetheless some sophisticated systems,which deal adequately with overlaping (Dessinoz and Cranlund(1979)) and complex shapes (Umetari(1979)),have been designed.

A major part of the research has concentrated on inspection.Many manufacturing processes involve inspection,which is carried by human inspectors who are prone to errors,because of boredom and fatigue.Inspection,generally,involves specific checks for specific defects,and reduces to a classification problem.At The City University two projects dealing respectively with metal plate (Noton-Wayne(1978)) and integrated circuit boards inspection,have been going on from the early seventies and are at present ,at the stage of being introduced in industry.

More complex image analysis systems have been developed to provide vision for robots. Research robot prototypes with visual sensors cover the whole machine vision from artificial intelligence to factory production. Some typical tried and tested systems with television cameras for industrial processes are SIRCH,a robot of Nottingham University (Hoeginbotham(1972)), and Mitsubish Electronic Corporation Robot (Tsubi(1976)). These are generally research tools rather than reliable systems ripe for industrial application.

The SIRCH system collected visual data through a VIDICON television camera which was coupled directly via multiplex control,to a small digital computer(Honeywell DDP 516) with a 12K store,and was restricted to two-dimensional profile components,which were however randomly orientated and presented. The three linear axis and the two rotational axis of the hand were operated by stepping motors.The hand had mechanically coupled gripping mechanisms. The most appropriate gripper was selected when the orientation and the position of the component had been determined,and,for the next part of the operation, the machine worked blind,without any kind of feedback.The one inch vidicon camera with amplifier was mounted above the manipulator in bearing concentric with the gripper axis.The camera control unit and synchronising pulse generator was mounted remotely from the machine.

The Mitsubishi robot used a camera in the center of the hand,with the image processing used for component shape recognition,and guided the hand to pick up the component.The system needed a feeder and was used for carbon brush application.

For systems ,working in less constrained environments,we can cite MARK 1.5, the Edinburgh University Robot (Ambler(1976)),HIVIP,The Hitachi Ltd Robot (Ejiri(1971)), and the recent Hitachi automated factory robot.

MARK 1.5 had an overhead one side view camera,and dealt with a small number of simple components,randomly orientated,but with good

background contrast.The system,which used an individual large program for each specific task,and which involved matching the processed visual data against internal models,had some learning capability but was very slow.

HIVIP ,which was used in assembly,had two cameras.One camera was used for viewing engineering drawings and one camera for viewing the assembly area. The system involved component shape recognition,matching of the drawing with the actual components,at random orientation,and handling the component. Its main disadvantages were that it only dealt with polyhedral components and straight line engineering drawings,and required a good background contrast. The system required very large programs and was very slow.

Among mobile robots,we can cite a computer controlled robot cart developed in Japan by Yoshiham Ambe(1972).It transported material in a factory or office ,and loaded and unloaded automatically.It had one front wheel,which furnished drive power and was used for steering,and two rear wheels.The CART-B model was controlled by a mini computer which received the modulated visual data from the cart through wireless communication devices.

Munson(1970) developed a system consisting of two parts,a computer and a mobile vehicle,connected by radio.Two stepping motors were used to drive independently wheels on either side of the vehicle.The vehicle carried a vidicon television camera,whose

focus,iris settings,and tilt and pan angles were controlled by motors,optical range finder and control logic which routes commands from the computer to the appropriate action sites on the vehicle.

Computer commands were used to control power switches,to request readings of the status of various vehicle register,and to control various interrupts. To monitor collision with obstacles,the vehicle was equiped with 'cat- whisker' touch sensors.Two special radio links,for narrow-band telemetring and television video signal transmission,were used.

The environment was very specific.It consisted of a laboratory open space with a number of simple geometrical shape three-dimensional objects,placed in various configurations.A more sophisticated mobile vehicle,with stereoscopic vision was developed by Moravec(1972).

3 IMAGE PROCESSING TECHNIQUES

3.1 Introduction

Image processing,which is constantly evolving,comprises a large set of techniques. These include representation,coding and compression,restoration and enhancement, segmentation,object three-dimensional reconstruction and recognition of visual data.It has mainly two aims.The first,which involves restoration and enhancement, is to improve certain image qualities so as to facilitate the extraction of specified information by a human viewer.The second aim,which could be seen as including the first,is to extract automatically specific information from an image with only the minimal help,or no help at all,from a human viewer.

Computer grahics,which deal with the problem of generating and displaying images synthesised from mathematical models using computers,and which concentrate on perspective views and shading (grey level determination) of surfaces of three-dimensional objects,could be seen as part of image processing,but is usually considered as a separate subject,and will thus not be included in this overview.

The history of computer image processing is relatively brief.It is in 1951 that a committee of the National Coal Board,chaired by Dr Bronovski,was set up to investigate 'the possibility of making a machine to replace the human observer'(Walton(1952)). In the last decades,a host of image scanning and image display devices have been

available, but,there remain many problems with digital image display devices and hardcopy output which are still quite slow and of low quality.

Image processing is part of data processing,threfore many factors responsible for the expansion in general data processing such as the hardware development,are contributing to the evolution of image processing. Because of the large amount of visual data involved,real time image processing succeeded the batch mode,only in the early 1970's,with the development of faster hardware and bigger solid state memories.

The Jet Propultion Laboratory,which was assigned,in the early 1960 by NASA,the task of providing television communication with the unmanned and manned space probes,largely contributed to the rapid development of image processing in the last decade. It provided television images, with acceptable quality,by scanning, digitising and transmitting the video signal to earth stations,where error correction and several restoration and enhancement techniques were used to obtain images with certain qualities that were sometimes missing in the transmitted signal.The development of these techniques was carried further with the increasing utilisation of remotely sensed visual data in many fields of science.Cal Tech,Lawrence Livermore Laboratory,Los Alamos and Jet Propulsion Laboratory have been the main pioneers which contributed to this development.

A major factor in the development of image processing have been the automation of certain cognitive processes such as counting and classification of simple objects such as particles,cells and chromosomes,with the sole use of visual data. As early as 1955,Causley and Young(1955) developed a flying spot microscope which counted and sized red cells at 1 micron resolution and generated a size histogram,going from 0 to 30 microns in steps of 5 microns,of all cells in a 500x500 microns field,with a classification performance of 98 per cent.

Image processing consists of a large number of techniques for transforming symbolic information into more suitable symbolic information or into semantic and pragmatic information for the more advanced image processing systems. Many techniques,in conjunction with efficient data base for image representation, help to reduce the massive symbolic information content of an image to an acceptable amount by eliminating redundant data and rejecting irrelevant information.

During image processing,images are first scanned,digitised and stored in a mass storage such as a magnetic tape or disc.Then,many techniques are used to improve image quality for human viewing,or to extract part,or all,the semantic information available in the image. Due to this availability of hardware,many image processing techniques have been developed and applied. Most of the research in the field has been application oriented.Although most techniques were developed for specific applications,many are also of general

application.

There are a large number of systems which perform various functions regarded as image processing. (Booth and Schroder(1977),Gambino and Schrock(1977) and Wells(1977)).Many of these systems are interactive. The field of application of these image processing systems is extensive.It includes applications as varied as document storage and retrieval,high energy physics,earth resources inventory,spatial exploration ,industrial visual automation and inspection,and medical application such as tomography and ultra-scan. Usually,a general image processing system involves preprocessing,which is concerned with the modification of the image in order to simplify subsequent analysis steps(enhancement and restoration),segmentation, region property measurement,classification and recognition.In general,this preprocessing reduces greatly the complexity of the data.

The understanding and solution of the problems of image processing would play a crucial role in the development of a general machine system capable of competing with the human visual system.Although many of these problems have been resolved,improved and optimal solutions are still needed.

It is not our purpose to describe all the available methods for image processing ,instead we will concentrate on the common methods that are in general use.In the rest of the chapter,we will therefore concentrate on providing an extensive,although not

complete,introduction to the general techniques commonly used in image processing.

3.2 Imaging

The imaging process,which preceeds image processing,involves projecting three-dimensional points of a scene onto a two-dimensional image plane. The projection,which,because of the process involved in the image formation,is called a perspective projection,involves viewing the scene through a camera lens. Although there has been a generalisation of the concept of an image to include two-dimensional arrays of signals possible gathered from sources such as radar and sonar which are not strictly imaging sensors,we will concentrate on images obtained from reflected light(electromagnetic waves with wavelength in the range of 350 microns,for violet,to 750 microns,for red).

Generally any material or system,which has any of its propreties influenced by irradiation with light,may be used for image reproduction and display.Recently many electronic imaging systems such as solid state imagers and image tubes, have been developed(SPIE(1978)).

Optical data recording and reproduction systems involve many different techniques of encoding continuous tone images.The encoding

could be stochastic or multilevel. In the case of stochastic encoding,image tones are achieved by a statistical distribution of activated and inactivated image elements,and the image,generally, suffers from noise.Whilst in the multilevel encoding,the structure of the image elements is more involved and consists of elements which can take a finite number of values(grey levels).This method,usually,leads to insufficient tonal resolution when only a low number of levels is used.But it does not suffer from noise.

The classical techniques for recording,storing and reproducing visual data are silver halide photography,half tone printing and television.Recently there has been an increase substantial increase of recording optical data in digital form,using computers with mass storage facilities such as magnetic tapes and discs.

Silver halide systems,involved in photography,consist of silver halide microcrystals of 0,1 to 1 micron serving as light sensors.The microcrystals are randomly distributed within a gelatine layer,and could be optically sensitised to enable the recording of colour images.For colour image representation,The silver crystals are sensitised different parts of a radiation,and are distributed at different sites of a suitably designed layer system.In the additive system,the layer system consists of colour screens containing equal areas of the principal colours (blue ,green and red).Whilst in the substrative system,it consists of different layers,which are in a multilayer arrangement,and are sensitive to different colours (blue,green and red).

3.3 Image Representation

The efficiency of image processing algorithms depends closely on image representation. In general,an image can be represented as a real valued non-negative function of two variables,f(x,y).The image function,f(x,y),has to be proportional to the light intensity impinging on the picture at the point (x,y).In general,for any image function ,there is an upper bound T to the possible brightness of all image points,and thus the image function is bounded as follows:

$$0 \leqslant f(x,y) \leqslant T$$

For black-and-white images,the value of the image function f(x,y) at the point (x,y) is sometimes refered to as a grey level or intensity reading. For colour images representation,vector valued image function or several image fuctions (blue,green and red) have to be used.

Generally,for images of real scenes,the function f(x,y) cannot be represented in a simple analytical form.This fact added to the fact that sensor technology is becoming more and more discrete,and that a great deal of image processing seems to result in fundamental numerical analysis,require that images must be represented in a discrete form.The conversion of a continuous range and tone image into a discrete and numerical form,is a two-step process involving sampling,followed by quantisation.

Sampling the image involves using the value of the image

function at a number of regularly spaced discrete points in the X_Y image plane.The simplest sampling method involves partitioning the image plane by a quadrupled grid,and sampling the image function at the center of each cell.

In the more general case the sampling theorem provides a mathematical basis for the sampling process.Provided that the sampling is sufficiently dense,the two-dimensional sampling theorem states that a bounded function f(x,y) may be represented in terms of a sequence of its sampled values at equispaced points (iX,jY):

$$f(x,y)=\sum_{i=-\infty}^{\infty}\sum_{j=-\infty}^{\infty} XYf(iX,jY)p(x-mX,y-nY)$$

The sampling theorem provides guidelines for the selection of discrete spatial values from the available continuous region.Sampling is normaly followed by quantisation which is the discrete representation of the magnitude of the image function,so it may be stored and processed in a computer system.The concepts involved in quantisation have been extensively studied (Panter and Dite(1951),Max (1966),Pratt(1970)).

During quantisation,the range of the image intensity,from black to white,is divided into uniform intervals and only a finite number of discrete grey level values are used.

For image representation ,the usual sampling resolution is of the order of 525x525 pixels,with the quantisation involving integer

values going from 0(black) to 255(white).For the purpose of computer image processing,images are represented ,not as continuous image fuctions,but as discrete,usually two-dimensional,array of non-negative integers (two-dimensional matrix). Image database systems are usually based on a raster (matrix) representation or a chain-coded vector representation of boundaries.The vector representation is more compact,but is only useful for the representation of line drawing images.

The conversion of computer images back into viewable form is.an important part of image processing.Interpolation theory and the modelling of the human viewing process are the determining tools in the design of the display devices, which are usually electronic devices.Benning(1969),Davis(1969,the IEEE (1971) proceedings and the conference record of display research(1980) give a good introduction to the devices,which are available, and the principles which are involved in their design.

In some cases,it is worthwhile noting that it is convenient to consider images as stochastic processes or,in the case of images for computer analysis,as samples of stochastic processes.One specific case is texture,which is an important component in human perception of the character of visual surfaces.It has been systematically analysed and many measurements of textural information have been devised, with a variety of others still being investigated (Harralick(1979)).Examples of textural measurements are coarsness,which measures the size of the texture elements ,and

directionality which measures the variation of coarsness as a function of direction.

3.4 Coding and Compression

With the present trend tawards using digital instead of analog techniques in transmission and storage of electrical signal in video-phones,teleconferencing and outer space telecommunications,coding and compression,which is concerned with data reduction and is used to remove redundancy inherent within the raw digitised data,are of major importance in data processing. In the case of image storage in digital form,where the number of information bits (symbolic information) involved is often very large,it is often desirable,if not necessary,to efficiently code and compress the raw data.

The earliest widely used encoding technique for digital signal transmission was Pulse Code Modulation (PCM),Which was first used in 1939 and considerably refined since then (Deloraine and Reeves(1965)).This method,generally distorted very little the image,and involved the sampling on a grid of points and the encoding of the samples in a binary code.The code words may be of unequal length with the shorter words used for the grey levels that occured more often (Jelineck(1968)).

The Pulse Code Modulation transmission of images involves ,first the sampling in the spatial domain to produce an array of discrete samples,and then the quantisation of the image intensity in,2 to the power k,levels. In general subjective tests are used to determine the optimal spatial and tonal resolution. In most cases,the optimal value of K,for a spatial resolution around 525x525 pixels, is about 8 bits per pixel.

If a slight degradation of the image is acceptable,a more efficient technique for data compression would be the Differential Pulse Code Modulation (DPCM), which is based on an invention by Cutler(1952).The Differential Pulse Code Modulation technique can reduce the bit rate from 8 to 1 or 2 bits per pixel,and preserve an acceptable image quality (O'Neal(1966)).

A further reduction is possible for sequences of dynamic images,with interframe coding techniques,where correlation between frame is used,frame differences are transmitted (Haskell,Mounts and Candy(1972)).

In the special case of binary images with man made symbols (documents,engineering drawings,weather maps and graphics),Run Length Coding is one of the most efficient techniques for data compression (Cherry,Kubbu,Pearson and Barton(1963)).

Most coding and compression scheme result in a degradation of image quality.In general,the degree of compression which can be

achieved depends on the nature of the image and is proportional to the image degradation which can be tolerated.

3.5 Restoration and Enhancement

Both restoration and enhancement,which are sometimes refered to, in scene analysis and machine vision,as preprocessing,aim at improving image quality for human viewing or futher processing.They involve the transformation of an image into a modified form more suited for subsequent analysis. The transformations involved in enhancement and restoration operate,usually, only on symbolic information.They involve different kind of filtering for correcting distortions by deblurring and noise removal.Although there is not a clear demarcation between restoration and enhancement,and even sometimes restoration is considered as a particular class of enhancement,generally restoration involves the removal or reduction of definite degradation such as defocussing,to obtain the ideal image wich should have been obtained in the abscence of the particular degradation.On the other hand,enhancement is broader in scope ,and involves transforming the image in a form suitable for a particular purpose,not necessarily the ideal form (original image without degradation). This could involve edge sharpening,thresholding,false colour utilisation and smoothing.

3.5.1 Restoration

Restoration plays a major role in image processing.It is involved in many fields such as outer space communication and fields involving imaging. One of the most stricking success in restoration,is the one achieved by the Jet Propulsion Laboratory with the transmission of good quality images from the moon ,Mars and lately Saturn,by compensating for various degradations such as random noise,interference,geometrical distortions,contrast loss and blurring. Research is still going on for possible compensation of blurring due to atmospheric turbulence.

Image restoration attempt to recover the maximum amount of information concerning the original scene,by the inversion of the degradation process involved in the imaging process,such as blurring introduced by optical systems and image motion,and noise due to electronic and optical sensors. When considerable knowledge,for correctly modelling the imaging system,is available ,restoration is easily achieved.But in the majority of cases,the modelling of the imaging system is not possible,or involves laborious and expensive calculations.

All restoration techniques require some form of knowledge concerning the degradation process.This knowledge consists,in general,of analytic models, or other a priori information,of the imaging system,the lighting and the spatial environment where the

imaging was performed.The different techniques model the degradation and try to inverse the process so as to recover the original image.

Restoration techniques could be seen as consisting of a priori techniques,which aim at designing devices giving images without degradations, and a posteriori techniques which aim at improving degraded images.The most recent a priori technique is adaptive optics where the degradations are measured, and deformable optics are controlled to eliminate the degradations in real time.

The a posteriori techniques,which are the relevant techniques in image processing, have concentrated on linear degradations.In many cases,the ideal image function ,f(x,y),and the corresponding degraded image function,g(x,y),are related as follows:

$$g(x,y)=\int\int h(x,y,a,b)f(a,b)dadb + v(x,y)$$

v(x,y): random noise

h(x,y,a,b): spatial degradation function depending on the position of the point,(a,b),in the ideal image.

When the noise is ignored and the degradation is position-invariant,if G,F and H are respectively the Fourier Transforms of g,f and h,we obtain the following equation:

$$G = F.H$$

By dividing G by H and taking the inverse Fourier Transform,we could recover the ideal image f(x,y).H,the transfer function of the imaging system,is usually obtained by degrading a one point image and taking the Fourier Transform of the result.

In general,linear spatially invariant inversion does not deal adequately with the majority of systems which are non linear.Linear spatially varying or non linear techniques,in conjunction with information about the image function, such as the fact that the image function is non negative,are,often,used to achieve superresolution.Several linear spatially-variant techniques are based on generalised matrix inverse,such as projection iterative method and singular value decomposition.The non linear techniques include maximum-entropy and Bayes estimation(Pratt(1978)).

One of the earliest deblurring technique,in which the high frequencies in an image were strengthened relatively to the low frequencies,was developed by Koraszany and Joseph(1955).It used a derivative operator and assumed that the degradation process satisfies the following equation:

$$\partial f/\partial t = k(\partial^2 f/\partial x^2 + \partial^2 f/\partial y^2) = kD^2 F$$

D^2F:Laplacian of f

k : constant

When the blurring process satisfies the above equation,the blurred image f1 would,to a first approximation,be related to the sharp image by the following equation:

$$f = f1 - KD^2 f1$$

K : constant

Therefore it is possible to attenuate the blurring process by substracting a multiple of the Laplacian of the blurred image from the blurred image itself. In the case of digital images the Laplacian is approximated as follows:

$$D^2 F = D_x^2 F + D_y^2 F = (\Delta_x f(x,y) - \Delta_x f(x-1,y)) + (\Delta_y f(x,y) - \Delta_y f(x,y-1))$$

$$D^2 F = (f(x+1,y) + f(x-1,y) - 2f(x,y)) + (f(x,y+1) + f(x,y-1) - 2f((x,y))$$

$$D^2 F = f(x+1,y) + f(x-1,y) + f(x,y+1) + f(x,y-1) - 4f(x,y)$$

$$A = \begin{vmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{vmatrix}$$

Thus if we convolve the blurred image with the template matrix A,we will eliminate some of the blurring in the image.

A particular non linear filtering techniques,which compensate for wide variations in illumination, is homomorphic filtering.This technique is based on the fact that the grey of f(x,y) at a particular point (x,y),is proportional to the

reflectivity,r(x,y), times the illumination,i(x,y),incident on that point.Because the illumination ,in general,changes only slowly from one point to the adjacent one,i(x,y) contains low spatial frequency,whereas r(x,y) contains higher frequency information,for many images. The effect of illumination variation could therefore be compensated for by taking the logarithm of f(x,y) as follow:

$$\log(f(x,y))=\log(i(x,y))+\log(r(x,y))$$

Then the linear filtering has to be used on log(f(x,y)).finally the exponentiation of the filtered image log(f(x,y)) is executed to complete the processing (Opeinheim(1968)).

Early digital image restoration techniques operated in the frequency domain. But,recently many of them were transfered to the spatial domain where derivation is replaced by differences and integration by summation.Spatially variant and non linear techniques perform much better than Fourier methods,but involve, generally,complex and lengthy computations.

3.5.2 Enhancement

Enhancement is the transformation of an image into a form more suitable for a specific application.Depending on the nature of the image under consideration,the enhancement process makes

selected components of an image stand out so as to facilitate the subsequent analysis.In general,enhancement is the process which improves image quality for human viewing,or facilitate automatic image analysis. The improvements may involve sharpening the edges,delineating boundaries or stretching grey level histograms.

Enhancement frequency domain techniques convolve the image function with a point spread function.They are based on the convolution theorem.If

$$g(x,y)=h(x,y)*f(x,y)$$

* :convolution symbol

f(x,y):original image

g(x,y):image formed by the convolution of f(x,y) and h(x,y)

h(x,y):position invariant operator

the convolution theorem states that

$$G(u,v)=H(u,v).F(u,v)$$

G,H and F are respectively the Fourier Transform of g,h and f,with H(u,v) refered to as the transfer function.

In a general image enhancement problem,the transfer function H(u,v) is chosen so that the output of the convolution function g(x,y)

$$g(x,y)= F^{-1}(H(u,v).F(u,v))$$

exibits the required improvements in f(x,y).Choosing H(u,v) to

highlight low frequencies in f(x,y) would blurr the image,whereas choosing H(u,v) to highlight high frequencies would provide the edges.

More and more spatial domain techniques,which are based on the direct manipulation of the pixel grey levels,in the image plane itself,are used.A particular enhancement technique will depend on the desired modification of the image,and , generally,involves a certain amount of trial and error.

Because of the lack of a precise definition of image enhancement,it is difficult to describe all the available techniques (Pratt(1979),Rosenfeld(1976)).We wil therefore concentrate on the description of just a few of them.

3.5.2.1 Image Enhancement by Histogram Modification

This enhancement technique is based on the modification of the grey level histogram,which provides a global description of the appearance of an image. The histogram is modified in a specific manner to achieve the desired image enhancement (Rosenfeld(1969)).For example the improvement of the contrast would involve spreading the histogram.

3.5.2.2 Image Smoothing

The enhancement techniques,involved in smoothing,dimunish spurious effects, present in a digital image,because of poor sampling or noise.The simplest spatial domain technique is the neighbourhood averaging technique.

When an image suffers from noise such as 'TV snow' or 'salt-and-pepper' noise ,where grey levels have been randomly altered,it is possible to eliminate the effect of this noise by simple local averaging.However,because this introduce some blurring,a non linear method,which replace the gray level of a point,when it differs from its neighbours by a given amount,by the average of these neighbours,could be used instead. A more refined method is to use the latter method with an edge operator.

3.5.2.3 Image Sharpening

The image sharpening techniques perform the opposite operation of smoothing, They highlight the edges in an image.To get the opposite effect of smoothing (sharpening),differentiation is used instead of the integration,which is used for smoothing.The frequency-domain

techniques involve using a high pass filter.

3.6 Image Segmentation

Segmentation is the first step in the automation of information extraction from visual data.It consists of partitioning the image into meaningful regions which are,in general,the images of the different objects in a scene.

Segmentation techniques could be categorised as point dependent techniques, which are based on the examination of the image pixel by pixel,and region dependent techniques which use regional propreties such as differences in grey levels and texture,for segmentation.

The basic segmentation techniques involve clustering,thresholding,and sometimes matching ,for point dependent techniques,and edge detection,boundary formation and region growing,for region dependent techniques. Detailed discussion of the many available image segmentation techniques can be found in the previous chapter (State of the Art).In this chapter we will mainly concentrate on two basic techniques involved in image segmentation:thresholding and edge detection.

3.6.1 Thresholding

Thresholding is a technique often used in segmentation for both monochromatic and coloured images.It is ususally preceeded by some preprocessing such as noise cleaning or smoothing.It involves very simple computations.

Thresholding transforms a grey level image into a binary image.If f(x,y) is a grey level image function,whose grey levels range from z(0) to z(n),t is the threshold which lies between z(0) and z(n),and g(x,y) is the binary image which is the result of thresholding f(x,y) at t,the threholding operation could represented by the following mathematical equation:

$$f(x,y) = \begin{cases} 1 & \text{if } f(x,y) > t \\ 0 & \text{if } f(x,y) < t \end{cases}$$

Thresholding could also be extended to map specified image grey level ranges into 1 and levels outside these ranges into 0.

One of the main difficulties of the utilisation of thresholding in segmentation is the selection of the threshold value.This selection could be achieved by trial and error ,or,more interestingly,automatically.The automatic selection of the threshold value makes extensive use of the frequency

distribution of the image grey levels (histogram).

In the special case of an object lying in a uniform background,the histogram,usually,shows a peak ,and the bottom of the two valleys on the two sides of the peak,could be used as threshold values to extract the object from the rest of the image.In the general case,involving natural scenes,thresholding could segment the image reliably,only in the case of colour images,or images with a high dimension feature space(grey level and texture) (Price(1976)).

3.6.2 Edge Detection

The edge detection techniques involve localising image intensity changes and characterising them by their magnitude,and in some case their direction. The change in intensity can be detected by using two-dimensional derivative operators.

Korasznay and Joseph(1955) were some of the first investigators to apply two-dimensional operators to edge detection.Their method is to consider a grey level image as a two-dimensional function,f(x,y),and using partial derivatives, f/ x and f/ y,which respectively measure the rate of intensity change in the horizontal and vertical directions, to obtain an operator that will be sensitive to changes in all directions.

One such operator is the magnitude of the grey level gradient:

$$\sqrt{(\partial f/\partial x)^{+2} + (\partial f/\partial y)^{+2}}$$

In the case of digital images,which consists of a discrete array of grey levels, the derivative is replaced by a difference equation:

$$\sqrt{(\Delta f)_x^{2} + (\Delta f)_y^{2}}$$

with

$$|\Delta f|_x = |f(x+1,y)-f(x,y)|$$

$$|\Delta f|_y = |f(x,y+1)-f(x,y)|$$

Although adequate edges are obtained with this simple operator,more sophisticated gradients have been proposed and applied with good results,in the last decade.A more detailed description of these techniques is available in standard books (Pratt(1979),Rosenfeld(1976) and journals(Davis(1975)).

3.7 Regional and Structural image descriptions

After segmentation,the final stage in image processing consists of characterising the different regions of the segmented image by sets of descriptors,and determining the different structural relations between those regions.This selection of descriptors,involving features which aid in the classification of regions with different attributes,and in the recognition of regions by comparing them with particular models,results in a big reduction of data(regions are represented by fewer bits of information than was the case in the segmented image). The selected regional descriptors must be resonably insensitive to variations such as change in size,and spatial translation and rotation.Standard classical pattern recognition techniques (Watanabe(1972),Tou and Gonzalez(1974),and Chen(1976)) are then used for the classification and recognition of the different regions.

Images can be normalised with respect to translation by using autocorrelation (Horwitz and Shelton(1961)) and Fourier transform.The normalisation with respect to rotation and magnification could be achieved by polar coordinate autocorrelation (Doyle(1962)).

3.7.1 Regional Descriptors

Among regional descriptors the topological ones are the simplest,and are usually not affected by translation ,stretching or rotation of the image. Topological descriptors could be holes or connected components(set of points where any two points can be jointed by a connected curve lying entirely within the subset).

Fourier descriptors could also be used to characterise a set of boundary points.They are very useful for region shape descrimination.A third kind of regional descriptors,used to characterise grey level regions,are known as moments.

Given an image function,f(x,y),the moment of order (i,j) is defined as follows:

$$m_{ij} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^i y^j f(x,y)dxdy$$

the moments (i,j) provide information about how the image gray levels are distributed about the origin.They are insensitive to variation in translation rotation.and scaling.

For digital images the integration is replaced by summations.For general images, the set of moment m(pq) uniquely determine,and are uniquely determined,by the image function,f(x,y).Therefore,they constitute a complete image

description.

3.7.2 Relational description

A relational description involves determining the various relationships among specified regions of an image.The different regions of a segmented image are arranged into a meaningful relational structure.The determination of the structure of an image is only viable for simple images (documents and images with simple geometrical objects)(Eden(1972) and Robert(1965)).One of the major method used to organise regions of an image is the grammatical approach which is based on grammatical concepts(Fu(1974)).

4 HUMAN VISUAL SYSTEM

4.1 Introduction

There has been an increase incorporation of knowledge about animal and human visual systems into machine vision research.Although it is not very clear how far machine vision will incorporate the processes of visual biological perception, there has been a general assumption that the design of machine vision systems would require understanding the behaviour of biological visual perception systems,and specially the human visual system.

The understanding of the human visual system has been approached from two different directions.The earliest of the two approaches was started by Herman Von Helmholtz who studied the physiological mechanisms of the eye and the central nervous system.This physiological approach examines the anatomical structure of the eye-brain system in terms of light sensors,various links and a network of processors,and the behaviour of the structure in terms of impulse communication.

The second approach is a psychological approach.This approach ignores the internal structure of the eye-brain system by regarding it as a black box. The main concern of this approach is the determination of the relationships between the possible combination of the input and output states.

Although machine vision has progressed enormously in the last

decade,it still remains inadequate for most applications.In the majority of cases,human visual perception remains vital for extracting information from visual data. But it is hoped that the understanding of the general principles in animal and human perception,will ultimately lead to the incorporation of some of these principles into machine vision systems. The human visual system has its roots in the visual systems of more primitive animals.Biological visual systems with high performance evolved before the advent of man.Because man's survival,in recent times,depended less critically on his visual ability,there has not been any progressive evolution of his visual system. The recent progress in optical aids can even be seen as stimulating a certain regressive evolution of the human visual system.This possible regression is largely compensated by the progressive evolution of the brain.

The performance of the human visual system is very good : humans appear to perform visual perception tasks effortessly and instantaneously.It would be of great interest to determine how the brain recognise complicated spatial and temporal visual patterns,and how it extract information from incomplete data.The answers to this questions will greatly help in the design of machines which perform as reliably as the human visual system.Recently there has been an impressive advance in electronics,computers and automation. This motivated and helped the research for developing machines which duplicate the intelligent facets of the brain,and specially one of its important abilities,refered to as visual perception.

There are three major sensory data input channels to the brain.These do not use the spinal cord.they carry directly visual,aural,and gustatory information into the lower portion of the brain.The central nervous system is the center of perception.All the functions of the nervous system are performed solely by one type of highly specialised cell: the neuron or nerve cell.Although neurons occur in many different varieties,they all conform to a basic generic form.A typical nerve cell has a limiting membrane,a nuclueus and a cytoplasm containing a variety of organelles,and,specifically,has two types of elements extending from the cell body: the dendrites and the axon.The dendrites and the axon are highly specialised to transmit electrical signals,and to secrete from synaptic terminals a transmitter substance which carry the message across the gap between one cell and the next.The cumulative effect of transmitter substances arriving at the dendrites and cell body produce a wave of electrical activity,the nerve action potential,when the substance exceeds a threshold value.

Basically,all nerve cells consist of:

1- A main central cell body,the perikaryon,which could be seen as performing some sort of analog computation in response to certain chemical or electrical inputs,and produces an electrical output which is functionally related to the particular input

2_ A set of input elements to the nerve cell,the synaptic junctions,which are attached to the cell body and to the

denderites(The dendrites are a group of extensively arborised branches growing out of the cell body,which gather data to be fed into the cell body.

3_ An output organ,the axon,which leaves the cell body as a single fiber and then,usually,arborise to transmit to many other nerve cells through the synaptic junctions.

The information transmitted along the axon,which is the nerve cell output organ, is coded as a continuous sequence of short constant amplitude pulses with variable repetition rate,at regular or irregular intervals (100mV of amplitude from a normal resting potential of -70mV,0.5mS of pulse width,and pulse repetition rates ranging from 0 to about 1000 per second). The information transmitted from one cell to another is always coded in digital form.But the nerve cells are analog and not digital in nature like digital computers where each pulse is important and timing and synchronisation of pulses is as important as their occurence.They just utilise pulses as a convenient way to transmit variable intensity signals through a leaky conductor,the axon,where an analog signal would be attenuated and corrupted.

Although all the processing involved in neurons is not completely understood, it appears that it involves coding intensity of stimulus into rate of firing. How this analog to digital conversion,in 10 to the power 10 neurons,enables us to think is not as yet well understood.Many intensity stimuli from specific sensory

receptors are coded into pulse repetition rate,for transmission to the central nervous system.The coding is,generally,not linear.For example,in the case of the eye,the coding is a logarithmic function of the light received by the sensors of the retina.

In the case of sensory data,the initial coded signal from the periphery,is transmitted to the cortex of the brain by a succession of transmission neurons through a variable number of relay of grouping of nerve cells refered to as ganglia. This transmission of the data from the primary receptors to the cortex of the brain ,involves a certain reduction of this data as close to the periphery as possible. In the case of the eye,the primary visual sensors in the retina code directly the two-dimensional visual patterns produced by the optical system of the eye,into impulse repetition patterns.This coding is performed by approximately 105 million retina sensor elements,about 100 million rods and around 5 million cones,functioning independently.

This output represents an enormous amount of data. Because of limited channel capacity in the optic nerve and the reception areas of the central nervous system,not all the data gathered at the retina is transmitted to the central nervous system.The available channels for transmitting visual data,in man,number around one million for each eye and consist of the output axons of ganglion cells.These axons,which are insulated from each other by myelin sheaths that surround each axon,are bundled together and travel out from the back of the eye to the brain as the optic nerve.There are

many more light sensors,rods and cones,than channels to the central nervous system. This means that convergence has therefore occured at some level.This convergence,which involve the reduction of data,occurs directly in the retina,with,on average, each ganglion cell deriving its original input from about 100 light sensor cells.

The degree of data reduction varies spatially across the retina.There are two fairly distinct regions in the human eye .The first is the fovea which collects complete high resolution data for about 2 degrees of sharp vision.The second region consists of the portion of the retina peripheral to the fovea,and accounts for all the rest of the visual field.In the foveal region,there is very little convergence.Each cone,or group of cones,excites one bipolar cell,which in turn excites one ganglion cell.By contrast,in the retina,away from the fovea,the outputs of more than 100 primary sensors are processed directly in the retina and relayed into a single channel.

The information gathered by the peripheral region of the retina is mainly concerned with motion detection.The colour and form information collected by this region are very poor. For the collection of more detailed colour and form information,the sharp fovial vision has to be used. The poor quality of the peripheral information is illustrated by the fact that ,if we concentrate on a letter in the center of a ten letter word,then we will hardly be able to resolve the letters at the begining and at the end.

The region of high acuity,represented by the fovea,account only for two degrees of the visual field whose total range is around 180 degrees horizontally,60 degrees vertically.The fovea,which uses a dense resolution,is used to resolve uncertainties which are caused by the low resolution of the rest of the retina.It uses about ten per cent of the total visual channel capacity,although it represents only 100,000 sensors,for 105 millions receptors for the total retina.Nearly every visual sensor element of the 100,000 fovial receptors is allocated one channel.When required,the fovea is used to collect high resolution data for particular portions of a scene.

The visual data gathered by the eye and preprocessed by the ganglia,is finally transmitted to a two-dimensional surface of the brain termed the cortex,which is the highest portion of the brain.During transmission from the retina to the central nervous system,the data is preprocessed by the retina or ganglia.This preprocessing results in a substantial reduction of the data,and unfortunately it also results in a certain loss of information.From the thalamus,the data is relayed to the cortex still in a conformal form.

The part of the cortex responsible for the processing of visual data occupies about 26 square centimeters,in the back of the brain which is called occiput. It is the processing in the cortex,which,ultimately,permit us to perceive the visual world,and output data concerning muscular movement. The cortex consists of six layers or laminae of nerve cells,and is essentially a

two-dimensional sheet like structure with an area of about 2200 square centimeters and 2.5 millimeters thick.Although the demarcation between the different layers is not very clear,fairly sharp variations,in the number and type of neurons,can be seen along the thickness of the cortex between the six layers.The basic form of the layer structure changes very little,but individual layers may vary in thickness.The overall structure is fundamentally the same for all mammals.The six layers consist of a molecular layer (plexiform layer),an outer granular layer,a pyramidal cell layer ,an inner grannular layer,a ganglionic layer and a fusiform cell layer.The basic pattern of connectivity between the cells is fundamentally invariant in the six layers.It has been established from anatomical evidence that the cortex is characterised by a vertical organisation.The data in the cortex is transmitted perpendicularly to the plane of the sheet of the cortex.The transmission of data across the sheet of cortex is essentially negligeable.Input data is received through axons perpendicular to the surface of the cortex,and the output corresponding to the particular input departs the cortex effectively at the same spot where the input went in.

In the case of the region of the cortex,which is concerned with the visual data processing,the fourth layer is involved in the reception of the data relayed from the thalamus.From the fourth layer,information,in digital form (variable pulse repetition rate) is transmitted to all other layers in the cortex either directly or after further relaying from other layers. After the processing of

the input data in the different layers,some data is transmitted to other regions of the cortical sheet or to the thalamus from the fourth, fifth and sixth layers.

4.2 General Structure of the Eye

The perfection of the eye as an optical instrument contributes greatly to the good performance of the visual system.The three coats,which enclose a transparent refractive media,form the globe of the eye.The first outermost layer of the globe is made up of the sclera and cornea which is transparent.The second layer of the globe is mainly vascular,and consists of the chorio,ciliary body and the iris.The third innermost layer is the retina.It contains the essential sensor elements responsible for vision,and is continued over the ciliary body as the ciliary epithelium.The sensor elements are rods and cones whose names are derived from the shape of their outer segments.The cones gather information about colour and function in daylight condition.On the other hand,the rods gather data illustrating only the shades of grey,and function under low illumination.

The dioptric element of the eye is made up of transparent components.The first transparent structure of the eye is the cornea,which is composed of connective tissues and is covered on both sides by epithelial cells.It is transparent and curved.It is

the clear portion of the globe surrounded on all sides by the white of the eye called sclera.The second structure which refract further the light is the crystalline lens which is supported by the zonule that is itself attached to the ciliary body.The ciliary body muscle fibre can contract to increase the refractive power of the lens.

The lens plays a vital role in accomodation which it achieves not as in a camera,by changing the lens position,but by changing its shape. For example,for near vision,it reduces the radius of its curvature. The light,which passes to the lens and on to the retina,has first to pass through a variable aperture formed by the iris (the pupil).

The spaces within the eye consist of the anterior chamber and the posterior chamber ,which consist respectively of the small space between the cornea and the iris, and between the iris and the lens,and a large space behind the lens and the ciliary body. The posterior and anterior chambers are filled by a clear fluid,aqueous humour. But the large space is filled by a jelly,the vitreous body. The iris is just behind the cornea and behaves as a diaphragm which controls the amount of light that reaches the retina.

The retina,which is often described as 'an outgrowth of the brain',consists of a light sensitive layer of specialised nervous cells on the back of the eye. The cells of the retina are arranged in layers,including a layer of cells that are sensitive to light.In the case of vertebrates,the retina is inside out with the light

having to pass through several layers of cells before reaching the light sensitive cells.The light sensitive cells are the first stage in the visual system,and form synapses upon the next layer ,which is composed of bipolar cells.In turn the bipolar cells synapse onto ganglion cells.The axon of the ganglion cells constitute the optic nerve.

On its way to the cortex,the optic nerve is lead through a canal in the bony orbit ,the optic foramen.The fibres in the optic nerve finally cross over in the optic chiasma in such a way that most of the fibres from the right eye go to the left hemisphere of the brain and the fibres from the left eye go to the right hemisphere of the brain. A very specialised portion of the retina is the fovea which is used for colour and high resolution data.

The nutrition of the cells,which compose the eye,is assured by the capillaries of the vascular coat.The choroid,which is essentially a layer of vascular tissues next to the retina,the ciliary body,and the iris are fed by a system of arteries derived from the ophtalmic.On the other hand,a separate vascular system derived from thr central artery of the retina,which enter the globe with the optic nerve,feeds the inner nervous elements of the retina.

To gather appropriate information about the environment and use the fovea ,the eye has to be slightly adjusted.The movement of the eye is performed by the contraction of six muscles. When searching for an object or following a moving object,both eyes move

continuously and in various ways. The movement is smooth,when following,but it consists of a series of small rapid jerks,when searching.Another important movement of the eye consists of continuous small high frequency tremors,which are thought to be necessary to ensure that cells in the retina are stimulated by a stationary scene.

4.3 Seeing

The process of seeing is not achieved by the eye but is mainly due to the processing of the information sent by the eyes to the central nervous system. In vision,the pattern of neural activity produced by the visual input data sent by the eye represent the object,and as far as the brain is concerned,is the object.Although there is little understanding of how the pattern of neural activity gives rise to perception,there is some evidence which suggests that the image is seen in terms of patterns of lines.The visual system has generally little difficulty to extract from sometimes few lines,all the objects depicted by those lines.

The sensation of seeing involves the utilisation of many sources of information beyond the visual information.This ,in genral,involves some a priori knowledge of the objects derived from previous experience,which is not limited to vision, and which may involve other senses such as hearing,smell,touch and taste. The

visual system is not perfect,and is sometimes subject to errors.A same visual data might lead,in rare cases to different interpretation.Perception is ,therefore,not determined solely by the visual data gathered by the eyes,but it involves a dynamic searching for the best interpretation of the available data.

The a priori knowledge,which is used in visual perception and perception in general ,is stored in two ways.The two kinds of storage,which are involved in memory, are respectively short term memory and long term memory.Inactive information is stored in the secondary or long term memory,whilst information in the temporarily activated state is stored in primary or short term memory.In primary memory the access of information seems instantaneous although it takes a short finite time.This memory has severe storage limitations.Another problem with primary memory is that without constant attention it soon fades and becomes to weak to sustain recall. The strength of the memory trace,finally left in the secondary memory ,depends largely on how long information is maintained in primary memory activation. The human memeory system is highly organised:similar units of information are linked and stored together.This organisation of the memory system facilitate enormously the reading of specific information from this memory.

Two types of information seem to be stored in the human secondary memory system.The first concerns information about specific events we have experienced or being told about. The second concerns generalised information about the structure of classes of

objects and events.The first kind of information is called episodic memory. The second kind of information,where concepts have been derived from our experience of specific objects and events through a process of abstraction or generalisation,is refered to as semantic information.The concept representations stored in the semantic memory determine what is to be stored,in future,in the episodic memory: it is the interpretation of the observed event and not the event itself,which is stored in the episodic memory.

The brain uses the two kinds of memory to process the data gathered by the eyes and to interpret the data to give the sensation of seeing.The process of visual data interpretation is recursive,iterative and converges,in the majority of cases,to a unique interpretation.But,in a limited number of cases the processing of the visual does not yield a unique interpretation.In some of these cases two interpretations are possible (vase and two faces illusion).When this occurs the visual system considers first one then the other interpretation without never reaching a conclusion.In this case the interpretation oscillates between two possibilities which are both equally acceptable and does not converge. Some other ambiguities in the interpretation of visual data can arise. Such ambiguities can lead to wrong conclusion such as when we suffer from hallucinations or illusions.

4.4 The Falliability of Perception

Perception can,sometimes,lead to wrong interpretation of the sensed data.For example,in drug induced states,or in mental diseases,an entirely fictitious world can be created and mistaken for reality.In these very rare cases,the interpretation of the sensed data departs altogether from reality,but in other cases surrounding objects are just perceived in a distorted way.These incorrect perceptions,which are sometimes refered to as illusions,are unaffected by the knowledge of the observer that the perception is an illusion.

Some of these illusions are illustrated by the following simple figures.The distortions caused by certain illusion can,in some cases,be quite large.For example ,as illustrated by figure 4.1,two segment of the same length,incorporated in different configurations of lines,appear twenty per cent too long or too short. In another example,illustrated by figure 4.2,straight lines are seen as curved lines,so that it is difficult to believe them to be really straight.The illusions are,in general,caused by certain perceptual mechanisms which are appropriate to normal viewing but produce an incorrect interpretation in artificial viewing situation.An example of such a perceptual mechanism is size constancy wich tends to compensate for changes in the retinal image with viewing distance.

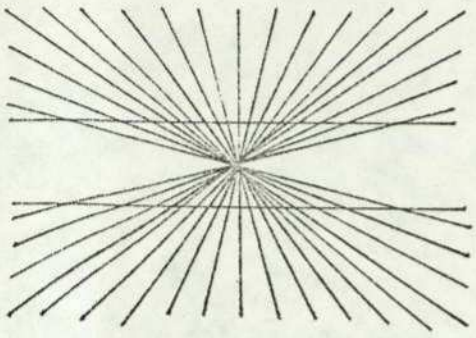


figure 4.2

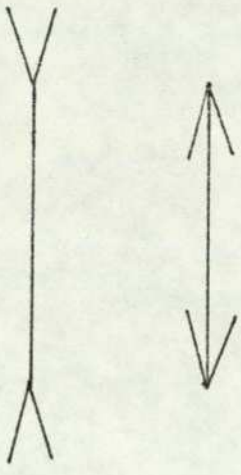


figure 4.1

Although illusions are inconvenient,they are becoming very useful,in understanding how the eye-brain system works.The existence of illusions shows that the visual process does not consist only of a data collection process using the eyes,but involves complex processings which use the memory to arrive at a semantically meaningful interpretation.

4.5 Visual System Interpretation of Brightness and Colour

The word brightness has been used in many different senses,but it usually refers to the subjective experience of observers when the eye is subjected to light.The brightness is related to the intensity

(energy) of light entering the eyes.Brightness,which is the simplest of the visual sensations,plays a very important role in visual perception.This is validated by the fact that black-and-white images are quite sufficient for perception.The relation between brightness and the light intensity,which generate the sensation,is a complex one, and it depends upon various conditions that determine the contrast of objects in a particular scene.The relation depends on the spatial and temporal distribution of the light which has reached the retina.Roughly,it depends on the intensity of light falling over a given region of the retina at a given time,on the intensity of light falling on the neighbouring regions of the retina,and finally on the light which fell on the retina in the recent past.

4.5.1 Contrast

An important factor in perception of brightness is the intensity of the light that the surrounding areas of the region of interest has been subjected to. A particular region will generally look brighter if its neighbouring areas are darker.The available evidence suggest that contrast enhancement seems to be related to the importance of edges in visual perception (in general a sketch provides sufficient information for visual perception). The evidence also suggests that it is primarily the information about the edges of the image which are transmitted to

the brain,and information about intensity is transmitted only when needed.This process of edge detection reduces the amount of data transmitted without great loss of information.Brightness is also a function of colour (Bartelson and Breneman(1967),Beck(1972)).When the eye becomes adapted to the dark,it exchange its acuity for an increase in sensitivity.

4.5.2 Sensitivity to Light

The rate of firing of the receptors in the retina depends directly on the intensity of light which reaches it.As the intensity of light increases the rate of firing of the sensors also increases.The relation between the rate of firing of the retina sensors and the intensity of light is approximately logarithmic in nature.The retina and the optic nerve are not entirely free of activity,when the eyes are completely in the dark. The permanent background activity of the visual system limts its sensitivity. To attenuate the effect of this permanent background noise the eye resorts to increases in the duration over which visual signals are integrated.It also adopts other measures such as demanding several signal confirmation from seperate receptors.

Sensitivity has been quantified by Weber's law which states that the smallest difference in intensity,dI,which can be

detected,is directly proportional to the background intensity I.The law could be represented by the following equation:

$$dI/I=\text{constant}$$

This law is an approximation,which is fairly accurate for a wide range of background intensity.But it fails to represent the behaviour of sensitivity for low intensity.It is thought that the inadequacy of of Weber's law for low intensity is mainly due to the permanent background noise which has been mentioned earlier. The absolute limit of intensity detection by the eye is determined by the smallest visual signal which can be seperated from the random background noise when no light is entering the eye.

4.5.3 Colour Vision

We have explained,above,The process of vison where information about the intensity of light is transformed into rate of neural impulses that are sent to the brain. This is,however not the only information transmitted to the central nervous system.Light entering the eye,being not only characterised by its intensity but also by its wavelength,is also characterised by variations of its wavelength. The variation in wavelength are perceived by our visual system as variations in colour. A normal

eye is usually sensitive to wavelengths varying from around 400 nanometers(violet or blue) to about 700 nanometers(red).

It is the cone system of the retina that is responsible for the perception of colour. When subjected to light,the retina uses its cones to generate a neural code from the wavelength of the absorbed light.There are at least three different types of cones.The first type specialises in the absorbtion of light with short wavelength (blue cones).The second type specialises in the absorbtion of light with medium wavelength (green cones).Finally,the third type specialise in the absorbtion of light with relatively long wavelength (red cones). Every pattern of wavelenght and intensity input will generate different configuration response from the cones.

Before being sent to the brain ,the wavelength information is pulse coded. The output of the cone system consists of three outputs.The first corresponds to intensity information. The second corresponds to the ratio of green to red light absorbed.Finally,the third corresponds to the ratio of blue to red light absorbed.It is from this coded information about intensity and wavelength that the central nervous system determine the colour sensation.The perception of colour involves complex processing where many variables,other than simply the visual data gathered by the eyes,are used

4.6 Binocular vision

The two eyes in the visual system do not function separately,but operate in co-ordination to achieve perception. The visual field of both eyes is usually filled with double images, which,because they are mostly peripheral and always out of focus,are not easy to observe.The utilisation of double images is the characterisation of what is usually refered to as binocular vision.Compared to binocular vision,monocular vision,which utilise single images,seems to lack the special quality of depth. To obtain this special quality of depth,double images,which carry more information about depth than a single image,have to be used.

The disparity of objects in the scene,in the central position of the binocular field ,is a direct function of their distance along the line of sight from the point on which the two eyes converge.The central portion of the binocular field represent the same spatial field viewed from two different positions.The disparity in this central binocular field,is the difference of the two retinal images, which are a projections of the same scene viewed from different positions. It is only the central binocular field ,where the two monucalar fields overlap,that is used in binocular vision.The neurological process originating in each eye seperately are fused to generate three-dimensional perception of objects.

Although binocular vision is very helpful for the appreciation

of depth and for three-dimensional reconstruction of scenes,it is not necessary as is supported by the fact that persons having vision in only one eye see the visual world in depth nearly as well as the persons using binocular vision. Binocular vision is just one of several mechanisms in visual perception,which are used to determine depth.There exist other mechanisms such as cues in monocular vision for the evaluation of depth.

5 COMPUTATIONAL FACILTIES

5.1 Introduction

The computational facilities,and in particular the input/ouput facilities play a role,although a minor one,in the determination of the overall strategy for solving the problem of concern for this research.At the start of the research ,a system based on a PDP11 was used to digitise the images,and the processing was carried out,using a very powerful system based at the University of London Computer Centre (ULCC).Although this system was very powerful,it could only be used in a batch mode.Because this mode of operation was slow,the work was later transfered to a new powerful interactive system based on a PRIME,and recently acquired by the Instrument Systems Centre.

The description of the various computational facilities,used during this research,will give a general idea of the tools which were available for solving our particular problem of vehicle guidance by automated scene analysis.In particular,the description of the image input and output devices will highlighten the specific problems of visual data input and output.

The computational facilities in general and the input-output capabilities in particular are of major inportance in image processing.The utilisation of the appropriate system could speed up the research considerably.The input-output capabilities of a system depends on the type of image processing system.In general,there are three types of image processing system.

Firstly,the system,which is most often used by researchers in the image processing field,is a general data processing system with sometimes a special image processing library,or a limited software package developed by a particular user for his particular problem.

Secondly,some general-purpose systems for image processing,which are primarily used in image enhancement and for the resolution of more specific image processing problems,have been developed(Wilson,Tenber,Thomas and Watkins(1977),Booth and Schroeder(1977), Gambino and Schrock(1977)).Usually,these systems consist of a dedicated computer and one or a number of graphic terminals for visual examination and other form of man-machine interaction.The various algorithms for data acquisition,analysis and display are,in general,provided by an integrated system composed of many modules,which can be run under the supervision of an executive program.The main function of these systems is to help the system user to resolve his particular problem more qickly by providing him with easily accessible algorithms for general operations in image processing(thresholding,filtering,edge detection and various operators). The primarily design requirements for general-purpose systems are flexibility and versatility.

The third type of image processing systems is used for special applications.The main function of such systems is to provide a computing facility for the resolution of a particular problem (surface inspection,industrial component recognition and handling). Both special systems and general-purpose systems are usually

provided with input data channels adapted for visual data acquisition.The form of this input data channel is a major component for the system specification.The first type of systems used for image processing,does not usually have a special input for visual data it could only process data which is already in digital form.

Many image processing systems are normally used interactively because this facilitates the debugging of programs,and reduces the time needed for program development. The majority of these systems do not operate in real-time. A real video interface to a simple black-and-white TV camera generates about 10 millions pixels of data every second,which is equivalent to a rate of 60 to 80 million bits per second.Most existing systems cannot accept such high data rate and do not have enough primary memory for storing the digitised images. It has been estimated that for achieving real-time operation a given system has to be capable of executing 1 to 100 billion operations per second,depending on image size and computational complexity of the operation to be performed.

Computational problem that arise in vision have many special properties that are a disadvantage for general purpose processors.Different computer architectures might alleviate the apparently insurmountable computational complexity involved in the processing of visual data.A very promising architecture is based on cellular logic (Duff(1980).The adoption of this architecture might lead to the design of computers which process data in parallel mode(instead of the present sequential mode), to process images in

real-time.Examples of theses systems,which have been designed,are SATARAN(Romrbacher and Potter(1972)),ILLIAC IV(Denemberb(1976)),and CLIP 4(Duff(1979)).

5.2 Computational Facilities Employed In The Investigation

5.2.1 Introduction

In the Instrument Systems Centre of The City University,during the last decade,a group,working on signal processing in general and particularly on acoustic and visual signals, has developed a small signal processing system based on a PDP11 and a rig for digitising acoustic signals and images.

The group is specially interested in recognition and classification of defects on surfaces(Hill(1977)),tomography(Cavouras(1981)),noise analysis and classification(Moukas(1976 and 1981)), silhouette recognition(Koulopoulos(1981)), inspection of printed circuit boards(West(1980)),and syntactic approach to classification of silhouettes(Babbra),and street scene analysis(Guentri (1979 and 1980)).

A computational system for signal analysis has to have specific features such as special hardware to input different kinds of signals.In the case of acoustic signals,a special interface to digitise analog signals from a tape recorder has been designed(Moukas(1981)),and a visual display was used to facilitate the display of the data.

The research in image processing is very varied,and involves the utilisation of different types of images such as images of steel surfaces(Hill(1977)),images of silhouette components(Koulopoulos(1981)),printed circuit board images (West(1980)),X-ray images of the heart(Cavouras(1981)),and complex street scene images(fig 5.1,fig 5.2,and fig 5.3). As in the case of acoustic signals,there was a need for a flexible input channel for visual data into the mini-computer memory or mass storage(magnetic tape or floppy disc).The input facility consisted of a rig which will be described latter in more details.

The system was originally intended for general applications and hence did not have the appropriate software for image processing. In the past most of the image processing was done at the Univesity of London Computer Center(ULCC) in a batch mode,with the output images displayed on microfilm.The PDP11 System of the Instrument systems Centre of the City University,was mainly used for digitising images and storing them onto 7-track magnetic tapes.Recently an alternative system to the

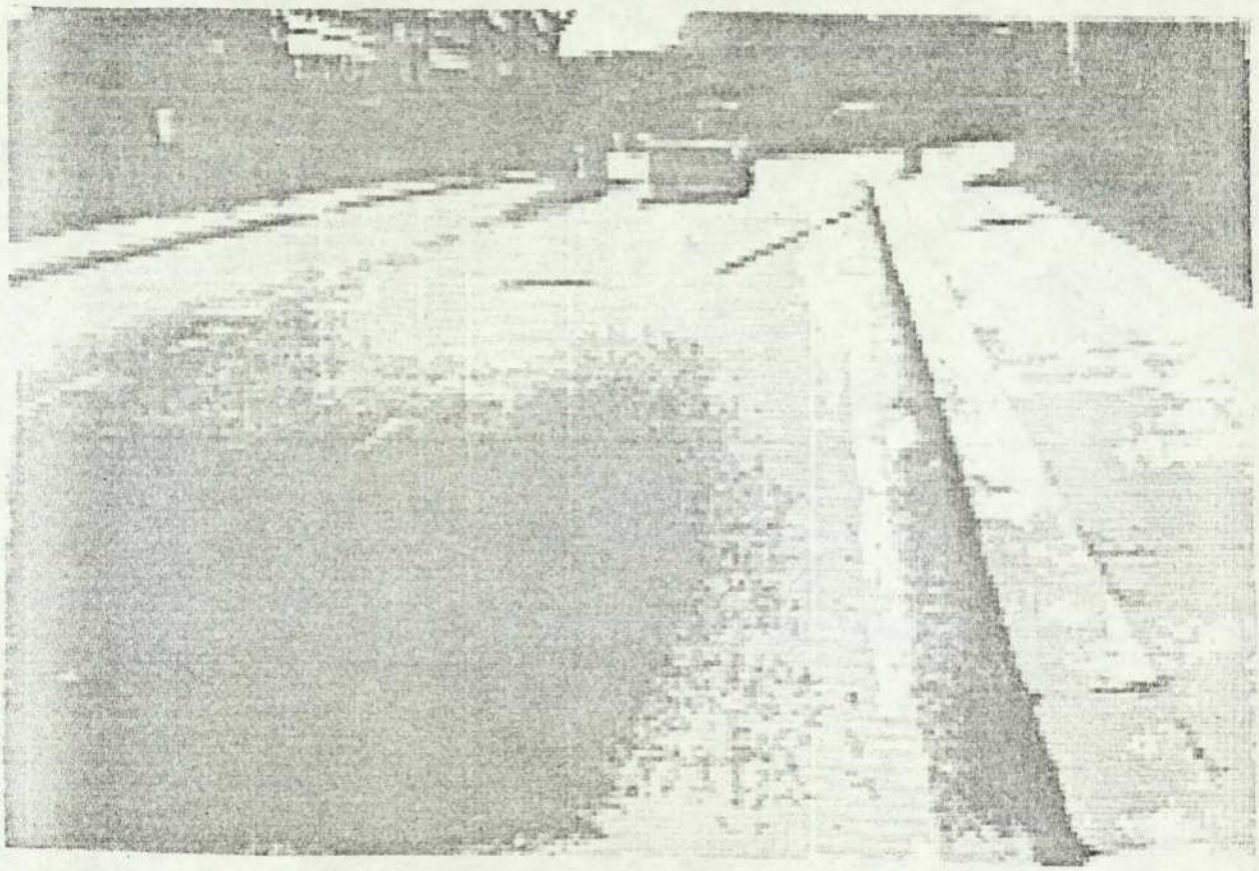


Figure 5.1. Street street image displayed on the GOC.

