A Machine Vision Approach to Rock Fragmentation Analysis

By Robert Charles Crida

Thesis presented for the Degree of Doctor of Philosophy,
Department of Electrical Engineering,
University of Cape Town.
Cape Town, September 1995

Abstract

This thesis is concerned with the development of an instrument for the purpose of performing online measurement of rock size distribution using machine vision. This instrument has application in the gold mining industry where it could be used to measure the fragmentation of gold ore on a conveyor belt feed to an autogenous mill, for the purpose of controlling the mill. The gold ore can range in size from fine material ($< 20mm$) *to very large rocks (0.5m).*

A machine vision approach is only capable of directly measuring the projected area of particles at the surface of the rock-stream. A volume distribution has to be estimated from this using a stereological method. These methods have been investigated previously and are typically error prone. They have not been investigated here.

An investigation of lighting demonstrates that a diffuse lighting arrangement is suitable for this application. This would have two advantages: specular reflection from wet material is suppressed; and intensity values can be used to predict the orientation of the surface of the particles.

A computational structure has been developed to identify and delineate rocks in an image for the purpose of measuring their areas. It is based on the human visual system in that it consists of a low-level preattentive vision stage and a higher-level stage of attention focusing.

Multiscalar image processing techniques have also been integrated in order to improve the detection of rocks across a wide range of sizes. A performance advantage can be obtained in this way because all the algorithms can be better matched to the size of the objects being detected.

Results have been obtained with an average true detection rate of 69% and a further close miss rate of 14%, with very few false alarms. The overall result is that the measured projected area distribution closely matches the true value for each test image.

Contents

List of Figures

List of Tables

Symbol Notation

One-dimensional Signals and Lists

$G(t;\sigma)$	one-dimensional Gaussian
$L_{ellipse}$	list of peaks in I_{peak} as the result of the first stage of preattentive vision
$L_{highlight}$	list of targets for attention focusing after the second stage of preattentive vision
L_j	an element of a list
$p(r)$	probability profile to bias segmentation
S_{edge}	profile of the radial distance to detected rock edge points
$S_{boundary}$	profile of the radial distance after processing
L_{edge}	list of edge points obtained by tracking around A_{label}
$L_{interior}$	list of all points in A_{label} that are not in L_{edge}
E_{edge}	list of the edge strength at each point in $S_{boundary}$
L_{class}	list of targets accepted by classification
L_{node}	list of nodes for tree creation at one level of the image pyramid
L_{rock}	list of rocks detected in the image

Two-dimensional Signals and Images

$G(x,y;\sigma)$	two-dimensional Gaussian
$K(x,y;t)$	discrete approximation to Gaussian
I_{camera}	input image of scene to be processed
P	image pyramid $(P_1,\ldots,P_n)$
I_{GRAD}	morphologically edge detected image
$I_{average}$	smoothed version of current P_j
I_{dilate}	dilated version of current P_j
I_{edge}	binary edge image used to detect ellipses
$I_{ellipse}$	Hough transformed version of I_{edge}
I_{clip}	$I_{ellipse}$ clipped at t_{clip}
I_{peak}	all the peaks in I_{clip}
A_{gap}	narrow gaps between rocks detected using tophat operator in a region of interest
A_{dir}	directionally edge detected image in a region of interest
A_{grad}	morphologically edge detected image in a region of interest
A_{label}	segmented region in a region of interest

SYMBOL NOTATION

Image operators

$*$	convolution operator
$\nabla^2 G$	Laplacian of Gaussian image operator
$\ominus$	morphological erosion operator
$\oplus$	morphological dilation operator
$\circ$	morphological opening operator
$\bullet$	morphological closing operator
$GRAD(f,g)$	morphological gradient of image f using structuring element g
$HAT(f,g)$	morphological tophat operator on image f using structuring element g

Classification

$\mathbf{x}$	feature vector
p_{opt}	optimal number of features for classification
k	number of samples considered for classification by k-nearest neighbour method

Acronyms

CCD	charge coupled device
DOLP	difference of low pass
ESIP	expert systems for image processing
GA	genetic algorithms
HVS	human visual system
IUS	image understanding systems
MTF	modulation transfer function
PMC	probability of misclassification
ROC	receiver operating characteristic
ROI	region of interest
RCM	radial contour model
ROM	run of mine
SE	structuring element

Ellipse Parameters

n_θ	number of different orientations
n_e	number of different eccentricities
a, b	major and minor axes
e	aspect ratio equal to the ratio $\frac{a}{b}$
θ	orientation of the ellipse
$\mathbf{p}$	coordinates of the centre of the ellipse
h	Hough transform value for the ellipse
p	average edge point probability for ellipse after binary segmentation

SYMBOL NOTATION

Lighting and Image Formation

E_i	incident irradiance
I_i	radiance of incident light
I_e	expected intensity of a rock
ρ	bi-directional reflectivity
ρ_d	reflectivity for diffuse reflection
ρ_s	reflectivity for specular reflection
k_d	diffuse reflection coefficient
k_s	specular reflection coefficient
I_r	reflected radiance
$\mathbf{n}$	surface normal
i	incident vector
$\mathbf{v}$	viewing direction
$\mathbf{r}$	reflected direction
α	angle between $\mathbf{r}$ and $\mathbf{v}$
ψ	angle between $\mathbf{n}$ and z-axis
μ	inclination of edge of light dome
P	projection plane
S	surface tangent plane

Factors and Parameters

$d_{feature}$	distance between detected rock and closest actual rock
σ_d	desired amount of smoothing
σ_e	amount of smoothing that is required
σ_i	current degree of smoothing in an image
$r_{expected}$	expected radius of rock based on position in scale-space
$levels$	depth limit for generating trees
t_{morph}	global threshold to get I_{edge} from I_{morph}
t_{edge}	intensity threshold to get I_{edge} from I_{dilate}
t_{clip}	Hough transform clip level
t_{gap}	minimum edge strength during segmentation
$t_{outlier}$	used to remove outliers from detected edge points
t_{prob}	threshold limit on edge probability
t_{moment}	moment-preserving threshold
$t_{feature}$	threshold for rule-based classification
t_{branch}	distance threshold when adding branches to trees
f_{scale}	size factor used to create image pyramid
f_{edge}	edge position uncertainty due to smoothing
f_{aspect}	factor used to deform a circle to create ellipses
$p_{crossover}$	probability of crossover for every pair of chromosomes
$p_{mutation}$	probability of mutation for each chromosome
f_{limit}	maximum scale change before distance becomes greater than one
$f_{overlap}$	allowable degree of overlap of circles representing two rocks
f_{depth}	proportion of depth of ellipsoid to minor radius of ellipse for calculation of volume

SYMBOL NOTATION

Optics and Image Acquisition

MTF_L	MTF of lens
MTF_I	MTF due to image smear caused by motion
MTF_D	MTF due to spatial integration of image detector
MTF_S	MTF of cascaded image acquisition system
f	focal length of a lens
m	image magnification
F/number	angular acceptance of a lens
D	entrance pupil diameter of a lens
λ	wavelength of light
μ	spatial frequency
γ	distance moved by conveyor belt during exposure
v	conveyor belt speed
T	exposure period
T_{alpha}	active dimension of a CCD pixel
T_{beta}	sample spacing of CCD pixels

Chapter 1

Introduction

1.1 Background

Sophisticated control systems are required in the gold mining industry for monitoring and controlling the milling of gold ore as part of the gold extraction process. The cost of producing gold is affected by the efficiency of the extraction process in at least two ways: milling requires large amounts of electrical energy and is therefore expensive, and obtaining a low percentage yield increases the amount of material which must be stored because it has too low a gold content to be processed.

It is therefore possible to reduce the cost of producing gold by improving the efficiency of the milling process.

It has been proposed that the efficiency of an autogenous mill depends partly on the size distribution of the material being milled. Stanley [85] states that "*In the autogenous mill, the grinding media are part of the ore to be ground. They have thus not only to be reduced in size themselves, but must also achieve useful breakage of the finer size fractions of the mill contents in the process.*" He further states that for successful autogenous milling there must be larger fractions in the mill which are capable of surviving in the mill in sufficient concentration to provide an adequate grinding load.

It therefore appears that if the size distribution of material being milled can be measured, it can be controlled, which will be beneficial to the milling process. This problem was highlighted in discussion with MINTEK personnel (Gunter Sommer, Tony Lange and Dave Hulbert. Ideally, some form of feedback could be used to control the distribution in the feed to the mill. Unfortunately, control of the feed size distribution is normally not possible except in occasional cases where some form of size selection is available by drawing material from different points of a silo. An alternative approach of preprocessing the material to correct the size distribution is undesirable for the following reasons: the philosophy for run of mine (ROM) milling is that the mill should take what it gets from the mining operation and should cope with it, preprocessing is expensive, and usually larger sizes are desirable anyway.

A more viable solution is to use a feed-forward system to control the mill to compensate for the size of material which is being fed into it. This information will also help to increase the understanding of the milling process by providing a measure of the mill feed disturbance.

A longer term *control* might be to feed back to the mining operation information about desired changes in the distribution. Changes to blasting procedures could then provide the necessary effect.

An automated solution will result in benefits for the mining industry. This will be achieved because better control of the milling process means more efficient milling (lower cost) and higher yield.

An instrument is therefore required to perform an on-line measure of the size distribution of material on the conveyor belt feeding the mill. The instrument is to be installed in the harsh environment of a mineral processing plant, where minimum maintenance is a high priority. The instrument must therefore have the following properties [46]:

- Perform a non-intrusive measure so that there is no contact with the rock-stream and therefore no wearing down of parts.
- The instrument must be reliable to reduce plant down-time.
- Production must not be hampered by failure of any part of the instrument.
- The instrument must be resistant to the following environmental factors: a temperature range of -10 to 40 degrees centigrade, a corrosive and dusty atmosphere, water from rain or high pressure hoses and electromagnetic interference from large motors found in such a plant.

1.2 A Machine Vision Approach

A machine vision approach to solving this problem provides several advantages given the adverse conditions of the environment and the nature of the problem. These include the following:

- Non-intrusive measurements can be performed on-line without stopping the conveyor belt.
- The instrument can operate for 24 hours per day.
- Increasing computer power will further increase the sampling frequency.
- Results are consistent for the same scene at any time given the same lighting conditions.
- Instruments at other locations will produce comparable results.

- There are no mechanical parts which dramatically reduces maintenance due to wear and tear.

In particular, machine vision offers advantages over sieving due to the significantly higher sampling rate and non-intrusive measure. It also benefits over visual estimation of fragmentation by removing the subjective element which characterises measurements made by humans from visual inspection.

The major disadvantage of machine vision is that it can only see two dimensional projection of the surface of the material so that it becomes difficult to measure the actual overall distribution of material. It has been found in previous research [56, 46] that it is still possible to obtain a useful measure of the overall size distribution under these conditions. Considering the many advantages that are offered by a machine vision approach, such a solution becomes viable.

1.3 Thesis Format

This thesis is divided into three main parts: firstly the problem is defined and a solution structure is proposed; secondly, the various stages of the solution are presented and thirdly, results and conclusions are given. The remainder of this section briefly describes the content of each of the remaining chapters.

Chapter 2 gives a background into the subject of rock size and its measurement. Previous attempts using machine vision are then examined and a new approach based on the human visual system and multiscalar image processing is then proposed as a framework for a solution.

Chapter 3 provides an introduction and theoretical background to several topics which are used later to develop the solution. Image processing techniques are used for image preprocessing and segmentation. Some background into image formation is given to facilitate the development of a suitable lighting model and the understanding of characteristics of captured images. Finally, aspects of pattern recognition theory are given as a basis for identifying rocks by means of a classification process.

Chapter 4 gives a complete overview of the instrument structure which is developed in this thesis. It consists of both hardware and software components which are identified and their significance analysed. The remainder of the body of the thesis (Chapters 5–9) is concerned with developing the various instrument components described in this chapter.

Chapter 5 investigates the various stages of image formation by analysing the effects of lighting on the visual appearance of the scene and identifying the effect of the image capture components on the characteristics of the resultant image. The aim of this procedure is to identify desirable attributes for a lighting arrangement and to determine whether or not an understanding of the lighting can be used to facilitate the identification of objects in the image.

Chapter 6 deals with the topic of image preprocessing and is split into two main sections. The first section describes the process of forming an image pyramid which is a representation of an image at different scales and forms the basis of a multiscalar approach. The second section develops a method based on the Hough transform for identifying possible positions in the image pyramid where rocks could be present.

Chapter 7 develops a segmentation algorithm which incorporates knowledge from several sources to identify the edges of rocks in regions of interest.

Chapter 8 explains how features can be measured from segmented regions of interest in order to identify which ones contain a rock and which do not by means of a classification process. This chapter also describes the classification process, topics regarding the selection of a representative training set and optimisation of results by means of feature selection.

Chapter 9 describes a process which is based on the use of tree structures for analysing results from the image pyramid in order to provide a final estimate of the positions of rocks in the image. From this information, a size distribution can be calculated by measuring the size of each of the identified rocks using the segmentation algorithm from Chapter 7.

Results are presented in Chapter 10. A wide range of images have been used to indicate the performance of the instrument under adverse conditions.

Conclusions and recommendations for future work are given in Chapter 11.

Chapter 2

Problem Statement

2.1 Rock Size Measurement

The aim of this thesis is to provide a method for measuring rock size. The following sections provide a background into the subject of rock size in terms of difficulties involved in the measurement environment, characteristics of rock material, quantification of rock size and methods for presentation of size data.

A review of previous work in the area of rock fragmentation analysis is presented in Section 2.2 followed by an introduction to the adopted research philosophy in Section 2.3. This will be critically compared with work from the literature review to determine that it identifies and tackles relevant issues to ensure a successful result.

2.1.1 Environmental Difficulties

A gold extraction plant provides a particularly unfriendly environment in which to operate a machine vision system [46]. The close proximity of mills and other heavy electrical machinery means that there is a significant quantity of vibrational and electrical noise. In addition, the constant movement of ore results in a high dust content in the atmosphere and a quick buildup of a thick layer on any exposed surfaces (this is a particularly important consideration for camera lenses). Another dangerous element is water which is used for hosing down the plant. A machine vision system must clearly be sufficiently robust to withstand these conditions.

Another very important consideration in the design of a machine vision system is the cancellation of lighting variation which can be caused by variable weather, different times of day, and different designs and conditions at the plant. In order to ensure good results from a machine vision application it is not only necessary to maintain constant lighting, but the conditions should be set to highlight solution oriented features and suppress irrelevant information.

Despite these contrary indications, using image processing for this application provides several advantages:

- The measurement is non-invasive which has two advantages: firstly, there is no interference with the rock-stream and secondly, there are no parts undergoing serious wear thus reducing maintenance.
- The material on the conveyor belt can be sampled fairly frequently compared to sieving (a couple of samples per day) and increasing computing power will ensure ever more representative results.
- Modern CCD array cameras provide sufficiently high shutter speeds to allow for accommodation of any conveyor belt speed while still producing unblurred images for processing.
- An image processing system will provide an objective measure of rock size distribution which will not vary with time nor location.
- Apart from some attention to the optical components of a machine vision system, maintenance is generally low due to the reliability of modern computing equipment and the lack of mechanical components.

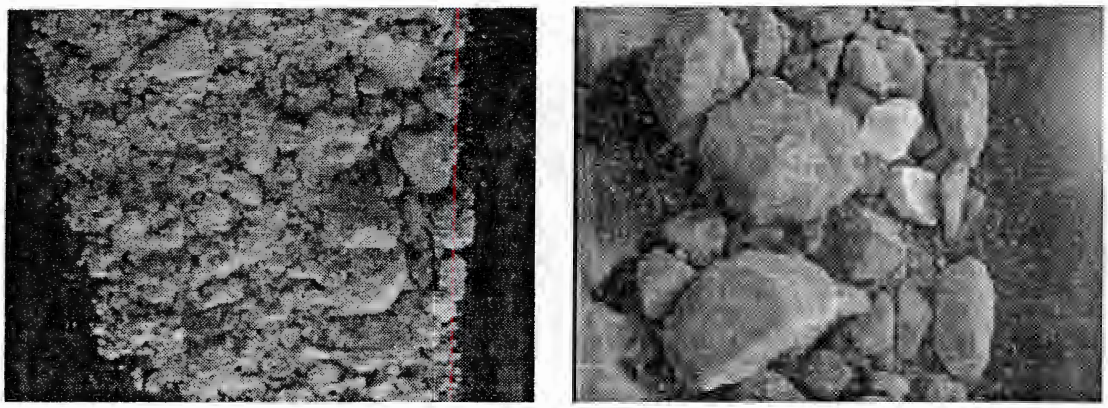
2.1.2 Rock Material

Rock material which is usually fed into an autogenous mill can be classified as run-of-mine material, that is, the material from the mine is fed directly into a mill without prior sorting. According to Lange [46], material from the mine is often stored in a bunker or storage heap in order to buffer the supply of rock material from underground to the mill. Another advantage of some rock storage systems is that it allows some degree of size selection due to the settling pattern of material on the storage heap.

Rock material can vary in size from very large (half a metre across) to very fine (tens of microns) and also in shape, colour and flow rate of the rock-stream on the conveyor belt. The material is usually very damp and under this condition fine material can form agglomerates which look like large rocks. The sticky nature of the mud also causes lumps of fines to stick to larger rocks.

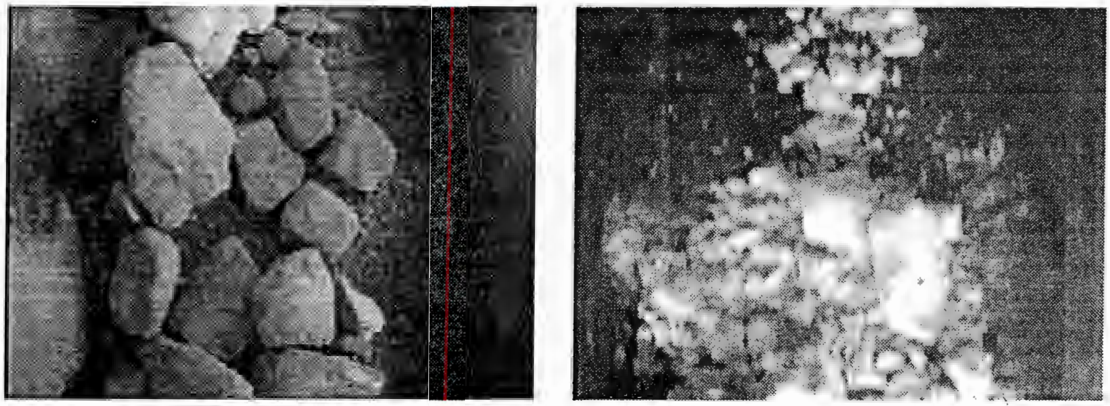
The characteristics mentioned above often vary from location to location so that two adjacent conveyor belts feeding mills can have rock-streams which appear different because they are taken from different parts of the stock pile. The various characteristics of rock material that can differ are: chemical and mineral content, size and shape distributions, colour, texture and moisture content.

The flow of material is uneven due to mechanisms for feeding material onto the conveyor belt. For example, the use of a vibrational feed results in lumps of material on the belt with spaces between them. Also, the speed of the conveyor belt influences the structure of



(a) Wet material covered with fines.

(b) Dry material reasonably closely packed with some occlusions.



(c) Dry material slightly separated with minimal occlusions.

(d) The edge of a lump of material caused by vibrational conveyor belt feed.

Figure 2.1: Different characteristics of rock material.

material. A slow moving belt is loaded heavily with material up to half a metre deep, while a high-speed belt can be loaded with a thin layer of fine material (< 10mm) with few large rocks (> 100mm) sparsely positioned on top of it.

There are therefore several complex parameters affecting the appearance of material on the conveyor belt which can be contributed to two basic influencing factors: the *structure* of the conveyor belt and loading equipment, and the *rock material* itself.

Figure 2.1 gives some examples of different visual appearances of rock material under various conditions described above.

2.1.3 Rock Size

According to Allen [2], the size of a spherical homogeneous particle is uniquely defined by its diameter. Similarly, the length of a side of a cube is characteristic, and for other regular shapes there are equally appropriate dimensions. It may be necessary to specify more than one dimension for some regular particles such as: cone, diameter and height etc.

He further states that derived diameters are determined by measuring a size-dependent property of the particle and relating it to a linear dimension. The most commonly used examples of this are the equivalent spherical diameters. For example, a unit cube has the same volume as a sphere of diameter $d_v = 1.24$ units and hence this is the derived volume diameter.

Other mechanisms can be used for determining the size of a particle such as comparing its terminal velocity in a liquid to a sphere in similar conditions. In this way the property of the particle can be expressed in terms of an equivalent spherical particle. See Appendix A for a list of some other equivalent diameters.

Allen also states that "*For irregular particles, the assigned size usually depends upon the method of measurement, hence the particle sizing technique should, wherever possible, duplicate the process one wishes to control.*" The performance of an autogenous mill is dependent on the relative masses of rocks in the mill. It is therefore desirable to measure the mass of the rocks which are fed into the mill. This can be achieved indirectly by measuring rock volume. Representation can then be achieved by means of the diameter d_v of an equivalent sphere with the same volume as the rock and known density.

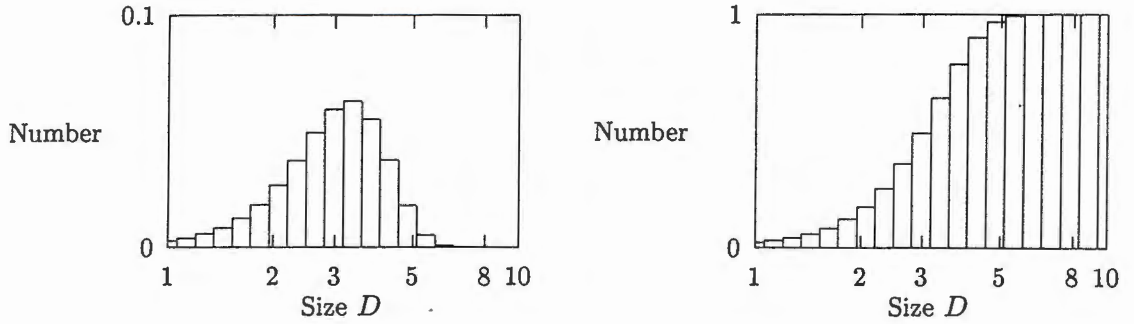
In the following discussion of size distribution representation (based on Lange [46]), it will be assumed that size has been measured with a characteristic dimension D, is classed into I size ranges and that D_{max} is the largest size measured.

2.1.4 Representation of Size Distributions

The result of the measurement process is typically a *number* distribution in that there are n_i particles counted in each of the size ranges D_i with a total of N particles. It is assumed here that the value D_i corresponds to the beginning of the size range and that the width, ΔD_i, increases exponentially with increasing i for display with logarithmic abscissa scale. The *number frequency* f_i of particles for each size range is given by the following equation.

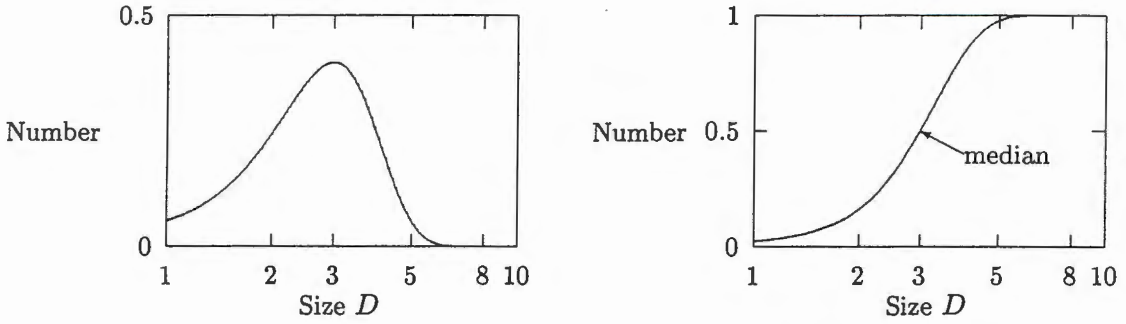
$$f_i = \frac{n_i}{N} \qquad \text{where} \qquad \sum_{i=1}^{I} n_i = N \qquad \text{and} \qquad \sum_{i=1}^{I} f_i = 1 \tag{2.1}$$

A histogram is commonly used for graphically displaying number frequency f_i and an example of this is shown in Figure 2.2(a). Note that as $\lim_{\Delta D_i \to 0}$ then $f_i \to f(D)$ which is a continuous function of D and $\int_0^\infty f(D)\, dD = 1$. The continuous frequency distribution curve



(a) Number frequency histogram

(b) Number frequency cumulative distribution



(c) Continuous frequency distribution

(d) Continuous frequency cumulative distribution

Figure 2.2: Various graphs used to represent size distribution. Note that the horizontal axes of the graphs are logarithmic.

is shown in Figure 2.2(c).

A mass or volume curve is usually required by the mineral processing industry because size data collected with the sieving methods are fractions by *mass* of material w_i which have passed through various sieves and has been caught by sieve D_i.

In general, the mass of a particle of size D is equal to

$$w = \alpha_{v,k} \rho D^3 \tag{2.2}$$

where k is the type of measurement, $\alpha_{v,k}$ is a volume shape coefficient [2, p108], ρ is the density of material and D is the equivalent diameter of a sphere of type k. For example, if the type of measure was projected area diameter (see Appendix A) then particle mass could be estimated using $\alpha_{v,a} \rho d_a^3$. Allen [2, p109] describes a method for calculating $\alpha_{v,a}$ by weighing particles of known size. In order to calculate a mass curve, it is necessary to determine the mass distribution $\frac{w_i}{W}$ from the number frequency distribution $\frac{n_i}{N}$. This can be achieved using

(2.2) as follows:

$$\frac{w_i}{W} = \frac{\alpha_{v,k}\rho n_i D_i^3}{\sum_{i=1}^{I} \alpha_{v,k}\rho n_i D_i^3} \tag{2.3}$$

Note that this method of conversion is particularly error prone due to raising the diameter D to the third power. The problem occurs especially when converting larger sizes of material because of the exponential nature of the conversion process. Consider material in a 1 to 10mm size range. A single error with a 10mm particle has the same weight as 1000 errors with 1mm particles. Errors can be reduced by increasing N and decreasing the width of size intervals, ΔD_i [2, p123].

A logarithmic cumulative distribution is the accepted method for representing size distributions in the mineral processing and quarrying industries and is therefore commonly used for representation of fragmentation results [38]. The fraction of particles by number less than size D_r is given by

$$F_r = \sum_{i=1}^{r} f_i \qquad < 1 \qquad \text{and} \qquad D_r < D_{max} \tag{2.4}$$

The fraction of particles by mass less than size D_r is given by

$$W_r = \sum_{i=1}^{r} \alpha_{v,k}\rho f_i D_i^3 \tag{2.5}$$

Cumulative distribution curves are shown in Figures 2.2(b) and 2.2(d) for the relative frequency curves shown there. When F_r and W_r are normalised to 1, they are referred to as the "S" curve or the "*cumulative percent passing frequency curves*".

Sometimes the distributions are represented by means of characteristic average diameters which measure a central tendency but are unaffected by the relatively few extreme values in the tails of the distribution [2]. In particular, a point which is commonly desired is the median of the curve. This point can be read off the "S" curve as the size at which 50% of the material is passed (See Figure 2.2(d)).

More generally, the "*cumulative percentage passing of some size*" is specified. This point, D_p, can be read directly off the "S" curve as the size at which $s\%$ of the material is passed or $\sum_{i=1}^{p} f_i = s\%$. The median occurs where $s = 50$.

2.2 Literature Survey

Several attempts have been made already to solve the problem of measuring rock fragmentation using computer vision with varying degrees of success. This section is divided into four parts which provide reviews of previous work and related topics in computer vision which are pertinent to the solution proposed in this thesis. Section 2.2.1 gives a review of recent work that has been performed in the field of rock fragmentation and related topics in the mineral processing industry. The subject of computer vision is also briefly reviewed in Section 2.2.2 as an approach to automatically extracting useful information about the world using images of it. Section 2.2.3 reviews the subjects of the human visual system and the related topic of perceptual organisation in computer vision which will be used as a computational framework in which to develop a new solution. Section 2.2.4 reviews the topic of multiresolution systems and multiscalar image analysis which can provide several advantages to many image processing applications and is to be used here to provide a mechanism for identifying and segmenting rocks independently of their size. Finally, Section 2.2.5 summarises the problem of measuring rock fragmentation with relation to the success and failure of previous attempts.

2.2.1 Rock Analysis Review

The aims of this review of rock fragmentation analysis and other similar computer vision processes in the mineral processing industry are:

- to give an indication of the requirements of a computer vision approach to measuring rock fragmentation,
- to investigate the types of problem and sources of error that can be introduced by such a system and
- to review approaches that have been attempted in the past with regard to their aims and how they have performed.

The main aim of this work is to provide an estimate of gold ore fragment size distribution for the purpose of control and optimisation of autogenous mills. This is not a new goal and has been attempted previously [5, 46, 71, 93]. Computer vision approaches have also been used for several other purposes when analysing images from the mineral processing industry, eg measuring the size of oversize rocks to facilitate breaking [13]; analysing muckpiles to improve blasting procedures [50, 56, 84, 96]; control of ore stock-pile height [39]; and monitoring of conveyor belt overburden at coal mines to maximise transportation rates [40].

Miles and Hall [58], Hunter *et al* [38] and Smith and Kemeny [84] review rock fragmentation analysis techniques using computer vision with an emphasis on the analysis of blast muckpiles. Paley [67] describes methods for determining rock size from images and expands on various errors which exist or are introduced when fragmentation is assessed by image analysis methods. Image analysis techniques for rock fragmentation analysis can be separated into three categories: manual, semi-automatic and automatic. The following list gives an indication of how each of the categories is achieved and the types of errors that are introduced.

Manual methods involve an operator performing the task of delineating fragments in an image for the purpose of digitisation. Although error can be reduced by allowing human interpretation of indistinct fragment boundaries, the process is very slow with the result that sampling errors are likely to increase because fewer images can be analysed [58].

Semi-automatic methods often involve a human operator who makes corrections after a computer has outlined fragments [58]. Another semi-automatic approach is described by Chung *et al* [14] in which a computer is used to correct projection transformations and provide a flexible setup in which various media may be analysed using a digitising tablet.

Automatic methods are achieved completely with the use of a computer which outlines rock fragments and calculates volume measurements using a stereological method to convert from the measured value of area or size [58]. Although automatic processing can proceed more rapidly than manual processing with the result that sampling errors are reduced, other measurement errors are introduced when the computer incorrectly outlines fragments.

Further sources of error [67] in the size measurement are the determination of fines which is usually difficult from an image; particle sorting errors which occur because the same particle may be measured to have a different size depending on what surface is being projected; segregation errors which occur because the fragments which occur at the surface may not be representative of the overall distribution; and errors which are introduced by the stereological method for converting to a volume measurement.

The most common approaches that are used for measuring fragmentation involve capturing greyscale images of the surface of material from a muckpile directly after blasting [44, 45, 56, 95, 96] or on a conveyor belt feed using a perpendicular view of the material [5, 46, 93] due to the low cost and robustness of video cameras. It is also possible to operate by measuring the profile of material using structured light [40] or where it passes over a head drum [58]. Cheung *et al* [13] have used rangefinding equipment to obtain a range image of the rock scene. A range image is easier to examine but requires expensive rangefinding equipment which is less robust than ordinary video cameras due to their mechanical scanning nature and are therefore less reliable and undesirable for harsh operating conditions.

A model of rock shape has been used to improve performance in some computer vision approaches. Cheung *et al* [13] use a superquadric surface of the form

$$\left(\left|\frac{x}{a_x}\right|^{2/\epsilon_2}+\left|\frac{y}{a_y}\right|^{2/\epsilon_2}\right)^{\epsilon_2/\epsilon_1}+\left|\frac{z}{a_z}\right|^{2/\epsilon_1}=1$$

which simplifies to an ellipsoid when $\epsilon_1 = \epsilon_2 = 1$ and allows a very wide variety of shapes to be modelled. This is applicable because they are working from a range image and therefore have a three-dimensional input image. When operating from a normal two-dimensional image, the ellipse has been used by Kemeny *et al* [45] to model the projection of a rock. They have also found that the best statistical relationship between the ellipse model and equivalent screen size or sieve diameter is linear with the following form

$$d_A = 0.45a + 0.73b \tag{2.6}$$

where a and b are the major and minor axes of the ellipse respectively.

Controlled illumination of the material is particularly important when automatic analysis of the image is to be performed [58] although it is common for images of muckpiles to be captured under natural light conditions because control of illumination above ground is difficult during daylight hours [38]. Controlled illumination has been suggested for conveyor belt systems [5, 46, 93] because image capture is usually made at the same place which can be readily protected from natural light. Structured lighting has been used for extracting profiles [40] and is used by a rangefinder for creating range images [13].

The remainder of this section gives a chronological survey of results and techniques that have been applied in order to measure a rock size or volume distribution. Although the actual motivation for the measurement may not be the same as for this research, each of the methods starts from a similar grey level image of a rock scene and attempts to measure rock size. Table 2.1 summarises all the information that is given in this section.

Nyberg *et al* [63] is the earliest system reviewed in this survey. They estimate partial gradient components using the Sobel directional gradient operator from which gradient magnitude and direction are calculated. Knowledge that the fragment has higher intensity than the background means that the gradient always points to the inside of a fragment. The "probability" of an edge point is initially defined to be the measured gradient magnitude which is normalised by the maximum value in the image. An iterative procedure based on the probabilities of adjacent points and the compatibility of their directions is then used to reinforce the edge contours in the image. A volume distribution is estimated by randomly measuring particle diameters in the edge detected image and transforming the diameter distribution using a stereometric method to a volume distribution. A major source of error is due to the use of a simplistic geometrical model which is used to define the transformation from diameters to volume.

Berger [5] argues that the use of a chord-length or diameter such as that measured by Nyberg *et al* is prone to error since it relies on an unchanging spectrum to enable statistical deduction of the size distribution. This is flawed since the actual particle shapes are not actually known. He presents a three stage process which is aimed at measuring particle area in the image. The first stage is based on a multiple level thresholding procedure which performs the dual tasks of separating foreground from background in the image while providing suitable primitive regions for a region merging and splitting algorithm. The second stage involves the merging of adjacent regions based on their relative intensities and to remove clusters

Name	Scene[a]	Purpose[b]	Source[c]	Image processing technique and comment
Teklogic [5]	R	A	E	Histogram adjustment is used to select initial regions then segmentation is achieved by region merging and splitting. Measures projected areas.
McGill [13]	G	O	I	Segmentation to separate rocks then fit to a superquadric shape model. It can compensate for overlapping but expects a low number of rocks. 3-D Shape information is used to guide a rock breaker.
BIAS [39]	S	H	E	Image sequence is used to reduce effect of dust cloud by averaging at the point of impact.
BIAS [40]	C	V	L	Coal stream profile analysis using triangulation to provide estimate of belt overburden and detect large rocks.
Mintek [46]	R	A	E	Image segmentation and random chord-length measurement.
Fudan [50]			E	Region growing and rule based correction to measure rock area.
Nottingham [56]	M	B	E	Image enhancement, thresholding and morphological processing to separate rocks. Measures diameter of an area equivalent sphere.
Luleå [63]			E	Edge detection reinforced by direction information to determine rock contours then finds a distribution of scan line diameters.
CSIRO Geomechanics [64]			E	Analysis of simple back-lit scenes with small separated fragments.
RAU [71]	R		E	Neural networks are used to determine predominant rock size from a histogram of shadow chord-lengths.
Wits [93]	M	B	E	Rule based image labelling which uses relative intensity and edge strength. Measures fragment area.
Tucson [95]	M	B	E	Shadow regions are connected between convex points and along lines of high gradient. Calculates an equivalent screen sieving distribution using the linear relationship between a fitted ellipse and screen size in (2.6)
Monash, CSIRO Geomechanics [96]	M	B	E	Region based automatic thresholding routine.

---

[a] R = rock stream, M = muckpile, C = Coal stream, S = stock pile, G = grizzly (metal sieve)

[b] A = control autogenous mill, H = control stock pile height, O = breaking oversize rocks, B = blast design, V = measure volume and overburden

[c] E = greyscale image, L = structured light, I = range image

Table 2.1: Summary of related contemporary rock size and mineral analysis techniques. A reference, in square brackets, is given for each of the techniques in the Name field.

of small regions from the interiors of larger regions. Errors do occur with the result that some particles become merged together and so a third stage of region splitting is used which identifies "bottlenecks" in regions and splits the region by placing a chord across the neck. Although Berger's method produced some good results the actual segmented region in the image does not appear to correspond very accurately with the underlying rock and there is a significant problem with fines being over-merged to produce large rocks.

Wigeson [93] has designed an algorithm which makes more efficient use of edge detection to improve its effectiveness. He uses a rule based approach to sequentially label each pixel of the image. At each point there is an option to merge the pixel to its neighbours. This depends on a high level decision made by a rule base and is based on the relative difference in intensity levels and the existence of strong edges. The major problem with this method is that since it comprises only a single pass across the image and only considers a close neighbourhood around each pixel, it does not consider the global context of the pixel so cannot be expected to perform well and will make mistakes especially in regions of low contrast.

Lu *et al* [50] propose a two step process to perform segmentation of images of rock scenes. In the first the image is preprocessed to remove clusters of fines by means of a thresholding algorithm. Regions are created in the remainder of the image using a partition mode test based on pixel intensities. The fines are removed because they would result in large regions which would be difficult to remove. The second stage of the algorithm uses shape rules to perform merging and splitting of the regions to obtain a final segmentation. They conclude that the precision is adequate for their particular application although their results indicate that the performance is reduced in regions of low contrast.

McDermott and Miles [56] report on a stereological approach to transform from an equivalent projected area sphere diameter distribution to a volume distribution in the case of overlapping fragments. Their image processing involves a procedure of image enhancement to eliminate noise and correct uneven illumination followed by a thresholding technique to produce a binary image which is then manipulated using morphological dilation and thinning to separate connected fragments. From this result fragment areas can be measured and the equivalent area circle distribution estimated. They identify two problems which restrict the performance of the system: (1) *disintegration* of larger sizes results when there is high resolution because of variations in grey level while small rocks can be successfully resolved. (2) *fusion* results from insufficient resolution so that large sizes can be seen but small particles are merged. In order to overcome these limitations, they measure the large and small sizes separately after preprocessing the image by smoothing it with a large or small Gaussian convolution kernel. They conclude that although size distribution can be calculated automatically from images of fragmented rock under laboratory conditions, the method might not produce an acceptable approximation of the size distribution in every case from real images.

Lange [46] argues strongly in favour of chord-length measurement instead of area measurement because he feels that the problem of accurate rock boundary detection is too difficult and will always be error prone. He states that experiments suggest that edge and segmentation errors are less severe for chord-length measurements than for area measurements. Rock boundaries are detected using a two stage process of low level edge detection followed by a higher level stage of edge correction based on knowledge of the most probable shape of rocks. The edge correction involves a process of spot removal followed by a joining algorithm to fill gaps in boundaries. He provides a theoretical background for calculating a sieve size distribution from a chord-length distribution and finds that it is not possible with great accuracy.

Yeo *et al* [96] have investigated the use of parallel processing to improve the speed of fragmentation assessment methods and make them practical for application. They describe a simple algorithm using a transputer array to calculate a threshold surface for the image using the Otsu automatic method [65] and find that processing time can be reduced from tens of minutes to around two minutes in the parallel implementation. It is observed that non-overlapping fragments are accurately segmented but when overlapping occurs, fragments are merged. A further erosion-dilation procedure is proposed as a solution to this problem. In addition, the algorithm responds poorly to darker fragments next to bright ones with the result that they are greatly reduced in size thus affecting the size distribution.

Wu and Kemeny [95] describe an automatic multi-connected particle segmentation system. Because the particles are touching they represent a multi-connected region in the image. Shadows between particles must then be simply-connected. They find a contour around each shadow regions in the image. Knowledge of rock shape is then used to propose that convex points in the shadow contour correspond to places where particles are touching. It is also assumed that image gradient will be stronger between particles than in a particle region. An heuristic search is performed using image gradient from convex points in the shadows to separate particles. A second algorithm is used to split remaining particles that are still connected because their intensity difference is small. They then use the result of Kemeny *et al* [45] to model each particle with an ellipse and estimate the sieve size using the linear relationship in (2.6). Their procedure produces very accurate results but has difficulty if the lighting is not uniform (for the purpose of detecting the shadows), if rock surfaces are very rough (because it causes strong gradient paths) and in the presence of fines.

2.2.2 Computer Vision Review

According to Haralick and Shapiro [32, 33], "*Computer vision is the science that develops the theoretical and algorithmic basis by which useful information about the world can be automatically extracted and analyzed from an observed image, image set, or image sequence from computations made by special-purpose or general-purpose computers.*" They further state that this information can be related to the recognition of a generic object, three-dimensional description of an unknown object, the position and orientation of the observed object, or the measurement of any spatial property of an object, such as the distance between two distinguishing points or the diameter of a circular section.

Kasturi and Jain [41, 42] give two collections of papers on various aspects of computer vision. They identify that typical computer vision systems consist of the following processing sequence:

- Image capture and enhancement;
- Segmentation;
- Feature extraction;
- Matching features to models;
- Exploitation of constraints and image cues to recover information lost during the image process; and
- Application of domain knowledge to recognize objects in the scene and their attributes.

Their two books give (1) a tutorial, (2) a guide to practical applications, and (3) a reference source on recent advances in computer vision research. A bibliography of selected papers is given which should act as a pointer to related literature. See also the January 1988 edition of the IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 10, Number 1 for a special edition on industrial machine vision and computer vision technology.

Rosenfeld and Kak [77, 78] also provide a source for many image processing techniques which could form part of a computer vision system.

See the beginning of Section 3.2 for a brief overview of image processing terminology and a background of techniques and algorithms.

2.2.3 Human Visual System and Perceptual Organisation

The human visual system is clearly capable of performing the task of detecting and delineating rocks in an image and with a lack of any suitable alternatives, it has been adopted as a basis for the development of this rock detection system.

The human visual perception system gives the impression that the visual world consists of coherently organised, recognisable objects [90], [92, Section 4.3.2]. One is not aware of the processes of detecting colours, edges, movement or distance and the resulting assembly process which produces the internal representation of objects for which identities and labels can be accessed from memory.

A model of the visual system containing two stages is accepted in the area of artificial intelligence research [90]. The first stage involves the extraction of features from patterns

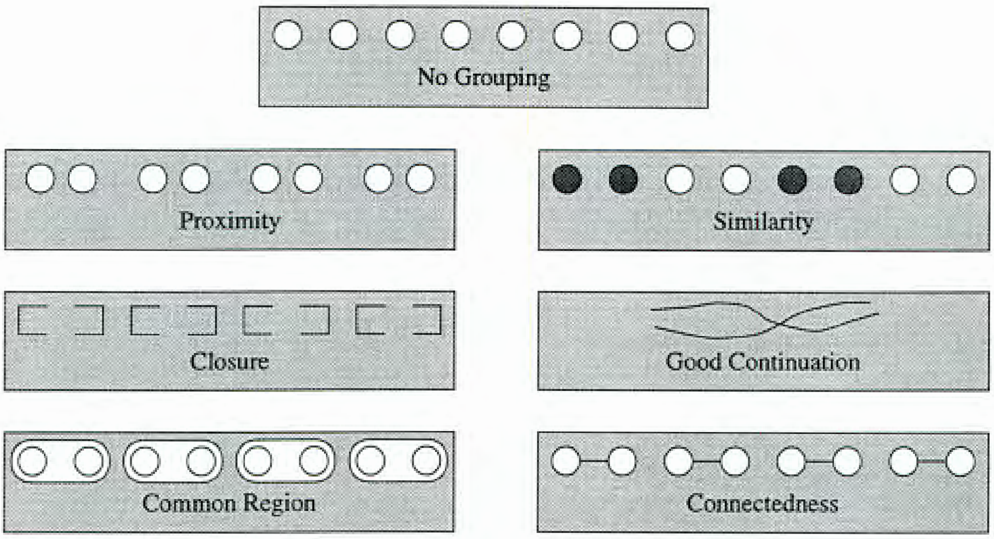


Figure 2.3: The Gestalt laws of grouping.

of light. This is preattentive vision and happens automatically and in parallel for the whole scene. Behaviour of the preattentive vision stage appears to be characterised by the Gestalt laws of grouping (Rock and Palmer [75]) which cause the perception that objects are grouped together if they share any of the following properties: proximity, similarity, closure, good continuation, common region, and connectedness. Figure 2.3 illustrates the Gestalt laws of grouping.

The second stage involves the identification of objects and their settings. This stage is referred to as attention focusing and uses prior knowledge which is typically specific to a domain to examine and classify regions of the scene which were highlighted by the feature detectors in the first stage.

Marr [54] provides a theory of vision that has received a large degree of acceptance and interest [61]. He finds that the analysis of perception can be separated into specialised parts or modules, such as analysis of motion and stereoscopic information, and can be processed independently (See also [94]). Furthermore, Marr proposed that there must exist an additional level of understanding of what is required of the information-processing tasks that is independent of the particular mechanisms and structures that implement them. He proposed a representational framework for deriving shape information from images which consisted of a three stage process. Initially, a primal sketch is obtained which detects the presence of simple features such as lines, edges and groups in the image. The primal sketch is converted to a $2\frac{1}{2}$-D sketch by determining the orientation and depth of visible surfaces and converting them to a viewer-centred coordinate system. The final result is a 3-D model representation of the scene which describes shapes and their spatial organisation.

According to Sarkar and Boyer [80], "*The evolution of perceptual organisation in biological* *vision, and its necessity in advanced computer vision systems, arises from the characteristic that perception, the extraction of meaning from sensory input, is an intelligent process. This is particularly so for high order organisms and, analogically, for more sophisticated computational models.*" They define perceptual organisation as the ability of a vision system to organise features detected in images based on viewpoint consistency and other Gestaltic perceptual phenomena. As with the human visual system, this has two components, a primarily bottom up preattentive part and a top down attentive part. In [80], Sarkar and Boyer provide a review of perceptual organisation in computer vision and in [81, 82] they develop a computational structure to facilitate preattentive perceptual organisation and use perceptual inference networks to manage vision processes for attentive analysis.

2.2.4 Multiscalar Image Analysis

The basic philosophy behind the use of multiscale for image processing applications is given by Lifshitz and Pizer [48], "*Multiresolution techniques attempt to gain a global view of an image by examining it at many resolution levels. The lower resolution provides a global view of the image, and the higher resolution provides the detail.*"

Several advantages are gained with the use of multiscale in many image processing applications. The majority of benefits are associated with the design of matched filters. A brief list is given below:

- A global view of the image is obtained with coarse and fine detail separated [48].
- No filter can be optimum for all scales of image content [53] and since *a-priori* information concerning object scale may not be available, it is often not possible to design an optimum matched filter for an application. This situation is worsened when object scales are different in the same image. Multiscale analysis provides a solution to this problem by allowing particular scales to be analysed separately.
- Multiscale analysis is supported by physiological evidence that low level biological vision systems incorporate multiscalar receptors [54, p54].

In its basic form, multiscale image processing makes use of a scale-space representation of an image. The image is represented by several band-limited images which are arranged in a stack. The bottom level of the stack has the high frequency detail and the top level has low frequency detail. A slight variation of the scale-space theme makes use of a pyramid representation for the stack. In this case, in addition to being a band-limited version of the previous layer, each layer is also subsampled to reduce its size.

These representations are used in a wide range of image processing applications ranging from image coding [1, 10] to image matching [1, 31, 60], shape representation [16, 17], hierarchical segmentation [26, 27, 48] and multiscalar edge detection [53, 69].

Background theory required for the generation of image pyramids is given in Section 3.2.4.

2.2.5 Problem Review

There are basically two problems which have to be solved to correctly obtain a representative size distribution of rocks from an image. The first problem is that of actually identifying rocks in the image and correctly segmenting the image so that their projected size/area can be measured accurately. The second problem is that only a two-dimensional projection of rocks at the surface of the rock-stream can be measured when the parameter of interest is actually the volume size distribution of all the material in the rock-stream.

The problem of determining volume distributions from low-dimension measurements, especially under the conditions where material is deeply piled on the conveyor belt, provides one of the strongest arguments against a machine vision solution to this problem. Lange [46] and McDermott and Miles [56] have investigated estimating volume distributions from random chord-lengths and two-dimensional sections which are found using image processing. It was found that this process can only be successful if the shape of particles is known. As will be found later in the thesis, images of rocks can be reasonably accurately modelled by means of ellipses and ellipsoids which means that volume estimation may not be completely unachievable although it still poses a serious problem which is not investigated here.

This work is primarily aimed at developing an instrument that is able to accurately identify rock boundaries and hence measure projected area from the image. The difficulty of this problem is influenced by the appearance of the rock material on the conveyor belt which is dependent on several factors. Some of the factors are described in the following list with the problems which can be associated with them.

- Material on the belt can be wet or dry. In the case of wet material, fines become attached to the surface of larger rocks which acts as camouflage and increases the difficulty of identifying rocks. A wet surface also increases the possibility of specular reflection which complicates the machine vision process by producing highlights or bright marks on object surfaces.

- The spatial distribution of material can vary from very sparse to a deep pile on the conveyor belt. Clearly, when the material is sparse, it is easier to identify rocks and a more accurate size distribution can be estimated because most of the material is visible. Once material begins to overlap, detection of rocks becomes more difficult due to occlusion and it is necessary to extrapolate to the actual size distribution from the measured one.

- The type of conveyor belt can affect the distribution of material. Wide belts tend to carry a shallow layer of material while narrow belts must carry a deeper layer of material.

- Conveyor belt feed mechanisms can result in different arrangements of material on the belt. For example, a vibrational feed which controls the amount of material loaded produces lumps of material on the belt with gaps between them. This results in the possibility of sampling the belt when no lump or only part of a lump is present. It is therefore necessary to distinguish between fines and conveyor belt surface in order to obtain an accurate estimate of the quantity of fines.

Although encouraging results have been reported by a few of the methods reviewed in Section 2.2.1, there are certain limitations and problems which have been consistently identified. In the first place, there appears to have been very little investigation into the effect of lighting with the result that uneven illumination or changes in contrast between adjacent rocks of different colour can result in incorrect measurements being made because rocks are not accurately segmented.

To further complicate this situation, almost all methods reviewed here aim to measure all sizes of rock at one time with the exception of McDermott and Miles [56] who use a two pass process to measure large and small rocks. This is a problem which is compounded by the fact that the segmentation algorithms generally consist of fairly low-level image processing procedures which are generally sensitive to scale and therefore not suitable for detecting a wide size range of objects [53].

Some high-level information in the form of knowledge has been used with a few of the methods but this is usually in the form of one or two simple rules regarding object shape and there is very little analysis done to check that the results that are being obtained are consistent with what is present in the image. One way of achieving this, which is supported by an understanding of the human visual system and work in perceptual organisation, is to have an internal representation of the object which can be compared to the image. Although Kemeny *et al* [45] use an ellipse to model their rocks this is only for the purpose of calculating and equivalent sieve diameter and is not used in any way to confirm or guide the segmentation process.

A result of these shortcomings is that when there is a high degree of complexity in the image caused by close or overlapping rocks with rough surfaces, non-homogeneity in colour and uneven lighting, many of these methods produce significantly degraded results. The errors can generally be characterised as falling into either of the following categories: splitting occurs when parts of a rock are identified as separate particles (this can either happen when a large particle is split into parts or only a highlight on a particle is labelled) and merging results in rocks which are clustered together being labelled as a single rock.

It can therefore be seen that there are several areas which can be investigated to improve the overall performance of a fragmentation analysis method.

2.3 Research Philosophy

2.3.1 Introduction

The aim of the remainder of this chapter is to give an introductory overview of the research and development achieved with this project. The overview begins in Section 2.3.2 with an analysis of the specific aims of the project which should be addressed in the solution. Based on the decision that the instrument should be modelled on the human visual system, a structure is set up to fulfill the various tasks required for a complete vision system in Section 2.3.3. Section 2.3.4 examines some limitations which are imposed on this work and Section 2.3.5 describes a performance verification and optimisation procedure based on receiver operating characteristic curves. Finally, the contribution of this thesis is discussed in Section 2.3.6.

2.3.2 Statement of Specific Aims

The main aim of this research is to provide a measure of size distribution of rock material as seen from a camera located above a conveyor belt feeding the material into a mill. In order to increase the robustness and accountability of the instrument, a measure based on identification and size measurement of each individual rock was adopted. This approach should provide advantages over some previous techniques which estimated a distribution from a random sample of some low-dimensional feature, such as chord-lengths, using a statistical relationship (for example Lange [46] or Nyberg *et al* [63]).

At the same time it was desired to incorporate as much physical knowledge as possible into the analysis and solution of the problem. This has been achieved in several areas: by selecting appropriate lighting conditions, based on a theoretical analysis, to highlight rock edges; design of a segmentation algorithm which finds rock boundaries based on an internal model of expected rock shape; selection of effective features for the purpose of identifying rocks using feature vector classification; and knowledge of the physical constraints imposed on the possible arrangement of rocks in the rock-stream.

Another way of improving resistance to noise is to include measurements from several different sources using a data fusion approach. Different sources that have been used to improve classification are edge detection (rock edge strength), thresholding (foreground and background) and the morphological tophat operator (highlighting narrow gaps).

It is desirable to obtain a size measurement based on rock shape and area since this could potentially quite significantly reduce errors in estimating volume compared to one-dimensional approaches (See Section 2.1.4). Since volume is proportional to D^3 and area is D^2 the weight of errors will be proportional to $D^{\frac{3}{2}}$. For the example rock-stream in Section 2.1.4 the relative weighting of errors changes from 1:1000 to 1:31 across the range of 1–10 size units.

The algorithms that are developed for detecting rocks should be adaptive in order to compensate for varying degrees of contrast and differences in individual rock intensities. Specifically, it should be capable of segmenting dark rocks correctly even when they are positioned close to much lighter rocks in the image since this condition often caused adaptive thresholds to be too high for the darker object. The instrument should also provide results which are consistent under rotation, reflection, magnification and translation of the image or rocks.

An extra aim is to find a solution which is not rigidly designed for this specific problem but which can be reorganised or the components changed to adapt to other similar applications. Another advantage of this is that a general purpose approach should be less adversely affected by minor changes in conditions and should be easier to correct. Specific operating conditions may be identified for optimum performance, but need not be provided for the instrument to work.

If these aims can be achieved, the instrument should be capable of providing an accurate measure of rock size distribution which is robust and capable of withstanding slight changes in operating conditions. The validity of the distribution measurement will benefit from the following two aspects: each rock will be measured individually and the projected area will be measured instead of a distribution of low-dimensional features.

2.3.3 Modelling the Human Visual System

This section gives a description of the mechanisms that have been used to implement an instrument for measuring rock size distribution. Although there is some mention of hardware, the overview is concerned mainly with the underlying principles of analysis. The basis for the development of this instrument is the fact that the human visual system (HVS) is capable of performing the task of identifying rocks. This section attempts to identify the relevant components and characteristics of the HVS and transform them to a suitable computational model.

Due to the complexity of the problem of detecting rocks in images, an empirical approach was taken when tackling the development of the solution. However, wherever possible, a theoretical foundation was investigated to ensure a sound basis. In the absence of this, a verification process described in Section 2.3.5 has been used to try to determine optimal operating conditions or parameters.

The first observation about the problem is that it is possible to solve it using only intensity information. This can be assumed because it is quite possible for a human to recognise and delineate rocks from a black and white image using only one eye. In this case there is no colour, motion or stereoscopic information so these elements of the HVS can be ignored and a monocular machine vision approach can be adopted, i.e. a single camera view of the rock-stream.

The second observation is that a human identifies each rock separately and delineates it before its size or area can be calculated with any accuracy. The size distribution should therefore be calculated by identifying each rock individually, delineating them to produce labelled regions using a segmentation algorithm, and finally calculating their sizes and shapes from the regions. In addition to increasing confidence in the instrument because it can outline all detected rocks, this can potentially result in more accurate estimates of rock volume since fully two-dimensional information is available. Equations for determining size distribution using shape and area are developed in Section 9.7.

The third observation about the HVS is that when viewing a scene to find an object, the expected size or scale of the object influences the perception of objects. For example, when viewing a scene with a very small rock lying on a large rock, the viewer will initially perceive either the small or the large rock depending on the expected size of the rock. Since rocks typically occupy a range of sizes, it has been necessary to adopt a multiscalar approach to recognising them. Once a definition exists of the appearance of a rock, it will be necessary to scan through several different size scales in order to find all the rocks.

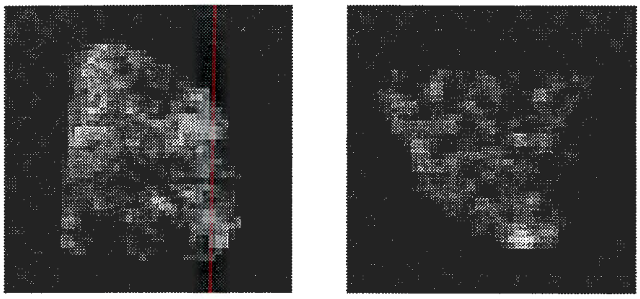
In order to facilitate the segmentation of the image to produce a labelled output, it is useful to provide a model of the expected shape of the object to be segmented. Based on observation of material in typical rock-streams it appears to be feasible to make the assumption that the rocks will be modularly shaped (See Appendix B). Furthermore, Kemeny *et al* [45] have found that it is possible to model the projection of a rock using an ellipse. This information is used to improve segmentation by biasing it toward a particular size and shape, although the segmentation algorithm that has been developed is not restricted to any particular shape.

The next section gives an explanation of the HVS from Section 2.2.3 which has been simplified and adapted to provide a foundation for the development of a computational framework for performing these tasks. After that, the application of a multiscalar approach is investigated and finally, the solution is separated into different components.

A Computational Framework of the Human Visual System

The basic structure of the software component of the system is based on the model of the human visual system which was described in Section 2.2.3 as containing two basic stages: preattentive vision and attention focusing which uses prior knowledge about the problem domain. In the current application, colour, distance and motion need not be considered. In addition, it is only necessary to produce a visual system which responds to one thing, i.e. rocks (this could be further refined to allow only rocks with the characteristics of gold ore, eg modularity).

The preattentive stage is accomplished in this application by identifying elliptically shaped target features using intensity and gradient information. Attention is then focused at these regions in order to locate actual rocks. This process involves segmenting the local region based on the orientation and position of the detected ellipse and incorporates knowledge about the visual characteristics of rocks. Finally, the segmented region is analysed using feature classification to determine whether it is a representation of a rock or not.



(a) (b)

Figure 2.4: An example of the fractal nature of rocks. The two rocks are not the same size, but it is very difficult to determine which is larger from these pictures because of a lack of contextual information. Rock (a) is more than twice as large as rock (b).

A restriction that has been imposed on the attention focusing stage is that it is limited to finding only one size of rock and is therefore effectively normalised. This is the basis of the next key element of the research philosophy which is the use of multiscalar image analysis.

Multiscalar Image Analysis

Section 2.2.4 gives an overview of multiresolution systems and the benefits that are obtained. A basic motivation for the use of a multiscalar representation in many of the applications described there is the analysis of each object in the image at several different resolutions in order to create a hierarchical representation which is invariant to size, position and orientation.

There is a significant difference in motivation in this rock segmentation application. Most applications use multiscale to view an object at several levels of detail. In this case, the desire is to find different sized objects by determining at what position in scale space they appear the same size. This can be explained by considering that each level of the pyramid is a image of the same scene from a greater distance. Given a matched filter for detecting a particular size rock, different rocks will be detected in each image. Their actual size can then be inferred from the scale of the image in which they were found. This method is justifiable due to the fractal or self similar nature of different sized rocks (without suitable contextual clues it is difficult to estimate the size of a rock from a picture of it). Figure 2.4 provides an example of this.

The use of a multiscalar approach to performing rock size measurement has been proposed previously by Lange [46].

Section 2.3: Research Philosophy

Computational Structure

An image processing system for identifying rocks is proposed based on the following four basic procedures: multiscalar image processing, macroscopic region analysis, feature classification and hierarchical analysis of multiscalar results. The following list provides a description for each of these terms:

Multiscalar image processing involves analysing the image at several different size scales in order to detect optimally rocks of different sizes.

Macroscopic region analysis is a process of local image segmentation to group pixels together based on knowledge of the visual appearance of a rock. This knowledge includes shape, the position of edges and edge characteristics, and incorporates contextual information. This is the first part of focusing of attention which delimits the region of interest.

Feature classification is used to discriminate between different regions of interest after the segmentation process. This is achieved by analysing the characteristics of each labelled region by measuring features to create a feature vector for each highlighted target. Representing knowledge about image objects in the form of a feature vector [33, Chapter 19] simplifies recognition and provides a direct interface to statistical pattern recognition which is described in Section 3.4.1. Classification is then used to compare them to an internal representation (which consists of a training set of preclassified samples) to decide which targets represent rocks. This is the second part of attention focusing which performs identification of the object in the region of interest.

Hierarchical analysis uses the knowledge about the layout of detected rocks to resolve any contradiction in the multiscalar results from the previous stages to produce a final list of rocks.

The macroscopic approach has been adopted to contrast with what shall be referred to as microscopic image analysis which has commonly been used in the past for segmenting rock scenes. In a microscopic system, each pixel is considered individually and labelled based on its value and a small neighbourhood. There are several arguments against this approach which is adopted in many applications and is only suitable for early vision processes:

- It is difficult to judge accurately the role of a single pixel without considering sufficient contextual information.
- No consideration is taken of the spatial relationships between various parts of any object being detected if it is larger than the neighbourhood described above.
- It is very difficult to make such a system responsive to any scale information which may require some form of adaption, i.e. the global view of the problem is completely lost.

An analogy for the microscopic approach is that it is like looking at an image through a very small hole. It would be impossible to see any structure in the image. The proposed solution would increase the size of the hole to be slightly larger than the object being detected to allow a complete view of the object and some peripheral or contextual information. In this situation, although analysis of the image is laborious, it is at least feasible.

Physiological evidence provides support for multiscale and macroscopic methods. Marr [54, p54] has reported on the detection of band-limiting filters within the early human visual system which perform smoothing and edge detection at various scales.

When the size of the desired object is known, valuable information about scale of features to be detected can be included in the design. A further advantage may be obtained by considering only one size that an object may have as this allows for the normalisation of the macroscopic features and simplifies the synthesis of matched filters. If objects do in fact occur at different sizes, it is simply a matter of analysing an array of resized images (image pyramid) because the object appears at the fixed size in one of the images.

It can be seen that the proposed solution to the rock recognition problem is significantly more complicated than any of the previous systems that were reviewed in Section 2.2.1. If the goals of the instrument design are achieved, it should provide a significant improvement in performance because of the improved accuracy of an area measurement, the confidence that can be associated with a system that can justify its results, and the incorporation of a multiscalar approach which should result in optimal detection of rocks within a reasonably flexible and general higher level vision structure.

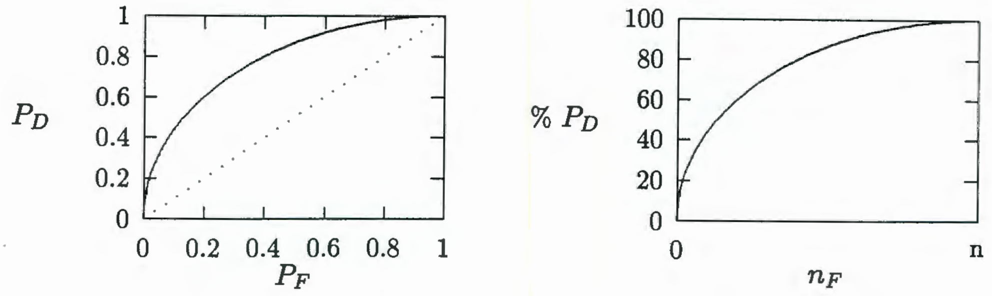
The next section provides a brief investigation into the limitations that have been placed on the performance of the instrument based on the approach that has been adopted. A complete description of the hardware and software or computational structure is given in Chapter 4.

2.3.4 Scope of Solution

As a result of the use of image processing and the computational structure that has been proposed, the following limitations will be inherent in the instrument:

- All that can be seen is the projected areas of the rocks at the surface of the rock-stream. Although the volume distribution is desirable, it is not possible to measure it directly from an image. The projection process also results in a loss of information which makes it difficult to actually estimate the volume distribution accurately and this has not been attempted.

- Because the instrument is designed to identify and label individual rocks, it will not be possible to get a measurement of the amount of fines in the image using the same method because at some point, rocks will be too small to have sufficient resolution to be



(a) Generic ROC

(b) Modified ROC

Figure 2.5: Two methods for displaying the receiver operating characteristic (ROC) curve.

recognised. It will therefore be necessary to consider another approach for estimating the amount of fines in an image but this has not been considered.

- Because more smaller rocks can fit into an image and therefore be detected at once, the measured distribution should be most accurate for the smallest sizes and decrease in certainty as the size increases. For this reason it may be desirable to operate on a range of input image sizes. If small images are used, results will only be obtained for large rocks. A carefully shaped pyramid of successive input images could result in uniform certainty of all sizes in the distribution. This would also have the advantage of decreasing execution time since a small image can be processed more quickly than a large one which would result in a higher average sampling rate.

Aside from these issues, the proposed instrument design should satisfy the aims of the project that were presented in Section 2.3.2. The next section describes a procedure that will be used for measuring the performance of the instrument.

2.3.5 General Verification and Optimisation Procedure

This section describes a mechanism that is used to estimate the performance of the instrument for the purpose of optimisation. Because the software component of the instrument which is described in this thesis is complicated and relies on the values of several parameters, it is necessary to have a means of measuring the performance given any particular set of values for the parameters. A method for achieving this that has been adopted is called the receiver operating characteristic (ROC) which is well known in one-dimensional detection theory [23, p80].

In detection theory there can usually be measured a probability of detection P_D and a probability of false alarm P_F which vary continuously as some threshold t is changed. Typically, when t is small, the resulting $P_D \approx 1$ and $P_F \approx 1$. As the threshold is raised,

P_D and P_F decrease monotonically until $P_D \approx 0$ and $P_F \approx 0$. It is therefore possible to plot P_D versus P_F as a function of the parameter t to summarise the performance of the detector. This locus of probability pairs is known as the receiver operating characteristic (ROC). Figure 2.5(a) shows a generic ROC curve as a solid black line. The dotted line represents the relationship $P_D = P_F$ which could be obtained by flipping a suitably biased coin and therefore indicates no improvement in detection. When the detector is suitable, the resulting ROC curve should always be above and to the left of $P_D = P_F$. The better the detector, the further its ROC curve will be from $P_D = P_F$.

Given a list of actual rocks in an image and a list of detected rock positions, it is possible to determine P_D using a distance measure which is described in Appendix D. If the minimum distance d from a detected rock position to the closest actual rock in the image is less than 1, then according to the definition of the distance measure, the detected rock corresponds to the rock and it has therefore been correctly detected. P_D is thus the proportion of the actual rocks that are correctly detected at least once.

If the minimum distance $d > 1$ then the detected position does not correspond to a rock and is thus a false alarm. It is possible to obtain a rough estimate of the maximum possible number of false alarms by summing all the possible positions at which a rock can be detected and subtracting the positions which are close to a rock. Given the size of the base of the pyramid is $n \times m$, the number of levels is l and the relative decrease in size between adjacent levels is a fraction f of n and m then the total number of positions is $\frac{nm(f^{2(l+1)}-1)}{f^2-1}$ which is obtained as the sum of a geometric progression. If a rock can be detected at o possible orientations, the number of detections increases by a factor o. Given that the number of rocks in the image is r and each rock can be correctly detected at approximately 15 positions then the number of true detections is $15r$.

The following list gives realistic values for each of the parameters: $n = 256$, $m = 170$, $l = 20$, $f = 0.9$, $o = 17$ and $r = 30$. Under these conditions it can be seen that the maximum number of possible true detections is $15r = 450$. The total possible number of detections is around 4 million which corresponds to an *a priori* probability of finding a rock of $p \approx \frac{1}{10000}$. Clearly, since the actual number of false alarms is only a few thousand after the first stage of preattentive vision and decreases to less than 10 for the final result, the normal way of calculating P_F will almost always result in a vanishingly small value and will not be particularly useful for analysing performance.

Because of this, a variation of the ROC curve has been used in this thesis. Instead of calculating a probability of false alarms P_F, the number of false alarms n_F is used instead. Although there is no longer an absolute reference in the form of the $P_D = P_F$ line, it is still possible to compare ROC curves obtained under different operating conditions or using different parameter combinations because good performance will still correspond to a curve which is above and to the left of other curves when the performance is not as good. Figure 2.5(b) shows a modified example of the ROC curve in which P_D as a percentage is plotted against n_F as a function of the threshold t.

This procedure will be used for determining the best operating parameters for various values during preattentive vision, classification and hierarchical analysis. The exact requirements of each of the modules are not all the same and so the selection of the best operating parameters may vary with the different modules by identifying different parts of the curve.

2.3.6 The Contribution of this Thesis

The work described in this thesis contributes to the field of particle size measurement using machine vision systems. Previous attempts at solving this problem are described in Section 2.2.1. The solution that is proposed here should provide an improvement in performance due to the innovative consideration of the following factors for this application:

- Using multiscalar image analysis via an image pyramid with the slightly different approach that the pyramid is used to make all rocks the same size rather than to view each rock at different sizes.

- Analysis of the effect of diffuse lighting illustrated that some useful information is contained in image intensity and not just gradient.

- A computational structure based on the human visual system is used and thus breaks away from more traditional low-level machine vision approaches to the problem of rock size measurement. This is characterised by a two stage process of preattentive target highlighting followed by attention focusing to identify the targets. This is achieved by hypothesising that each target can be modelled by a particular shape corresponding to how it was detected. The model is then compared to the image in order to determine whether or not it actually maps to the image content at that point. Internal knowledge in the form of a training set and implicitly encoded in algorithms is also used.

- The solution is not designed to operate under very specific conditions and due to flexibility introduced after analysis of effects of lighting and the combination of several stages of processing, the system performs reasonably under a variety of conditions. Certain artifacts which are introduced under different conditions such as shadows will be detrimental to the performance of the system but are not as severe as with simpler approaches.

- A segmentation algorithm has been used which uses not only image gradient information but also object size, shape and intensity to improve its performance by imposing constraints on the resulting labelled regions. This is possible due to the multiscalar approach and the fact that an internal object model is hypothesised and then compared to the image for verification.

As a result of the factors just described, it is expected that the solution will be better for the following reasons:

- It provides a robust size distribution in that the value is based on identifying individual rocks and calculating their size and so can easily be verified by comparing the detected rocks to those that are actually in the picture.
- The use of diffuse lighting is shown to provide advantages over the more conventional approach of using directional illumination in that it reduces shadows therefore decreasing the chance that facets of a rock will have different intensities and reduces the effect of specular reflection from wet material. Both of these benefits tend to decrease the chance of splitting a rock because the spurious features will be absent from the image.
- The multiscalar nature of the solution allows the instrument to be more precisely matched to the problem with the result that the performance should also improve.

The following chapter gives background theory which is required for the development of the remainder of the thesis. After that, in Chapter 4, the actual structure of the solution is presented and motivated.

Chapter 3

Background Theory

3.1 Introduction

This chapter is concerned with providing theoretical background for various techniques which are used in later chapters for the complete description of the rock detection system and ultimate goal of measuring rock size distribution.

It is divided into three broad sections. Section 3.2 is concerned with developing a background of basic image processing techniques and algorithms. Theory for the analysis of image formation is described in Section 3.3 and finally, Section 3.4 provides a theoretical background to important issues related to vector classification which is required for discriminating between selected regions of interest which could contain rocks.

3.2 Image Processing Fundamentals

An image can be represented by means of a two dimensional function such as $f(x, y)$ where x and y are horizontal and vertical spatial coordinates. In the case of a black and white image, the function returns the image intensity, brightness or grey level for the given point. It could also represent a colour vector or depth information.

A digital image is created from a continuous image by a process of digitisation which involves two steps [77, Chapter 4]: (1) Sampling is used to extract a discrete set of real image points or samples. (2) The discrete samples are quantised to the possible range of values. This results in an integer array or digital image. The elements of this array are referred to as picture elements, pixels, pels or points.

More complete introductions to image processing terminology and representation can be found in [30, Chapters 1,2], [62, Section 1.2], [70, Section 1] and [77, Section 1.2]

Figure 3.1 shows an example image of a rock scene captured under controlled diffuse lighting in the laboratory. The material was dry and no fines were included. This image will be used for demonstrating various techniques and algorithms throughout the thesis.

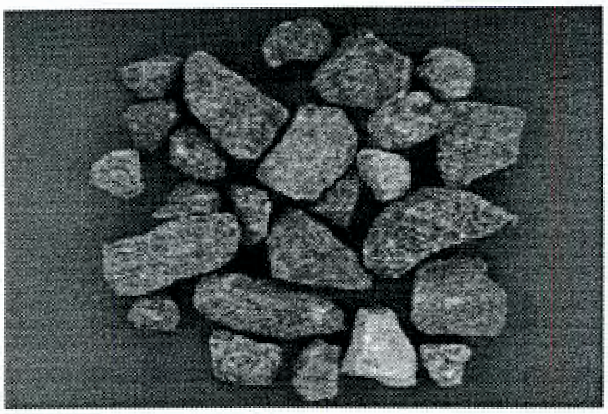


Figure 3.1: Example image of a rock scene captured in the laboratory with controlled diffuse lighting.

The remainder of this section provides background to edge detection, image thresholding and greyscale morphology which are common methods applied in image processing. They are described in Sections 3.2.1, 3.2.2 and 3.2.3 respectively. The creation of image pyramids for multiscalar image processing is described in Section 3.2.4. The Hough transform is described in Section 3.2.5 with special emphasis on the detection of ellipses. A brief introduction to knowledge based image processing systems is given in Section 3.2.6.

3.2.1 Edge Detection

Edges are lines or curves in the image where intensity changes rapidly. Horn [36, Chapter 8] suggests that edge detection is complementary to image segmentation since detected edges may be used to divide an image into different parts corresponding to different surfaces.

Changes in intensity which represent significant surface features are of interest in this project. These could be discontinuities in surface orientation, one object occluding another, the line around the edge of a shadow, or an abrupt change in surface reflectance properties.

A simple edge can be thought of as a border between two regions, each of approximately uniform intensity. It would be natural to expect a cross section of the intensity perpendicular to the edge to have a step discontinuity but this is not usually the case as blurring, noise and limitations in imaging devices tend to cause the edges to be more ramp like.

Noise places a limit on the amount of information about edges which can be retrieved by edge detectors. Edge detectors can respond to noise thus producing spurious output. Usually, edges corresponding to noise are small and can be removed by ignoring edges with a strength below a threshold. If the sensitivity or threshold below which edges are ignored is too high, then weak or short edges may be missed altogether. On the other hand, if the threshold is too low then a lot of spurious edges may be included which mask the location of real edges.

Canny [11] proposed the following criteria that should be observed when designing an edge detecting filter:

Good detection — the filter should detect edges where there are edges and not where there are not. This corresponds to maximising the signal to noise ratio.

Good localisation — the detected position of an edge should be as close as possible to the true position of the edge.

Single response — it is undesirable for the edge detector to have multiple responses to the same edge as this complicates the task of selecting the correct edge since only one can be optimal with regard to localisation. The filter should thus have a single peak response at an edge.

In addition the filter should have zero dc component so as not to respond to areas of constant intensity.

Edge detection is usually achieved in practice using the first or second derivatives of the image (See [30, Section 7.1.3], [70, Chapter16] and [77, Section 10.2] for a description of several edge detection operators.)

Simple edge detection filters such as the magnitude of the gradient and Sobel operators have been investigated but it was found that, given edge direction information and the particular requirements of the segmentation algorithm, directional nonlinear operators would provide better performance. The following sections describe two directional edge detection operators that are used for detecting rock edges while attempting to adhere to the Canny criteria.

Directional Gradient Detection

A directional edge detection operator has been developed based on a nonlinear curve detection operator described by Rosenfeld and Kak [78, Section 10.3.1 and 10.3.2]. The simplest case to consider is that of a vertical edge detector where the intensity increases from left to right. Given the image is g and the point of interest is (x, y) then a simple nonlinear edge detector has output

$$g(x+1,y-1)-g(x-1,y-1)+g(x+1,y)-g(x-1,y)+g(x+1,y+1)-g(x-1,y+1) \quad (3.1)$$

whenever

$$g(x+1,y-1) > g(x-1,y-1).g(x+1,y) > g(x-1,y).g(x+1,y+1) > g(x-1,y+1) \quad (3.2)$$

and is zero otherwise[1]. This operator only produces an output when the intensity is increasing from left to right and then it is proportional to the first derivative of the slope.

---

[1] Note that the point . is used to denote the logical AND operator

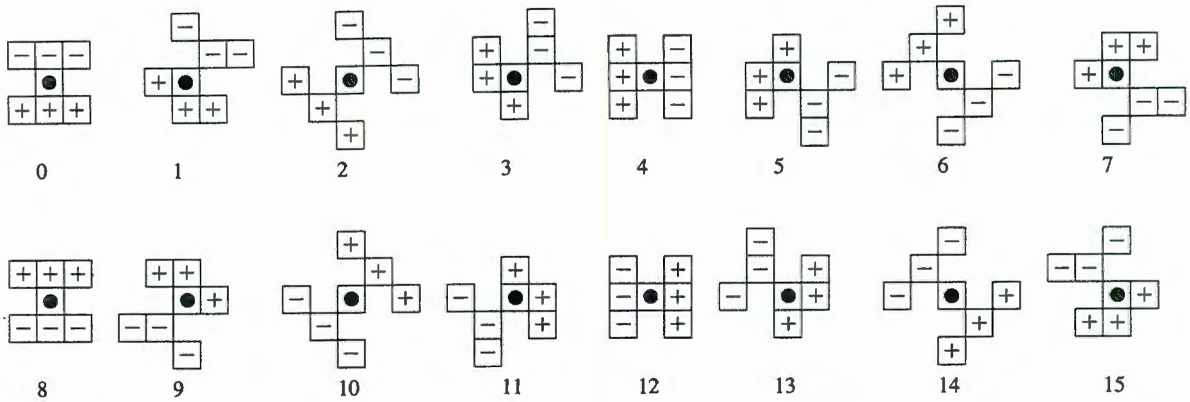


Figure 3.2: Edge detection masks for detecting directional intensity gradient. The shaded circle is the point for which edge strength is to be calculated and this is achieved by adding the positive points and subtracting the negative ones only if an edge is present according to an equivalent form of (3.2).