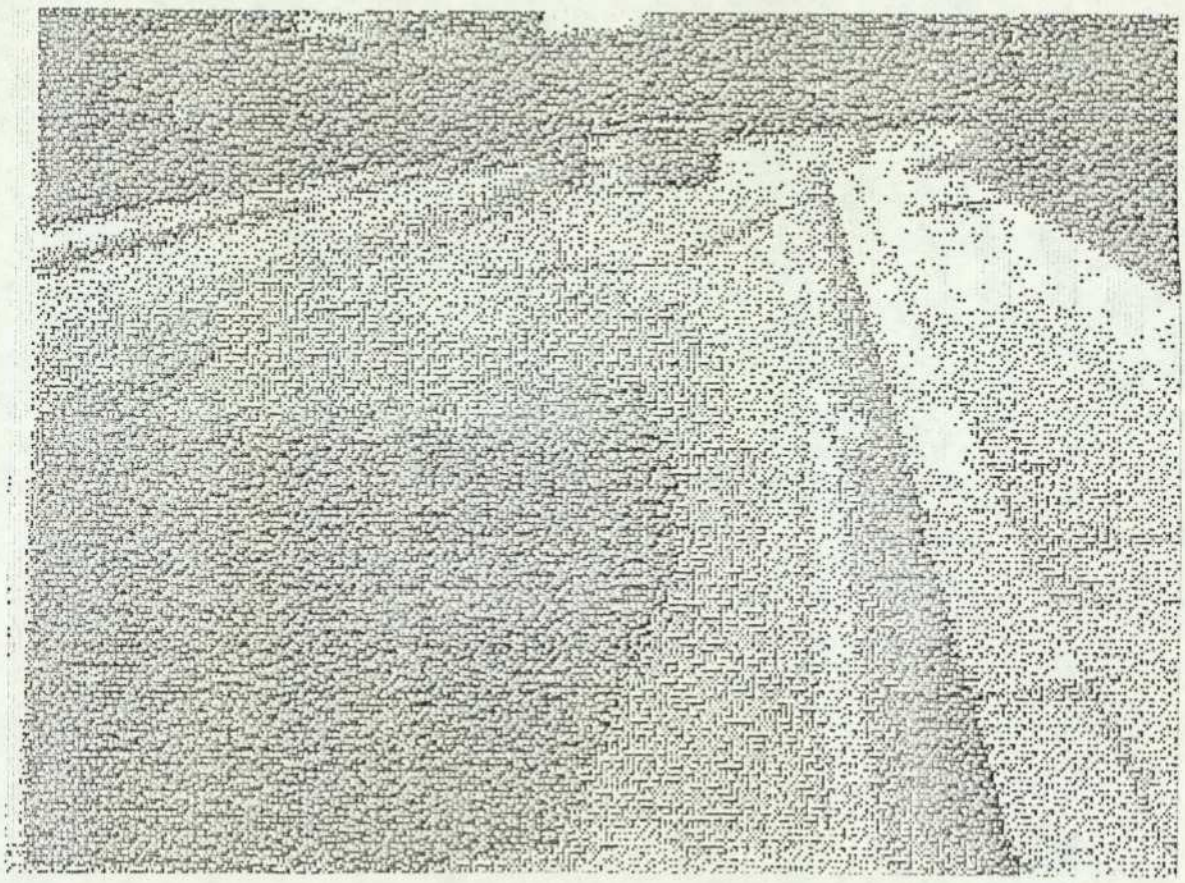


Figure 5.2. Street scene image displayed with a line printer.

Figure 5.3. Street scene image with grey level displayed in the third dimension.

ULCC system was acquired by the Instrument Systems Centre. This is an interactive system based on a PRIME.Therefore it was possible to process images interactively,with the output of the processed image displayed on a colour graphic terminal.the acquisition of this system facilitate considerably the research by reducing the overall processing time by approximately a factor of 5.

Because the computing facilities were intended for general data processing,the software for image processing was designed by individual researcher for their specific problem. Because it has been designed by different users for specific tasks,the present software consists of a collection of varied packages.The author personnally designed a package for general edge detection,thresholding,and line detection using a Hough transform.Recently,there has been an attempt to incorporate the different packages into an image processing library.With the recent linkage of the PRIME system to the PDP11 system,it would be possible to build up a complete versatile image processing system.This would probably be achieved in the near future(Ellis).

5.2.2 ULCC System

The computing system at ULCC was the first system used in this research.It was used in batch mode to process visual data stored into 9-track magnetic tapes held in a permanent library at ULCC.It is based on a CDC 6600,a CDC 6400,a CDC 7600, and a CDC Cyber 72.The 6600,6400 and Cyber 72,which are collectively refered to as the 6000 computer,each have a central processor to carry out mainly arithmetic work,and ten peripheral processors to handle the input and output devices such as card readers,line printers and magnetic tape drives.The 7600 has no input/output devices directly attached to it,but it is in communication with other computers,which act as stations for the purpose of handling input/output and magnetic tape transfers.It has a central processor and peripheral processors.The overall configuration of the system at ULCC is diagrammatically shown in figure 5.4.

One of the most essential parts of the computer linkage at ULCC is the arrangement of the magnetic disc hardware.Most disc packs are removable,but the 7600 has fixed discs. The 6000 permanent file base is accessible to all 6000 jobs,and can also be accessed by jobs on the 7600 via the permanent file station on the Cyber 72.

The 7600 is the most powerful machine at the centre in terms of speed of operation.It is approximately four times faster than

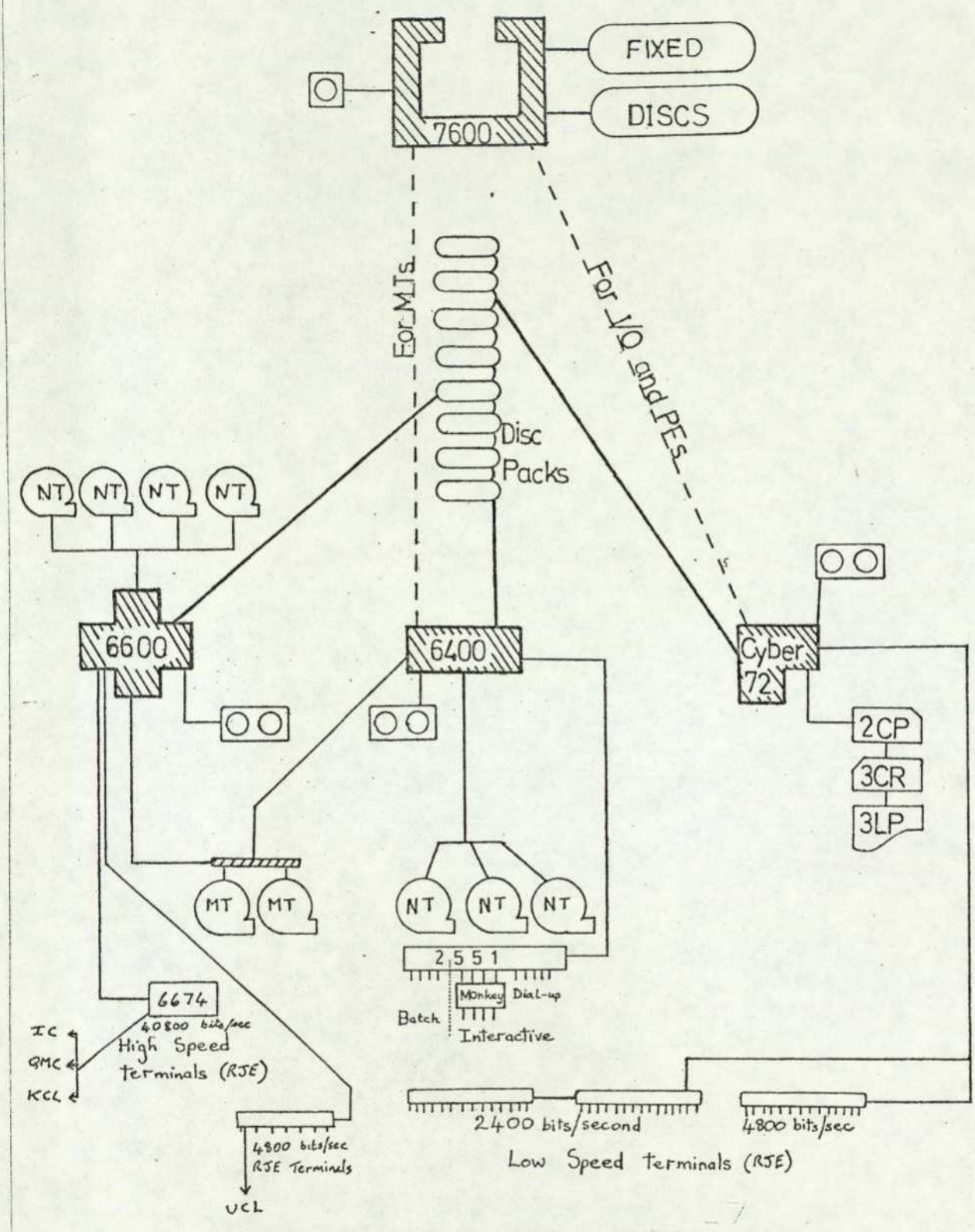


Figure 5.4. Diagramatial Representation of the ULCC system.

the 6600,12 times faster than the 6400 and 20 times faster than the Cyber 72.It has two kinds of core memory,small core memory(SCM) and large core memory(LCM),with the SCM used to store the program code which the computer is to execute.It runs using SCOPE 2 operating system which can manipulate disc files and magnetic tape files.

The CDC 6600 is the largest machine in terms of executable core available for effective use,and is the second fastest computer in processing speed at ULCC. As the CDC 6600,the CDC 6400 runs under the NOS/BE operating system,and handles magnetic tapes,disc files and all the interactive terminals.The Cyber 72 is not used for the execution of user jobs.It has connected to it the line printers,card readers,card punches,a magnetic tape drive and a number of telephone lines.

An important output device which is of special inportance in image processing, is the microfilm equipment.The production of graphical output on microfilm is done on a CalComp 1670 with its associated tape drive,tape controller, monitor,and a camera.There is also a dark room for developing the film and a machine for making readable copies from the developed film.Graphical output can be produced on 35 mm unsprocketted film or 16 mm sprocketted film.

5.2.3 PRIME/PDP11 System

As it was acquired just recently, the PRIME system was not available at the start of the research. It was used interactively to process images stored into a disc storage module. As it stands,at the present,the PRIME/PDP11 system is illustrated by figure 5.5. It is based on two computers,a PDP11/10 and a Prime 550 multi-user mini-computer, a multitude of terminals,a rig for digitising acoustic signals and images,and various mass storage devices.

In the PDP11 several devices are interfaced to the unibus:an 8Kw core and 16 Kw of semiconductor memory,a VT11 vector refresh display with dedicated display processor,a Tektronix 4006 graphics VDU(storage tube),an RX01 single sided,single density twin floppy disc unit,a TU60 dual cassette tape unit,a Racal T 7000 magnetic tape deck(7-track,200/556/800 BPI),a DRI6 123 matrix printer(106 CPS bidirectional),and an ASR 33 teletype.The PDP11/10 is mainly used for data acquisition from the rig and for very limited image processing.A DR11-B DMA port and a DR11-C serial interface are used to interface the rig and the acoustic signal digitiser to the PDP11.

The PRIME 550 multi-user minicomputer is connected to the PDP11.It is part of the Science Research Coucil's interactive computing facility and is mainly used for SRC-supported research

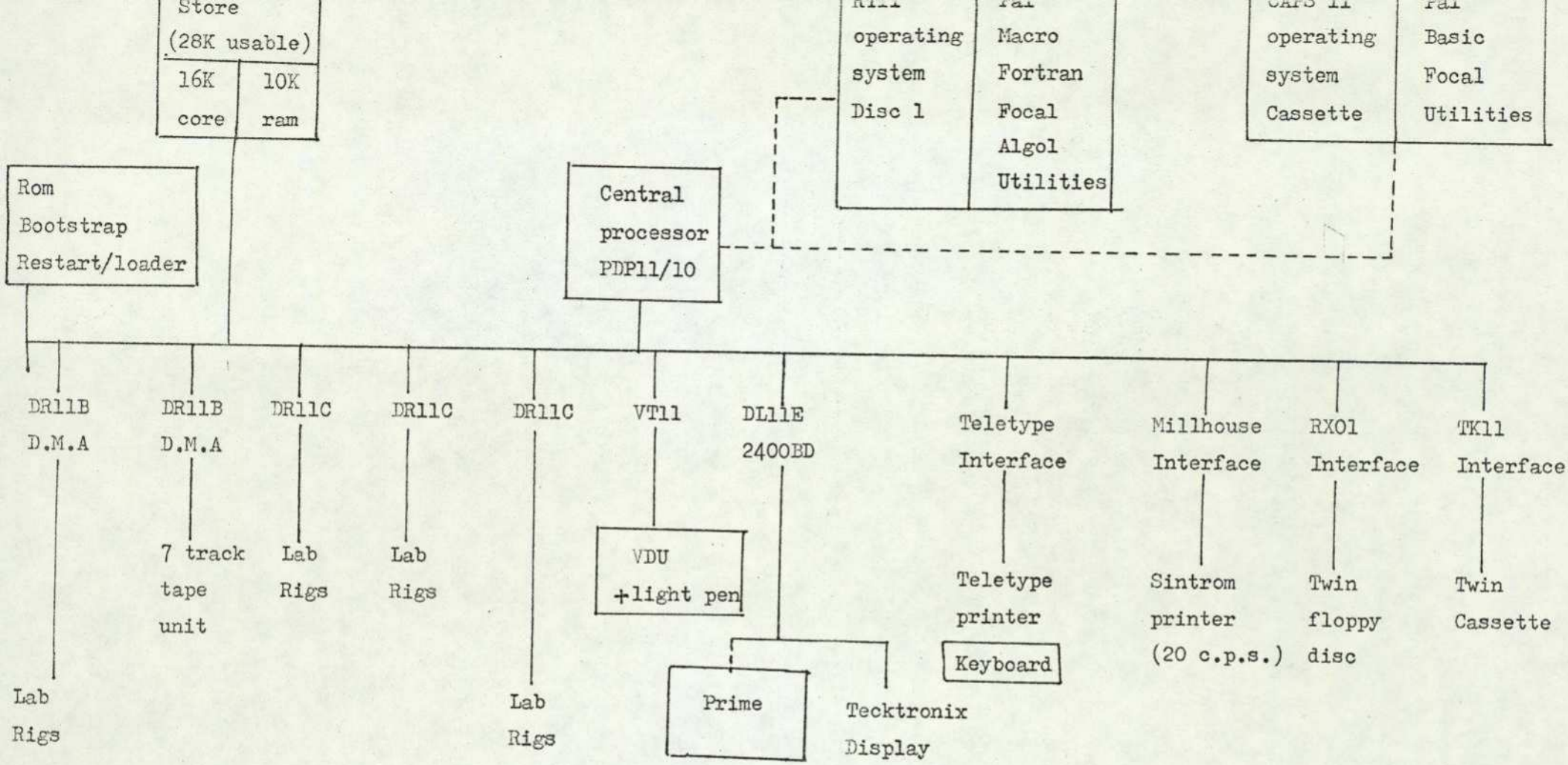


Figure 5.5. Prime / PDP11 system.

using interactive computer graphics. It has a 512 Kb MOS memory,an 80 Mb removable disc storage module and a magnetic tape deck(9-track,800/1600 BPI).It is interfaced to a line printer(PRINTONIX P300,300 LPM printer/plotter), a fast printer with keyboard(Tally 1612,160 CPS,bidirectional matrix printer),and a high quality printer with keyboard(Diablo 1640,45 CPS Daisy wheel printer). For graphics output,it uses a drum plotter(Benson 1302,3 colours,93cm maximum width),and 3 graphic VDUs(A Sigma 5600,black-and-white,raster scan refresh; a Sigma 5660 colour/grey scale,raster-scan refresh; and a Tektronix 4010,black-and-white storage screen).A Tektronix 4632,for refresh terminals,with enhanced grey scale,and a Tektronix 4631,for storage screen,hard copy units are used to obtain hard copy images displayed in various graphic VDU screens.

After this brief description of the computational systems,used during the research, we are going to discuss the input/output devices for visual data,emphasizing the devices which were built at the department or whose software was developed during the research.

5.3 Image Input Devices

The research involves feeding 'raw' visual data to the computer,processing the data,and finally displaying the output data into a form suitable for human viewing.In the following sections we will briefly describe the concepts involved in the imput of this visual data,and the various available image input devices.Finally,we will describe the actual device (rig) used to input the images at various stages during the research.

5.3.1 Introduction

In general images are based on the reflectance or emissivity of the surfaces contained in a particular scene. The energy of the electromagnetic wave is modified by the environment and the sensor characteristics,and is turned by the sensor into a permanent record such as a photograph or a video image.For computer use,the data is digitised and stored in digital form.The image is represented by a two dimensional matrix of numbers,each of which defines the intensity of the image at a point.The number of points in the matrix,which is,on average,about half a million,defines the resolution of the digitised image. In the case of photographs,the visual data has first to be transformed into analog electrical signals before digitisation.In many

applications in image processing the data is not collected on real timee.But the data collection is in the form of photographic transparencies which are digitised later using an image digitiser.A variety of devices have been used for visual input to computer systems.The four main types of devices currently used,are Vidicons,solide state array cameras,random access devices and laser scanners.

5.3.1.1 Vidicons

One of the most usual devices for visual data input is the Vidicon camera.Because of their high volume of production for the entertainment industry, vidicons,which are devices for TV signal generation,are unexpensive.But,for industrial application,due to their limited tube life and fragility,vidicons require constant readjustment for drift and ageing.

5.3.1.2 Solid State Cameras

Recently,cameras containing arrays of photosensitive elements have been used to overcome the problems of vidicons.The photosensitive elements,in these solid state

cameras,are either charge-coupled devices (CCD) or charge injection devices (CID).At present,a major problem with these cameras is their limited resolution (256x256).However newer cameras,with a spatial resolution equal to the conventional video resolution (512x486),are expected to be widely available in the near future. Another problem with the solid state cameras is the non uniformity of the response between the elements of the array;but it has been greatly improved recently.

When the scene to be scanned is in continuous linear motion,as in conveyor belts,instead of a two-dimensional array,a linear array can be used. One important advantage in the utilisation of the linear array is the high resolution which can be obtained.Currently,available linear arrays have a resolution of up to 2084 elements and good operating characteristics (uniformity,scanning rate and light levels).For equipments using linear arrays,the camera scans a line across the conveyor,and the motion of the conveyer produces the orthogonal direction of the scan.

5.3.1.3 Random Access Cameras

A random access camera is an image dissector which can provide high resolution images.In this device,small portions of interest in the image can be selected and hence there is no need to store the entire image.With a random access device a variety of scans such as circular scan or radial scan,can be selected.One main advantage of an image dissector is that the spatial resolution can be an order of magnitude better than that of a vidicon.

5.3.1.4 Laser Scanners

For particular applications,such as inspection of web materials,such as paper, cloth and plastic,laser scanners are frequently used.These devices,which are capable of extremely fast operation,involve an arrangement of rotating mirrors,which move the laser beam across the material perpendicular to the direction of the motion of the inspected material.Reflected,scattered light from various angles,is sensed by strategically placed photodetectors.Simple threshold detectors or a computer are used to provide sophisticated indication of the condition of the inspected material.

After this brief introduction concerning the various devices used in imputing visual data to computer,we are going to concentrate on the description of the rig ,which was used during this research. This rig was developed and built in the Instrument Systems Centre of the City University,and is used to digitised various images,which are then stored onto magnetic tapes.

5.3.2 Description of the rig

The rig (figure 5.6 and 5.7) was used for digitising photographs and storing the data onto 7-track magnetic tapes. It consists of a main frame,a drive mechanism,and a control signal board which interfaces the rig to the PDP11 computer.The rig was designed specifically for automatic inspection of surfaces. In outline and concepts,the rig is a scaled down version of a rig based at SIRA Institute which uses a laser scanning system.Instead of the laser system,the rig of the systems department uses for scanning a FAIRCHILD monolithic,self-scanned 1728 CCD element image sensor,mounted in a converted 35mm cammera. The amount of charge accumulated at each CCD element is a linear function of the incident illumination intensity.The output signal is in analog form.It is then digitised and preprocessed before being fed to the computer.

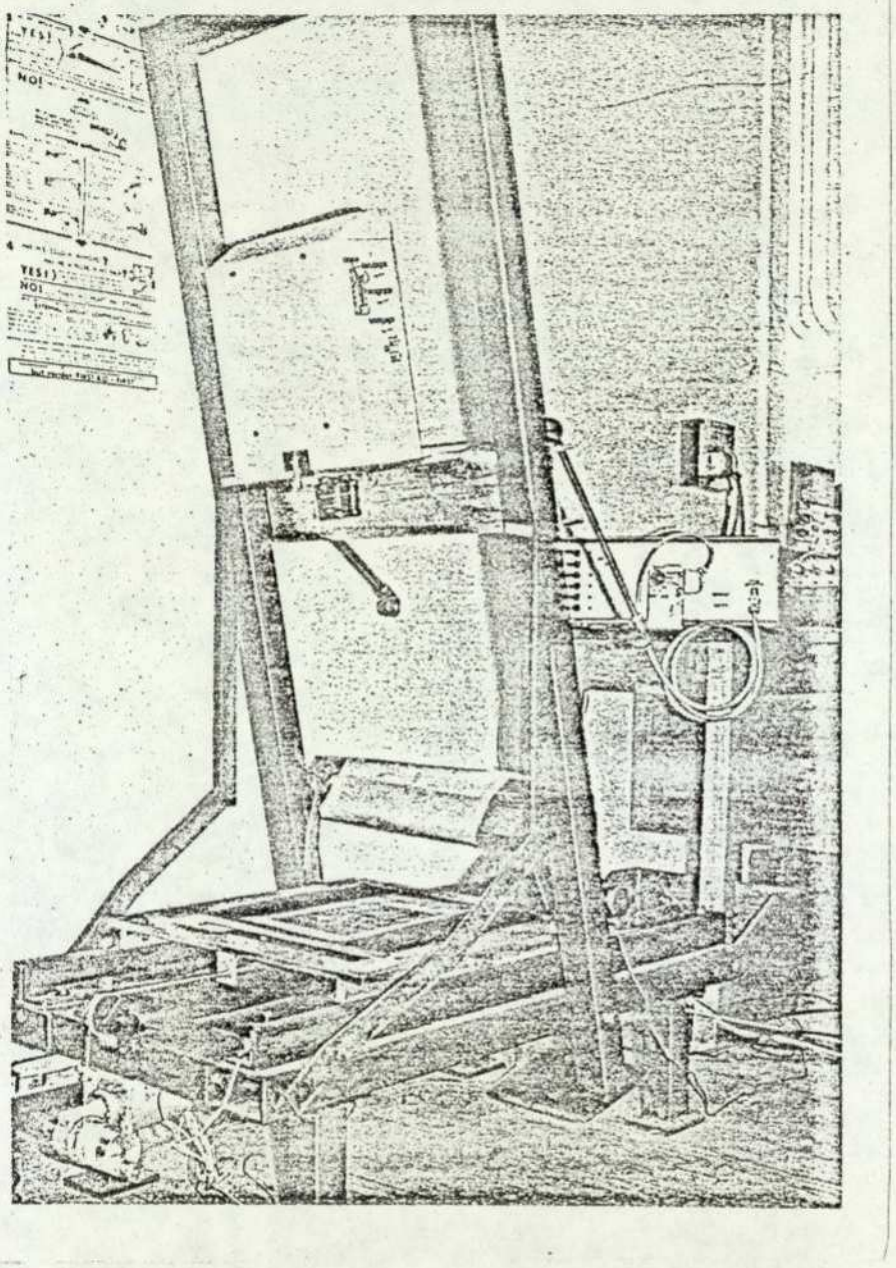


Figure 5.6 General view of the rig.

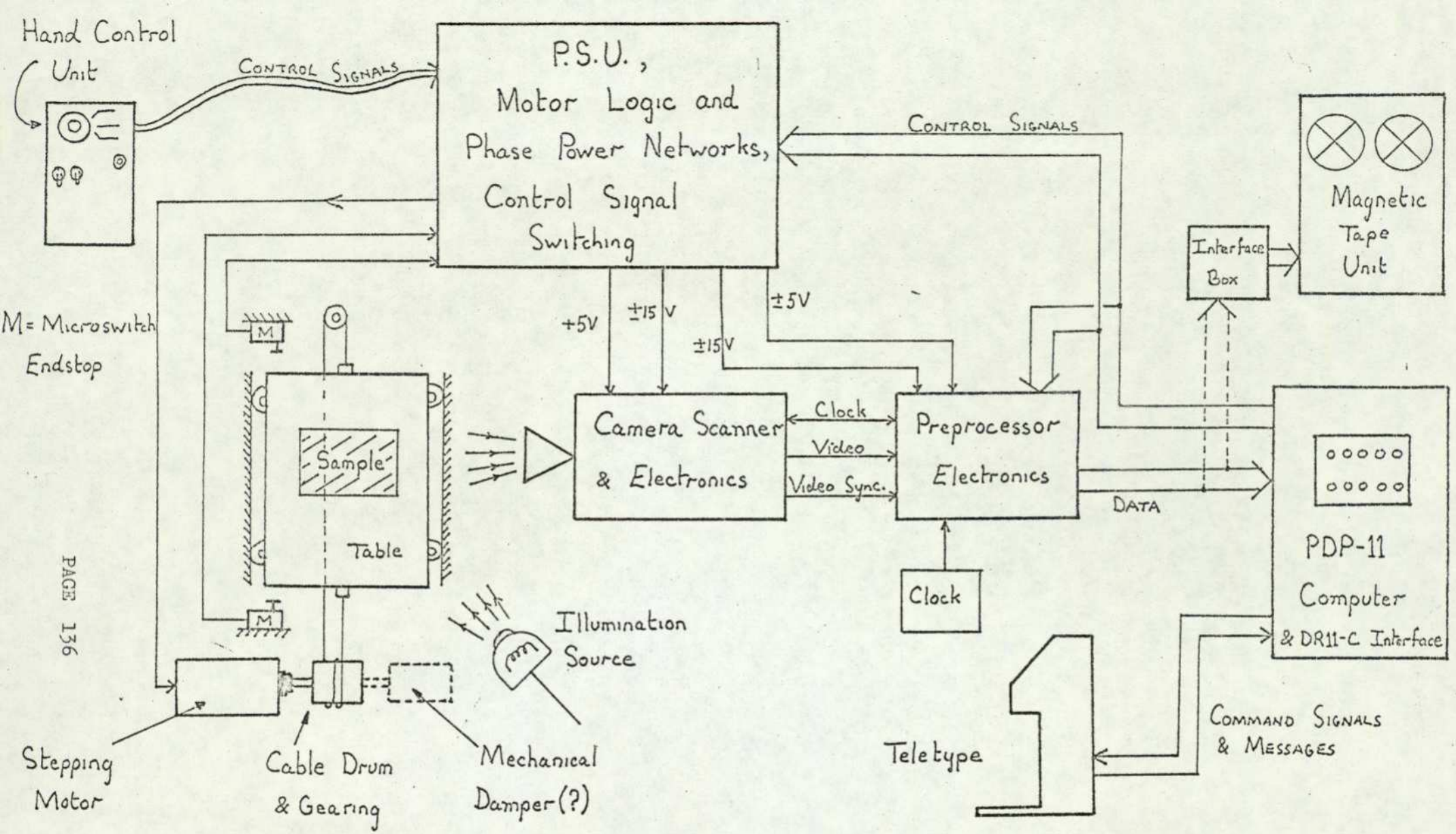


FIG.5.7 SIMPLE SCHEMATIC OF MAJOR COMPONENTS OF : SYSTEM.

The main frame of the rig consists of a horizontal section made of box section steel girders and supported by four legs,and of a vertical section made of bolted girders and capable of being pivoted with respect to the horizontal for altering the angle of incidence of light to the camera. The camera is fixed to the vertical section.The horizontal section supports a table made of thick Dural with its top capable of being replaced by a sheet of plate glass for illumination of samples from below.The table runs,by means of ball race bearings, on two runnners bolted to the horizontal section.The table can be illuminated from above or below.

The table is driven by a stepping motor with a maximum working torque of 1600gm.cm, through a 15.5 reduction worm and wheel gear,which enables table steps of 0.10325 mm for every pulse to the steping motor ,and a steel cable which is wrapped around the drum and attached to both ends of the moving table by tensioning clamps.The table is moved in one diresction or the other by rotating the drum,where a clutch mechanism is used for disengaging the motor drive from the gearing and drum,and which causes the cable to wind on one side and unwind from the other.For preventing the table from being driven hard against its end stops,microswitches,which sense that the table is at the end stop, are used to inhibit the drive pulses to the stepping motor.

The stepping motor has been used because of the ease with which it could be interfaced with the computer.The motor used is

an eight phase motor which has four pairs of coils that are switched sequentially.It is controlled through a standard electronic drive unit.The rig is interfaced to a manual handheld control unit,with facilities such as Fast,Slow and single shot movement,and to the computer by a control signal switching board.Both the stepping electronic board,the control switching logic board and the power supplies for the rig are housed in the main chassis.

The scanner assembly receives surface data from the sample in the form of reflected light,which is then passed to a preprocessor before being fed to the PDP11 minicomputer via a DR11-B interface units.A clock pulse generator provides some synchronisation between the scanner electronics and preprocessor,but,the whole scanning and motor control coordination are under computer control.The control unit accepts control signals fron the PDP11 computer,for controlling the motor,via a DR11-C interface unit.

Recently,Koulopoulos(1981) added a special feature for speeding up the transfer of data from the rig to the computer.This special feature is only useful for the scanning of silhouette component images.

5.4 Image Output Devices

5.4.1 Introduction

Although the representation of images by two-dimensional integer arrays is necessary for digital processing,this representation is not adequate for human viewing Therefore,to be inspected by human viewers,the two-dimensional integer arrays have to be displayed back as intensity readings.

A large number of computer graphics devices,with some suitable for image display, have been designed and built for particular use.The considerable expansion in the design of display devices was due mainly to the utilisation of microprocessors and of the microminiaturisation technology in general.The application of the display devices have steadily increased in various fields such as medecine,chemistry,engineering and image processing. In what follows we are going to concentrate on the image output devices used during the research,which were a microfilm plotter and a graphic option controller(GOC).

5.4.2 Microfilm Plotter

A microfilm plotter,based at the University of London Computer Center (ULCC),was used to display images,at the start of the research.The device was a CalComp 1670,which produces microfilm plots by photographing the path taken by an electron beam over the screen of a cathode ray tube.

On the CalComp 1670,lines are constructed of a series of dots consisting of sensitised spots on the CRT,with adjustable incremental distances between spots.The choice of the incremental distance determines the resolution and speed of plotting of the image.lines are not drawn with a simple beam on movement,rather the beam is unblanked for a controlled period at each of the incremental points along the plotted line.The unblanked period determines the intensity at a particular dot.The intensity varies from 0 to 30 with zero intensity when the beam is off. The screen of the CRT comprise a mesh of 16383 by 16383 points,although only a subset of these is used for a particular camera.

An image is generated by the combination of the beam movements and intensity variations under the control of the 1670 which interprets commands from a plot tape. The plot tapes are produced through the medium of a Fortran program utilising the CalComp 1670 Host Computer Basic Software (HCBS) provided on the

ULCC CDC 6/7000 machines.The commands are submitted by means of subroutine calls for various plotting functions such as drawing lines,symbols,advancing frames,formatting the plotting directives into a structure acceptable to the 1670.The principal routines,within which variables concerning the plot are defined according to the particular requirements for the display of the image,are concerned with information concerning beam movement and intensity.

for photographing the CRT screen,there is a choice between two cameras.The option for camera and film combination available at ULCC are a 35 mmcamera with unsprocketed film and 16 mm camera with sprocketed film.

5.4.3 Graphic option controllers

The image output device used in the PRIME system for displaying images is a microprocessor based 5660 graphic option controller.For displaying grey level images,I had to develop from scratch all the necessary software for controlling the microprocessor.

The device is a microprocessor based raster scan display generator.Graphical data generated by the Prime is interpreted by the GOC and mapped into four graphic stores each of 512 by 768

bits. Lines,dots,characters,symbols and blocks can be written or selectively erased in the store matrix for subsequent output to a display. Figure 5.8 shows a block diagram of the device illustrating how the basic functional blocks are interconnected.

A microprocessor forms the basis of the system monitoring messages sent by the main computer and initially passing all data to and from the downstream interface.Connected to this interface is a VDU. The GOC is initially transparent with all 'transmit' and 'receive' functions behaving as if the VDU was directly connected to the main computer. If the microprocessor detects a sequence of characters(+-*/) transmitted from the main computer which it recognises as a graphic control string it prevents this and subsequent data from being passed to the downstream interface port. Instead,it interprets the data as graphical or control information and processes it accordingly.

All of the functional graphic commands and graphical instructions utilise the printable ASCII character set. Control characters are in general excluded.Further,coordinate information is transmitted in simple numerical form equivalent to the actual cartesian coordinates of the display area.

Having interpreted any commands the microprocessor writes or erases the data in the pixel store or sets internal flags as appropriate.Associated with the processor is ROM for program storage and 1K of RAM used as a line buffer storing the

Figure 5.8
GOC Block Diagram (p. 143)
has been removed for copyright reasons

data on first-in first out basis if GOC is busy,for storing the programmable symbol set data and as a general scratch pad area.

The process of writing to the pixel store is controlled by a hardware vector generator which interpolates the position of the dots that fall on or near the straight line requested by the microprocessor in response to a command. The pixel store consists of single semiconductor memory planes each containing 366,592 bits arranged in a 512 by 716 square matrix.The word formed by the planes at each bit position corresponds to a picture element(pixel) for output to the display.

Asynchronously with pixel writing by the vector generator the scanner circuitry serially outputs the parallel pixel store contents to the video processor prior to output to the display.The scanner,which reads 16 bit pixel string from the stores for serialising,is synchronised with either an optional externally derived frame and line sync or with the interval sync generator.

In addition to controlling the graphic data the scanner also generates a cross hair cursor at coordinates defined by the microprocessor.The position of the cursor can be modified by the user using a graphic keypad.

The video processor receives the data for each pixel location from the parallel planes and inputs it to a

transformation table in a PROM.The table has a 16 bit output which is used to define the analogue signal levels for the three possible video outputs,Suitable transformation tables can produce (red,green,blue) information for colour or produce grey levels.The PROM has two sections permitting two alternative tables to be switch selected and there are facilities for blinking individual pixels on or off or between two states.

5.5 Conclusion

In this chapter,a brief survey of the computational facilities has been given to provide the necessary background for understanding the tools for processing,data acquisition,and data display which were available for developing algorithms for street scene analysis in order to automate vehicle guidance.

Some of the devices for data acquisition and display,which have been frequently used during this research,have been fully described.The description of the University of London Computer Centre system,which has been used in a batch mode at the beginning of the research,and the PRIME system,to which the work was transferred when the system was acquired by the Instrument Systems Centre,and which was used in an interactive mode,is also provided.

The processing systems used various operating systems.But when

using these facilities,FORTRAN was exclusively used as the programming language for the various algorithms for scene analysis.

6 LOCATION OF A STREET IN A STREET SCENE

6.1 Introduction

In the previous chapters we concentrated mainly on the general techniques which are used for extracting information from visual data in general. But in our particular research,we are mainly concerned with the analysis of street scenes representing the view seen by the driver of a vehicle,so as to guide the vehicle without human intervention.

In vehicle guidance,events,which take place inside the street,are of prime concern, whilst events happening outside are of secondary importance. Hence,although some objects lying completely outside the road(e.g. traffic lights) are relevant in a sophisticated analysis,the main region of interest ,in the street scenes used in this project,comprises chiefly the road and,in particular,its boundaries. The determination of the inside of the street and,in particular,its boundaries, is important because it identifies the bounded surface inside which the vehicle should always remain.Another important advantage of locating the inside of the street is that it permits to discard,early in the processing stage,the information in the scene which is redundant as far as the required analysis is concerned. The location of the inside of the street permits to divide the scene into two regions:the first containing information of interest for the subsequent analysis and consisting of the inside of the street,the second containing information of no interest and consisting of everything else,which

is not inside the street.

Our first concern is therefore to try and isolate the street from anything else contained in the street scene image.In this research ,we restricted ourselves to the analysis of scenes containing streets with borders that are approximately straight and described by the model in figure.1.3.When success is achieved with these kinds of street scenes,then it would be reasonnabely easy to extend the analysis to general street scenes.The important step is to develop a methodology for analysing particular street scenes which could be extended, when needed,to street scenes in general,and in a further stage,pavements and traffic lights could be included in the analysis.

The problem of locating the street reduces to a problem of segmentation, in which we are required to divide the street scene into regions characteristic of specific objects within the scene.Our first approach consisted of using thresholding,assuming that,because of the viewing position (see figure.1.4),the street filled a large part of the field of view,and that the grey level intensity of the street was more or less constant,and different from that outside its boundaries. Unfortunately this simple approach did not work(figure.6.1 and figure.6.2).

The second method,that we tried,was to detect edges in the image and try to locate the boundaries of the street,which manifest themselves as discontinuity in intensity. We tried many method of

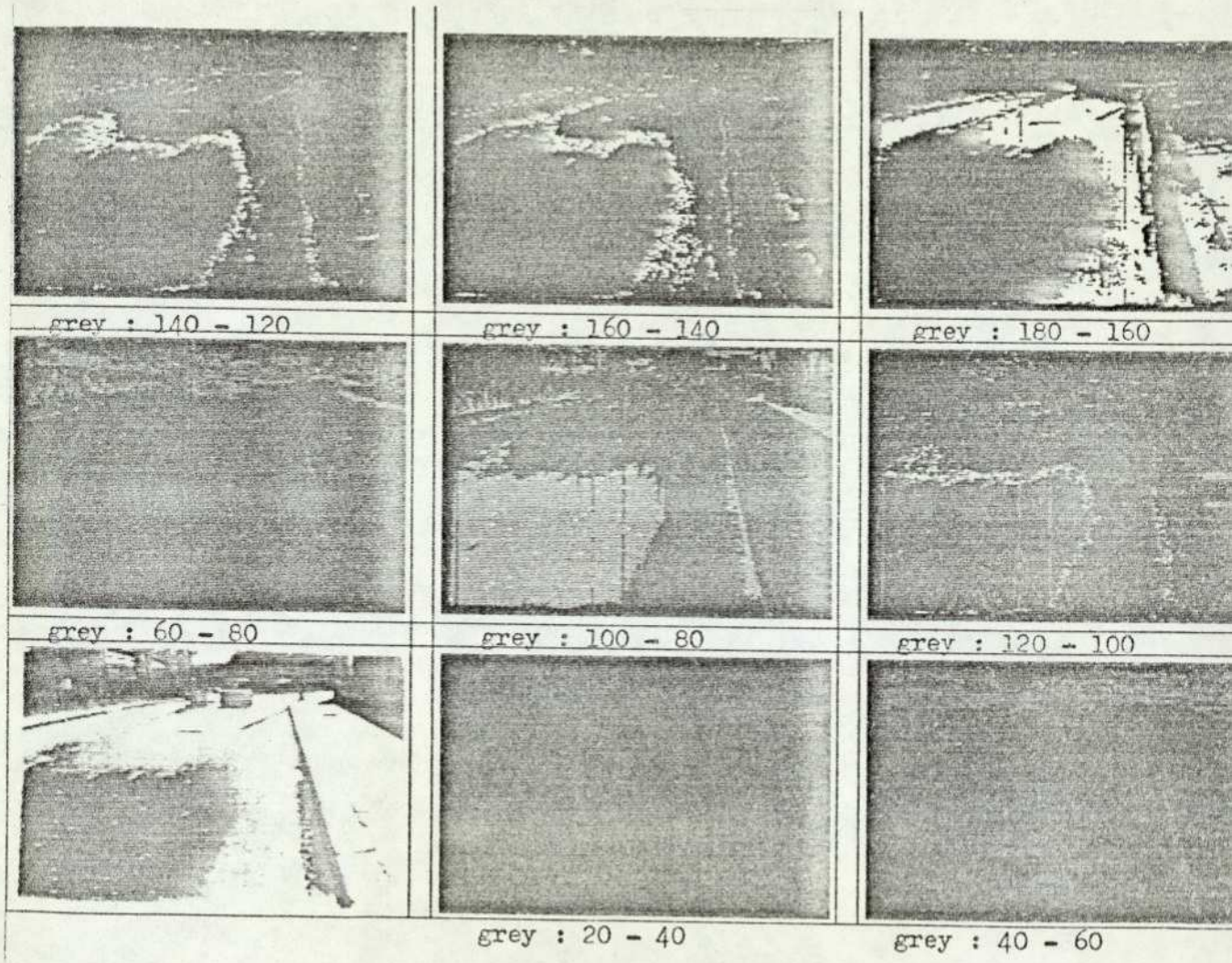


Figure 6.1: Original image,and the same image thresholded at different grey level values.

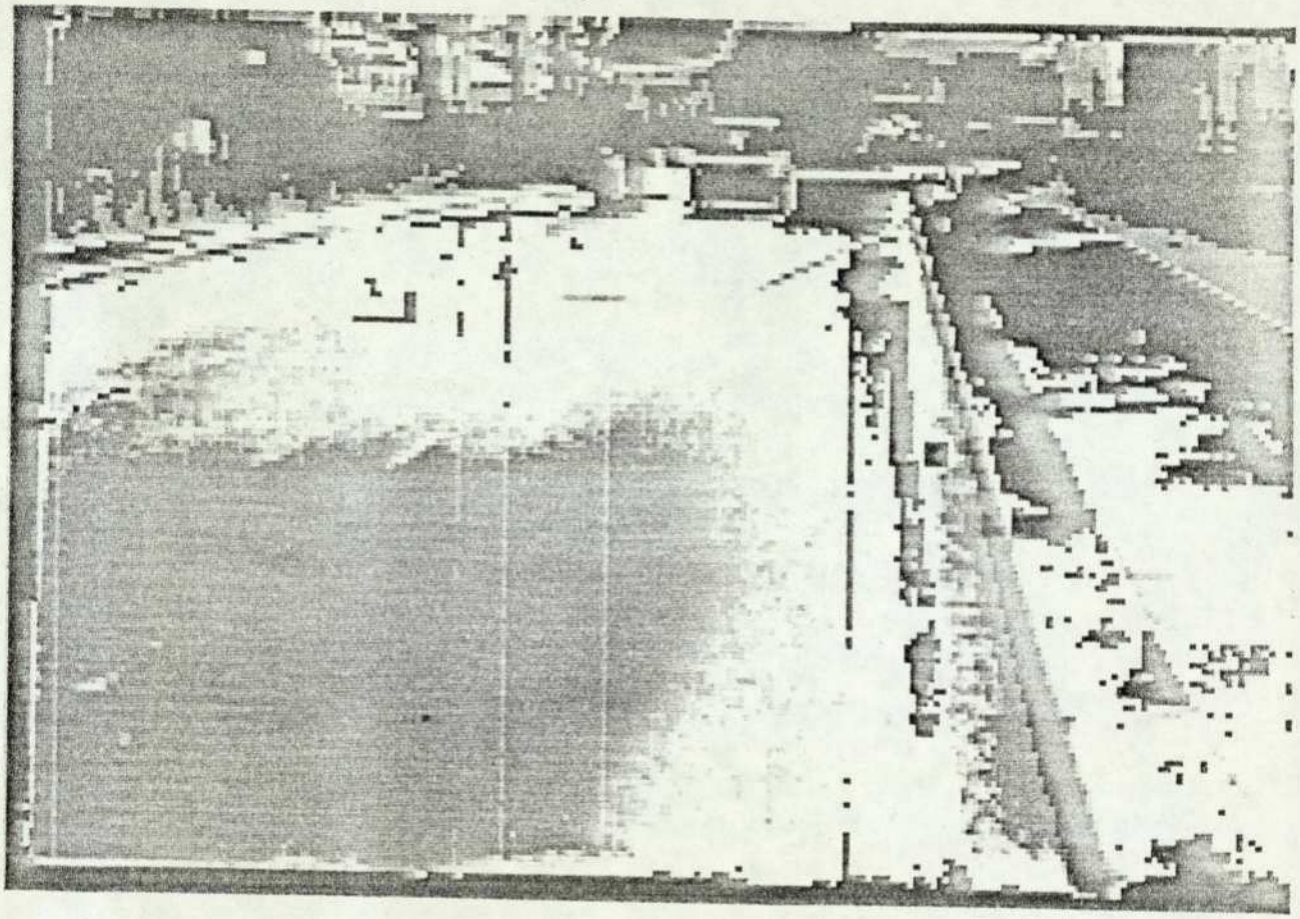


Figure 6.2: Thresholded image (starting grey value :80,

range:100)

edge detection based on convolving the image with masks, with the value of the convolution sum giving the magnitude of the gradient.

The method based on edge detection was still not adequate for street boundaries extraction,thus we had to combine thresholding method with edge detection and a final filtering operator,known as Hough Transform,for finding the boundaries. Combining the simple thresholding with edge detection and Hough transform we were capable of extracting the street from the whole street scene image.

This chapter describes how the data was obtained,how the optimal resolution was determined,and finally how the location of the street in the scene has been achieved.A complete description of the techniques,used during this stage of the research,is also provided.

6.2 Data Preparation

The selection and gathering of the data,used in this part of the project, was part of the research. It consisted of black-and-white photographs of street scenes (figures :6.3,6.4,6.5,6.6,and 6.7), which were obtained using a conventional 35mm camera.

As mentioned earlier in chapter 4,a picture can be defined by a mathematical function which represents the variation of the brightness or colour of points of a two-dimensional space(flat

6.2 Data Preparation

The selection and capturing of the data,used in this part of the project, was part of the research. It consisted of black-and-white photographs of street scenes (figures 6.3,6.4,6.5,6.6,and 6.7), which were obtained using a conventional 35mm camera.

As mentioned earlier in chapter 1,a picture can be defined by a mathematical function which represents the variation of the brightness or colour of points of a two-dimensional space(flat

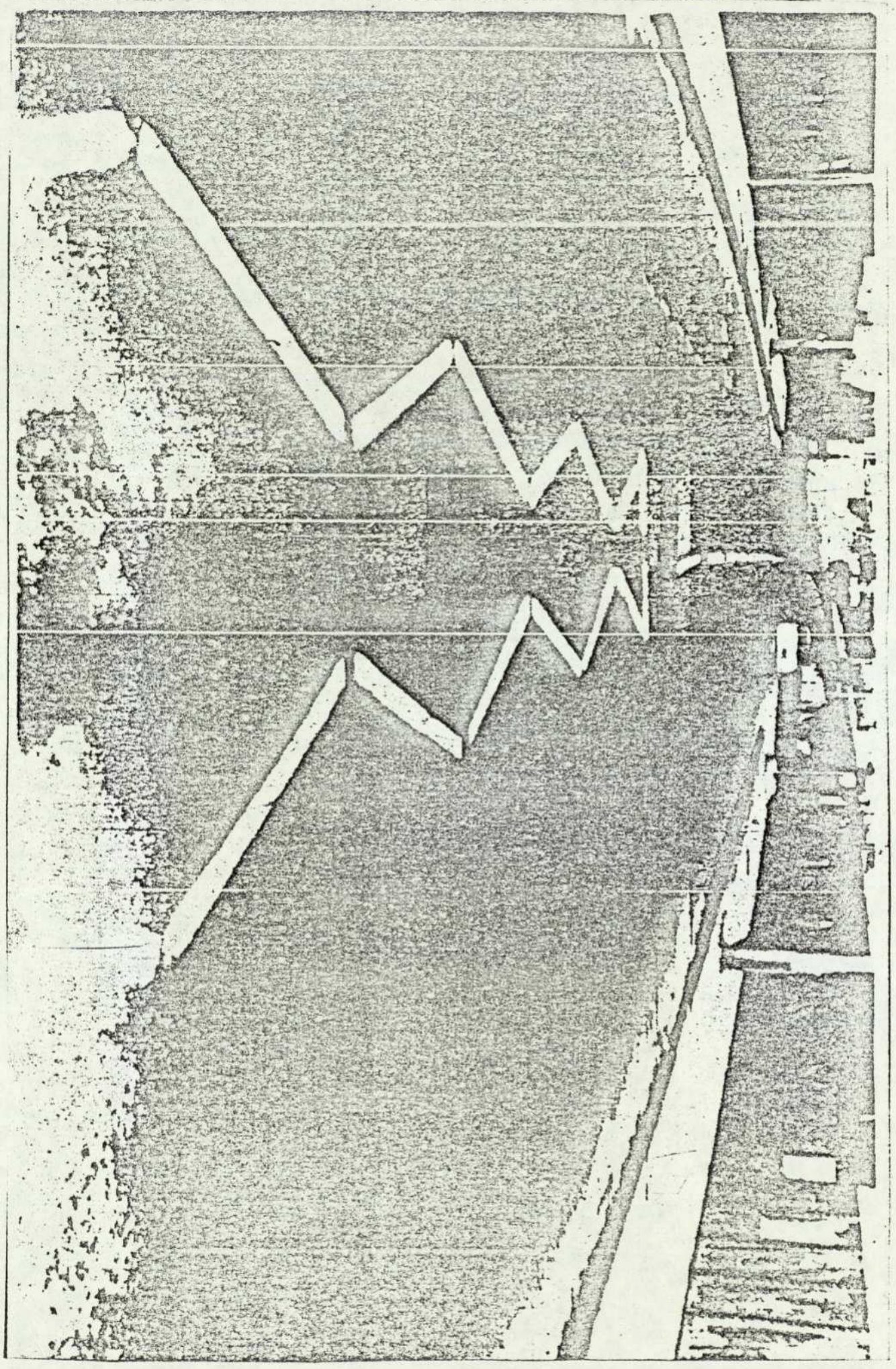


Fig. 6.3 Street scene image.

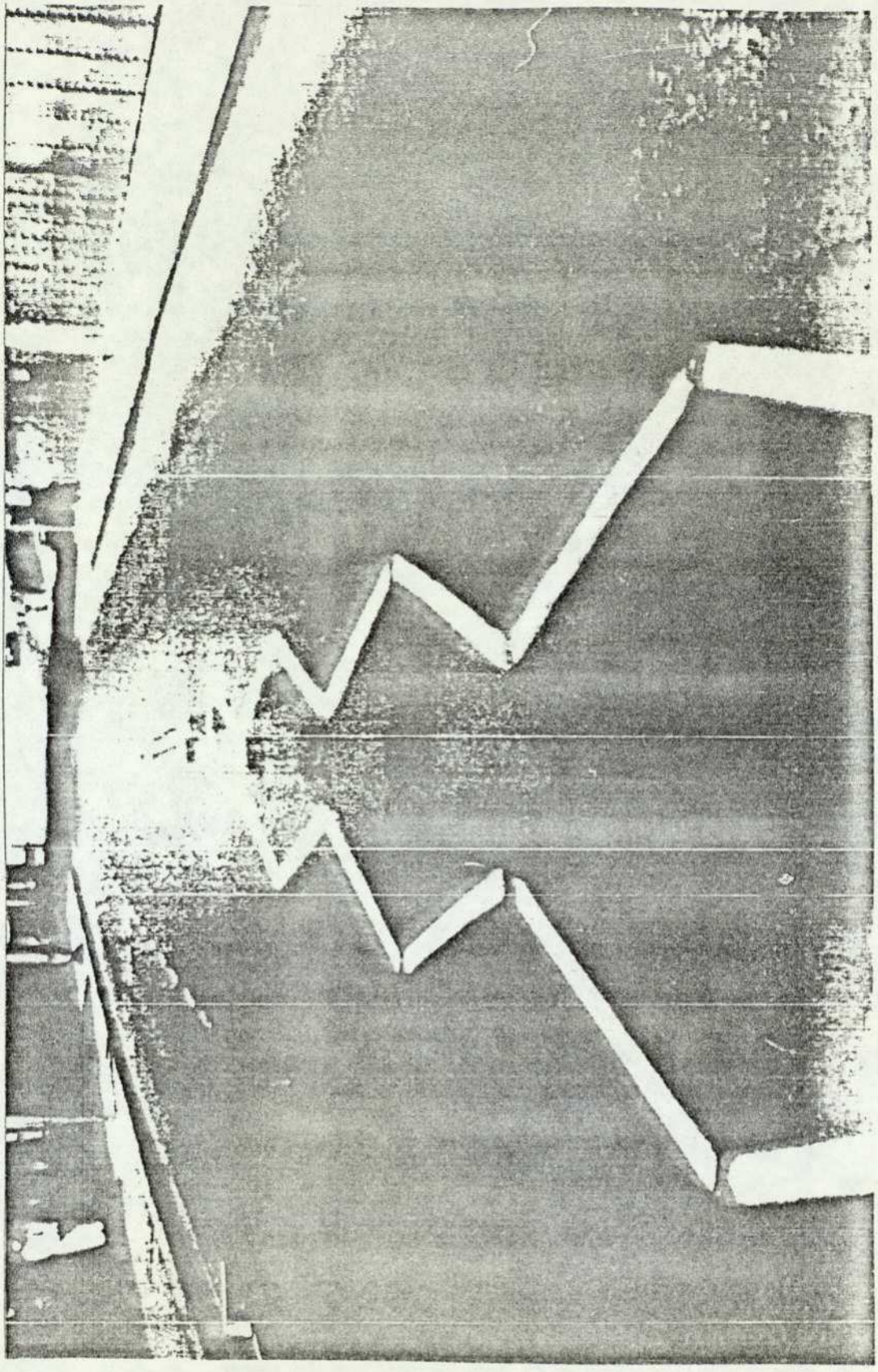


Figure 6.4 : Original street scene image.