This definition for an edge detection operator attempts to satisfy the first two Canny criteria. Only detecting edge strength where the condition is satisfied and using three adjacent points improves detection by detecting edges only where they are present and improving signal to noise ratio by averaging. The localisation of the operator is good because only a small region around the point (x, y) is considered. In terms of single response, the operator will produce its highest output at the location of the steepest gradient, but this will be spread over the whole region where the intensity is increasing.

The operator was generalised by defining 16 directions in which an edge may be detected. This is shown in Figure 3.2 where the point (x, y) is marked with a circle and the value of the output is produced by adding the points marked with a $+$ and subtracting those with a $-$.

The edge detector could be used by considering all of the directions and recording the highest edge strength for each point and its direction. In this application, the desired edge direction is available since the shape of the rock being detected will be known. This information is used to create an edge direction map which informs the edge detection function which direction to use at each point in the image. A variety of shapes can be optimally detected by presenting different direction maps tailored to their shape.

Orientated Valley Detection

Pearson and Robinson [68] describe the use of valley detection operators for finding useful edges in images of humans based on an analysis of edges created under a simple lighting arrangement. They observed that most relevant features occur at intensity minima which form valleys in a two-dimensional image.

Detecting valleys in an image using convolution with a valley-shaped operator corresponds to taking the second derivative of the image. It is possible to create isotropic and oriented valley detectors for two-dimensional use by either rotating the valley on its axis or sliding it perpendicular to the section.

A simple valley detector will be described here which uses the same orientations from Figure 3.2 as the directional gradient operator. An example of output for the case of detecting a vertical valley using a nonlinear valley detector is

$$\begin{aligned} & g(x+1,y-1)+g(x-1,y-1)+g(x+1,y)+g(x-1,y)+g(x+1,y+1) \\ + \quad & g(x-1,y+1)-2(g(x,y-1)+g(x,y)+g(x,y+1)) \end{aligned} \tag{3.3}$$

whenever

$$\begin{aligned} & g(x+1,y-1)>g(x,y-1).g(x-1,y-1)>g(x,y-1). \\ & g(x+1,y)>g(x,y).g(x-1,y)>g(x,y). \\ & g(x+1,y+1)>g(x,y+1).g(x-1,y+1)>g(x,y+1) \end{aligned} \tag{3.4}$$

and is zero otherwise. This operator only produces an output when the intensity is lower in the middle than at either side and then it is proportional to the second derivative of the valley.

The same arguments concerning adherence to the first two Canny criteria apply to this operator as for the directional gradient operator since it has the same area of support; however it will produce an edge in the region of a local minima rather than steepest gradient. Again, this operator may be used with a direction map which specifies the probable orientation of an edge given the position in the image.

3.2.2 Thresholding

Thresholding is a basic image segmentation process which is necessary in many image analysis applications, and is discussed in many texts [29, Section 7.3], [32, Section 2.2], [62]. The simplest form of thresholding, *bilevel thresholding*, which is investigated here, involves classifying pixels of an input image into two groups, generally foreground and background. In this form a threshold is calculated and then all pixels with their grey values above the threshold level become foreground and the rest are background.

There are a couple of problems associated with this method. The first is the selection of the threshold level. The second is the fact that when a threshold level is applied globally to the whole image the result is usually inadequate due to varying lighting which causes global changes in image intensity. This can be overcome by the use of local area thresholding which is described later in this section.

A solution to the problem of determining a suitable threshold level given the region to be thresholded is the moment-preserving threshold algorithm.

Moment-Preserving Threshold

The moment-preserving threshold is a method presented by Tsai [91] of calculating the appropriate threshold to use for a given image. According to Symmonds [87], Tsai's algorithm automatically computes the optimum threshold level using the first three moments (mean, variance and skewness) of the image. The threshold is chosen so that the moments of the original image are preserved in the thresholded image.

The grey level histogram of an image with 8 bit quantisation is $[n_0, n_1, \ldots, n_j, \ldots, n_{255}]$ where the n_j entry denotes the number of pixels with grey level j. If the total number of pixels in the image is N, the grey level probability histogram can be calculated using the following expression:

$$P_j = \frac{n_j}{N} \text{ for } j = 0, \ldots, 255$$

The i'th moment of the grey level probability histogram is given by:

$$m_i = \sum_{j=0}^{255} P_j\,(j)^i \tag{3.5}$$

From (3.5) the first four moments are calculated using:

$$m_0 = \textstyle\sum_{j=0}^{255} P_j = 1 \quad m_1 = \sum_{j=0}^{255} P_j(j)$$

$$m_2 = \textstyle\sum_{j=0}^{255} P_j(j)^2 \quad m_3 = \sum_{j=0}^{255} P_j(j)^3$$

For bilevel thresholding, a single threshold T will be used to classify image pixels into two classes with mean grey levels z_0 and z_1. If p_0 is the fraction of pixels below T and p_1 is the fraction above T, then the moments for the thresholded image are given by:

$$m'_i = \sum_{k=0}^{1} p_k(z_k)^i$$

which on expanding yields:

$$\begin{aligned} m'_0 &= p_0(z_0)^0 + p_1(z_1)^0 \\ m'_1 &= p_0(z_0)^1 + p_1(z_1)^1 \\ m'_2 &= p_0(z_0)^2 + p_1(z_1)^2 \\ m'_3 &= p_0(z_0)^3 + p_1(z_1)^3 \end{aligned}$$

The moments of the original image are preserved in the thresholded image by equating

m_i and m_i' (for i=0,1,2,3), which produces four equalities which are known as the moment-preserving equations for bilevel thresholding:

$$p_0 + p_1 = 1 \tag{3.6}$$

$$p_0 z_0 + p_1 z_1 = m_1 \tag{3.7}$$

$$p_0(z_0)^2 + p_1(z_1)^2 = m_2 \tag{3.8}$$

$$p_0(z_0)^3 + p_1(z_1)^3 = m_3 \tag{3.9}$$

The equations of (3.6), (3.7), (3.8) and (3.9) have to be solved to obtain p_0 and p_1. The threshold T can then be calculated by determining the p_0-tile from:

$$p_0 = \sum_{j=0}^{T} P_j$$

Since this cannot usually be evaluated exactly, the required threshold value is taken as the nearest grey level value.

Solving The Moment-Preserving Equations

Equations (3.6), (3.7), (3.8) and (3.9) need to be solved for the four unknowns p_0, p_1, z_0 and z_1. Expressions for solving p_0 and p_1 are found by rewriting equations (3.6) and (3.7) in matrix format.

$$\begin{bmatrix} 1 & 1 \\ z_0 & z_1 \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \end{bmatrix} = \begin{bmatrix} 1 \\ m_1 \end{bmatrix}$$

The unknowns p_0 and p_1 can be solved using Cramer's rule to give:

$$p_0 = \frac{1}{p_{det}} \begin{vmatrix} 1 & 1 \\ m_1 & z_1 \end{vmatrix} \tag{3.10}$$

$$p_1 = 1 - p_0 \tag{3.11}$$

where:

$$p_{det} = \begin{vmatrix} 1 & 1 \\ z_0 & z_1 \end{vmatrix} = z_1 - z_0$$

Tsai has shown that the moment-preserving equations can be solved indirectly by following these two steps:

Step 1 Using moment values m_0, m_1, m_2 and auxiliary values c_0, c_1 and c_2, a set of linear equations can be established:

$$c_0 m_0 + c_1 m_1 + c_2 m_2 = 0$$

$$c_0 m_1 + c_1 m_2 + c_2 m_3 = 0$$

Setting c_2 to 1 and rewriting in matrix form gives:

$$\begin{bmatrix} m_0 & m_1 \\ m_1 & m_2 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} -m_2 \\ -m_3 \end{bmatrix}$$

The unknowns c_0 and c_1 can be solved for using Cramer's rule to give:

$$c_0 = \frac{1}{c_{det}} \begin{vmatrix} -m_2 & m_1 \\ -m_3 & m_2 \end{vmatrix} \qquad c_1 = \frac{1}{c_{det}} \begin{vmatrix} m_0 & -m_2 \\ m_1 & -m_3 \end{vmatrix}$$

Where:

$$c_{det} = \begin{vmatrix} m_0 & m_1 \\ m_1 & m_2 \end{vmatrix} = m_0 m_2 - m_1^2$$

Step 2 The grey levels z_0 and z_1 are obtained by solving the following second order polynomial:

$$z^2 + c_1 z + c_0 = 0$$

Using the standard formula for the solution of second order polynomial equations gives the two results for z as:

$$z = \frac{-c_1 \pm \sqrt{c_1^2 - 4c_0}}{2}$$

Figure 3.3 shows the result of calculating the moment-preserving threshold for Figure 3.1 and using that value to segment the image into foreground and background.

Local Area Thresholding

Local area thresholding is used to overcome the problem with global image thresholding. A threshold value is calculated for each pixel in the image based only on the histogram of a relatively small window or neighbourhood around the point. In this way the threshold value can adapt to the changing intensity levels in different regions of the image.

The moment-preserving threshold can be used in a local area thresholding program as the algorithm for calculating each of the threshold levels for the given histogram. In practice, however, this is extremely computationally intensive and an alternative solution was chosen. The moment-preserving algorithm is only used to calculate the threshold values on a grid of points across the image. All intermediate threshold values are calculated by interpolating between the grid points using a bi-linear interpolation routine (See Section 3.2.4).

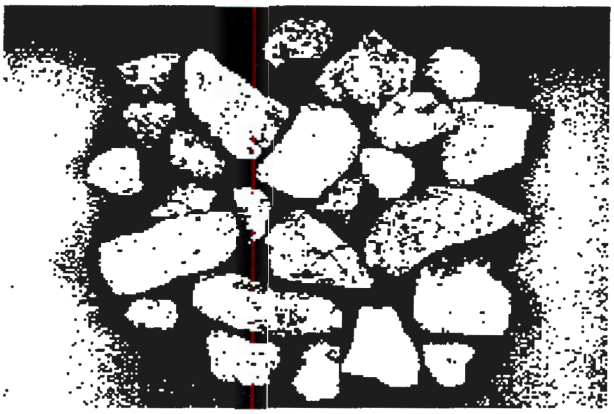


Figure 3.3: Rock image segmented by globally applying the moment-preserving threshold determined from a histogram of the whole image. A global thresholding technique suffers from the problem that it is not able to adapt to regions of the image where another threshold value would be more suitable.

Figure 3.4 shows the result of applying the moment-preserving threshold in a local area method with a window size of 15 × 15 pixels to image 3.1. It also shows the results obtained when interpolation is used. The grid spacing is 16 pixels. As can be seen there is very little visible difference between the two solutions and for a window size of 15 × 15 the interpolation method is about 250 times faster.

In the local area thresholding method, far fewer rock boundaries have been corrupted. Large rocks do tend to become fragmented in their interiors but this is only when their size is large relative to the local window size. The window size must therefore be chosen carefully to achieve the best results for the rock size which is to be detected.

3.2.3 Greyscale Morphology

Introduction

Mathematical morphology provides an approach to image processing which is based on the shape of geometrical structures in the image which is considered to be a landscape. This is in contrast to traditional linear systems which do not address this issue and are therefore often limited for image-like signals. When used appropriately, morphological operations tend to simplify image data by preserving their underlying shape characteristics and eliminating small irrelevant features in a nonlinear manner.

Mathematical morphology is based on set theory and nonlinear superpositions of signals. Morphology has been defined for both binary and grey-scale images. Dougherty [19] and Dougherty and Astola [20] give introductory tutorials in background theory of mathematical morphology and examples of applications and techniques. Maragos and Schafer [52] provide a review of morphological systems to show how they can enrich the theory and applications

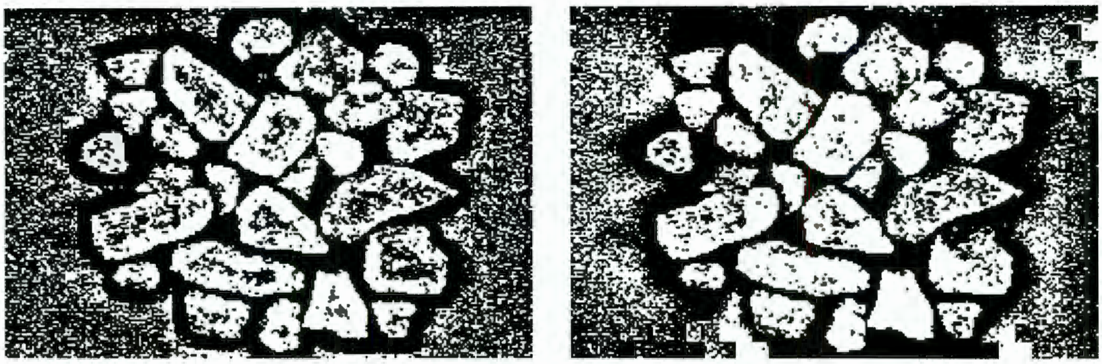

(a) (b)

Figure 3.4: Rock image segmented by locally applying the moment-preserving threshold determined from the histogram of a 15 by 15 window in (a). For the second image (b) the threshold was only calculated for a grid with spacing of 16 pixels. The intermediate threshold values were interpolated using bi-linear interpolation. Local thresholding is matched to objects with a size proportional to the window and is able to adapt across the image. The second form of local thresholding gives comparable results to ideal local case (a) with a dramatic improvement in computational cost. Note that the blocking to the right and bottom of (b) is caused by the interpolation routine resorting to the nearest neighbour method because required pixels are out of the image.

of multidimensional signal processing. A thorough tutorial on the basis and theory of mathematical morphology is given by Haralick *et al* [34].

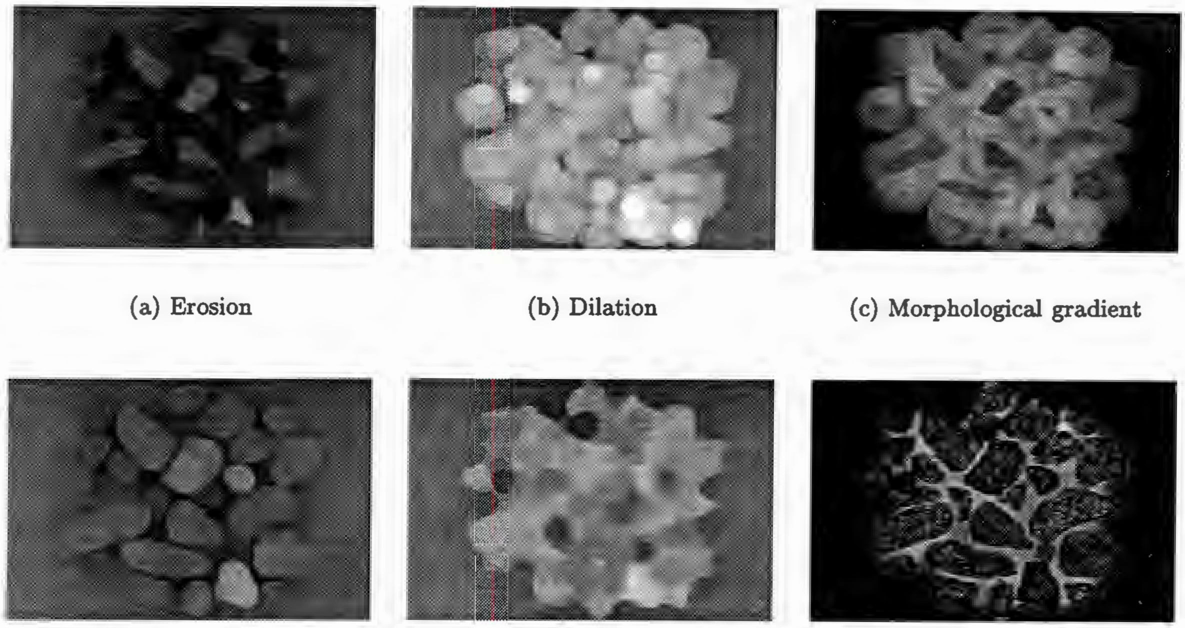
Mathematical morphology has been applied to a number of image processing problems. Several examples are cited in [34], but in addition, Maragos [51] has defined a measure of match which is based on the ℓ_1 norm and is shown to be related to morphology. Sternberg [86] uses greyscale morphology and the concept of image umbra to perform image enhancement for representation.

The next section gives a brief overview of some greyscale morphological image processing operators.

Morphological Operators

Six morphological operators are described in the following subsections. The first two: erosion and dilation, form the basis from which the other operators are derived. Examples of the application of these operators to a sample image are given in Figure 3.5.

Erosion and Dilation The representation of a function's *umbra* has been used by Sternberg [86] to define the morphological operators. Given a *d-dim* function $f(\mathbf{x})$ where $\mathbf{x}$ is a *d-dim*



(a) Erosion (b) Dilation (c) Morphological gradient

(d) Opening (e) Closing (f) Tophat transform

Figure 3.5: Examples of the effects of greyscale morphological operators on the sample image. The morphological structuring element used was a disc of zero height and radius 7. The sample image has dimensions 256 × 170. Note the tophat transform has been applied as a valley detector and the resultant intensities have been brightened for both (c) and (f).

vector, then its umbra is represented by a *(d+1)-dim* set

$$U(f) = \{(\mathbf{x}, a) : a \leq f(\mathbf{x})\}$$

The umbra is the set of points below the surface represented by $f(\mathbf{x})$. The umbra set extends to $a = -\infty$ and the function can be reconstructed from its umbra since

$$f(\mathbf{x}) = T[U(f)] = \max\{a : (\mathbf{x}, a) \in U(f)\}$$

is the surface of the umbra. Erosion and dilation are then defined to be the surface of the umbra of the image eroded or dilated by the umbra of the structuring element, i.e. $f \ominus g = T[U(f) \ominus U(g)]$ and $f \oplus g = T[U(f) \oplus U(g)]$ respectively.

Erosion and dilation can be computed directly from the following formulae

$$(f \ominus g)(\mathbf{x}) = \min\{f(\mathbf{x}+\mathbf{z}) - g(\mathbf{z}) : \mathbf{z} \in G \text{ and } (\mathbf{x}+\mathbf{z}) \in F\}$$
$$(f \oplus g)(\mathbf{x}) = \max\{f(\mathbf{x}-\mathbf{z}) + g(\mathbf{z}) : \mathbf{z} \in G \text{ and } (\mathbf{x}-\mathbf{z}) \in F\}$$

where $G = \mathrm{supp}(g)$ and $F = \mathrm{supp}(f)$. In this context, the *support* of f means the set of $\mathbf{x}$ at which $f(\mathbf{x}) \neq -\infty$. The function g which plays the part of a structuring element usually has a small support.

Examples of the effects of erosion and dilation are shown in sub-figures (a) and (b) of Figure 3.5 respectively.

Opening and Closing The opening of an image f by a structuring element g is denoted by $f \circ g$ and is defined as $f \circ g = (f \ominus g) \oplus g$. An intuitive explanation of the operation of the opening operator is to consider the image f to be a landscape, where the value $f(x, y)$ is height. Opening is then the operation of sliding the structuring element along beneath the image and determining for each position how high the structuring element may be pushed up.

Similarly, the closing of the image is denoted by $f \bullet g$ and is defined as $f \bullet g = (f \oplus g) \ominus g$. Closing is equivalent to the inverse of opening the inverse image with the inverse structuring element i.e., $f \bullet g = -[(-f) \circ (-g)]$. Intuitively, it is equivalent to determining how far down the structuring element can be pushed at each point when slid along the top of the image.

Opening and Closing are illustrated as sub-figures (d) and (e) of Figure 3.5 respectively.

Morphological Gradient Morphological gradient in one form is defined by $GRAD(f, g) = (f \oplus g) - (f \ominus g)$ where f is the image and g is a flat structuring element centred about the origin [19, Section 7.1].

The dilation and erosion by the flat structuring element produces maximum and minimum filters, respectively, at each point. The morphological gradient therefore determines the difference between the maximum and minimum values in the neighbourhood of each point defined by the structuring element. An example of the effect of the morphological gradient operator is shown in sub-figure (c) of Figure 3.5.

Tophat Transform The tophat transform of an image f by a structuring element g is denoted by $HAT(f, g)$ and is defined as $HAT(f, g) = f - f \circ g$. This is clearly the difference between the image and the opened image. Due to the nature of the opening operator, the tophat transform is guaranteed to be non-negative for all points [19, Section 7.2].

The effect of the tophat transform is to detect intensity peaks in the image by removing the smooth opened image. Only features which are too small for the structuring element to fit into from below will be preserved and all others will be completely removed. The tophat transform can be used to detect valleys instead of peaks. This is achieved with the formula $(f \bullet g) - f$ which can be shown to be equivalent to $HAT(-f, g)$ [19, Section 7.2]. The valley detection operation of the tophat transform is shown in sub-figure (f) of Figure 3.5.

3.2.4 Image Pyramid Generation

Multiscalar image processing was introduced in Section 2.2.4 and its use was motivated in Section 2.3.3. Although it was found that the motivation for using image pyramids appears different for this application, the same process is used for generating them and is described here.

Typical image pyramid generation involves an iterative process of smoothing followed by resampling to reduce the size. This results in a scale-space representation of the image. Usually, the Gaussian G is used as a smoothing kernel, and in most cases a dyadic pyramid is generated, i.e. the resolution of each level is an integer power of two. For that case resampling is trivial since each sample falls exactly on alternate samples in the smoothed image.

For image shape description, zero-crossings in a Laplacian pyramid are often used. The ∇^2 operator is useful because it is the only orientation independent second order differential operator. A computationally efficient method for creating a Laplacian Pyramid involves calculating the difference of low pass (DOLP) filtered images. This provides an approximation to $\nabla^2 G$ which has a bandpass characteristic. Motivation for the use of the $\nabla^2 G$ operator is that there is strong evidence [54, p54] that the receptors in the human eye perform immediate filtering of this form. In addition, Rogers and Kabrisky [76, p30] state that implementing a similar system for edge enhancement substantially improved the performance of their neural network pattern recognition system.

A paper by Zuerndorfer and Wakefield [97] attempts to generalise scale-space theory to include sensors with non-Gaussian characteristics and to extend zero-crossing detectors to general function-crossing.

The requirements of the current problem include an arbitrary scale factor between pyramid levels and similar visual characteristics for each level. In effect, each level should appear like a picture of the same scene taken from increasing distances so that it is not possible to differentiate between scales when examining detail of similarly sized objects. The following two sections discuss the smoothing and resampling requirements for pyramid generation.

Smoothing

A Gaussian kernel is usually used for smoothing images in scale-space applications. The main motivation for this is based on *causality*, that is, no new extrema are created as the image is smoothed. This result was achieved by Lifshitz and Pizer [48] after solving the diffusion equation for an input image. Furthermore, Marr and Hildreth [53] suggest that both the spatial and frequency responses should be small and find that the optimum distribution for this conflicting requirement is the Gaussian.

$$G(x, y; \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.12}$$

Non-Gaussian kernels have been used in some cases. For example, Meer *et al* [57] designed an optimal low pass filter using the *equiripple approximation.* For a certain width of transition band, several of the filter weights become negligibly small and the obtained solution becomes similar to a *half-band* filter. This can lead to particularly computationally efficient results. Note that this filter does however tend to contradict the requirement of Marr and Hildreth [53] for a low spatial response. Perona and Malik [69] also make use of non-Gaussian smoothing by varying the diffusion constant across the image in order to preserve strong edges while decreasing noise. Due to the forward nature of this anisotropic diffusion process, the result is stable.

Lindeberg [49] argues that image processing applications are discrete and that the predicted Gaussian behaviour in continuous space with respect to the causality of extrema does not commute to discrete space when a kernel is generated by sampling a Gaussian. He develops axioms of scale-space and solves a discrete diffusion equation which results in the following separable kernel:

$$K(x,y;t) = T(x;t)T(y;t) \qquad t > 0 \tag{3.13}$$

where $T(n;t) = e^{-t}I_n(t)$ and I_n are the modified Bessel functions of integer order. This is the discrete equivalent of the Gaussian kernel.

Minimal overhead is required to use discrete Gaussian and it has a more robust theoretical background. The convolution mask size will be the same as that calculated for the Gaussian.

Resampling

Due to the arbitrary scale factor requirement for this problem, it is necessary to be able to subsample the smoothed image using a sample distance which results in sample points that do not coincide with a sample position in the smoothed image. It is therefore necessary to perform some sort of interpolation [62, Section 6.4] in order to estimate the sample values.

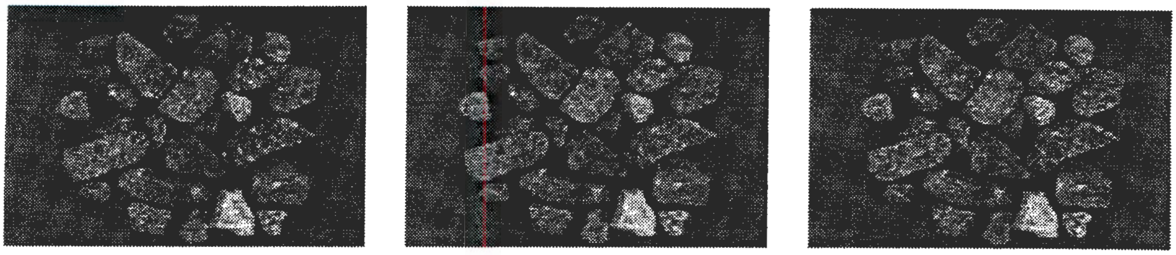
Interpolation involves determining the value of a function $f(x,y)$ for any value of x, y, given the values $f(x_i, y_j)$. Three interpolation methods for estimating $f(x,y)$ are described in [62]:

Nearest Neighbour (NN) This is a form of zero order interpolation which assigns the value of $f(x_i, y_j)$ where (x_i, y_j) is the closest to x, y.

Bilinear Interpolation (BL) A first order interpolation algorithm which linearly interpolates between the closest four data points.

Cubic Convolution (CC) Cubic convolution is a third order interpolation technique which attempts to approximate the sinc function associated with ideal low pass filtering.

The three types of interpolation have been listed by increasing order and provide correspondingly better performance. This comes at the cost of increased computational complexity,



(a) Nearest neighbour (b) Bilinear interpolation (c) Cubic convolution

Figure 3.6: A comparison of the three types of interpolation used for scaling Figure 3.1 by a factor of 2/3. Since image size is being reduced and no rotation is taking place, there is very little advantage in using the more computationally expensive techniques for a negligible performance advantage.

i.e. NN requires 1 data point, BL 4 and CC 16.

In this particular application all resampling is required for reducing the size of the image. Accurate resampling is more critical when an image is being enlarged or rotated in which case aliasing effects become more noticeable. Because the image is being reduced and the solution is not dependent on exact visual characteristics, the importance of interpolation diminishes and computational cost becomes more relevant.

This is illustrated in Figure 3.6 which shows the same figure reduced by a factor of $\frac{2}{3}$ using the three different interpolation techniques. As can be seen, there is no visible difference between the images.

3.2.5 The Hough Transform

The Hough transform [32, Section 11.6] is a technique used for the detection of well defined lines or curves within images. The image is transformed to a parameter space which can describe all the possible lines or curves in the image. Peaks in the parameter space correspond to lines or curves in the image.

Duda and Hart [21] describe the Hough transform applied to finding straight lines. A straight line can be represented using the *normal parameterisation*:

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

where θ is the angle of the line's normal and ρ is the algebraic distance from the origin to the closest point on the line. θ is restricted to the interval $[0, \pi]$ and the parameters are therefore unique. The Hough transform involves transforming all points in an image to the $\theta - \rho$ plane in such a manner that all collinear points transform to curves that have a common point of intersection. In this case, each point (x_i, y_i) is transformed to the sinusoidal curve in the $\theta - \rho$ plane defined by

$$\rho = x_i \cos\theta + y_i \sin\theta$$

Finding lines is reduced to detecting maxima in the transform space. Unfortunately, this process is exhaustive but tremendous computational savings can be made by quantising the parameter space into a grid. The grid can then be treated as an array of accumulators in which the transform curves are drawn by incrementing the bins. After all the points in the image have been transformed, if the count in a cell (θ_i, ρ_j) is k then exactly k points in the image lie along the line (to within the quantisation error) with normal parameters (θ_i, ρ_j).

For measuring rock fragmentation, ellipses with the following form must be detected:

$$\frac{((x-x_0)\cos\theta + (y-y_0)\sin\theta)^2}{a^2} + \frac{(-(x-x_0)\sin\theta + (y-y_0)\cos\theta)^2}{b^2} = 1 \tag{3.14}$$

Note that a parametric equation for the ellipse in (3.14) can be written as follows:

$$\begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} = \begin{pmatrix} a\cos\theta\cos\phi + b\sin\theta\sin\phi \\ b\cos\theta\sin\phi - a\sin\theta\cos\phi \end{pmatrix} \tag{3.15}$$

The orientation (θ) and eccentricity (a, b pairs) are both quantised to allow only N ellipse shapes which can occur anywhere within the image. The Hough transform has been implemented by considering the transform space to consist of N two dimensional spaces or levels which correspond to the possible positions for each shape of ellipse. Thus for each level, the values of a, b and θ are known as well as each point (x_i, y_i) in the image. The parametric ellipse equations can now be solved for x_0 and y_0 to give:

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} x_i \\ y_i \end{pmatrix} - \begin{pmatrix} a\cos\theta\cos\phi + b\sin\theta\sin\phi \\ b\cos\theta\sin\phi - a\sin\theta\cos\phi \end{pmatrix} \tag{3.16}$$

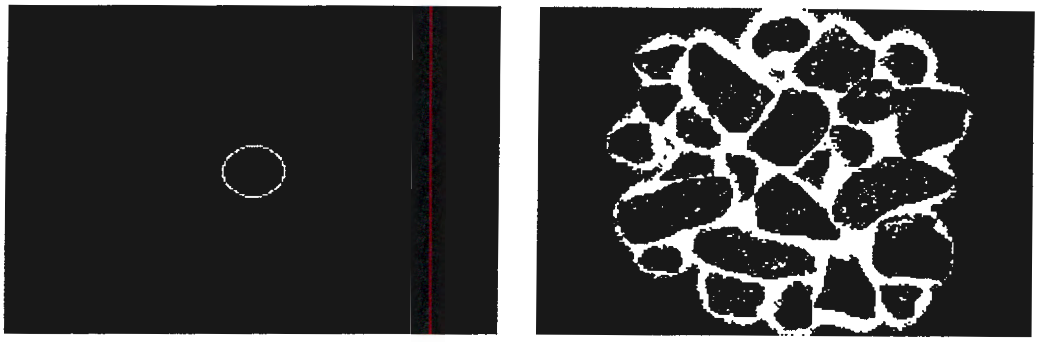
The points (x_0, y_0) are then incremented in the transform space in the same manner as for straight lines described above. Note that ellipses are then detected by finding all peaks in the transform space above a threshold.

Examples of the performance of the Hough Transform for detecting ellipses are shown in Figure 3.7.

3.2.6 Knowledge Based Systems

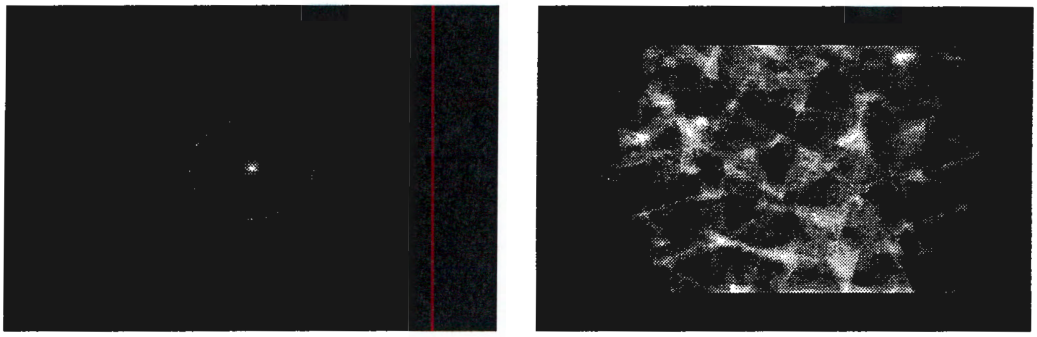
Knowledge based systems make use of knowledge about the problem to solve it. The knowledge can be expressed in two ways:

Implicit knowledge is encoded in algorithms for performing image processing operations. An example of this is an edge detection algorithm which assigns edge strength based on



(a) Ellipse

(b) Rock edges



(c) Hough transform of ellipse

(d) Hough transform of rock edges

Figure 3.7: The performance of the Hough Transform is shown here for two example images. (a) shows the exact ellipse which is to be detected. There is a clear peak in the transformed image (c) at the corresponding position. The second example (b) shows the sample rock image in which the edges have been detected and thresholded. The strongest response in the transform image occurs where the edge image contains structural features similar to the ellipse in (a).

image intensity gradient.

Explicit knowledge is stored as facts and relationships. This form of knowledge is typically used within an expert system environment. The advantage of explicit knowledge is that it can be readily examined, corrected or extended. In addition, the knowledge base can be completely changed for a different problem domain thus providing a more general system.

Haralick and Shapiro [33, Chapter 19] provide a basis for knowledge based vision systems ranging from knowledge representation, control strategies and information integration through to examples of techniques and applications.

Section 3.2: Image Processing Fundamentals

Sarkar and Boyer [80] give a review of knowledge based vision systems. They attempt to explain perceptual organisation and Gestalt behaviour (i.e. the detection of subjective contours and grouping) by operating at four levels: signal, primitive, structural and assembly levels. Image operators are examined for each of these levels for different types of image, i.e. 2D, 3D, 2D+time and 3D+time.

Knowledge based systems fall into two basic categories: Image Understanding Systems (IUS) and Expert Systems for Image Processing (ESIP).

An IUS requires diverse sources of knowledge to interpret visual scenes [55]. The knowledge can be classified into the following three levels:

1. Scene domain knowledge: geometric models of objects in scene and relations, etc.
2. Mapping from scene to image: physical properties of the imaging system (eg focal-length, angle of camera, spectral properties)
3. Image domain knowledge: types and relations among image features such as edges, lines and regions.

An ESIP is used to generate effective image analysis processes by composing primitive image processing operators. The knowledge used by an ESIP is concerned with how to use different image operators. The discrimination between an IUS and an ESIP is that the ESIP makes full use of image processing techniques to achieve a goal, while an IUS is developed to improve visual perception capability. Binford [6], Clément and Thonnat [15] and Matsuyama [55] describe expert systems for automatic analysis and solution of image processing problems.

The following knowledge is available for use in the current application:

1. Scene domain knowledge: Individual rocks are assumed to have reasonably homogeneous reflectivity. The shape of a rock is modular and can be roughly modelled by means of an ellipsoid. Rocks are of various sizes and are heaped along the centre of a conveyor belt.
2. Mapping from scene to image: Image characteristics with respect to lighting and formation can be modelled. The camera is assumed to be vertically perpendicular to the conveyor belt. The distance of the camera is only necessary for a final size scaling factor.
3. Image domain knowledge: The two dimensional projection of an ellipsoid is an ellipse. The lighting and shape of a rock defines the intensity profile in the image of a rock, i.e. interior of approximately uniform intensity and edges corresponding to steep image gradient and low intensity. Lighting also causes shadows which appear as very dark regions. Different sizes of rock will be present.

3.3 Image Formation Fundamentals

This section provides an introduction to some background theory and terminology which is required for analysis of the spatial frequency characteristics of the image acquisition system and the development of a lighting model.

3.3.1 Image Sensing Fundamentals

The modulation transfer function (MTF) of a device in an imaging system is its sine-wave spatial-frequency response [73, Section 8.3]. The spatial frequency (μ) is the number of cycles per unit length. The particular use of this measure of imaging performance is the ability to cascade the effects of several components in the imaging system to determine a measure of the resolution of the overall system. The MTF of the overall system is the product of the MTFs of the elements if the system components are assumed to be linear and shift invariant. The MTF can be written as

$$\mathrm{MTF}_S(\mu) = \mathrm{MTF}_L(\mu)\mathrm{MTF}_I(\mu)\mathrm{MTF}_D(\mu) \tag{3.17}$$

where

- $\mathrm{MTF}_S(\mu)$ is the modulation transfer function of the system,
- $\mathrm{MTF}_L(\mu)$ is the modulation transfer function of the lens,
- $\mathrm{MTF}_I(\mu)$ is the modulation transfer function due to image smear and
- $\mathrm{MTF}_D(\mu)$ is the modulation transfer function of the detector.

Note that the spatial frequency μ has units cycles.metre^{-1}.

The definitions of MTF for lenses, image smearing and detectors are given in the following three sections.

MTF for Thin Lenses

A thin lens [73] is either a converging or diverging lens which has negligible thickness compared to the object and image distances. These lenses have relatively small apertures. Only a convergent lens is considered here. The following characteristics of thin lenses are required for the calculation of lens MTF.

Focal length f is the image distance when the object is infinitely distant.

Thin-lens formula Given that the object is a distance o from the lens, the image distance, i, is given by

$$\frac{1}{\mathrm{f}} = \frac{1}{o} + \frac{1}{i} \tag{3.18}$$

Image magnification m is given by

$$\mathrm{m} = \frac{i}{o} \tag{3.19}$$

F/number or "speed" of a lens is a measure of the angular acceptance of the lens where

$$\mathrm{F/number} = \frac{\mathrm{f}}{\mathrm{D}} \tag{3.20}$$

and D is the entrance pupil diameter of the lens (assuming the lens has circular cross section).

The MTF for the lens can now be calculated. For a perfect, diffraction-limited circular lens, the MTF is a function of the wavelength, λ, of light and the F/number of the lens for incoherent radiation [73].

$$\mathrm{MTF}_L(\mu)\big|_{\mu=\sqrt{\mu_1^2+\mu_2^2}} = \frac{2}{\pi}\left[\arccos(\lambda \mathrm{F}\mu) - \lambda \mathrm{F}\mu\sqrt{1-(\lambda \mathrm{F}\mu)^2}\right] \tag{3.21}$$

The spatial cutoff frequency is $\mu = \frac{1}{\lambda \mathrm{F}}$ where μ is in cycles per metre at the image plane.

MTF for Image Smearing

Image smear in this application is caused by conveyor belt motion during camera exposure. The derivation here is limited to the case where the conveyor belt motion has only a vertical component. Assuming that the conveyor belt corresponds to the object plane and it moves a distance γ during exposure then the image moves a distance $\gamma\mathrm{m}$ and

$$\mathrm{MTF}_I(\mu_1, \mu_2) = \frac{\sin(\frac{\mu_2 \gamma \mathrm{m}}{4\pi})}{\frac{\mu_2 \gamma \mathrm{m}}{4\pi}} \tag{3.22}$$

where μ_2 is the vertical spatial frequency in cycles per metre at the image. γ is calculated as $\gamma = Tv$ where T is the period of exposure, v is the speed of the object plane and m is the image magnification.