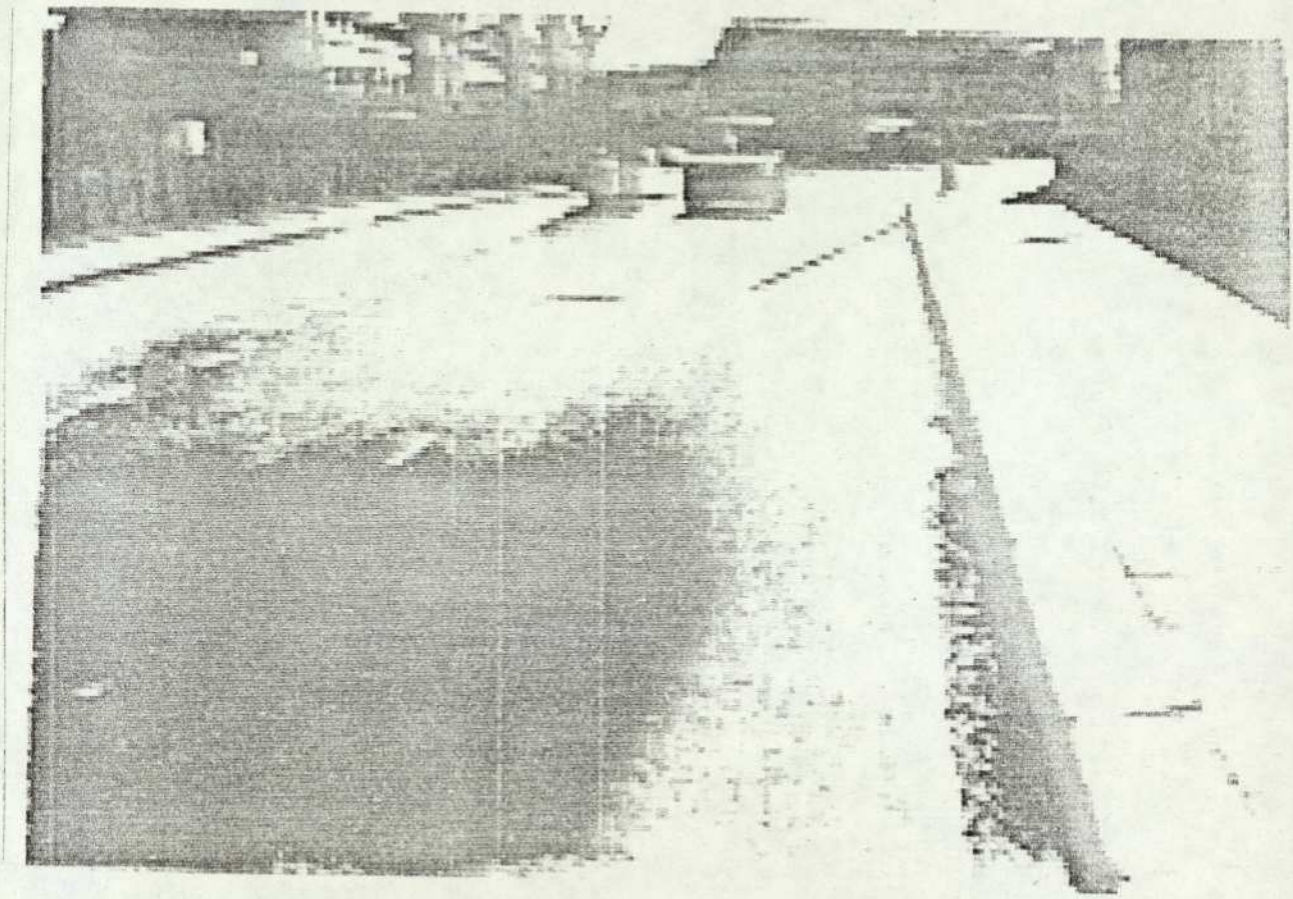


Figure 6.5 : Original street scene image.

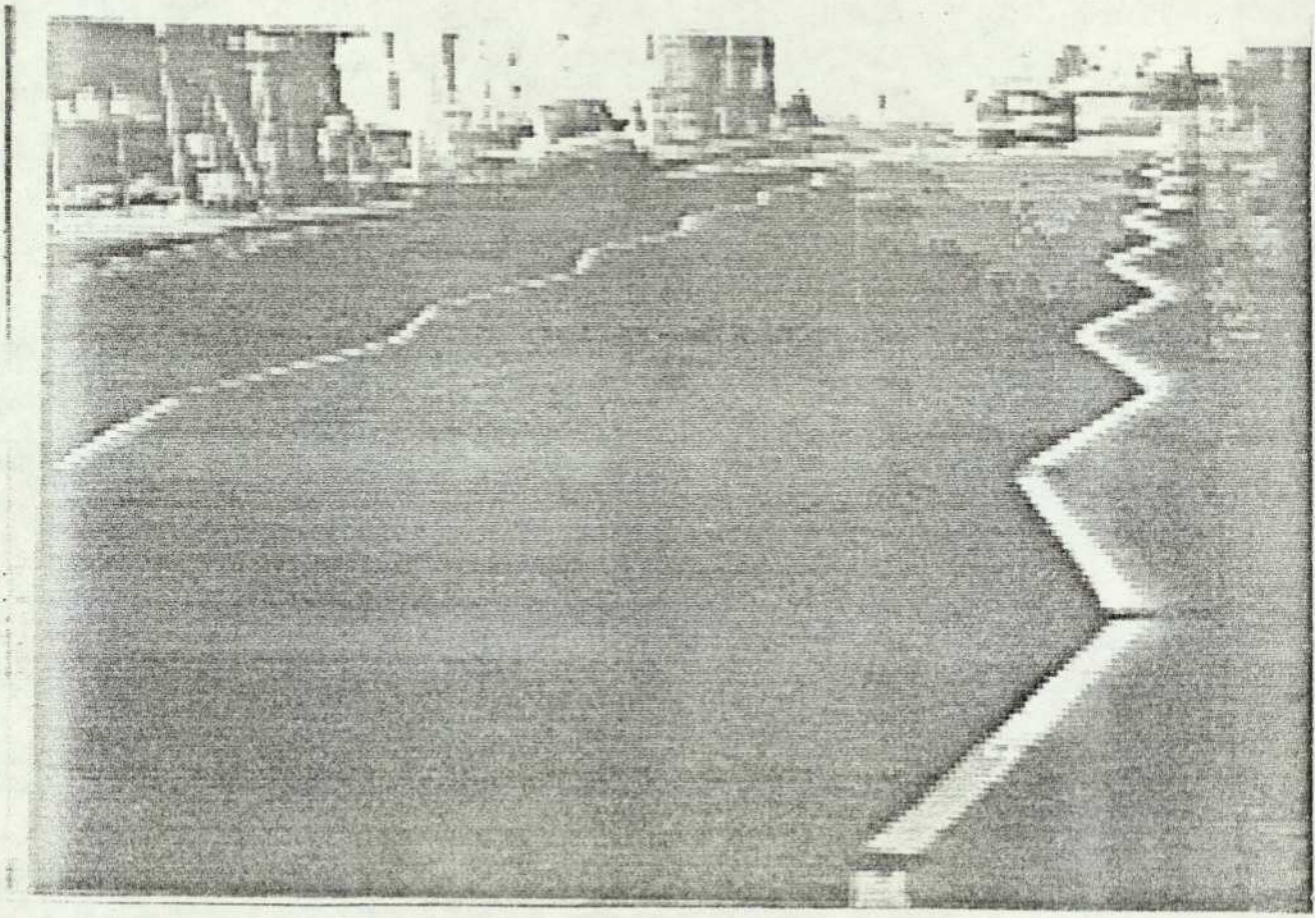


Figure 6.6 : Original street scene image.

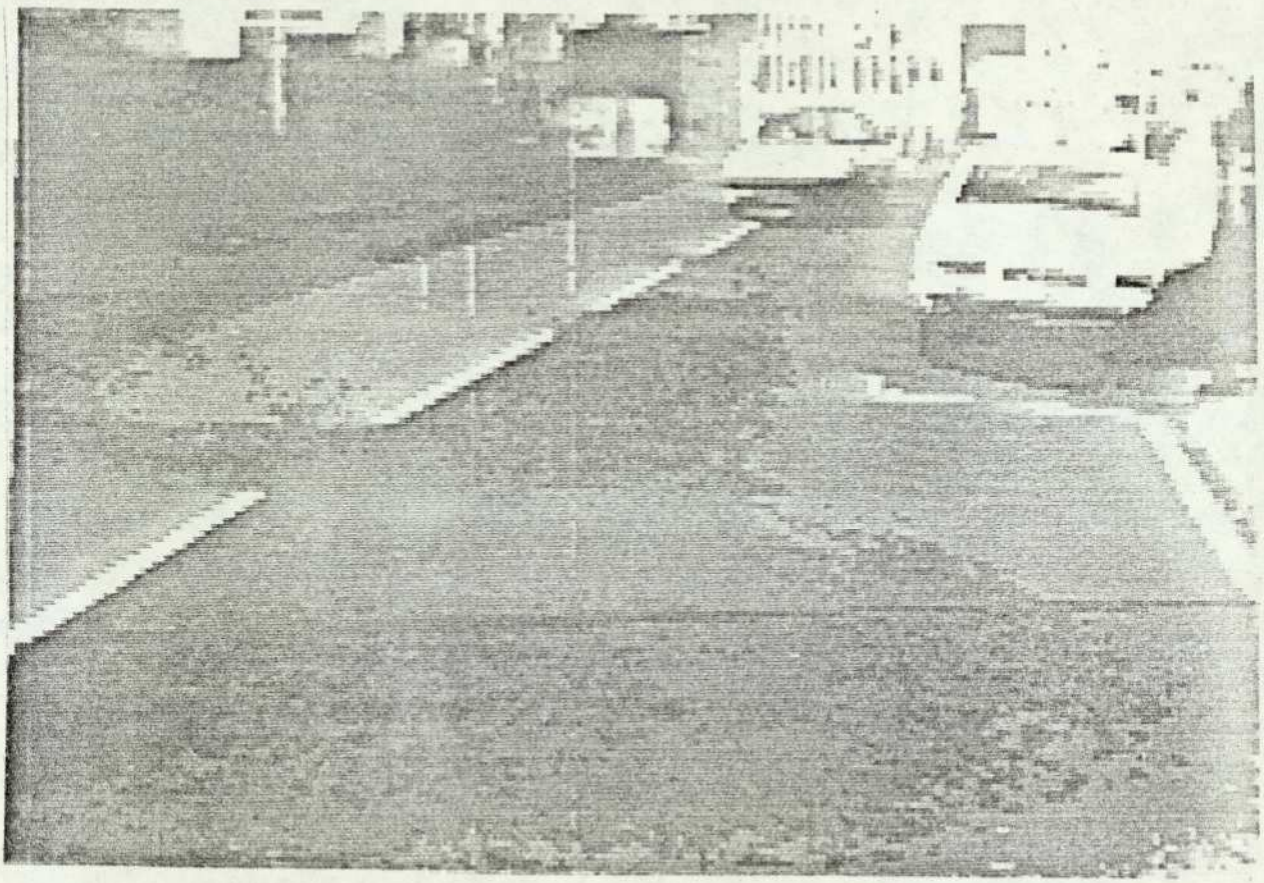


Figure 6.7 : Original street scene image.

surface).In our case,the flat surface was,originally,a film whose points are image points of the object points which lie in a three-dimensional space in the field of view of the camera.Each object in the scene,which is illuminated,reflects part of the light in different directions.Therefore,as illustrated by figure 6.8,every point can be considered as being the centre of a bundle of reflected light rays.The rays from a bundle,which strike the lens of the camera,are refracted by the lens and then intersected in a point which lies on the film and becomes an image point of the object.

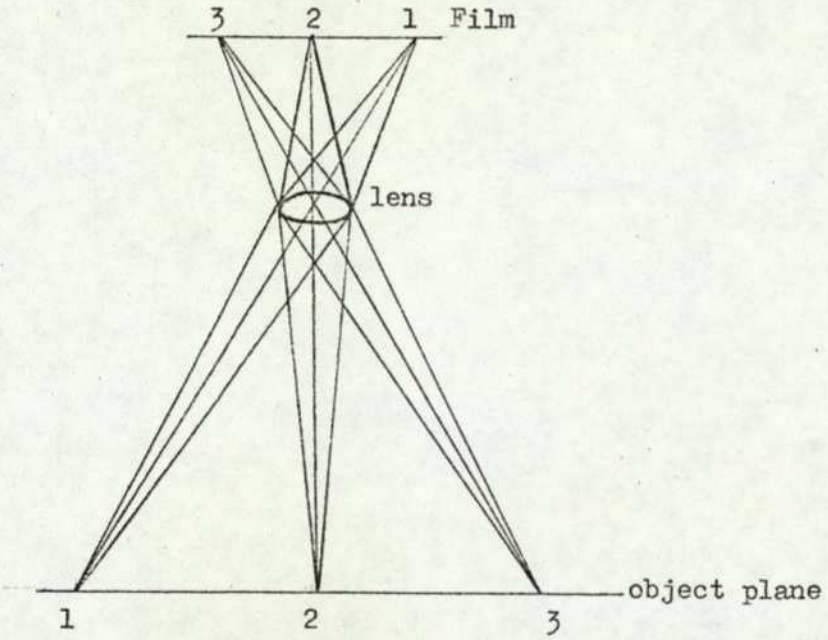


Figure 6.8. Imaging Process of a camera.

Because we are using black-and-white images,the mathematical function, representing the image,is a single real valued function whose value at a point is called grey level of the picture at that point.Because all our processing was done with a computer,the picture has to be digitised.The image is then represented as a

rectangular array of real numbers that are ,subsequentely,quantised to a set of equally spaced grey level values,which,in this research,go from 0 to 255 (0 for black and 255 for white).

Using the PDP11 and the rig described in chapter 5,five black-and-white photographs of street scenes were digitised and stored on three magnetic tapes with seven tracks.The tapes contained five files,each containing 1120 records and each record containing 1728 12 bit-words.The three tapes were subsequentely transfered to ULCC system and copied into three 9-track tapes,which are stored in a permanent library.

The resolution of an image,which is represented by the size of the array of the digitised image(1120x1728),is an important factor in processing,because the processing time is a direct function of the resolution.Hence,it is desirable to take the smallest resolution possible.The first task,in this project,was therefore to determine the optimal resolution.

6.3 Determination Of The Optimal Resolution

The lower resolution used,the less amount of information would be available from the given visual data.However a low resolution image requires less data for its specification than a high resolution image and is thus more ecnomical to process and store.The

optimal resolution is the lowest resolution which contains sufficient information for adequate analysis.The optimal resolution will depend on the kind of information to be extracted from the image(for example the size of the smallest object to be recognised in the image will determine the resolution to be used). It would be very tedious to devise a methodology to determine the optimal resolution in general.As it is not the object of this project,a simple criterion was used. This criterion was that the optimal resolution is the lowest resolution for which a human observer can clearly distinguish and recognise all relevant objects in the image.This has been carried out for five street scene images,and an optimal resolution was determined.

This has been achieved by displaying each picture six times with decreasing resolution.This involves repeated 2x2 averaging of the original image,so that a 2 to the power n by 2 to the power m image is successively reduced to 2 to the power n-1 by 2 to the power m-1,2 to the power n-2 by 2 to the power m-2,...yielding an 'exponential pyramid' of images at successively lower resolution(figures 6.9, 6.10,6.11,6.12,6.13,6.14).So we can work at the lowest resolution and go up the resolution pyramid when needed.

A viable strategy would be to use lower resolution images to generate 'plans' for analysing the full resolution image.This would be very much like the strategy which is used by human beings, with the retina divided into two fairly distinct regions,the fovea through which sharp vision (high resolution) is obtained for about

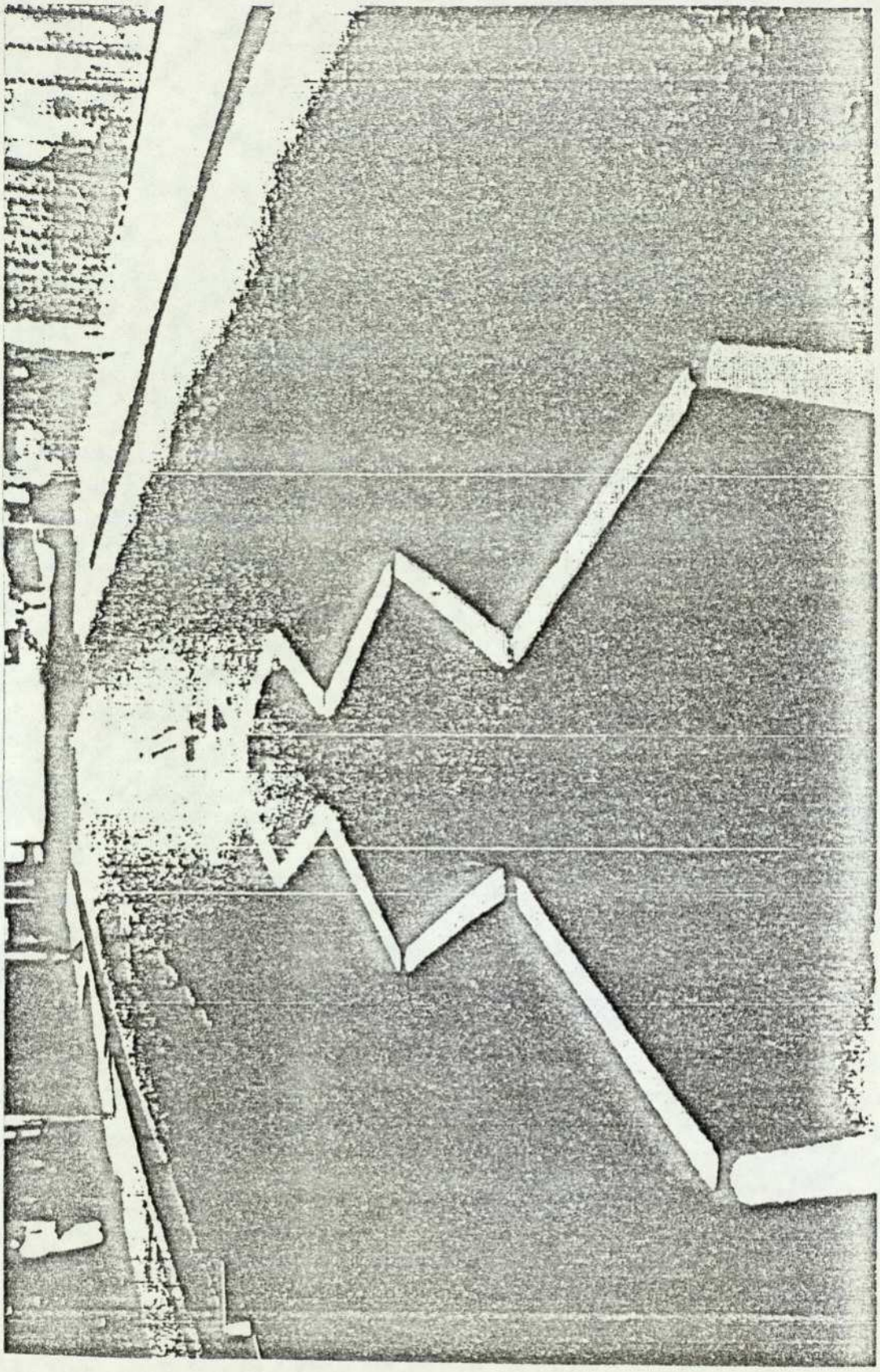


Figure 6.9: Image with resolution :1728x1728 .

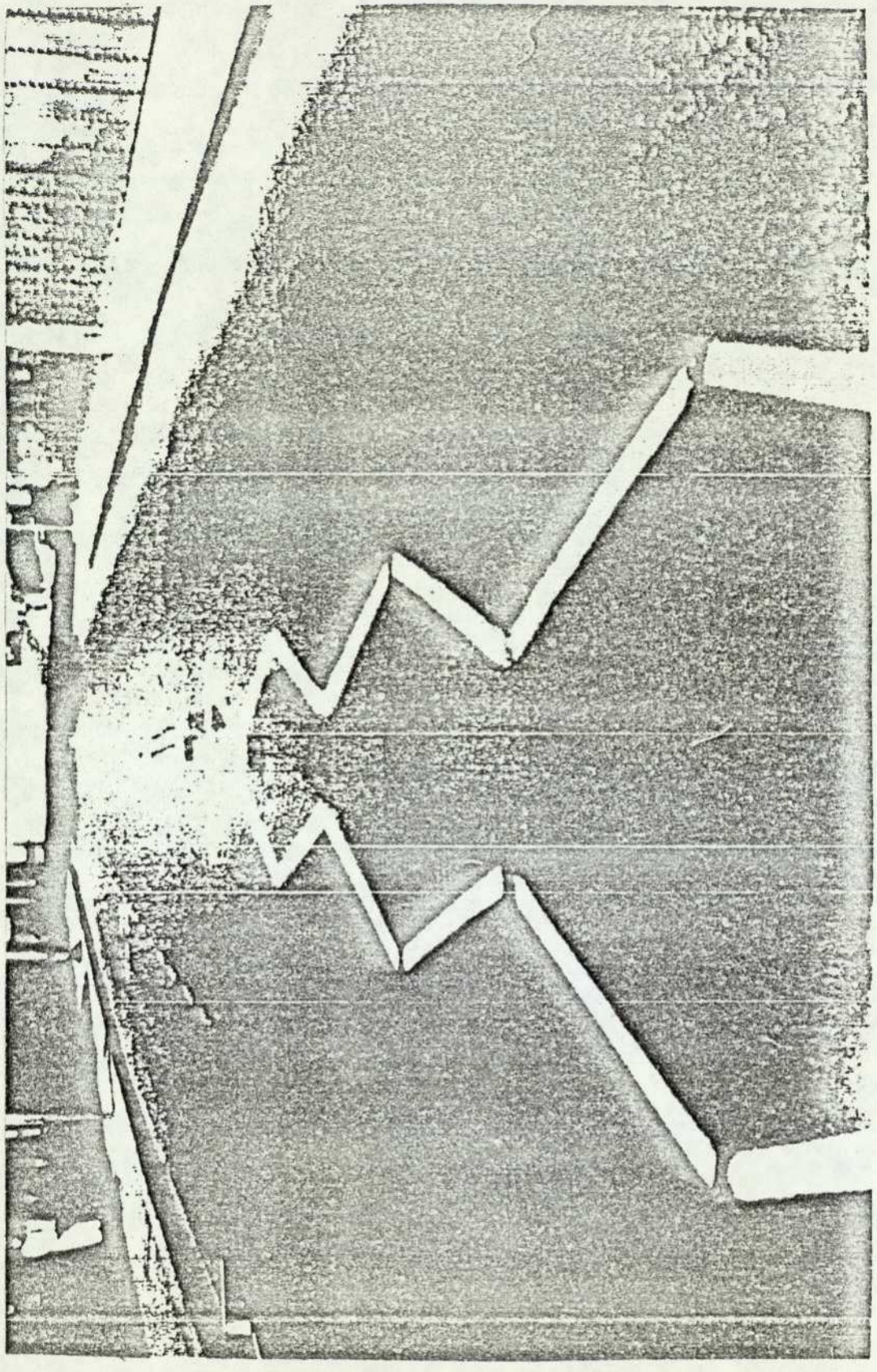


Figure 6.10: Image with resolution :864x560 .

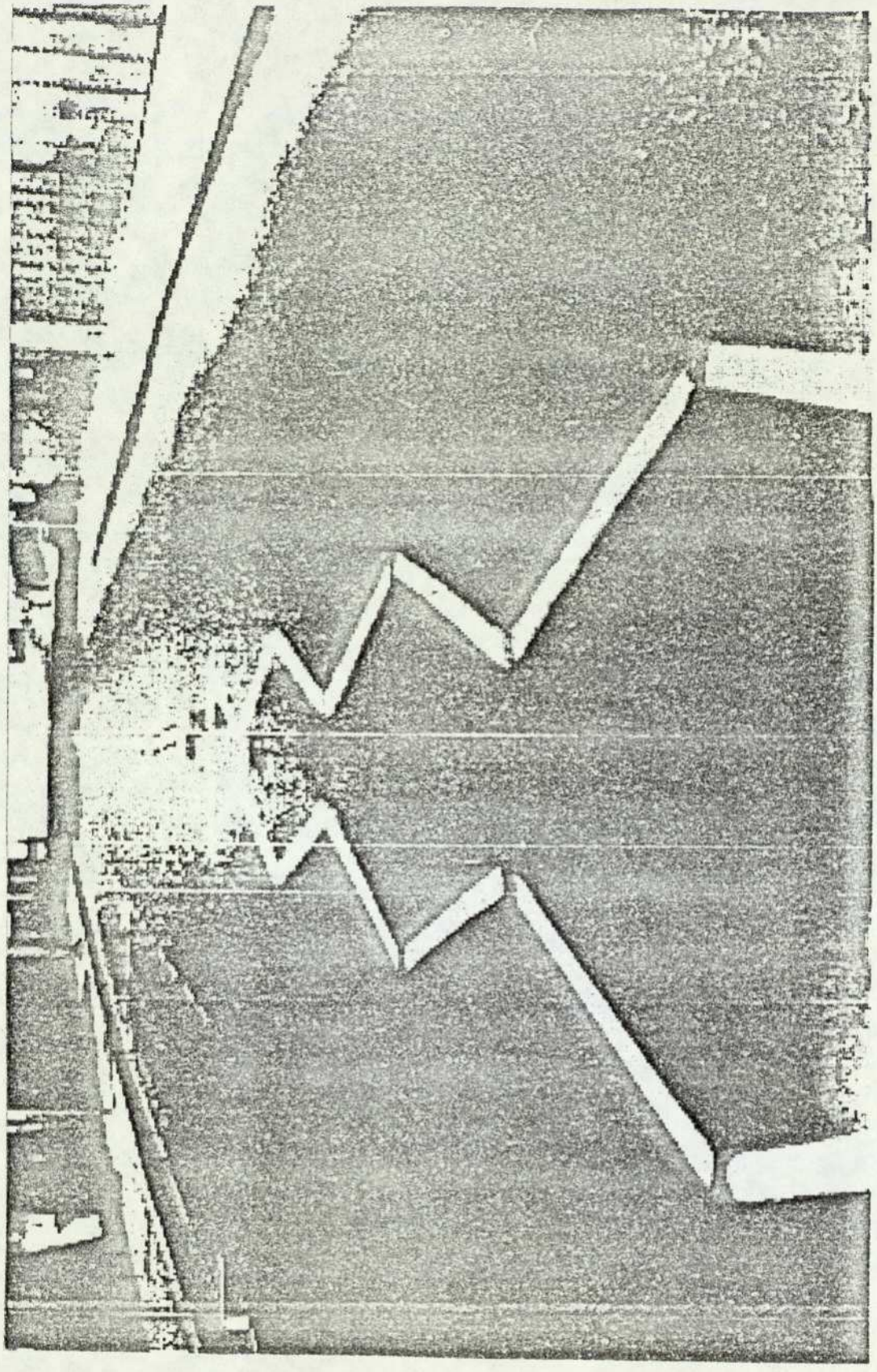


Figure 6.11: Image with resolution :432x280 .

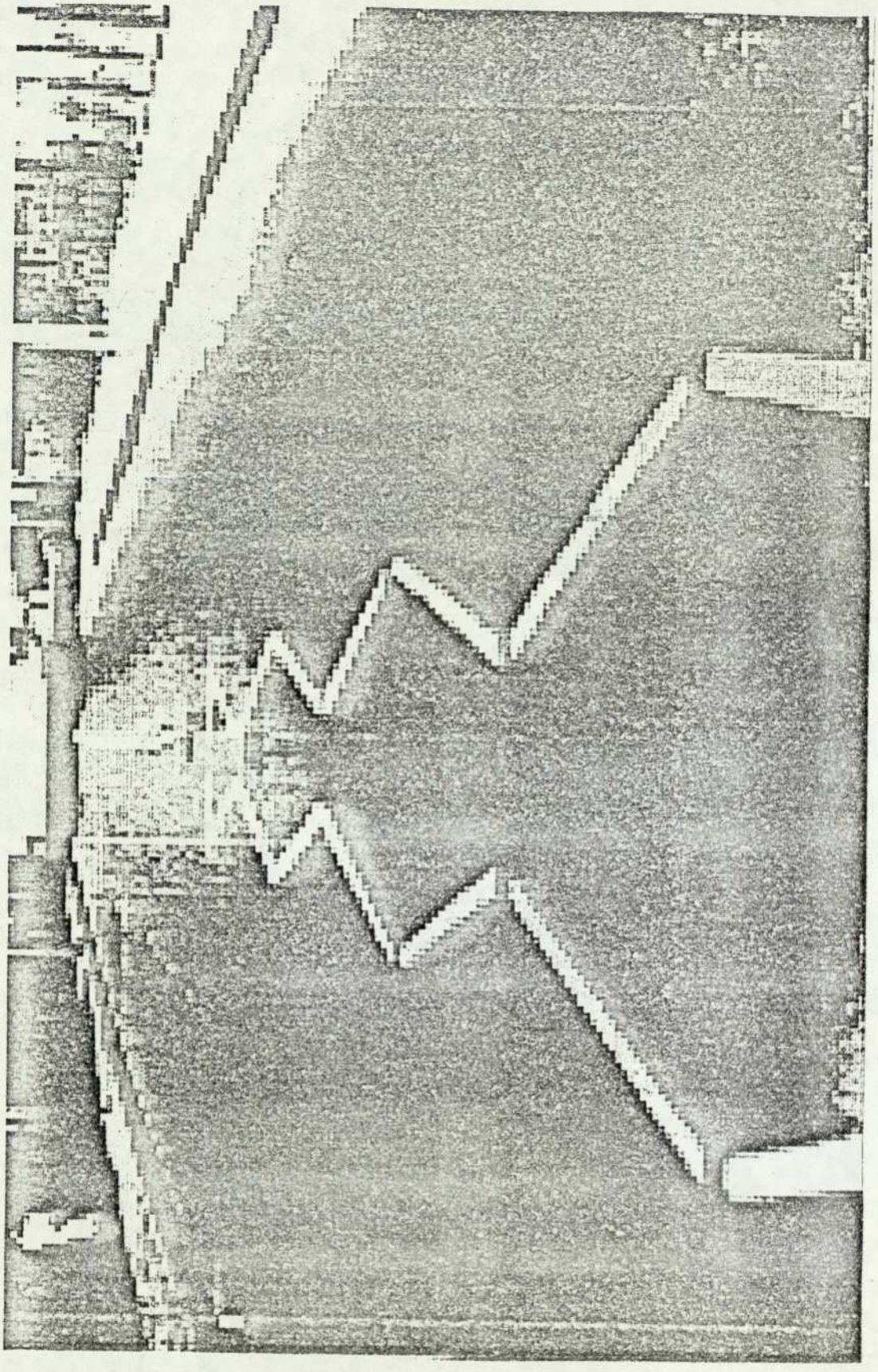


Figure 6.12:Image with resolution :216x140 .

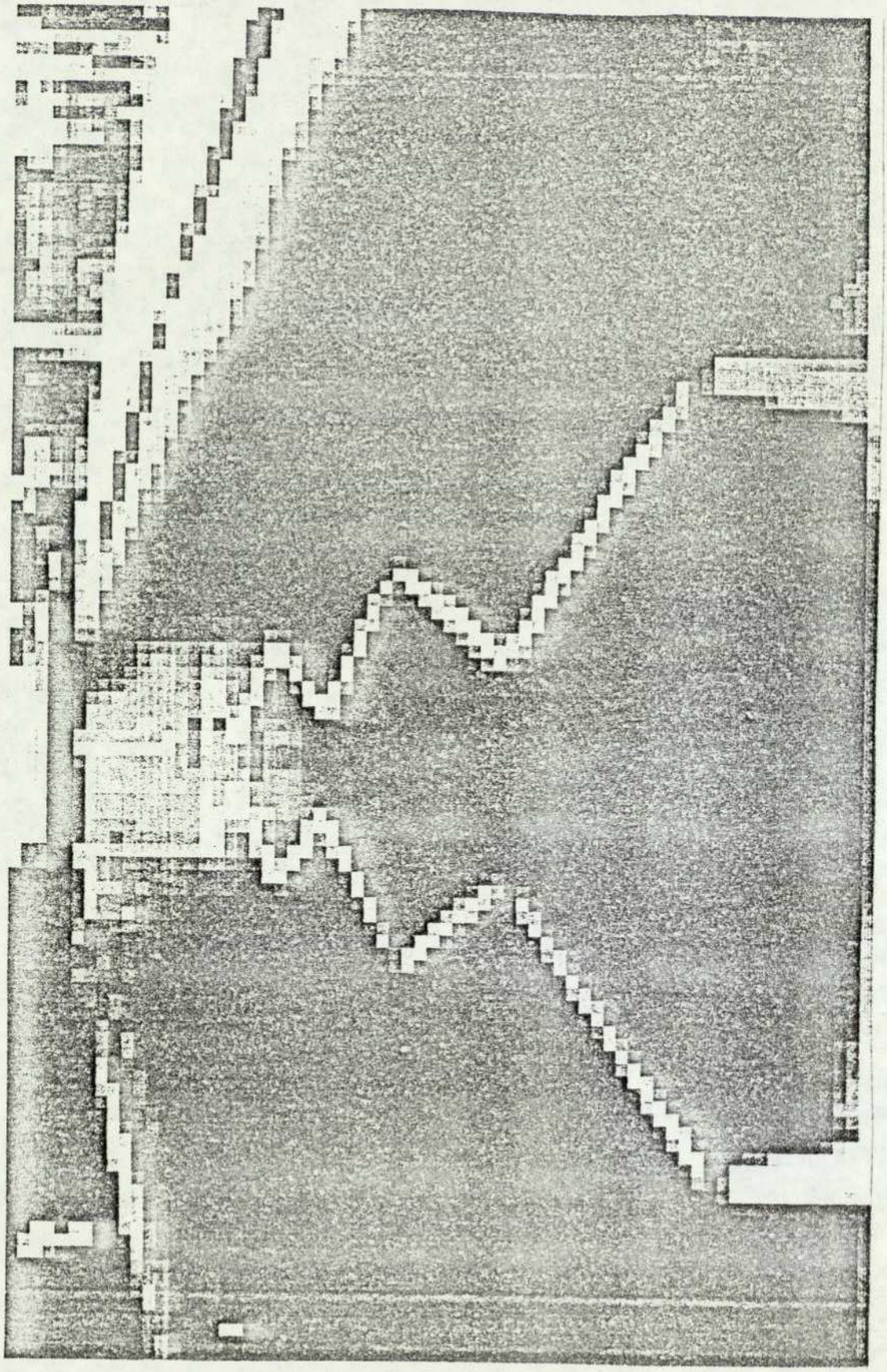


Figure 6.13:image with resolution :108x70 .

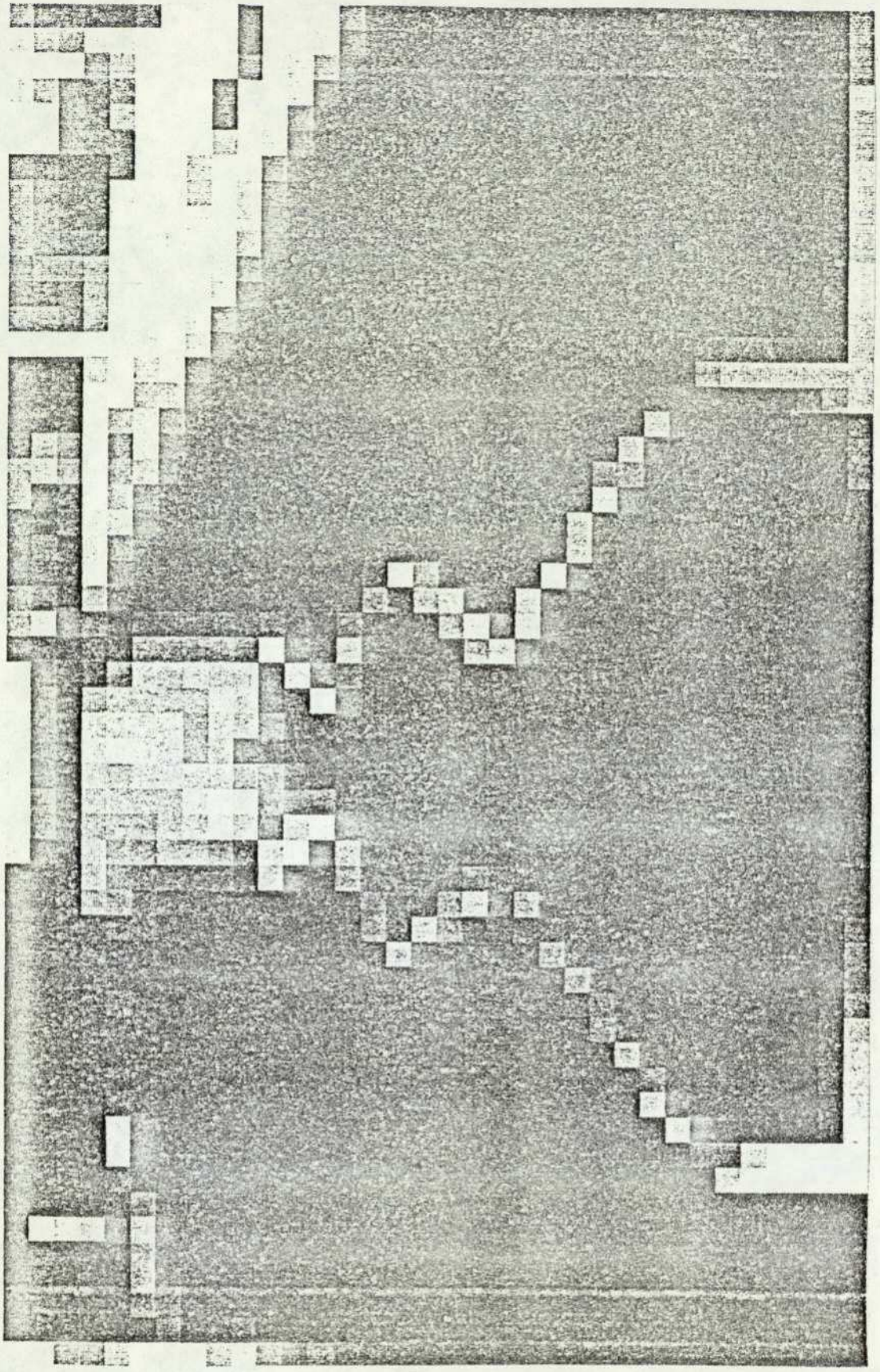


Figure 6.14: Image with resolution :54x35

two degrees of the total visual field,and the portion of the retina peripheral to the fovea (low resolution) ,which accounts for all the rest of the visual field (180 degrees horizontal and 60 degrees vertical). In the peripheral portion of the retina,the outputs of more than 100 primary receptors are processed directely in the retina and relayed into a single channel. In the fovea region,however,there are many optic nerve channels as there are receptors. So when needed the brain can direct the fovea,which represents a high resolution small window,on the region of interest whose 'description'by the peripheral retina to the fovea,which represents lower resolution,does not provide enough information.

Figures 6.9,6.10,6.11,6.12,6.13 and 6.14 show images of a street scene in which the resolution has been reduced successively by a factor of 2.It is seen that a threshold resolution exists beyond which a vehicle cannot be identified specifically as being a vehicle.This effect has been designated as being worthy of a more detailed examination.The results of the examination will be outlined in what follows.

To quantify the loss of information,which results from the reduction of the resolution,we used a special measure of information termed entropy. The concept of entropy was derived from thermodynamics.It follows from the second law of thermodynamics that any physical system in the natural world left to itself tends towards a condition of minimum coherence or maximum randomness. This condition is often described as having maximum entropy,with

entropy being essentially a measure of disorder or of uncertainty in our knowledge.

If the finite n possible outcomes of an experiment X have probabilities p1,p2,...,pn,the amount of information generated by experiment X is defined by

$$H(X)=\sum_{i=1}^{n}-(pi)x\log(pi)$$

This sum bears a formal resemblance to a quantity called entropy in statistical mechanics first introduced by Boltzman(Levine and Tribus(1978)).For this reason the function H(X) is called the entropy function of p1,p2,...,pn. The fact that this function is a maximum when outcomes are equally probable and zero when there is complete certainty provides some justification for considering it as a measure of uncertainty.

A major significance of entropy,for practical application in information theory, comes from the source coding theorem which states that if H is the entropy of a source letter for a discrete memoryless source,then the sequence of source outputs cannot be represented by a binary sequence using fewer than H binary digit per source digit on the average,but it can be represented by a binary sequence using as close to H binary digits per source digit on the average as desired.This is widely used in the determinationof the optimal number of digit for coding source outputs.

As far as this research is concerned,the entropy of eighteen images (three images with six different resolution) were obtained.The results are tabulated in table 1:

Resolutions / Images	1728x1120	864x560	432x280	216x140	108x70	54x35
1	2.79	4.03	5.18	5.89	6.21	6.39
2	3.99	5.15	5.84	6.14	6.33	6.45
3	3.84	5.00	5.69	5.98	6.16	6.23

Table 6.1. Entropy of digital images with decreasing resolution.

The optimal resolution,which was chosen by visual inspection,is the resolution 216x140 with an entropy around 6.Hence an entropy around 6 seems to be a threshold. For a human observer the images of resolution 1728x1120,864x560,432x280 and 216x140 are nearly all similar.For the resolution 108x70 a human observer can recognise the objects,but the objects are not as clear as for the above resolution.For the resolution 54x35 a human observer cannot recognise any object at all.

The optimal resolution is therefore 216x140 and would be used in all further processing when possible.

6.4 Thresholding

Many objects in an image have generally a reasonably uniform brightness, with a background of differing brightness.In some specific cases,such as handwriten or typewriten text,and bright objects against black background(metaliic component on a conveyer belt),the brightness is a distinguishing feature that can be used to locate the objects of interest.In trivial cases the object can be characterised by a single grey level value. However,for practical problems,the image is subject to noise,and a broad range of grey scale is needed to characterise an object.

However,for non-artificial images the problem is much more complicated because we do not have a uniform background,and the objects themselves could have such large range of grey level,as to make the brightness a non acceptable feature for characterisation.

Although thresholding cannot be used on its own to segment a complexe image, it could be used in preprocessing,with preprocessing taken to mean the rejection of data which is irrelevant to the problem we set out to resolve.

Having thresholded street scenes (figures 6.1 and 6.2) and realised that it cannot be used on its own for locating the street,and having decided to use thresholding as a tool for preprocessing we were faced with two problems:

1.How to determine the width of the range of grey levels which characterise the object of interest?

2.How to determine the lowest grey level inside the range?

In our particular case,where the concern is to analyse a specific street scene configuration,we used the fact that the street will be the biggest region in the image.To determine the width of the range of grey levels characterising the street ,some experiment were made with the five images of street scene at our disposal. The experiment involved visual inspection of films of the images (figures 6.15,6.16,6.17, and 6.18) displayed with 3 ranges whose width were respectively 20,40,and 60 using images with grey levels going from 0 to 255.The lower grey levels were determined as follows. First the histogram (figure 6.19) of grey levels was computed,then the biggest region with a specific width of the range was determined by integrating the histogram.For a range with a given width we calculate the number of cells contained in the range starting from every gray level.Finally we determined which grey level,from where the range start,gave the biggest region.

From the analysis of different thresholded street scene images (figures 6.15,6.16,6.17,and 6.18), and the analysis of their histograms,the street was found to be characterised by a range of width of about 40,and starting at different grey levels depending on the image.

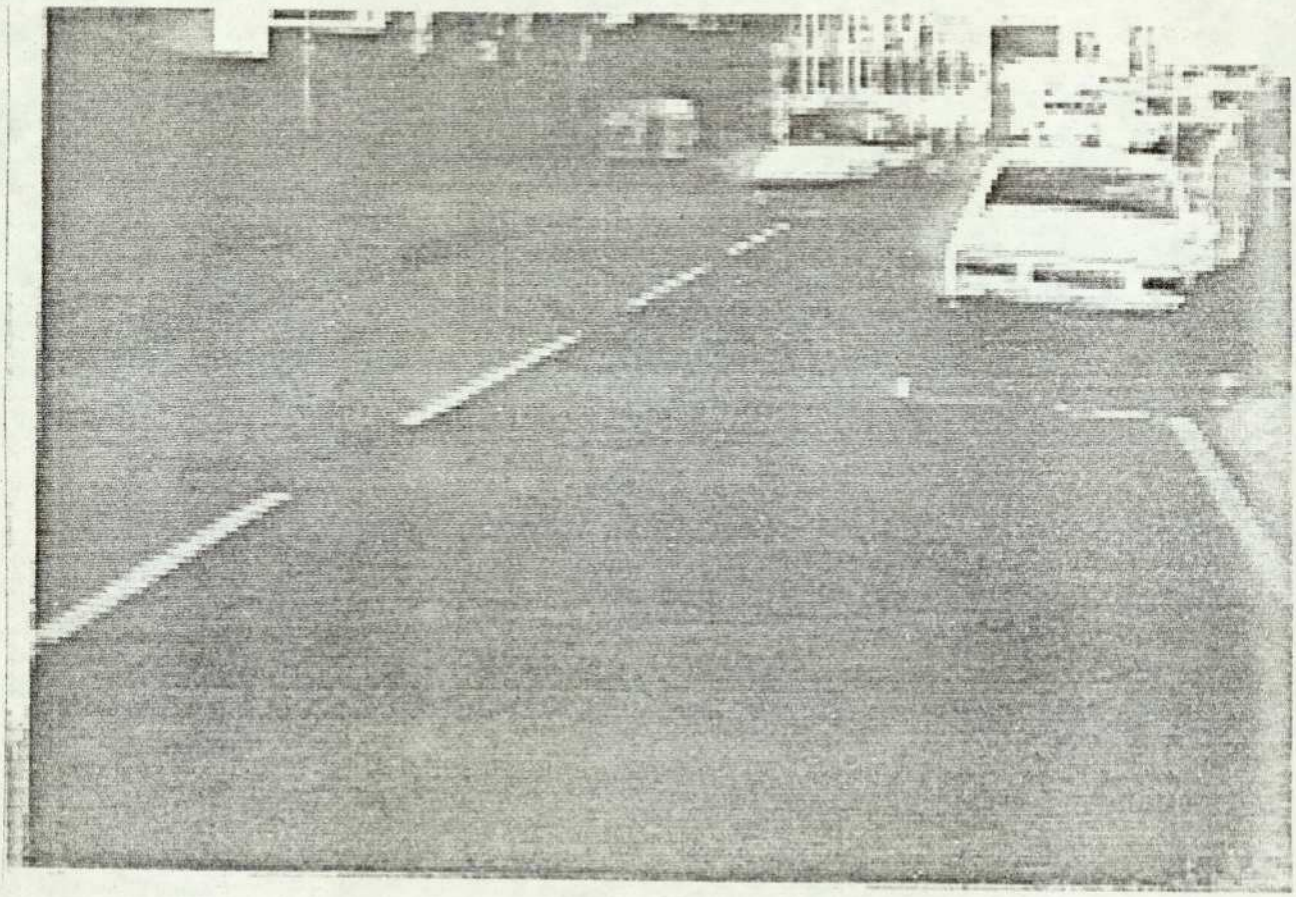


Figure 6.15 : Original stret scene image.

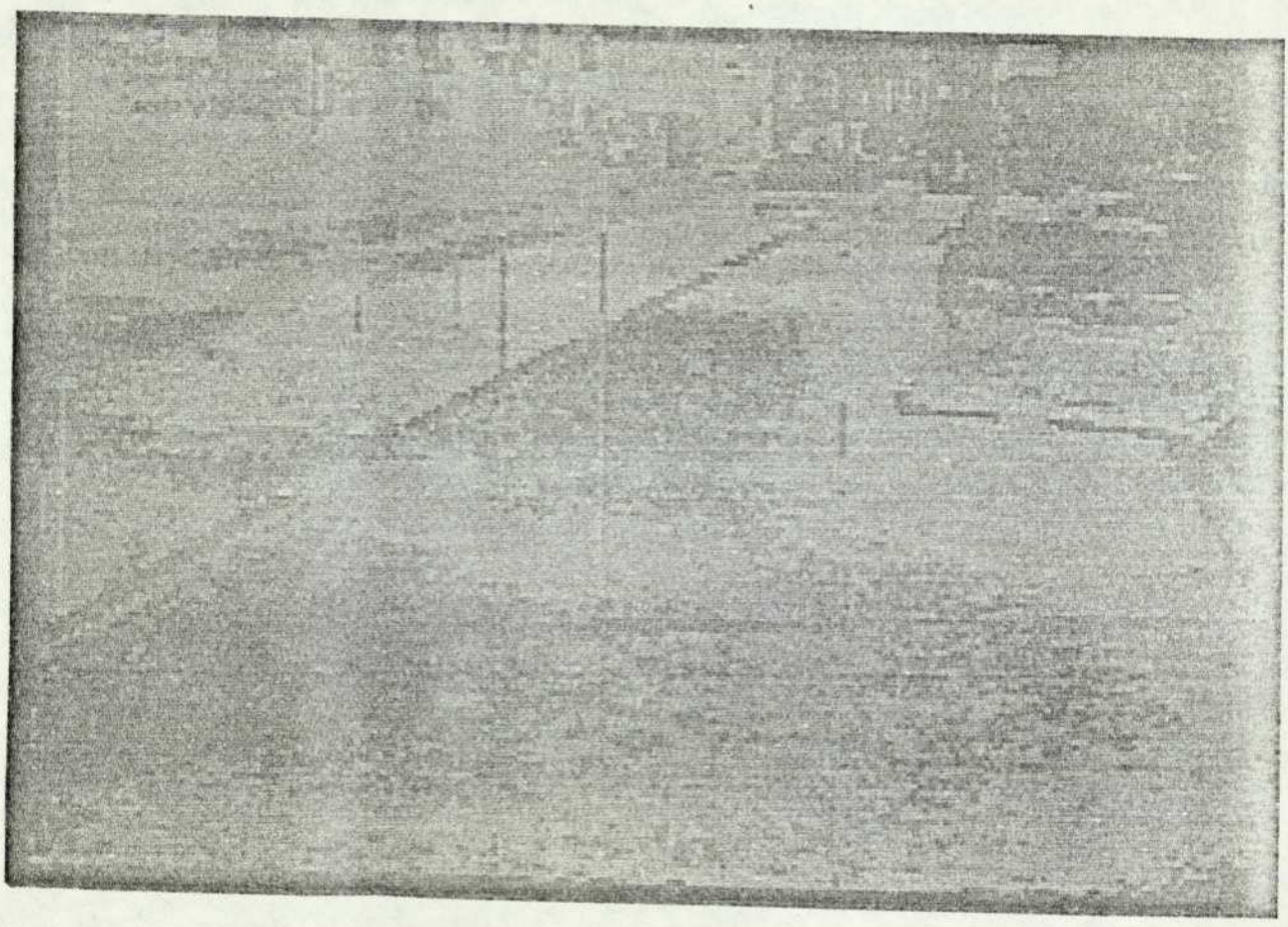


Figure 6.16 : Thresholded image

(starting grey value:37,range :20)

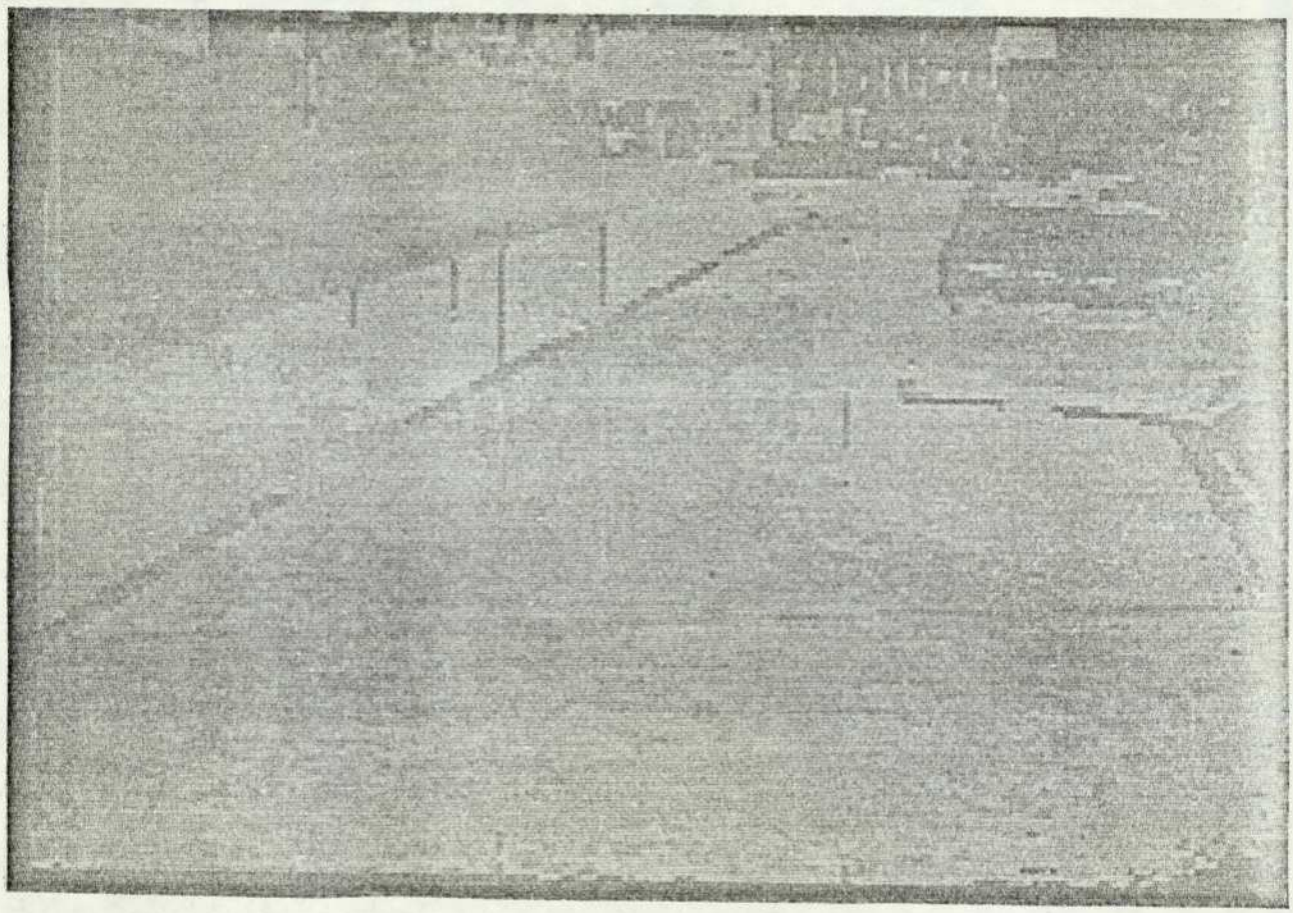


Figure 6.17 : Thresholded image

(starting grey value:17,range :40)

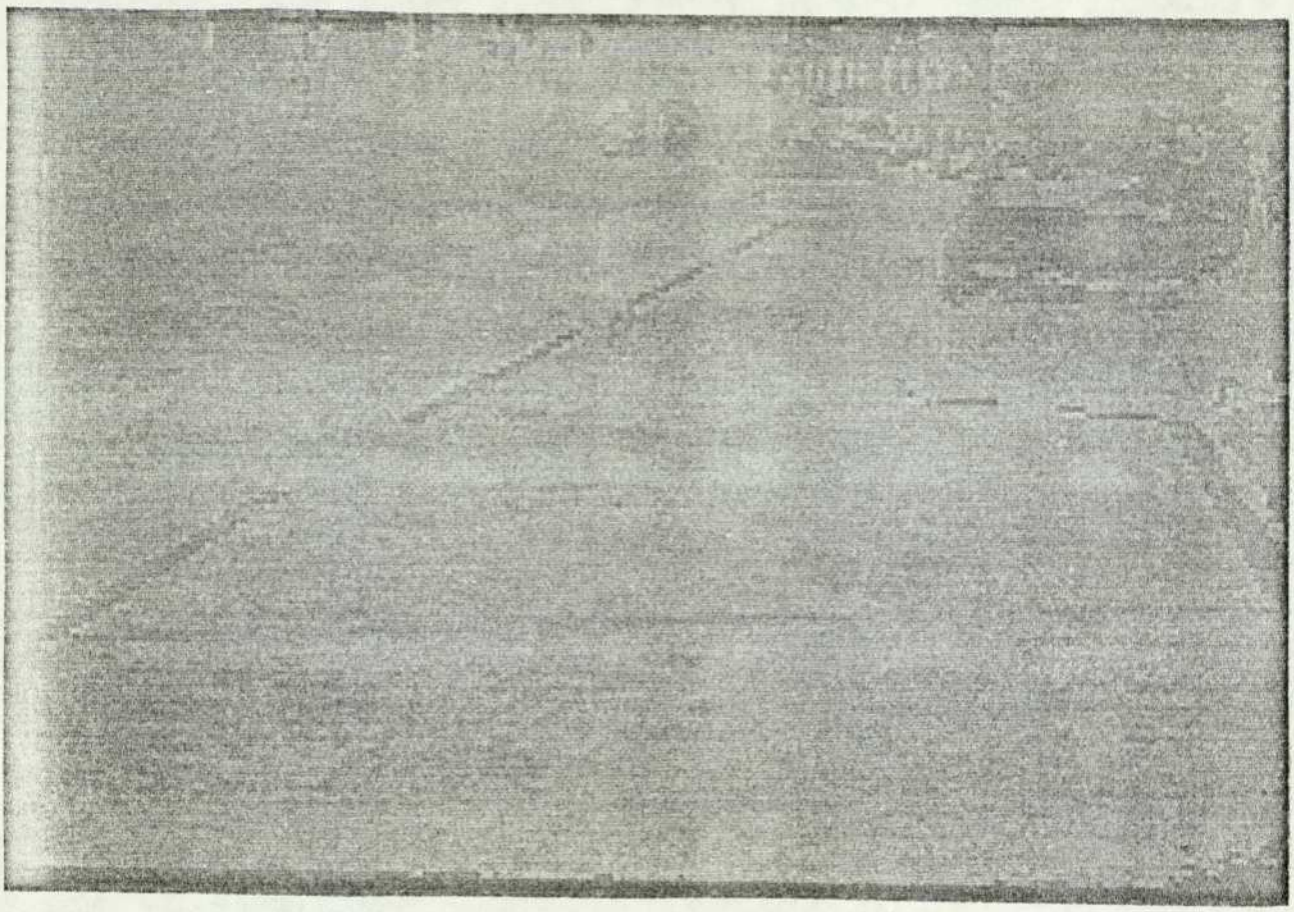


Figure 6.18 : Thresholded image

(starting grey value:7,range :60)

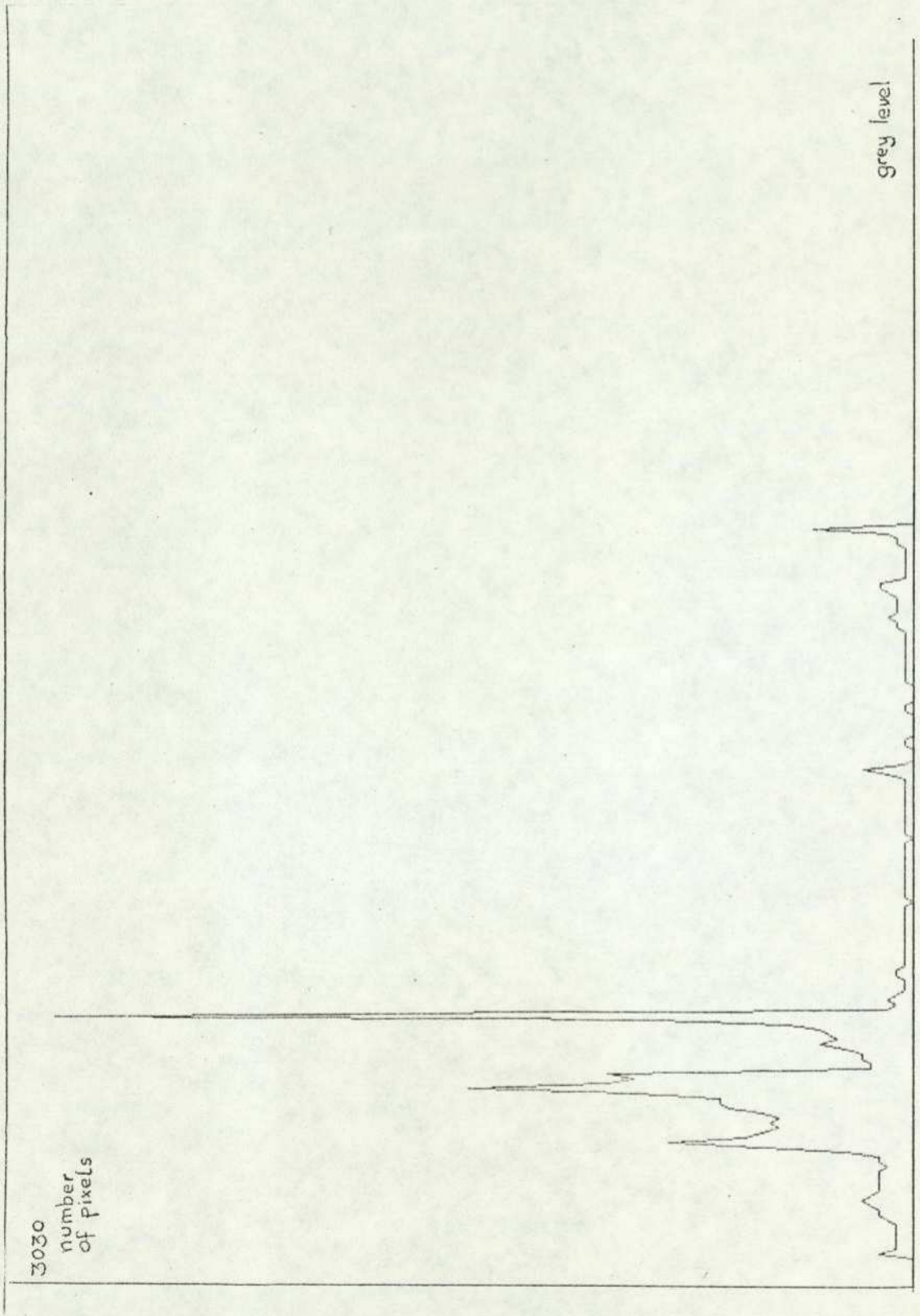


Figure 6.19:Histogram of image in fig. 6.15.