MTF for Detectors

A *charge coupled device* (CCD) camera is proposed as the electro-optical imaging detector for this application. A CCD consists of an array of pixels in the form of potential wells which act as photon detectors. During operation, each pixel acts as an integrator of electrons which are activated by incident photons for the period of exposure. After exposure, the charge from each pixel is transferred sequentially to the output.

Due to the sampling nature of a CCD, it is not shift invariant (the MTF becomes dependent on the alignment of the CCD with respect to the image) and the calculation of MTF becomes more complicated (See Feltz [24] or Moon [60]). Essentially, the MTF can be divided into two components. The first effect is due to spatial integration and can be multiplied with the other system component MTFs. The second effect is caused by the sampling nature of the CCD and introduces a dependence on the relative alignment of the image and the CCD.

The MTF due to spatial integration is given by

$$\mathrm{MTF}_D(\mu_1, \mu_2) = \frac{\sin(\frac{\mu_1 T_\alpha}{4\pi})}{\frac{\mu_1 T_\alpha}{4\pi}} \frac{\sin(\frac{\mu_2 T_\alpha}{4\pi})}{\frac{\mu_2 T_\alpha}{4\pi}} \tag{3.23}$$

where the active part of a CCD pixel is assumed to be square with dimension T_α. μ is the spatial frequency in cycles per metre at the image plane.

The second effect caused by sampling is addressed by Feltz [24] and Moon [60]. As stated, a dependence on the relative alignment of the CCD array and the image plane is introduced which causes variable attenuation of the higher frequency components in the image. Since there can be no control over this effect, it is not investigated here. Reichenbach *et al* [74] describe a method for measuring the optical (or modulation) transfer function of a digital imaging device such as a CCD camera while explicitly dealing with the fundamentals of a sampled system.

3.3.2 Optical Modelling

The basis for understanding the interaction between light and objects is the study of thermal radiation [25, Section 16.7]. The following radiometric terminology is required for the development of an illumination model.

Flux is the rate at which light energy is emitted and is measured in watts (W).

Radiant intensity is the flux radiated into a solid angle in a particular direction and is measured in W/sr.

Radiance is the radiant intensity per unit foreshortened surface area and has units W/(sr $\cdot$ m^2). *Foreshortened surface area* or *projected surface area*, is the projection of the surface onto the plane perpendicular to the direction of radiation. It is calculated by multiplying the surface area by $\cos\theta_r$, where θ_r is the angle of radiated light relative to the surface normal.

Irradiance or *flux density* is the incident flux per unforeshortened unit surface area and is measured in W/m^2.

In order to model the illumination of a simple rock scene, it is necessary to determine the relationship between the light incident on a surface and the light reflected from that surface. The irradiance of incident light shown in Figure 3.8 is

$$E_i = I_i \cos\theta_i d\omega_i$$

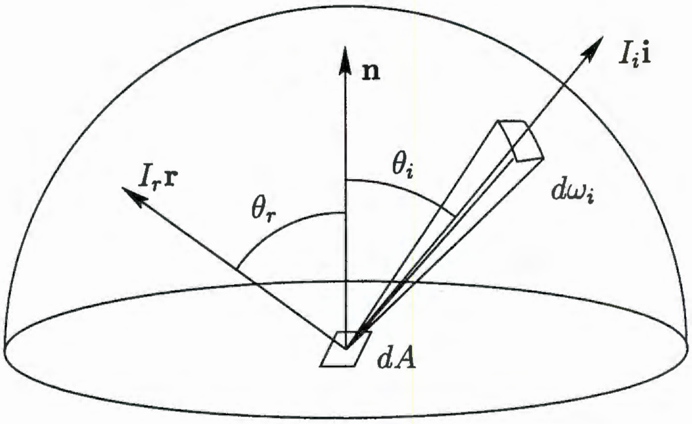


Figure 3.8: Reflected radiance and incident irradiance.

or

$$E_i = I_i \mathbf{n} \cdot \mathbf{i} d\omega_i \tag{3.24}$$

where I_i is the radiance of the incident light and $\cos\theta_i = \mathbf{n} \cdot \mathbf{i}$ converts from the foreshortened area of radiance to unforeshortened area for irradiance.

The radiance of reflected light can be calculated from the measure of irradiance of the surface using the following equation

$$\rho = \frac{I_r}{E_i} \tag{3.25}$$

where ρ is the *bidirectional reflectivity* which is a function of the directions of incidence and reflection. ρ can be approximated by separating the diffuse and specular components to give

$$\rho = k_d \rho_d + k_s \rho_s \tag{3.26}$$

where ρ_d and ρ_s are the diffuse and specular bidirectional reflectivities respectively, and k_d and k_s are the diffuse and specular reflection coefficients respectively. Note that $k_d + k_s = 1$.

The reflected radiance of a surface can then be approximated by substituting (3.26) into (3.25) and solving for I_r to give

$$I_r = (k_d \rho_d + k_s \rho_s) I_i (\mathbf{n} \cdot \mathbf{i}) d\omega_i \tag{3.27}$$

A Lambertian surface can be modelled by setting $\rho_d = c$ and $k_d = 1$ where c is a constant object bidirectional reflectivity coefficient. This surface has the characteristics of no specular reflection and uniform diffuse reflection in all directions.

Specular reflection [25, Section 16.1.4] occurs on shiny surfaces where bright highlights are caused by specular reflection. Specular reflection is characterised by the colour of the incident light and not the colour of the surface. The position of a specular highlight on a

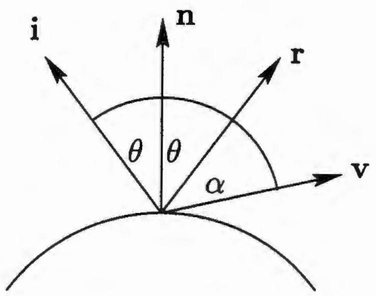


Figure 3.9: Specular Reflection

surface will appear to move as the viewpoint is shifted. This is so because the surface is not a perfect mirror and reflects light unequally in different directions. The amount of reflection is a function of the angle α in Figure 3.9. A perfect mirror will only reflect light in the direction where $\mathbf{v} = \mathbf{r}$ and $\alpha = 0$.

The Phong illumination model is a popular model in computer graphics for specular reflection. It assumes that the maximum reflectance occurs when $\alpha = 0$ and decreases as α increases. The falloff is approximated by $\cos^n \alpha$, where n is the material's *specular-reflection exponent*. The value of n varies from 1 for dull surfaces to hundreds for very shiny surfaces.

Phong's model can be incorporated into (3.27) by making $k_s > 0$ and

$$\rho_s = W(\theta) \cos^n \alpha \tag{3.28}$$

where $W(\theta)$ is the amount of incident light that is specularly reflected as a function of the angle of incidence.

In practice, $W(\theta)$ is typically set to a constant value and $\cos\alpha = \mathbf{r} \cdot \mathbf{v}$ if $\mathbf{r}$ and $\mathbf{v}$ are normalised. This can be calculated as [25, p730]

$$\mathbf{r} \cdot \mathbf{v} = (2\mathbf{n}(\mathbf{n} \cdot \mathbf{i}) - \mathbf{i}) \cdot \mathbf{v}) \tag{3.29}$$

Incorporating specular reflection into (3.27) results in the following equation

$$I_r = (k_d \rho_d + k_s(\mathbf{r} \cdot \mathbf{v})^n) I_i (\mathbf{n} \cdot \mathbf{i}) d\omega_i \tag{3.30}$$

Note that while Phong's model produces reasonable results it is not based on a physical model of the object surface. The Torrance-Sparrow model [25, Section 16.7.1] based on a microfacet distribution function provides a physically based model of a reflecting surface. Since a diffuse light source is to be used under which specular reflection is significantly less prominent due to the lack of any point sources, any improvement in accuracy from the physically based model will be reduced and the simpler Phong's model has been used.

3.4 Classification Theory

The subject of classification theory is divided into two parts in this section. In Section 3.4.1 the subject of pattern recognition is introduced to provide a mechanism for differentiating between different objects. Section 3.4.2 describes an optimisation approach called Genetic Algorithms which will be used to improve classification performance by selecting optimal features for classification.

3.4.1 Pattern Recognition

Introduction

Tou and Gonzalez [89] state that recognition is a basic attribute of human beings and other living organisms. A pattern is a description of an object. The problem of pattern recognition is that of discriminating between various input data to determine not their individual pattern but which population they belong to.

The design of pattern recognition systems involves three stages [7]. The first stage is that of sensing the data. This is achieved by means of focusing attention on a region of interest in the image. The second stage concerns the extraction of characteristic features from the input data to reduce the dimensionality of the problem. The features can be represented as a feature vector $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ which represents the pattern as a point in multidimensional space.

The third stage is that of vector classification which is a process by which a feature vector is classified or labelled by partitioning the multidimensional space into regions corresponding to the different populations; in this case, a region for rocks and the rest for non-rocks. Ideally, the regions should be such that the decisions are never wrong. Since this is not always possible, one should try to minimise the probability of error, or the average cost of errors if some are more expensive than others.

Typically the classification is determined by using information encoded in probability distributions or more commonly in a set of training samples. The subject of pattern recognition and classification is described in [7], [22] and [89] where several forms of discriminant functions are described for performing vector classification under different conditions.

The Bayesian classifier is capable of providing optimal results [22] and is described next. Unfortunately, it cannot be used in this application so the k-Nearest-Neighbour method is proposed as an alternative which is relatively easy to implement and is applicable here. That is followed by a brief analysis of classification error and how to reduce it using optimal feature selection.

Bayesian Classifier

The Bayes decision theory is described in Duda and Hart [22, Chapter 2]. The approach is based on the assumption that the problem is posed in probabilistic terms, that all relevant probability values are known and provides optimum classification. In this case, each image must be classified as rock or non-rock. Let ω denote the *state of nature*, with $\omega = \omega_1$ for rocks and $\omega = \omega_2$ for non-rocks. The state of nature is unpredictable and so ω is considered to be a random variable.

Now we assume that there is some *a priori* probability $P(\omega_1)$ that the next sub-image is a rock and some *a priori* probability $P(\omega_2)$ that it is a non-rock sub-image. These *a priori* probabilities reflect our prior knowledge of how likely we are to see rock or non-rock sub-images before they actually appear. $P(\omega_1)$ and $P(\omega_2)$ are non-negative and sum to one.

The additional information that we have is the feature vector $\mathbf{x}$. Different samples of sub-images yield different values for $\mathbf{x}$, and these can be expressed in probabilistic terms; consider $\mathbf{x}$ to be a vector valued continuous random variable whose distribution depends on the state of nature. Let $p(\mathbf{x}|\omega_j)$ be the *state-conditional probability density* function for $\mathbf{x}$, the probability density function for $\mathbf{x}$ given that the state of nature is ω_j.

If both the *a priori* probabilities $P(\omega_j)$ and the conditional densities $p(\mathbf{x}|\omega_j)$, and the value of $\mathbf{x}$ have been measured, this information can be used to determine the certainty of the true state of nature. *Bayes Rule* can be used for this purpose:

$$P(\omega_j|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_j)P(\omega_j)}{p(\mathbf{x})} \tag{3.31}$$

where

$$p(\mathbf{x}) = \sum_{j=1}^{2} p(\mathbf{x}|\omega_j)P(\omega_j)$$

Bayes rule has shown how finding the value of $\mathbf{x}$ changes the *a priori* probability $P(\omega_j)$ to the *a posteriori* probability $P(\omega_j|\mathbf{x})$. Hence, if we have an observation $\mathbf{x}$ for which $P(\omega_1|\mathbf{x})$ is greater than $P(\omega_2|\mathbf{x})$, it indicates that the true state of nature is more likely to be ω_1. Similarly, if $P(\omega_2|\mathbf{x})$ is greater than $P(\omega_1|\mathbf{x})$, ω_2 is more likely to be the true state of nature. This decision can be justified by analysing the probability of error for each decision. When a particular $\mathbf{x}$ value is observed:

$$P(error|\mathbf{x}) = \begin{cases} P(\omega_1|\mathbf{x}) & \text{if we decide } \omega_2 \\ P(\omega_2|\mathbf{x}) & \text{if we decide } \omega_1 \end{cases}$$

Using the previously described decision rule (3.31), the average probability of error will be minimised. The average probability of error is then given by

$$P(error) = \int_{-\infty}^{\infty} P(error, \mathbf{x})d\mathbf{x}$$

$$= \int_{-\infty}^{\infty} P(error|\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

which must be as small as possible if for every $\mathbf{x}$, $P(error|\mathbf{x})$ is as small as possible.

Bayes decision theory hinges on the knowledge of the conditional probability densities $p(\mathbf{x}|\omega_j)$. Usually, the probability densities are not known for every possible value of $\mathbf{x}$ which is needed for a general solution. This problem can be overcome if the density is known to have a standard form like the multivariate normal density. In this case, the probability density at any point is calculated parametrically and these parameters can be derived or determined using a suitable design set. Unfortunately, in most pattern recognition applications, and indeed in this one, the form of the underlying density function is unknown and so a *non-parametric* technique such as the k-nearest-neighbour classifier was investigated.

k-Nearest-Neighbour Classifier

Non-parametric techniques for classification are described in Duda and Hart [22, Chapter 4] and the k-nearest-neighbour technique is described in [22, Section 4.7]. This rule classifies $\mathbf{x}$ by assigning it the label most frequently represented among the k nearest samples in a design set. A decision is made by examining the k nearest neighbours and taking a vote. k is chosen as an odd number to avoid ties.

As a motive for considering the k-nearest-neighbour rule consider that if k is fixed and the number n of samples is increased to infinity, then all of the k nearest neighbours will converge to $\mathbf{x}$. The labels on each of the neighbours are random variables, independently assuming the value ω_i with probability $P(\omega_i|\mathbf{x})$, $i = 1, 2$. If $P(\omega_m|\mathbf{x})$ is the larger posteriori probability, then the Bayes decision rule always selects ω_m. The k-nearest-neighbour rule selects ω_m if a majority of the k nearest neighbours are labelled ω_m, which has a probability of

$$\sum_{i=(k+1)/2}^{(k-1)/2} \binom{k}{i} P(\omega_m|\mathbf{x})^i[1 - P(\omega_m|\mathbf{x})]^{k-i} \tag{3.32}$$

This generally gives a greater probability of ω_m being chosen for a larger value of k. If k is odd, the large-sample two-class error rate is bounded above by the function $C_k(P^*)$, where C_k is defined to be the smallest concave function of P^* greater than

$$\sum_{i=0}^{(k-1)/2} \binom{k}{i} [(P^*)^{i+1}(1-P^*)^{k-i} + (P^*)^{k-i}(1-P^*)^{i+1}]$$

As k increases, the upper bound determined above gets progressively closer to the lower bound, i.e. the Bayes rate. In the limit as k tends to infinity, the bounds meet and the k-nearest-neighbour rule becomes optimal. In use, a large value of k is desirable. However, it is also desirable that the k nearest neighbours $\mathbf{x}'$ be very close to $\mathbf{x}$ to ensure that $P(\omega_i|\mathbf{x}')$

is almost the same as $P(\omega_i|\mathbf{x})$. This forces k to be a small fraction of the number of samples and so the rule can only be optimal as n tends to infinity.

Classification Error

Since perfect classification cannot be achieved, it is necessary to perform an analysis of the classification error. Raudys and Jain [72] provide several methods for estimating classification error (Devroye [18] provides another reference on the subject of classification error). The methods can be studied using the following two factors:

1. The way in which multivariate samples are used to design or train the classifier and to test its performance.
2. The pattern error function that determines the contribution of each sample of the test set to the estimate of the probability of misclassification (PMC).

Two approaches that can be used for training and testing the classifier are:

Hold-Out Method Given the total number of samples is n^*. One portion of the set of samples (the *training set* containing N samples) is used to design the classifier, and the remaining $(n^* - N)$ portion (the *test set*) is used to estimate the error rate.

Cross-Validation Method In this case, $\begin{pmatrix} n^* \\ k \end{pmatrix}$ classifiers are designed. Each classifier is determined using k of the n^* samples as a training set, and its error rate is estimated using the remaining $(n^* - k)$ samples. This is repeated for all distinct choices of k patterns and the average error rate is computed. A common choice for k is 1, yielding the leave-out-one method.

Either of the above error estimation procedures can be used with different pattern error functions $h(\hat{g}(\mathbf{x}))$, where $\hat{g}(\mathbf{x})$ is the sample-based discriminant function:

$$\hat{P} = \frac{1}{n_t} \sum_{j=1}^{n_t} h(\hat{g}(\mathbf{x}_j))$$

and $\mathbf{x}_1, \ldots, \mathbf{x}_{n_t}$ are the test samples. Two possible pattern error functions to be used here are:

Error Counting

$$h^{EC}(\hat{g}(\mathbf{x})) = \begin{cases} 1 & \text{if } \hat{g}(\mathbf{x}) \neq \omega_1 \text{ and } \mathbf{x} \in \omega_1 \\ 1 & \text{if } \hat{g}(\mathbf{x}) \neq \omega_2 \text{ and } \mathbf{x} \in \omega_2 \\ 0 & \text{otherwise} \end{cases}$$

In this case, all incorrectly classified samples contribute equally to the estimate of PMC.

Weighted Error Counting

$$h^{WE}(\hat{g}(\mathbf{x})) = \begin{cases} f(\mathbf{x}) & \text{if } \hat{g}(\mathbf{x}) \neq \omega_1 \text{ and } \mathbf{x} \in \omega_1 \\ f(\mathbf{x}) & \text{if } \hat{g}(\mathbf{x}) \neq \omega_2 \text{ and } \mathbf{x} \in \omega_2 \\ 0 & \text{otherwise} \end{cases}$$

Here, a weighting function is used which determines how much each misclassified sample should contribute to the PMC. This allows for characterisation of variable expenses for different samples.

Feature Sensitivity Analysis

Raudys and Jain [72, Section V] state that a fundamental problem in statistical pattern recognition is the selection of features for optimal classification performance. The purpose of feature sensitivity analysis is to identify features which are important for discriminating between the classes.

Ideally, when determining the optimal classification, the *best* features are added first and the less useful ones added later. This results in a decreasing rate at which the probability of misclassification decreases. Unfortunately, the addition of new features involves measuring new parameters. Due to inexact parameter estimation, a point is reached at which the addition of further features actually decreases classification performance. This occurs when the increase in error caused by measurement error is greater than the decrease in PMC achieved by adding the new feature. The number of features which produces the lowest expected PMC is called the *optimal* number of features, p_{opt}.

If the features are ordered *a priori* as assumed above, p_{opt} can be determined using a simple procedure: Given p is the dimensionality or number of features, find *unbiased* estimates $\hat{P}_{(i)}$ of the expected PMC for $i = 1, \ldots, p$ using hold-out or leave-one-out methods. Select the subset which produces the lowest estimate PMC.

Because features are not usually ranked according to their effectiveness, sample information must be used to compare and rank them. These estimates are not exact and so the features cannot be ranked properly. The errors in ranking order cause bias in the estimates of the best subsets containing $i = 1, \ldots, (p-1)$ features. The estimates of p_{opt} also become biased. According to Raudys and Jain, this problem of determining p_{opt} for empirically ordered features is unsolved. They continue to say that the only way to guarantee which subset of features is optimum is to perform a complete inspection of all subsets, which is computationally prohibitive. Suboptimal methods do exist for this task which require less computation. They generally use the sequential addition or deletion of features or a combination of the two.

An alternate approach to solving this problem was taken by using a genetic algorithm to determine the optimal subset of features. In this case, each feature is represented by a bit in the chromosome string and the genetic algorithm is programmed to manipulate the chromosome string to find the minimum PMC.

3.4.2 Genetic Algorithms

Introduction

The Genetic Algorithm (GA) is an optimisation algorithm based on population genetics which has been shown to consistently outperform both gradient methods and random searches for difficult optimisation problems. Holland [35] provides a standard reference. An optimisation problem is difficult when the cost function is noisy, multi-dimensional and has many local optima. Many real optimisation problems exhibit these characteristics and have been tackled using this method, for example see Katz and Thrift [43], Roth and Levine [79].

GA uses a population P of n individuals denoted by $\{I_1, \ldots, I_n\}$ which is *evaluated* in an environment E which returns a fitness $e_1, \ldots, e_n$ for each individual. Each individual I_j is represented by a genome g_j which encodes its genetic information. The standard representation for a genome is a bit string which is called a chromosome. In basic operation, the GA evolves a new population P' from P by means of the processes of *selection*, *mutation* and *crossover*. Selection is an operation which determines the relative representation of individuals I_j in P' based on their fitness e_j. Mutation changes small random parts of g_j and the crossover operation takes pairs of selected individuals and produces children by means of exchanging portions of their genomes.

Two basic replacement methods can be used for generating P' from P. In the first case, called *steady state* replacement, selection is used to choose pairs for crossover which are mated to produce children which then possibly undergo mutation before replacing another pair in the population (possibly the weakest). The second case, called *generational* replacement, uses selection and crossover enough times to fill the new population completely at each generation. The generational method has been used for these experiments.

The following sections describe the order of events which are required to implement the generational strategy as well as elaborating on the actual algorithms for evaluation, selection, mutation and crossover.

Generation

A generation consists of an iterative process of evaluation of fitness $e_1, \ldots, e_n$, followed by a process of selection which ranks $\{I_1, \ldots, I_n\}$ by e and determines the rate of survival for each I_j in P'. There is then a small random chance that each individual in P' will undergo mutation followed by a chance that all pairs of individuals will undergo crossover. These operators are described in more detail in the following list:

Evaluation is a process whereby an individual I_j is assigned a fitness e_j which represents its performance in the environment E. This is the only part of the GA which is dependent on the actual problem being optimised and the mechanism for determining the fitness is not known to the algorithm. E is typically implemented by providing a function which takes a genome g_j as its parameter and returns a scalar value e_j.

Selection Initially, the population is ranked with respect to e. At this stage, the probability of selecting a particular individual is proportional to its index in the population. The best individual will be I_n and the worst, I_1. The probabilities are represented by means of segments on a wheel where the size of the segment is proportional to the relative probability of selection of the individual. Performing n spins of the wheel will result in the selection of the new population P' which is randomly ordered.

Mutation For each individual I_j in P' there is a probability $p_{mutation}$ that it will undergo mutation. In the case that it does, a randomly selected element of the genome g_j is inverted and the chance of further mutation is tested. If the test indicates mutation, the process is repeated until no mutation is indicated. This allows for mutation of more than one gene to occur with geometrically decreasing probability. Typically, $p_{mutation}$ is a small value.

Crossover All pairs of individuals in the population have a probability $p_{crossover}$ of undergoing crossover. In the event of crossover, a randomly selected segment of the two genomes will be interchanged provided that they are different both inside and outside the selected segment. If they are the same, the crossover operator would not produce a new individual. The probability of crossover is usually high.

Elitism

Elitism is a variation of the generational replacement method in which the best individual I_n is copied from P to P'. I_n is also preserved from mutation and crossover to ensure its survival. The benefit of this approach is that the performance will increase monotonically with no chance of losing a good solution before it is superseded by a better one.

3.5 Conclusion

The theoretical background presented in this chapter provides a basis for the development of a rock detection system. A complete overview of the proposed instrument structure is provided in the next chapter from which the body of the thesis is developed.

Chapter 4

Instrument Structure

4.1 Introduction

The aim of this chapter is to provide an overview of the complete instrument structure for converting a scene of rocks on a conveyor belt to a measurement of size distribution. The design consists of both a physical and a software component, both of which are introduced here.

The physical component performs the transformation from a scene with rock material on a conveyor belt to an internal representation in the form of a digitised image. Elements of the physical system are the conveyor belt supporting the rock-stream and the image acquisition system which consists of an arrangement for controlling the lighting, a camera and a frame-grabber inside the computer.

The software component is responsible for analysing the digitised image to calculate a size distribution of material in the image. Information about the physical conditions is used to convert the units of the measurements back from the internal representation to the actual dimensions of the scene.

Each component of the system will be divided into its constituent elements in the remainder of the chapter in order to provide a global view of the system. Section 4.2 gives an overview of the physical component and Section 4.3 gives an overview of the software component.

4.2 Physical Structure

An outline of the physical structure is shown in Figure 4.1. It consists of three basic modules whose responsibility it is to convert from a scene of rock material on a conveyor belt to a digitised image in a computer in which the software can be executed. The software will be described in Section 4.3. The remaining modules will be divided into their constituent elements and their possible influences will be discussed in the following two sections.

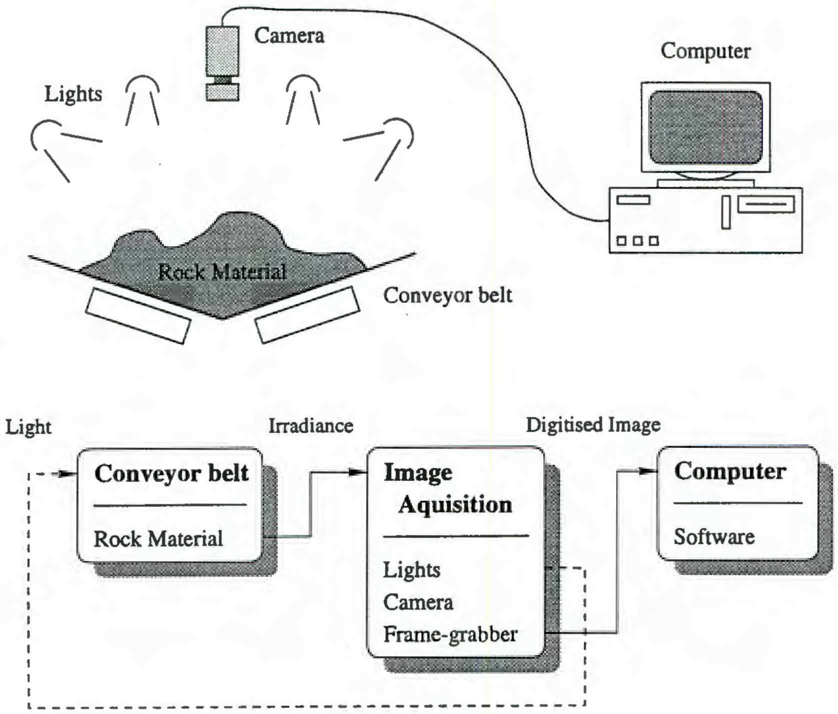


Figure 4.1: Physical overview.

4.2.1 The Scene

The scene consists of the object plane which is imaged by the camera. In this case, only the conveyor belt and rock-stream are of any concern for calculating rock size distributions. The "output" of this stage is the visual object plane which is viewed by the camera. It is affected by the lighting arrangement. Conveyor belt and rock material characteristics are discussed in the following two sections.

Conveyor belt

The conveyor belt has three parameters which could affect the complexity of measuring rock size distribution. Each of the parameters is described in the following list with a method of compensation and its implications:

Speed Conveyor belt motion will result in image blurring unless the camera operates at a sufficiently high shutter speed. Higher shutter speeds necessitate brighter lights which can result in heating problems under extreme conditions. Belt speed is used in Section 5.3.1 to determine the extent of blur caused by belt motion.

Size The size of the conveyor belt determines the minimum possible rock size that can be detected given the resolution of the image acquisition system. Regardless of the distance of the camera from the belt, if the focal length of the lens can be adjusted to result in a complete view of the width of the belt in the image, then small rocks will appear the same size. A minimum rock size is determined in Section 6.2.3 and the relationship between belt width and minimum rock size is investigated.

Wetness A wet conveyor belt can cause specular reflections under unsuitable lighting conditions which would increase the complexity of a rock detection system. This is offset by the advantage that it will be easier to discriminate between a belt cleaned with water and fines for the purpose of calculating the amount of material below the size resolution limit. The effect of the lighting model on specular reflection is examined in Section 5.2.4.

Rock-Stream

The material in the rock-stream can have a number of characteristics which could affect the performance of a machine vision system. The characteristics are listed below with an indication of their implications:

Density The density of material on the belt can vary from sparse through closely packed to a deep pile. A vibrational feed of material results in reasonably deep lumps of material on the belt with intermediate spaces of no material. Sparse material is the simplest to measure and a volume distribution can be estimated since all material can be seen. Closely packed material can result in occlusions which can result in incorrect size estimates and increased difficulty in estimating the volume. The occlusion situation is aggravated in a deep pile and volume cannot be easily estimated since most of the material is not visible.

Size Rock size can vary from fines which are too small to be resolved right through to occasional large rocks which occupy a large area of conveyor belt space. Depending on the field of view of the camera, large rocks can be difficult to see since they are very unlikely to lie completely within the area of perception of the vision system at the instant of exposure. While fines are usually less important in this measurement application, the detection of large rocks is important due to their relatively low frequency. Possible solutions to the problem of missing large rocks are to increase the sampling rate or increase the field of view of the camera. A simple solution to this problem is presented in Section 5.3.5.

Wetness Fines do not stick to dry rocks which makes them easier to identify. Wet material is sticky and results in a layer of fines which tends to obscure the underlying structure of the material. Specular reflection can also be caused by wet surfaces and this will be investigated further in Section 5.2.4.

4.2.2 Image Acquisition

The image acquisition system consists of a controlled lighting arrangement, a camera and frame-grabber card for generating a digitised image in the computer memory to represent the scene described in the previous section.

Lighting

Lighting must be controlled to provide consistent operating conditions regardless of weather and varying light conditions during the day and from plant to plant. The lighting arrangement should emphasise object features which are relevant to the solution while not introducing irrelevant artifacts. It should also provide sufficient light to allow the camera to operate at a sufficiently high shutter speed to reduce image blur to an acceptable level.

The design of a lighting model is presented in Section 5.2.

Camera

The characteristics of the camera must compensate for conveyor belt speed and space restrictions.

Shutter speed Camera shutter speed must be high enough to reduce image blur to an acceptable level. This is analysed in Section 5.3.1 where the effect of shutter speed is incorporated with conveyor belt speed to determine acceptable operating settings.

Lens It is necessary to be able to view the entire width of the conveyor belt given any space restrictions. The focal length, f, of the lens determines the field of view given the distance, o, of the object from the camera and the size of the image plane (See Section 3.3.1).

Aperture The aperture setting in conjunction with the lighting must allow sufficient light for proper exposure given the desired shutter speed.

An analysis of the spatial frequency response for determining the visual characteristics of the images based on camera characteristics and conveyor belt speed is given in Section 5.3.

Frame-grabber

A frame-grabber is used to sample a frame from a video signal to generate a digitised image in computer memory. Images created using a frame-grabber are characterised by the following features:

Image resolution This is the size of the image which is commonly 512×512 pixels.

Pixel quantisation Pixels are represented in the computer by means of a string of bits. The number of grey levels that can be assigned to a pixel is 2^{n_p} where n_p is the number of bits used to represent the pixel. It is therefore necessary to quantise the actual grey level of the pixel to one of the 2^{n_p} values. It is common to have 8 bit representation, however, some more stringent applications require 10, 12 or 16 bits.

Aspect ratio A video display has an aspect ratio which is defined as the ratio of display width to display height. Since the shape of the image may not correspond to the shape of a video display, the image and its pixels appear distorted. This distortion is also called aspect ratio. Given a standard video aspect ratio of 4:3, a square sampled image would have an aspect ratio of 3:4

Interlacing Frames in a video signal generally consist of two fields which have been exposed at an interval of $\frac{1}{50}$sec. When the image is created, the fields are interlaced. If there is motion of the scene during exposure, since the two fields are exposed at different instants in time, adjacent lines in the image will be offset from each other. It is therefore necessary to use only one of the fields thus halving the vertical resolution of the image and changing the aspect ratio to 3:2.

4.3 Software Structure

The software structure that is described in this thesis is concerned with the design of algorithms and flow of information through the system rather than an actual analysis of implementation in code. The rest of this section provides an overview of the various components of the software structure which are responsible for the detection and measurement of rocks in an image.

Figure 4.2 shows a flow diagram for the software component of the instrument. Given an input image, an image pyramid is generated and then each level of the pyramid can be processed independently. This involves the two stages of the simplified visual system described in Section 2.3.3, i.e. a preattentive stage which highlights points in the image where an elliptical object has been detected and an attention focusing stage which performs local area segmentation and classification to identify each of the highlighted objects. The results from each level are then combined in a hierarchical analysis to produce a final list of rocks from which a size distribution can be calculated.

4.3.1 The Image Pyramid

The creation of an image pyramid is influenced by the selection of a scaling factor and how smoothing is performed. The choice of these factors can influence the operation of the rock detection system (image pyramid generation was introduced in Section 3.2.4):

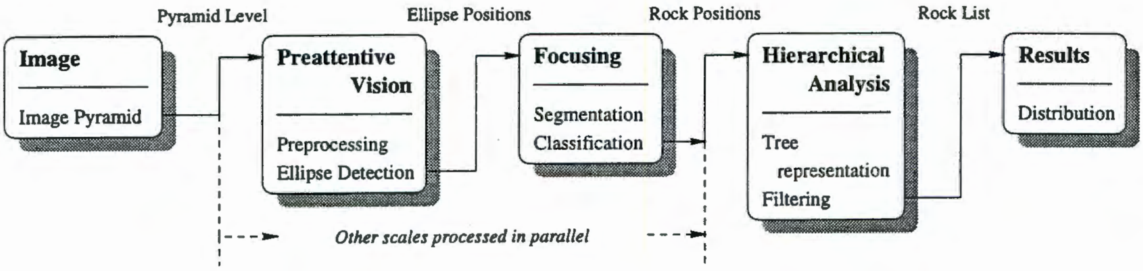


Figure 4.2: Software overview.

Scaling Factor A scaling factor, f_{scale} determines the difference in size between one pyramid level and the previous one. A value of f_{scale} close to one gives a fine step size which results in increased processing time. Coarse step sizes corresponding to a low value of f_{scale} decrease the continuity between levels and can result in rocks being missed.

Smoothing All levels of the pyramid should have similar visual characteristics to the starting image whose characteristics are determined by the image acquisition system. Smoothing should be applied before scaling to result in all levels of the pyramid containing similar spatial frequency content.

The selection of parameters and generation of the image pyramid is explained in Section 6.2.

4.3.2 Preattentive Vision

Preattentive vision involves marking positions in the image where targets could be present. Sarkar and Boyer [80, 81] review previous work in the computation of preattentive perceptual organisation which is the process of forming meaningful groups of image features based largely on the Gestaltic laws of grouping. They then propose a three level structure for perceptual organisation which initially detects lines and curves, then organises them into increasingly complicated structures such as ribbons, closed figures and finally groups of features and well defined shapes such as circles and polygons.

Since the current application is well defined, the preattentive vision stage could be designed to suit the requirements of the problem and is not nearly as generalised as the structure of Sarkar and Boyer. It is achieved in this application by means of a two level procedure. The first level involves edge detection followed by application of the Hough transform for finding ellipses. Positions where ellipses are found with sufficient strength are then considered in a second level which compares the shape of the ellipse to the shape in the image using a simple segmentation algorithm. If there is sufficient similarity between the two, that ellipse can be highlighted and will be investigated further during attention focusing. Image processing steps required for the two levels of preattentive vision are described in Sections 6.3 and 7.6 respectively.

4.3.3 Attention Focusing

In a model of the human visual system, Treisman [90] proposes that focused attention selects and integrates information about features within an attention spotlight region or region of interest. This information is then compared with an internal database of representations for the purpose of identification. The two basic steps of attention focusing are present in the rock detection system:

1. Perform segmentation of a region of interest around the specified point to find the extent of the object at that location.

2. Perform feature classification based on features calculated from several information bands using the labelled region from the segmentation for guiding feature measurement. Note that in this application knowledge about possible rock features is incorporated into the system to aid the process of segmentation and provide a model for comparing the detected objects to.

The first part of focusing uses the knowledge of rock images to postulate rules which are appropriate when a rock is present in the region of interest which corresponds in size and shape to that predicted by the preattentive stage to facilitate segmentation. It is hypothesised that each point marked by the preattentive stage is the centre of a rock which is modelled by an ellipse of known size, shape and orientation. The rules are then used to determine the extent of a blob in a region of interest around the point by locking onto image features which could correspond to the expected rock. A blob is defined as a bright region within the grey level image, the extent of which is the segmented region. The algorithm for segmenting the region of interest is given in Chapter 7.

The second part is presented in Chapter 8 and involves testing the hypothesis for each point that has been segmented previously. A feature vector is calculated for the segmented blob and classification is performed to decide if the rules were valid for that particular region of interest. This corresponds to comparing the detected object to an internal model for identification. Feature calculation and the method of classification are described in Sections 8.3 and 8.5 respectively.

4.3.4 Hierarchical Analysis

A final stage of multiscalar analysis is required to resolve any contradictions which could have developed in the internal representation of the image after attention focusing. It is responsible for determining the most probable collection of rocks in terms of a weighting function by deleting rocks which may have been detected in overlapping positions to result in a physically realisable list of rocks. The hierarchical analysis procedure is described in Chapter 9.

4.4 Summary

From the broad overview of the instrument components given here, each of these areas will now be developed in the body of the thesis. Analysis of the physical component is concerned with the design of a lighting arrangement and the characteristics of the image acquisition system. This is tackled in Chapter 5.

The software component which consists broadly of four main sections is then described: image preprocessing which incorporates pyramid generation and preattentive vision in Chapter 6; the first stage of attention focusing which performs segmentation of a highlighted region of interest in Chapter 7; the second stage of attention focusing is responsible for identifying which segmented regions correspond to the expected rock by means of a pattern recognition process in Chapter 8; and finally multiscalar analysis is performed to resolve contradictions in the representation to produce a list of rocks in Chapter 9.

Chapter 5

Image Formation

5.1 Introduction

This chapter is concerned with the mechanisms by which the scene is created and converted to a digitised image which can be analysed using a computer. The two factors which are considered in this chapter are the lighting arrangement and image sensing equipment. The effect of lighting on the visual appearance of the rock scene is described in Section 5.2 and the characteristics of the imaging system are analysed in Section 5.3.

The chapter concludes with an overview of observations in Section 5.4 of the lighting and image capture processes with emphasis on specifying suitable operating conditions for the instrument.

5.2 Lighting

5.2.1 Introduction

The role of lighting is to provide controlled conditions under which important image features such as rock edges are highlighted and lighting variations across the image are minimised.

It is important to control lighting because one cannot rely on natural light sources for several reasons: (1) Direct sunlight is effectively a point source which causes serious specular reflection problems. (2) A covered system will result in indirect light which is better than direct light because the lack of strong point sources decreases highlighting caused by specular reflection. The light is, however, likely to be directional which will result in angled shadows and the facets of different orientations having different intensities which will result in false edges in the interior of rocks. (3) Variable weather and time of day can result in changed lighting intensity and characteristics.