As the thresholding was only used as a preprocessing tool,and because it is acceptable to accept more information than is necessary,but not acceptable to loose necessary information,a range of width of 60 grey levels was chosen to characterise the street.

As a conclusion to thresholding,mention must be made of the work of Ohlander(1975) who succeded in segmenting a scene by the use of thresholding.Unfortunately he worked with colour images and we only work with black-and-white images. When working in a multidimensional feature space(colour images:3 colour,grey levels,and texture),thresholding can be used very effectively for segmentation purposes.But,when working in a unidimensional space(grey levels),it cannot be used on its own for segmentation of non-artificial images.Thresholding of black-and-white images is not a very powerful technique.But it should be said that in specific cases,objects in a uniform background,for example,it becomes a very powerful technique and could solve on its own the problem of segmentation.

6.5 Edge Detection Techniques

The importance of edge detection is illustrated by the fact that for human beings a drawing containing only lines conveys nearly as much information (with much less data) about the image as does the same grey scale image,and sometimes more information than a

blurred black-and-white image.For human beings edge detection plays a major role in visual perception.

In general,an object location in a grey scale image could be achieved by a comparaison with representative object masks (matching).Since however size,orientation and brightness of the objects may vary strongly,a large number of masks would be needed to identify even one object.Thus this approach would be exceedingly laborious.

It is important to reduce picture information,and to generate suitable invariant features.The most important features of an objects are its boundaries and the contour-lines which seperate its different regions,and not the grey values of the regions.Hence a first step in processing images for segmentation purposes would be to generate all contour-lines of the grey scale image.

There are many edge detection techniques,but none of them is perfect.Thus a selection and testing of these techniques was needed.In what follows,description of these techniques,their performance and the choice between them is discussed.

6.5.1 Simple Edge Detection Technique

Changes or discontinuities in an image attribute,such as brightness,are fundamentally important features of an image since they often provide an indication of the physical extent of the objects within the image.Edges could be presented as a ramp increase in image amplitude level,which is characterised by its height,the slope and the x coordinate of the slope midpoint.An edge is considered to exist if both the slope and height are larger than a critical value.This is important because a method for automatic thresholding has to be devised.In practical cases,only the threshold of the height matters.

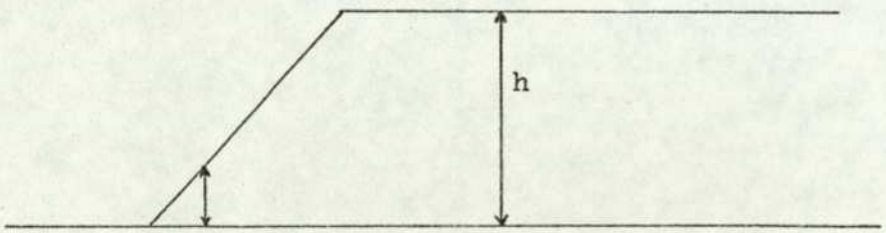


figure 6.20 : Edge in a unidimensional space.

The simplest way of representing the edges would be a binary matrix,where elements which belong to a boundary are marked by one and the other by 0. However the use of such a representation would not give any information about the magnitude of the

edge(boundary between two areas with slightly differing grey values would be represented in the same way as boundary between two areas with considerably differing grey values).Also there would be no information about the direction of the edge.To overcome these problems,a measure of the difference of the grey level on both sides of the contour element(magnitude of the edge) and a measure of the direction of the edge must be used.For calculating the magnitude of the edge,the first derivative of the image function could be used.The first derivative of an image may be accomplished using local operations,which must be performed in each image element.Because an image is not given as an analytic mathematical function but as a grid of discrete numbers forming a rectangular matrix,the first derivative may be evaluated by a diffence method.For this reason the neighbouring elements of an element (i,j) is summarised in a submatrix with (i,j) as the central element.

i - 1	i	i + 1	
$a_{i-1,j-1}$	$a_{i,j-1}$	$a_{i+1,j-1}$	j -1
$a_{i-1,j}$	$a_{i,j}$	$a_{i+1,j}$	j
$a_{i-1,j+1}$	$a_{i,j+1}$	$a_{i+1,j+1}$	j + 1

figure 6.21 : 3x3 mask.

The first derivative may be evaluated using the following equation:

$$\Delta F_{i,j} = \frac{1}{2}(\Delta F^{x}_{i,j} + F^{y}_{i,j})$$

$$\Delta F^{x}_{i,j} = |b_{i,j-1} - b_{i,j+1}|$$

$$\Delta F^{y}_{i,j} = |b_{i-1,j} - b_{i+1,j}|$$

This method has been tested with the five street scene images and gave satisfactory results (figure 6.22).But a more sophisticated method giving the magnitude and the direction was required if we wanted to design an edge detection system which would be powerful enough to help us resolve the problem of automatic driving.

6.5.2 Edge Detection Using Directional Masks

In this method,the edge detection system uses 3x3 masks,and edge angles are quantised to eight equally spaced directions as represented below: in figure 6.23.

There are many masks which could be used for this

Fig.6.22 Edge magnitude map using simple differentiation.

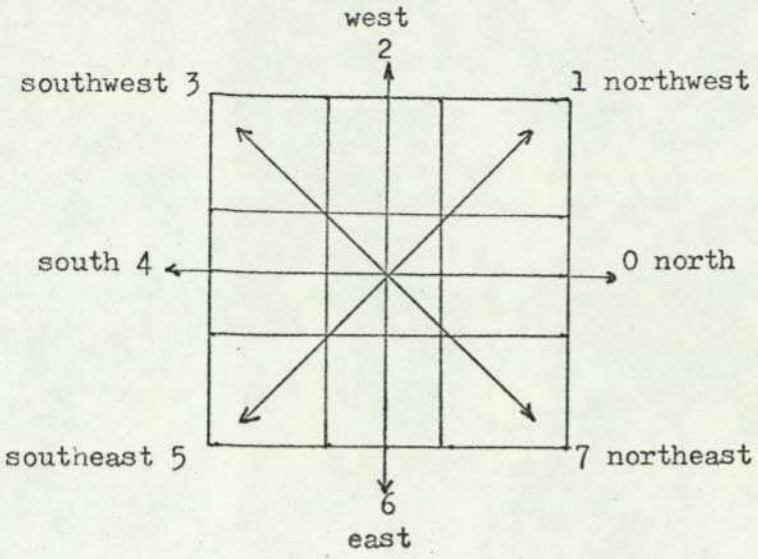


Figure 6.23 : Quantised Direction measure.

discrete differentiation can be performed by convolving a grey scale image with compass gradient masks.The compass name indicates the slope direction of maximum response.Edge directions corresponding to eight compass directions are determined in such a way that the bright side of the edge is always to the left as one moves in the direction of the edge.The number 0,1,...,7 were used for the eight principal directions in a 3x3 grid by Freeman in his chain coding scheme.Types of masks which could be used are the 8 Prewitt operators,8 Kirsh operators,8 three level simple masks or 4 five level simple masks(Pratt(1977), Prewitt(1970),Kirsh(1971)).Among these operators we will mainly concern ourselves with the five level simple masks.

the set of five-level simple directional masks give the edge magnitude of the image and the edge direction in a simple manner.These masks contain five integer weights between -2 and +2.The reason for our choice of these operators is that the masks in direction 0 and the mask in direction 2 approximate the

partial derivative in the x-direction and y-direction,That the 0 weight in the center of the masks suppress the jitter of the line and that only 4 masks are needed instead of 8 as in the case for kirsh and Prewitt operators.The structure and integer weights of the five-level mask make them especially suitable for fast digital computation of gradient magnitudes and directions.They are also very compatible with locally adaptive threshold.The four masks give the eight directions by using the sign of the convolution sum (for example if mask 0 yields the maximum response and is positive the direction is then 0,but if it is negative,then the direction is 4) are listed below:

These operators were tested and gave satisfactory results after visual inspection of the differentiated five images.It must also be reported that there was not any major noticeable difference between the edge magnitude obtained by this method and that obtained by the previous simple method described earlier. But,obviously,this method provides additional information about the direction of the edge,which the previous simple edge detection did not provide.

At this stage the system is not complete because we did not resolve the problem of the threshold and some improvement could be obtained by using a connectivity map ,which is going to be described in what follows.

				if positive	if negative
1	2	1			
0	0	0	direction	0	4
-1	-2	-1			
2	1	0			
1	0	-1	direction	1	5
0	-1	-2			
1	0	-1			
2	0	-1	direction	2	6
1	0	-1			
0	-1	-2			
1	0	-1	direction	3	7
-2	-1	0			

Figure 6.24 : Five-level Masks.

6.5.3 Edge Detection System

By applying the four five-level simple masks to a 3x3 grid surrounding a picture element,we obtain the gradient magnitude and direction.The gradient magnitude map is obtained by taking the maximum gradient magnitude at each point,given by the four masks.The masks which yields the maximum gradient value determines the direction of the edge,hence we get an edge direction map,which is a two-dimensional array of integers which range between 0 and 7.

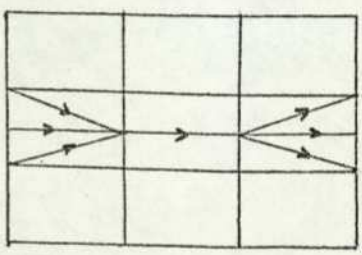
Using the edge direction map we can generate a local connectivity binary map as follows.If the direction at the centre of the 3x3 grid,(i,j),is k and if the direction of the preceeding and succeeding edge vectors are k-1,k,or k+1 mod(8) then the edges are connected.This procedure is illustrated below.If there is a connection a 1 is put in the connectivity map,and a 0 otherwise.

- direction k = 0

  If $k(i,j) = 0$ and $k_{(i-1,j)} \in \{0,1,7\}$

  $k_{(i+1,j)} \in \{0,1,7\}$

  then (i,j), (i - 1,j) and (i + 1,j) are connected.

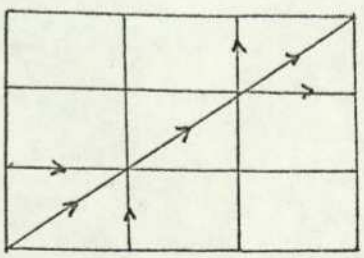


- direction k = 1

  If $k(i,j) = 1$ and $k_{(i-1,j+1)} \in \{0,1,2\}$

  $k_{(i+1,j-1)} \in \{0,1,2\}$

  then (i,j), (i - 1,j + 1), (i + 1,j - 1) are connected.

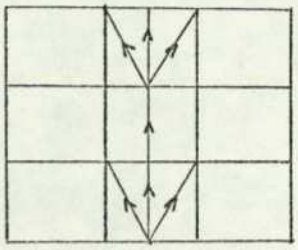


- direction $k = 2$

  if $k(i,j) = 2$ and $k(i,j-1) \in \{1,2,3\}$

  $k(i,j+1) \in \{1,2,3\}$

  then (i,j), $(i,j+1)$ and $(i,j-1)$ are connected and 1 is put in (i,j).

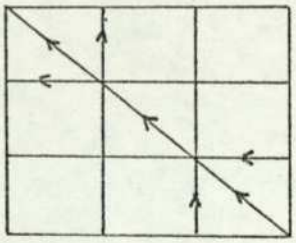


- direction $k = 3$

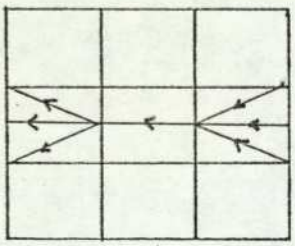
  if $k(i,j) = 3$ and $k(i-1,j-1) \in \{2,3,4\}$

  $k(i+1,j+1) \in \{2,3,4\}$

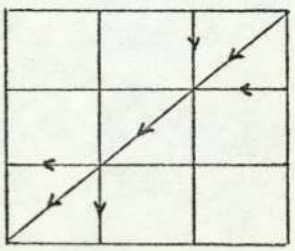
  then (i,j), $(i+1,j+1)$ and $(i-1,j-1)$ are connected.



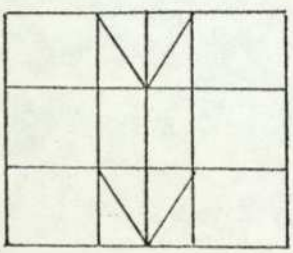
- direction $k = 4$



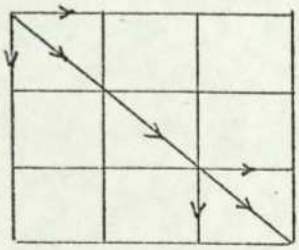
- direction $k = 5$



- direction $k = 6$



- direction $k = 7$



A problem,which we are faced with in edge detection,isthe choice of a suitable threshold.We experimented first with a fixed threshold.For the edge magnitudes going from 0 to 255,we used a threshold value equal to 40,every edge giving a magnitude above 40 was accepted as an edge.The fixed threshold worked for three images,but did not work for two others were we lost edges which were needed,and we had to lower the threshold to obtain them.From the gradient picture we got a histogram of an exponential form (figure 6.25).

Because the usage of a fixed threshold was a nuisance (it had to be changed for each image),it has been decided to devise an automatic method for the choice of the threshold.

For the choice of the threshold,the following assumptions were made. The magnitude of the edges,which form the lines of interest (borders of the street), belong to the biggest region of magnitude edges,the region having a range of 40 grey levels.Another assumption,which is justified by the exponential form of the histogram,is that the regions,which have a number of cell exeeding a certain limit must be rejected.The regions are determined in the same way ,which is described in the paragraph about thresholding. After a certain number of trials we determined the maximum number of cells ,which the region might have.For a range of 40 grey levels this limit was set to 35xNX,NX being the length of the image in pixels (number of image elements in a scan line).This was used simultaneously with fixed threshold

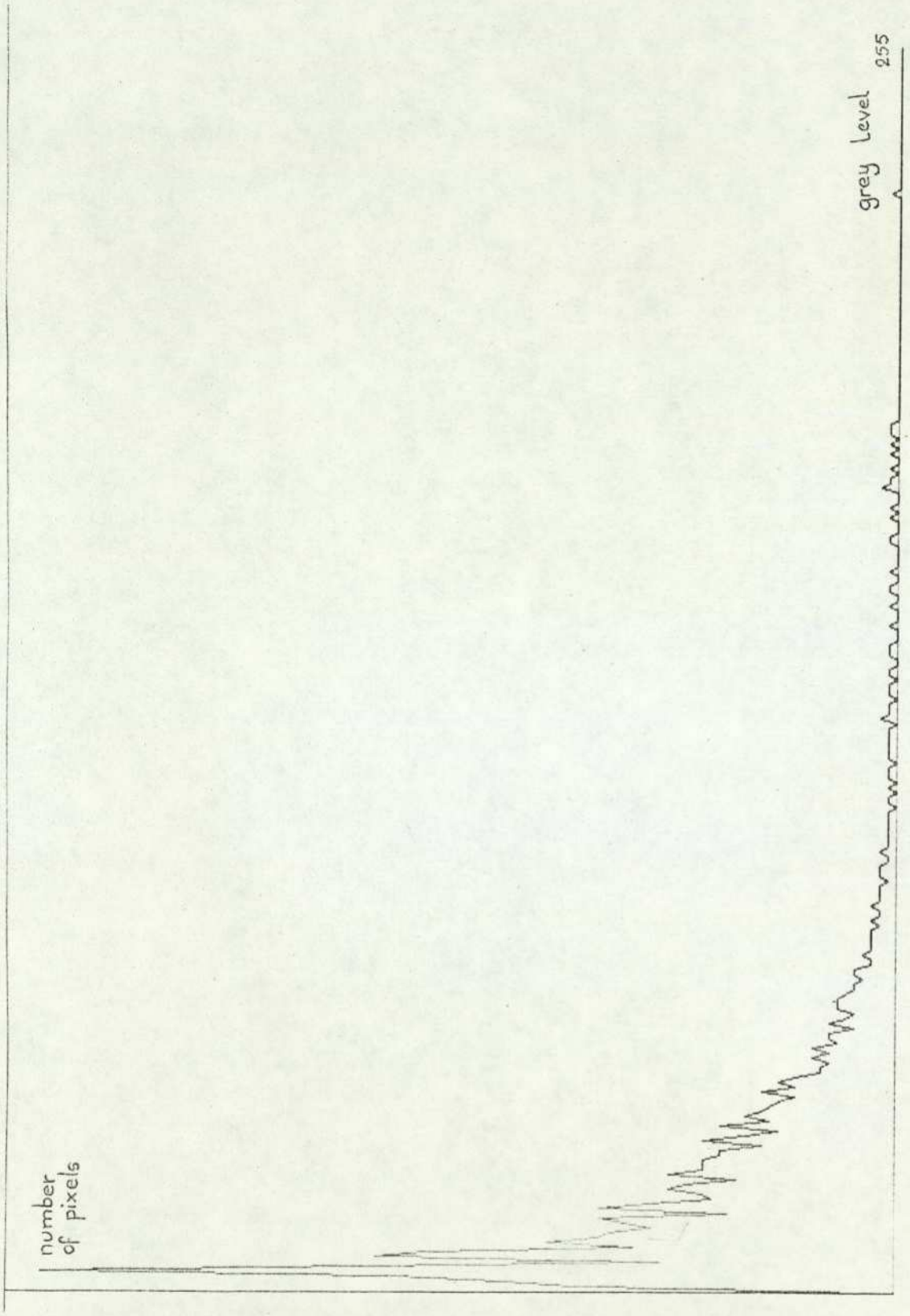


Figure 6.25:Histogram of the edge magnitude image.

of 12.

Some experiments were also caried out,with a locally adaptive threshold. With a fixed threshold,when the threshold is too low,too many edge points are obtained,and when the threshold is too high,then some significant edge points are lost.Some researchers (Pratt(1977),Rosenfeld(1969)) suggested that the use of local threshold can give noticeable improvement. So,it was decided to look at the problem and to test some of the local thresholds.

Using a edge activity index (EAI) defined as the ratio of the maximum gradient magnitude at an image point to the average magnitude of gradient in eight compass directions,improvement of the analog gradient image can be obtained.If the eight compass gradient values at a pixel (i,j) were y1,y2,...,y7,then EAI is defined as:

$$EAI = \frac{\text{Max}\left\{|y_k|, k = 0,1,\ldots,7\right\}}{\sqrt{\frac{1}{8}\sum_{k=0}^{7} y_k^2}} - 1$$

This expression becomes simpler for the simple mask since only the first four masks are enough to obtain gradients in all eight compass directions.If there exists no prefered orientations in the image then EAI is equal to 0,meaning there is no edge activity.The analog image gradient can be improved by imposing a

threshold on EAI.If EAI is greater than some threshold,i.e. if the edge activity is considerably superior in the direction of the maximum gradient,then the maximum gradient value is taken,otherwise the gradient value is set to zero.This operation results in a sharper histogram for the analog gradient image.

Another adaptive threshold,is the unsharp masking obtained by comparing the edge magnitude image with a blunted version of the original image,which was obtained by a low pass filter on the image.The blurred image is obtained by using the following mask,B:

$$B = \begin{array}{|c|c|c|} \hline 1 & 2 & 1 \\ \hline 2 & 4 & 2 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

A local thrshold measure is defined as follows:

$$L.T.M._{i,j} = \frac{\text{Mask}\{|y_k|,\ k=0,1,\ldots,7\}}{\text{Output of the low pass filter at pixel } (i,j)}$$

By setting a threshold on L.T.M we get an improved gradient image.

In our case,the improvement which was obtained was not enough to justify the computation time,which was needed for

calculating the local threshold measure.So,it was decided not to use the local threshold.

In summary the edge detection system which was devised could be represented by diagram 6.1 .

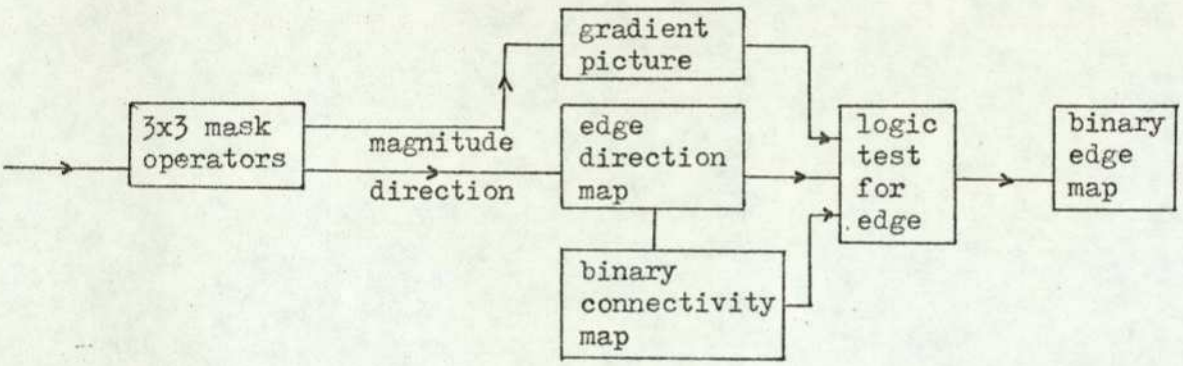


Diagram 6.1 : Edge detection System.

Another important improvement,which was carried out on this system was that ,as we are concerned with the extraction of the borders of the street,the quantisation of the direction into the 8 compass directions was dropped and instead we employed a technique to get an 'analog' direction measure.For this end,the following two orthogonal masks which approximate the partial derivative in the x-direction and y-direction were used.

$X =$

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

$Y =$

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

Figure 6.26 : Orthogonal masks.

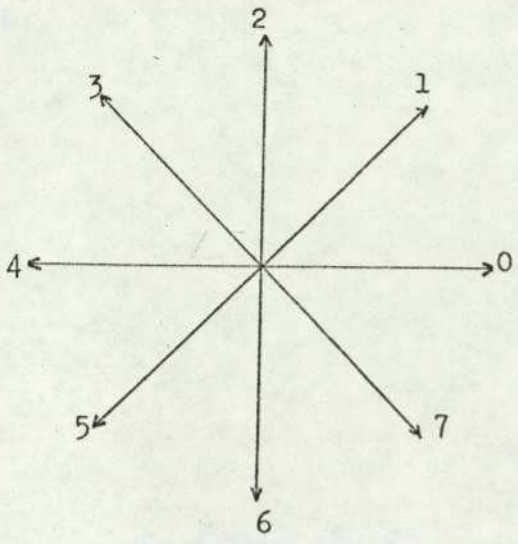
The edge magnitude M was obtained by taking the magnitude of the two orthogonal mask outputs.

$$M = \sqrt{X^2 + Y^2}$$

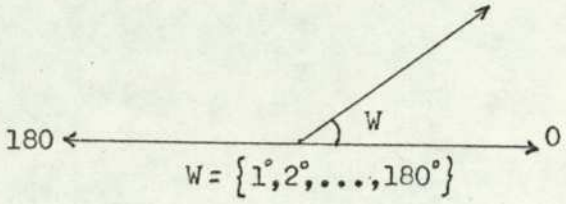
The direction of the edge was obtained by an arctangent operation on the two orthogonal masks.

$$W = \text{Arctan } \frac{Y}{X}$$

This method was tested and gave more information about the direction.The direction was quantised into 180 elements instead of the 8 for the previous method.



previous method



above method

6.6 Hough Transform

The system,as it stands,was still not adequate for street boundary detection because there was too much noise(redundant picture elements) in the image. As a final filtering operation for finding the street bounndaries the Hough Transform (Hough(1962)) was used.This takes each boundary point,(x,y),and expresses it in parametrical form:

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

Assigning the point as belonging to a straight line whose normal distance from the origin is R and for which the normal makes an angle W with the positive x-axis.Evidently,a whole family of straight lines can pass through a given point, (x,y).The objective

of the transform is,given a sequence of points (x1,y1), (x2,y2),...,(xn,yn),to find straight lines which pass through at least k of the points.

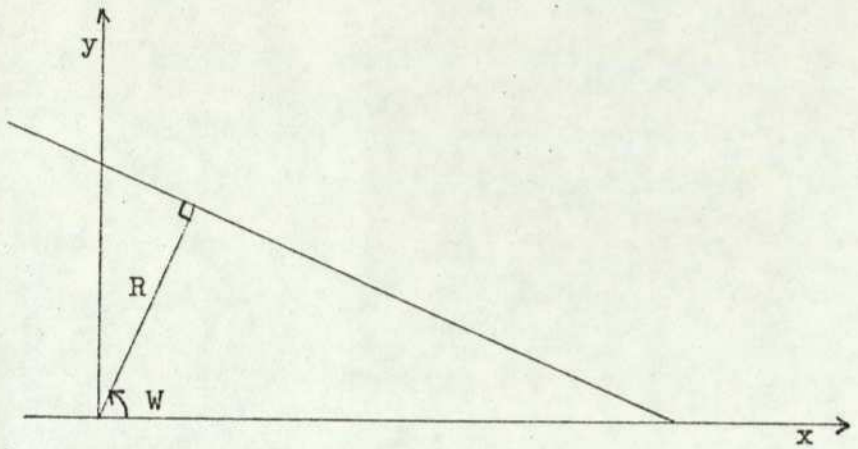


Figure 6.27 : Parametrisation of a line.

The Hough transform of a line is then a point at coordinates (R,W) in the polar domain.

We restrict W between 0 and 2 right angles,and let R be positive for y positive and R negative for y negative.In this case every line in the x-y plane is represented by a unique point in the W-R polar plain.

Every point,(xi,yi),of the x-y plane is transformed into a sinusoidal curve in the W-R plane defined by:

$$R = xi.\cos W + yi.\sin W$$

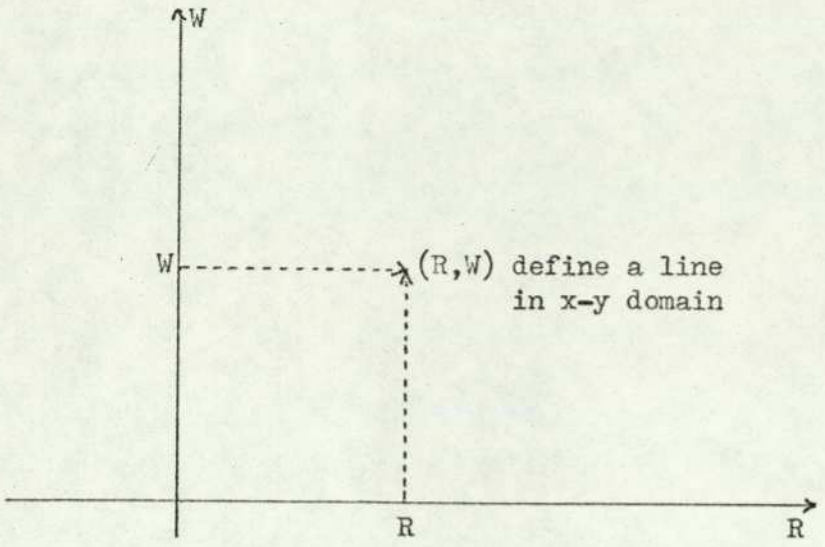


Figure 6.28 : Lines in a polar domain .

The sinusoidal curves of colinear points have a common point of intersection. The point of intersection defines the line passing through the colinear points.

Using this technique,we effectively look in a given direction and count the number of points in that direction.And so we built up a table giving the number of points for each given W and R.The counters W-R with big values are accepted as lines,and the others discarded.

Depending on the desired accuracy,the plan W-R is quantised into a quadruled grid. The quantisation is confined to the region $0<W<II$ and $-Rmax<r<Rmax$ where Rmax is the largest distance from the origin,O,in the x-y plane,to any point in the image.This is because points ouside this range correspond to lines,which cannot exist in

the image.

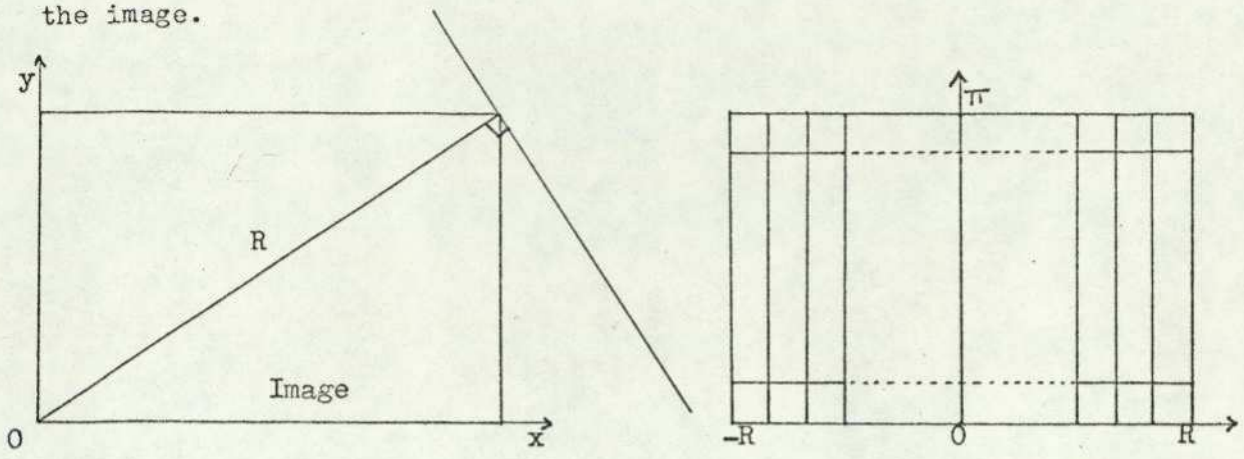


Figure 6.29 : Illustration of the Hough Transform.

The quantised region is treated as a two dimensional array of point counters. For each point,(xi,yi),in the picture plane,the corresponding curve given by

$$R = xi.\cos W + yi.\sin W$$

is entered in the array by incrementing the count in each cell along the curve. Thus a given cell in the two-dimensional array registers the total number of curves passing through it.After all edge points have been treated,the array is inspected to find cells having high counts.If the count in a given cell, (W,R), is k,then precisely k edge points lie(to within quantisation errors) along the line whose normal parameters are (W,R).

6.7 Complete System For Locating The Street In The Scene

6.7.1 System For Street Location

The system starts by generating from the grey scale image of resolution 1728x1120 the same image with smaller resolution which is 216x140.This involves 2x2 averaging of the original image,so that the original image of resolution 1728x1120 is successively reduced to 864x560,432x280 and finally 216x140.This image can either be reproduced in film using the Pic-Pac,available at the University of London Computer Centre,or be reproduced with a line printer using an overprinting routine,which I have developed.

From the grey scale image of reduced resolution (216x140),the system generate three maps which are:

1. An edge direction map with the value from 0 to 8 or from 0 to 180 (Figures 6.30 and 6.31).

2. A binary connectivity map with 1 for connected edges and 0 for non-connected edges (figure 6.32).

3. An edge magnitude map with values going from 0 to 255.the edge magnitude could be obtained by three methods:

a. A simple mask (figure 6.22)

$$\begin{matrix} 0 & 1/2 & 0 \\ 1/2 & 0 & -1/2 \\ 0 & -1/2 & 0 \end{matrix}$$

b. 4 five-level masks (figure 6.33)

c. Two orthogonal masks (described previously)

The generation of these three maps could be done for any resolution. Hence,at the end of this process we have at our disposal 4 two-dimensional array which are:

1. A gray scale image,array of dimension 216x140

2. An edge magnitude map,array of dimension 214x138

3. An edge direction map,array of dimension 214x138

4. A connectivity map,binary array of dimension 214x138

Using these four two-dimensional arrays,we attempt to extract the border of the street,rejecting everything else.For achieving that,we carried out the following steps illustrated by diagram 6.2 .

We take first the edge magnitude array and threshold it as described previously.If a cell is inside the threshold range we leave it as it is,if not a 0 is put instead of the previous

ig. 6.30Direction map using four masks to obtain the 8 compass direction (the direction measure is codified as light intensity in the image).

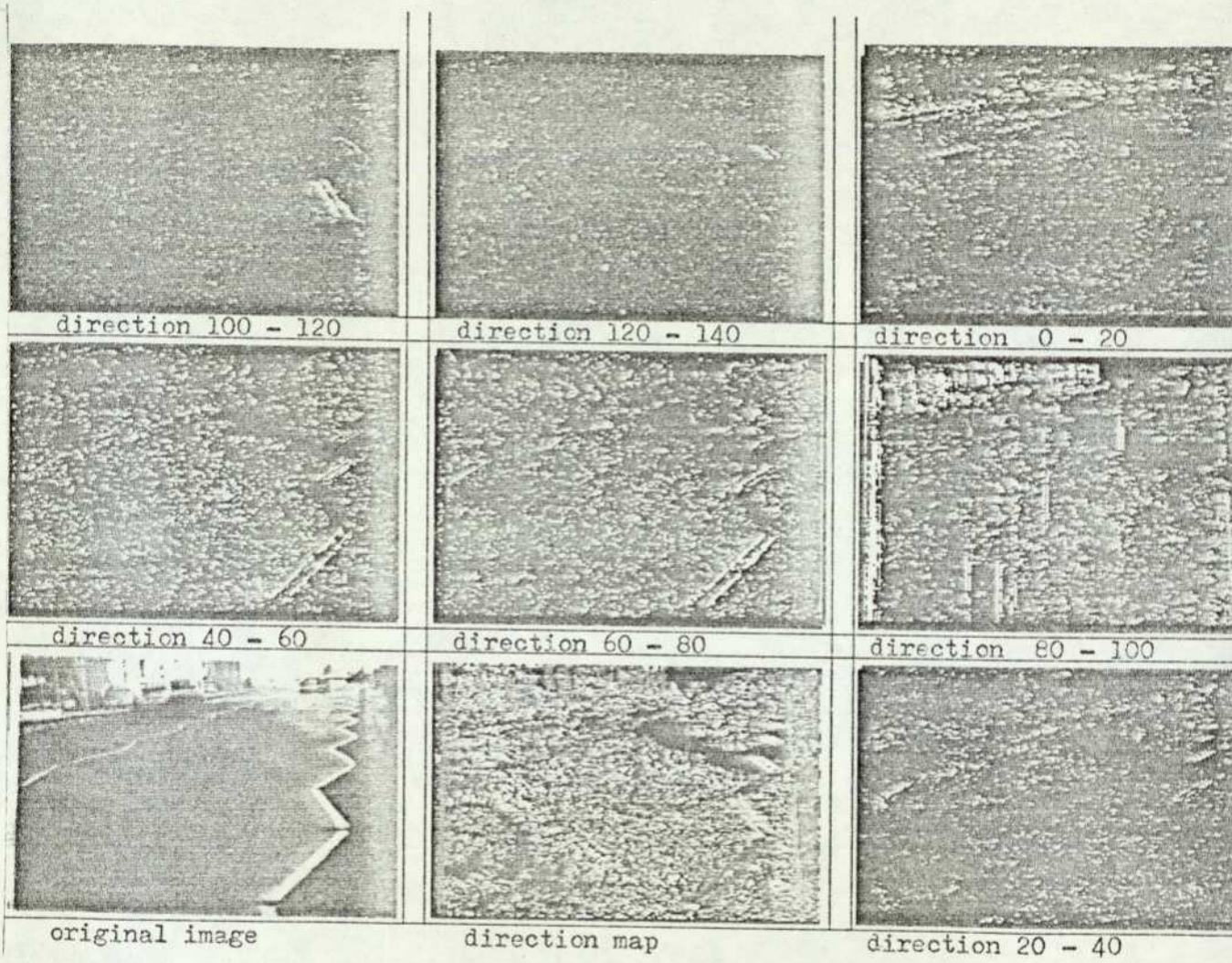


Figure 6.31 : Edge direction map (values from 0 to 180)

thresholded at different values.

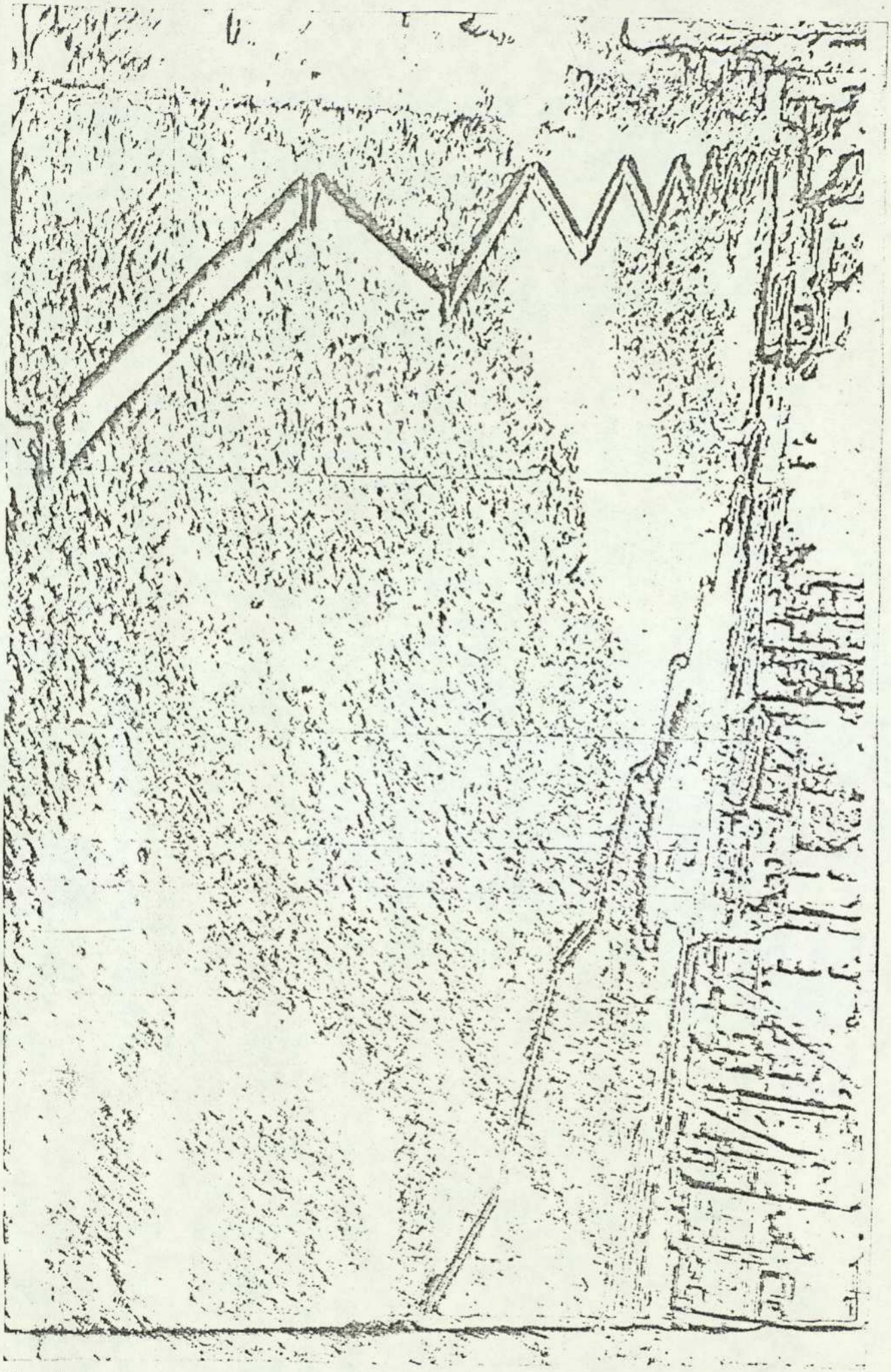


Fig. 6.32 Connectivity map.

ig.6.33 Edge magnitude map using 4 masks.

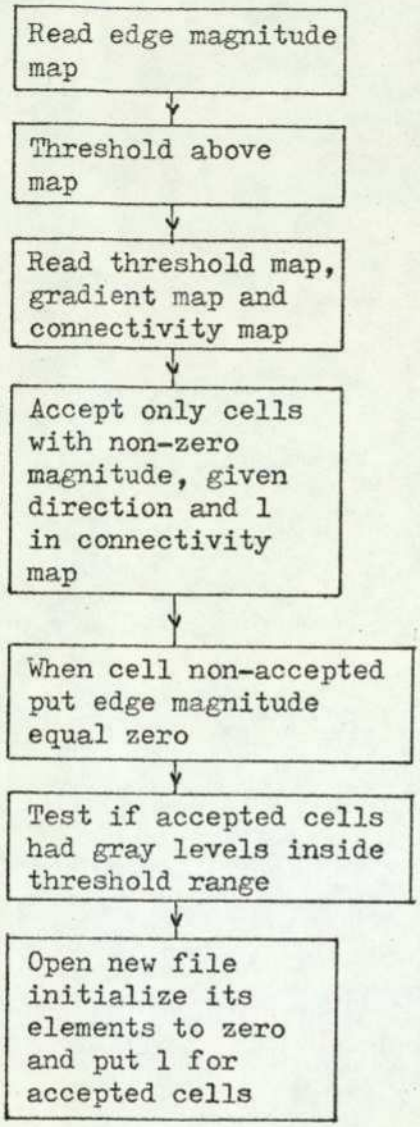


Diagram 6.2: Algorithm for street location.

file and the connectivity file,which is binary,are read simultaneously and every cell which does not have either a non-zero magnitude or a given direction and which is connected (1 in connectivity file),is rejected.For every rejected cell we change its magnitude to zero.The direction,which are accepted,are as follows: 1,2,3 and 5,6,7.

This is justified by the fact that the kinds of street scene,which are analysed, cannot have their borders horizontal.

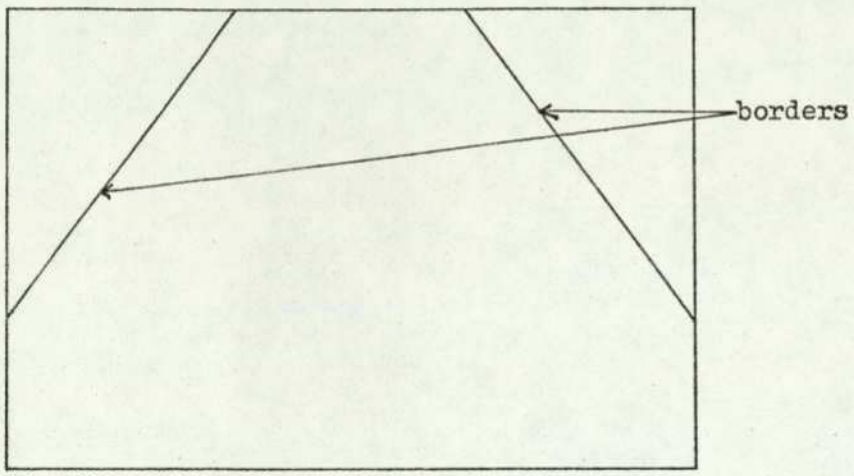


Figure 6.34. : Border of the street.

Finally,we read the grey level file and we test if the accepted cells had grey levels inside the threshold range.We open a new file and put 1 for every accepted cell and 0 for non accepted cells.

At the end of the procedure,described by the above diagram,we will obtain a binary file with 1 for cells,having satisfied the tests.Usually this file will contain the border of the street but also zigzag lines and yellow lines which are inside the street and some noise.but the border lines are generally the most important lines and applying the Hough transform to this image,the borders of interest can be extracted.The W-R plane was quantised as follows:

$0 < R < 275$

and $0 < W < 70$ degrees

and $100 < W < 120$ degrees.

This quantisation is justified,as illustrated by the following model,by the kind of images which were analysed (figure 1.4).

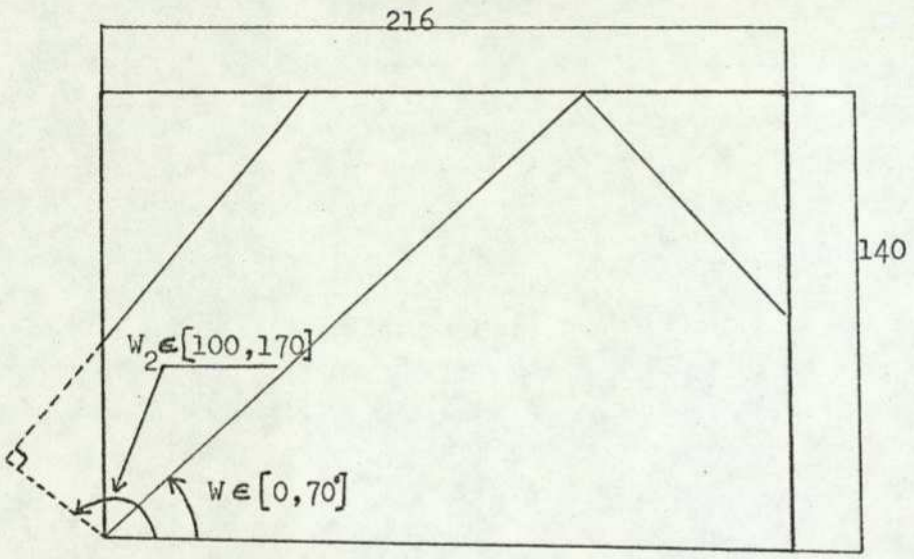


Figure 6.35: Illustration of the quantisation of the polar plane.

Using the system as described above we succeded in locating the street boundaries of three images.One of the reason why the system failed for the two other images,is the presence of zigzag and yellow lines,which were more important than the street borders.

Having located the boundaries of the street,we then read the

grey scale image and fill the outside of the street with zero and the inside with the original grey level values.

Because the system did not work for two images,many improvement of the system have been carried out,to try and make it work for the remaining images.The first of these improvements is the use of the direction measures which are more accurate.Before we were using 8 compass direction,but the new system uses direction given in degrees going from 0 to 180.With this information we could detect more effectively the lines of interest.

Another improvement was the implementation of the following test.Before applying the Hough transform,we tried to improve the image by rejecting the zigzag and yellow lines.The test consists in taking every non-zero cell of the final file,reading the grey scale image and taking the ten nearest cells which lie horizontally to the left and the right of the tested element,and counting the number of cells inside the threshold range to the left (LC).and to th right (LR).If LC and LR exceed a certain threshold value the cell is rejected.The test is based on the fact that the zigzag and yellow lines will have more cells,in the proximity,which lies in the threshold range,than the border of the street.

Unfortunately this involved the use of threshold values.LC and LR had to be thresholded and the determination of this

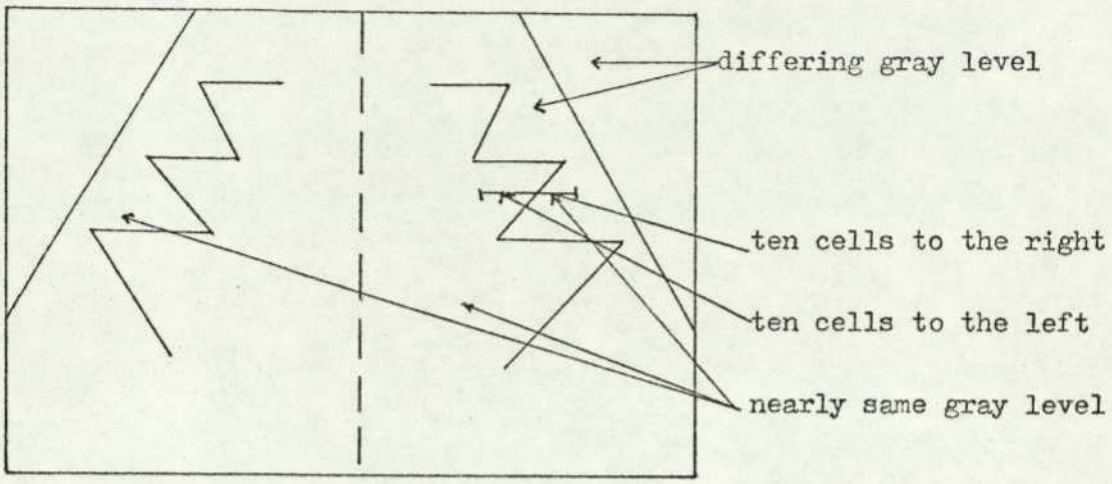


Figure 6.36 : Street Scene image with zigzag lines .

threshold could not be achieved by trial and error.

Another attemp was to use the Hough transform and to determine instead of one border the most important fifteen lines in the image.Then,calculating for each line some features,which,hopefully,will discriminate between the borders and the inside lines.The features,which were chosen,were:

1. Number of points in the line

2. Number of points in proximity of the line (range:ten to the left and ten to the right) which were inside the threshold value.

3. Total difference in grey level between successive points in the line.

Feature 2 was designed to reject zigzag and yellow lines,and feature 3 was designed to reject lines due to noise points.If the line is the border of the street it will have low values for feature 2 and feature 3.After many trials,it has been decided to abondon this strategy because the features used did not yield suitable results.

As it stands the system,whose performance is illustrated by the series of figures 6.38,6.39,6.40,6.41,6.42, and6.43,can be summarised by the diagram in figure 6.37.

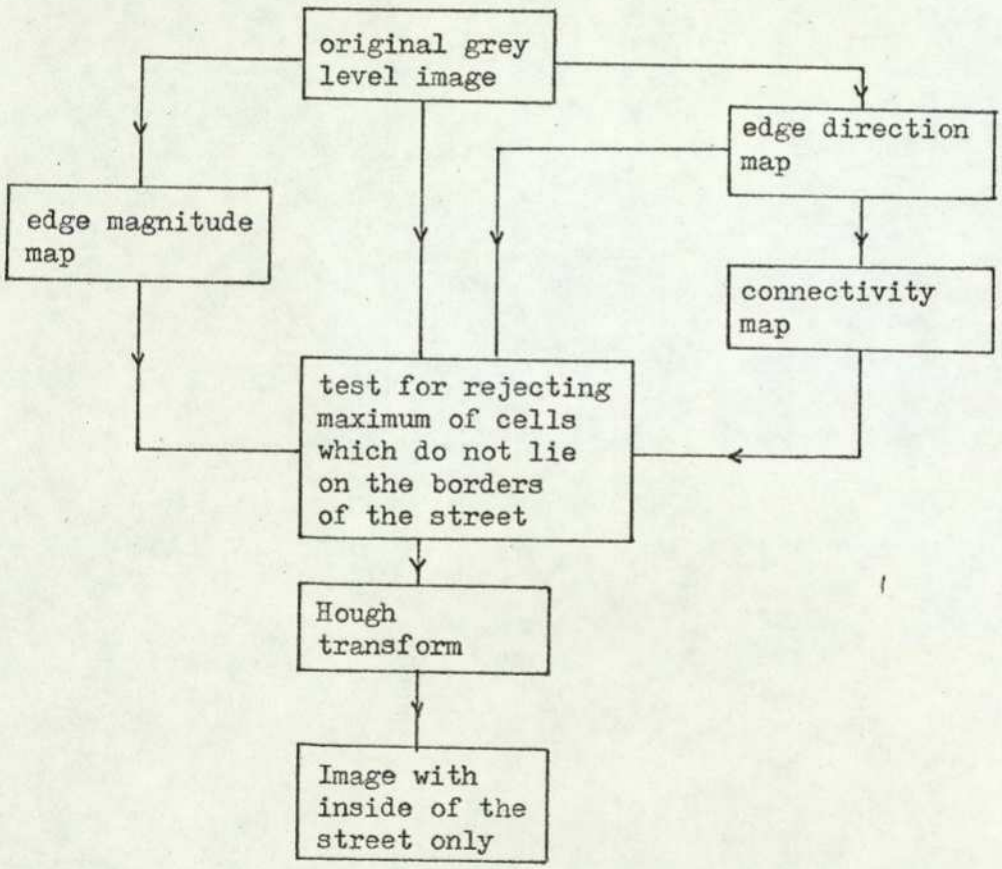


Figure 6.37 : System for street location.