Figure 5.1 provides three example images of rock material illuminated by direct sunlight, indirect sunlight and controlled laboratory light. Bright reflections are dominant in the direct

(a) Direct sunlight (b) Indirect natural light (c) Controlled lighting

Figure 5.1: The effect of various forms of lighting on the appearance rock material. Direct sunlight in (a) results in overexposed rocks and reflections from the wet conveyor belt. Directional shadows and different intensities of facets of a rock are visible in (b). Controlled lighting in (c) results in the clearest image where individual rocks appear to be more homogeneous with respect to intensity and there are only shadows in the gaps between rocks.

sunlight example and while the indirect light provides reasonable results, there is significantly more clarity in the controlled image due to the relatively constant intensity across each individual rock.

All the problems associated with uncontrolled light place unnecessary additional burdens on a machine vision system which may be insurmountable and will definitely compromise performance. They can usually be overcome by controlling the lighting to obtain desirable conditions.

One approach that can be used to provide controlled lighting conditions is the application of structured light. For example, Ivanov *et al* [40] project a light strip across a conveyor belt so that when it is viewed from an angle, an instantaneous profile of the rock-stream can be obtained. Cheung *et al* [13] make use of a laser rangefinder to create a range image which is analysed instead of an intensity image. Although structured light methods offer advantages, a major argument against their use is the fact that a mechanical arrangement is usually required for their implementation. Not only does this result in a considerably more expensive system, but there is also increased maintenance due to the harsh nature of the operating environment.

Since this instrument has been designed to operate on intensity images and due to the disadvantages of structured lighting, a system based on general scene illumination was considered more appropriate and is investigated further in this section.

When the whole scene is illuminated, it is necessary to determine the presence of features by observing patterns in the image intensity. In order to achieve this, one must understand what features are significant and how their appearance is effected by the form of illumination [37].

Surface orientation is responsible for many perceptually important features in an image.

Pearson and Robinson [68] have used a study of the effect of lighting on surface orientation to determine the appearance of perceptually important features such as occluding contours in images of humans for the purpose of low bit-rate video communication.

They proposed that important points in an image of a human occur where the surface normal vector is perpendicular to the camera viewing direction and found that these features correspond to image intensity discontinuities. These points are also important in a rock scene since they correspond to the edges of rocks. Since rock edges can occur at any orientation in the image, it is necessary to ensure that the lighting is circularly symmetrical to give consistent edge intensity characteristics regardless of their direction in the plane of the image.

Gennert *et al* [28] describe a criterion for uniform surface illumination using circular, linear and point light sources which set to zero two or three terms of a two-dimensional Taylor expansion of the surface illumination. Unfortunately, their approach is inadequate for this application since it does not consider reflections or surface orientation.

For the purpose of rock detection it is undesirable to have a lighting arrangement which can create image artifacts such as specular reflections when the material is wet because the resultant highlight features can obscure the underlying shape of the rock which is what is desired. Bartlett *et al* [3] also found that specular reflection is a problem when inspecting solder joints and proposed a diffuse lighting arrangement as a possible solution.

A diffuse light source appears to provide an optimal solution since it is perfectly symmetrical and produces relatively weak highlights caused by specular reflections from shiny surfaces (such as wet rocks) due to its diffuse nature. As a result of these considerations and the observation of Bartlett *et al* that a diffuse light source should reduce the problems of specular reflection, a diffuse light source has been modelled and its effect on a simple scene has been calculated to determine its suitability.

A diffuse light source model is proposed in Section 5.2.2. A simple scene consisting of a sphere on a horizontal plane is used for all derivations of lighting effects. These are separated into dry and wet material which can be modelled by only diffuse reflection for dry material and a combination of diffuse and specular reflection for the wet material. The appearance of dry material under diffuse lighting is examined in Section 5.2.3 and wet material in Section 5.2.4.

In each case, an equation for the reflected radiance I_r is derived independently for the surface of a convex solid object which is used to model the sphere, and then for the background plane by considering the shadow which is cast by the sphere. At the ends of Sections 5.2.3 and 5.2.4, results in the form of cross-sections and images are presented. Finally, Section 5.2.5 provides a discussion of observations which can be made from the lighting model.

Derivations in the following sections are based on theory introduced in Section 3.3.2.

5.2.2 Lighting Model

The lighting model that is described here is used during the remainder of Section 5.2

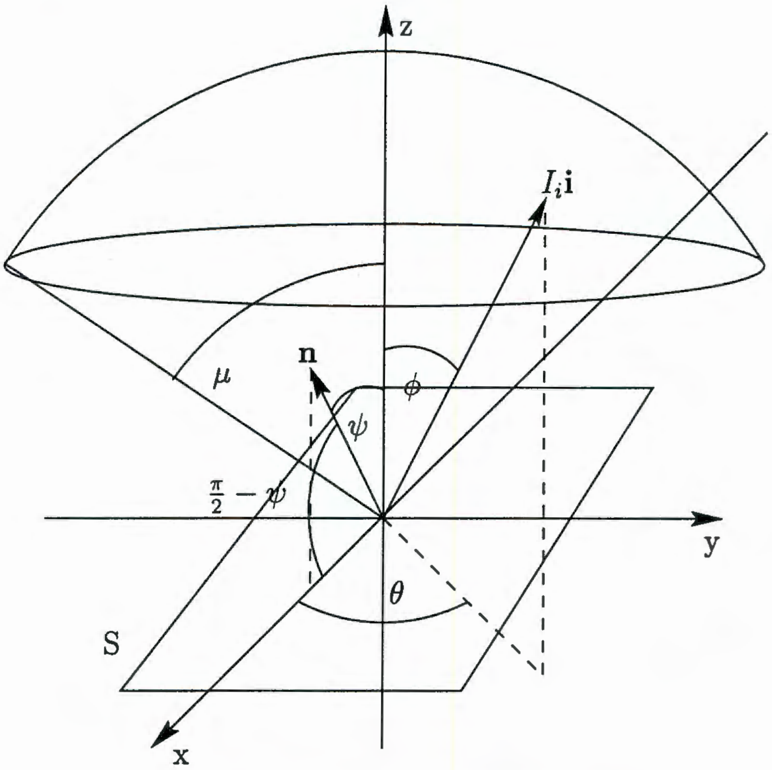


Figure 5.2: Coordinate system axes and angles. The surface is labelled with an S and its normal vector which is at an angle ψ to the vertical is shown as **n**. An arbitrary light ray is shown as I_i**i**. In spherical coordinates, its length is I_i, and its direction is specified by the angles θ and ϕ. The light dome is shown to drop to an angle μ below the vertical.

for calculating the effect of lighting on a simple scene of a sphere representing a rock on a horizontal plane (the conveyor belt). The light model is in the form of a dome positioned above the origin and descending to an angle μ from the vertical as shown in Figure 5.2. The radiance of incident light is I_i and is equal from all parts of the dome. It is to be assumed that the dome is infinitely large. This means that each point on the surface can be considered to be at the origin with respect to the incident light.

The value of I_i is determined to be the inverse of the area of a unit dome of the same shape. This means that the integral of all the light that is emitted by the dome is unity regardless of the value of μ and so comparisons can be more easily made of the relative effect of changing that value.

A Lambertian surface model, in which light is reflected equally in all directions, is used for the diffuse reflection component. Phong's illumination model, in which reflected radiance is determined by the cosine of the angle between the viewing direction and the reflection vector, is used for the specular reflection component. Note that, for simplicity, no attempt is made in the derivations to account for mutual illumination.

5.2.3 Dry Material with Diffuse Lighting

Very little specular reflection will occur on a dry rock material surface and so it can be approximately modelled using Lambertian diffuse reflection. Under this model there is no specular reflection so that the specular reflection coefficient $k_s = 0$. Since all incident light is reflected diffusely, the diffuse reflection coefficient $k_d = 1$ and because it is reflected equally in all directions, the surface diffuse bidirectional reflectivity $\rho_d = c$ where c is a constant which determines the relative brightness of the object. Equation (3.26) shows the relationship between these values and their effect on the resultant reflected radiance.

The reflected radiance of a convex surface and background plane are calculated in the next two sections by integrating the incident irradiance at each point of the surface. This is possible since ρ_d is a constant and can be moved out of the integral.

Results of the derivations are presented for diffuse reflection under three variations of the lighting model at the end of this section.

Surface Irradiance

The irradiance of the surface of a convex object under a diffuse lighting arrangement is now derived. Due to the circularly symmetrical lighting conditions it is only necessary to determine the irradiance of the surface given the angle between its normal vector at the point of interest ($\mathbf{n}$) and the vertical (ψ) which is shown in Figure 5.2. The surface normal vector is represented as:

$$\text{Surface Normal: } \mathbf{n} =< -\sin\psi, 0, \cos\psi > \tag{5.1}$$

Light rays come from all parts of the light dome and intercept the surface at the origin. An arbitrary light ray can be represented in Cartesian coordinates by:

$$\text{Incident Ray: } I_i\mathbf{i} = I_i < \sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi > \tag{5.2}$$

where I_i is the incident radiance of the light, $\mathbf{i}$ is a unit vector representing the direction of the particular light ray and θ and ϕ are standard spherical coordinate system angles shown in Figure 5.2.

Light which is incident on a Lambertian surface is reflected equally in all directions. It is therefore possible to integrate the irradiance caused by all incident rays before determining the reflected radiance. The irradiance caused by a single incident ray is found from (3.24) to be:

$$\text{Incident Irradiance: } E_i = I_i\mathbf{n}\cdot\mathbf{i}\,d\omega_i = I_i(\cos\psi\cos\phi - \sin\psi\sin\phi\cos\theta)\,d\omega_i \tag{5.3}$$

In the first simple case, the slope of the surface is sufficiently low that it does not obstruct any of the incident light rays. This condition occurs when $\frac{\pi}{2} - \psi \geq \mu$. In this case the total irradiance is calculated by integrating (5.3) over the whole dome. Note that the Jacobian is used to convert the differentials $d\omega_i = \sin\phi d\phi d\theta$.

$$\begin{aligned}
\text{Surface Irradiance} &= \int_0^{2\pi}\int_0^{\mu} I_i(\cos\psi\sin\phi\cos\phi - \sin\psi\sin^2\phi\cos\theta)\,d\phi\,d\theta \\
\text{Substitute: } a &= \sin\phi \\
\text{Then: } da &= \cos\phi\, d\phi \\
&\Rightarrow I_i\int_0^{2\pi}\left(\cos\psi\int_0^{\sin\mu} a\,da - \sin\psi\cos\theta\int_0^{\mu}\frac{1}{2}(1-\cos 2\phi)\,d\phi\right)d\theta \\
&= I_i\int_0^{2\pi}\cos\psi\left[\frac{a^2}{2}\right]_0^{\sin\mu} - \frac{1}{2}\sin\psi\cos\theta\left([\phi]_0^{\mu} - \left[\frac{1}{2}\sin 2\phi\right]_0^{\mu}\right)d\theta \\
&= I_i\int_0^{2\pi}\frac{1}{2}\cos\psi\sin^2\mu - \frac{1}{2}\sin\psi\cos\theta(\mu - \frac{1}{2}\sin 2\mu)\,d\theta \\
&= I_i\left(\frac{1}{2}\cos\psi\sin^2\mu[\theta]_0^{2\pi} - \frac{1}{2}\sin\psi(\mu - \frac{1}{2}\sin 2\mu)[\sin\theta]_0^{2\pi}\right) \\
E_{i,total} &= \pi I_i\cos\psi\sin^2\mu \qquad (5.4)
\end{aligned}$$

From (3.25), the reflected radiance from the unobstructed surface can be calculated using:

$$I_r(\mu) = k_d\rho_d E_{i,total} = k_d\rho_d\pi I_i\cos\psi\sin^2\mu \qquad (5.5)$$

In the second case when $\frac{\pi}{2} - \psi < \mu$, a wedge of light from the dome is occluded by the surface. The amount of reduction of irradiance can be calculated by integrating (5.3) over the occluding wedge. The total irradiance is then calculated by subtracting the occluded irradiance component from (5.4). In order to perform the calculation, it is necessary to determine the limits of integration for both θ and ϕ for the wedge.

Consider a plane S which is tangential to the surface at the origin and a vertical projection plane P which is oriented in the direction θ as shown in Figure 5.3. The plane equations for S and P are:

$$\begin{aligned}
\text{Surface Plane (S):} \quad \sin\psi x + \cos\psi z &= 0 \qquad (5.6) \\
\text{Projection Plane (P):} \quad \sin\theta x + cos\theta y &= 0 \qquad (5.7)
\end{aligned}$$

Figure 5.4 shows the line of intersection between S and P which is found by equating (5.6) and (5.7). The limit for ϕ given θ can be found by determining the slope of the line of intersection. This is calculated in the following way:

$$\begin{aligned}
\text{Line of Intersection:} \quad &\left(\begin{array}{ccc|c} \sin\psi & 0 & \cos\psi & 0 \\ \sin\theta & \cos\theta & 0 & 0 \end{array}\right) \\
\sim\ &\left(\begin{array}{ccc|c} 1 & 0 & \cot\psi & 0 \\ 0 & 1 & -\cot\psi\tan\theta & 0 \end{array}\right)
\end{aligned}$$

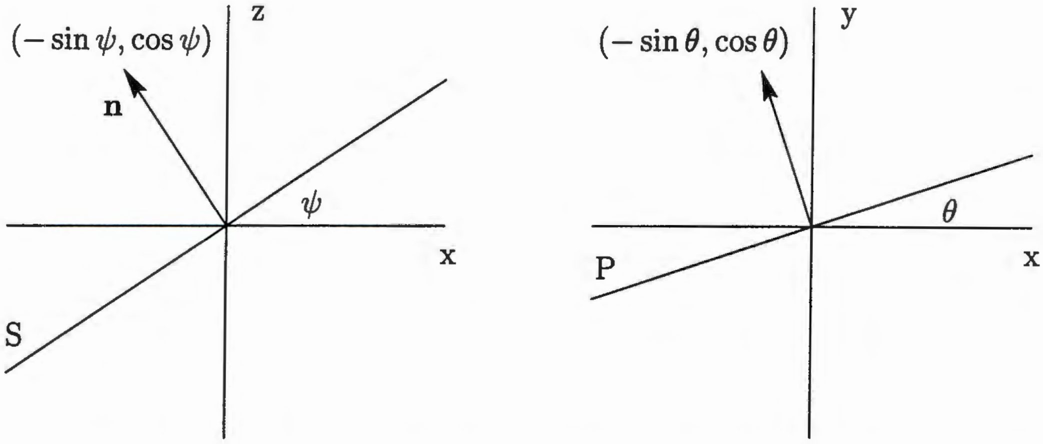


Figure 5.3: Planes used for calculating the limits of integration for ϕ in obstructed light. The left figure shows a side view of a tangential surface plane with normal vector **n**. The right figure shows a vertical plane through the rotating angle θ.

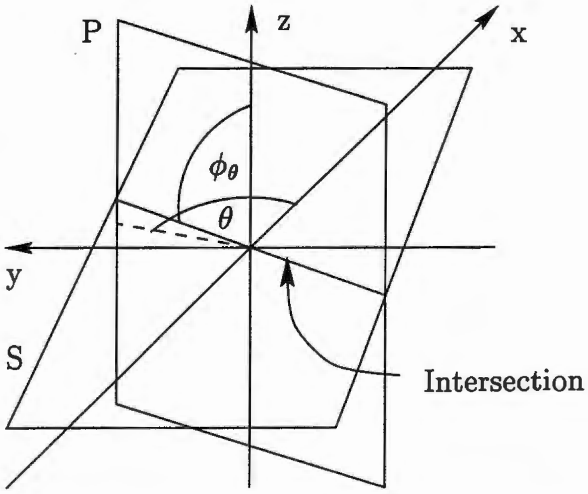


Figure 5.4: The line of intersection between the surface plane S and P. As the angle θ is increased, the slope of the line of intersection will trace the upper edge of the obstructed wedge of light.

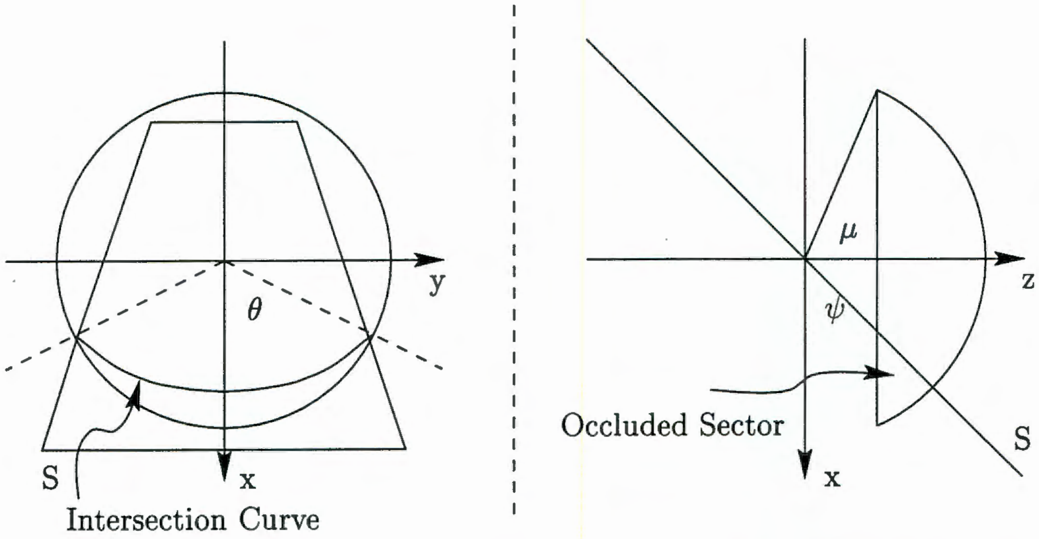


Figure 5.5: The location of the blocked wedge of light is used to determine the limits of integration for θ.

$$\begin{aligned}
\text{Letting: } z &= t \\
\text{Then: } x &= -t\cot\psi \\
y &= t\cot\psi\tan\theta \\
\text{Limit for } \phi &= \frac{\sqrt{x^2+y^2}}{z} \\
&= \arctan\left(\sqrt{\cot^2\psi+\cot^2\psi\tan^2\theta}\right) \\
&= \arctan\left(\cot\psi\sqrt{1+tan^2\theta}\right) \\
\phi_\theta &= \arctan\left(\cot\psi\sec\theta\right) \qquad (5.8)
\end{aligned}$$

The limit of integration for θ is determined by μ and is denoted by θ_μ. It is found by equating the slope (5.8) of the line of intersection between S and P with μ and solving for θ.

$$\begin{aligned}
\text{Limit for } \theta\text{: } \mu &= \arctan\left(\cot\psi\sec\theta_\mu\right) \\
\text{Therefore: } \tan\mu &= \cot\psi\sec\theta_\mu \\
\cos\theta_\mu &= \cot\psi\cot\mu \\
\theta_\mu &= \arccos\left(\cot\psi\cot\mu\right) \qquad (5.9)
\end{aligned}$$

Figure 5.5 shows the occluded wedge of light. The component of irradiance which is blocked by the occluding wedge is calculated by integrating (5.3) over the range θ from $-\theta_\mu$

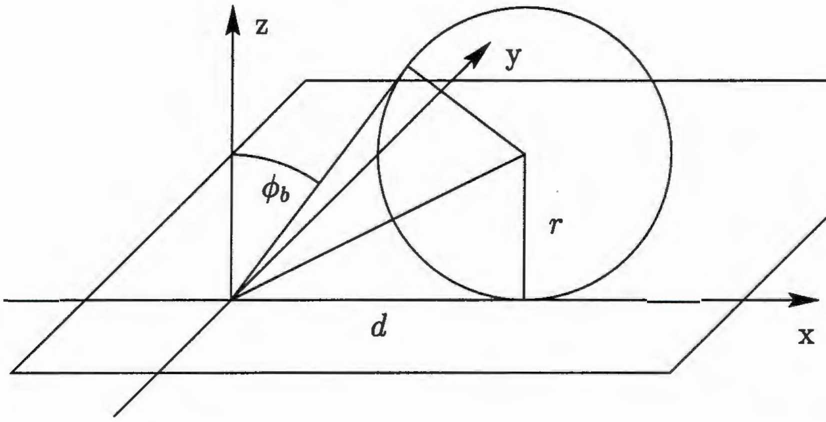


Figure 5.6: The inclination to the highest visible point of the ball is used as the limit for ϕ.

to θ_μ and ϕ from ϕ_θ to μ. This results in the following integral:

$$\text{Occluded Surface Irradiance: } E_{i,occluded} = \int_{-\theta_\mu}^{\theta_\mu} \int_{\phi_\theta}^{\mu} I_i(\cos\psi \sin\phi \cos\phi - \sin\psi \sin^2\phi \cos\theta)\, d\phi\, d\theta$$

$$= \int_{-\arccos(\cot\psi \cot\mu)}^{\arccos(\cot\psi \cot\mu)} \int_{\arctan(\cot\psi \sec\theta)}^{\mu} I_i(\cos\psi \sin\phi \cos\phi - \sin\psi \sin^2\phi \cos\theta)\, d\phi\, d\theta \qquad (5.10)$$

Since no closed form could be found for (5.10), it has to be computed numerically.

Now, using (3.25), (5.4) and (5.10), the reflected radiance of the surface can be calculated for the case when $\frac{\pi}{2} - \psi < \mu$ using:

$$\begin{aligned} I_r(\mu) &= k_d \rho_d (E_{i,total} - E_{i,occluded}) \\ &= k_d \rho_d \left(\pi I_i \cos\psi \sin^2\mu - \int_{-\theta_\mu}^{\theta_\mu} \int_{\phi_\theta}^{\mu} I_i(\cos\psi \sin\phi \cos\phi - \sin\psi \sin^2\phi \cos\theta)\, d\phi\, d\theta \right) \\ &= \pi c I_i \cos\psi \sin^2\mu - \int_{-\theta_\mu}^{\theta_\mu} \int_{\phi_\theta}^{\mu} c I_i(\cos\psi \sin\phi \cos\phi - \sin\psi \sin^2\phi \cos\theta)\, d\phi\, d\theta \qquad (5.11) \end{aligned}$$

Background Irradiance

The irradiance of the background plane is now derived for the case where a spherical object is positioned on the background. The sphere has a radius r and is positioned a distance d from the point of interest. Due to the symmetrical nature of the problem, the equations are only formulated for the case where the sphere is positioned along the x-axis as shown in Figure 5.6 without loss of generality. Because the background is flat and horizontal, its normal vector at the point of interest can be represented as:

$$\text{Surface Normal: } \mathbf{n} = < 0, 0, 1 >$$

It is assumed that as in the previous case, the point of interest is at the origin and the light source is infinitely far away. The representation of an arbitrary light ray in Cartesian coordinates is:

$$\text{Incident Ray: } I_i\mathbf{i} = I_i < \sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi >$$

The irradiance of the background caused by each incident ray is calculated using (3.24) which gives:

$$\text{Incident Irradiance: } E_i = I_i\mathbf{n}\cdot\mathbf{i}\,d\omega_i = I_i\cos\phi\,d\omega_i \tag{5.12}$$

From the point of interest, the smallest angle ϕ at which the sphere can be observed is shown in Figure 5.6 and can be calculated using:

$$\phi_b = \frac{\pi}{2} - 2\arctan\left(\frac{r}{d}\right) \tag{5.13}$$

The effect of the sphere is to cast a shadow on the background. The depth of the shadow at any point on the background is calculated by determining the amount of incident light that is blocked by the sphere for that point. The simplest case is that in which the ball is sufficiently far from the point of interest so that the sphere has no effect on incident light. This occurs when $\phi_b > \mu$ where μ is the angle of the light dome in Figure 5.2. The total irradiance is then calculated by integrating (5.12) over the whole dome to give:

$$\begin{aligned}
\text{Background Irradiance} &= \int_0^{2\pi}\int_0^{\mu} I_i\cos\phi\sin\phi\,d\phi\,d\theta \\
\text{Substitute: } a &= \sin\phi \\
\text{Then: } da &= \cos\phi\,d\phi \\
&\Rightarrow I_i 2\pi\int_0^{\sin\mu} a\,da\,d\theta \\
&= 2\pi I_i\left[\frac{a^2}{2}\right]_0^{\sin\mu} \\
E_{i,total} &= \pi I_i\sin^2\mu
\end{aligned} \tag{5.14}$$

From (3.25), the reflected radiance from the background with no shadow from the sphere can be calculated using:

$$I_r(\mu) = k_d\rho_d E_{i,total} = k_d\rho_d\pi I_i\sin^2\mu \tag{5.15}$$

For the case where $\phi_b \leq \mu$ the situation depicted in Figure 5.7 occurs. In this case, the sphere obscures some incident light from above the angle μ. The quantity of irradiance that is occluded by the sphere is found by integrating the incident irradiance over the projection of the sphere from the point of interest onto the light dome. This is achieved by integrating horizontal slices in the range $\phi : \phi_b \rightarrow \mu$. The width of the slice is used as the limit of θ and

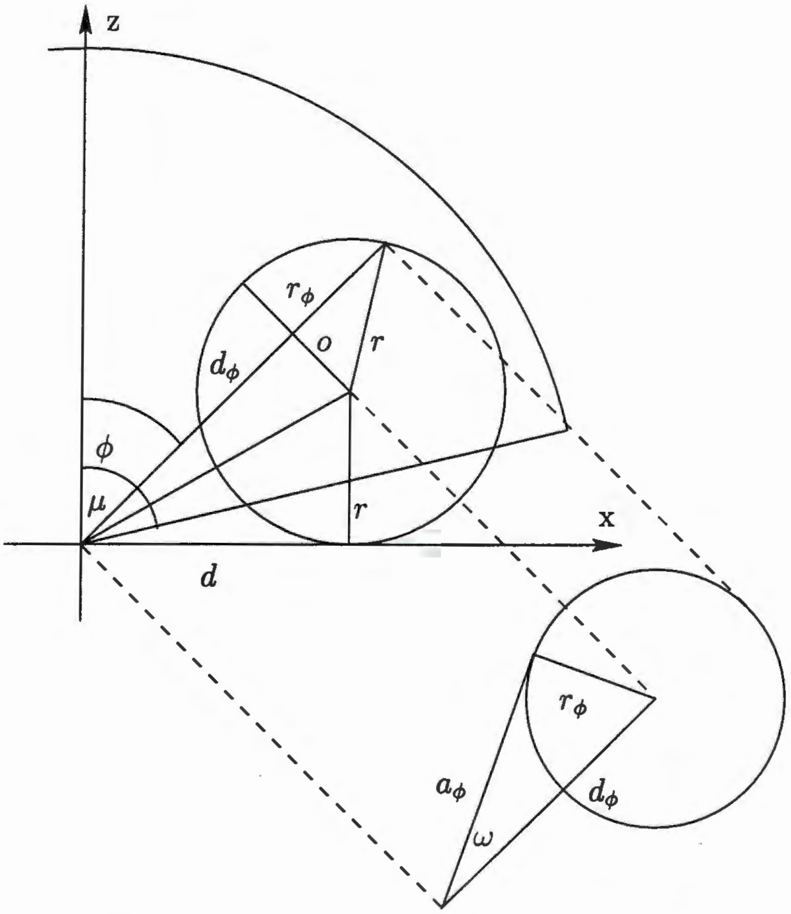


Figure 5.7: Determining the limits of integration for θ given ϕ.

is found using the following method:

If the radius of the circular slice is found to be r_ϕ then the following values can be obtained from the geometry of the problem:

$$\begin{aligned}
o &= \sqrt{r^2+d^2}\sin\left(\frac{\pi}{2}-\phi-\arctan\left(\frac{r}{d}\right)\right)\\
&= \sqrt{r^2+d^2}\cos\left(\phi+\arctan\left(\frac{r}{d}\right)\right)\\
d_\phi &= \sqrt{r^2+d^2+(r^2+d^2)\cos^2\left(\phi+\arctan\left(\frac{r}{d}\right)\right)}\\
r_\phi &= \sqrt{r^2-(r^2+d^2)\cos^2\left(\phi+\arctan\left(\frac{r}{d}\right)\right)}\\
a_\phi &= \sqrt{r^2+d^2+(r^2+d^2)\cos^2\left(\phi+\arctan\left(\frac{r}{d}\right)\right)-r^2+(r^2+d^2)\cos^2\left(\phi+\arctan\left(\frac{r}{d}\right)\right)}\\
&= \sqrt{d^2+2(r^2+d^2)\cos^2\left(\phi+\arctan\left(\frac{r}{d}\right)\right)}
\end{aligned}$$

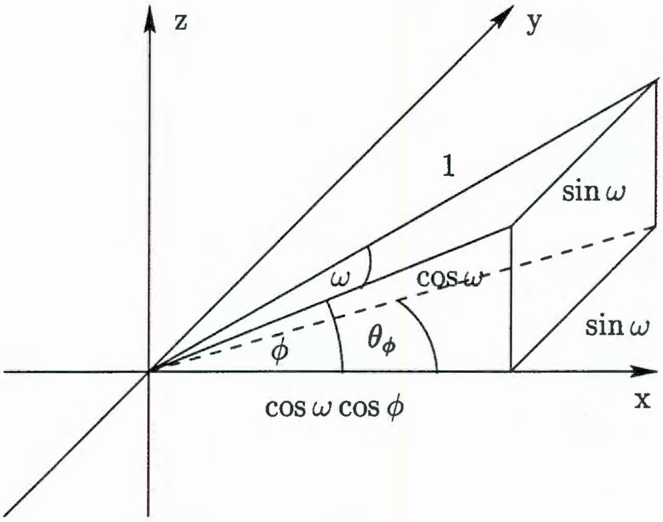


Figure 5.8: Projection of an angle at inclination ψ onto the x-y plane.

The circular slice is at an angle ϕ to the vertical and the angular limit of view of the slice is ω. Because the slice is inclined at an angle ϕ, its projection onto the x-y plane must be found in order to determine the limit of θ. The solution to this problem is depicted in Figure 5.8. Given an angle ω at inclination ϕ, θ_ϕ is found as:

$$\begin{aligned}\text{Limit for } \theta\text{: } \theta_\phi &= \arctan\left(\frac{\sin\omega}{\cos\omega\cos\psi}\right)\\ &= \arctan(\csc\phi\tan\omega)\\ &= \arctan\left(\csc\phi\frac{\sqrt{r^2-(r^2+d^2)\cos^2\left(\phi+\arctan\left(\frac{r}{d}\right)\right)}}{\sqrt{d^2+2(r^2+d^2)\cos^2\left(\phi+\arctan\left(\frac{r}{d}\right)\right)}}\right) \qquad (5.16)\end{aligned}$$

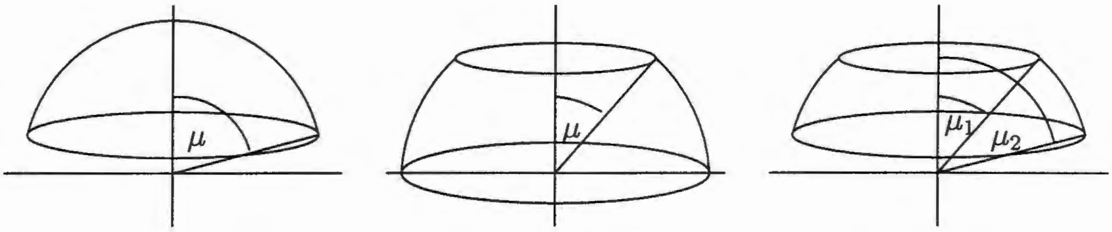
The component of the irradiance that is occluded by a sphere when $\phi_b \leq \mu$ is determined using the following integral which will be evaluated numerically:

$$\text{Occluded Background Irradiance: } E_{i,occluded} = 2\int_{\phi_b}^{\mu}\int_{0}^{\theta_\phi} I_i d\theta\cos\phi\sin\phi\, d\phi$$

$$= 2I_i\int_{\frac{\pi}{2}-2\arctan\left(\frac{r}{d}\right)}^{\mu}\arctan\left(\csc\phi\frac{\sqrt{r^2-(r^2+d^2)\cos^2\left(\phi+\arctan\left(\frac{r}{d}\right)\right)}}{\sqrt{d^2+2(r^2+d^2)\cos^2\left(\phi+\arctan\left(\frac{r}{d}\right)\right)}}\right)\cos\phi\sin\phi\, d\phi \qquad (5.17)$$

The reflected radiance of the background surface can now be calculated using (3.25), (5.14) and (5.17) for the case when $\phi_b \leq \mu$.

$$I_r(\mu) = k_d\rho_d(E_{i,total} - E_{i,occluded})$$



(a) Top (b) Side (c) Band

Figure 5.9: The three possible lighting models achievable using the surface and background irradiance equations.

$$= \pi c I_i \sin^2 \mu - 2cI_i \int_{\phi_b}^{\mu} \int_0^{\theta_\phi} d\theta \cos\phi \sin\phi \, d\phi \tag{5.18}$$

Dry Material Results

This section presents several results in the form of cross-sections and images of the simple scene under various diffuse lighting conditions. Some examples of real objects under an approximation of diffuse light are also given.

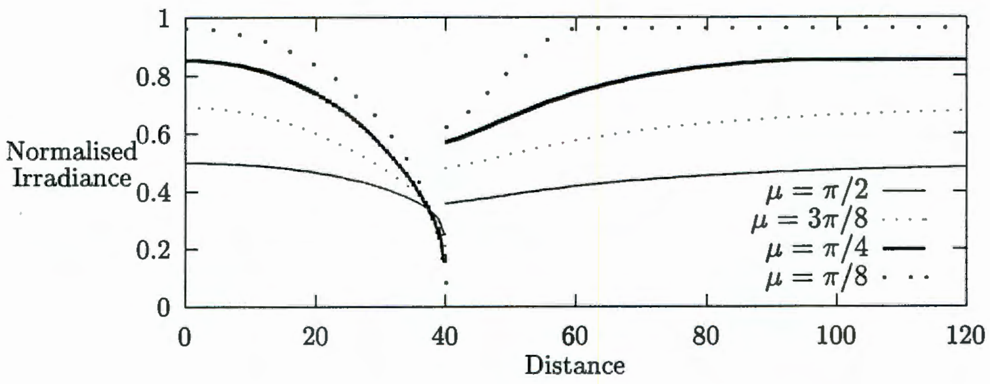
Three types of diffuse lighting arrangement can be modelled using (5.11) and (5.18) and they are portrayed in Figure 5.9. The first case is a direct application of the formulae to model a diffuse light source in the form of a dome extending downwards from the top to the given angle μ. In this case, the reflected radiance at a point on the surface or background is calculated using $I_r(\mu)$ from (5.11) or (5.18) where appropriate. Figure 5.10(a) shows example cross-sections of reflected radiance for four different values of μ using the top diffuse lighting model[1].

In the second lighting model shown in Figure 5.9(b), only light from the sides illuminates the scene. The reflected radiance is calculated by finding the difference between a complete hemisphere of light and a dome extending down to μ. This is achieved by calculating reflected radiance using $I_r(\frac{\pi}{2}) - I_r(\mu)$. Figure 5.10(b) shows cross-sections for three different values of μ using the side lighting model. Note that for the case of $\mu = \pi/4$ the radiance from the surface of the ball is uniform until very near its edge thus resulting in a very localised edge in the graph. This would appear to be a very good model of lighting to implement with only diffuse reflection since image intensity would only change near the edge of objects, thus reducing the chance of detecting internal features due to shape. However, image contrast would not be very high.

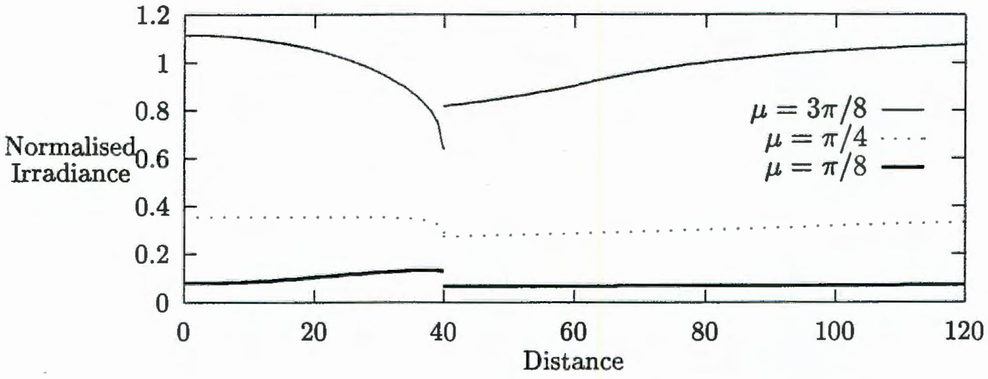
Note that a problem that is associated with side lighting is the casting of shadows by rocks on the edge of the rock-stream thus changing the appearance of adjacent rocks which

---

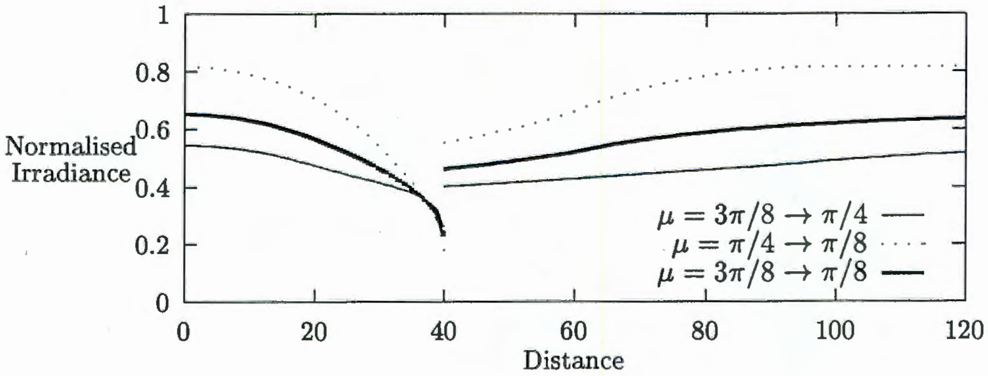
[1] In Figure 5.10 the ball and surface are both Lambertian with the same bidirectional reflectivity



(a) Top Lighting

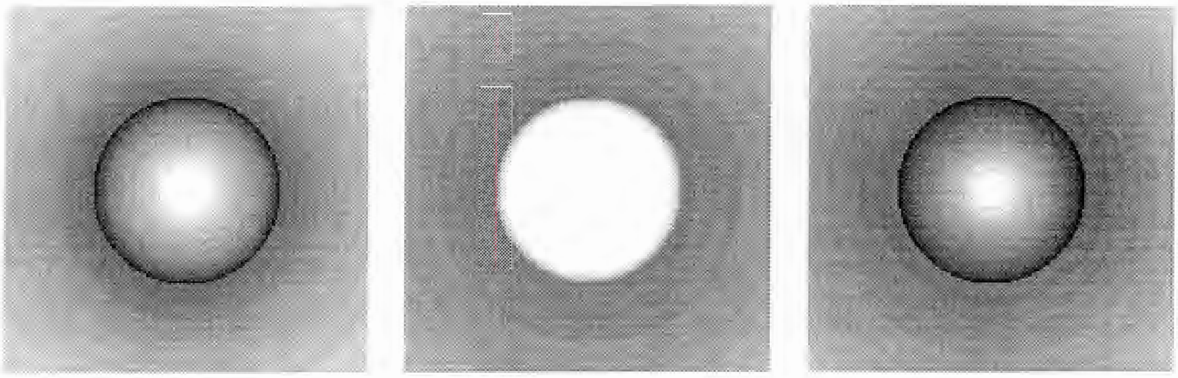


(b) Side Lighting



(c) Band Lighting

Figure 5.10: Cross-sections of reflected radiance given a ball of radius 40 on a flat plane under diffuse lighting. Only the diffuse reflection component is shown.