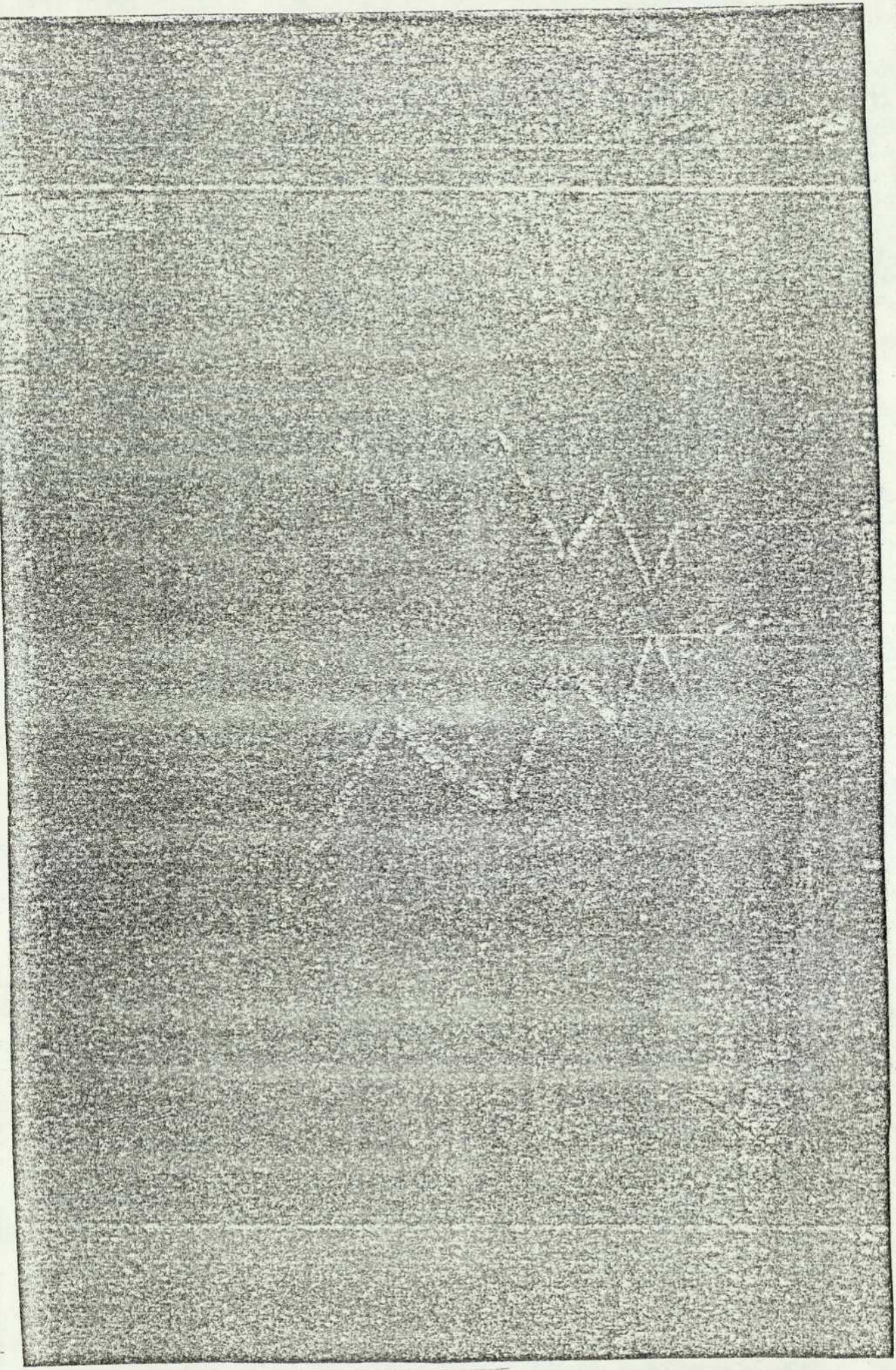
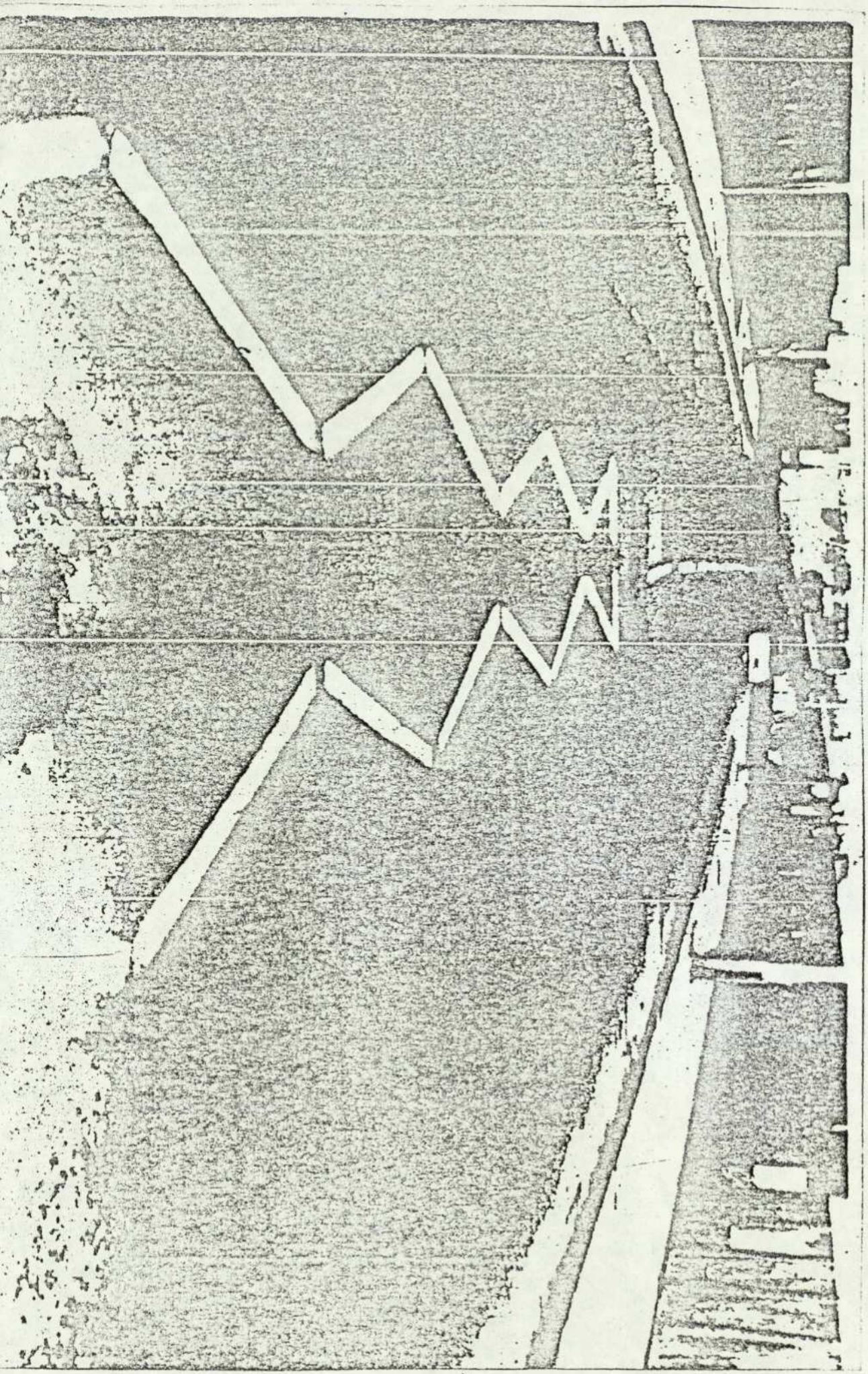


Fig. 6.38 Street scene with poor contrast.



ig. 6.39 Street scene image.

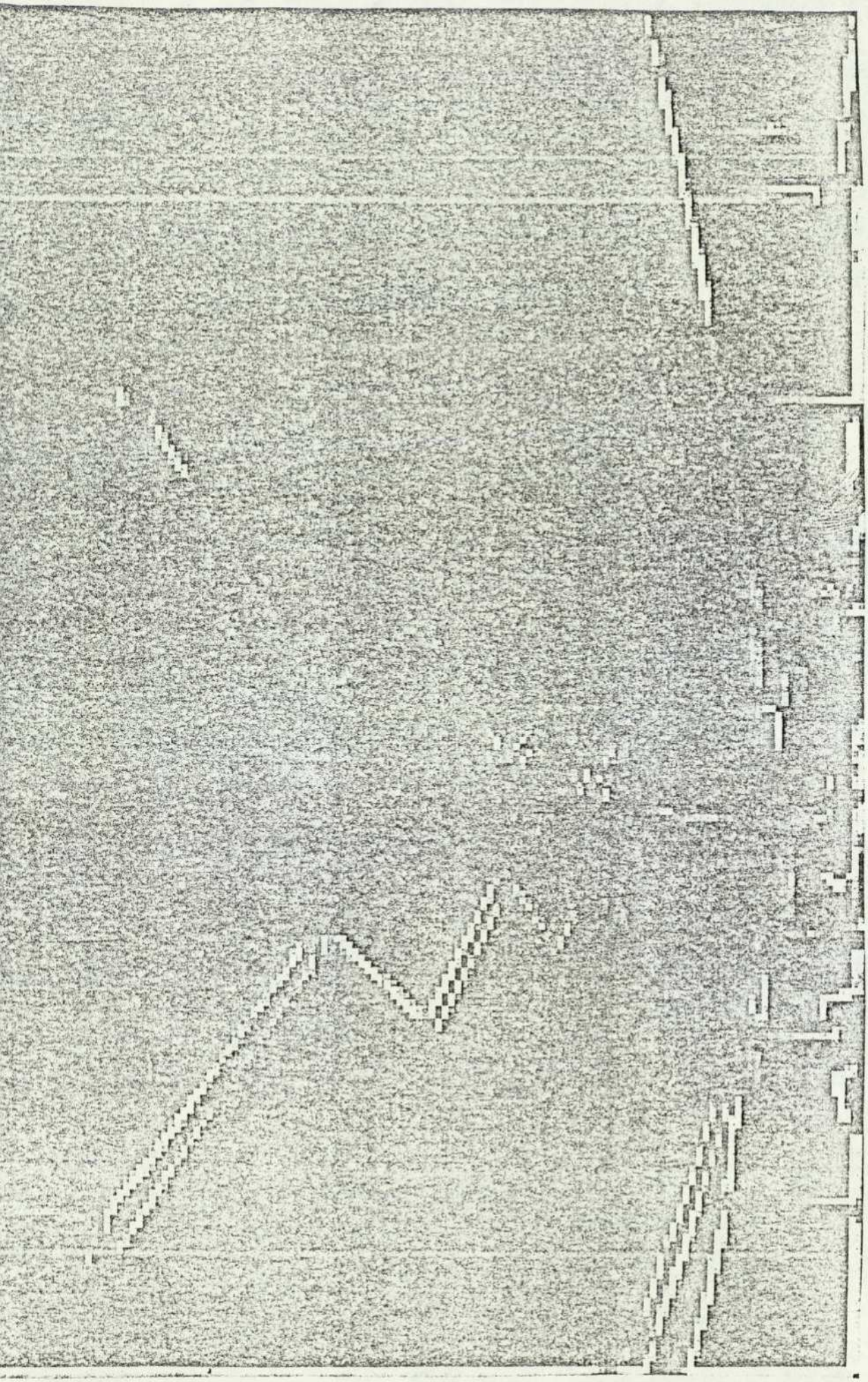
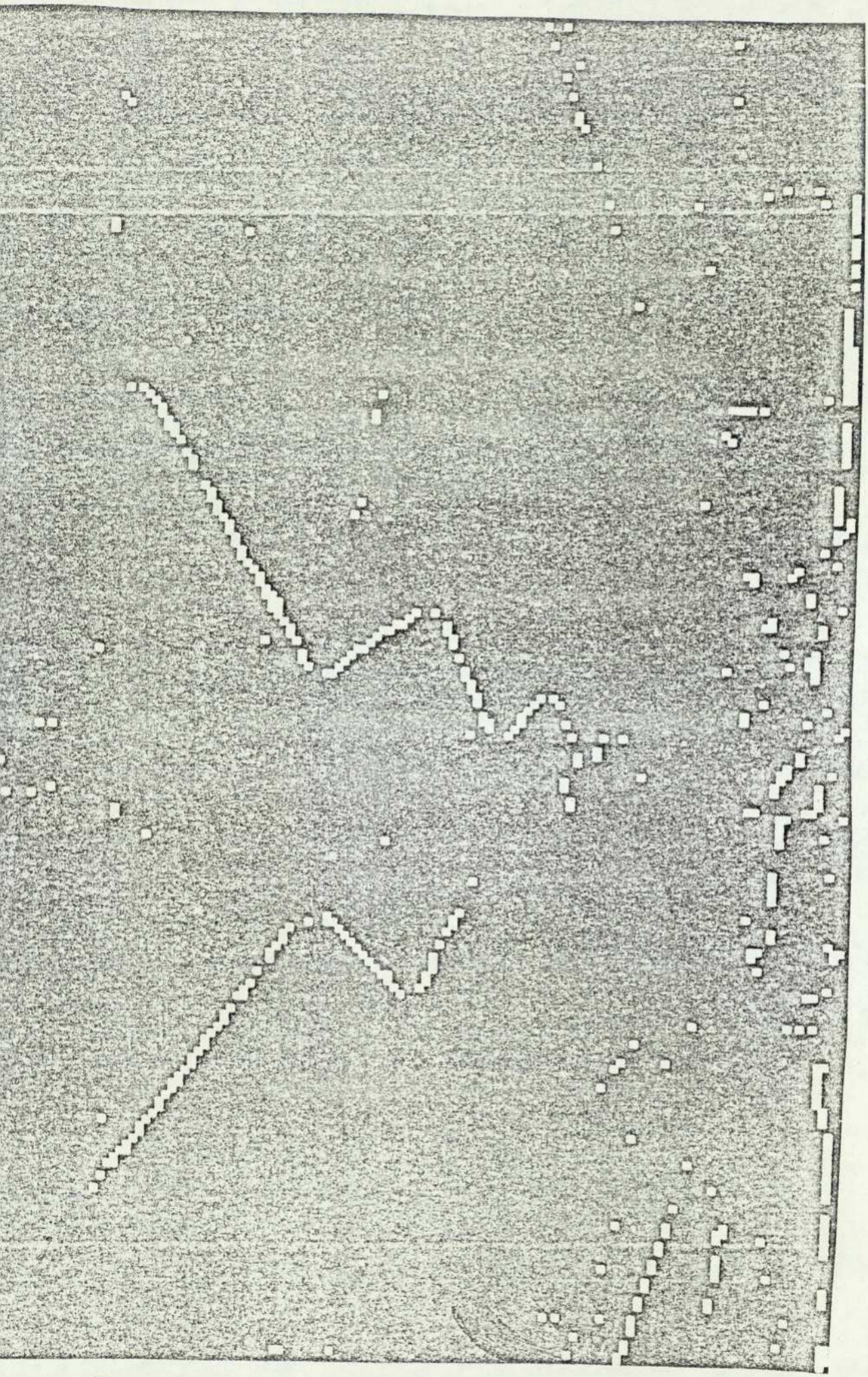
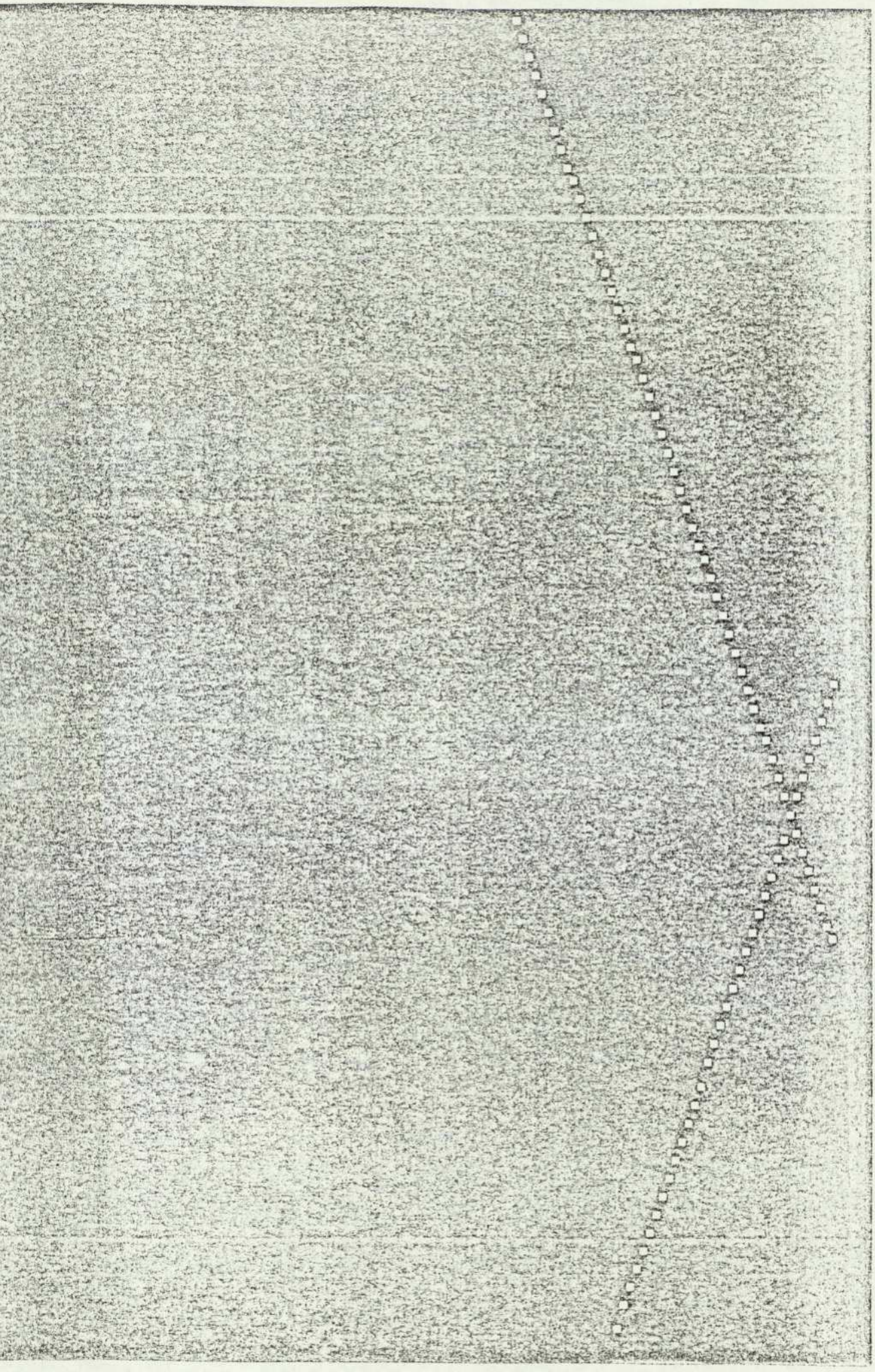


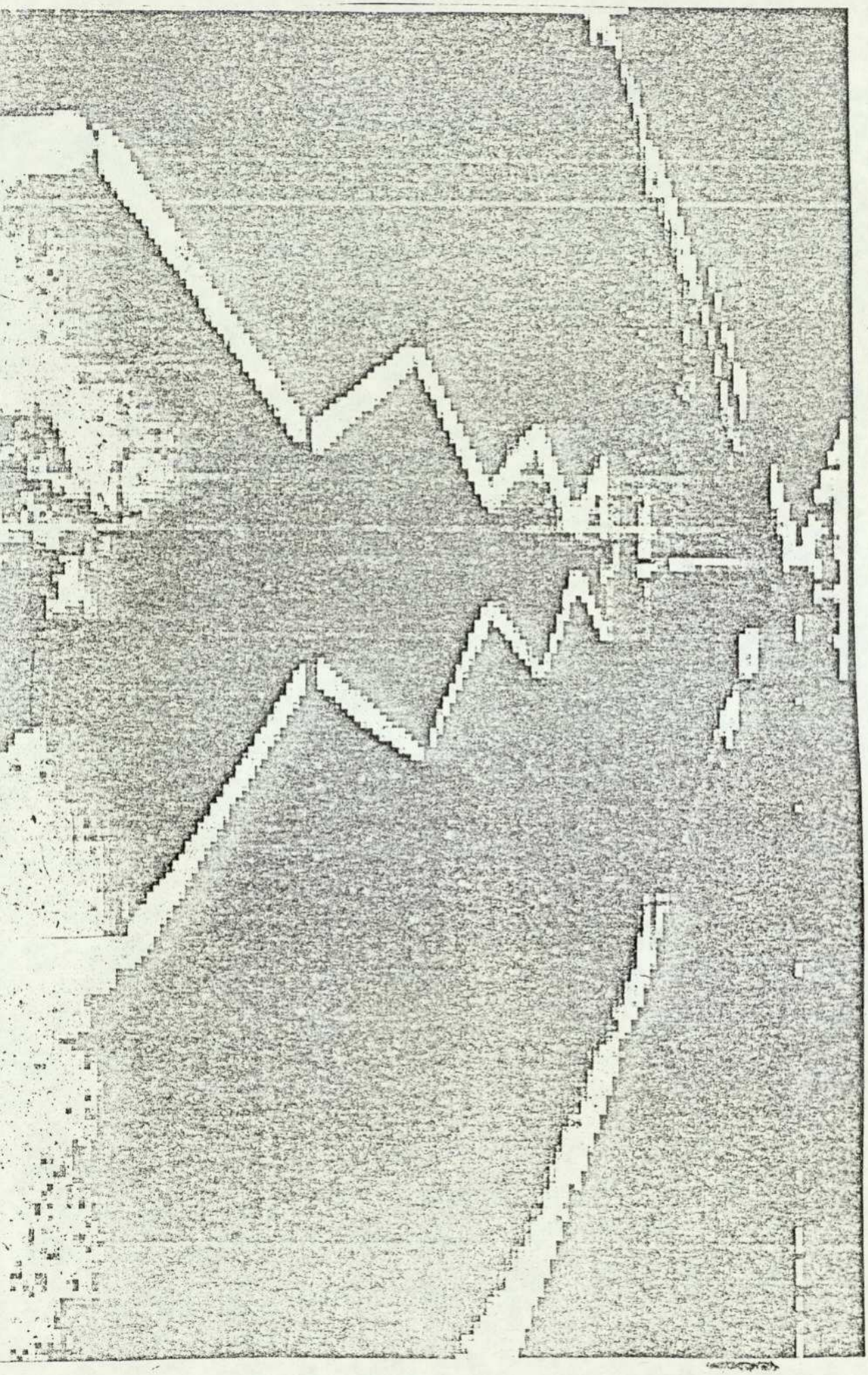
Fig. 6.40 Edge magnitude map using four masks.



.g. 6.41 Image obtained by combining the 3 maps and thresholding.



g. 6.42 Hough transform applied to image in fig. 6.41



ig. 6.43 Inside of street scene obtained by the complete procedure (see fig. 14,15,16,1

6.7.2 Improvement in the system

The previous system,using simultaneously an edge magnitude map,an edge direction map and a connectivity map,and different kinds of tests and the Hough transform,was successful only for three among five images,hence steps for improve the performance of the system had to be taken.

one of the major reasons why the previous system did not succed in locating the street in two of the five images,was the presence of yellow and zigzag lines which were mistakingly taken for street boundaries.Because of this,it was decided to remove the zigzag and yellow lines before starting the processing. The removal of these lines was achieved by locating regions in the image,which have high grey level values.The processing consisted in locating white regions which have a given spatial width,and replacing them by the average of the grey level values of their surrounding elements (Figure 6.44 and 6.45)

For locating the street in the two remaining images,we modified slightly the previous system.The new system is still based on edge detection and Hough transform.The first step in the system consists of the removal of the zigzag and yellow lines as described above (figure 6.45).The second step consists of computing the edge magnitude and edge direction maps using two orthogonal masks and thresholding the two maps as was carried out in the previous system (Figures 6.46, 6.47, 6.48 and 6.49)

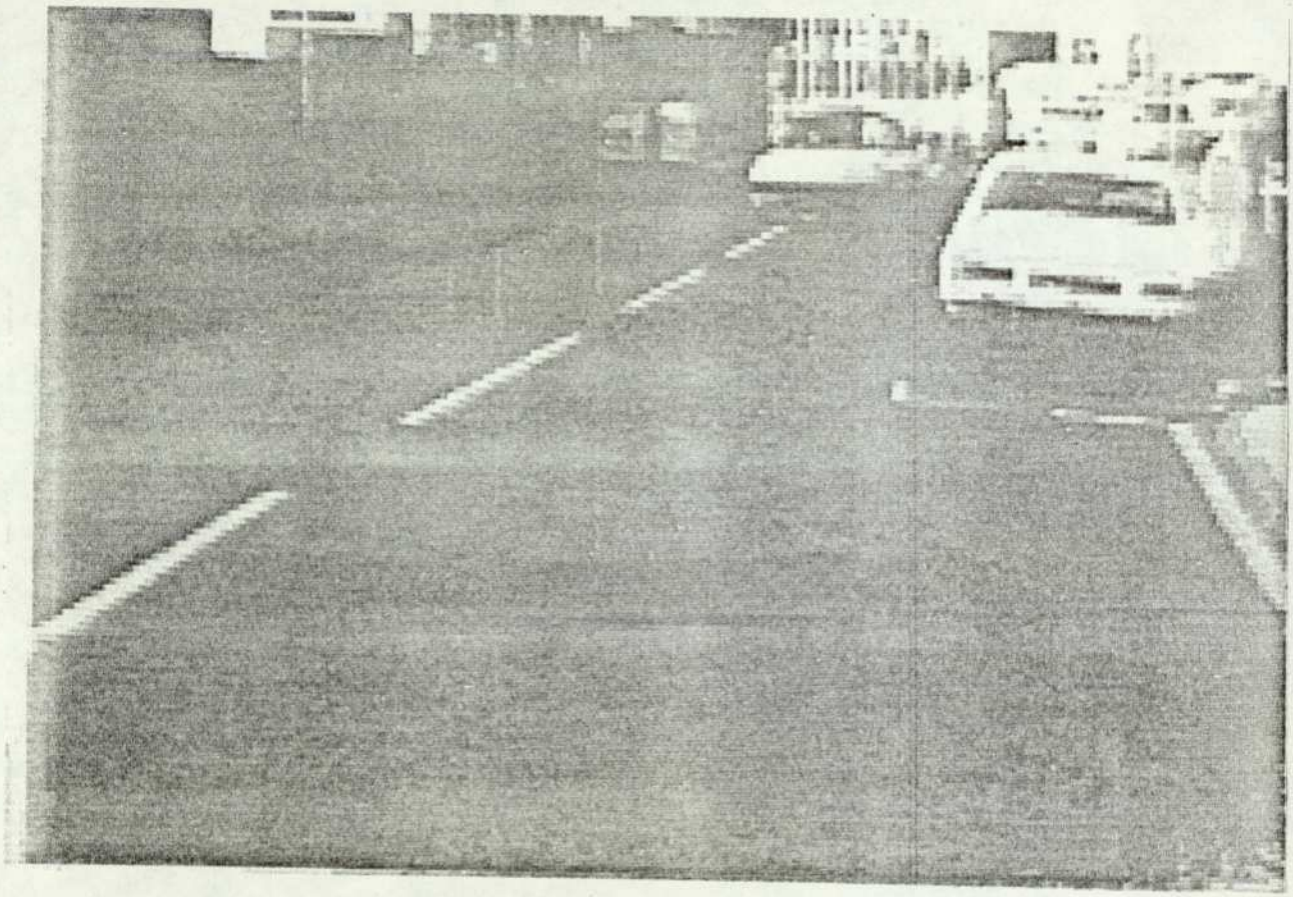


Figure 6.44 : Original street scene image.

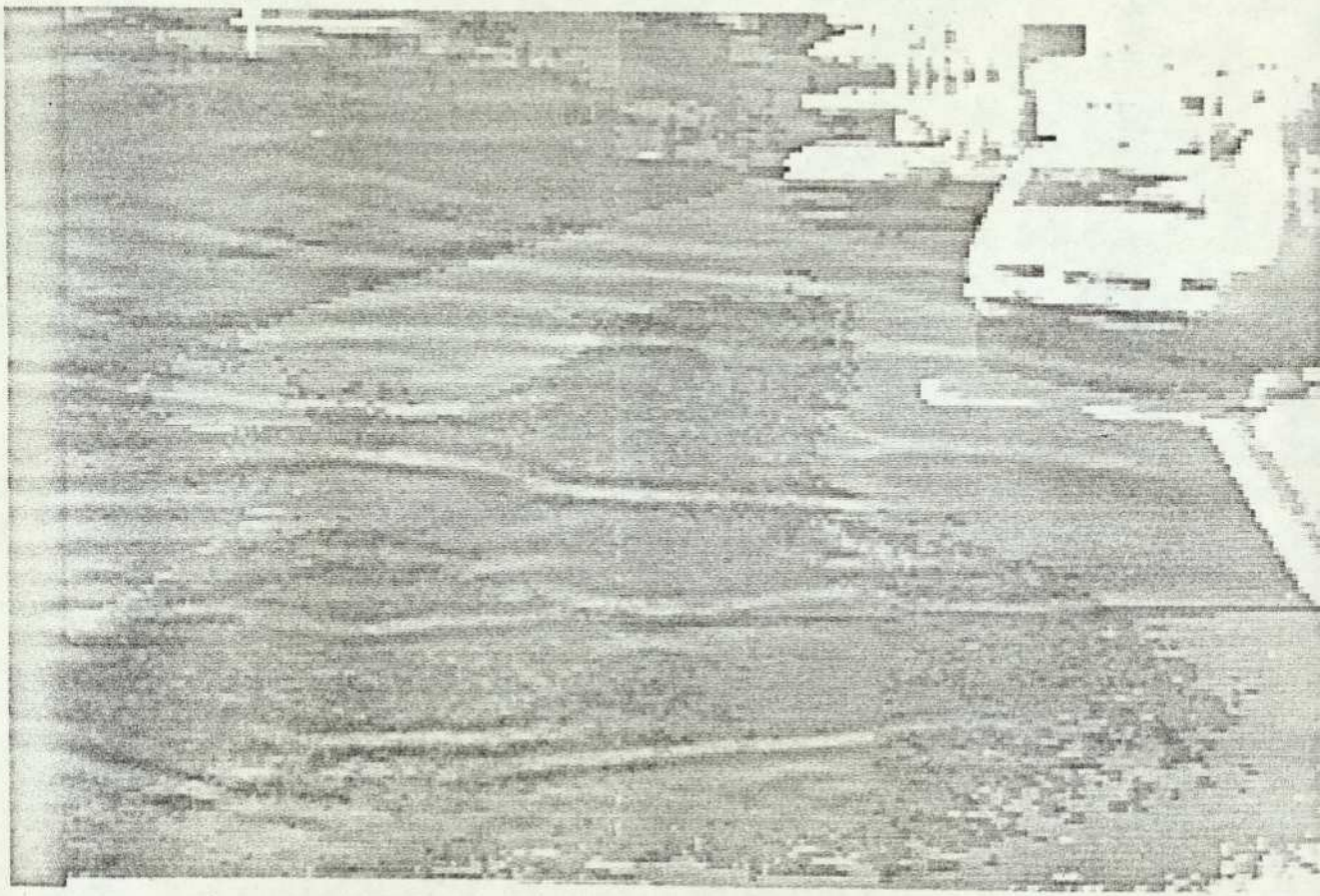


Figure 6.45: Image of fig. 6.44 with central white line removed.

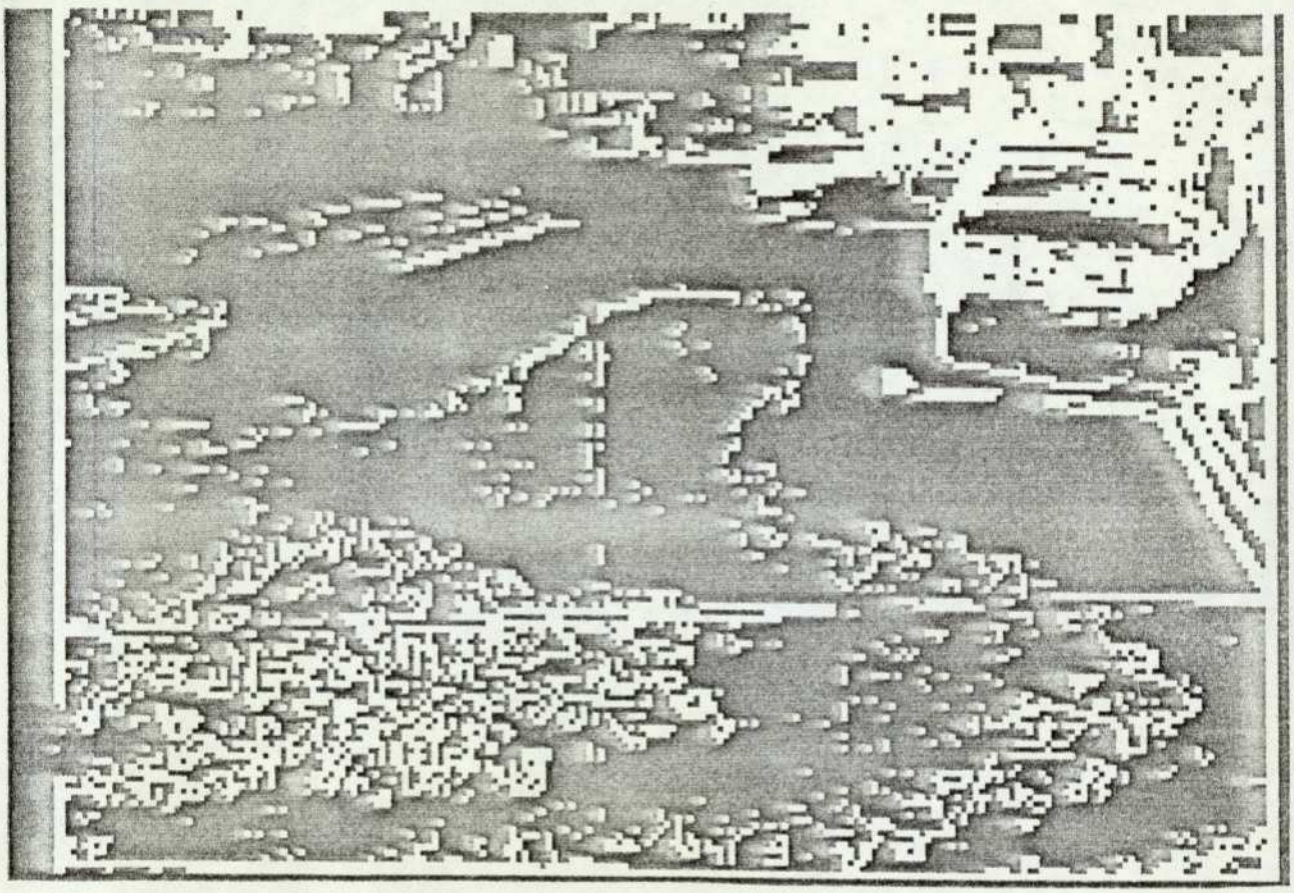


Figure 6.46 :Edge map (with orthogonal masks).

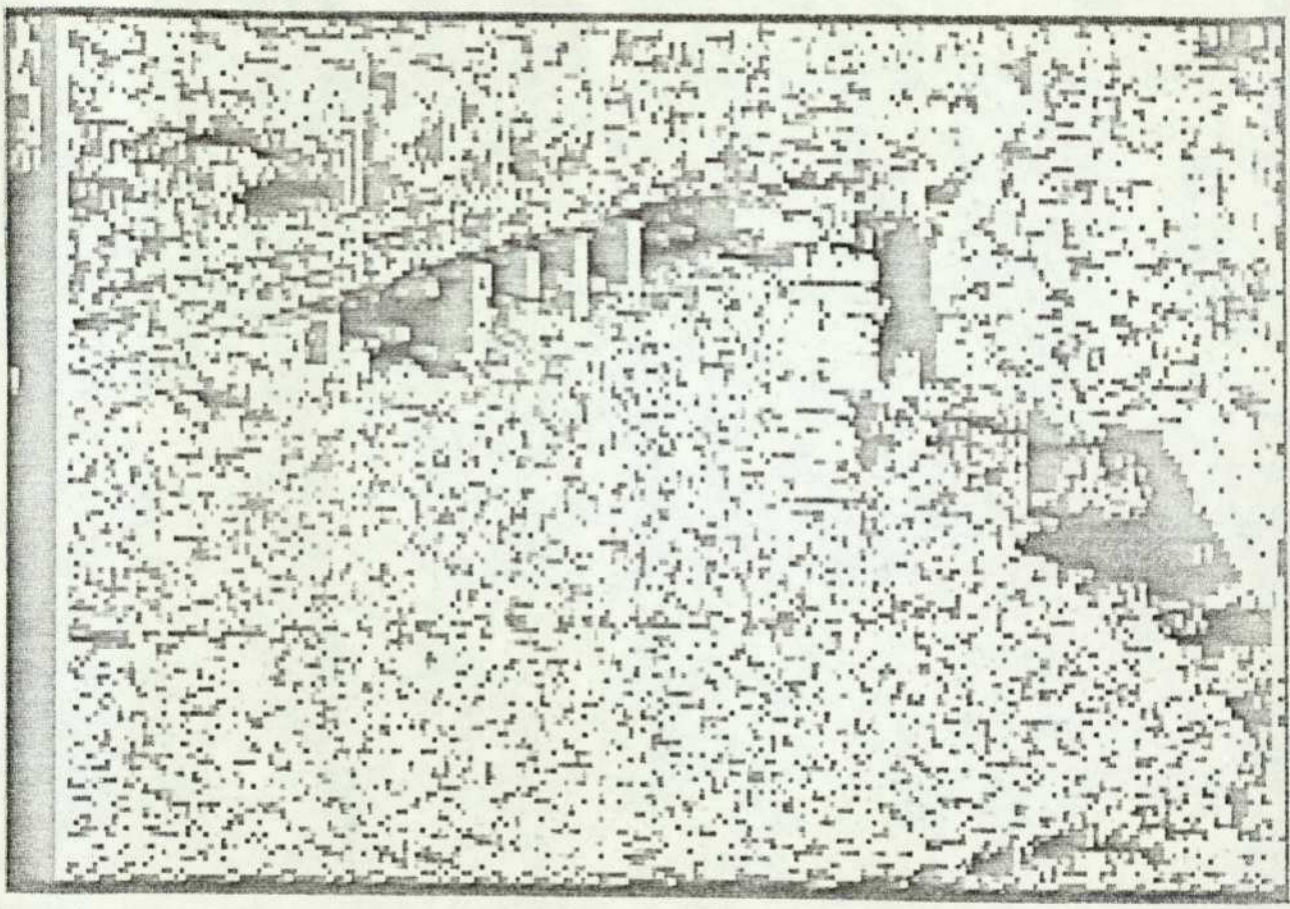


Figure 6.47 : Direction map (with orthogonal masks).

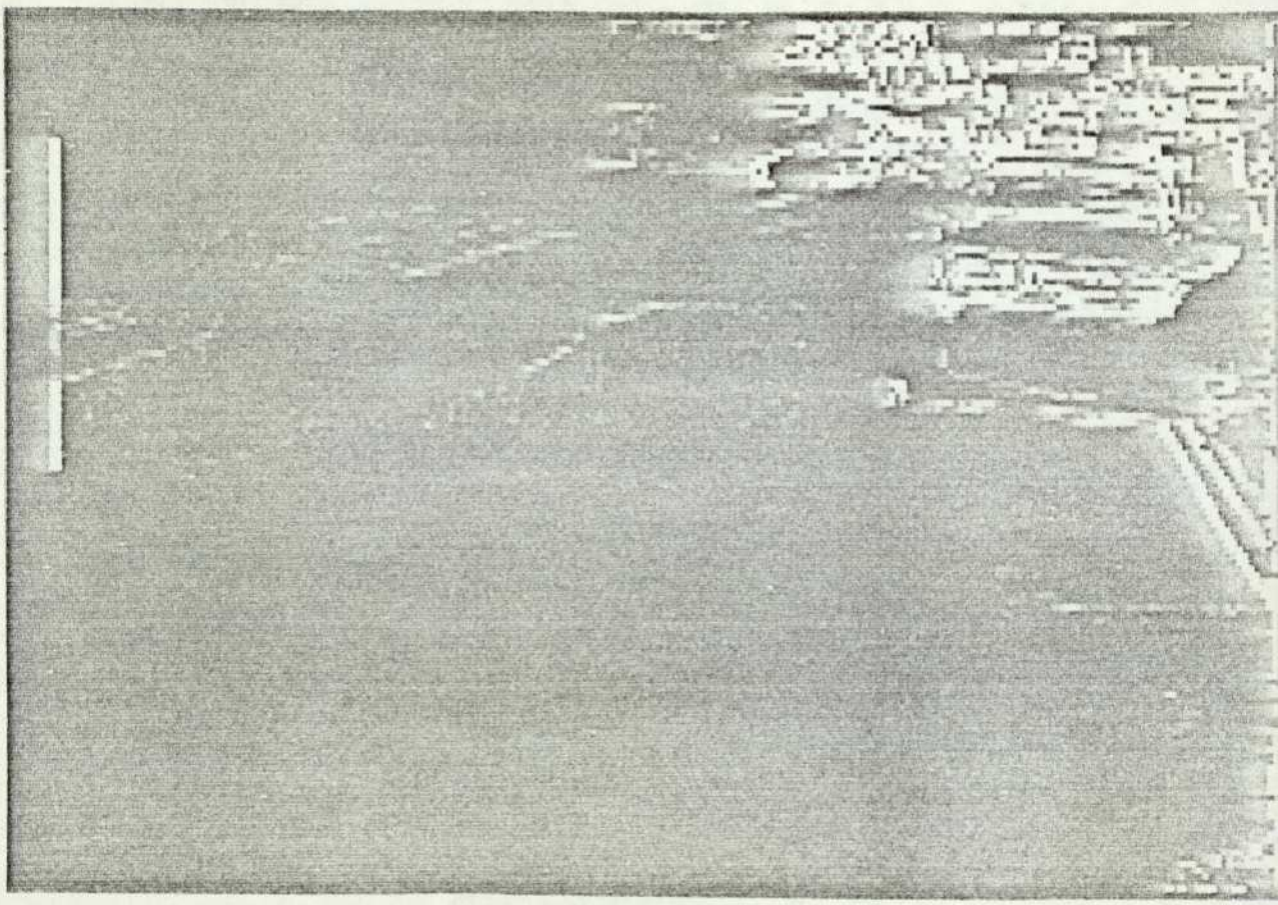


Figure 6.48 : Thresholded edge map.

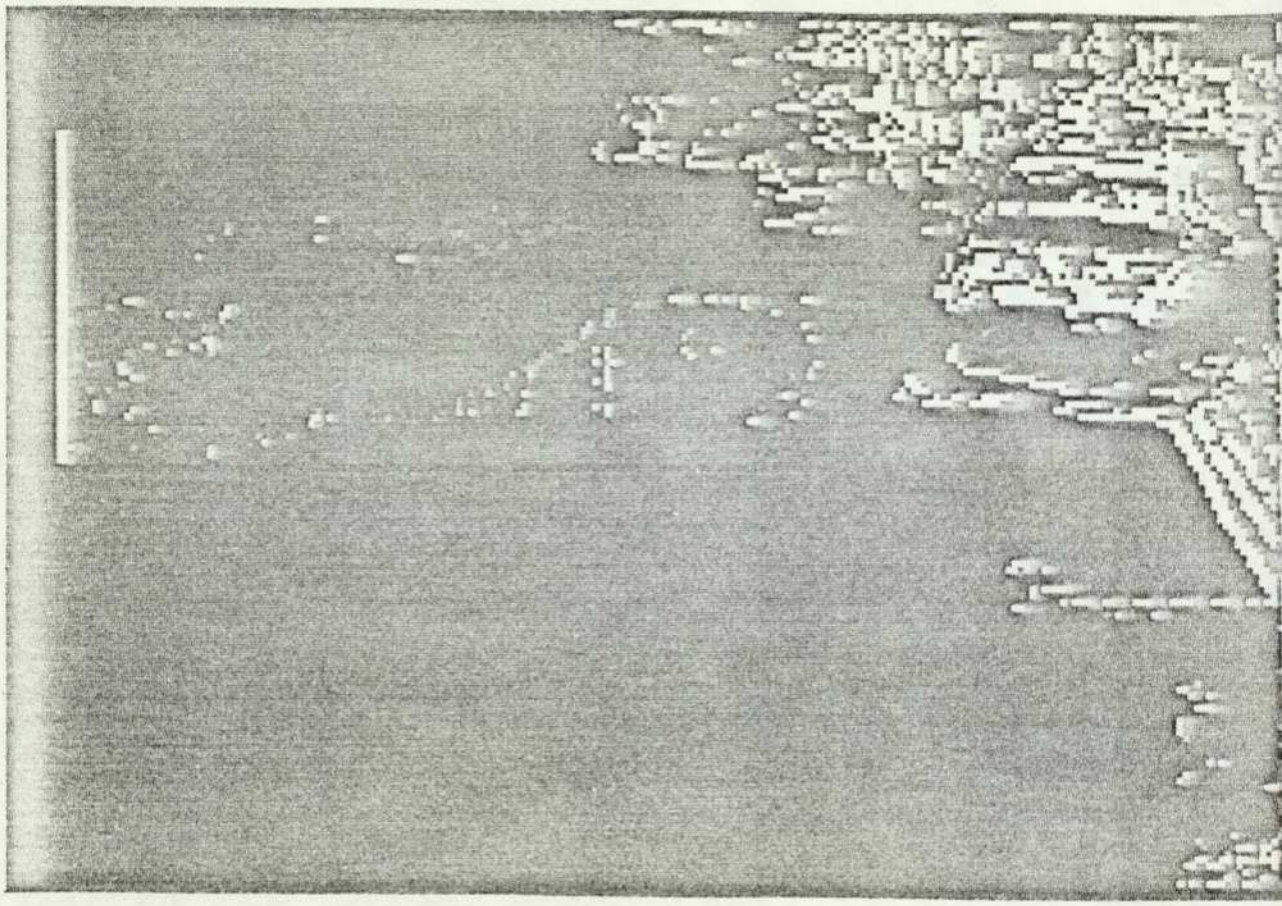


Figure 6.49 : Thresholded direction map.

The third step consists of using the Hough transform to reduce the noise (edges not needed in subsequent analysis) in the image.This filtering procedure consists of generating from the edge magnitude map 26 images by thresholding the original image consecutively between 0 and 9,10 and 19,20 and 29,...,230 and 239,240 and 249,250 and 255.In each of the 26 images,we use the Hough transform to determine the major five lines (lines with the highest number of points).If a line has less than 20 points it is rejected,if it has 20 or more points it is accepted and all its points are recorded in a new image (filtered image),(figure 6.50).The same procedure is applied to the edge direction map,from which we generate 18 images by thresholding the original image consecutively between 0and 9,10 and 19,..., 160 and 169,170 and 180.

The final step in the procedure is to compare the lines obtained from both the edge magnitude and edge direction maps and delete allthe points of the lines which were not present in both filtered images.Using the Hough transform and one of the two final filtered images (edge magnitude and direction maps),we were able to locate the street for the 2 images where the previous system failed(figures 6.51 and 6.52).

The filtering procedure is justified by the fact that we are only interested in the major lines contained in the street scene image,and that the points (edges) of the lines of interest (boundaries of the street)have,in general,approximately a same

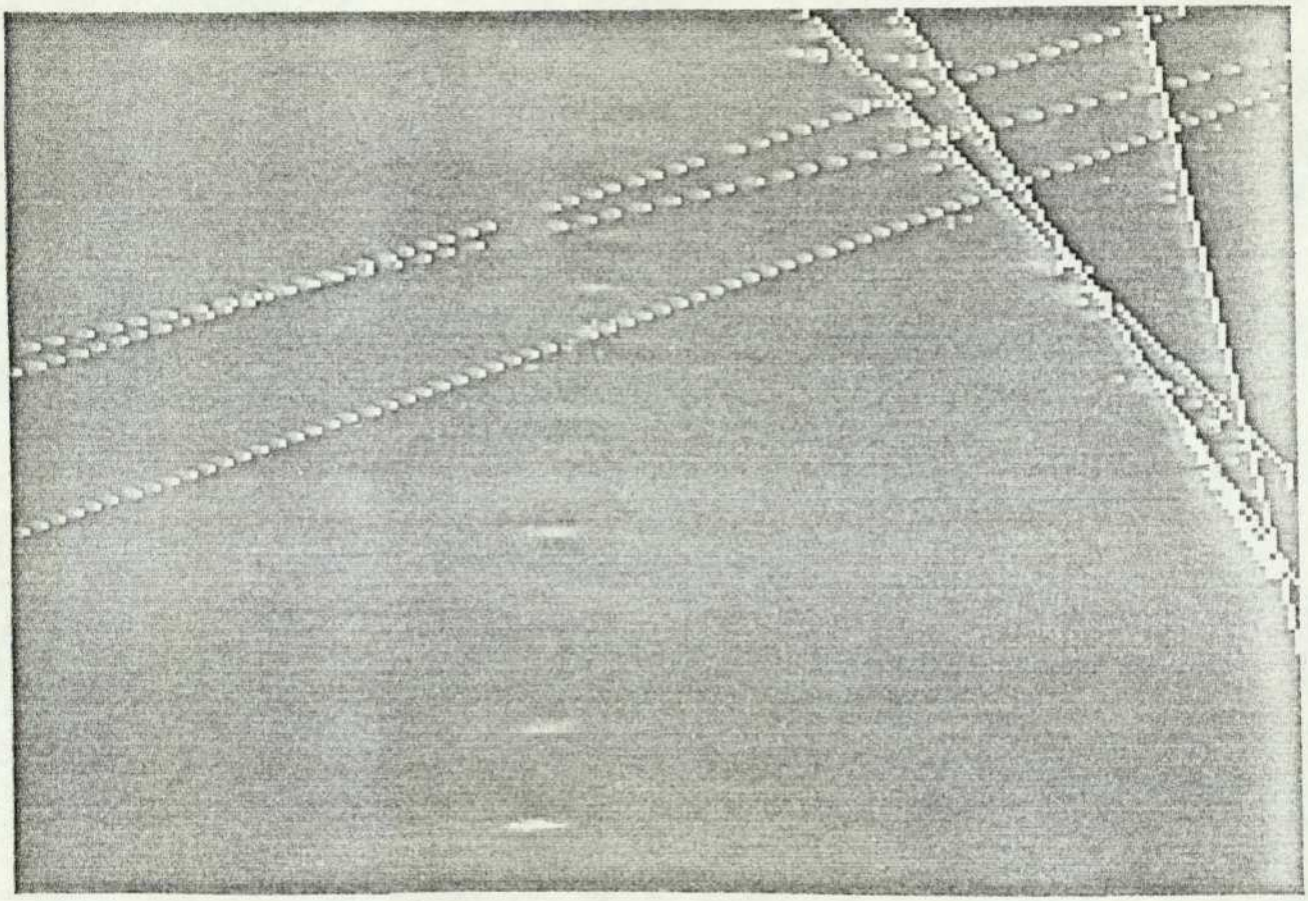


Figure 6.50 : Filtered image (with lines)

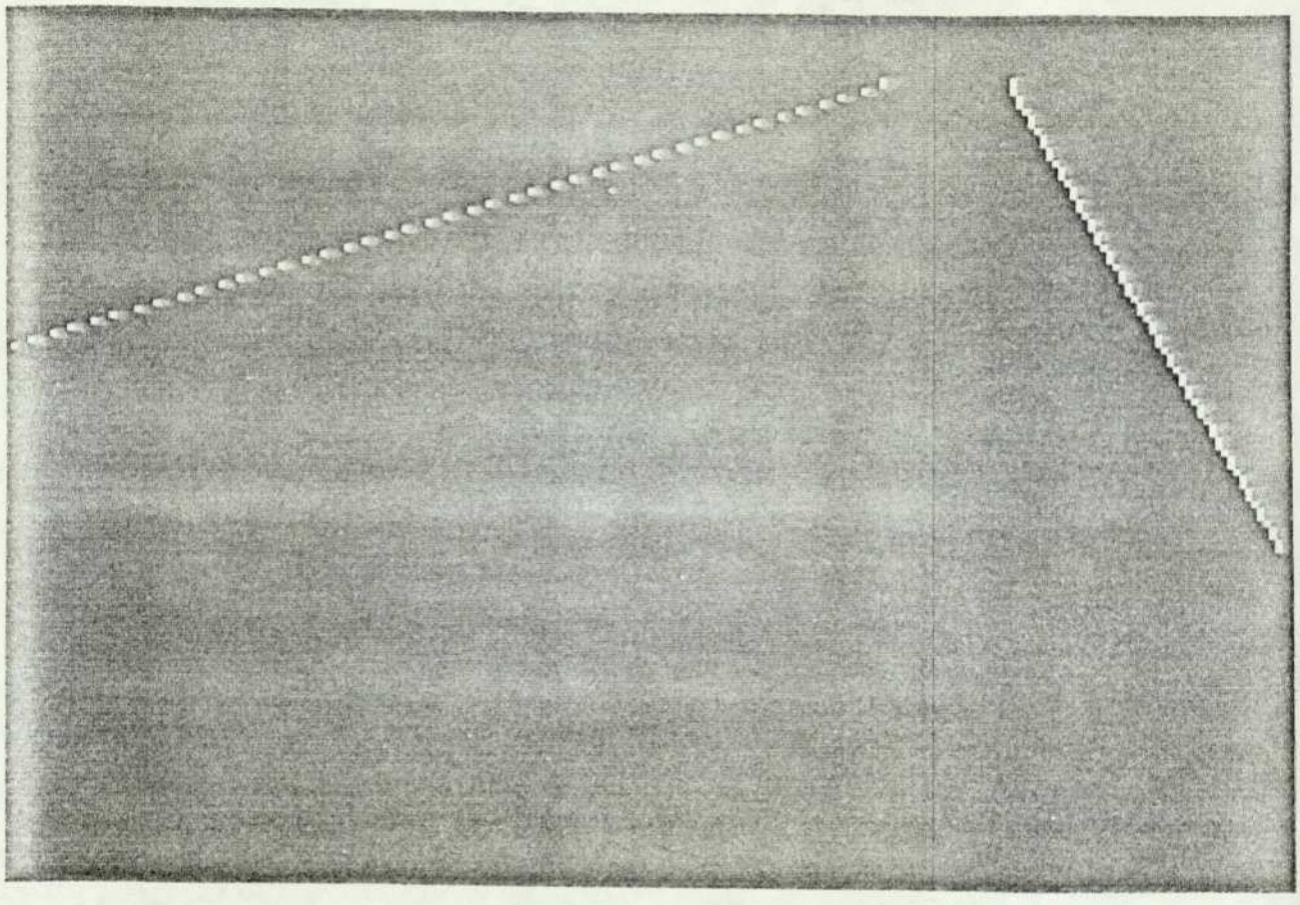


Figure 6.51 : Street boundaries obtained from fig. 6.50

by applying the Hough transform.

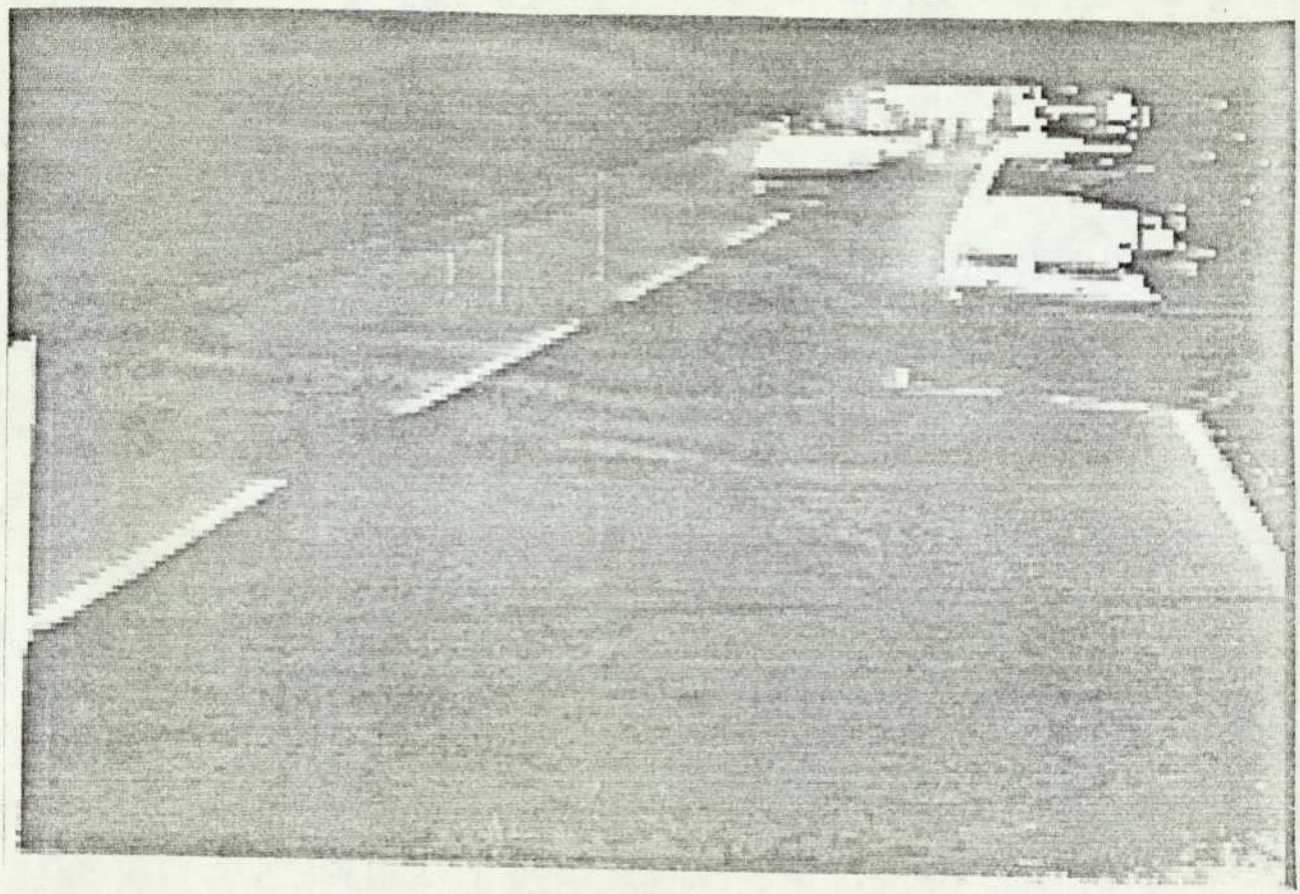


Figure 6.52 : Inside of the street scene image.

points,we were able to reject edges,which did not lie on the street boundaries, and hence we were able to 'filter' the thresholded maps.