(a) Top $\mu = \pi/2$ (b) Side $\mu = \pi/4$ (c) Band $\mu = 3\pi/8 \rightarrow \pi/4$

Figure 5.11: Three images of a white ball on a white background generated by revolving the indicated diffuse component of the reflected radiance cross-section about the origin.

is ignored in the simple model shown here. It would therefore appear that light from a very low angle is not desirable as it can only affect rocks at the edge of the pile.

For the third model shown in Figure 5.9(c), light comes from a band between two values of μ. The reflected radiance is calculated by finding the difference between a dome extending down to μ_2 and a dome extending down to μ_1. This is achieved by calculating radiance using $I_r(\mu_2) - I_r(\mu_1)$. Figure 5.10(c) shows cross-sections of radiance for three different pairs of μ using the floating band lighting model.

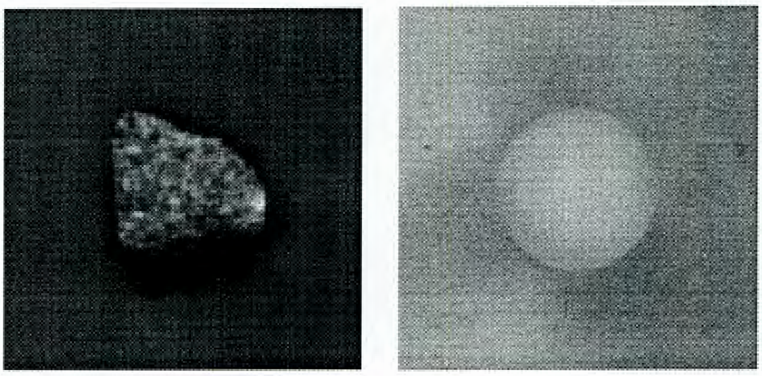
Figure 5.11 shows three example images of the ball on a plane under different lighting conditions. In each case, the pixel intensity has been normalised to make the brightest point white in order to facilitate comparison of the images[2]. It can be seen that in Figure 5.11(b), the edge of the ball corresponds very closely to the position of an edge in the image. Although the gradient is not very strong due to low contrast, it should be well suited to the directional gradient detector described in Section 3.2.1.

For Figures 5.11(a) and (c), the intensity decreases over a wide range before the actual edge of the ball so that although the correct edge should be found using the directional gradient detector by locating the maximum gradient, the presence of intensity gradient over the whole ball could increase the chance of finding the incorrect edge point under conditions of noise or non-ideal shape. For this reason, the valley detection operator described in Section 3.2.1 would appear to be more appropriate.

Figure 5.12 shows two example images of real objects and Figure 5.13 shows a cross-section through the centre of each object. It can be seen that especially with a light background, achieving uniform diffuse light is difficult with the result that there are faint shadows and

---

[2] Note that this would require different intensities of light source in practice



(a) Rock on a dark background.

(b) A table-tennis ball on white paper.

Figure 5.12: Two example images of real objects under an approximation of a diffuse light source. Four lights are positioned at an inclination of approximately 45 degrees. A white cover is placed over the arrangement to increase illumination from all angles and a white diffuse reflector is placed between each source and the object to disperse the light and prevent direct rays causing shadows. Note that especially in (b) there is an intensity gradient from left to right.

an intensity gradient across the scene. The lack of contrast in the ball image is probably partly caused by the point spread function of the camera which has decreased the depth of the intensity valley as a result of smoothing. Despite these problems, both objects exhibit characteristics predicted by the model. Valleys are clearly visible at the edge of the rock and the white ball is seen to have an intensity which varies with surface orientation.

The effect of specular reflection from a wet surface is modelled in the next section.

5.2.4 Wet Material and Specular Reflection

The effect of specular reflection under diffuse lighting is investigated in this section. Specular reflection can occur on a wet surface where incident light is reflected, thereby creating highlights. Phong's illumination model (See Section 3.3.2) provides a popular approximation of specular reflection which will be used here to model the properties of a wet surface under the diffuse light model. As with diffuse reflection, the remainder of this section is divided into modelling the reflection first from a convex surface and then from a horizontal background plane. Finally, a section of results presents the specular reflection component of reflected radiance under the same lighting conditions as Section 5.2.3. The results of incorporating both diffuse and specular components are also shown.

Although the formula used here does not provide a very accurate model of specular reflection since it does not actually model the physical characteristics of the surface, it does

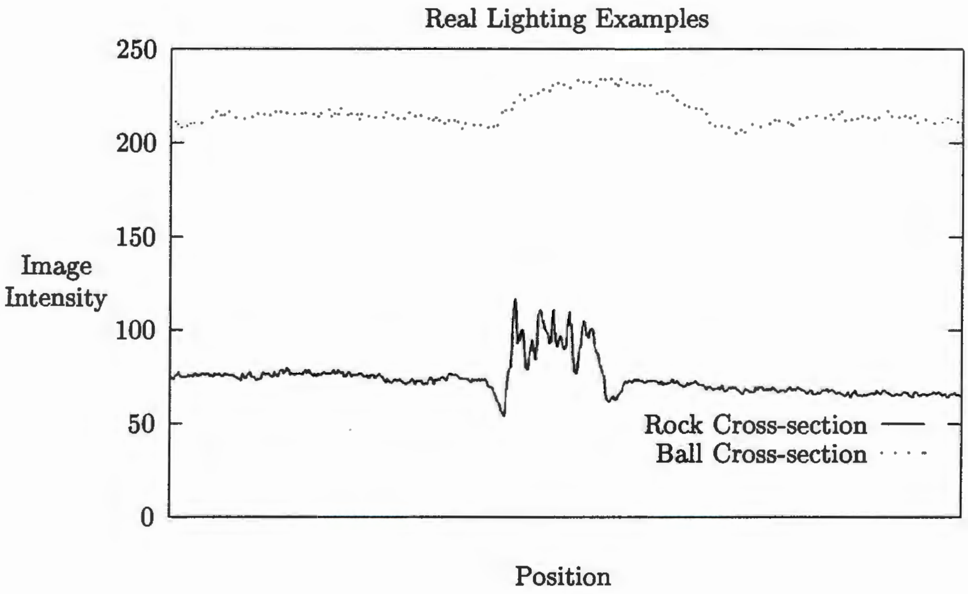


Figure 5.13: A horizontal cross-section through the rock and a vertical cross-section through the ball. Notice the valleys at either edge of the rock and the change in shading of the ball with surface orientation.

show that under appropriate diffuse lighting conditions, highlighting is significantly reduced and therefore less of an obstruction to machine vision than under discrete source conditions. This is because the highlights caused by all the incident rays occur in different positions and therefore tend to even out across the surface.

To incorporate the effects of specular reflection it is necessary to make $k_s > 0$ and define ρ_s. In this case Phong's model will be used to give $\rho_s = \cos^n \alpha$.

Since ρ_s is not constant, it is necessary to integrate an expression of reflected radiance rather than surface irradiance. However, since the same physical arrangement of a sphere on a horizontal plane is used, the same limits for integration can be used and it is only necessary to determine the expression for reflected radiance for each situation.

Surface Reflection

The reflected radiance due to specular reflection from a convex surface under the diffuse lighting arrangement is now derived. Specular reflection is a function of the viewing direction, surface normal and direction of incident light. It is assumed here that the camera is positioned perpendicularly above the conveyor belt so that the viewing direction is vertical. In that case:

$$\text{View Direction: } \mathbf{v} = < 0, 0, 1 >$$

The same surface normal (5.1) and incident ray (5.2) are used as in the diffuse case to give

$$\begin{aligned}\text{Surface Normal: } \mathbf{n} &= <-\sin\psi, 0, \cos\psi> \\ \text{Incident Ray: } I_i\mathbf{i} &= I_i <\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi>\end{aligned}$$

where I_i is the radiance of the incident light.

The specular reflection component of reflected radiance is extracted from (3.27) to give the following equation for reflected radiance from a specular surface:

$$I_r = k_s\rho_s I_i(\mathbf{n}\cdot\mathbf{i})d\omega_i \tag{5.19}$$

The component of the incident ray along the normal, $\mathbf{n}\cdot\mathbf{i}$ was determined in (5.3) to be

$$\mathbf{n}\cdot\mathbf{i} = \cos\psi\cos\phi - \sin\psi\sin\phi\cos\theta$$

and ρ_s is determined using Phong's model which is introduced in Section 3.3.2. With this model, ρ_s is estimated to be a power n of the cosine of the angle α between the direction of reflection $\mathbf{r}$ and the viewing direction $\mathbf{v}$ which can be expressed using (3.29) as

$$\begin{aligned}\rho_s &= (\mathbf{r}\cdot\mathbf{v})^n \\ &= ((2\mathbf{n}(\mathbf{n}\cdot\mathbf{i}) - \mathbf{i})\cdot\mathbf{v})^n \\ &= (2\cos\psi(\cos\psi\cos\phi - \sin\psi\sin\phi\cos\theta) - \cos\phi)^n\end{aligned}$$

Since ρ_s is a function of θ and ϕ it is necessary to integrate reflected radiance rather than surface irradiance which results in the following integral when there is no obstruction of incident light, i.e. $\frac{\pi}{2} - \psi \geq \mu$.

$$\text{Surface Reflection: } I_{r,total}(\mu) = \int_0^{2\pi}\int_0^{\mu} k_s\rho_s I_i(\mathbf{n}\cdot\mathbf{i})\sin\phi\, d\phi\, d\theta \tag{5.20}$$

For the second case when $\frac{\pi}{2} - \psi < \mu$, the component of reflection due to the occluded wedge is found using ϕ_θ from (5.8) and θ_μ from (5.9) as integration limits and the result is subtracted from (5.20) to give the amount of Occluded Surface Reflection:

$$I_r(\mu) = k_s I_i\left(\int_0^{2\pi}\int_0^{\mu}\rho_s(\mathbf{n}\cdot\mathbf{i})\sin\phi\, d\phi\, d\theta - \int_{-\theta_\mu}^{\theta_\mu}\int_{\phi_\theta}^{\mu}\rho_s(\mathbf{n}\cdot\mathbf{i})\sin\phi\, d\phi\, d\theta\right) \tag{5.21}$$

Background Reflection

As with the background plane in the dry material case, the presence of a sphere on the background will cast a shadow which will result in a reduction of incident light and hence in the amount of specularly reflected radiance from the background. Given the same horizontal background plane used in the diffuse reflection model, the following vector values can be defined:

$$\begin{aligned}
\text{View Direction: } \mathbf{v} &= <0,0,1> \\
\text{Surface Normal: } \mathbf{n} &= <0,0,1> \\
\text{Incident Ray: } I_i\mathbf{i} &= I_i <\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi> \\
\mathbf{n}\cdot\mathbf{i} &= \cos\phi \\
\rho_s = (\mathbf{r}\cdot\mathbf{v})^n &= \cos^n\phi
\end{aligned}$$

For the simple case when $\phi_b > \mu$, the ball is sufficiently far from the point of interest so that the sphere does not cast any shadow. ϕ_b was defined in (5.13) as $\frac{\pi}{2} - 2\arctan\left(\frac{r}{d}\right)$. In this case, the reflected radiance is integrated over the whole dome to give:

$$\begin{aligned}
\text{Background Reflection: } I_{r,total}(\mu) &= \int_0^{2\pi}\int_0^{\mu} k_s I_i \cos^{n+1}\phi \sin\phi \, d\phi \, d\theta \\
&= k_s I_i \frac{2\pi}{n+2}\left(1-\cos^{n+2}\mu\right) \qquad (5.22)
\end{aligned}$$

In the second case where $\phi_b \leq \mu$, the component of reflection which is occluded by the projection of the sphere is found using the limits of integration for θ which were given in (5.16) as θ_ϕ. The occluded component is subtracted from (5.22) to give the amount of Occluded Background Reflection:

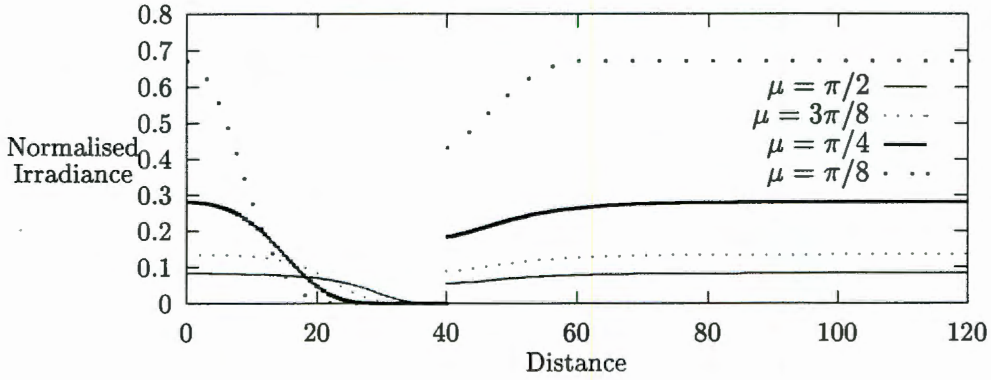
$$I_r(\mu) = k_s I_i \left(\frac{2\pi}{n+2}\left(1-\cos^{n+2}\mu\right) - 2\int_{\phi_b}^{\mu}\int_0^{\theta_\phi} \rho_s(\mathbf{n}\cdot\mathbf{i})\sin\phi \, d\theta \, d\phi\right) \qquad (5.23)$$

Wet Material Results

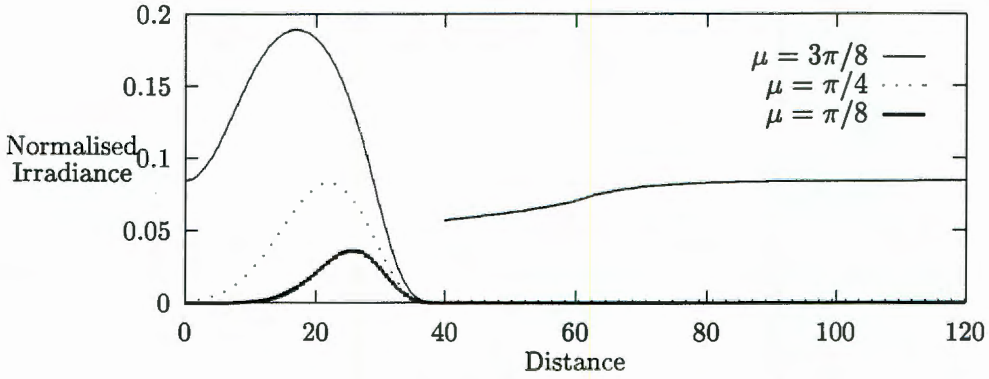
This section presents cross-section and image results of the specular reflection component and combined radiance effects under the same diffuse lighting conditions as the dry material results in Section 5.2.3.

Figure 5.14 shows example cross-sections of reflected radiance due to specular reflection for the three lighting models, i.e. top dome, side strip and floating band models. The reflected radiance was calculated using I_r from (5.21) for the sphere and (5.23) for the background. Figure 5.15 shows images produced by revolving the cross-sections around the image centre. Note that a material specular-reflection exponent value $n = 10$ was used to correspond to a fairly shiny surface in these results.

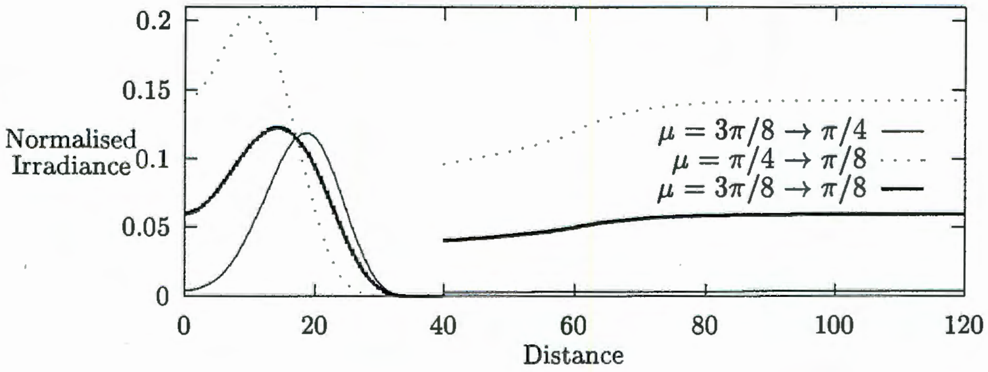
It can be seen from the graphs and images that when there is no light from the sides, there is a predictably low degree of reflected light at the edges of the ball. Correspondingly, when there is no light from above, the top of the ball appears dark with the result that there is a ring of higher intensity between the centre and the edge of the sphere.



(a) Top Lighting

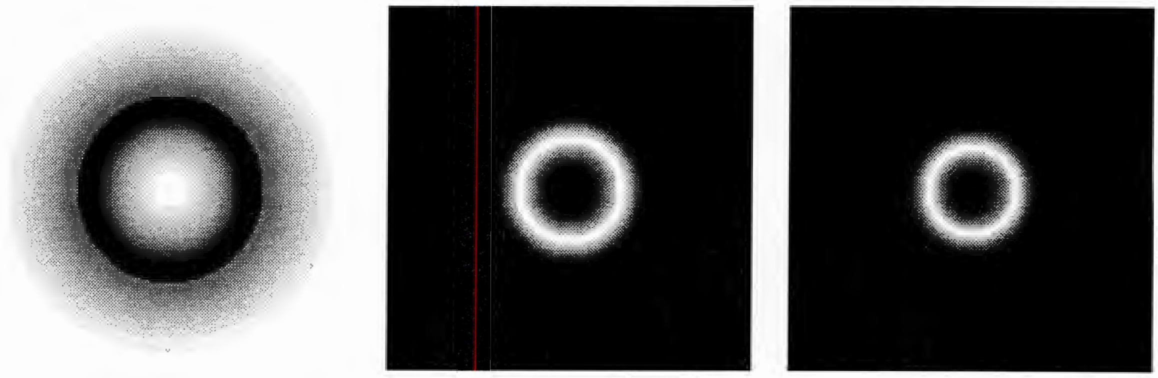


(b) Side Lighting



(c) Band Lighting

Figure 5.14: Cross-sections of reflected radiance given a ball of radius 40 on a flat plane under diffuse lighting. Only the specular reflection component calculated using Phong's illumination model with $n = 10$ is shown.



(a) Top $\mu = \pi/2$ (b) Side $\mu = \pi/4$ (c) Band $\mu = 3\pi/8 \rightarrow \pi/4$

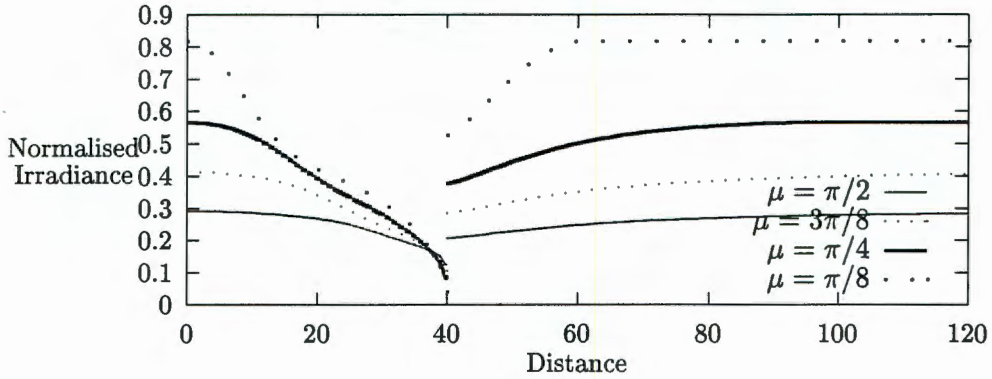
Figure 5.15: Three images of a white ball on a white background generated by revolving the indicated specular component of the reflected radiance cross-section about the origin.

Clearly, the case in which $\mu = \frac{\pi}{2}$ and the dome is a hemisphere produces a reflection component which is least likely to cause undesirable artifacts in the image, since it is most uniform and falls off in a similar manner to the diffuse component near the edge of the rock.

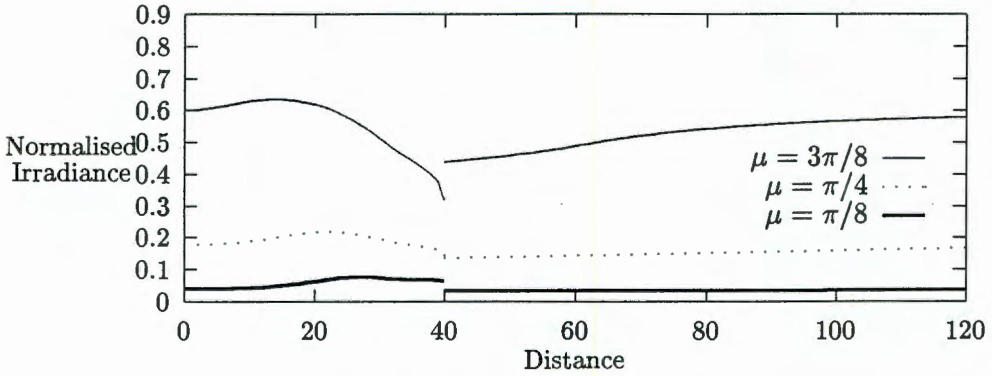
The results of combining the diffuse and specular components of reflected radiance are shown in Figures 5.16 and 5.17. For this example the diffuse and specular reflection coefficients were set to $k_d = k_s = 0.5$ for an equal weighting. It can be seen that while in Figure 5.17(a), the results are only slightly different to the diffuse only results, for the other two cases, rings are clearly visible thereby detracting significantly from their suitability as lighting sources.

The effect of n in the Phong model is to determine how shiny the surface is or how diffuse the highlights are. A low value results in very diffuse rings in the specular component of reflected radiance, while higher values correspond to increasingly narrow and intense rings. Regardless of the specular-reflection exponent, a hemispherical diffuse light source will produce diffuse reflections thereby decreasing the chance of undesirable artifacts and it would therefore appear that that would be the most suitable condition in which to operate this machine vision system.

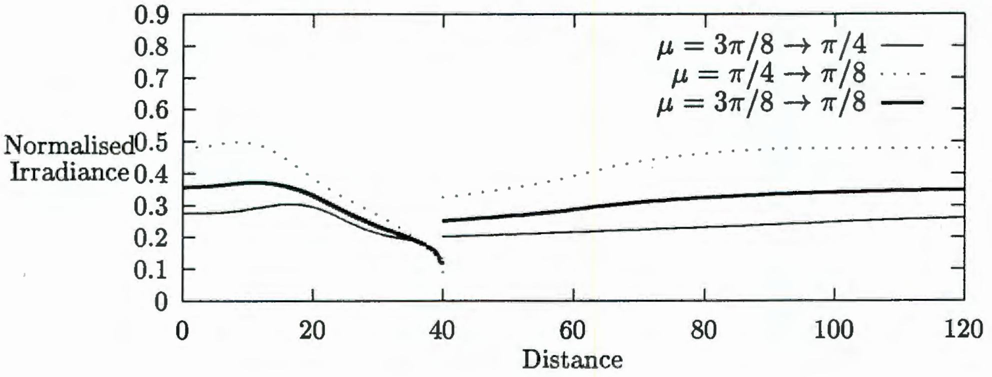
As stated earlier, Gennert *et al* [28] describe how to position circular, linear and point light sources to achieve uniform illumination of the origin. Figure 5.18 shows two ray traced images of the sample scene using a combination of diffuse and specular reflection predicted using Phong's model with an n value approximately equal to 10. It can be seen that although it may be possible to obtain reasonable results for the surface of the sphere in the four point source case, in both cases, there is quite a large amount of structure introduced to the image in the form of overlapping shadows. Using a circular light source would improve this situation; however, the results would be similar to those predicted for a thin band of diffuse light which



(a) Top Lighting

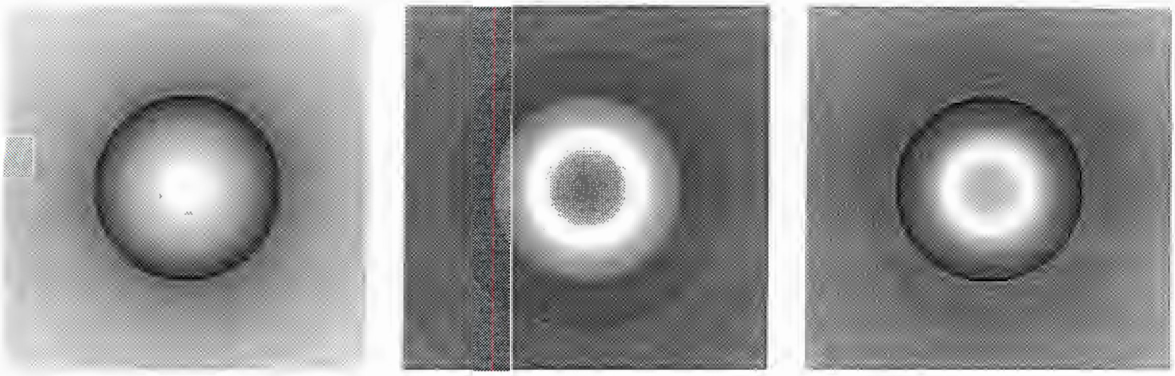


(b) Side Lighting



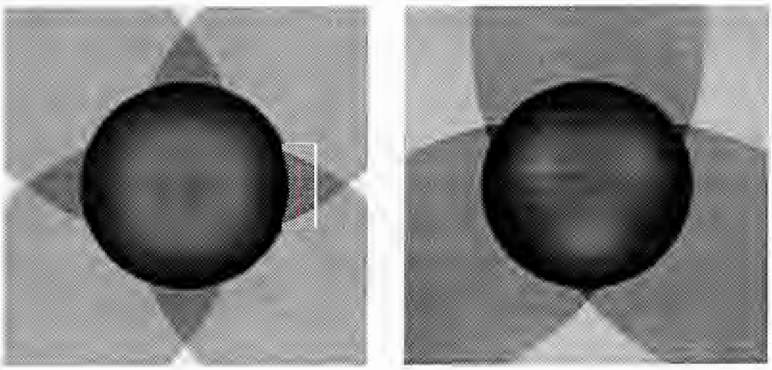
(c) Band Lighting

Figure 5.16: Cross-sections of reflected radiance given a ball of radius 40 on a flat plane under diffuse lighting. Both the diffuse and specular reflection components are included with equal weighting, i.e. $k_d = k_s = 0.5$.



(a) Top $\mu = \pi/2$ (b) Side $\mu = \pi/4$ (c) Band $\mu = 3\pi/8 \rightarrow \pi/4$

Figure 5.17: Three images of a white ball on a white background generated by revolving the indicated reflected radiance cross-section about the origin.



(a) Four point sources. (b) Three point sources.

Figure 5.18: Ray traced scenes using point source configurations from Gennert *et al* [28] to obtain uniform illumination. The two arrangements shown here were taken from their Figures 9 and 15 for four and three point sources respectively.

has been shown not to be optimum.

5.2.5 Discussion of Lighting

This section provides observations that can be made from the lighting model and results. It also gives some discussion of how this information has been incorporated into the algorithm design to improve rock detection and identification.

From the previous results when considering both dry and wet material, it would appear that the most suitable light model is that in which diffuse light comes from a dome which extends down to an angle of $\mu = \frac{\pi}{2}$ since it is least affected by an increase in the specular reflectivity of the material or conveyor belt and therefore provides most consistent conditions.

It can be inferred from the equations for reflected radiance that the bidirectional reflectivity ρ_d of a rock (its shade or colour) has a linear effect on its visual appearance or intensity in the image. Having a high ρ_d value will make the rock bright and a low value will make it relatively dark but dividing the intensity by ρ_d will result in a representation with normalised contrast.

This information can be used if ρ_d can be estimated. ρ_d could have been approximated here by finding and measuring local maxima in the image which would probably correspond to horizontal surfaces and hence the tops of rocks. Unfortunately, surface texture does bias this measurement and a more representative result is obtained if the local area mean value is used.

Given that image intensity decreases with surface orientation when light is incident from all directions, then assuming that ρ_d is known for a particular rock, an intensity threshold can be obtained for that rock below which any points have a high probability of being shadows or the edge of that rock[3].

Using the same argument, it is also possible to determine an intensity threshold above which any points are definitely part of that rock.

Clearly, these methods are very local and are only applicable within the bounds of the given rock. That is not a problem since the size of the rock will be known because a multiscalar approach has been adopted. It is therefore possible to determine threshold values and their approximate scope within the image based on the intensity of bright areas and the expected sizes of rocks. Since a threshold is only applicable within the bounds of the rock, it will be necessary to perform segmentation from the middle of the rock outwards to its boundary.

This knowledge has been incorporated into the algorithm design. It is used in Section 6.3.2 to provide a basis for preattentive vision and is later used as part of a feature to facilitate discrimination between regions of interest which have been segmented in Section 8.3.

---

[3] With top lighting and $\mu = \frac{\pi}{2}$ the minimum possible intensity should be half of that at the centre of the rock.

5.3 Image Sensing Characteristics

The aim of this section is to determine the effect of image sensing equipment on the visual characteristics of the image using the concepts of MTF and spatial resolution which were introduced in Section 3.3.1. Physical characteristics of imaging equipment and the actual spatial arrangement of the problem will be used to estimate the overall imaging modulation transfer function for an image acquisition system which is practical for this application. The following system components are considered in the next four sections: motion smear, spatial frequency response of a diffraction limited lens, CCD array and the frame-grabber.

5.3.1 Image Smear MTF

Motion of the conveyor belt is responsible for creating image smear which is caused by the object moving during exposure. Given that the conveyor belt moves at a speed of $v = 1\text{m.s}^{-1}$ and the period of exposure is T, then the distance moved is $\gamma = Tv$.

A JVC GR-AX107 video camera was used during image capture at the processing plant which has a focal length of $\text{f} = 6.7\text{mm}$. The camera was positioned at an approximate height of $o = 1.2\text{m}$ above the conveyor belt. This limit was imposed by the presence of the roof over the conveyor belt and resulted in almost complete coverage of the width of the belt.

Under these conditions, which are reasonably typical, the image distance $i = 6.74\text{mm}$ is determined using (3.18) and from (3.19) the image magnification, $\text{m} = \frac{0.00674}{1.2}$.

Figure 5.19 shows the vertical component of image smear MTF calculated using (3.22) for several different exposure times under the described conditions. The sampling rate that is indicated is equal to the inverse of the pixel spacing T_β on the CCD array which is defined in Section 5.3.3. Note that a desirable sampling rate will decrease the amount of image blur to negligible levels.

5.3.2 Lens MTF

Given the focal length of the lens found in the previous section and a reasonable estimate of pupil diameter, D, to be 3mm, the F/number of the lens is given by (3.20) as $\text{F} = \frac{6.7}{3}$. Green light will be used for estimating the lens MTF_L which has a wavelength $\lambda = 520\text{nm}$.

The MTF_L curve given by (3.21) for a perfect diffraction limited lens under the given conditions is shown in Figure 5.20.

5.3.3 CCD MTF

The use of a CCD array for detecting the image results in a low pass filter action as a result of the integration of photons in the potential wells which have a non-zero size. A CCD detector commonly consists of a 512 × 512 array of pixels with a typical size of 5mm square. The

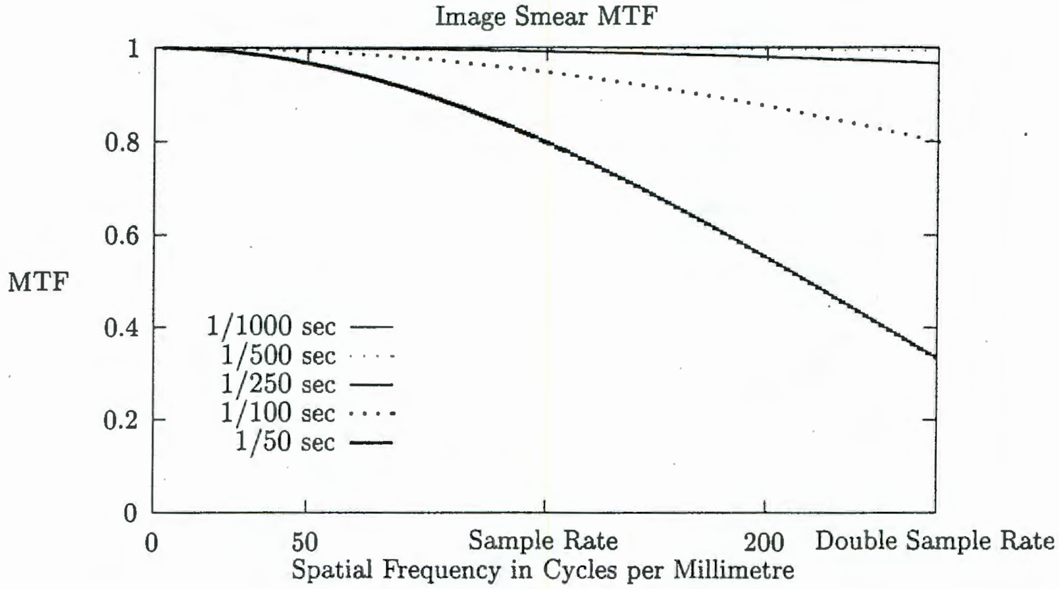


Figure 5.19: Image smear MTF for 5 exposure periods. Vertical component only is shown because horizontal component is unity due to direction of motion.

sample spacing of the CCD array can then be calculated to be $T_\beta = \frac{5}{512}$mm. Assuming that the active portion of each pixel is equal to three quarters of the pixel size, $T_\alpha = 0.75T_\beta$.

The MTF$_D$ curve given by (3.23) for the described CCD array is shown in Figure 5.20.

5.3.4 Combined Sensing MTF

The combined image acquisition system MTF$_S$ is determined using (3.17). Figure 5.20 shows MTF$_S$ which is the result of cascading the other curves by calculating their product at each spatial frequency. Note that the lens MTF$_L$ provides the dominating factor in the degradation of image quality while motion smear has a small effect and detector integration is negligible.

The shutter speed of $\frac{1}{250}$sec which is used for this graph appears to provide satisfactory results. Higher speeds will not significantly improve image quality while lower shutter speeds will be detrimental to image quality and image smear will become the dominating factor in image degradation for T $\geq \frac{1}{50}$sec.

5.3.5 Frame-grabber

A frame-grabber is used to sample a frame from a video signal to generate a digitised image in computer memory. The basic characteristics of frame-grabber features were introduced in 4.2.2. A typical frame-grabber, such as the MATROX PIP-EZ card, samples a frame at a resolution of 512×512 pixels where each pixel is represented by 8 bits. The image then consists of two interlaced fields which occupy every alternate line. Due to conveyor belt motion, only

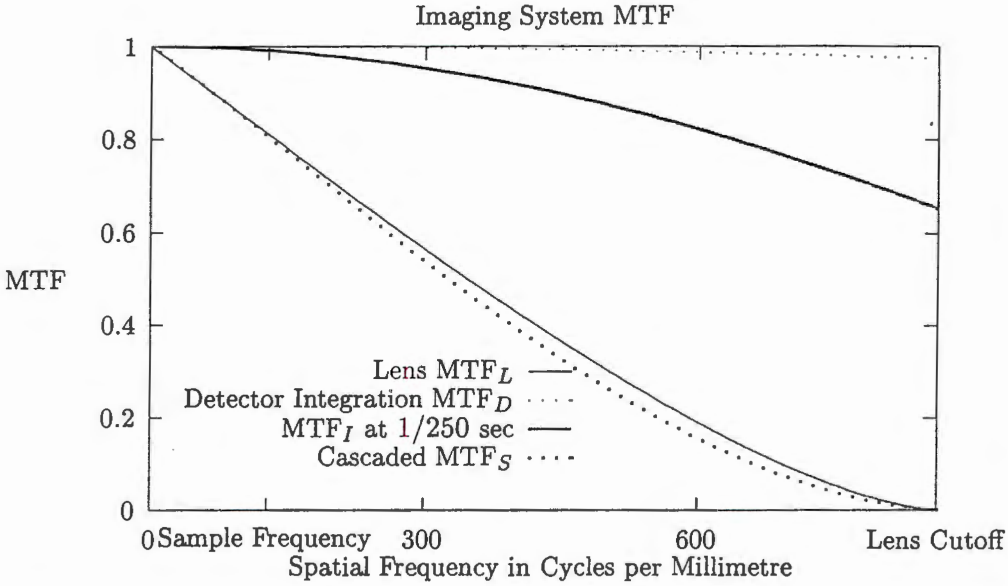


Figure 5.20: Combined image sensing MTF is determined as the product of MTF due to image smear, lens and CCD integration. A vertical cross-section of MTF$_S$ is shown for $\mu_1 = 0$. The horizontal cross-section is virtually identical to the lens MTF$_L$ due to the very small contribution of detector integration and absence of motion smear.

one field can be used thus reducing image resolution to 512 × 256.

Standard video displays have an aspect ratio of 4:3 and since the frame-grabber array is square, it has a distorted view of the scene, i.e. pixels have a 3:4 aspect ratio. Sampling the image to produce a 512 × 256 image results in an aspect ratio of 6:4 or 3:2 which could be corrected by resampling and interpolating (See Section 3.2.4) to give a final image with resolution 384 × 256 or 256 × 171 and an aspect ratio of 1:1. This correction is necessary as the distortion introduces a directional bias to the system which relies on symmetry.

An advantage of the aspect ratio of the image acquisition system is that it can be used to effectively increase the useful field of view of the vision system [5]. Orientating the camera so that the height of the image corresponds to the width of the conveyor belt and the material moves horizontally, results in a maximum length object of $\frac{4}{3}$ times the conveyor belt width. Under this condition, since a large object will be in the image for longer than if the camera were orientated so that the material moved vertically, it is far more likely to be correctly detected by the system.

5.4 Overview of Operating Conditions

This section provides a summary of the findings of this chapter with the aim of specifying desirable operating conditions for the hardware component of the instrument.

Algorithm for calculating suitable camera shutter speed for given physical conditions of use.

Given:		
	o	Camera to scene distance (m)
	v	Conveyor belt speed (m.s^{-1})
	f	Focal length of the lens (mm)
	pix_{num}	Number of pixels on a line
	CCD_w	Width of the CCD array (mm)

Then from (3.18), the image distance $i = 1/\left(\frac{1000}{f} - \frac{1}{o}\right)$.
Now using i, the image magnification can be calculated using (3.19) to give $\mathrm{m} = \frac{i}{o}$.
The size of a pixel on the CCD array is equal to $T_\beta = \frac{\mathrm{CCD}_w}{\mathrm{pix}_{num}}$ mm. From this, the sample spatial frequency is determined as $T_{sample} = \frac{1}{T_\beta}$ cycles.mm^{-1} or $\frac{1000}{T_\beta}$ cycles.m^{-1}.
A desirable exposure time can be found by making the spatial frequency $\mu_2 = T_{sample}$ cycles.m^{-1} and finding a value of T which will make (3.22) close to unity, i.e.

$$\frac{\sin\left(\frac{\mu_2 T v \mathrm{m}}{4\pi}\right)}{\frac{\mu_2 T v \mathrm{m}}{4\pi}} > 0.95$$

Figure 5.21: Algorithm for calculating exposure time.

It was observed in Section 2.2.5 that a thinner layer of rock material on the conveyor belt would be fundamentally easier to measure using machine vision, with the result that measurements obtained would be more accurate and reliable than those obtained where there is a deep pile and volume measurements have to be determined by extrapolation.

Analysis of lighting further indicates that dry material should be easier to analyse due to the lack of specular reflection. Since it is not generally possible to ensure this, the instrument should be capable of operating successfully on wet material and the lighting conditions should be designed appropriately.

Based on the results of Sections 5.2.3 and 5.2.4, a diffuse light source in the form of a dome extending down to $\mu = \frac{\pi}{2}$ should provide reasonably similar results for both wet and dry material. In practice, this will involve a lighting arrangement in which the scene is illuminated evenly from all directions. The advantage of this arrangement is a reduction in shadows in the images and a reduction in highlighting caused by specular reflection if the material or conveyor belt is wet.

A shutter speed or exposure time should be selected so that (3.22) gives a value close to unity given the conveyor belt speed, v, and image magnification, m, at the sample frequency of the CCD array. Figure 5.21 presents a procedure which can be used for the purpose of selecting a suitable shutter speed.

As indicated in the previous section, the aspect ratio of the imaging system can be used to advantage to increase the useful field of view of the image by orientating the camera so the the rock-stream moves horizontally across the image. This will increase the likelihood of a complete large rock being contained in the image and therefore improve measurement accuracy.

The following four chapters are concerned with the development of suitable algorithms for the purpose of identifying the rocks in the images which have the characteristics described in this chapter.

Chapter 6

Image Preprocessing

6.1 Introduction

The aim of this chapter is to describe the image preprocessing that is applied before the main analysis of the image can proceed. Image preprocessing is used to prepare an image for processing by attempting to highlight important features (eg outline extraction [66]), normalising the representation of data by compensating for variations (eg contrast equalisation, noise reduction and uneven illumination correction [56]), or converting the image into a form which will be more efficient for processing (eg Laplacian pyramid formation for image coding [10]).

In Section 2.2.3 it was found that an accepted model of the human visual system contains two stages. The first stage is preattentive vision which is responsible for automatically highlighting important features in the image for further analysis. This is a form of preprocessing and is therefore described later in this chapter.

Preprocessing is split into two parts in this system. An explanation of the process of image pyramid formation is described in Section 6.2 with special attention given to the specification of suitable smoothing before image scaling. A preattentive vision stage which is applied at each level of the image pyramid in order to detect regions of interest for subsequent attention focusing is then described in Section 6.3.

6.2 Image Pyramid Formation

6.2.1 Introduction

Multiscalar image processing and the use of an image pyramid were discussed in Section 2.3.3. It was found that due to the fractal nature of rocks, filters could be designed which are matched to a specific rock size. The use of an image pyramid therefore provides a foundation for the system.

Section 6.2: Image Pyramid Formation

This section discusses in detail the process of creating the image pyramid. The most important factor in this process is that of smoothing the image before resampling in order to obtain desirable visual characteristics of each level of the pyramid. An initial definition for desirable visual characteristics is that they are easy for a human to understand. It will be seen later that this may not be a suitable definition.

The use of an image pyramid in multiscalar image processing was introduced in Section 3.2.4 where it was found that a Gaussian smoothing kernel is commonly used for generating image pyramids because no new extrema are created during image smoothing. Application of the Gaussian kernel is further investigated here.

For the purpose of analysing properties of Gaussian smoothing kernels, the continuous Gaussian will be used. Once appropriate parameters are determined a discrete approximation created using (3.13) attributed to Lindeberg [49] will be used for actual filtering. This approach is justifiable since the exact properties of the filter are not critical in the application because it does not rely on exact behaviour of the filter. Since the Gaussian is circularly symmetrical and separable because $e^{x^2+y^2} = e^{x^2}e^{y^2}$, the analysis of the Gaussian will take place in one-dimension only for the sake of simplicity.

As described in Section 3.2.4, generating an image pyramid involves generating a stack of images each of which requires two steps [1, 10]: initial smoothing of the image and resampling to the desired size. In order to achieve this for any level of the pyramid it is necessary to know the following information: the amount of image size reduction that is required and the amount and type of smoothing.

6.2.2 Size Reduction

How much size reduction should there be at each level of the pyramid? Most applications described in Section 3.2.4 use a dyadic pyramid, i.e. the size of each level is an integer power of two and so the size reduction is a factor of a half.

In this system, it is important to have less size reduction between each level of the pyramid because the attention focusing procedures have a response which is localised in scale-space, i.e. they only respond to objects in a narrow range of sizes. The scale distance between pyramid levels must be less than the range of attention focusing to ensure that no rocks are accidentally missed because they fall in a size range between the acceptance of two adjacent pyramid levels. In addition to the response of attention focusing, it is necessary to give a good input to the hierarchical analysis routine for acceptable results. This is characterised by continuity of detected points through scale-space which correspond to rocks and can be grouped based on proximity.

If the size reduction factor is represented by f_{scale} which can have a value in the range 0–1, then from the above arguments, a value close to 1 is desirable to ensure that the scale of adjacent levels is similar and the pyramid is almost a continuous solid. Note, however, that a high value of f_{scale} results in increased processing time due to a large number of pyramid levels. After investigation of system performance, it was found that below a minimum value for f_{scale} of 0.8, the sampling distance between levels was too great and performance was degraded due to a high probability of missing rocks or determining an incorrect segmentation resulting in inaccurate size results. Most of these problems are reduced when f_{scale} is increased to the range 0.9–0.95.

6.2.3 Smoothing

Based on the arguments for Gaussian smoothing given in Section 3.2.4, a Gaussian kernel, or discrete equivalent of it, is used to perform smoothing. Assuming that it is necessary to reduce the image by a factor M, i.e. multiply the size of the image by a factor $1/M$ (Note that $M = 1/f_{scale}$), it is necessary to decide on the amount of smoothing to perform.

The remainder of this section is split into three parts: firstly, some useful properties of Gaussian smoothing are introduced, secondly the degree of smoothing is determined by specifying a standard deviation, and finally, windowing of the Gaussian kernel for computational purposes is discussed.

Properties of Gaussian Smoothing

Since performing smoothing basically corresponds to low-pass filtering, all the elements of the filter kernel should sum to one for a unity DC response. It is therefore necessary to normalise all smoothing kernels which results in the first property of Gaussian kernels.

Property 6.1 *A Gaussian smoothing kernel is uniquely specified by its standard deviation,* σ, *i.e.* $G(t;\sigma_1) \neq G(t;\sigma_2)$ *only if* $\sigma_1 \neq \sigma_2$.

The second property relates to the Fourier transform of a Gaussian kernel and states:

Property 6.2 *The Fourier transform of a Gaussian smoothing kernel,* $G(t;\sigma)$, *is also a Gaussian,* $G_{FT}(\omega;\sigma)$, *with the inverse standard deviation, because*

$$\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-t^2}{2\sigma^2}} \xrightarrow{\text{FT}} e^{\frac{-(\sigma\omega)^2}{2}}$$

Next, it is necessary to prove a cascading property of Gaussian smoothing kernels, i.e. the effective smoothing after first convolving with $G(t;\sigma_1)$ and then with $G(t;\sigma_2)$ is equivalent to convolving once with $G(t;\sqrt{\sigma_1^2+\sigma_2^2})$. This can be stated as:

Property 6.3 *Convolving two Gaussians with standard deviations of* σ_1 *and* σ_2 *respectively results in another Gaussian with a standard deviation of* $\sqrt{\sigma_1^2+\sigma_2^2}$, *i.e.* $G(t;\sigma_1) * G(t,\sigma_2) = G(t;\sqrt{\sigma_1^2+\sigma_2^2})$.

Proof:

$$\begin{aligned}
\textit{Convolution property: } G(t;\sigma_1) * G(t;\sigma_2) &\xrightarrow{\text{FT}} G_{FT}(\omega;\sigma_1)G_{FT}(\omega;\sigma_2) \\
&= e^{\frac{-(\omega\sigma_1)^2}{2}} e^{\frac{-(\omega\sigma_2)^2}{2}} \\
&= e^{\frac{-(\sigma_1^2+\sigma_2^2)\omega^2}{2}} \\
&= G_{FT}(\omega;\sqrt{\sigma_1^2+\sigma_2^2}) \\
\textit{and: } G(t;\sqrt{\sigma_1^2+\sigma_2^2}) &\xleftarrow{\text{IFT}} G_{FT}(\omega;\sqrt{\sigma_1^2+\sigma_2^2}) \\
\textit{Therefore: } G(t;\sigma_1) * G(t;\sigma_2) &= G(t;\sqrt{\sigma_1^2+\sigma_2^2})
\end{aligned}$$

From this property, it can be determined that if an image has already been smoothed with a Gaussian kernel, as in the case where a level of the pyramid must be reduced further, the effective amount of filtering that is still required can be expressed as σ_e where

$$\sigma_d = \sqrt{\sigma_e^2 + \sigma_i^2} \tag{6.1}$$

Here, σ_d is the overall desired amount of smoothing and σ_i is the amount of smoothing that has already taken place. This is a formula for calculating the smoothing kernel if no scaling is to take place.

If the image is to be scaled by a factor of $1/M$ it must be resampled with a new spacing of M. In this case, the desired value of σ_d is specified for after scaling. In order to calculate σ_e, the value of σ_d must be known before scaling. This is found using another property of a Gaussian:

Property 6.4 *Scaling a Gaussian with a standard deviation of σ_s by a factor $1/M$ in the space domain produces another Gaussian with a standard deviation of σ_s/M, i.e. $G(Mt;\sigma) = G(t;\sigma/M)/M$.*

Proof:

$$\begin{aligned}
G(Mt;\sigma) &= \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(Mt)^2}{2\sigma^2}} \\
&= \frac{1}{M}\frac{1}{\sqrt{2\pi}\frac{\sigma}{M}} e^{\frac{-t^2}{2(\frac{\sigma}{M})^2}} \\
&= \frac{1}{M} G(t;\frac{\sigma}{M})
\end{aligned}$$

This is illustrated graphically in Figure 6.1 which shows the effect on the shape of a Gaussian of changing the sampling spacing and rescaling the axes.

It can be seen from Property 6.4 that if the desired sigma value is σ_d after scaling with a factor of $1/M$, then it is $M\sigma_d$ before scaling. Modifying (6.1) to calculate effective smoothing

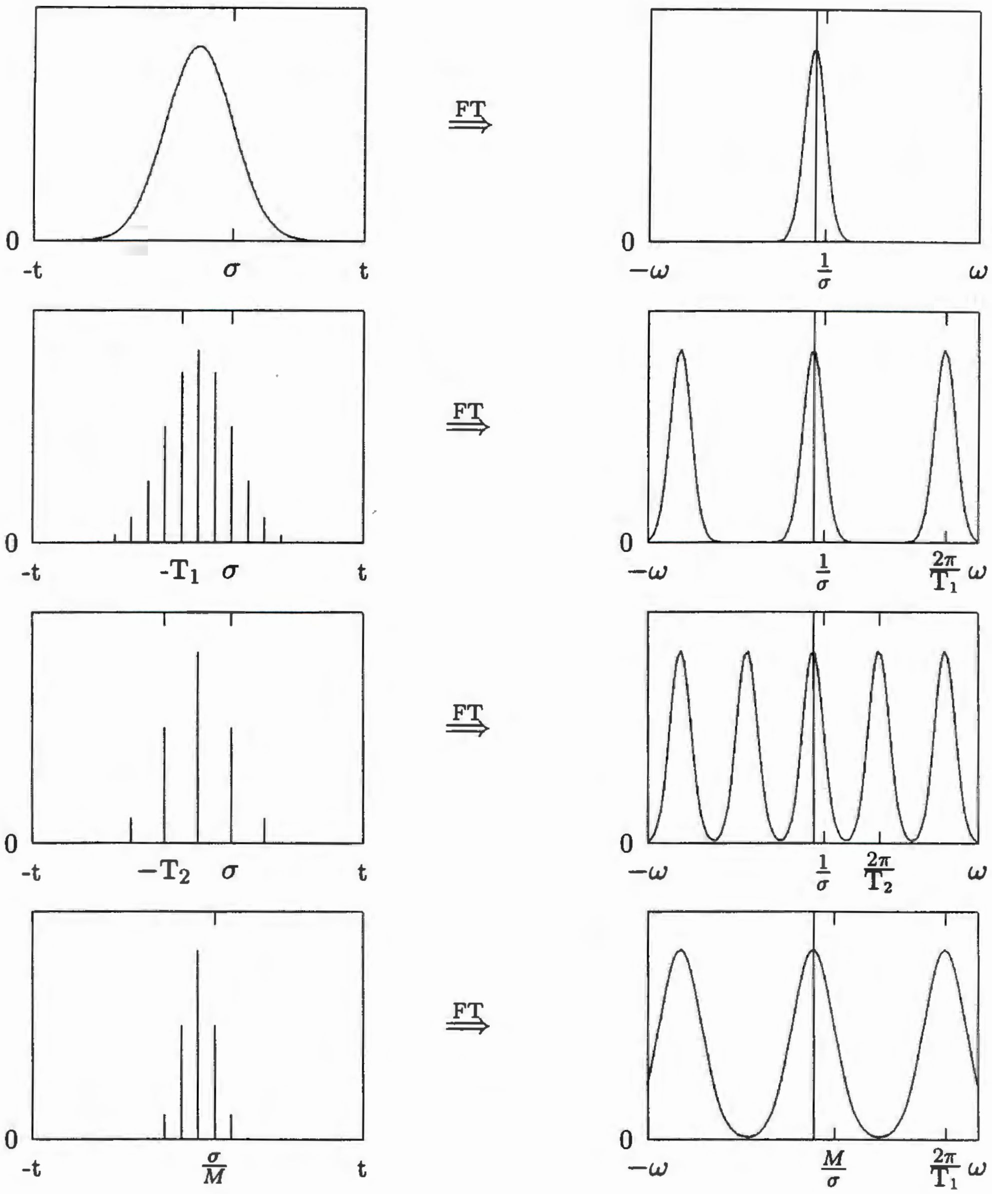


Figure 6.1: Fourier transform pairs showing the effect on a Gaussian of sampling and rescaling to reduce the size of a sequence. The top pair gives a Gaussian curve with its Fourier Transform. Sampling the Gaussian with a spacing of T_1 results in the second pair. Subsampling the second sequence at every second point is equivalent to sampling the Gaussian with a spacing of $T_2 = MT_1$. Rescaling the third pair by a factor of $1/M$ results in a new Gaussian form with a new sigma value of $\sigma_n = \sigma/M$.

value while including a scale factor of $1/M$ results in the following equation:

$$\sigma_e = \sqrt{(M\sigma_d)^2 - \sigma_i^2} \tag{6.2}$$

Determining the Degree of Smoothing

As was stated in Chapter 4, the width of the conveyor belt places a limit on the smallest rock size which can be detected. As a wider belt must be accommodated in the image, rocks will appear smaller as the camera is moved away from the belt or the focal length is decreased. In conjunction with this, the degree of smoothing further limits the resolution of the system. A large amount of smoothing results in a large region of interest for sufficient detail whereas a smaller region of interest can be used when there is less smoothing. Since the size of the region of interest is the same for each pyramid level, it determines the smallest size rock that can be detected at the base of the pyramid.

It is now necessary to decide on the desired amount of smoothing in order to specify σ_d. Figure 6.2 shows the effect on aliasing in the frequency domain of three possible values of σ. Since $\sigma = 1$ has marginal aliasing, it would appear that the desired value must be greater than or equal to 1 because it will result in the maximum image detail without introducing aliasing.

The other consideration when choosing σ_d is the size and detail of rocks to be detected. It would be best to capture just the macro shape of the object without having to contend with complicating surface texture detail. Therefore a higher value of sigma with a lower cutoff frequency could be chosen.

If the expected (or desired) rock size is expressed as a radius $r_{expected}$, then given an expected rock diameter of $n = 2r_{expected}$ pixels across, and an allowable uncertainty in the position of its edges as a fraction f_{edge}, then the cutoff frequency should correspond to spatial features which are smaller than or equal to nf_{edge}. The cutoff frequency (-3dB point) of a Gaussian smoothing kernel is found by solving the expression $G_{FFT}(\omega;\sigma) = 1/\sqrt{2}$ for ω resulting in $\omega = \sqrt{2\ln\sqrt{2}}/\sigma \approx 1/\sigma$. This corresponds to a spatial wavelength of $2\pi\sigma$ or a minimum feature size of $\pi\sigma$. The size of the smallest possible feature in the image is related to the highest spatial frequency. The Nyquist criterion states that for a signal to be unambiguously detected, it must be sampled at intervals corresponding to half the period of this frequency. Equating this with the desired image detail and solving for σ gives the following equation:

$$\sigma_d = \frac{nf_{edge}}{\pi} \tag{6.3}$$

The dynamic range of the system is the maximum range of rock sizes which can be detected in an image. The smallest rock that can be detected occurs at the base of the pyramid. The largest rock that can be detected occurs at the highest level of the pyramid which is equal in size to the rock size, n. Given an image width of i pixels, the relative size of the largest

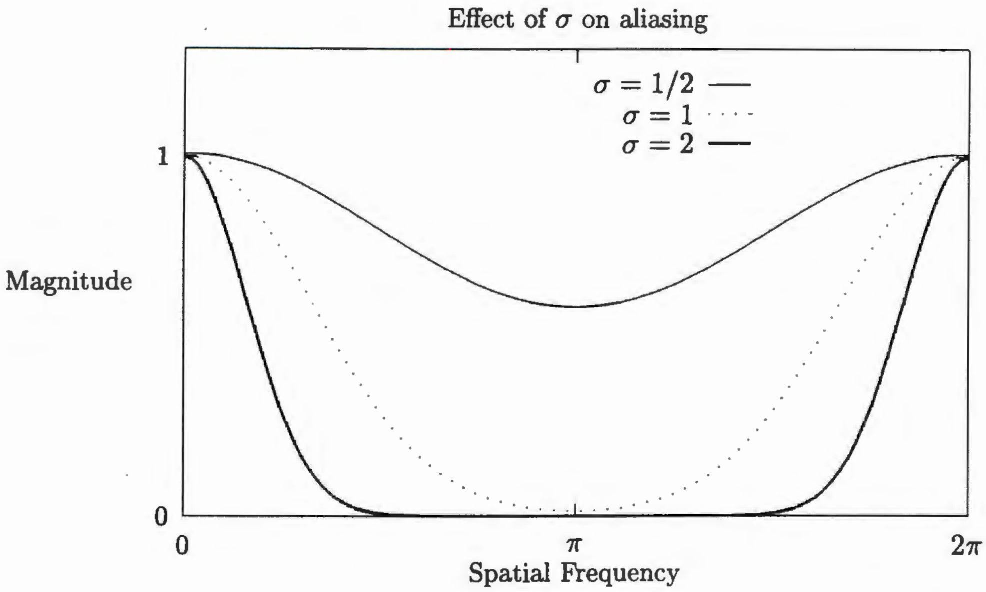


Figure 6.2: Choosing desired sigma value for smoothing. Graph shows the effect of choosing a sigma value in the space domain on the amount of aliasing in the frequency domain. The curves are created by summing Gaussians at the origin and 2π which is equivalent to a sample spacing of 1. Clearly, the most smoothing which corresponds to the value of $\sigma = 2$ has the lowest frequency cutoff and no aliasing. $\sigma = 1$ has marginal aliasing and $\sigma = 0.5$ has severe aliasing.

detectable rock compared to the smallest rock is equal to i/n. This is a possible representation of the dynamic range of the system.

The speed of the segmentation algorithm is proportional to the square of the rock size, n, since it operates on the area of the rock. It can therefore be seen that there is a trade-off between the following four parameters of the system: the amount of aliasing determined by choice of σ_d, edge and hence area accuracy determined by f_{edge}, processing time, and the dynamic range of the system.

Although there is a fair deal of flexibility with each of the parameters, all should be limited to within reasonable constraints. Figure 6.3 shows a possible solution to the problem of selecting operating parameters. In this figure, aliasing has been suppressed by setting $\sigma_d = 1$. The dynamic range is calculated using a typical image width of $i = 256$ as $R_d = \frac{256}{n}$. As stated previously, processing time is proportional to n^2 and is in the order of 10 minutes for $n = 15$ on a Sun Sparcstation 20. This results in processing time $T_p = \frac{n^2}{22.5}$. Edge accuracy as a percentage is $f_{edge} = 100\frac{\sigma\pi}{n}\%$.

It is desirable to maximise dynamic range while minimising processing time and f_{edge}. Normalising the cost of range to 10, processing time to 10 minutes and uncertainty to 5% and

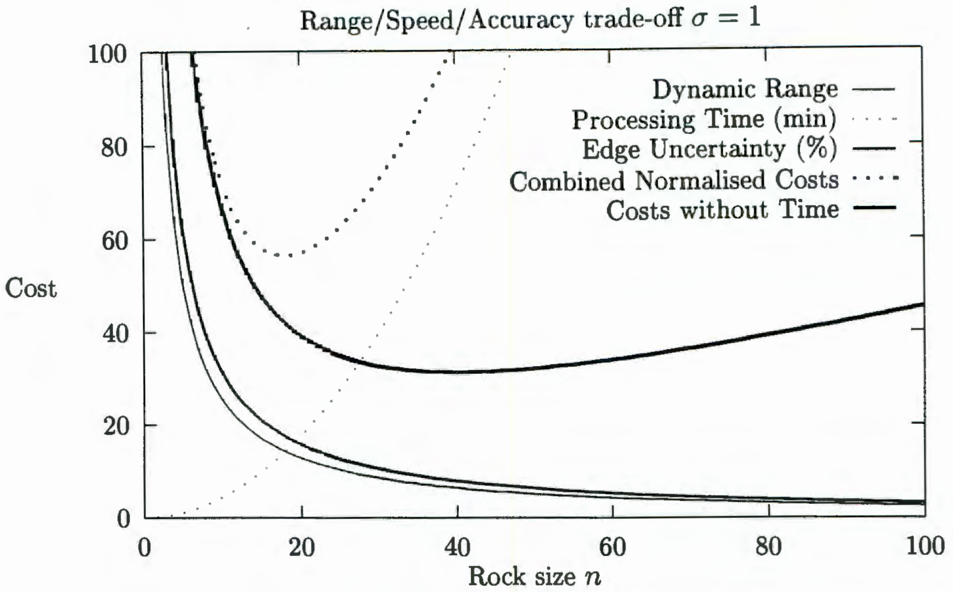


Figure 6.3: Graph showing the tradeoff between different effects associated with rock size when the value of σ is fixed at 1 in an image with a width of 256 pixels. There are three effects: processing time which is $O(n^2)$, dynamic range which is inversely proportional to n, and edge uncertainty which is inversely proportional to n. The optimal operating position is determined by finding the cost of the three curves by normalising and summing them. Equal costs correspond to a processing time of 10 minutes on a Sun Sparcstation 20, a dynamic range of 10 which would be suitable for available test material, and edge uncertainty of 5%. A second curve is shown which does not include processing time for another possible operational setting. Note that the cost associated with dynamic range is inversely proportional to dynamic range and is hence linear.

summing gives the following weighted equation (the weights make each term comparable):

$$\text{Cost} = \frac{100}{R_d} + T_p + 2f_{edge}$$

Finding the value for n which minimises this cost in Figure 6.3 and calculating the other parameters from it results in the following operating setting: $\sigma_d = 1$, $n \approx 18$, $R_d \approx 14$, $T_p \approx 14$ minutes and $f_{edge} \approx 17\%$.

When an image is smoothed using $\sigma_d = 1$, the result appears blurred. This is due to the fact that normally there is an element of aliasing in natural images. Figure 5.20 shows the combined MTF of the imaging system used to capture images. From this graph it can be noted that there will be severe aliasing when the image is sampled at the shown sampling frequency since the frequency response of the system is only down to approximately -3dB at the sample frequency. This effect can be simulated by selecting a Gaussian smoothing kernel with a cutoff frequency that corresponds to the sample frequency. This corresponds to a value of $\sigma_d = 1/3\pi$. A new operational setting based on the new value of σ_d is $n \approx 8$ and

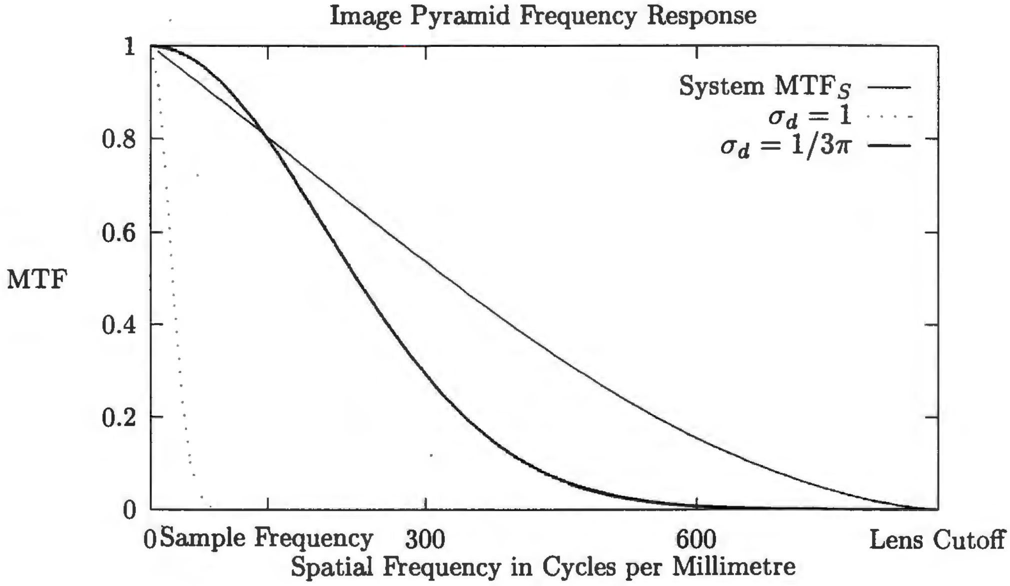


Figure 6.4: MTF of system with Gaussians showing Low Pass filtering. Graph of comparison between frequency response of the imaging system (MTF$_S$) and that of Gaussian smoothing. Smoothing curves are shown of Gaussians with standard deviations of 1 and $1/3\pi$ in sample space. Note that when $\sigma = 1$, there is negligible aliasing but the frequency response is completely different to the imaging system. The alternate option where $\sigma = 1/3\pi$ attempts to model the frequency response of the imaging system by having a similar cutoff frequency (-3dB) and hence results in a dramatic level of aliasing.

$f_{edge} \approx 5\%$. Another alternative obtained by discarding processing time from the weighting function is $n \approx 15$ and $f_{edge} \approx 3\%$.

Figure 6.4 shows each of the values of σ_d compared to the system MTF. Two reasonably extreme values have been selected here for σ_d since it is not certain whether or not the machine vision system will be able to cope with un-bandlimited images directly from the imaging system. It is therefore necessary at this stage to to follow both options and make a comparison once the performance of the whole system can be measured.

Windowing the Gaussian Kernel

The final task before an image pyramid can be created is determining how the convolution kernel should be created. Because the Gaussian has infinite domain, it is necessary to perform windowing to produce a computationally achievable kernel. A rectangular windowing scheme has been selected for this purpose because of its ease of implementation. Investigation of tables of the normal distribution function [59, p490] show that $F(z) = \frac{1}{\sqrt{2\pi}} \int_{-z}^{z} e^{-\frac{t^2}{2}} dt$ when $\sigma = 1$. Selecting a cutoff point as $z = \sqrt{2\pi}\sigma$ results in a value of $F(z) = 0.988$ which is a measure of the energy of the Gaussian which is contained below z. Since the two-dimensional

Gaussian is separable, the energy can be calculated by squaring $F(z) = 0.988$ which results in 0.976. This is very nearly all the power in the Gaussian which means that there should be only minimal effect due to windowing if this cutoff is used.

6.2.4 Example Image Pyramid

From Figure 6.4, it can be seen that smoothing using a Gaussian does not result in an image with the same characteristics as that from the camera. When $\sigma_d = 1$ is used there will not be aliasing which results in very blurred images. When $\sigma_d = \frac{1}{3\pi}$ there is a large degree of aliasing but a much closer frequency response to that of the camera with the result that the image is more visually acceptable.

This does not, however, mean that a machine vision system will be successful when applied to the aliased image pyramid or visa versa and so both possibilities are tested in Chapter 10.

Based on the findings of this section, image pyramids will be created for testing using the following information:

- The size reduction factor f_{scale} can have a value in the range 0.8–0.95.
- A Gaussian smoothing kernel will be used with a desired standard deviation of $\sigma_d = 1$ or $\frac{1}{3\pi}$.
- The smoothing kernel will be truncated at $\sqrt{2\pi}\sigma_d$.

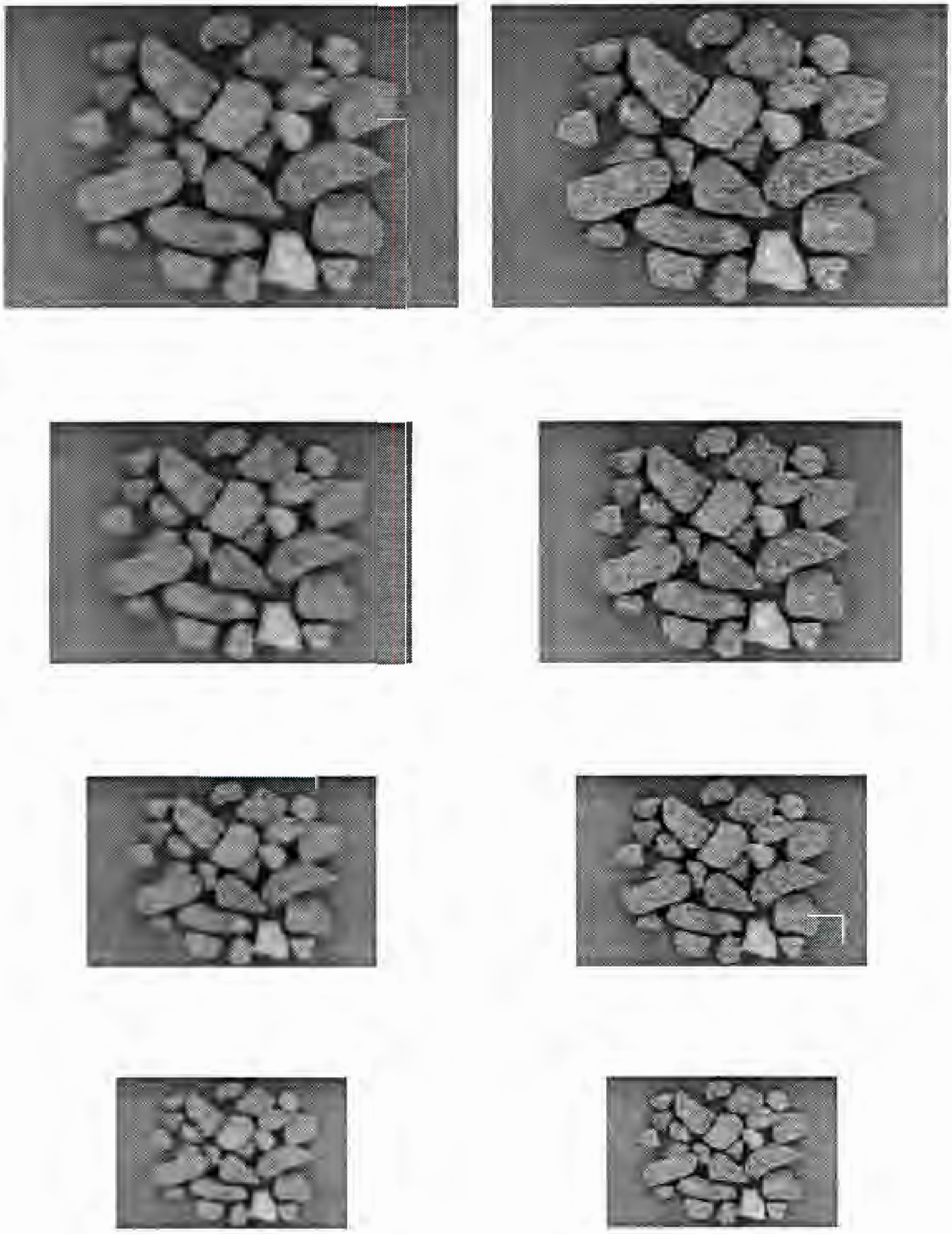
Given the input image, I_{camera}, and a discrete approximation of the Gaussian smoothing kernel, $K(x, y; \sigma_d)$, with a standard deviation equal to σ_d, then the base level of the pyramid is created using $P_1 = I_{camera} * K(x, y; \sigma_d)$ since the camera image has not yet been smoothed.

Successive levels of the pyramid can then be generated in one of two ways:

1. Each level is generated from the previous one so that $P_n = Rescale(P_{n-1} * K(x, y; \sigma_e), f_{scale})$ where $\sigma_e = \sqrt{(\sigma_d / f_{scale})^2 + \sigma_d^2}$.
2. Each level is generated from the initial level such that $P_n = Rescale(P_1 * K(x, y; \sigma_e), f_{scale}^{n-1})$ where $\sigma_e = \sqrt{(\sigma_d / f_{scale}^{n-1})^2 + \sigma_d^2}$.

Note that $Rescale(I, f)$ is a function which scales an image I by a size factor of f using the nearest neighbour method of resampling described in Section 3.2.4.

Two example image pyramids which were created using this process and the two values of $\sigma_d = 1$ and $\frac{1}{3\pi}$ are shown in Figure 6.5. Note that when $\sigma_d = \frac{1}{3\pi}$, although there is significant aliasing, the results appear to be subjectively perfectly acceptable to the human eye.



(g) $\sigma_d = 1$

(h) $\sigma_d = 1/3\pi$

Figure 6.5: Example image pyramids created using two different desired values of standard deviation. The size reduction factor used here is $f_{scale} = 0.8$ and only the first four levels are shown. In practice, the pyramids continue until the level size is equal to the rock size, n.

6.3 Preattentive Vision

6.3.1 Introduction

This section describes the development of the first level of the preattentive vision process which was introduced in Section 4.3.2. Its role is to highlight positions in each level of the image pyramid where target features in the form of ellipses are present, which could indicate the presence of a rock, for further analysis by the second level of preattentive vision which is described in Section 7.6.

Preattentive vision or perceptual organisation processes have been investigated previously in machine vision applications, usually as a means of modelling some aspect of the behaviour of biological vision systems. Marr [54] proposed a primitive description of the image called the primal sketch in which image features such as lines are detected and composed into simple structures based on their proximity and similarity.

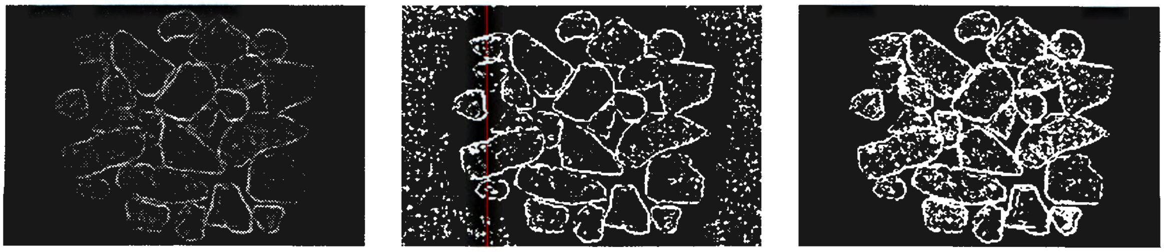
Sarkar and Boyer [80, 81] review perceptual organisation in computer vision and find that the organisation in images observed by Gestalt psychologists and the Gestalt principles and laws of grouping provide a strong basis for its implementation.

Because preattentive vision operates on the whole image with no regard for image context or high level syntactic information, it can usually be accomplished using parallel processes for high performance.

Since the human visual system has been adopted as a basis for the rock measurement instrument, a preattentive visual process based on perceptual organisation has been investigated and is described here. It was found, however, that for this specific application it was not necessary to develop a flexible preattentive process as the image content was restricted to a particular type, i.e. rocks. Therefore, a reasonably simple process based on the identification of elliptical objects in the image is presented here which satisfies the requirements of this system. Note that in this case, knowledge about the specific problem domain could be used to restrict the search to features of a particular size and range of shapes with no loss of functionality due to the multiscalar method adopted and the nature of the objects to be identified.

The main goal of preattentive vision is to ensure that all rocks in an image are highlighted using an appropriate ellipse since if they are missed at this stage, they will not be found later. Clearly, this can be achieved using an exhaustive search, but that would be prohibitively expensive so the second, and lesser goal, is to reduce the number of false hits to as low a value as possible.

Target or feature detection is achieved using two processes in the first level of the preattentive vision stage: firstly the edges of rocks are highlighted in the image and then patterns are detected which could correspond to one of a number of expected rock shapes which are modelled as ellipses. The goal of this process is to produce a list of highlighted points for



(a) Morphological Gradient

(b) Moment-Preserving local area threshold.

(c) Global threshold $t_{morph} = 30$.

Figure 6.6: Highlighting rock edges using morphological gradient detection followed by thresholding with $r_{expected} = 10$.

each level of the image pyramid P_j where an elliptically shaped object may be present.

6.3.2 Highlighting Rock Edges

This section presents two methods which are used to identify rock edges in the image by creating an edge image I_{edge} for the purpose of ellipse detection in Section 6.3.3. This process is similar to the edge enhancement preprocessing which occurs in the human eye and is responsible for the appearance of Mach bands [70].

In the first case, a simple approach using an edge detection algorithm based on intensity gradient is described. The second alternative incorporates knowledge obtained from the lighting model to produce a more robust algorithm based on image intensity that is matched to optimally highlight the edges of rocks which fall into the expected size category. Note that each of these methods results in a thresholded image since the Hough transform operates on a binary image.

Morphological Edge Detection

In this case, image gradient is used to find the edges of rocks in the image. Edge strength is determined using the morphological gradient detection operator $GRAD(f, g)$ (described in Section 3.2.3), because of its isotropic response. Alternatives such as the magnitude of the gradient could also be used.

Given the input image P_j, the edge strength is calculated as $I_{GRAD} = GRAD(P_j, D)$ where D is a disc structuring element with unit radius (this corresponds to a point and its four-connected neighbours). The output edge image is then found by thresholding I_{GRAD} to give I_{edge}.

Figure 6.6 shows the edge strength image I_{GRAD} in (a) and two thresholded versions of it in (b) and (c). In (b) the local area moment-preserving threshold described in Section 3.2.2

is used with a window size equal to the expected rock size n and a grid spacing of $r_{expected}$ pixels. In (c) a global threshold $t_{morph} = 30$ has been applied. The threshold of 30 was selected for display as that threshold at which the rock edges are all visible. An ROC curve is used in Section 6.3.4 to determine the most suitable value for t_{morph}.