This new system, which succeeded in locating the streets in the two last street scene images, can be illustrated by figure 6.53, and diagram 6.3

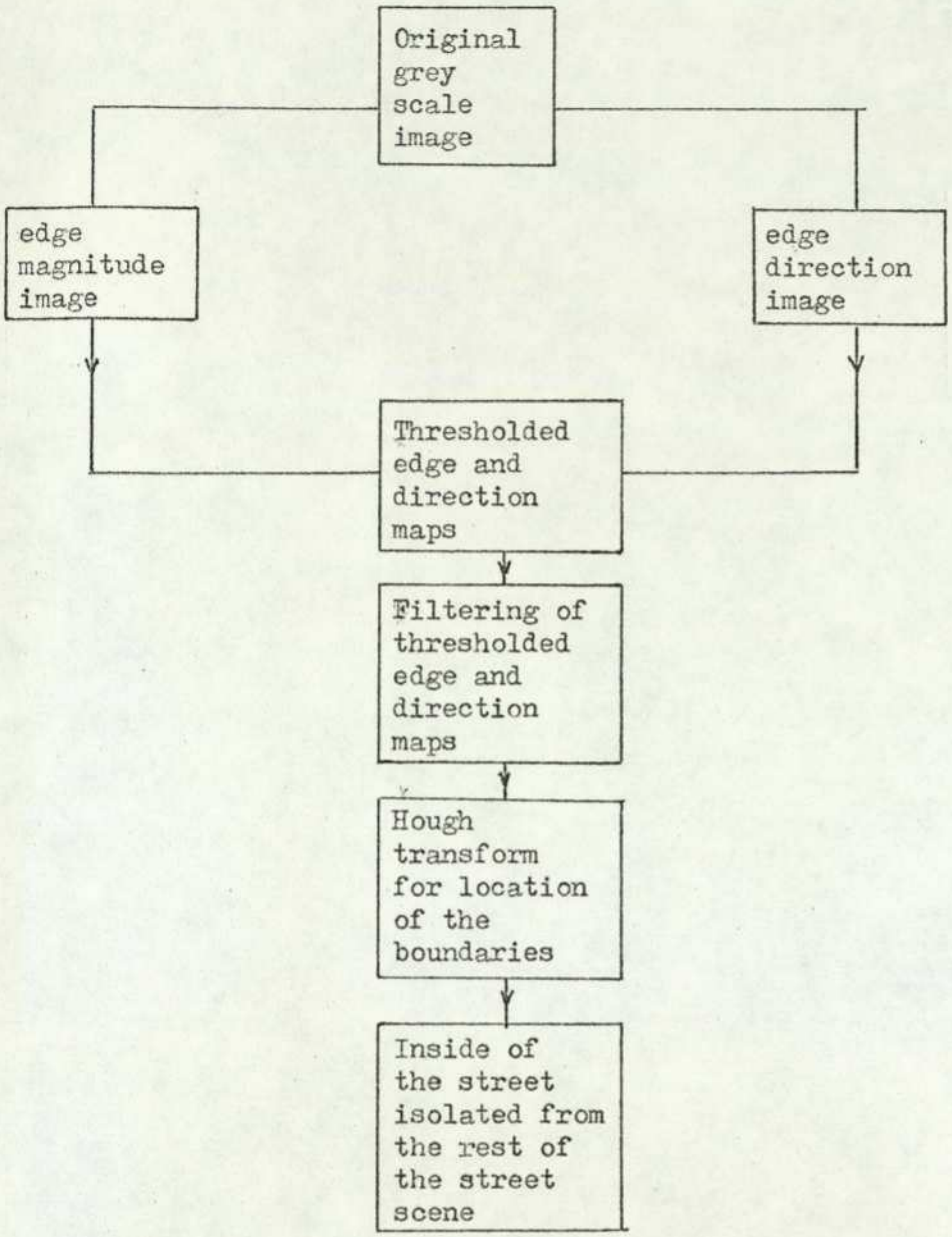


Diagram 6.3 :New system for street location.

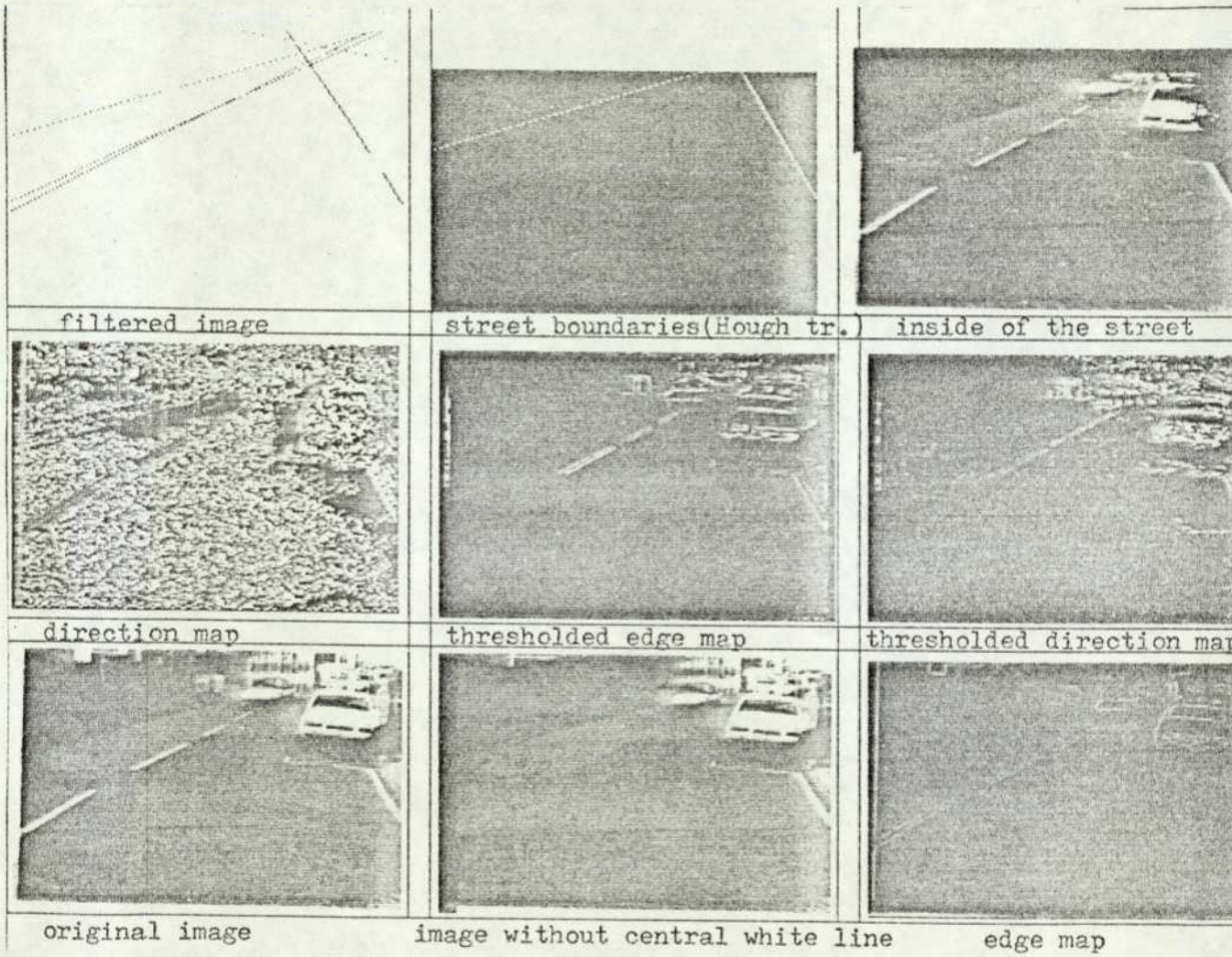


Figure 6.53 : Illustration of the system for extracting the road from a street scene.

7 CALCULATION OF DISTANCES

7.1 Introduction

In this chapter,we will discuss the different concepts involved in the calculation of distances using visual data.An introduction of the basic of photogrammetry will be given.A mathematical discussion of the way to calculate distances, using photographs,will also be provided.

The special relationship between the coordinates of the three-dimensional real world and the coordinates of the two-dimensional photograph will be described in depth.Finally,a discussion of how the photogrammetric techniques could be adapted to resolve our special problem,which consists of evaluating distances between different obstacles inside the street and the guided vehicle, will be provided.

7.2 Basic Of Photogrammetry

Photogrammetry is the science of making precise and accurate measurements from photographs.The basic function of such measurements is to provide metric information on distances,angles,shapes,volumes and movement of objects or structure represented in a given image.It has its beginning in France,in 1859,with Aime Laussedat's surveying camera.Within the past 3

decades,with the adoption of analytical methods, photogrammetry has experienced its greatest advance.

The use of photography as a source of information for various fields of study, among which are engineering,planning,geography,geology,forestry and many others, has increased considerably in recent years. The interpretation of these photographs,sometimes,necessitates the calculation of metric imformation about objects in the image,by using photogrammetry.

Photogrammetry still finds its main use in its original purpose which was topographic mapping.But the number of non-conventional applications of photogrammetry,particularly close range photogrammetry,is rapidily increasing. There are a number of fields in which photogrammetry can be used as an effective measuring technique.Among these fields are traffic investigations,structural deformation studies,medical measurement,archeological surveys,underwater mapping, architectural studies,and many other industrial applications.

Photogrammetry is being used increasingly in the field of industrial metrology as an alternative to traditional measuring techniques.One of the main reason for the adoption of photogrammetry for measurement purposes is that it renders very complex ,and very large or very small objects manageable by virtually replacing the real object by a three-dimensional optical model (two stereoscopic photographs), which is much easier to handle,and which can be

related accurately to any three-dimensional cartesian reference system allowing the measurement of the coordinates of points,lines or surfaces of interest.

The photogrammetrical process consists of a data acquisition phase and a data reduction phase.The data acquisition phase consists of taking the necessary photographs of the object to be measured.The data reduction phase consists of reducing the photographs,which represent perspective projections, into spatial coordinates,which are further processed to obtain the surface area or the volume of the photographed object.

There are two alternative approaches for data reduction:analog and analytical. The analog approach uses special instruments (stereoplotters). The analytical approach,which is increasingly used, uses stereocomparators or monocomparators for data acquisition and computers for processing the collected data.The analytical approach has some important advantages over the analog approach,such as increased accuracy of measurements and superior flexibility.

In photogrammetry,photography is generally considered as a central projection of three-dimensional objects onto a plane.This is only true when the lens does not distort the bundle of rays (collection of rays spreading out in three dimensions in an unbrella rib fashion)relating the object to the image,and when the photographic emulsion (film) lies in a plane. These conditions are satisfied by the use of specially built cameras,or by mathematical

correction when an analytical method is used.

The conversion of data from the three-dimensional object space into a two-dimensional image space is achieved by the use of physical means which represent the projecting system (camera),where the rays are broken in the effective projection centre. A wide variety of cameras,ranging from the very small 8 format,through 16mm, 35mm and 70mm format to 230x230mm format most commonly used for conventional air survey photography,have been used in photogrammetric applications. Because of the relatively low cost of their films and the long period of coverage afforded,16mm cameras have been very popular.But 35mm cameras can offer certain advantages over the 16mm format,because a frame of 35mm film is bigger than the area of a frame of 16mm and thus requires less magnification for study of a given area at a given scale.

Cameras used for close-range photogrammetry are classified in two groups: metric and non-metric cameras.Karara (1975) defines a non-metric camera as one that has not been designed especially for photogrammetric purposes.Faig (1975) gives a more precise definition which states that a non-metric camera has an interior orientation that is completely or partially unknown,and that it does not have fiducial marks.

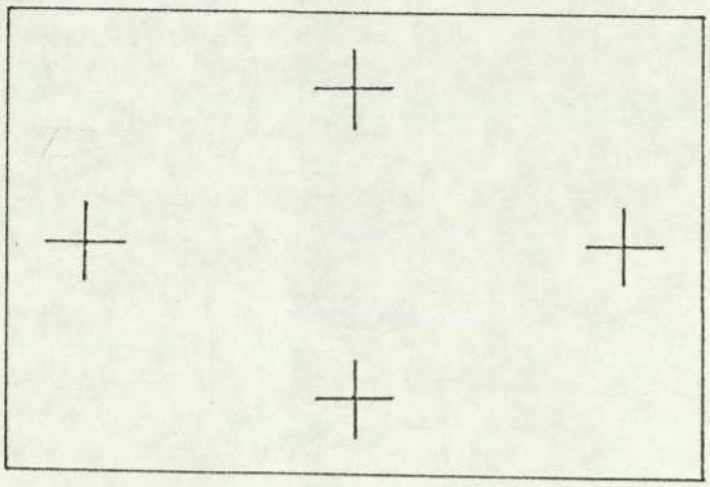


Figure 7.1. Schematic representation of a photograph from a metric camera.

7.3 Close-Range Photogrammetry

We use the term 'non-topographical photogrammetry' to refer to the expanding fields of application of photogrammetry outside the field of topographical mapping.It covers close-range photogrammetry,terrestrial photogrammetry and underwater photogrammetry.Close-range photogrammetry is generally used when the distance between the camera and the photographed object is about 300 metres.

The recent surge of publications on various aspects of non-topographical photogrammetry emphasizes the growing importance

of the special area of photogrammetry.From its conventional use as a mapping tool,photogrammetry has developed,as illustrated by its various applications outside the topographical field, into a precise measurement technique.

Recently,significant progress was made in various aspect of non-topographical photogrammetry.With the substantial increase in the flexibility and versatility of close-range and metric cameras,there has been a sustantial increase in new fields of applications,particularly in biomedical,industrial and architectural photogrammetry.

Among the recent applications,we can mention Burch and Forno (1975) who,at the National Physical Laboratory in collaboration with the Transport and Road Research Laboratory carried out measurements of the deflections of a large beam under load.Among the commercial applications there has been measurements of dams and rock outcrops (Cheffins and Rushton (1970)),and another interesting application described by Cheffins (1975) was the determination of distortions in the area adjacent to the static vent of a BAC1-11 aircraft.

There has also been interest,in a number of countries,in measuring the effect of erosion (Butoutt et al (1974)),for example The City University with the collaboration of the Institute of Hydrology used terrestrial photogrammetry to analyse snow distribution (Blyth et al (1974)).Finally,we can also mention the utilisation of photogrammetry in the analysis of models.An

interesting example described by El Bud(1973) was to obtain precise measurements of surface movement of soil models which rotate in a centrifuge.

The data reduction approaches in close -range photogrammetry are similar to those used for aerial mapping.For metric photography,with little or no modifications,many of the computational methods developed for topographic mapping can be applied in close-range photogrammetrical methods.For non-metric photography,special approaches,such as the Direct Linear Transformation approach (Abdel-Aziz and Karara (1974)) and the Analytical Self Calibration method (Faig (1976)),have been developed.

7.4 Fundamental Concepts In Photogrammetry

In the present section we will concentrate on the general analytical treatment of photogrammetry,and we will be concerned with the general geometric understanding of the basic of photogrammetry.We will start with the definition of the perspective projection,and the definition of the coordinate systems of the camera and the ground.We will conclude with the determination of the equations which relate the coordinates of the image points on the photograph to the real coordinates of the object points.

There is often a need for interpreting the information on a two-dimensional image of a three-dimensional world in order to determine the location of the three-dimensional objects that have been imaged.This interpretation task requires the mathematical understanding of the perspective transformation which describes the transformation of the three-dimensional metric information onto the two-dimensional image.Therefore, before looking at the particular application of the perspective transformation in this research,we will discuss this transformation,which is the fundamental mathematical basis for photogrammetry,in detail.Finally ,we we will concentrate on developing the equations for the single camera.

7.4.1 Perspective Transformation

Photogrammetry may be seen as the science of converting the perspective projection of the image plane to an orthogonal projection of an object.Conventially a point is represented in space by its cartesian coordinates (orthogonal projections) as shown in Figure 7.2 .

The position of the point A (Figure 7.2) is completely defined when the orthogonal distances of A from each plane is known. The space coordinates of A,which are X on the YZ plane,Y on the XZ plane,Z on the XY plane,define completely the point A in space.

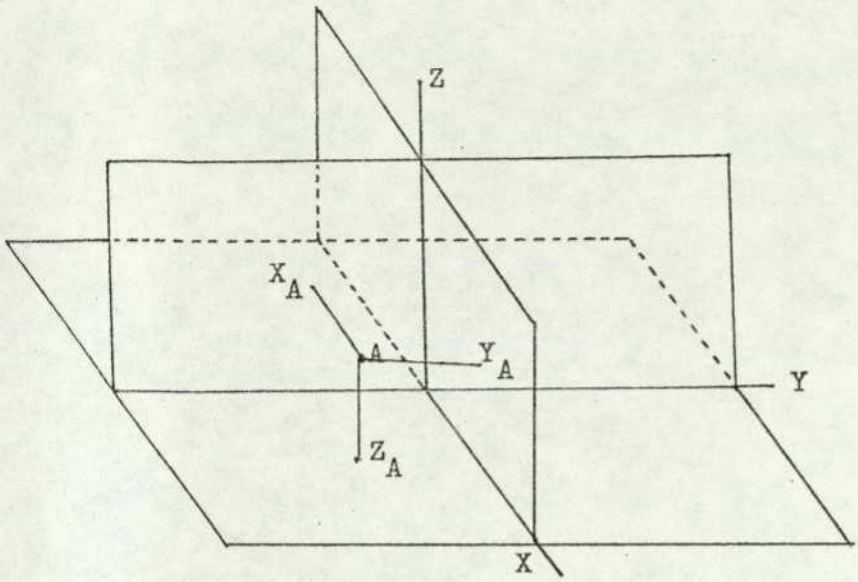


Figure 7.2 : Orthogonal definition of a point.

A projection of an object on a plane is basically the construction of straight lines through every point of the object according to some system (orthogonal projection and perspective projection) and cutting these lines by a plane so as to form an image on the plane which corresponds point for point with the original object.A special projection is the point or perspective projection (Figure 7.3) in which the projecting rays pass through a common point,O.

A camera is a device based on perspective transformation.Basically,a camera is a dark box containing a covered aperture on on side and a light sensitive material (usually film) on the opposite inner side. When the aperture is opened light reflected by the objects outside the camera ,enters through the aperture and records an image (negative) on the light sensitive surface opposite (schematic representation :Fig 7.3).

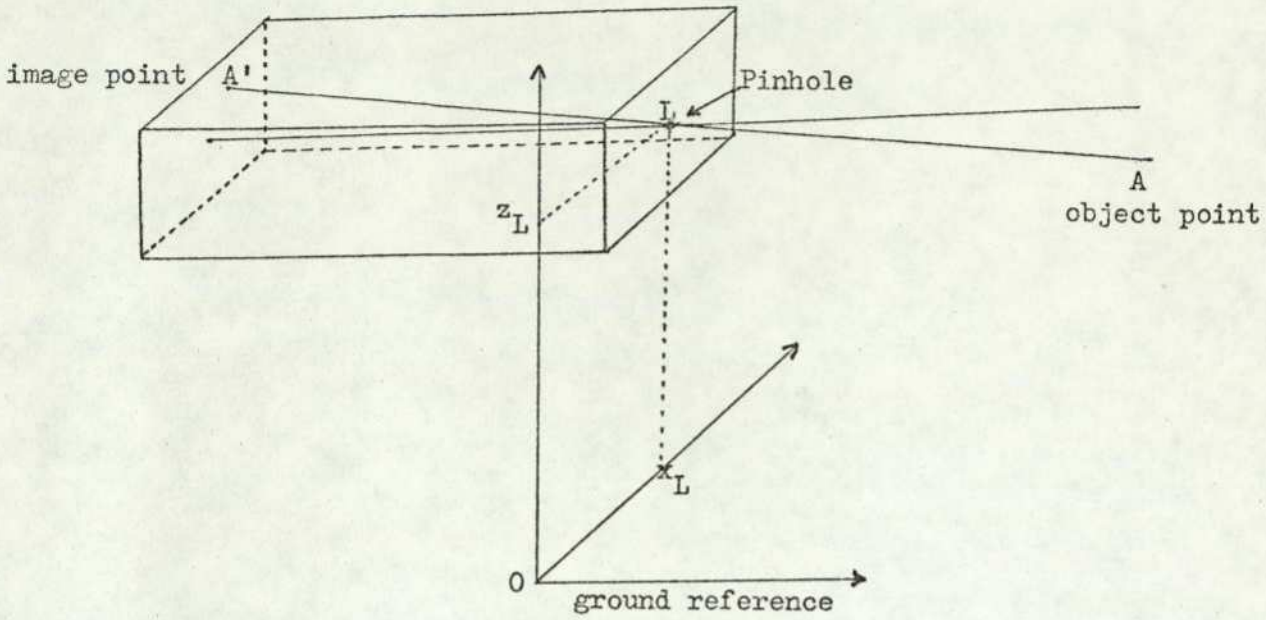


Figure 7.3.Pinhole Camera.

The light sensitive plane on which images of the objects of the viewed scene are recorded is referred to as the image plane,the space within the box is referred to as the image space,and the space outside the box is referred to as the object space.

7.4.2 Image-Ground Relationship

The image-ground relationships are defined with respect to two coordinate systems: The image or camera coordinate system,and the ground coordinate system. The camera coordinate system,referenced to the focal length and the image plane, defines the inner orientation and is used as a reference system

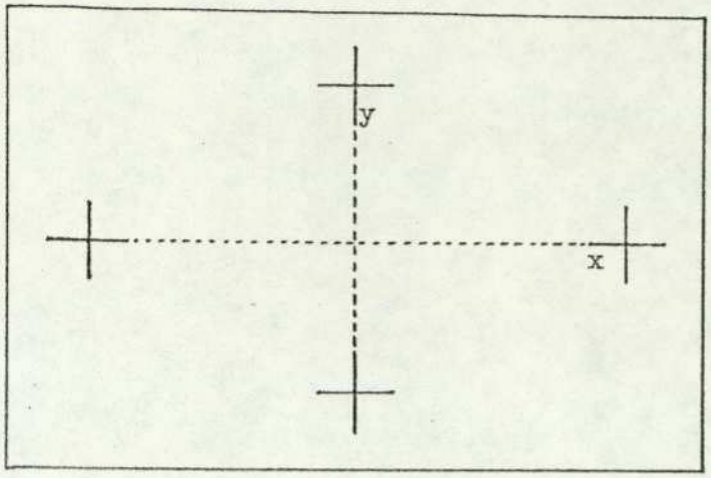


Figure 7.4 : Image coordinate system.

for the image space.The space coordinates (Xo,Yo,Zo) of the camera (perspective centre) referenced to the ground coordinate system and the angular orientation with respect to the object space (ground space) coordinate system,define the exterior orientation.

7.4.2.1 Image Coordinate System

The image plane is usually referred to as the xy plane of the camera. the perspective centre of the lens forms the geometrical centre of the photograph and is referred to as the principal point which is the intersection of the lines that join the opposite pairs of fiducial marks,registered on the sides of the metric photograph.This principal point is,in

general,taken as the origin of the image coordinate system,with the x axis parallel to one edge of the image plane,and the y axis parallel to the other edge of the image plane.The two axis x and y are respectively obtained by joining the opposite fiducial marks (Figure 7.4).The third axis z of the image coordinate system,which coincides with the optical axis of the camera,passes through the principal point and is perpendicular to the image plane.

The cartesian coordinate system of the image space is therefore defined by the x and y axis,which lie in the image plane and intersect at the principal point,and the normal distance f from the centre of projection to the xy image plane.

7.4.2.2 Ground Coordinate System

Similarly to the image points whose positions,in the image space are established by their coordinates (x,y,f),the object points have their coordinates (X,Y,Z) in the ground coordinate system which establish their positions in the object space.In general,the coordinates of object points are based on an orthogonal projection referred to an arbitrary origin with the XY plane parallel to a datum plane (sea-level or ground),and the Z axis perpendicular to the XY plane.The

coordinate Z on the Z axis are in general referred to as the elevations or heights of the object points.

The coordinates (Xo,Yo,Zo) of the centre of projection with respect to the ground coordinate system,represent the space coordinates of the camera in the object space.The coordinates Xo and Yo are referred to as the plane coordinates of the camera,and the coordinate Zo as the elevation of the camera.The object space coordinates of the camera (Xo,Yo,Zo) could be calculated from photogrammetric equations, if the space coordinates of three object points and their image points are given.

7.4.2.3 Image-Ground Relationship And Angular Orientations

The image space is referenced to a plane and a point not in the plane at some fixed location with respect to the plane.The plane is the xy image plane which is at some distance f from the centre of projection,and the point is the centre of projection.Similarly the object space is referenced to a plane and a point not in the plane itself,but at a fixed position with respect to the plane. The object space plane is the XY datum plane,and the point is the centre of projection which is a fixed point at a distance Zo from the XY datum plane.The space coordinates of the centre of

projection,(Xo,Yo,Zo),are the three coordinates which define the position of the camera.

However the three coordinates do not determine,on their own,the complete spatial position of the camera.When the camera is tilted the geometrical relationship between the camera and the ground will be changed.To specify the complete spatial position of the camera,the tilt angles have therefore to be specified (Figure 7.5).

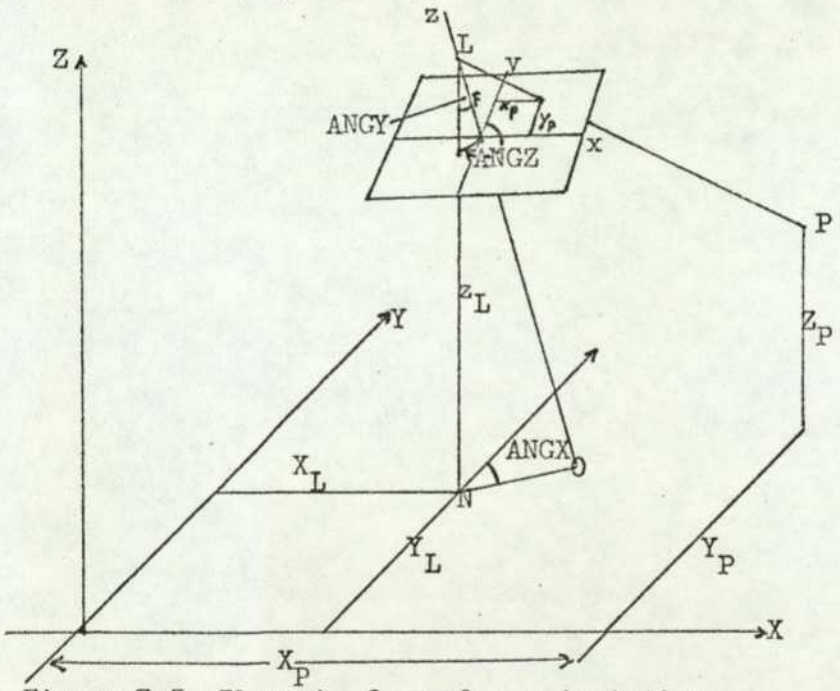


Figure 7.5 :Element of angular orientation.

When the xy image plane is not parallel to the XY ground reference plane owing to a rotation of angle ANGY,taken about the Y axis,the angle t is referred to as the tilt of the image plane with respect to the reference plane. With the presence of tilt the optical axis of the camera is no longer pointing vertically,and the point on the image plane vertically beneath

the camera is referred to as the nadir point.The optical axis and the vertical line together constitute a plane referred to as the principal plane.

A second rotation of angle ANGX,taken about the X axis,could be introduced. The angle ANGX is referred to as the azimuth of the principal plane. A third rotation of angle ANGZ,taken about the Z axis,could be introduced. The angle ANGZ is referred to as the swing angle. clockwise in the xy plane from the y axis to the principal line.

The tilt,the swing,and the azimuth of the principal plane are the elements of the exterior orientation of the photograph.These three angles and the coordinates (Xo,Yo,Zo) of the centre of projection define completely the position and orientation of the photograph in space with respect to the ground reference system.

When the location of the perpendicular to the image plane and the focal length are known the interior orientation of the camera is known.When the space coordinate (Xo,Yo,Zo) and the angular orientation ANGY,ANGZ,and ANGX are known then the exterior orientation of the camera is known.When the interior and exterior orientation of the camera are known,the angular and linear relation of image and object are completely established provided all objects lie in the XY plane. But this seldom occurs.In general,objects do not lie on a

plane,and the elements of interior orientation and exterior orientation of a single camera do not define the angular and linear relation of image and object,unless the elevation of the objects with respect to the XY plane are known.But, when two cameras are used,and the elements of the interior and exterior orientation of the two cameras are known,then the spatial configuration of the surface in object space,represented in the overlaping image planes, may be completely defined.

7.4.3 Perspective Transformation Equations

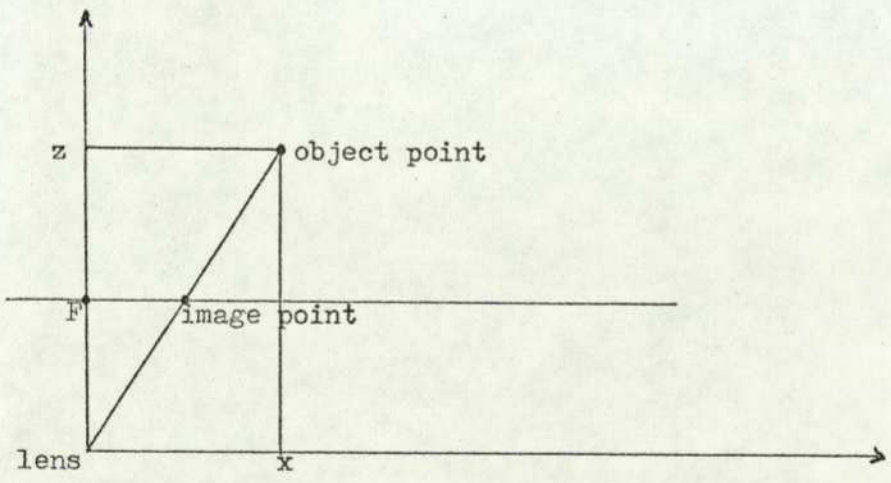


Figure 7.6: Illustration of a simple perspective transformation in one dimension

In order to keep the image in a positive orientation,we

assume that the image plane is at a distance f (principal distance) in front of the camera lens,and that the lens project forward to it.This eliminates the problem of left-right reversal in an image behind the lens.

Let us assume that the camera lens is at the origin and points directely down the z axis as illustrared by figure 7.6 . Then we have the image plane xy parallel to the plane XY. According to the geometric rays optics model of the lens,the lens will focus an object point,(X,Y,Z),on the image plane,which is parallel to the XY plane at a distance f directly in front of the lens,at the intersection between the line from the object point to the origin,and the image plane. Hence the perspective projection will have the coordinates ((Xf/Z),(Yf/Z),f) in the original coordinate system.Because both the numerator and the denominator of (Xf)/Z and (Yf)/Z are linear combination of X,Y and Z,we can compute them by an appropriate linear transformation by using homogeneous coordinates and taking ratios of the components of the transformed vector. The transformation could therefore be represented by the product of a translation ,which translate the object point (X,Y,Z,1) down the Z axis by a distance f and a perspective transformation to the image plane.Hence

perspective translation

$$\begin{vmatrix} x' \\ y' \\ p \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 1 \end{vmatrix} \times \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -f \\ 0 & 0 & 0 & 1 \end{vmatrix} \times \begin{vmatrix} X \\ Y \\ Z \\ 1 \end{vmatrix}$$

The image coordinate are the obtained by the following equations:

$$x = x'/p \quad , \quad y = y'/p$$

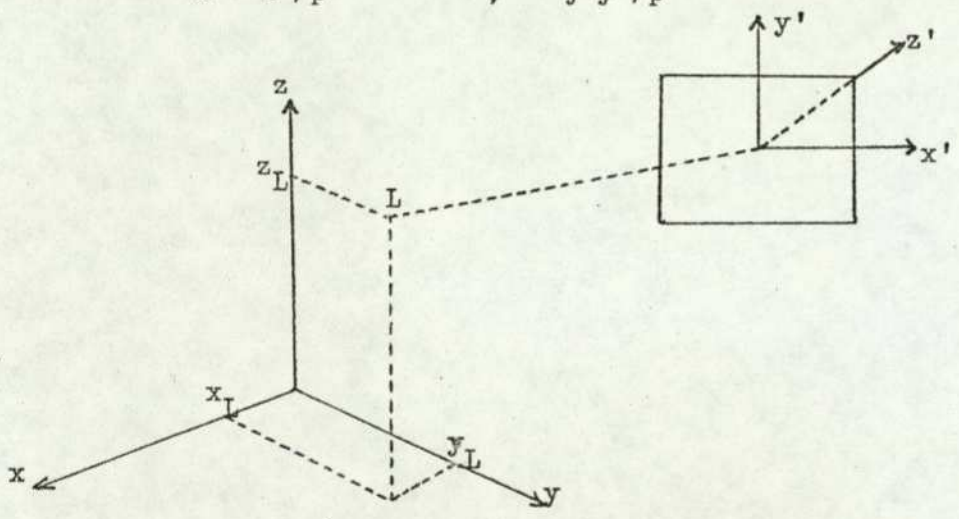


Figure 7.7 : Ground and camera coordinate system.

If the camera,which takes two-dimensional images of a three-dimensional world, is set as illustrated by Figure 7.7,then to obtain the image frame coordinates for a given point in three-dimensional space,we first translate this point to a

three-dimensional coordinate system centred at the lens of the camera. Then we rotate the coordinate system so that its XY plane is parallel to the image plane.Finally,the coordinates in the image are then obtained by translating the rotated coordinate system along its Z axis to the desired location of the image and taking the perspective transformation to it.

We will take (X,Y,Z) as the original coordinates of a point in a three-dimensional space and (x,y) as the coordinates of the perspective projection of the object point in the image,and (Xo,Yo,Zo) as the position of the lens in the ground reference system.If we assume that the lens is pointing down the Z axis in a new coordinate system obtained by rotating the XY plane through an angle ANGX(Azimuth angle) rotating the XZ plane through an angle ANGY (tilt angle),and rotating the XY plane through an angle ANGZ (swing angle).The perspective transformation involved is then defined by first translating the original ground reference system to the centre of the lens.Then the Y and Z axes are rotated by an angle ANGY (tilt angle),the X and Y axes are rotated by an angle ANGX (azimuth angle),and the X and Z are rotated by an angle ANGZ (swing angle).These coordinates are then translated to the image plane and the perspective is taken to it.The following equation defines the process involved in the perspective tansformation.

$$P.T.R.T' \cdot \begin{vmatrix} X \\ Y \\ Z \\ 1 \end{vmatrix} = \begin{vmatrix} x' \\ y' \\ p \end{vmatrix} \qquad (1)$$

And $x = x'/p$, $y = y'/p$

R : Perspective projection matrix

T : Translation (from centre of the lens to the image) matrix

R : Rotation (by ANGZ,ANGY,and ANGX) matrix

T : Translation (to the centre of the lens) matrix

$$P = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/f & 1 \end{vmatrix}$$

$$T = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -f \\ 0 & 0 & 0 & 0 \end{vmatrix}$$

$$T' = \begin{vmatrix} 1 & 0 & 0 & -Xo \\ 0 & 1 & 0 & -Yo \\ 0 & 0 & 1 & -Zo \\ 0 & 0 & 0 & 1 \end{vmatrix}$$

The rotation matrix is the product of the matrices R',R'',R''' which represent respectively the rotation by angle ANGX,around the X axis,the rotation by angle ANGY,around the Y axis,and the rotation of angle ANGZ,around the Z axis.

$$R=R'.R''.R'''$$

$$R' = \begin{pmatrix} \cos ANGZ & \sin ANGZ & 0 & 0 \\ -\sin ANGZ & \cos ANGZ & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$R'' = \begin{pmatrix} \cos ANGY & 0 & -\sin ANGY & 0 \\ 0 & 1 & 0 & 0 \\ \sin ANGY & 0 & \cos ANGY & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$R''' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos ANGX & \sin ANGX & 0 \\ 0 & -\sin ANGX & \cos ANGX & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

```
   ! cosANGYcosANGZ  cosANGXsinANGY          sinANGYsinANGZ         0!
   !                +sinANGXsinANGYcosANGZ -cosANGXsinANGYcosANGZ 0!
   !                                                               !
R=!-cosANGYsinANGZ  cosANGYcosANGZ          sinANGXcosANGZ         0!
   !                -sinANGXsinANGYsinANGZ +cosANGXsinANGYsinANGZ 0!
   !                                                               !
   ! sinANGY         -sinANGXcosANGY         cosANGXcosANGY         1!
```

From equation (1) we get :

```
x=(-F).((X-XO).Cos(ANGY).Cos(ANGZ)+(Y-YO).(Cos(ANGX).Sin(ANGZ)
   +Cos(ANGZ).Sin(ANGX).Sin(ANGY))+(Z-ZO).(Sin(ANGX).Sin(ANGZ)
    /((X-XO).Sin(ANGY)
+(Y-YO).(Sin(ANGX).Cos(ANGY))+(Z-ZO).Cos(ANGX).Cos(ANGY))
y=(-F).((X-XO).(-Cos(ANGY).Sin(ANGZ))+(Y-YO).(Cos(ANGX).Cos(ANGZ)
     -Sin(ANGX).Sin(ANGZ).Sin(ANGY))+(Z-ZO).(Sin(ANGX).Cos(ANGZ)
    +Cos(ANGX).Sin(ANGY).Sin(ANGZ)))/((X-XO).Sin(ANGY)
+(Y-YO).(-Sin(ANGX).Cos(ANGY))+(Z-ZO).Cos(ANGX).COS(ANGY))
```

Two interesting properties of the perspective transformation,which relate lines in the three-dimensional world to lines in the image plane,are that lines in the three-dimensional world transform to lines in the image plane ,and that parallel lines in the three-dimensional world meet in a vanishing point in the image plane.

7.5 Distance Measurement

It is possible to measure distances using a single monochrome image providing the orientation of the image with respect to the scene is known ,that the scene is assumed to comprise a horizontal plane,and that the hight of the camera is known.By moving a cursor inside the street image,our program can read the absolute (real) coordinates of any point of the street image pointed at by the cursor (Figure 7.8).

To define the positions of the objects and of the camera two reference systems are used.The main reference system is taken in such a way that the XY plane lies on the street ,the origin O lies at the intersection of the plane of the street and the perpendicular to the street,which passes through the center of the camera lens.The system associated with the camera has the zx plane (the axis z coincides with the focal axis) parallel to the street,and the y axis perpendicular to the XY plane,as shown below:

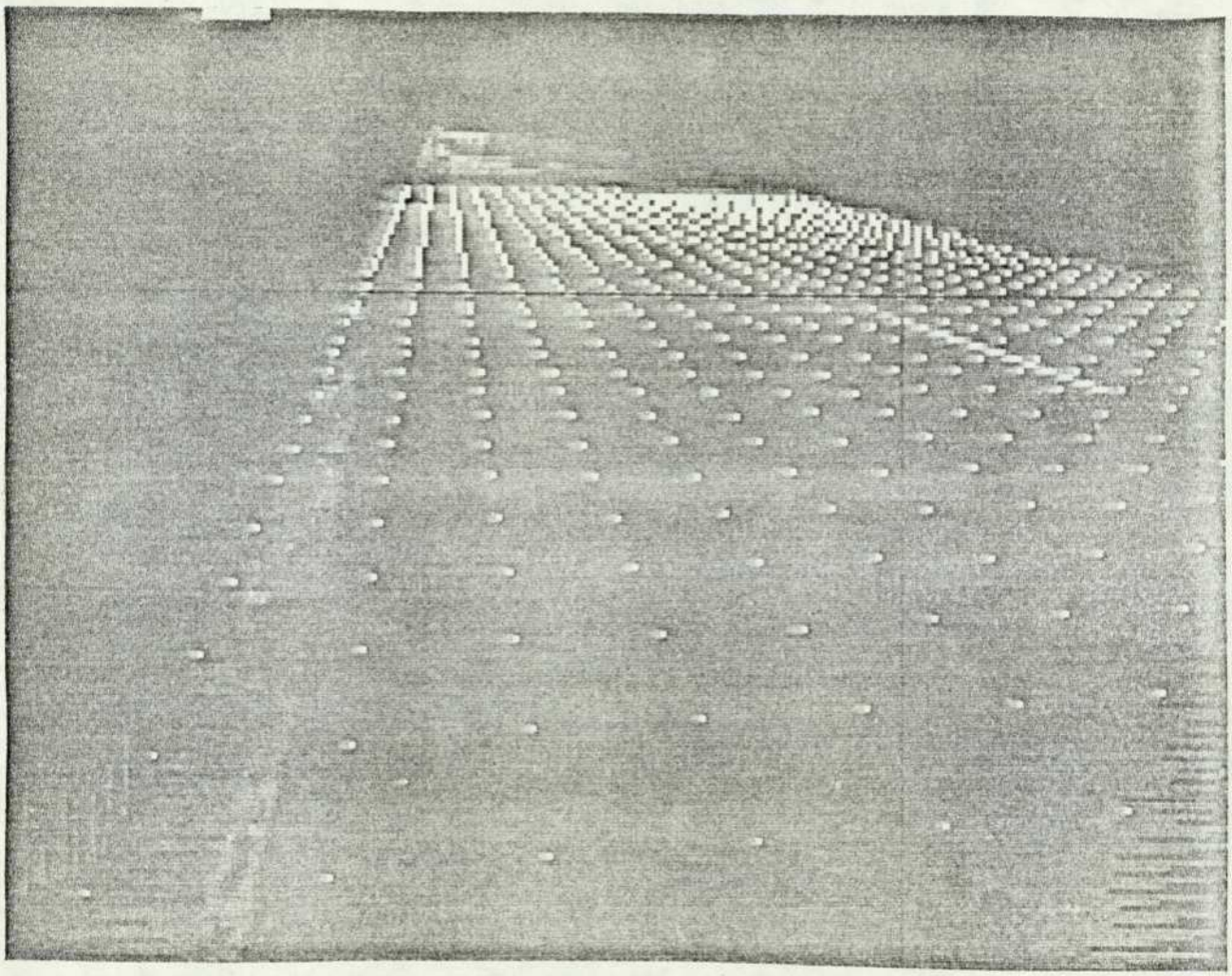
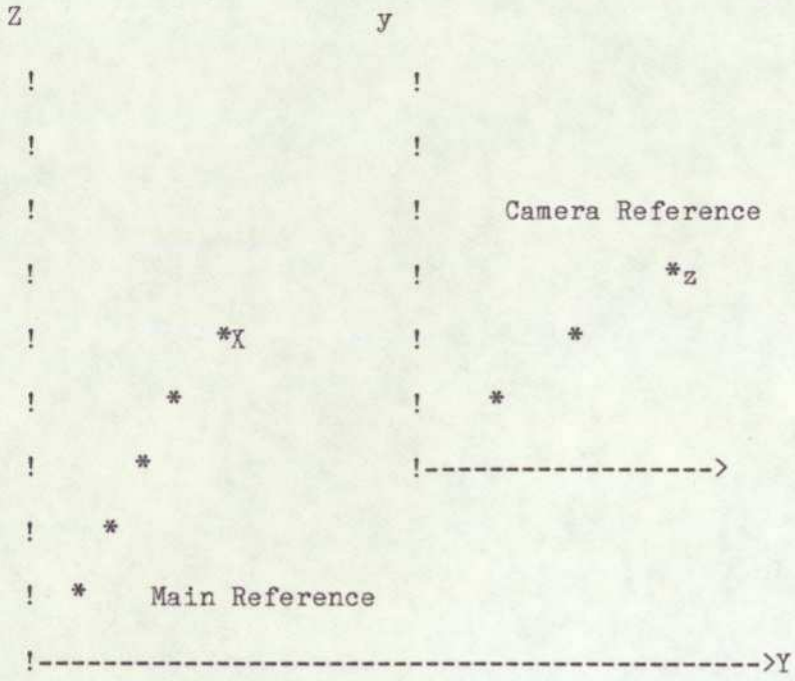


Figure 7.8: Inside of a street with grid (0.5x0.5 m).



The image of the street is projected on a plane parallel to xy,using as center of projection the origin of the camera reference system.Only a portion of the projection plane (the proportion depends on the particular camera used) is used,and called the film plane.The distance F between the film plane and the center of projection could be considered asthe focal length of the camera.The camera used was a WILD P32 terrestrial camera with principal distance 64mm,angular field:long side 64 degrees(32+32),and short side 49 degees(32+17),and image format 60mmx80mm.

Knowing the distance from the center of projection to the street and the tilt angles of the camera ANGX,ANGY,ANGZ respectively around the axis x,y,z, we can project any point ,inside the field of the camera,and in the plane XY of the street ,which is supposed to be plane,on the film plane.Hence we can associate to every point on

the film plane a point on the street and its coordinates in the real three dimensional world.The procedure is therefore to take every point of the plane which lies on the street and the field of the camera,project it on the film plane,and at the point of projection write the coordinates of the projected point.By completing this 'distance map ' we can read from the position of the projected point on the film its real coordinates in the real world. If x and y are the coordinates in the film frame,X and Y the coordinates of the points on the street,ANGX,ANGY,ANGZ the various tilt angle of the camera ,H the height of the camera from the street,and F the focal distance of the camera,the the projection laws are as follows:

```
x=(-F).((X-XO).Cos(ANGY).Cos(ANGZ)+(Y-YO).(Cos(ANGX).Sin(ANGZ)
   +Cos(ANGZ).Sin(ANGX).Sin(ANGY))+(Z-ZO).(Sin(ANGX).Sin(ANGZ)
    /((X-XO).Sin(ANGY)
+(Y-YO).(Sin(ANGX).Cos(ANGY))+(Z-ZO).Cos(ANGX).Cos(ANGY))
y=(-F).((X-XO).(-Cos(ANGY).Sin(ANGZ))+(Y-YO).(Cos(ANGX).Cos(ANGZ)
     -Sin(ANGX).Sin(ANGZ).Sin(ANGY))+(Z-ZO).(Sin(ANGX).Cos(ANGZ)
     +Cos(ANGX).Sin(ANGY).Sin(ANGZ)))/((X-XO).Sin(ANGY)
+(Y-YO).(-Sin(ANGX).Cos(ANGY))+(Z-ZO).Cos(ANGX).COS(ANGY))
```

XO,YO,ZO are the coordinates of the origin of the system of reference of the camera in the main system of reference which is associated to the street,and ZO=H (h :height of the camera).

7.6 Conclusion

In this chapter,the role of photogrammetry in providing three-dimensional metric information from two-dimensional photographs has been sressed.The application of non-topographical and close range photogrammetry in various fields have been underlined to place our work in a wider context.The coordinate systems involved in photogrammetry have been described.

The geometrical relationship between the photograph and the scene viewed by the camera have been explained.The utilisation of these relationships to measure distances by using a single photograph and stereoscopic pairs of photographs have been noted.

Special attention has been drawn to the analytical approach,with the determination of the equations establishing the relationship between the scene and the image.Finally the utilisation of the perspective transformation ,for determining distances in street scenes,is fully described and illustrated by figures 7.9,7.10,7.11,7.12, and 7.13.

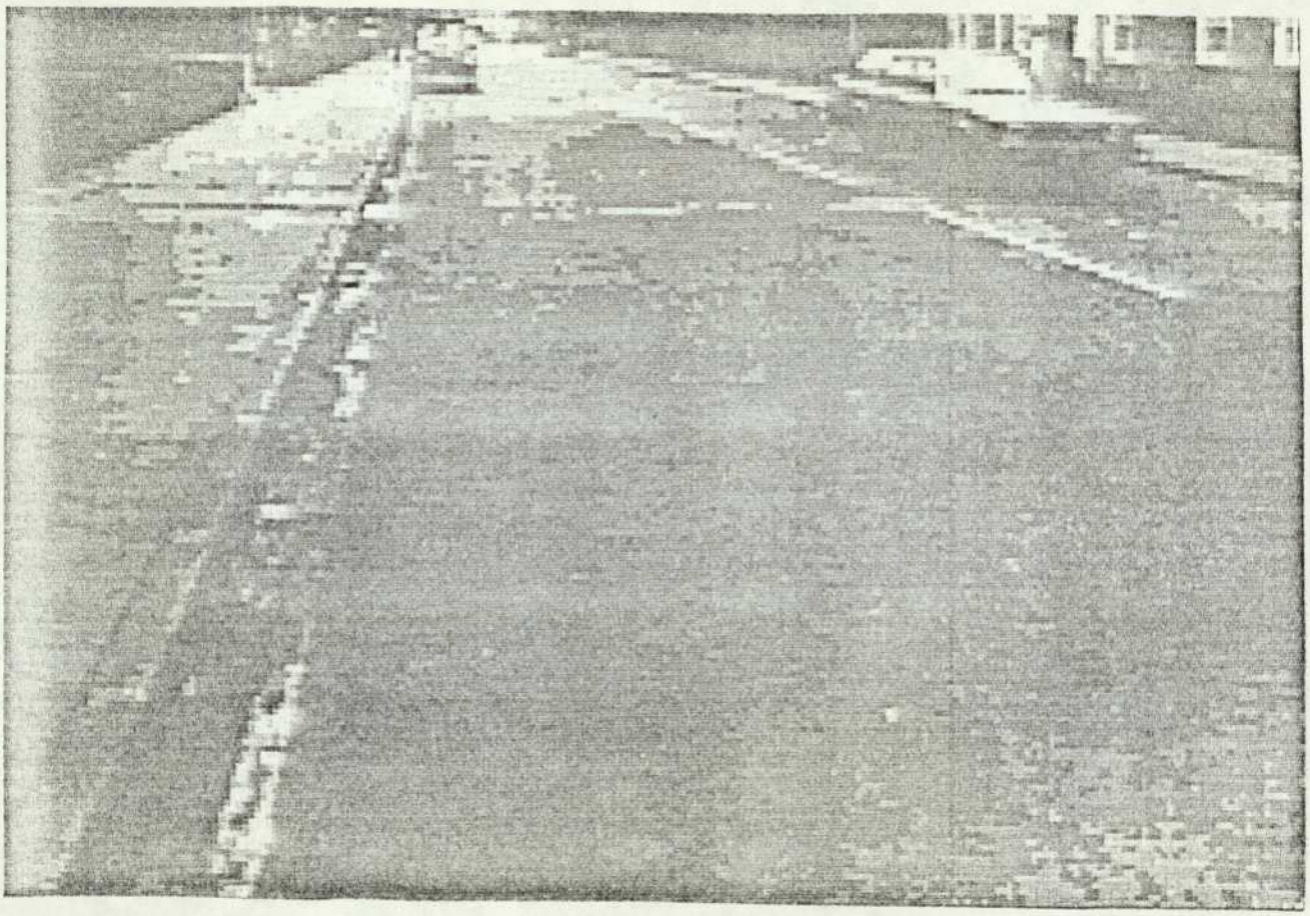


Figure 7.9 : Original street scene image.

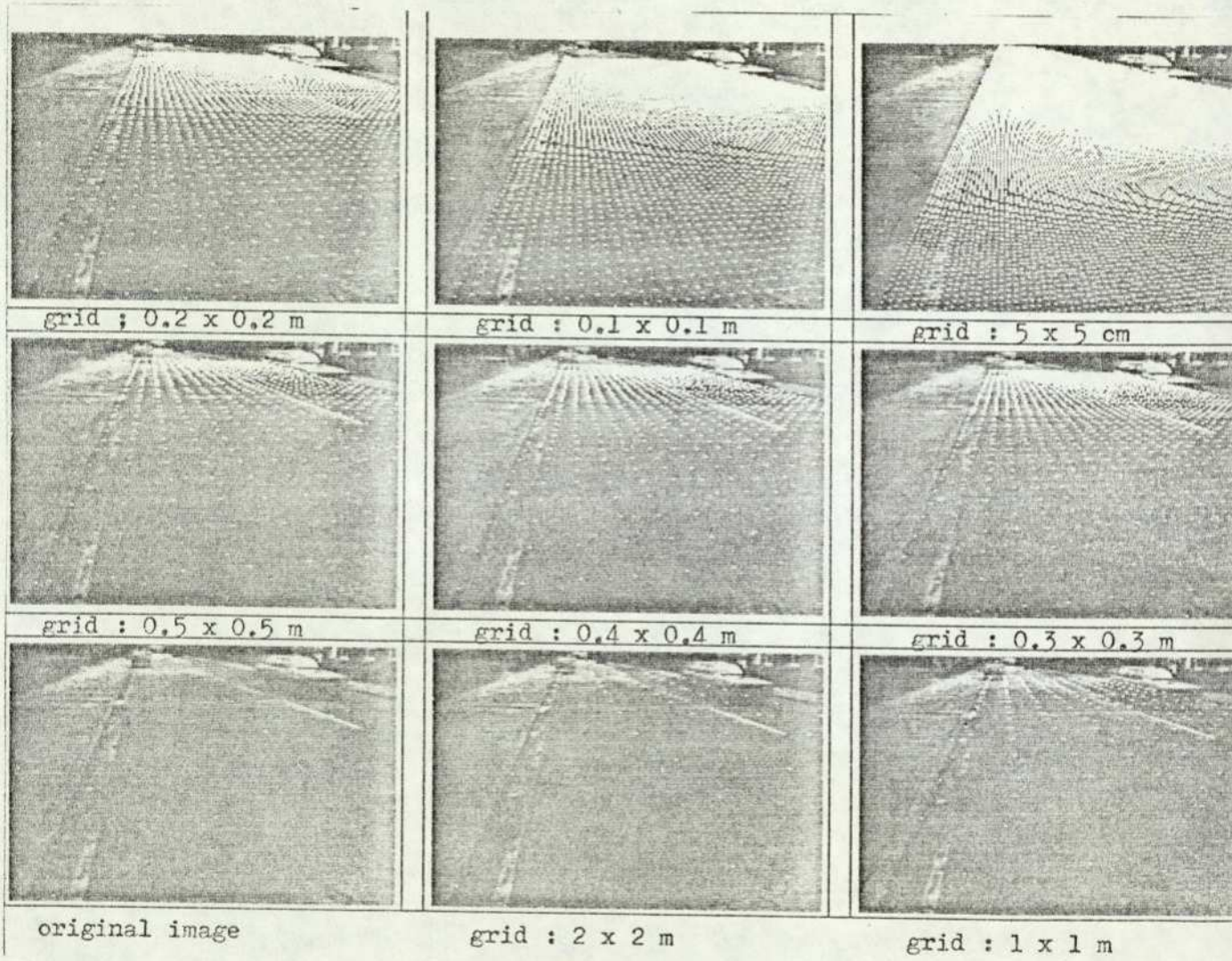


Figure 7.10 : Images with variable grids on the street.

Figure 7.11 : Image of fig. 7.9 with grid (1x1 m).

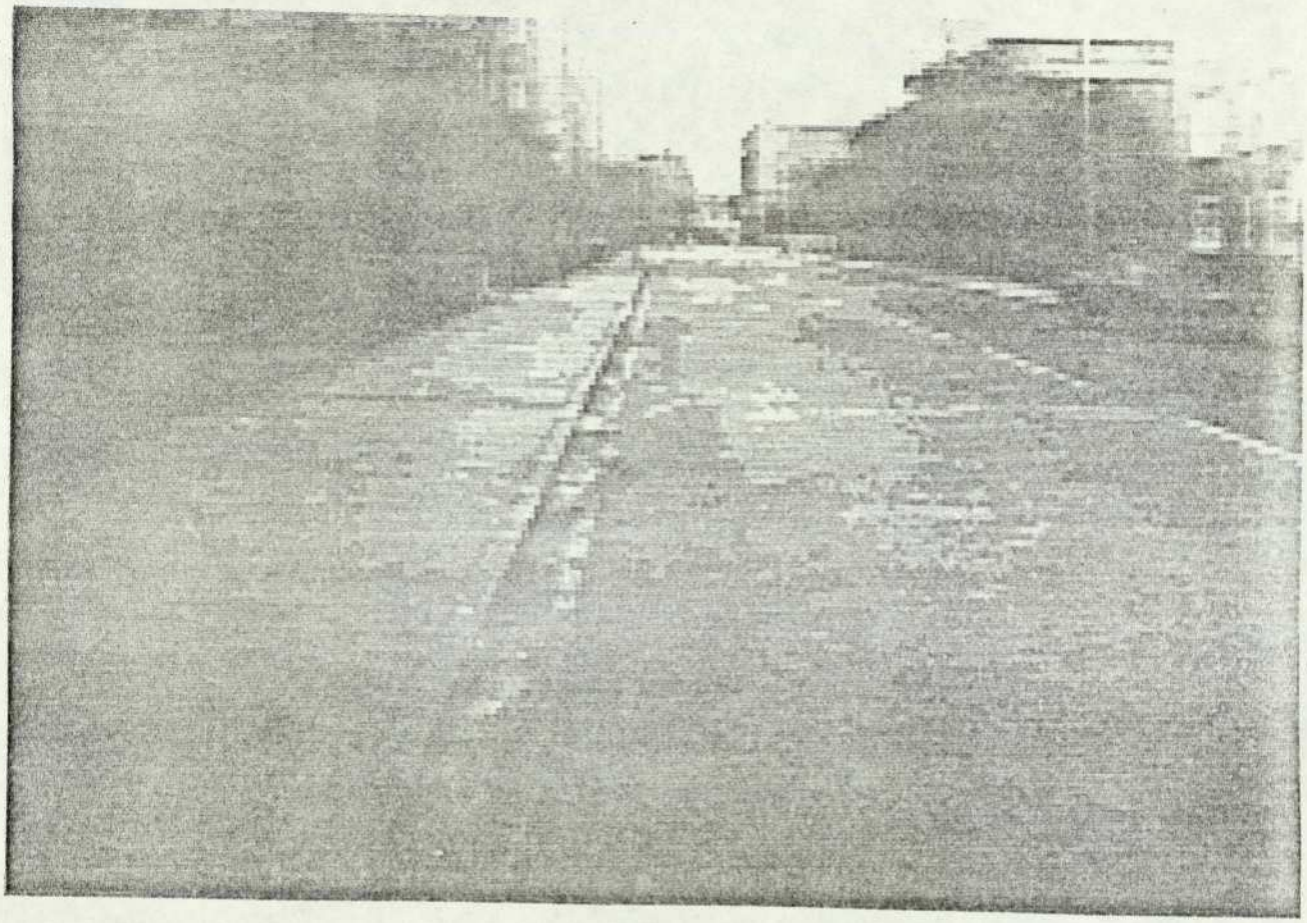


Figure 7.12 : Original street scene image.

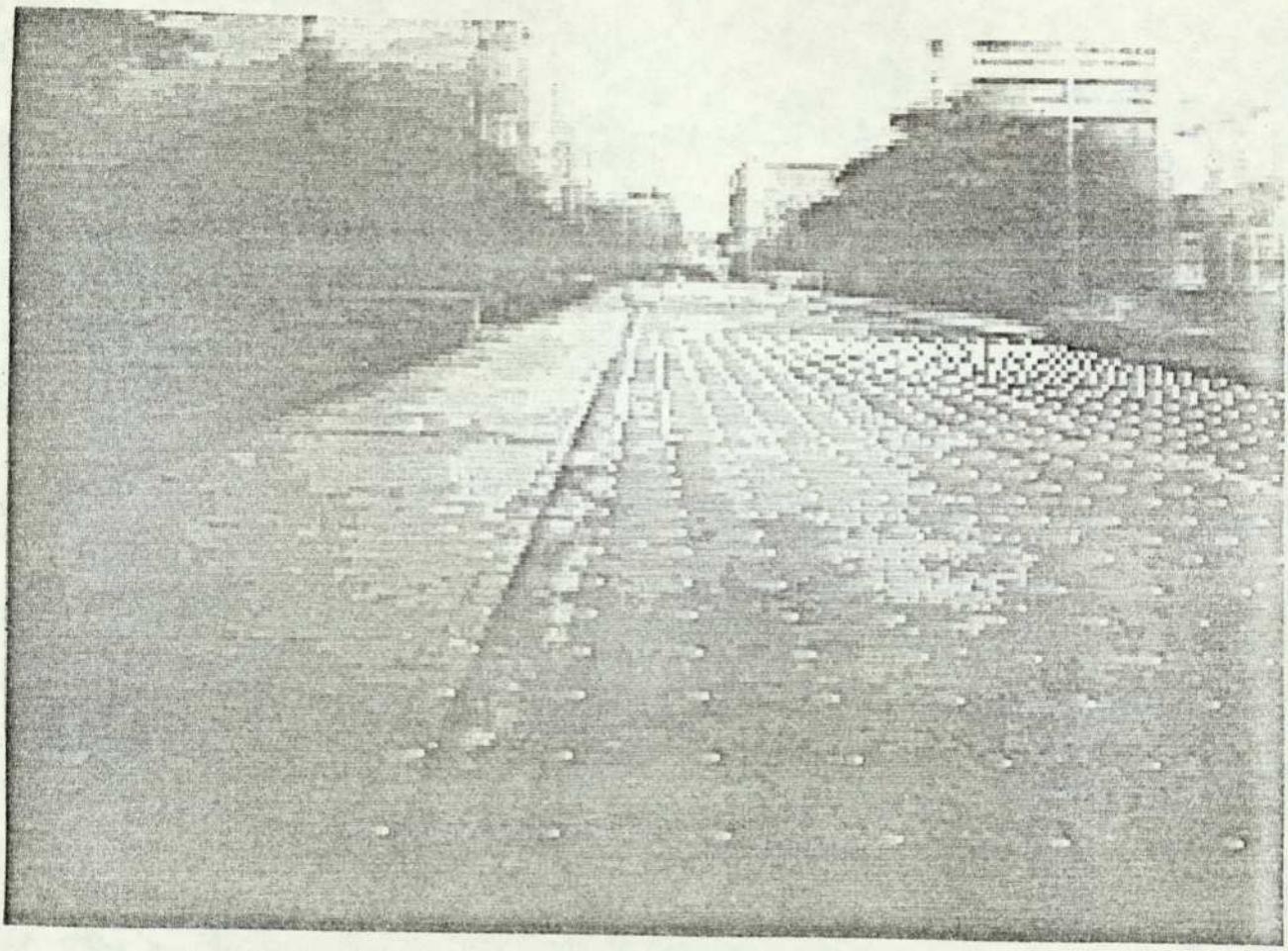


Figure 7.13 : Image of fig. 7.12 with grid (0.5x0.5 m)

8 LOCATION OF OBSTACLES

8.1 Introduction

After having located the boundaries of the street,and thus isolated the street from the rest of the street scene,an important task is to locate all the obstacles in the street,or at least all the obstacles which lie in a critical zone.The obstacles are located randomly in the street.Because the occlusion problem can greatly complicate the problem of location,we restrained our research to street scenes which contain obstacles the images of which are isolated and do not overlap.

Since precise outlines are not necessary to locate the obstacles in the scene (as opposed to identifying them) a simple procedure could be used to find them in order to speed up the processing.Although,in this research,we will concentrate mainly on the location of obstacles,it is worthwhile noting that,when all the obstacles inside the street have been found,each of them could be analysed separately using a high resolution image of the narrower region containing the selected obstacle,for the purpose of identification and classification.

This chapter describes the final part of the research which concentrated on the development of methods of locating automatically particular objects that are inside the street and which are relevant to the guidance problem. The different methods are transformed into algorithms which are programmed using Fortran.The processing was

carried out,off line, using the PRIME interactive system.

The first approach,which was tried,involved assuming that the obstacles lie on a background (road surface) defined by a specific range of grey level. The points,which have grey level outside the range of grey levels of the background,were considered to belong to the obstacles.

Because the important factor in driving is the determination of the nearest boundary of the obstacle,so as to avoid it,a second approach based on edge detection was also investigated and eventually adopted.Given the kind of street scenes,which were used in our research, the front or the rear of the obstacles,which ,in this investigation,were limited to vehicles,were delineated by horizontal lines.This second approach involved the detection of horizontal lines in the images and the development of a procedure which will determine among these horizontal lines the ones which delineate the rear for the lead vehicles,and the front for the vehicle coming towards the controlled vehicle.

However,before describing the different algorithms which can be used or were used to locate the obstacles in the street scene,we will first try to analyse the different factors involved,and the different information required in the longitudinal control of the guided vehicle.

8.2 Longitudinal Control In Driving

In all types of locomotion,collisions with various obstacles lying on the planned trajectory,are avoided by changing the direction of the motion or by stopping the vehicle.Driving is a special form of locomotion where the driver ensures that the path of the vehicle coincides with the part of the road which is not occupied by other objects(vehicles,humans,animals,or any object lying on the street).

The control of the vehicle is achieved by the variation of two variables of the motion of the vehicle,which are the speed of the vehicle and the direction of the motion.The speed of the vehicle is controlled by using the throttle for acceleration,and the brake for reducing the speed and halting.The directional motion of the vehicle with respect to the road,which is usually determined from visual information, is controlled by turning the steering wheel.The direction of the vehicle is determined by the angle through which the steering wheel is turned.The mathematical function relating the direction of the vehicle to the angle,through which the steering wheel is turned,can be determined for any vehicle.

The possibilities of automating traffic systems have been considered but up to the present there seems to have been more attention devoted to the problems of following the lead vehicle with steering and distance between vehicles automatically controlled.The

first step for the automation of driving is the automation of the avoidance of the vehicles and other obstacles which are in the path of the controlled vehicle.For a number of years,many people concerned with urban traffic systems,have done work on devices and systems which can monitor the location,speed and direction of vehicles operating in urban environment.This research was not directed towards the automation of driving,but was primarily concerned with gathering data which can be used to improve urban traffic.

The data about the location and status of the vehicles is gathered through a collection of electronic and electromagnetical devices.The various types of vehicle location techniques can be classified in two groups,namely,on-board locators,where the necessary devices for determining a vehicle location,with respect to a reference coordinates system,are contained within the vehicle, and off-board locators,where the devices for location are all or in part external to the located vehicle. In the following section we are going to review in more details the different types of detectors which are available for vehicle location.

8.2.1 Detectors

Because of traffic problems,there has been,since the late 1920s,a need for automating the gathering of the data necessary for traffic analysis,which involves vehicle location.A wide range of physical phenomena,suitable for vehicle location,including acoustic,radio frequency,optical and magnetic phenomena,have been thorouly investigated.

The vehicle detectors,which do not require aditional equipment in the vehicle to be located,range from simple detectors,which indicate when a vehicle passes through a selected point (speed can be measured by the elapsed travel time between two detectors longitudinally displaced in the line of travel of the vehicle),to detectors which lend themselves to direct speed measurements.

Many detectors involve the beaming of an electromagnetic or acoustic waves on an area of the roadway,and the detection of the reflected energy by the vehicle. The acoustic devices use ultrasonic waves.The doppler effect,which involves sensing the shift in the frequency of a tone that occurs when the tone is reflected from a moving vehicle,can be used for the detection of vehicle motion.

Although these detectors can be used for vehicle

location,they are expensive. Because the main aim of this research is to concentrate on visual information extraction by scene analysis,none of these detectors was used,and the location of the vehicles was achieved by analysing the street scene containing the vehicle. However,it must be noted that the utilisation of such detectors,in the future, can help to improve the performance of the system.

8.2.2 Longitudinal Control Of Vehicles

The longitudinal control of the vehicle is determined by the frontal obstacles such as other vehicles,people or animals crossing the road,and traffic signals.The control is achieved by steering in such a way as to avoid collision with the different obstacles,and to keep the vehicle moving into the middle of the safest path.

A system for automatic longitudinal control is required,in any system for automatic driving.Two general types of strategy for automatic longitudinal control can be adopted.The first strategy relies solely on the state of the controlled vehicle with respect to the nearest lead vehicle.The second strategy depends on the state of all vehicles which are in the field of view.

Research on the first type of system was started in the late 1950 s at the General Motors Corporation,where an automatic system for speed control and variable spacing was examined.a large amount of theoritical work has been done on the optimum control of a string of vehicles.For example,Levine and Athaus (1970) investigated the optimum control of a string of high speed vehicles moving at relatively large separations,and Levis and Athaus (1970) investigated optimum sampled data control for such a string.

Any automatic longitudinal control system for a vehicle must be able to operate in the open-road,where the controlled vehicle is not close to other vehicles, and the denser urban traffic,where the contolled vehicle is influenced by the behaviour of nearby vehicles.Some of the requirements of such a system is that the control system must not be required to respond so as to exceed the response capabilities of the vehicle,and that the average separation between adjacent vehicles must not be excessive.

In order to achieve satisfactory longitudinal control,we must have a means for measuring the headways and relative velocities between the controlled vehicle and the other vehicles which are inside a critical zone.There are a number of techniques based on electromagnetic and acoustic signals,which can be used to measure the headways and relative velocity. For example Ford Motor Company used a narrow-band suboptical frequency beam which

emanates from the controlled vehicle. When a lead vehicle is within the zone of influence of the beam,it reflects the signal back to the controlled vehicle.The reflected signal contains information about the distance between the controlled vehicle and the lead vehicle,and the relative velocity of the lead car.

In our research,the necessary information for the longitudinal control of the vehicle is extracted from the street scene.First the obstacles lying inside the road are located,and then the distance map which is obtained, by using photogrammetric techniques as described in the previous chapter, is used to calculate the distances between the controlled vehicle and the different obstacles.The relative speed of the moving obstacles (vehicles) can be estimated from the variation of the distances in successives frames. As,in this investigation,we are only concerned with the extraction of information from a single street scene image,we will not investigate the determination of the speed,but we will concentrate on the determination of the distances.

An important factor in the longitudinal control of a vehicle is the minimum breaking distance required to stop the vehicle.This critical zone is the zone within which the controlled vehicle could stop when required.The size of the minimum stopping zone is dependent on the speed of the vehicle,on the condition of the road-surface and the wheels,and finally on the state and the nature of the breakes.Human drivers,even when

following a vehicle travelling at a given speed in identical surroundings,choose different headways.Although this headway,or minimum stopping zone,is an important factor in driving it is difficult to measure.

The psychological research done on human drivers determined that a human car driver first looks as far along the street as possible,then back to the vehicle he is driving,then again along the road in front,and so on alternatively. At a speed of 25 Km/h a human driver finds it necessary to spend 60 per cent of his time looking 50 m ahead,and thus anticipates by 7.5 s,and 80 per cent of his time looking 30 m or more ahead,and thus anticipates by about 5 s.It also seems that the human drivers uses his central vision for obstacle detection and his peripheral vision as a guideline for general detection.This suggests that if we want to detect obstacles it would be advisable to use high resolution.This is one of the reason why we finally decided not to detect the whole obstacle but just the line which indicates where the obstacle start,and which is the nearest to the controlled vehicle.The above procedure,adopted by the human driver,also suggests that a system,which locate the obstacles,that lie about 30 m would be acceptable in this preliminary investigation.

8.3 Obstacle Location By Scene Analysis

8.3.1 Introduction

Obstacle location from street scene images is a preliminary step in the automation of driving using visual information.The problem of obstacle location, considered as an image processing problem,reduces to a segmentation problem for which many algorithms have been developed.In image processing terms,the obstacle location problem could be seen as the location of a connected subset of picture elements (pixels) in a larger set of data (image) subject to a given criteria.

Although,obstacle location,from a street scene image,appears to be an easy segmentation problem,once the street have been extracted from the rest of the street scene, many problems in the automatic location of these obstacle,by digital image processing,persist. The problems in the location of obstacles arise because of the noise in the image, because of quantisation anomalies,and because the grey level characterising the road (background)is not uniform as it appears to be.

The segmentation of one object from its background is a major part of image processing.In this part of the research,the

problem is more complex because it involves the location of not one object but of a number of objects randomly scattered on a background,with the objects being the obstacles,and the background consisting of the road surface. The location of potential obstacles in the street can be accomplished by using thresholding and edge detection.The obstacle location can be achieved by computing the background statistics,and then comparing every point in the image with the background statistics,and labelling as object points,the points that do not match the background.However this method is limited for monochrome images,and requires additional features,such as texture and colour, for accurate segmentation.

Many image segmentation techniques have been developed and utilised successfully for segmenting a variety of images.In the following part of this chapter,we will describe some of these segmentation techniques which can be used for locating obstacles in the street.

8.3.2 Segmentation techniques

The two important segmentation techniques are edge based and region based techniques.The region based techniques are based on global attribute of the image.They involve finding areas in the image over which one attribute ,usually grey level,or more

attributes,among which colour and texture,are constant. The edge detection based techniques depend on local attributes,and involve the detection of local discontinuities in grey level or texture of an image,and the construction of object boundaries from these edges. It can also be noted that an increasing number of methods use both edge detection based techniques and region based techniques for the segmentation of various images.

For both edge detection and region based techniques,perfect image segmentation is however rarely obtained.Although both methods have been,and are still being improved,at the current state of development their performance is limited in many cases.

The principal problem with the use of region based techniques is that certain false regions may be detected with the true object regions,or some part of the true object regions may be rejected depending on the range of the threshold. The choice of a suitable threshold for the segmentation of objects from a given background is difficult,if at all possible.This imperfection and difficulty of the threshold selection stems from the variation of the attributes of the object of interest,which is due to variations in scene illumination ,noise in the different sensors,and imperfections in the road.

The principal problem with the edge based techniques are the missing edges, and in noisy images,the false edges.The false edges are detected at points which are not part of region

boundaries.These false and missing edge make it necessary,for the formation of boundaries,to use tracking techniques which tolerate gaps in the edges,and which reject false edges.

8.3.2.1 Histogram Segmentation

Segmentation based on histogram-directed thresholding is adequate for various images,and remains the best method at the present. One major advantage of histogram segmentation is that it is computationally simple.The method involves histogram smoothing,mode searching,and threshold setting.It is based on the analysis of the histogram of the image.

In this investigation,which involves street scene images with obstacles randomly located on the road,the images comprises regions of interest composed of various obstacles,and a region of no interest (background region) constituted by the road surface. The prominent regions in the image correspond to the prominent modes in the histogram of the image,whilst the antimodes of the histogram correspond to the grey values that occurs relatively unfrequently and could be used as the thresholds which seperate the different regions.

For a great number of images the histograms contain

noise,which makes it difficult to identify the various antimodes of the histogram,and makes it necessary to filter the histogram.The noise in the histogram can be removed by smoothing the histogram so that the threshold for the segmentation of the image can be reliably determined by the antimode of the given histogram.

A main problem with histogram smoothing is that there are often too many regions segmented if the histogram has not been smoothed enough,and too few regions if the histogram has been smoothed too much.This results in the image either having important regions missing or many extra regions,depending on whether the histogram has been over smoothed or undersmoothed.

Another method of filtering the histogram for removing noise,involves mode sharpening. It consists of mapping the grey values close to the modal values into these modal values.The resulting histogram,which has a sharpened or spike like appearance, could then be used to segment the image.The same problems encountered with histogram smoothing,are also encountered with histogram sharpening.

8.3.2.2 Edge Detection Based Segmentation

Edge detection based segmentation is an important technique in image segmentation. But it often requires additional techniques to edge detection,for filling gaps and removing false edges.This thechnique has been described in great details,in the state of the art chapter.Hence we will just note that it is often used in conjunction with thresholding techniques (Ohlander (1975)).It is also worthwhile noting that the main advantage of the techniques based on edge detection is that,because region boundaries are naturally closed,it is suitable for shape description.

8.4 Vehicle Location In Street Scenes

8.4.1 Introduction

In this investigation,our main is to extract the maximum possible amount of information from monochrome images of street scenes,using the various techniques of image processing,so as to automate the guidance of vehicles. Although the complete location of obstacles inside the street could be possibly

included in a procedure for collision avoidance,it is not necessary. As a first step in the automation of vehicle guidance,the location of just the nearest points of all the obstacles would be sufficient to take avoidance actions such as reducing the speed or stopping.

We have described above methods for locating obstacles lying on a uniform background.However due to the the non-uniformity of the background (street), shown in figure 8.1,8.2,8.3,and 8.4,it would be difficult to determine the threshold values for this background.Therefore we decided to try another method involving smoothing.This first method,which was investigated, involved averaging picture elements comprised in a square of variable size. Taking a square with dimensions,respectively of 4x4,8x8, and 12x12 pixels,has the effect of oversmoothing the image,as illustrated by figures 8.5,8.6, and 8.7.Using this oversmoothed image,it is much easier to isolate obstacles inside the street.But unfortunately ,the oversmoothing of the image distort considerably the image of the obstacles ,which become hardly recognisable.Nevertheless,this method does not necessitate considerable processing,and could therefore be used to give an indication of the location of obstacles.A major problem with this method was the choice of a suitable threshold.

Because of the difficulty of choosing a suitable threshold,ad because the first method did not permit the precise delineation of the boundaries of the obstacles,we decided to

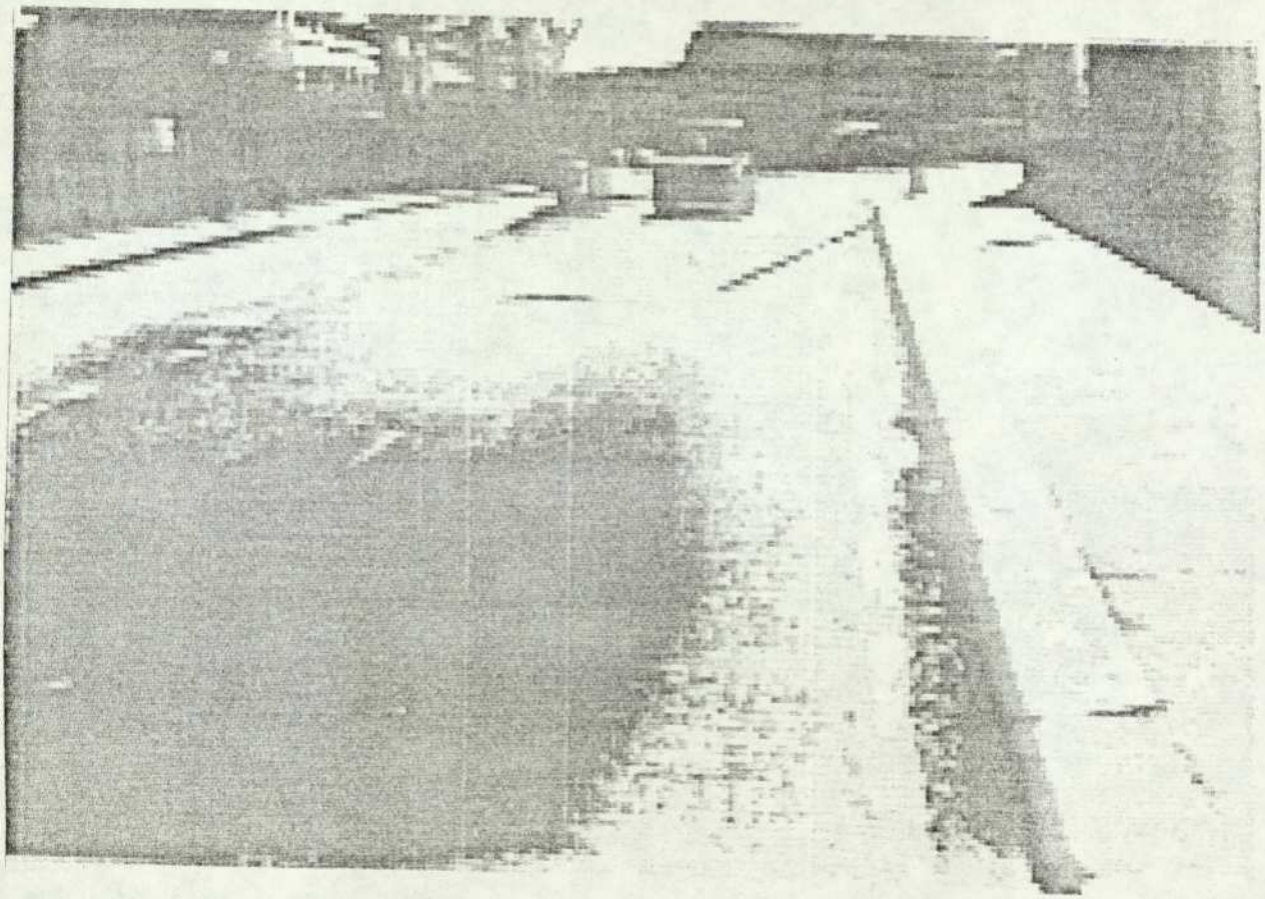


Figure 8.1 : Original street scene image.

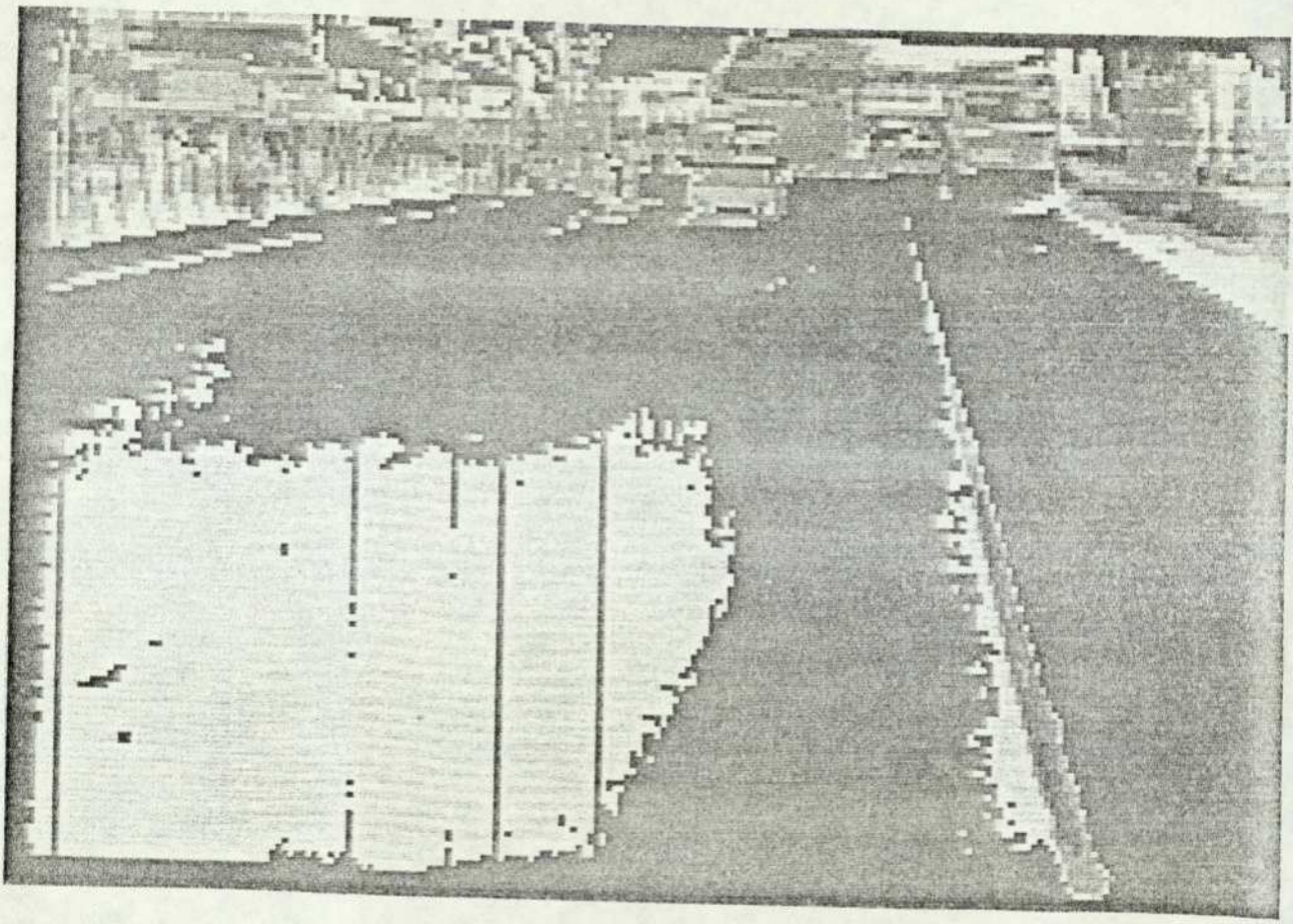


Figure 8.2 : Pixels of image in fig 8.1 having grey level values between 84 and 100.

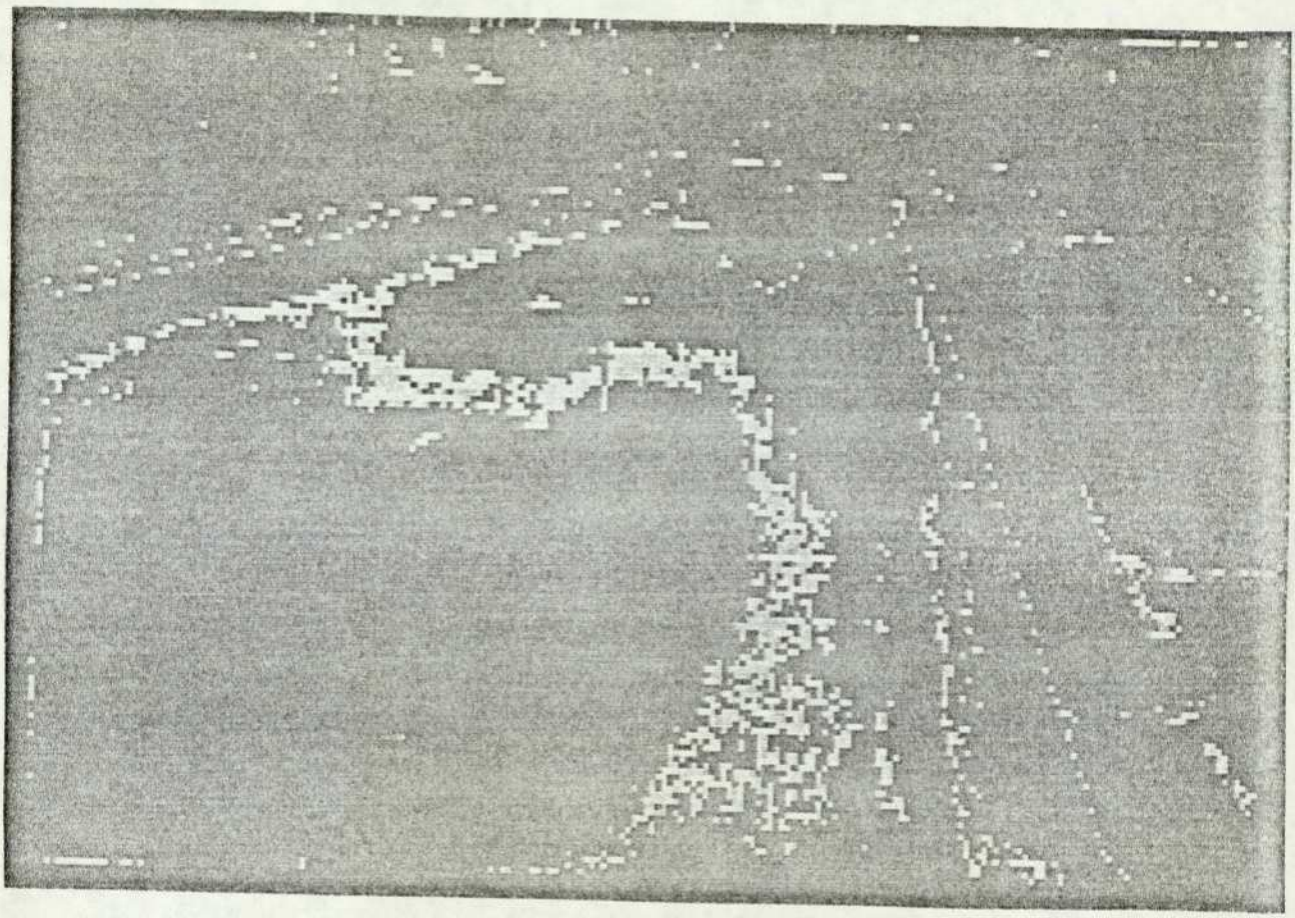


Figure 8.3 : Pixels of image in fig 8.1 having grey level values between 144 and 160.

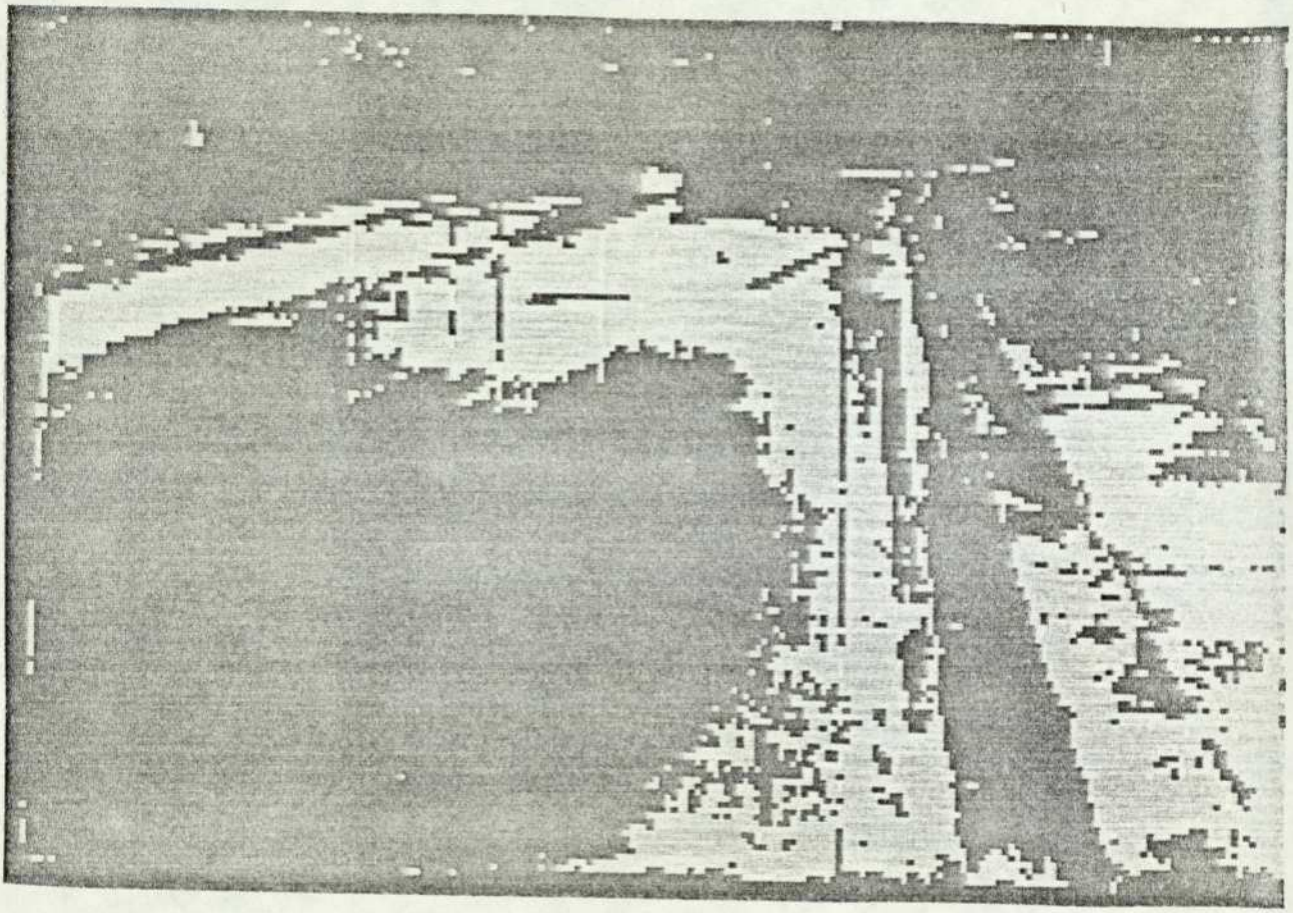


Figure 8.4 : Pixels of image in fig. 8.1 having grey level values between 164 and 180.

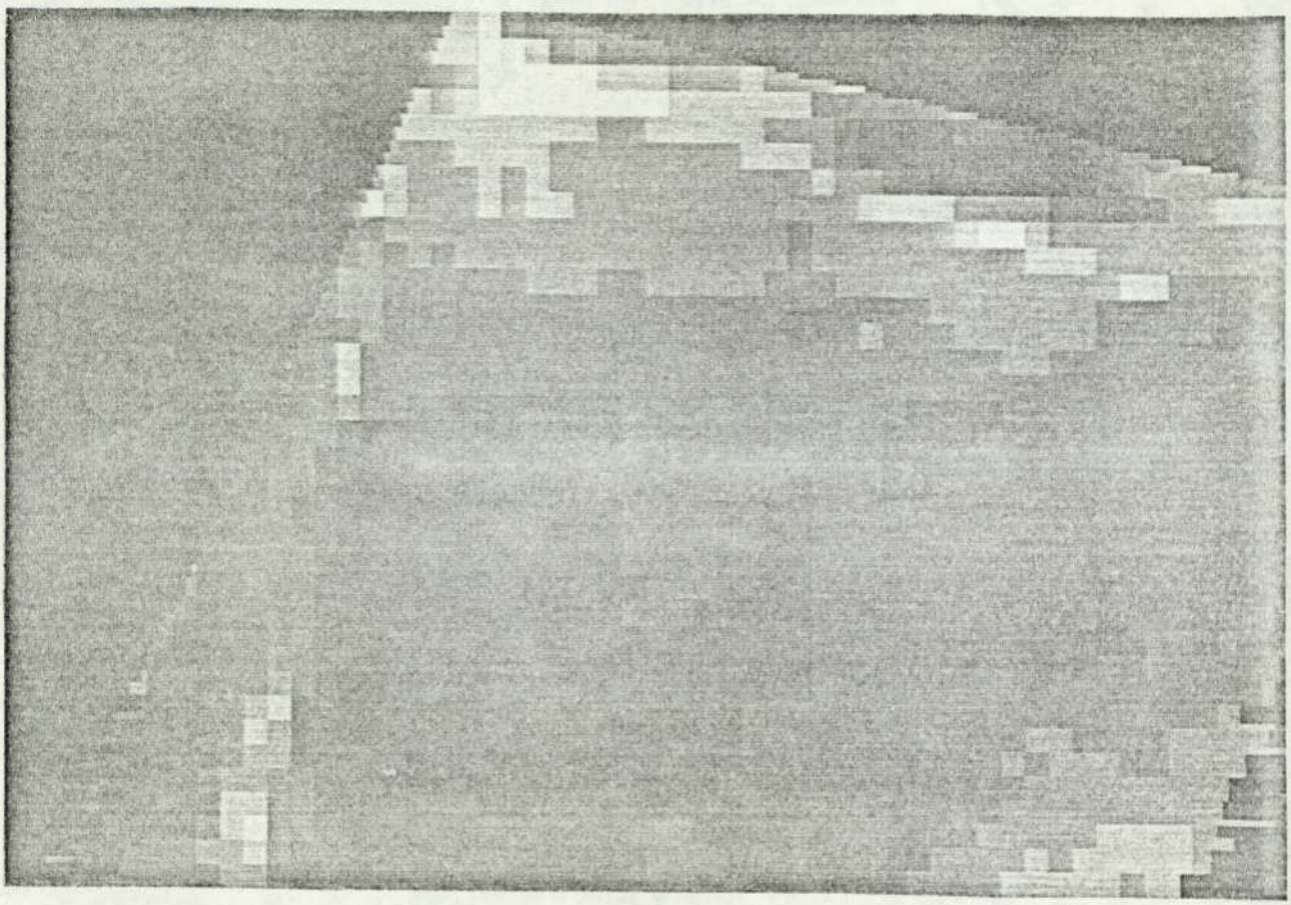


Figure 8.5 :Oversmoothing of image in fig. 8.13 with square of dimension 4x4 pixels

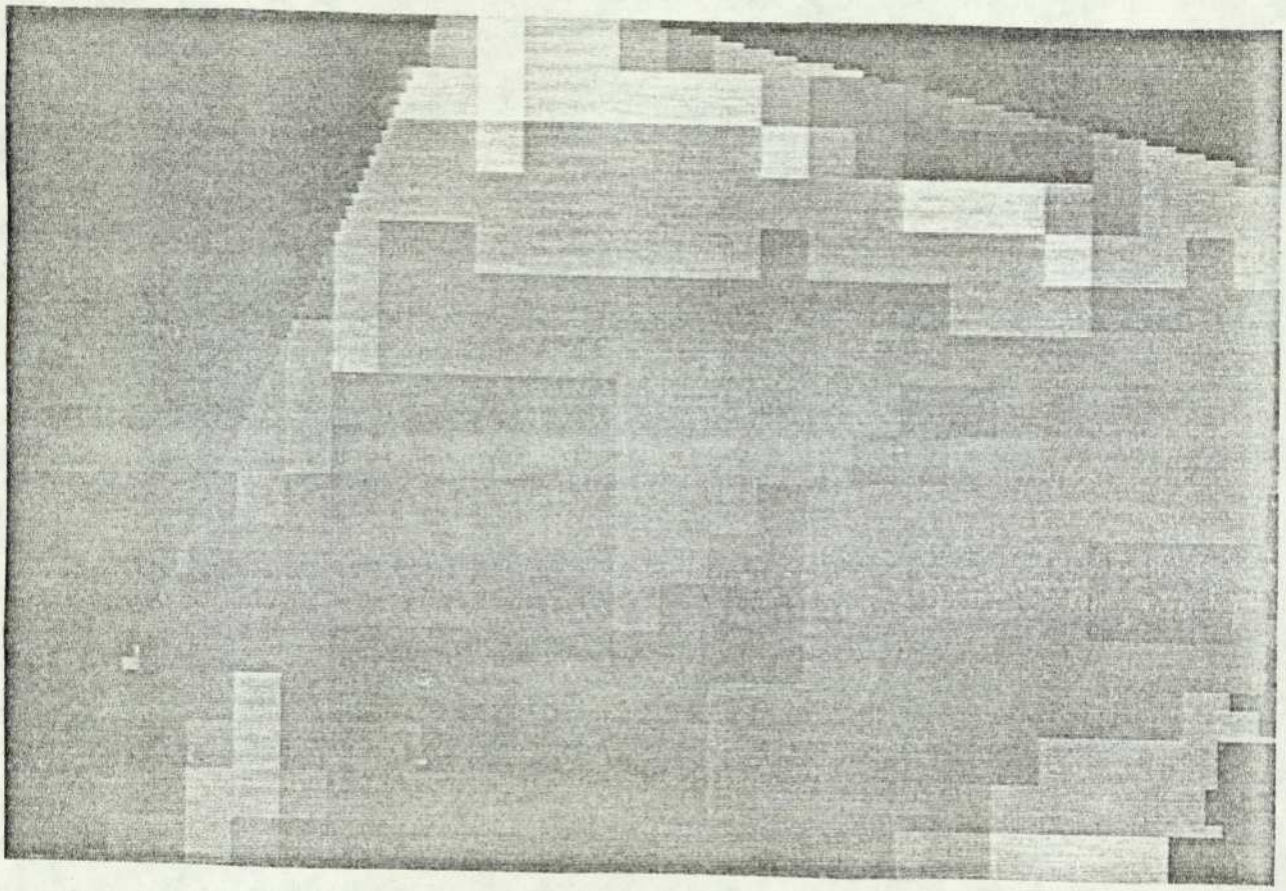


Figure 8.6 :Oversmoothing of image in fig. 8.13 with square of dimension 8x8 pixels.

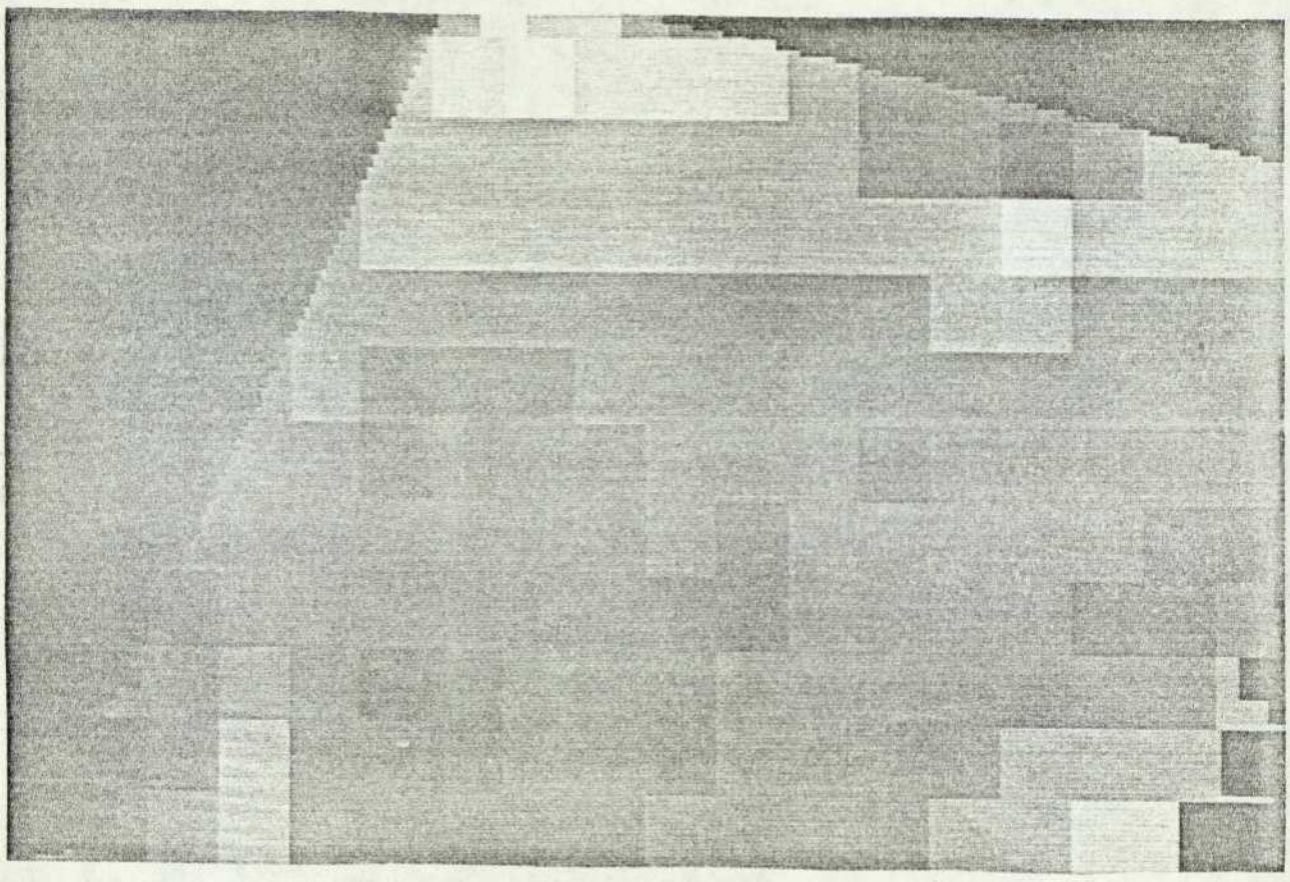


Figure 8.7 :Oversmoothing of image in fig. 8.13 with square of dimension 12x12 pixels.

abondon it,and to develop a new method based on edge detection.Because our main aim is to measure distances between the controlled vehicle and the different obstacles inside the street,it is not necessary to locate the whole vehicle boundary,but just the nearest points of the boundaries of all the obstacles to the controlled vehicle.Therefore we decided to locate just the lines,which indicate where the images of the obstacles start. Because of the special nature of the analysed street scene images,where the obstacles were only vehicles,the back of the lead vehicles and the front of the vehicles coming in the opposite way would be represented in the image by horizontal lines,as illustrated by figure 8.8.

The second method involved locating all the approximately horizontal lines (figure 8.9) in the image,and the development of a procedure for choosing amongst these horizontal lines the ones which represent the front or the back of a vehicle. To obtain these approximately horizontal lines,orthogonal masks,which were described in chapter 6,were used to determine the direction of the edges in the image (figure 8.10),and then the edges,which had a direction between 0 and 10 were accepted as points belonging to the approximately horizontal lines (figure 8.11).The determination of the nearest point of an obstacle to the controlled vehicle is sufficient for the evaluation of the distance from this obstacle to the controlled vehicle.

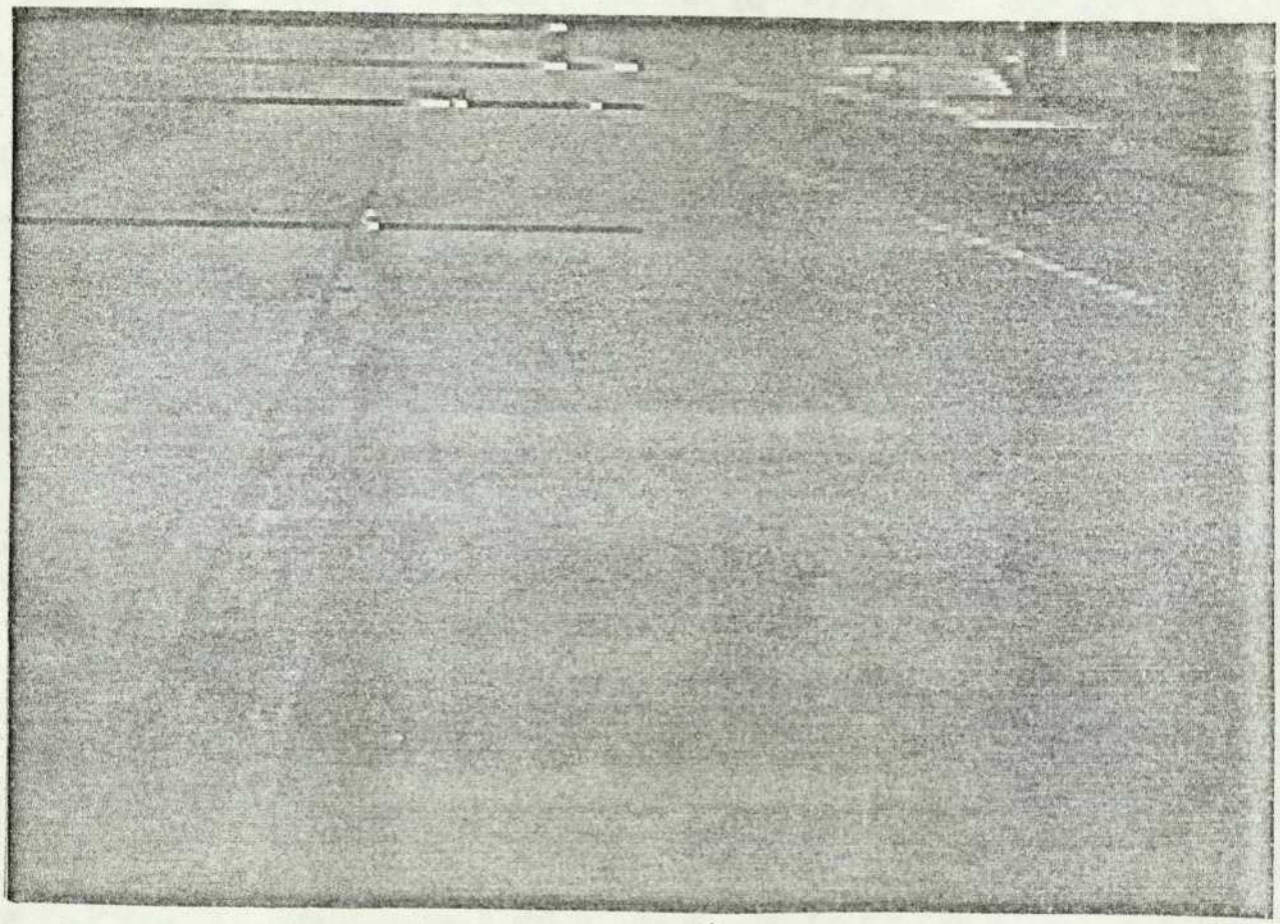


Figure 8.8 :Vehicles located by horizontal white lines.

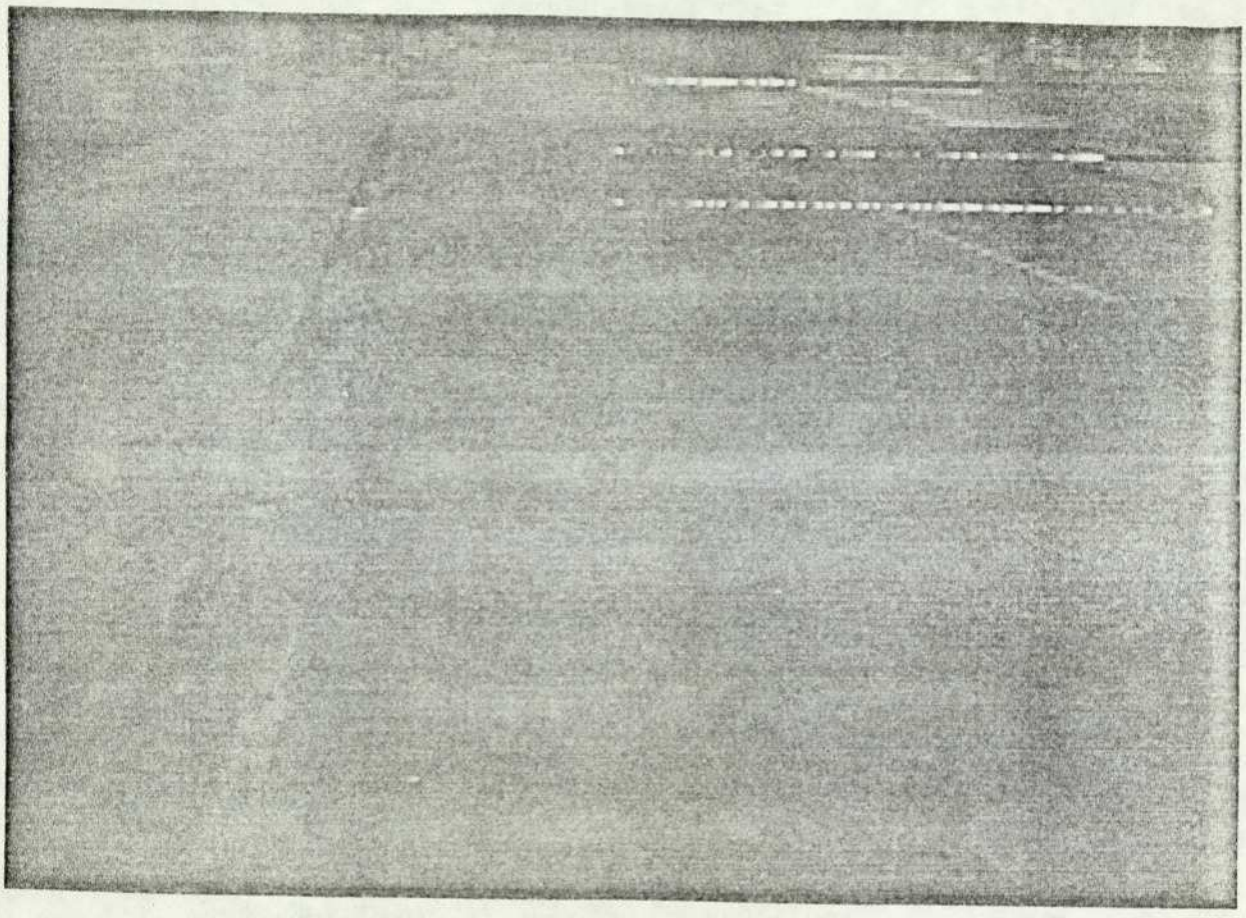


Figure 8.9 :Horizontal lines of image in fig. 8.13.

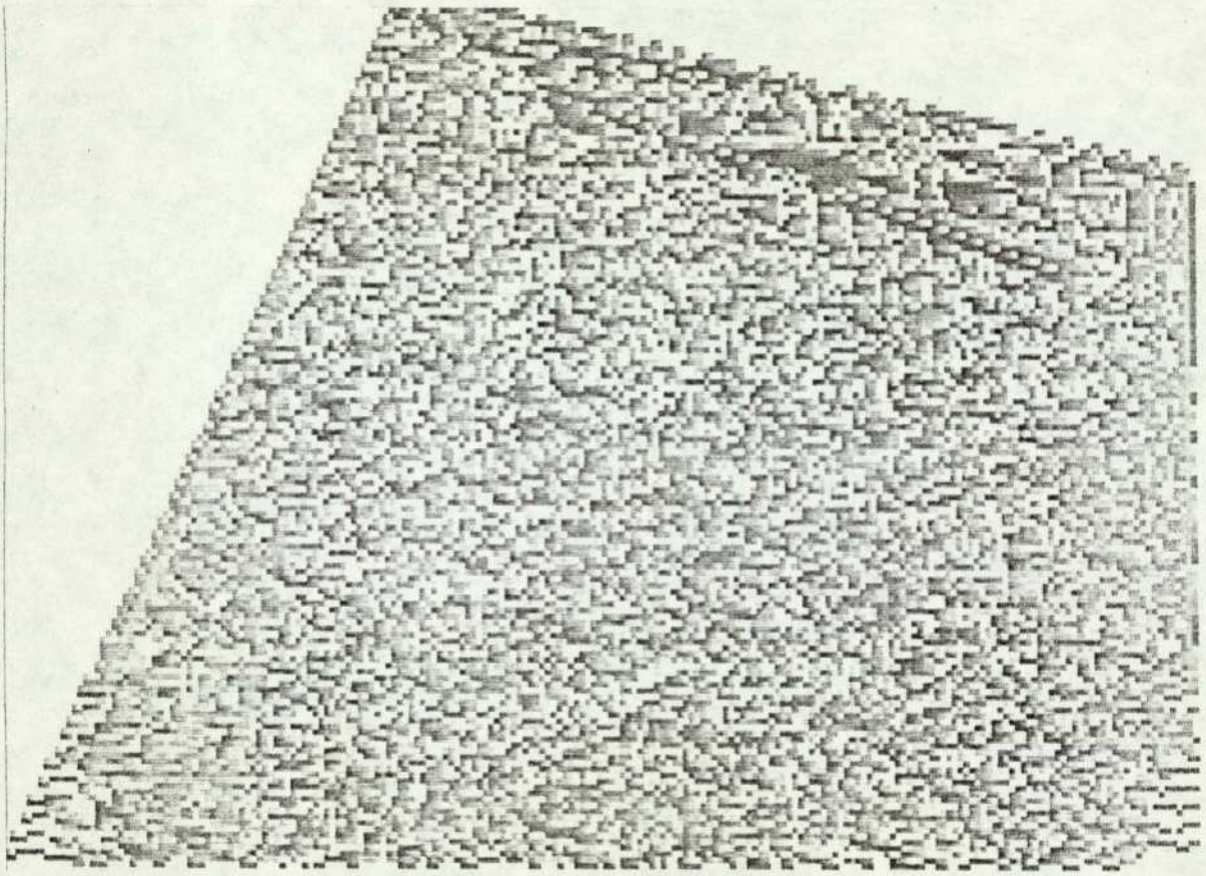


Figure 8.10 : Direction map of image in Fig 8.13.

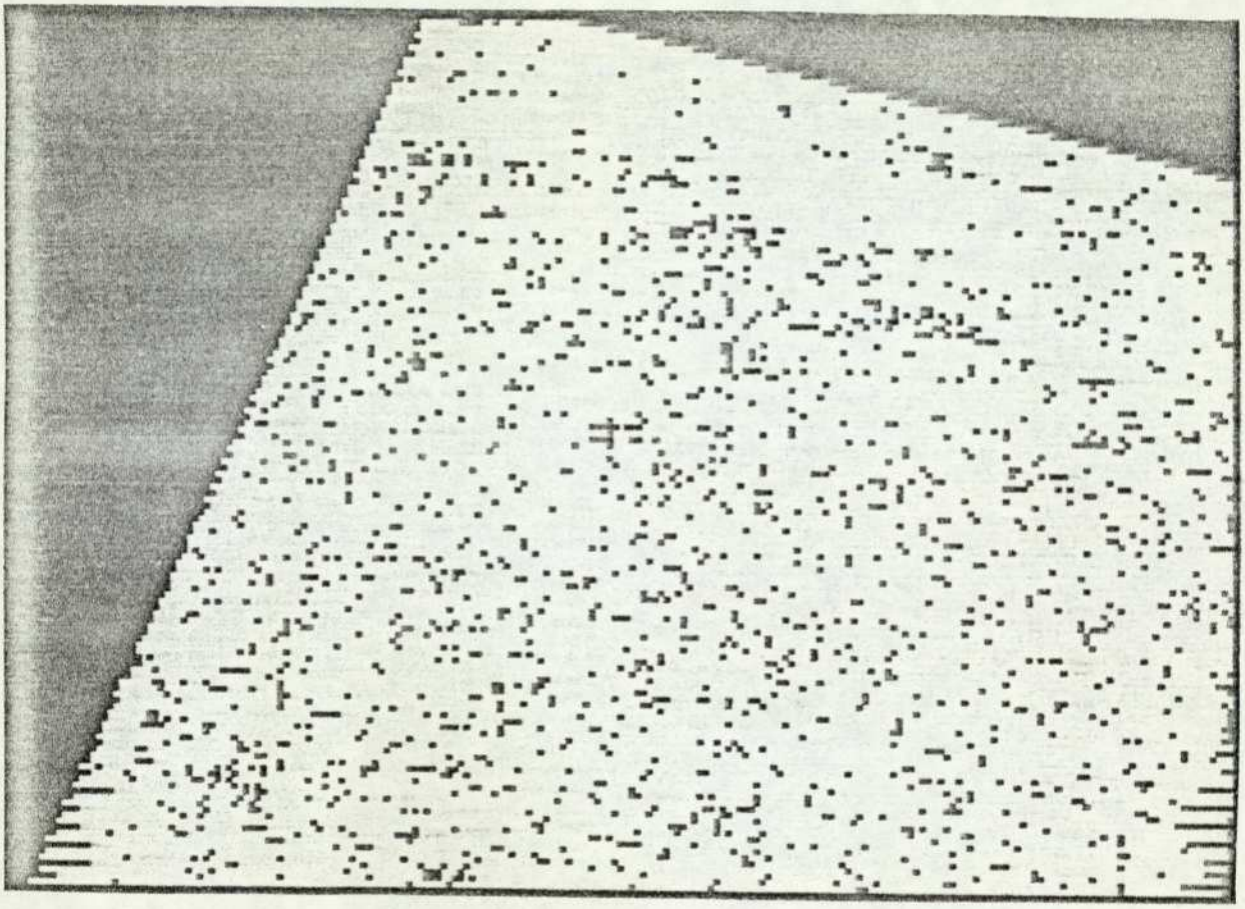


Figure 8.11 :Thresholded direction map of fig. 8.10.

8.4.2 Smoothing Method

In this part of the investigation we start with an image containing just the image of the road,with the grey values of the picture elements,which are not inside the road,changed into 0 (figures 8.12 and 8.13).Due to the noise in the image ,it is not easy to locate the obstacle by just thresholding as described above. Therefore we tried a new technique which involves smoothing the image,before trying to locate the different obstacles inside the street.

The image is divided into a variable number of squares,and the smoothing involved averaging the grey values of the picture elements comprised in each square. It was hoped that by varying the size of the squares,it would be possible to determine an optimal square size for the determination off each obstacle. However the adoption of such a procedure,which will determine the optimal size of the square for the location of each obstacle would be complicated and would require a lot of computer processing.Therefore,it was decided to choose the size of the square,so as to be of the order of the size of of the smallest image of all the obstacles inside the street.For example, for the image illustrated by figure 8.5,the square having the same size as the smallest image of all the obstacles is of the order 4x4.

Having chosen the size of the square,we then procede to

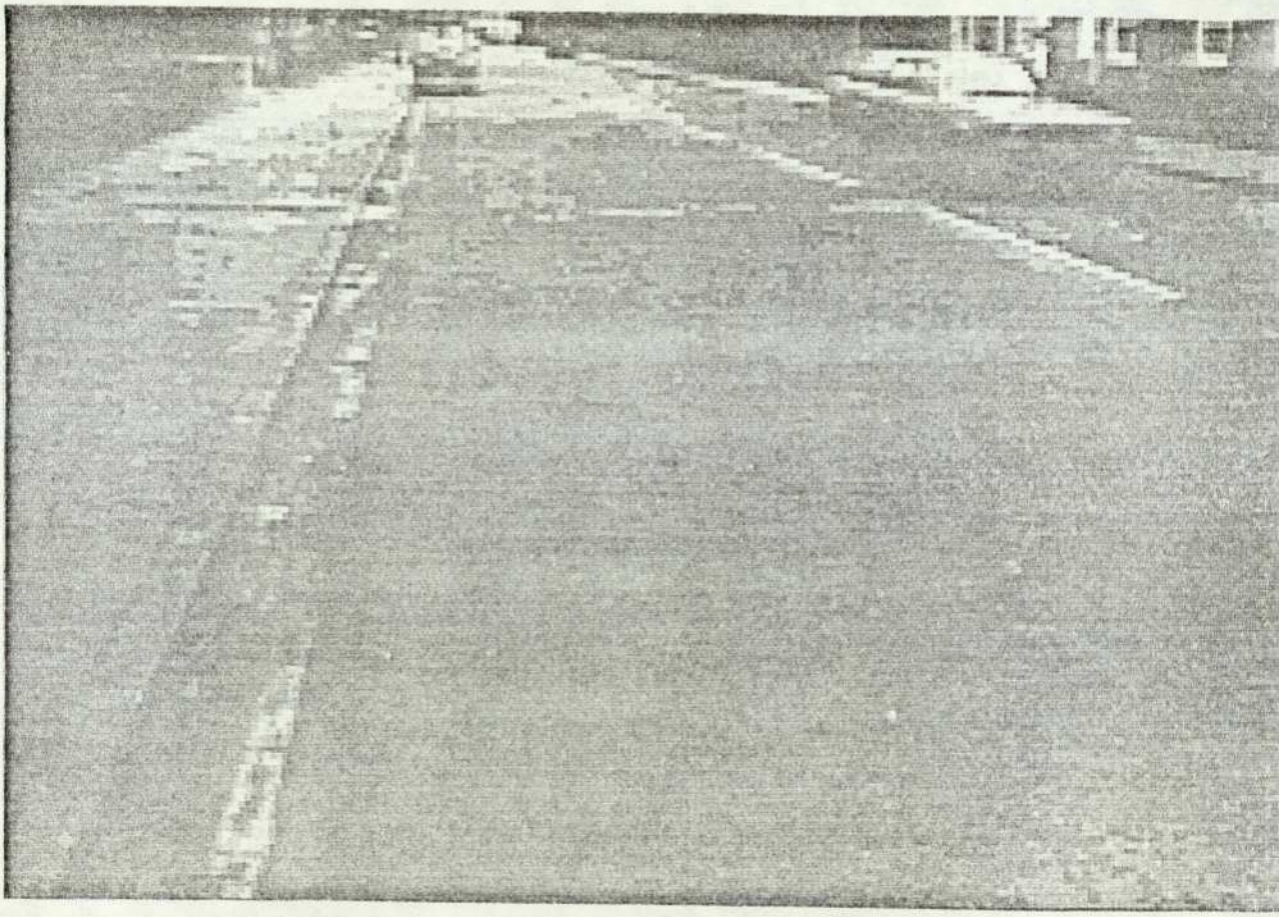


Figure 8.12 :Original street scene image.

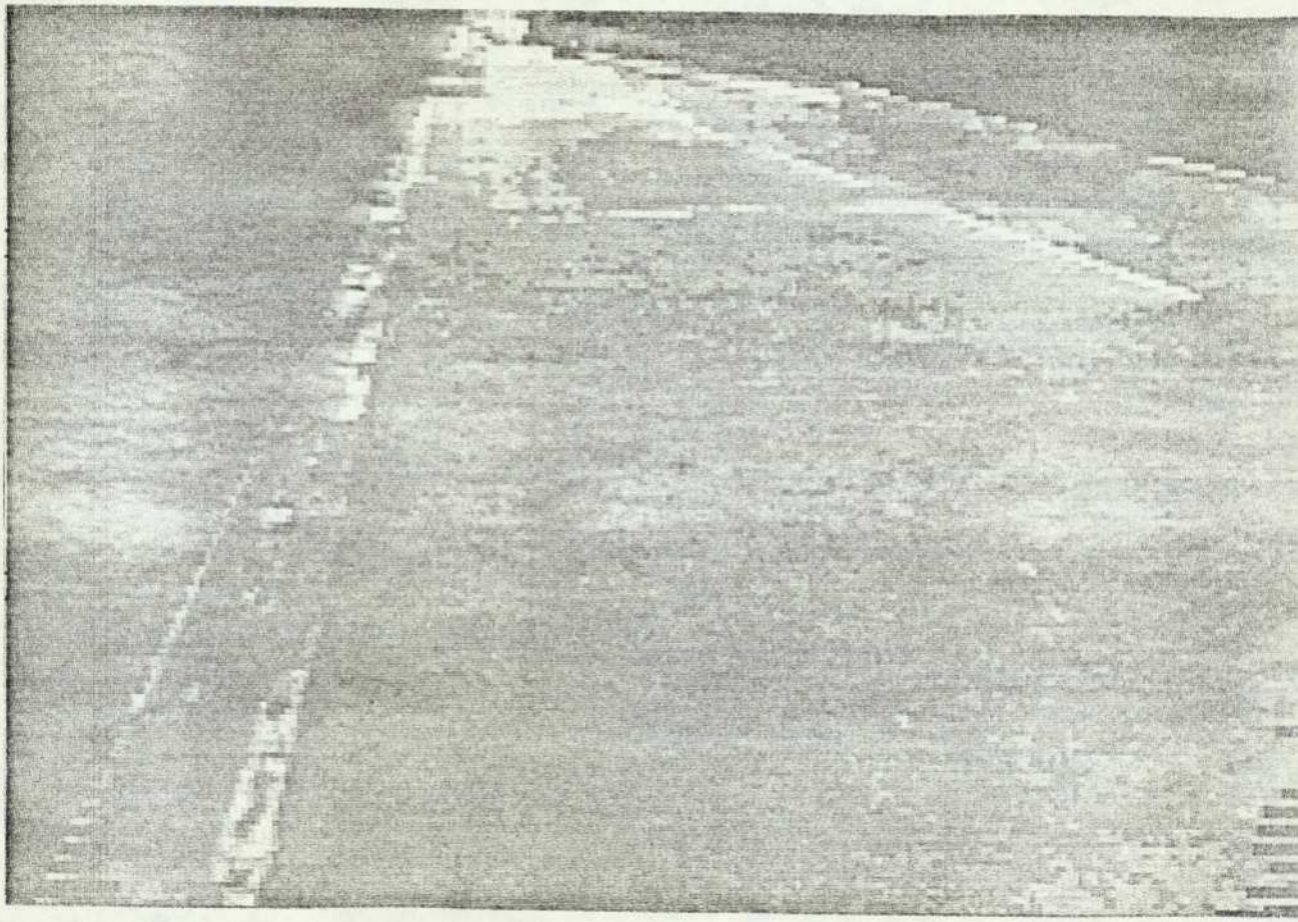


Figure 8.13 : Inside of the street of image in fig 8.12.

locate the obstacles inside the street by trying to determine homogeneous regions in the image,which could be possible obstacles.

Although this method was computationally easy,its performance was not satisfactory.The main problem with this method was that it does not permit the determination of the contour of the obstacles,which are necessary for the determination of the distances between the controlled vehicle and the different obstacles.A second problem with this method is that oversmoothing could completely destroy the images of very small,or far away obstacles.Another major problem is the determination of the threshold.Although this method could be used to give an indication of the region where the obstacles are,it was abondonned and not used in the final algorithm,which was used to locate the obstacles so as to measure distances.

8.4.3 Edge Detection Technique

Because we encountered some problems with the determination of the threhold in the method involving the smoothing of images,we decided to try an edge detection method for locating the obstacles. The main aim of this research being the automation of vehicle guidance,which involves the determination of distances from the controlled vehicle to the different obstacles,makes the complete location of the obstacles not necessary.Just the

determination of the nearest point of an obstacle to the controlled vehicle is therefore required for the calculation of distances.This fact and the fact that all our obstacles are vehicles,which are man-made objects whose images comprise many straight lines,make it possible to develop a simple algorithm for locating the obstacles.

As for the previous method,we start with the images containing just the inside of the street,with the rest of the street scene picture elements being replaced by dark pixels (picture elements with 0 grey value).We then procede to detect the edges.The edges are detected by using orthogonal masks fully described in chapter six.

```
        ! 1  2  1!              ! 1  0 -1!
   X =  ! 0  0  0!         Y =  ! 2  0 -2!
        !-1 -2 -1!              ! 1  0 -1!
```

Orthogonal Masks.

These orthogonal masks permit the determination of both the magnitude and the direction of the edges.Having detected all the edges in the image,and using the direction map we detect all approximately horizontal lines (figure 8.9).Amongst these approximately horizontal lines,there are lines which represent the backs of the lead vehicles or the front of the vehicles

coming in the opposite direction to the controlled vehicle.

After having located all the approximately horizontal lines in the image,the next step consists of determining the lines which represent the nearest boundary lines of all the obstacles.The technique is based on the fact that,because our obstacles were vehicles,they have at least two horizontal lines very close to each other.These two lines represent the boundaries of one of the vehicle bumper.By locating lines which have other lines very close to them,it was possible to locate the obstacles in many images,as illustrated by figures 8.14 and 8.8.

8.5 Conclusion

In this chapter,we described in some detail the longitudinal control of vehicles, and gave someexample of existing systems.We also described existing detectors for locating obstacle in traffic situations.Concerning this investigation the longitudinal control of the vehicle is determined by the calculation of the distances from the controlled vehicle to the different obstacle,by street scene analysis.When the controlled vehicle is very close to an obstacle it can reduce its speed,halt,or as a last resort change direction.

We then considered the problem as an image processing problem,and described ,in some detail,different method for locating

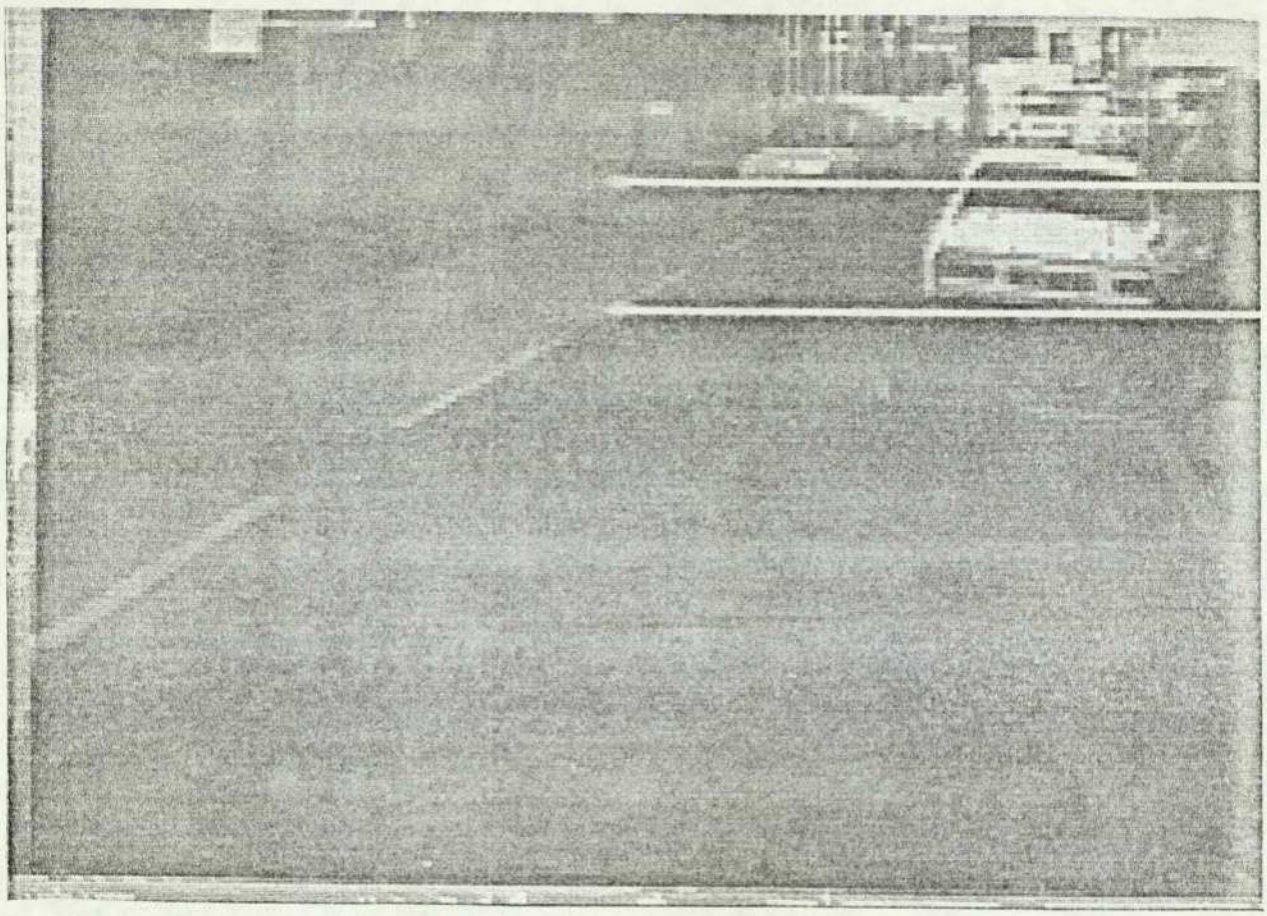


Figure 8.14 : Image with located vehicles (by white lines).

objects lying on a homogeneous background.Although those methods were not specifically used,they summarise the present different possible approaches for locating obstacles inside a street.

Finally we described the two approaches to solve the problem,which were investigated.Although the first approach not finally used,because of some difficulties with thresholding,it can however be used to determine approximately the area where an obstacle is located.A higher resolution of the area can then be used to locate precisely the obstacle.The utilisation of the high resolution would however considerably lengthen the processing.So we developed a second method based on edge detection.The method is described in some detail.Although this method was designed specifically to locate vehicles,it was felt that this is adequate ,because in many traffic situation the obstacles are mainly composed vehicles. The extension of the method for locating other types of obstacles could be attempted in the future,by improving the first method or trying new thresholding methods.

The results given by the edge detection technique are quite acceptable,as illustrated by figures 8.14 and 8.8,where the obstacles have been located.

Once the obstacles have been located ,the distances,necessary for the longitudinal control of the vehicle,can then be calculated by using the distance map described in the previous chapter.The determination of these distances could then be used to automatically

guide the vehicle.

9 DISCUSSIONS AND CONCLUSIONS

9.1 Discussions

Given the actual state of the art in image processing and scene analysis,it is not possible,yet,to replace completely human drivers by machine vision systems. But,even with just the available image processing techniques,in some particular cases ,the automation of visual information extraction from street scenes for automating vehicle guidance is a project which is feasible.One such particular case is the modern one-floor factory.With its spread-out layout, one of the critical problem in such a factory is to connect together all areas within the plant,by moving materials between different departments along the production line,and between receiving docks and the manufacturing areas. Although there are a number of alternative solutions to this problem such as conveyor belts or mounting the vehicle on rails,the most flexible would be an autonomous vehicle capable of operating in the semi-constrained factory environment.

The research,described in this thesis,has been motivated by the awarness of the need for the automation of vehicle guidance in space exploration because of time delays in signal transmission,and in underwater exploration and inspection(oil rigs),in material handling in radioactive or polluted environments (nuclear reactors,mines,...) because of the risks involved for the human operators. With the expansion of robotics,where,in its actual state of the art,the robot stands in a single station,a logical improvement in this field would

be to give them a flexible mobility.To be flexible an automatic guidance system has to get the robot where it is going by using the most direct route.This flexibility would require the vehicle to be selfcontained,and thus would not use external homing beacons or underground wire grid such as the one used by Bendix Compay at Aun Arbor in the USA.The roving robot should choose the sortest path and should not have to touch the obstacles surounding it.Although many detectors,such as the radar,ultrasonic,and laser can be used for gathering the information necessary for vehicle guidance,they were discarded from the initial project due to their complexity and cost. Thus all the information needed for guidance has to be gathered from visual data.

The advance in robotics,and machine vision in particular,make it possible to examine the problems associated with the new automation capabilities to city traffic.Investigation in many related areas,such as the development of techniques for automatically guiding unmaned vehicles for use in hazardous environments, and the investigation of highway guidance techniques,are underway.In order to obtain user acceptance of a fully automated vehicle,it is necessary to retain the major feature of the personnal vehicle: flexibility.For maintainning this flexibility,autonomous traffic control may be divided into two functions: vehicle routing and vehicle guidance.The vehicle routing function selects the overall path for the whole trip from begining to end.The vehicle guidance function ,which is the main concern of this research,ensures that the vehicle follows the desired path

safely without any collision.This division of the traffic control task imposes a hierarchical structure where a central large off-line computer can be used to solve the stochastic routing problem,and a smaller on-board computer can be used for real-time vehicle guidance by processing in situ the sensor data.

The routing and guidance problems can be approached independently.The routing problem has many other applications,such as communication,warehousing, and railroad,and there has been an extensive body of research in the field which can easily be adapted to traffic control.However little research has been done on vehicle guidance.Many of the investigations on automatic vehicle guidance are aimed at maitaining control over a precise fixed path and thus reducing the problem to an equivalent problem of railroad control.

In order to maintain the flexibility of urban traffic,a vehicle guidance system must allow path selection flexibility,must not require a high degree of precision,and finally must sense and respond to unexpected events.A solution fulfilling all these requirements is available ,and is the one based on human vehicle control.A human does not drive precisely,but rather drives relatively to external references,such as distance to the near obstacle,and the distances to the boundaries of the road.Thus,vehicle guidance in traffic does not require precision,but operates by using relative position of the order of two to three feet for lateral control,and ten to twenty feet for longitudinal control.Also,although identification of obstacles is desirable ,it

is not necessary for vehicle guidance.The solution of both problems of guidance and path selection will depend on the ability of the vehicle to sense (visually) its environment and to determine its position relative to that environment.

The research,described in this thesis,was concerned with the extraction of information from street scene monochrome images so as to guide a vehicle, in an urban environment,requiring no (or possibly only occasional) action on the part of a human operator.Although some research have been done for automating vehicle guidance in very restricted environment (Gennery(1979)), as far as it can be gathered from the various publications in the field of machine vision, there has not been any research for automating vehicle guidance in unrestricted urban environments by the sole use of visual data.

The research undertaken in the field of driving has been mainly concerned with establishing some criterion for measuring the performance of human drivers so as to compare them with each other. The need for measuring the performance of drivers similarly lead to the treatment of the human driver as a servomechanism,with the hope that a transfer function for the human driver can somehow be obtained,and thus giving a complete description of the driver's performance. Because the driving task (in urban situations) is too complex to be amenable to a simple servomechanistic description,the approach did not succeed.But the analysis of driving in terms of a control system,was not entirely fruitless,because the adoption of

such an approach implied tat driving can be automated.

Although drivers have often been concidered as elements in control systems (Mac Ruer and Krendel (1949)),no attempt was done to replace them by systems based on the processing of visual signals.It has even been suggested that the task of automating driving is too complex to be amenable to anything practical in the near future. This view was justified before the development of the field of scene analysis and image processing,but,given the actual state of the art in image processing, the automation of the task of driving is becoming a feasible project.

A system,capable of replacing a human driver,which has to extract the necessary information,for the automation of vehicle guidance,from visual data consisting of monochrome street scene images,reduces the probleme of driving to a problem in scene analysis which involves the adequate interpretation of the visual data.This interpretation involves finding a correspondence between elements in the real world and their two-dimensional representation within the camera image.For this to be possible the system requires a model of its environment which guides the interpretation.In this research the model, which represents the street scene,consist of a two-dimensional bounded surface (inside of the street) where obstacles (mainly vehicles) are distributed randomly.

During the research,a detailed analysis of the task of driving was undertaken,so as to define the necessary information for the

automation of this task.Driving is achieved chiefly by using vision.The driver examines his visual field and determines the optimal path such that obstacles are avoided and the destination ultimately reached.The analysis of the driver's visual field is selective in that the elements of the field which are pertinent to driving are attended to,whilst non-pertinent elements such as scenery,normaly receed into the background. The most important part of the region included in the field of interest is the road. Within the boundaries of the road lies a bounded surface which constitutes the region of safe operation,that should not be left by the vehicle. Within this region lie all the possible paths which the vehicle may take unimpeded. The possible paths are dictated by the presence of randomly distributed obstacles inside the road.At any given moment the set of possible path constitutes the safe region where the car can operate.

From the analysis of driving it appears that the important elements in the street scene are:

1- the boundaries of the road

2- the obstacle inside the street.

Having located the boundaries of the road and the various obstacle inside the street,the next step in the automation of driving is to evaluate the relative distaces (and if possible speeds) of the different obstacles with respect to the controlled vehicle. When the boundaries of the road have been located,and when the distances to the different obstacles have been calculated,the control of the vehicle will then consist of staying inside the road,of moving in a

parrallel fashion to the boundaries of the road (lateral vehicle control),and of taking avoidance action when the distance of the controlled vehicle to the nearest obstacle becomes critical (longitudinal vehicle control).

Because bends in the road require steering action on the part of the driver ,it is possible to regard them as obstacles.A simple model of driving will then consists of examining the visual field for obstructions and bends,taking the appropriate avoidance action by adjusting speed and direction,and continually monitoring the directional motion with reference to the boundaries of the road. The mechanism of obstacles avoidance can be considered in terms of the field of safe operation (Gibson and Crooks(1938)),or in terms of information and subjective risk (Hollingdale).

As far as our investigation is concerned,the problem have been divided into 3 independent parts.The first part of the problem is concerned with locating the boundaries of the road and hence extracting the road from the rest of the street scene.The second part is concerned with the development of a method based on photogrammetric techniques for measuring distances on the road so as to evaluate the distances between the controlled vehicle and the different obstacles inside the street.Finally,the third part involves the location of the different obstacles inside the street.The combination of these three parts constitute a complete algorithm for automating vehicle guidance.

In the first part of the research,the boundaries of the road were located by using various image processing techniques.The various image processing techniques involved in this research were:

1- Thresholding

2- Edge Detection

3- Hough transform.

Thresholding was used for data reduction.Edge detection was used for locating the various lines which constitute the street scene.Finally the hough transform was used for locating the boundaries of the street amongst the various other lines of the image.The boundaries of the road have been successfully located for five street scene images.The performance of the system can be considered to be very encouraging,because the street scene images,which were analysed,were taken with an ordinary camera without any special lighting,and thus contained a lot of noise.Therefore the performance of the system can be expected to improve, when a special sensitive camera is used.The utilisation of a higher resolution could also improve the performance of the system.The system works even when the boundaries of the road are partially hidden by various obstacles,and its performance could be expected to improve when only small parts of the road boundaries are hidden.The white and Zig-zag lines caused major difficulties,thus their abscence could also be expected to improve the performance of the system.In summary,The system as it stands,is quite an acceptable system,but it must be tested with a much larger number of street scene images before been considered for the final implementation in hardware.

In the second part of the research,a system for calculating distances was devised. The system was based on photogrammetric principles and involved knowing the orientation of the camera.As far as we know,this the first time photogrammetry has been applied to measure distances for vehicle guidance.The important feature of this part of the research is,that it can be applied directly for measuring distances on any flat surface.Another important feature of the system is that any desired accuracy can be achieved by choosing an appropriate resolution for the image.The performance of the system,which was achieved during testing was reasonably adequate for vehicle guidance.Because this system is completely independent of the other systems required for guidance,it can therefore be implemented independentely.Thus it can operate in parrallel with the other systems and speed up the processing (hence facilitating real time implentation of the overall system).

In the third part of the research,a technique for detecting obstacles inside the street was devised.Due to the short time,which is available for this last part of the research,we limited our research to the location of obstacles consisting exclusively of vehicles present on the road.The technique involved the location of horizontal lines in the images,and determining which of these lines represent the boundaries of an obstacle.The technique used edge detection, and was effective for locating obstacles in many images.Once the obstacles have been located,the technique developed in the second part of the research can be used for calculating the distances from the controlled vehicle to the other vehicles inside

the street.For extending the system so as to locate obstacles other than vehicles,various segmentation techniques could be used. The system can also be applied to the location of any object delimited by horizontal lines and lying on an approximately uniform background.

As a final conclusion,it can be said that the performance of the three systems developed in this research is quite encouraging,given that no special hardware have been used for obtaining the data or for processing it.The reseach have been judged to be sufficiently successful for the extention of the work to the task of identification of road signs which have just started. Although we can say that,given the actual state of the art in image processing and scene analysis,it is feasible to extract from images of street scenes some of the information necessary for driving,more research is needed before the final conclusion on the automation of driving can be reached.Amongst the possibilities of the future research,we can cite research in colour street scenes,and research in stereoscopic street scenes.We feel that the research in these two fields can improve the performance of the system and facilitate the procedures which were used.

9.2 Suggestion For Future Research

A major contribution that can be made in machine vision,would be the development of a general vision system,which is not dependent on a particular set of heuristics,and which can be used for a large number of different applications.During this research,we tried,when possible to develop systems which could be easily adapted and incorporated,in the future, in a general vision system.We feel,that the way to develop a general system,is to develop machine vision systems for particular applications,which would be integrated, when possible,in a general system.Therefore,we are suggesting that the main thrust of future machine vision research,should concentrate on particular applications, and particularly on the automation of information extraction from visual data, so as to automate the guidance of vehicles in all kinds of environments.

Within the scope of the research of this thesis,there are several extentions and enhancements to be suggested.The first suggestion involves the introduction of colour.During all the research,only monochrome images have been used.This fact make it difficult to use thresholding for segmenting the image.Thus the introduction of colour can greatly facilitate the isolation of the road in the image,and the location of the different obstacles inside the street.The research,undertaken by Ohlander (1975) can be easily adapted for this purpose.Although the introduction of colour will multiply by three the data to be analysed for a given resolution,we

think that,when all the parameters of the problem are taken into account,this will be a favourable innovation.It would also be advisable to investigate texture as a possible feature for segmentation.

The second suggestion,which we intended to investigate,but did not have the necessary time for it,is the essential ability to focus on particular area of the image. A desirable system would be a system,comparable to the human visual system,and capable on focusing on given areas of the scene as directed by the program. This would permit the utilisation of an adjustable resolution,for all the image processings such as thresholding,edge detection,and Hough transform. A related hardware enhacement would be the utilisation of a special array processor,which would allow the investigation of the performance of the system in real-time.

The final suggestion is concerned with the use of dynamic stereoscopic vision.Real world scenes are nearly always dynamic and spatially three-dimensional.It seems than reasonable that a machine vision system should take these features of the real world scenes into account.The perception of moving objects and the perception of depth can be obtained from motion parallax,as achieved by the human visual system. The utilisation of a dynamic stereoscopic vision system,which integrates diffences in images,would allow the synthesis of various kind of information from visual data,and thus could lead to the development of a general machine vision system which will have a performance approaching that of humans.

9.3 Conclusion

This thesis reports an initial investigation into the problems of automating the extraction of the information necessary for automatic vehicle guidance from street scene images.The analysis methods used and the results obtained are described.These studies have demonstrated the feasibility of the automation of the extraction from street scene images so as to permit the automation of driving,given particular restrictions of the environment.But much more extensive research is needed before the limits of of validity or usefulness of the techniques can be finally established.

An essential part of the analysis has been the isolation of the street from the remainder of the street scene,the development of a technique for measuring distance inside the street,and the location the different obstacles inside the street.

In the first part of the research,we have been able to isolate the street from the remainder of the street scene.This was achieved by combining thresholding, edge detection and the Hough transform for locating the boundaries of the street.The main problems of this part of the research,were due to the presence of yellow and zig-zag lines in the majority of urban street scenes,and to the obstruction of the boundaries of the street by parked vehicles.An essential feature of this research has been the extensive investigation of the techniques for the location of straight lines in images.

During the second part of the research,an original method based on photogrammetric principles have been devised for measuring distances inside the street. This method involved knowing the different parameters specifing the spatial location of the camera which was used to image the scene.An important feature of this part of the research is that it can be used for measuring distances on any plane surface,when the spatial coordinates and orientations of the imaging device are known.

The third part of the research involved the location of obstacles inside the street.We mainly concentrated on the location of obstacles which represent vehicles.This involved the location of the lines which delimit the boundaries of the rear or the front of the obstacles.As illuustrated in chapter 8,we were able to locate all the obstacles in one image.

For automating vehicle guidance,the algorithm developed in the first part,will be used to determine the bounded area which should not be left by the controlled vehicle.The location of the boundaries of the street will also permit the lateral control of the guided vehicle.The algorithms developed in the third part of the research can be used for locating the different obstacles inside the street.Finally the algorithms developed in the second part of the research can be used to calculate the distances between the different obstacles and the controlled vehicle. The determination of these distances will permit the longitudinal control of the guided vehicle,which will involve taking avoidance action,when

appropriate, so as not to collide with the obstacles.

REFERENCES

Y.I.ABDEL-AZIZ and H.M.KARARA (1971) :'Direct linear transformation from comparator coordinates into object-space coordinates',ASP Symposium on close range photogrammetry,1971.

G.J.AGIN (1975) :'An experimental vision system for industrial application.',Proc. of the 5th int. symp. on industrial robots,1975,pp 135.

A.P.AMBLER and R.J.POPPLESTONE (1976) :'Inferring the positions of bodies from specified spatial relationships.',Artificial Intelligence,vol 6,!1976, pp 157-174.

R.H.ANDERSON (1968) :'Syntax directed recognition of handprinted two dimensional mathematics',Ph.D thesis,Division of Engineering and applied physics,Harvard university,1968.

M.L.BAIRD (1977) :'Image segmenteation techniques for locating automotive parts on belt conveyors'Proc. of the 5th int. joint conf. on artificial intelligence,Cambridge,Massachusetts,aug,1977.

C.J.BARTELSON and E.J.BRENEMAN (1967) :'Brightness perception in complex fields',J. Opt. Soc. Amer.,vol 57,pp 953-957.

J.BECK (1972) :'Surface color perception',Cornell University Press,Ithaca, N.Y,1972.

K.BLYTH et al (1974) :'Snow depth measurement with terrestrial photos', Photogrammetric Engineering,vol 40(8),1974,pp 937-942.

J.M.BOOTH and J.B.SCHROEDER (1977) :'Design concideration for digital image processing systems',Computer,vol 10(8),Aug 1977.

C.R.BRICE and C.L.FENNEMA (1970) :'Scene analysis using regions', Artificial Intelligence,vol 1,1970,pp 205-220.

J.M.BURCH and C.FORNO (1975) :'A high sensitivity Moire grid technique for studying deformation in large objects',Optical Engineering,vol 14(2),1975, pp 178-185.

C.BURT (1955) :'The evidence for the concept of intelligence',Brit. J. Educ. Psych.,vol 25,1955,pp 158-177.

D.CAVOURAS (1981) :' ',Ph.D thesis,The City University,1981.

O.W.CHEFFINS (1975) :'Some practical applications of non-topographic photogrammetry',Photogrammetric record,vol 8 (46),1975,pp 505-520.

O.W.CHEFFINS and J.E.M.RUSHTON (1970) :'Edinburgh castle rock:a survey of the north face by terrestrial

photogrammetry',Photogrammetric Record,vol 6, 1970.

C.W.Chen (1976) :'Theory and applications of imagery pattern recognition', Proc. of the 4th Int. Cong. Stereology,National bureau of standard,Publ 431, Washington,D.C,1976.

E.C.CHERRY,H.H.KUBBU,M.B.BARTON and D.E.REARSON (1963) :'An experimental study of the possible bandwidth compression of visual image signals',Proc. IEEE,vol 51, 1963,pp 1501.

Y.S.CHING et al (1978) :'Advances in recognition of handprinted characters', Proc. 4th Int. Joint Conf. Pattern Rec.,Kyoto,Japan,Nov 1978.

M.B.CLOWES (1972) :'Scene analysis and picture grammars',Graphics Languages, EDS Nake and Rosenfeld,pp 70-80.

C.C.CUTLER (1952) :'Differential quantization of communication signals',U.S. patent 2,605,361,July 19,1952.

L.S.DAVIS (1975) :'A survey of edge detection techniques',COmputer Graphics and Image Processing,vol 4,1975,pp 248-270.

L.S.DAVIS (1979) :'Shape matching using relaxation techniques',IEEE Trans. Pattern Anal. Machine Intell.,vol PAMI 1(1),1979,pp 60-72.

E.M.DELORAINE and A.H.REEVES (1965) :'The 25th anyversary of pulse code modulation',Spectrum,vol 2(5),May 1965,pp 56-63.

J.D.DESSIMOZ et al (1979) :'Recognition and handling of overlapping industrial parts',Proc. Int. Symp. Ind. robots,Washington,D.C,March 1979.

W.DOYLE (1962) :'Operations useful for similarity-invariant pattern recognition', J. Ass. Comp. Mach.,vol 9,pp 259-267.

R.O.DUDA and P.E.HART (1972) :'Use of the Hough transform to detect lines and curves in pictures',Comm. A.C.M.,vol 15,pp 11-15.

M.J.B.DUFF (1976) :'CLIP 4:A large scale integrated circuit array parallel processor',Proc. 3th Int. Joint Conf. Pattern Recog.,Coronado,California,nov 1976.

M.J.B.DUFF (1980) :'Propagation in cellular logic arrays',Proc. Works. IEEE 1980 Pic. Data Desc.,Asilomar,California,Aug 1980.

M.EJIRI et al (1971) :'An intelligent robot with cognitiion and decision making ability',Proc. 2nd Int. Joint Conf. Art. Intell.,London,1971,pp 350-358.

M.EJIRI et al :'A process for detecting defects in complicated pattern', Comp. Graph. Image Proc.,vol 2,1973,pp 320-339.

A.H.A.EL-BEIK (1973) :'Photogrammetry in centrifugal testing of soil models', Photogrammetric Record,vol 7(41),1973,pp 538-554.

T.G.EVANS (1969) :'Descriptive pattern analysis techniques',Automatic interpretation and classifications of images,Eds Grasselli,Academic Press, NewYork,1969,pp 79-96.

G.FALK (1972) :'Interpretation of imperfect line data as a three dimensional scene',Artificial Intelligence,vol 3(2),pp 101-144.

W.FAIG (1975) :'Photogrammetric equipment systems with non metric cameras', ASP Symposium on close range photogrammetric systems,1975.

W.FAIG (1975) :'Photogrammetric potentials for non metric cameras', Photogrammetric Eng. Remote Sens.,vol 42(1),pp 47-49.

W.FAIG (1976) :'Calibration of close range photogrammetric systems-Mathematical formulation',Helsinki,1976,(PE and RE,dec 1975).

C.L.FENNEMA and W.B.THOMPSON (1979):'Velocity determination in scenes containing several moving objects.',Computer Graphics and Image Processing,1979, Vol 9,pp 301_315.

K.S.FU (1974) :'Syntactic method in pattern

recognition',Academic Press,1974.

K.S.FU and B.MOAYER (1975) :'A syntactic approach to fingerprint pattern recognition',Pattern recognition,vol 7,1975,pp 1-23.

K.S.FU and Z.CHEN (1975) :'On the connectivity of clusters',Information Science 8,1975,pp 283-299.

D.GABOR (1961):'Light and Information.',E.Wolf,Ed.,Progress in Optics, Amesterdam,1962.Vol 1.

L.A.GAMBINO and B.L.SCHROCK (1977) :'An experimental digital interactive facility',Computer,vol 10(8),Aug 1977.

D.B.GENNERY (1979):'Object detection and measurement using stereo vision.' Proc. of the 5th International Joint Conference in Artificial Intelligence, Tokyo,Japan,August 20_23,1979,pp 320_327.

GIBSON and CROOKS (1938),American Journal of Psychology,51,pp 453

D.GUENTRI (1979) :'Visual system to extract information from street scenes', The City University Systems Science Department,DSS/DG-LNW/185.

D.GUENTRI (1980) :'Extraction of information from images of

street scenes', The City University Systems Science Department,DSS/DG-LNW/208,july 1980.

GUENTRI and NORTON=WAYNE (1980):'Automatic guidance of vehicles using visual data.',Proc. of the 5th International Conference on Pattern Recognition,Miami ,1980,pp 146-149.

A.GUZMAN (1971) :'Analysis of curved line drawings using context and global information',Machine Intelligence,vol 6,Melster and Michie,Ed Edinburgh,Scotland,Edinburgh University Press,1977.

A.GUZMAN (1968) :'Decomposition of a visual scene into three dimensional bodies',A.F.I.P.S. Proc. Fall Joint Comp. Conf. 33,pp 291-304.

R.W.HAMMING (1950) :'Error detecting and error correcting codes',The Bell System Technical Journal,vol 29,1950,pp 147.

R.M.HARALICK (1979) :'Statistical and structural approaches to texture', Proc. IEEE,vol 67,1979,pp 786-804.

C.A.HARLOW et al (1975) :'Automated inspection of electronic assemblies', IEEE Comp.,vol 8,April 1975,pp 36-45.

B.K.P.HORN (1975) :'Obtaining shape from shading information',The Psychology of Computer Vision,P.H.Winston (Ed),Mc Graw Hill,1975.

B.K.P.HORN (1977) :'Understanding image intensities',Artificial Intelligence, vol 21(11),1977,pp 201-231.

L.P.HOROWITZ and G.L.SHELTON (1961) :'Pattern recognition using autocorrelation', Proc. IRE,vol 49,1961,pp 171-185.

P.V.C.HOUGH (1962) :'Methods and means for recognising complex patterns', U.S. Patent 3069654,Dec 18,1962.

D.A.HUFFMAN (1962):'A method for the construction of minimum redundancy codes.',1962,Proc.IRE,40,1098-1101.

C.JACOBUS and R.T.CHIEN (1978) :'Variable neighborhood computations in scene analysis',Univ of Illinois,Urbana,CSL Rep T-60,1978.

F.JELINECK (1968) :'Probabilistic information theory.Discrete and memoryless models',Mc Graw-Hill,NewYork.

H.M.KARARA (1975) :'Industrial photogrammetry',Close range photogrammetric systems,American Society of Photogrammetry,Falls Church,Virginia,1975,pp 97-141.

R.A.KIRSH (1964) :'Computer interpretation of english text and picture patterns', IEEE Trans. on Electronic Computers,E.C.B,pp 363-376.

R.A.KIRSH (1971) :'Computer determination of the constituent structure of biological images',Computer and Biomedical Research,vol 4,1971.

L.S.G.KORASZNAY and H.M.JOSEPH (1955) :'Image processing',Proc. IRE,vol 43,1955, pp 560-570.

B.KRUSE (1973) :'A parallel picture processing machine',IEEE Trans. Comp., vol C-22,1973,pp 1075-1086.

L.S.LEDLEY (1964) :'High speed automatic analysis of biomedical pictures', Science 146,pp 216-223.

L.M.LESTER,H.A.WILLIAMS,B.A.WEINTRAUB and J.F.BRENNER (1978):'Two graph searching techniques for boundary finding in white blood cell images', Comput. Biol. Med.,vol 8,1978,pp293-308.

M.D.LEVINE and J.LEEMET (1976) :'A method for non-purposive picture segmentation', Proc. 3rd Int. Joint Conf. Pattern Rec.,Coronado,California,Nov 1976.

M.D.LEVINE and S.I.SHAHEEN (1979) :'A modular computer vision system for picture segmentation and interpretation.',Proc. of IEEE Computer Society Conference on Pattern Recognition and Image Processing,Chicago,IL,August 6_8,1979,pp 523_539.

W.S.LEVINE and M.ATHAUS (1966) :'On the optimal error

regulation of a string of moving vehicles',IEEE Trans. Automatic Control,Vol AC11,July 1966,pp 355-361.

R.D.LEVINE and M.TRIBUS (1978) :'Maximum Entropy Formalism',Conf at Massachusetts Institute of Technology,May 1978,MIT Press.

A.H.LEVIS and M.ATHAUS (1968) :'On the optimal sampled-data control of strings of vehicles'Transportation Sci.,Nov 1968,pp362-382.

R.L.LILLSTRAND (1972) :'Techniques for change detection',IEEE Trans. on Computers,vol C21,1972,pp654-659.

MACKWORTH (1973) :'Interpreting pictures of polyhedral scenes',Artificial intelligence,vol 4,1973,pp 121-137.

A.MARTELLI (1976) :'An application of heuristic search methods to edge and contour detection',Comms. ACM 19,1976,pp 73-83.

D.MARR,G.PALM and T.POGGIO (1978) :'Analysis of a cooperative stereo algorithm', Biological Cybernetics,vol 28,1978,pp 223-239.

W.N.MARTIN and J.AGGRAWAL (1978) :'Dynamic scene analysis',Comp. Graph. Image Proc.,vol 7,1978,pp 356-374.

J.MAX (1966) :'Quantizating for minimum distortion',Trans.

IRE,vol IT-6, 1960,pp 7-12.

J.W.Mc KEE and J.K.AGGRAWAL (1975) :'Finding edges of the surface of three-dimensional curved objects by computer',Pattern Recognition,vol 7,1975, pp 25-52.

U.MONTANARI (1971) :'On the optimal detection of curves in noisy pictures', Communs. ACM,vol 14,1971,pp 335-345.

HANS.P.MORAVEC (1977) :'Towards automatic visual obstacle avoidance.', Proc. of the 5th International Joint Conference on Artificial Intelligence, Cambridge,MA,USA,August 22_25,1977,pp 584.

H.P.MORAVEC (1979) :'Visual mapping by a robot rover',Proc. 6th Int. Joint Conf. Art. Intell.,Tokyo,Japan,Aug,1979.

P.MOUKAS (1976) :'Identification of sources of noise pollution',The City University Department of Systems science,DSS/PNM/111,July 1976

P.MOUKAS (1981) :'Application of feature space pattern recognition to identification of nuisance sounds',The City University Department of Systems Science,DSS/LNW-PM/219,March 1981.

J.H.MUNSON (1970) :'The SRI intelligent automation program',Proc. 1st Int. Symp. Ind. robots,ITT Research Institute,Chicago,1970,pp 113-117.

H.H.NAGEL (1976) :'Experience with Yakimovsky's algorithm for boundary and object detection in real world images',Proc. 3rd Int. Joint Conf. Pattern Rec.,Coronado,California,Nov 1976.

H.H.NAGEL (1978) :'Formation of an object concept by analysis of systematic time variations in the optically perceptible environment.', Bericht nr 27,University of Hamburg,July,1976. ', IFI-HH-M-79/80,Hanburgh University,19980.

P.NAGIN (1979) :'Studies in image segmentation algorithms based on histogram clustering and relaxation', Ph.D Thesis,Department Comp. Inform. Sci.,Univ. Massachusetts,Amberst,1979.

R.NARASIMHAN (1966) :'Syntax-directed interpretation of a class of pictures', Communs. ACM,vol 9(3),pp 163-173.

D.NITZAN,A.E.BRAIN and R.O.DUDA (1977) :'The measurement and use of registered reflectance and range data in scene analysis,Proc. IEEE,vol 65,1977,pp 206-220.

C.S.NOVAK and V.R.TAYLOR and J.LEESE (1970) :'The determination of cloud pattern motion from geosynchronous satellite image data',Artificial Intelligence,vol 2,Dec 1970,pp 279-292.

L.NORTON-WAYNE (1979) :'The City University contribution to E.C.S.L. final report',The City University Department of Systems Science,DSS/LNW--WJH/183.

R.OHLANDER,K.PRICE and D.R.REDDY (1978) :'Picture segmentation using recursive region splitting methods.',Computer Graphics and Image Processing,1978,Vol 8,pp 313_333.

R.OHLANDER (1975) :'Analysis of natural scene.',Ph.D. Thesis,Carnegie_ Mellon University,April 1975.

J.B.O'NEAL (1966) :'Predictive quantizating systems(differential pulse code modulation) for the transmission of television signals,BSTJ,vol XLV, 1966,pp 689.

J.T.OLSZTYN et al (1973) :'An application of computer vision to a simulated assembly task',Proc. 1st Int. Joint Conf. Pattern Rec.,Oct 1973,pp 505-513.

A.V.OPEINHEIM,R.W.S.HAFER and T.G.STOCKHAM (1968) :'Non linear filtering of multiplied and convolved signals',Proc. IEEE,vol 56,1968,pp 1264-1291.

P.F.PANTER and W.DITE (1951) :'Quantization distortion in pulse count modulation with non-uniform spacing of levels',Proc. IRE,vol 39,1951, pp 44-48.

T.PAVLIDIS (1976) :'Syntactic feature extraction for shape recognition', Proc. 3rd Int. Joint Conf. Pattern Recog.,Coronado,California,Nov 1976.

W.A.PERKINS (1976) :'Multilevel vision recognition system',Proc. 3rd Int. Joint Conf. Pattern Rec.,Coronado,California,Nov 1976,pp 739-744.

W.A.PERKINS (1978) :'A model based vision system for industrial parts', IEEE Trans. Comput.,vol C-27,1978,pp 126-143.

W.A.PERKINS (1980) :'Area segmentation of images',IEEE P.A. Mach. Int., vol PAMI-2(2),Jan 1980.

J.POTTER (1975) :'Scene segmentation by velocity measurements obtained with a cross-shaped template',Proc. 4th Int. Joint Conf. Art. Intell., Tblisi,Georgia,USSR,Sept 1975.

J.M.PRAFGER (1980) :'Extracting and labelling boundary segments in natural scenes',IEEE Trans. Pattern Analysis and Machine Intelligence,vol PAMI-2(1),Jan 1980.

W.K.PRATT (1977) :'Digital image processing',John Wiley and Sons,NewYork,1977.

W.K.PRATT (1979) :'Image transmission techniques',Academic Press,NewYork,1979.

J.M.S.PREWITT (1970) :'Object enhancement and extraction',Picture Processing and Psychopictorics,Academic Press,NewYork,1970.

K.E.PRICE (1976) :'Change detection and analysis in multispectral images', Ph.D Thesis,Carnegie-Mellon Univ,Pittsburgh,PA,1976.

E.U.RAMER (1975) :'The transformation of photographic images into stroke arrays',IEEE Trans. Circuits Syst.,vol CAS-22,1975,pp 363-374.

RISEMAN and ARBIB (1977) :'Computational techniques in the visual segmentation of static scenes',Computer Graphics and Image Processing,vol 6,1977,pp 221-276.

L.G.ROBERTS (1965) :'Machine perception of three dimensional solids', Optical and electro-optical Information Processing,J.Tippett et al (Eds), MIT Press,1965,pp 159-197.

D.ROHRBACHER and J.L.POTTER (1977) :'Image processing with the SATARAN parallel computer',Computer,vol 10(8),Aug 1977.

A.ROSENFELD (1969) :'Picture processing by computer',Academic Press,NewYork,1969.

A.ROSENFELD and A.C.KAK (1976) :'Digital picture processing'Academic Press,NewYork,1976.

A.ROSENFELD (1980) :'Picture processing 1979',Computer Graphics and Image processing,vol 13(1),May 1980.

A.ROSENFELD,S.W.ZUCKER and R.A.HUAMMEL (1977) :'An application of relaxation labelling to line and curve enhancement',IEEE Trans. Computers,vol 26,1977, pp 394-403 and 922-926.

A.ROBERTS and B.MATHEWS (1980) :'Charge-coupled device (CCD) camera/ Memory optimisation for expandable autonomous vehicle.',Proc. of the Society of Photo-Optical Instrumentation Engineers,Vol 219,pp 69_76, February 6_7,1980.

A.ROSENFELD and M.THURSTON (1971) :'Edge and curve detection for visual scene analysis',IEEE trans. Comput.,vol C-20,1971,pp 562-569.

T.SATO (1979) :'Automotive stereo vision using deconvolution techniques', Proc. 6th Int. Joint Conf. Art. Intell.,Tokyo Japan,Aug 1979.

B.S.SCHATER,L.S.DAVIS and A.ROSENFELD (1976) :'Scene segmentation by cluster detection in color spaces',SIGART newslettres,n:58,1976,pp 16-17.

C.E.SHANNON (1948) :'A mathematical theory of communcation',Bell Syst. Tech. J.,vol 27,1948,pp 379-423 and 623-656.

A.C.SHAW (1969) :'A formal picture description scheme as a basic for picture processing systems',Information and Control,vol

14,1,9,58.

S.D.SHAPIRO (1975) :'Transformations for the computer detection of curves in noisy pictures',Computer Graphics and Image Processing,vol 4,pp 328-338.

R.B.STANTON (1972) :'The interpretation of graphics and graphics languages', Ed F.Nake and A.Rosenfeld,1972,pp 144-159.

J.M,TENENBAUM and H.G.BARROW (1976) :'IGS:A paradigm for integrating image segmentation and interpretation',Proc. 3rd Int. Joint Conf. Pattern Recog.,Nov 1976, pp 504-513.

J.M.TENENBAUM and H.G.BARROW (1977) :'Experiments in interpretation guided segmentation',Artificial Intelligence,vol 8,June 1977,pp 241-274.

L.M.TERMAN (1921) :'Symposium on intelligence',Journal of Educational Psychology,1921.

R.L.THORNDIKE (1921) :'Symposium on intelligence',Journal of Educational Psychology,1921.

J.T.TOU and R.C.GONZALEZ (1974) :'Pattern recognition principles',Reading, MA:Addison-Wesley,1974.

Y.TSUBOI and T.INOUE (1976) :'Robot assembly system using TVV

camera',Proc. 3rd Conf Ind. Robot Technology and 6th Int. Symp. Ind. Robots,Univ. Nottingham,UK,March 1976.

A.M.TURING (1950) :'Computing machinery and intelligence',Mind,vol 56(236), 1950,pp 433-460.

M.S.ULSTAD (1973) :'An algorithm for estimating small scale differences between two digital images',Pattern Recognition,vol 5,1973,pp 323-333.

Y.UMETANI and K.TAGUCHI (1979) :'Feature properties to discriminate complex shapes',Proc. 9th Int. Symp. Ind. Robots,Soc. Manufacturing Engineering,1979.

W.H.WALTON (1952) :'Automatic counting of microscopic particles',Nature, n:169,1952,pp 518-520.

D.L.WALTZ (1972) :'Generating semantic descriptions from drawing of scenes with shadows',M.L.C.AI-TR-271,MIT ,Cambridge,Massachusetts,also in Winston,1975

S.WATANABE (1969) :'Methodology of pattern recognition',Academic Press,NewYork,1969.

D.WECHSLER (1958) :'The measurement and appraisal of adult intelligence', Baltimore:Williams and Wilkins,1958.

D.C.WELLS (1977) :'Interactive Image for astronomers',Computer,vol 10(8),Aug 1977.

N.WIENER (1948) :'Cybernetics',J. Wiley and Sons,NewYork,1948.

P.M.WILL and K.S.PENNINGTON (1971) :'Grid coding :a preprocessing technique for robot and machine vision',Artificial Intelligence,vol 2,1971,pp 319-320.

R.M.WILSON,D.L.TENBER,D.T.THOMAS,J.R.WATKINS and C.M.COOPER (1977) :'The MSFC image data processing system',Computer,vol 10(8),Aug 1977.

A.WOODROW (1921) :'Symposium on intelligence',Journal of Educational Psychology,1921.

Y.YAKIMOVSKY and J.A.FIELDMAN (1974) :'Decision theory and artificial intelligence :A semantic-based region analyser',Artificial intelligence,vol 5,1974,pp 369-371.

YODA et al (1973) :'A hand-eye system for selection process',Proc. 5th int. Symp. Ind. Robots,Chubu Automations Society,Japan,1973.

Y.ANBE et al (1972) :'A computer controlled robot cart',Proc. 2nd Int. Symp. Ind. Robots,ITT Research Institute,Chicago,Illinois,May 1972.

S.W.ZUCKER (1976) :'Relaxation labelling and the reduction of local ambiguities', Proc. 3rd Int. Joint Conf. Pattern Rec.,Nov 1976,pp 852-862.