Figures 6.6(b) and (c) illustrate two common problems which occur when edge strength is used for edge detection purposes. In (b), rock edges are present but the threshold adapts to the mean in outlying areas of the image where no rocks are present resulting in noise in the binary image. In both (b) and (c), a lot of noise is caused by internal edges which correspond to surface texture and is significantly detrimental to the usefulness of these images.

As can be seen from Figure 6.6, when rock edges are highlighted using image edge strength, the surface texture of the rocks results in considerable amounts of spurious noise being detected. This problem was also noted by Yeo *et al* [96] who proposed a region-based thresholding approach as a desirable alternative. As a result, another approach based on image intensity and thresholding using observations from the lighting model has been developed and is described next.

Adaptive Thresholding based on Intensity

Image intensity is determined by reflected radiance which is proportional to the diffuse bi-directional reflectivity of the rock ρ_d and varies with surface orientation as found in Section 5.2. A local intensity maxima in an area of the same size as the expected rock probably corresponds to a horizontal surface of a rock assuming that a rock is there (Note that this assumption is tested in Chapter 8). The intensity value is proportional to ρ_d which is assumed to be constant (apart from texture) over the whole of that particular rock. This assumption is valid for all material that has been investigated thus far.

In Section 5.2.5 it was found that given a diffuse lighting arrangement in which light is incident from all directions equally, the intensity of the rock will decrease below a threshold t_{edge} towards the edge of the rock where surface orientation increases or as shadows are caused by the close proximity of two rocks. It was shown that for that particular light model, edge points which are perpendicular to the camera will have half the reflected radiance and hence half the intensity of the maximum intensity from a horizontal surface.

Due to unmodelled light effects such as mutual illumination and reflection, $t_{edge} = \frac{1}{2}$ may not be appropriate in practice and experimentation was required to determine a more suitable value under the actual conditions. A process of investigation was followed during which it was found that a low value of 0.6 results only in relatively deep shadows being detected which give only partially connected edges, while a higher threshold around 0.75 increases the area of detected shadows and produces almost complete edges around rocks without introducing internal edge points. ROC curves can be created to facilitate the selection of t_{edge} to maximise the performance of the preattentive vision stage. This is done in Section 6.3.4 once all the components of the stage are available.

The following three stage technique has been implemented to use t_{edge} for highlighting rock edges to determine I_{edge}:

Smooth In order to reduce the effect of surface texture resulting in fine scale intensity patterns on the surface of a rock, the image is initially smoothed using a moving average window. This is achieved using the expression $I_{average} = P_j * K$ which can be calculated efficiently. K is a relatively small averaging convolution kernel with a size of $\frac{2}{3}$ expected rock diameter. This value has been used so as not to decrease the maximum intensity estimate by including edge points where the intensity has decreased due to surface orientation and not texture.

Dilate The morphological dilation operator described in Section 3.2.3 is then used to enlarge the maxima to the expected size of a rock by using a disc structuring element D with a radius of $r_{expected}$ and applying the following expression: $I_{dilate} = I_{average} \oplus D$. It is not usually necessary to have D as large as the expected rock size since the rock material usually has an area of almost uniform intensity in the centre and intensity only decreases near the edge of the rock. Using a smaller disc would therefore result in a slightly lower intensity value at the edge of the rock, but would decrease the computational load. Therefore, the reduced computational burden may be obtained with a slight increase in the value of t_{edge}.

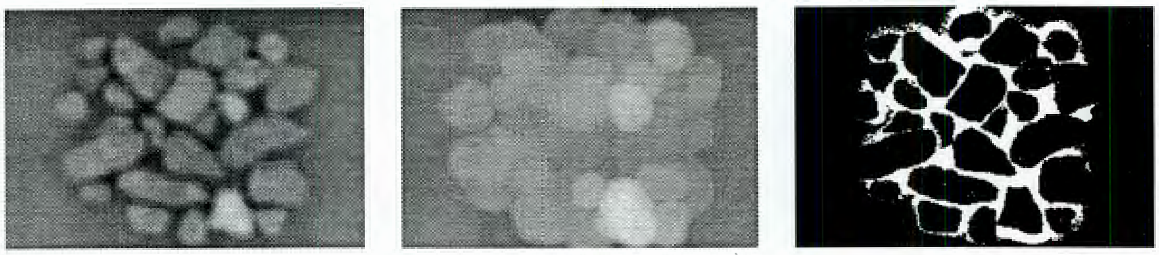
Threshold The image I_{edge} is finally obtained using an adaptive threshold routine operating on P_j and I_{dilate}. Each point in I_{edge} is calculated using the expression:

$$I_{edge}(x, y) = \begin{cases} 1 & \text{if } P_j(x, y) < t_{edge} I_{dilate}(x, y) \\ 0 & \text{otherwise} \end{cases}$$

Figure 6.7 shows the result of each of the three stages for determining I_{edge}.

6.3.3 Preattentive Ellipse Detection using the Hough Transform

The use of ellipses as a simple model for modularly shaped rocks has been suggested by Kemeny *et al* [45] who found a linear relationship between the screen size of a rock and its major and minor axes. In addition, since ellipses can readily be detected in a parallel manner using the Hough transform they appear to be suitable for this application. The aim of this section is to detect ellipses in I_{edge} to produce a list $L_{ellipse}$ of detected ellipses which could correspond to elliptically shaped objects which must be checked in the second level of preattentive vision. The check is necessary since it is possible for the Hough transform to produce a large output in a position which does not correspond to what would be perceived



(a) Smoothed image $I_{average}$ using local mean.

(b) Dilated image I_{dilate} to make intensity maxima as large as expected rock size.

(c) Thresholded rock image I_{edge} using 0.7 times the dilated image value.

Figure 6.7: Highlighting rock edges using knowledge from lighting model to determine non-rock points based on local area intensity maxima with $r_{expected} = 10$.

as an ellipse for this application which is only interested in ellipses which have nothing in their interiors.

The next section defines what ellipses are allowed and then the Hough transform is used to detect them after which peaks in the transformed image are identified to produce a list of ellipses corresponding to the highlighted positions.

Defining the Ellipse Model

This sections describes the process of selecting suitable ellipses for reasonably comprehensively representing rocks. Because the particular expected rock size is known, all ellipses should have a size that relates to $r_{expected}$. With respect to what? Two possibilities have been investigated: (a) Make all ellipses with equal area since that is the parameter which is closest to rock size and allows rocks detected on one level of the pyramid to be grouped with respect to size. (b) Make all ellipses with equal minor radius since they will be better matched to filters which are matched for rock size since the rock width will be constant.

Apart from the size of the rock, the ellipse has to model its shape and orientation. This can be achieved with the ellipse model by adjusting the eccentricity and orientation of the ellipse. A flexible scheme is used for determining the set of ellipses to use. The set is specified by three parameters: the number of orientations n_θ and eccentricities n_e and an elongation factor f_{aspect} to change the aspect ratio $\frac{a}{b}$ of the ellipse.

This results in a two dimensional array of ellipses $E(n_e, n_\theta)$ where E(0,0) is a circle of radius $r_{expected}$. The remaining ellipses are determined by distorting the circle using f_{aspect} and rotating it.

The major axis of the ellipse can be calculated using f_{aspect} and the index of eccentricity i as $a = r_{expected} f^i_{aspect}$. The minor axis is calculated in one of two ways depending on whether area or minor axis is being preserved. Given that the area of an ellipse is πab and the desired

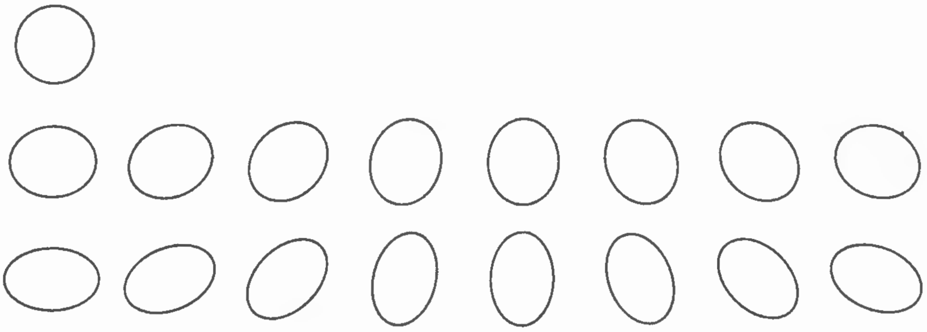


Figure 6.8: Examples of the ellipse shapes that are detected using the Hough transform operating on a rock edge image.

area is $\pi r^2_{expected}$, then in order to preserve area, the minor axis can be calculated using $b = \frac{r^2_{expected}}{a}$. If the minor axis is to be preserved then $b = r_{expected}$.

The orientation of each ellipse is incremented by n_θ steps of $\frac{\pi}{n_\theta}$. Note that it is only necessary to rotate the ellipse through a maximum of 180 degrees since the remaining ellipses are produced by symmetry.

Figure 6.8 shows 17 ideal ellipses that can be specified using this system using $n_\theta = 8$, $n_e = 3$ and $f_{aspect} = 1.1$. In practice, the relatively low $r_{expected}$ value results in a degree of quantisation but it does not significantly affect the results. For notation purposes, each ellipse is characterised by its aspect ratio $e = \frac{a}{b}$ and its orientation θ.

Hough Transform

In practice there will be N possible ellipse shapes and they will be detected in I_{edge} using the Hough Transform as described in Section 3.2.5 to result in N transform images. Figure 6.9(a) shows the resulting transform space $I_{ellipse,1.2,0}$ after using the Hough transform to detect ellipses with $e = 1.2$ and $\theta = 0$ in I_{edge} created using the intensity method (Figure 6.7(c)).

Since ellipses are only actually present where there is a high value in the transform space, the image must be clipped to give $I_{clip,e,\theta}$ before peaks can be detected. Each point in $I_{clip,e,\theta}$ is obtained using the following expression:

$$I_{clip}(x,y) = \begin{cases} 0 & \text{if } \frac{I_{ellipse}(x,y)}{c} < t_{clip} \\ I_{ellipse}(x,y) & \text{otherwise} \end{cases}$$

where t_{clip} is a clipping threshold and c is the maximum possible response of the Hough transform to an ellipse which is equal to its circumference $c = \pi(\frac{3}{2}(a+b) - \sqrt{ab})$. In practice, due to a rock having gaps in its edges and not being exactly elliptical, the Hough transform can produce a lower peak corresponding to it. It has been found by experimentation that the response is also affected by the form of preprocessing that is used to produce I_{edge}. Clearly,

(a) The Hough transform $I_{ellipse,1.2,0}$ of the rock edge image.

(b) Clip image $I_{clip,1.2,0}$ which only keeps strong ellipse responses.

(c) Detection of peaks in $I_{peak,1.2,0}$ for final point selection.

Figure 6.9: Ellipse detection using the Hough transform. Note that in (b) and (c), the images have been photometrically inverted to facilitate printing.

if rock edges are thicker, then the Hough transform is more likely to result in a higher value. It is therefore necessary to determine by experimentation exactly what clipping threshold to use. ROC curves are used in Section 6.3.4 to determine the best value for t_{clip} given the type of edge highlighting used and parameters of pyramid creation[1].

Figure 6.9(b) shows the result of clipping $I_{ellipse,1.2,0}$ at $0.6c$ (in reverse polarity).

Peak Detection

In order to further reduce the search space required for the second level of preattentive vision and hence attention focusing, only peaks or local maxima in I_{clip} are accepted. Due to the integer nature of the Hough transform incrementing bins, it is possible for several adjacent bins to accumulate the same value thus resulting in local maxima in the form of plateaus.

This section describes a peak detection algorithm which overcomes this problem by operating in a manner similar to a thinning algorithm [77]. All pixels are checked from the cardinal directions to check that they are not local maxima. This is true if a point adjacent to it in the current direction is less than it and the three eight-connected neighbours in the opposite direction are also less than or equal to it. If the adjacent point in the current direction is less than it and the opposite points are greater than or equal to it, the point cannot be a local maximum and its value is decreased. The whole image is iteratively scanned from the different directions until no points are decreased in four successive scans.

The result is that plateaus are reduced to a skeleton of points instead of all the points on plateau. The resultant points are central relative to the shape, i.e. a disc will reduce to one point in its centre, whereas two overlapping discs should reduce to the centre points of the

---

[1] Note that it is important to err on the side of leniency with the selection of t_{clip} since rejecting a valid rock at this stage is irrevocable, but false alarms can be detected and removed later.

An algorithm for determining the position of peaks in an image by finding the centres of plateaus if they exist. Note that $\mathbf{o}$ is an array of direction vectors such that

$$\mathbf{o} = (\mathbf{o}_1, \ldots, \mathbf{o}_9) = (NW, N, NE, E, SE, S, SW, W, NW)$$

```
Function:  Peak_Detect( $I$)
    define $\mathbf{o}_x$ = (-1,0,1,1,1,0,-1,-1,-1)
    define $\mathbf{o}_y$ = (-1,-1,-1,0,1,1,1,0,-1)
    define $N = 2, E = 4, S = 6, W = 8$
    $T_1 = T_2 = I$
    repeat
        $count = 0$
        for( $i = N, E, S, W$)
            for( all image points($\mathbf{p} : \mathbf{p} \neq$ boundary point))
                $T_2(\mathbf{p}) = t = T_1(\mathbf{p})$
                $n_1 = T_1(\mathbf{p} + \mathbf{o}(i))$
                $n_2 = T_1(\mathbf{p} - \mathbf{o}(i))$
                $n_3 = T_1(\mathbf{p} - \mathbf{o}(i + 1))$
                $n_4 = T_1(\mathbf{p} - \mathbf{o}(i - 1))$
                if( $n_1 < t.(n_2 \geq t + n_3 \geq t + n_4 \geq t)$)[a]
                    $T_2(\mathbf{p}) = n_1$
                    $count = count + 1$
            $T_1 = T_2$
    until( $count = 0$)
    return $T_1$
```

---

[a] Note that + is the logical OR operator and . is AND.

Figure 6.10: The peak detection algorithm.

two discs[2]. Figure 6.10 shows the peak detection function in algorithmic form. The output after peak detection of $I_{clip,e,\theta}$ is $I_{peak,e,\theta}$ and is shown in Figure 6.9(c). Note that detected points are shown as grey corresponding to their Hough transform value while the remainder of the image has been set to white for clarity.

Possible Ellipse list for Refinement

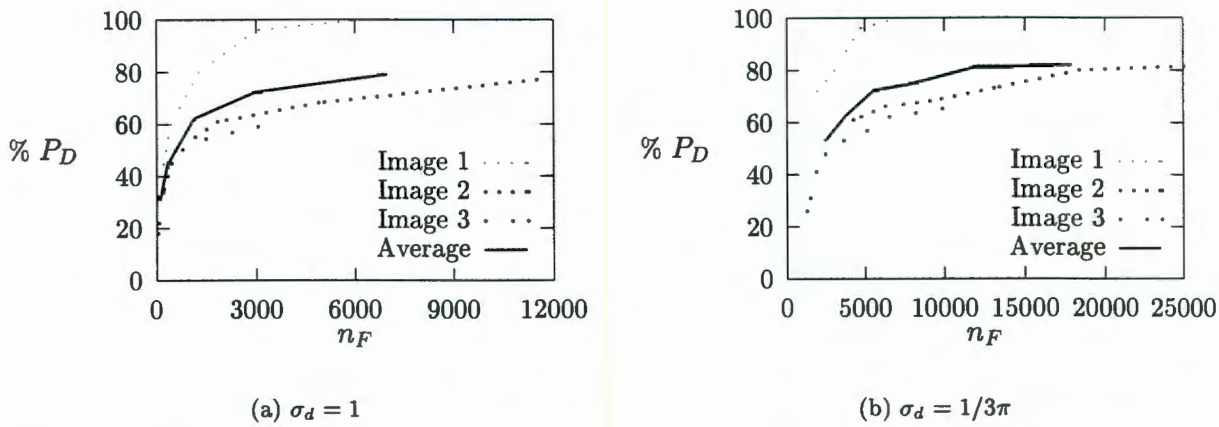
The final task of the first level of preattentive vision is to compile a list of highlighted points in the image to act as input to the second level of preattentive vision for further refinement. This is achieved by finding all the peaks in each of the $I_{peak,e,\theta}$ images and compiling them into a list $L_{ellipse} = (L_1, \ldots, L_n)$.

Each element of $L_{ellipse}$ is a structure containing the following information: $L_j = (\mathbf{p}, e, \theta, h)$ which is required for attention focusing and hierarchical analysis.

$\mathbf{p}$ = The coordinates of the centre of the ellipse

---

[2] It is possible that due to quantisation, a shape may result in more peaks than expected



(a) $\sigma_d = 1$

(b) $\sigma_d = 1/3\pi$

Figure 6.11: ROC curves obtained when varying t_{morph} in steps of 2 from 14–22 for $\sigma_d = 1$ and 28–38 for $\sigma_d = 1/3\pi$ with $t_{clip} = 0.8$. A high value of t_{morph} corresponds to a low % P_D.

e = The aspect ratio of the ellipse from which a and b can be determined

θ = Ellipse orientation

h = The grey level peak value of the Hough transform normalised by the circumference

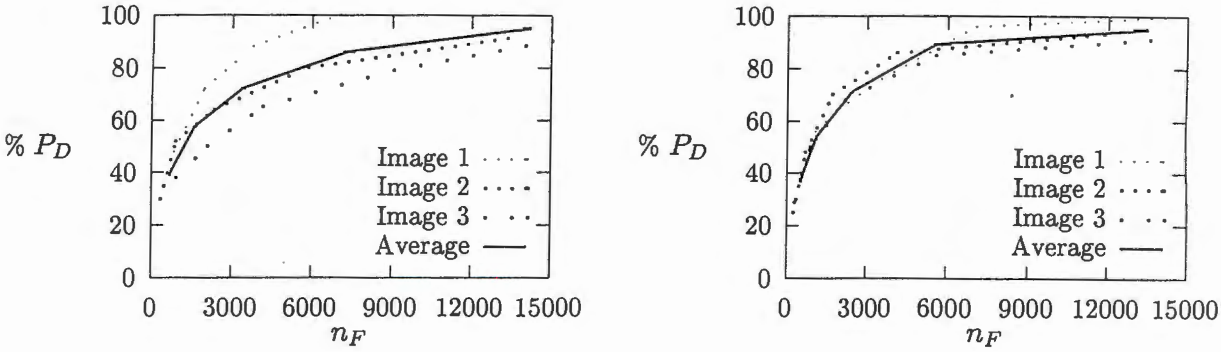
6.3.4 Review of Edge Detection Goals

The goal of the preattentive vision stage is to ensure that all rocks are highlighted at least once, because any rock missed here will not be found later. At the same time, due to the computational burden and the chance for mistakes imposed by false hits, the number of false hits should be minimised.

ROC curves are determined here for the purpose of selecting good values for t_{morph}, t_{edge} and t_{clip} which affect the performance of this level of the preattentive vision stage. The performance is measured using the procedure described in Section 2.3.5 using all the points detected by this level of preattentive vision and disregarding the classification and hierarchical analysis stages.

Normal edge detection such as the morphological method are inadequate due to the presence of spurious edges corresponding to internal edges, or adaption to the background resulting in noise. Figure 6.11 shows an ROC curve in which t_{morph} is adjusted for both forms of the image pyramid. It can be seen that for the three images, there is significant difference between the effectiveness of this type of edge highlighting, making it unreliable and on average, not particularly accurate.

The poor performance of the gradient approach resulted in the development of a new method which finds thick edges around rocks, with minimal response to texture and the background, by using intensity information determined from the lighting model. Figure 6.12



(a) $\sigma_d = 1$

(b) $\sigma_d = 1/3\pi$

Figure 6.12: ROC curves obtained when varying t_{edge} in steps of 0.05 from 0.6–0.8 with $t_{clip} = 0.6$. A high value of t_{edge} corresponds to a high % P_D.

t_{clip}	$\%P_D$	n_F
0.75	51	476
0.7	69	1232
0.65	81	2949
0.6	89	5563
0.55	95	11219

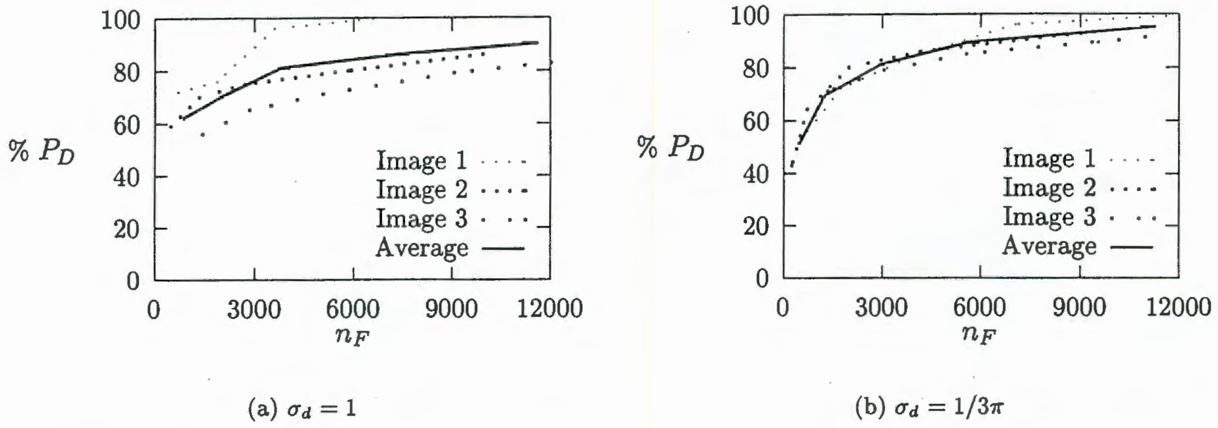
Table 6.1: Data for the average ROC curve for the case when $\sigma_d = 1/3\pi$ and $t_{edge} = 0.75$.

shows ROC curves in which t_{edge} is adjusted. It can be seen that especially for the un-bandlimited image pyramid, there is very similar response for each of the test images. This indicates that the intensity method of edge highlighting is more reliable than the gradient method.

Figure 6.13 shows the result of varying the clipping threshold t_{clip} when the intensity edge highlighting method is used with $t_{edge} = 0.75$. It can be seen again that when $\sigma_d = \frac{1}{3\pi}$ the results are very consistent between the three test images.

A comparison of the average ROC curves for each of the situations in Figures 6.11, 6.12 and 6.13 is given in Figure 6.14. From this graph, it can be seen that the first level of the preattentive stage provides the best performance when the intensity method is used and the image is un-bandlimited. Since true hits must be maximised and reasonable processing time can be achieved for about 10000 total hits, the following thresholds would appear to be suitable: $t_{edge} = 0.75$ and $t_{clip} = 0.6$ or 0.55 (See Table 6.1).

Ellipses must be detected in either of these edge detected images. Thick edges increase the response of the Hough transform but also increases chance of falsely detecting ellipses where the edge is very thick around large rocks. It may therefore be better to thin the edges



(a) $\sigma_d = 1$ (b) $\sigma_d = 1/3\pi$

Figure 6.13: ROC obtained when varying t_{clip} in steps of 0.05 from 0.55–0.75 with $t_{edge} = 0.75$. A high value of t_{clip} corresponds to a low % P_D.

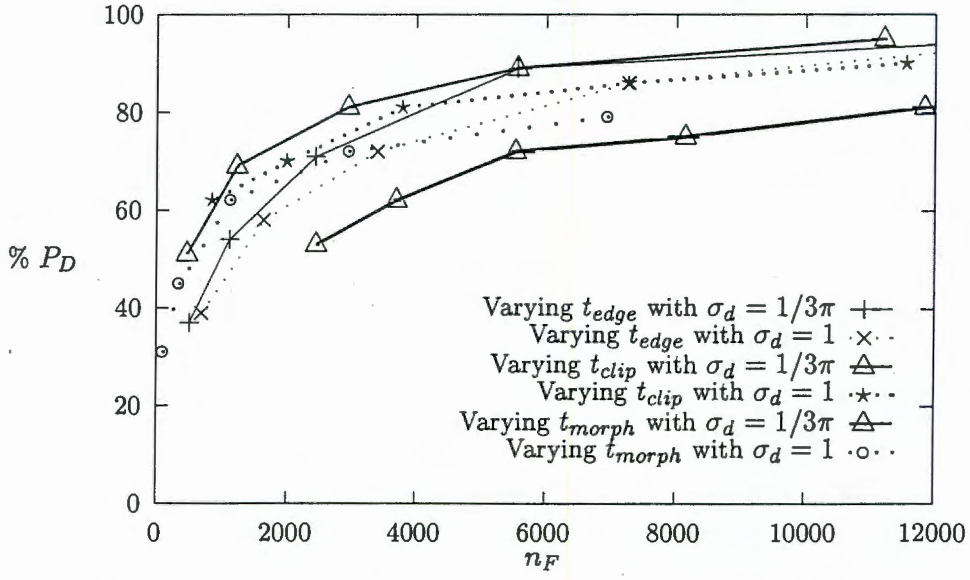


Figure 6.14: This graph shows the mean ROC curves for all six cases of σ_d with t_{morph}, t_{clip} and t_{edge}. From this graph, the optimum value for t_{edge} can be seen to be 0.75 since for all values other than the intersection point, it is below the t_{clip} curve for both cases of σ_d. The selection of t_{clip} depends on the relative costs of true and false hits. It can also be seen that morphological edge detection does not produce as good results as the adaptive thresholding technique, and that $\sigma_d = 1/3\pi$ gives better results than $\sigma_d = 1$.

and use a more sophisticated form the of the Hough transform for a decrease in the number or proportion of false hits.

An alternative approach to achieving this is presented in Section 7.6 which performs a comparison of the ellipse with the actual content of the binary image in order to detect these undesirable situations and thereby reduce the proportion of false hits produced by the preattentive vision stage.

6.4 Summary

A procedure for creating the image pyramid, P, using Gaussian smoothing and an arbitrary scale factor between levels was described in Section 6.2 with the goal of producing similar visual characteristics for all levels of the pyramid, $(P_1, \ldots, P_n)$.

Subsequently, the first stage of image analysis is preattentive vision. Given P, each level P_j is investigated using a parallel operation for the whole image to detect positions where elliptical objects might be present. This is achieved by first highlighting edges to form a binary image I_{edge} and then using the Hough transform to detect ellipses in that image. The output of this process is a list of possible rock positions, $L_{ellipse}$ which must be further refined to remove false hits in the following chapter before being passed on to attention focusing to decide if a rock is present or not at each remaining ellipse.

Chapter 7

Macroscopic Blob Segmentation

7.1 Introduction

A macroscopic segmentation algorithm is described in this chapter which is used for three purposes. Initially, a simplified form of the algorithm will be used to refine the list of ellipses $L_{ellipse}$ produced by the first level of preattentive vision to produce a list $L_{highlight}$ of highlighted ellipses or target positions in the image which predict the positions of rocks. These must be further investigated in attention focusing. The complete macroscopic blob segmentation algorithm is then used as the first stage of the attention focusing process. Its aim there is to analyse each area of the image that corresponds to a predicted rock in order to identify and label the object at that point. The third use is to determine the extent of detected rocks once their position, size and shape are known, for the purpose of calculating the size distribution of the rocks accurately.

Although the first use of the segmentation algorithm is as a refinement level for the preattentive stage, it is only a simple variation of the algorithm used by the attention focusing process. For this reason, the complete segmentation algorithm will initially be described then the simplified form will be introduced and its application as the second level of the preattentive vision stage will be described.

In order to achieve the goal of segmenting the image, the algorithm uses knowledge of the expected rock — its size, position and shape — which is encoded by the ellipse which was detected by the first level of the preattentive vision stage. Note that since a valid rock may not actually be present in a region of interest, whatever is there will be referred to as a blob.

An important goal in the design of this segmentation algorithm is to ensure that if a rock matching the predicted size and shape *is* present, the labelled blob will accurately correspond to it.

According to Haralick and Shapiro [32], an *image segmentation* is a partitioning of an image into a set of non-overlapping regions whose union is the entire image. The purpose of image segmentation is to decompose an image into parts that are meaningful with respect to a particular application, while not being concerned with what the regions represent [62]. In this application, segmentation is performed to separate a two-dimensional blob from its background in a region of interest.

It is very difficult to describe what actually constitutes a meaningful segmentation for the purpose of writing a computer program. It is therefore common for segmentation procedures to obey the following rules [32]:

1. Regions of the segmentation should be uniform and homogeneous with regard to some characteristic, such as grey level or texture.

2. Region interiors should be simple and without many small holes.

3. Adjacent regions should be significantly different with respect to the characteristics on which they are uniform.

4. Boundaries of each region should be simple and spatially accurate.

Several examples of segmentation based on thresholding and region growing which observe these rules are described by Niblack [62], Haralick and Shapiro [32], and Rosenfeld and Kak [77].

Segmentation has previously been used in various forms for the purpose of delineating particles with varying degrees of success (see Section 2.2.1). The previous approaches can be divided into three basic groups: finding edges in the image [63, 93], performing region merging operations [5, 50], and using thresholding algorithms [56, 96]. It was found in Section 2.2.5 that these segmentation algorithms had some shortcomings which could be attributed to not considering the following approaches: using knowledge obtained from analysis of illumination; incorporating shape information in the form of a predicted shape model; and by restricting detection to only one size of object.

A particularly difficult problem which must be overcome by the segmentation algorithm, and which is not handled particularly well by the previous methods, is that of regions of low contrast. It has been found in medical image analysis that common segmentation methods such as edge following and region growing are inadequate for soft tissue objects where the contrast is low [9]. These problems can be overcome by incorporating knowledge of the problem domain which can be used to compensate for image ambiguities. According to Brinkley [9], one of the more important kinds of knowledge is spatial knowledge of anatomic shapes. He has used a radial contour model (RCM) which is a type of geometric constraint network to model anatomical shapes. This model encodes spatial knowledge about anatomic shapes and is used to guide segmentation by imposing constraints on the resulting region. An advantage of a shape constraint is that it allows the shape to be predicted where edges are not found.

There are two significant differences between the segmentation of rocks and organs by Brinkley. Firstly, it is not certain whether or not a rock is in fact present and secondly, the shape of a rock is not as characteristic as an anatomical shape.

Based on observations of the shortcomings of the low level segmentation methods that have been applied previously to the rock fragmentation problem, and the advantages that can be obtained by incorporating knowledge of the problem domain to guide segmentation in regions of low contrast, a segmentation algorithm has been developed which incorporates knowledge from two sources: Firstly, the size, position and shape of a rock is predicted in each element L_j of $L_{ellipse}$ and it is initially assumed that a rock *is* present in the form of an ellipse with known e and θ values. Secondly, knowledge concerning the appearance of rocks is used to determine rules to guide the segmentation process. Segmentation will be used to determine the extent of the blob that is present by labelling its interior and exterior in the region of interest.

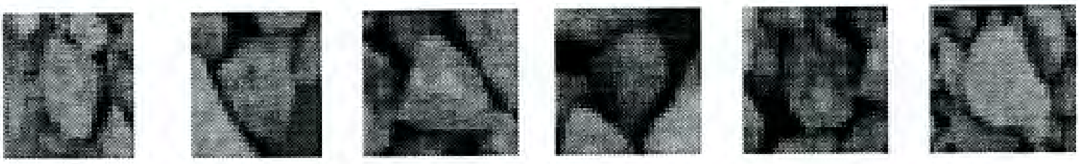
The actual region of interest R that is to be segmented is centred at $\mathbf{p}_j$ which is the centre of the ellipse identified in L_j. It is necessary to make the region of interest larger than the size of the predicted rock that it contains. Doing this provides some contextual information which is important for performing accurate segmentation. Specifically, since the elliptical model of the rock may not be completely accurate, some edge points of the rock may occur outside the ellipse. If these edge points are not contained in R, then they cannot be detected and the segmented region will be inaccurate.

In practice, the size of the region of interest has been calculated to be 1.5 times the size of the bounding box of the ellipse defined by L_j. This value has consistently been sufficient to contain all rocks being segmented. Due to the use of a mask which specifies possible edge positions, there would be no benefit in increasing the size of R. Reducing R would result in decreased processing time, but the segmentation of badly matched rocks could be compromised.

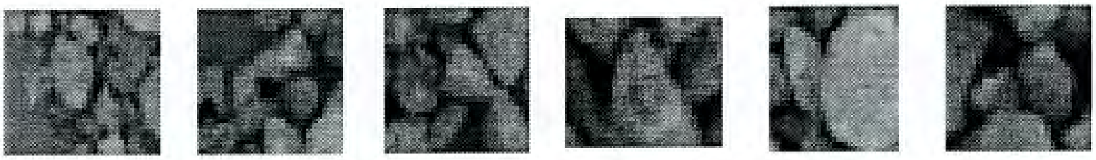
The segmentation algorithm has been called a macroscopic segmentation algorithm because it attempts to perform the labelling of a region of interest in a global manner which is consistent with respect to the rock model. Furthermore, the segmentation algorithm is supposed to operate only on a single whole rock, while incorporating information from three sources: (1) the image of the region of interest in the form of edge strength, (2) the preattentive stage in the form of an expected shape and size, and (3) knowledge about the appearance of rocks in the form of rules.

Sample regions of interest which will be used to demonstrate the segmentation process throughout this chapter are given in Figure 7.1. Note that half of the regions of interest do contain a rock and the remainder do not.

Section 7.2 describes how knowledge about rocks is encoded as rules which are then used in a three stage algorithm for segmenting the region of interest. The three stages are given



(a) Regions of interest containing suitable rocks.



(b) Regions of interest not containing suitable rocks.

Figure 7.1: Sample regions of interest for segmentation.

in Sections 7.3, 7.4 and 7.5. Section 7.6 describes how the first stage of the segmentation algorithm can be modified to refine the selection of target ellipses by the preattentive vision stage in a second level.

7.2 Modelling Rocks

Based on the assumption that a correctly sized rock is present in the region of interest (i.e. its size at the current level of the image pyramid is equal to $r_{expected}$), it is necessary to ensure that the segmentation process labels the blob that corresponds to the rock. For this reason, several characteristics of the rock are proposed:

1. The rock will have a known size and shape that is predicted by the preattentive vision process (See Figure 7.2). Initially, only circular rocks were detected but the process was extended to detect elliptical rocks as well, providing increased flexibility. It is possible to approximate the rocks using ellipses in this context since the ellipse only forms a guide to the segmentation which allows some flexibility in the shape of the final blob. It is therefore possible to correctly segment a rock which is not exactly modelled by the ellipse.

2. A rock is lighter than its immediate background due to the lighting conditions which cast a shadow around objects (See Section 5.2.3).

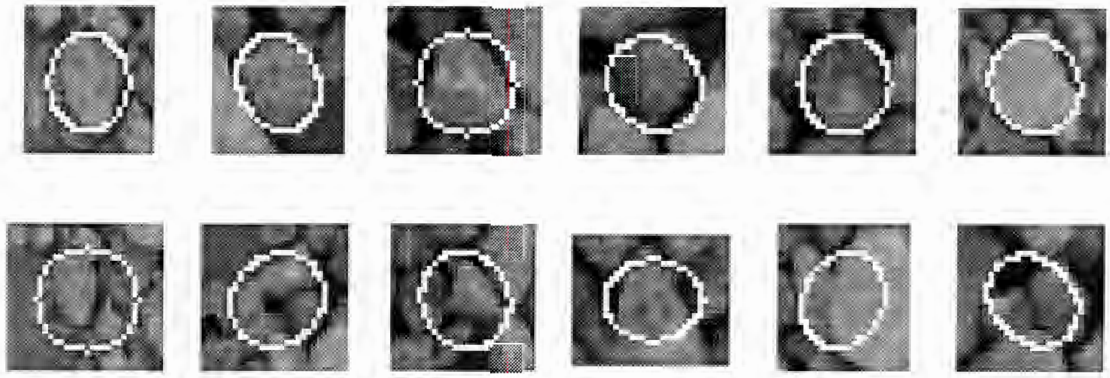


Figure 7.2: The expected rock size and shape that was found during preattentive vision is indicated using an ellipse.

3. A rock is a physical object and must therefore have a continuous boundary.
4. The rock will have a simply connected interior (the rock cannot have any holes within itself).

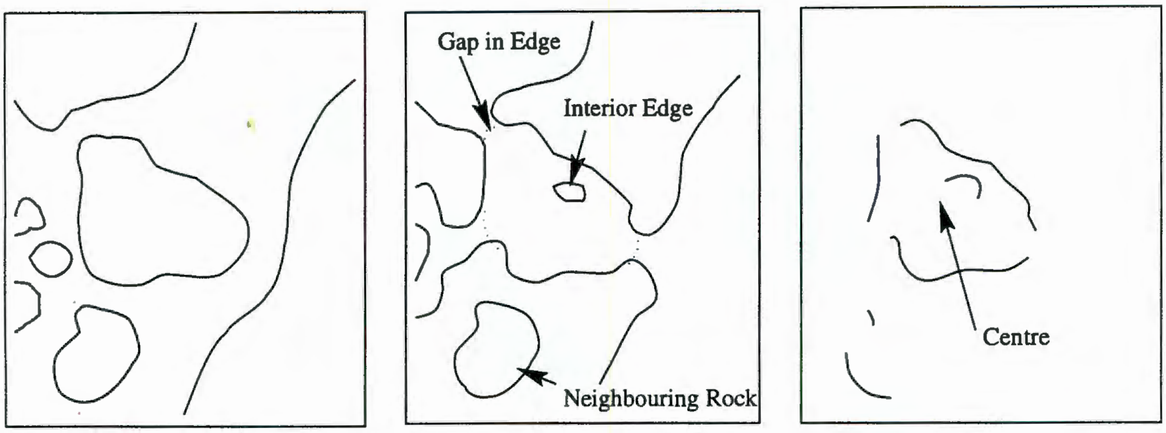
A segmentation algorithm was developed which encodes the knowledge contained in the four rules just given. The algorithm is split into 3 sections:

1. Determine blob edges while observing rules 1, and 2.
2. Produce a closed and continuous boundary to satisfy rule 3.
3. Fill the interior of the boundary to produce a final labelled region which satisfies rule 4.

These three steps are presented in Sections 7.3, 7.4 and 7.5 respectively.

7.3 Determining Blob Edges

An edge strength image is used for determining the position of blob edges. Figure 7.3 shows ideal, typical and directional edge detection diagrams. With ideal edge detection, each rock is disconnected from each neighbour and there are no spurious edges within the rocks. In a typical edge image gaps appear in the rock boundaries where adjacent rocks have merged. Internal edges appear within rocks due to surface marking and close neighbouring rock edges can confuse the detection of rock boundaries. A directional edge detector was described in Section 3.2.1 which only responds to edges of a particular orientation. Given the centre of the object of interest, many of the spurious edges are not present in the directional edge



(a) Ideal Edge (b) Typical Edge (c) Directional Edge

Figure 7.3: Ideal, typical and directional edge images. (a) represents ideal edges from a rock image. (b) shows typical contamination that is usually present within an edge detected image. (c) shows the result of using directional edge detection given the centre point of the rock of interest.

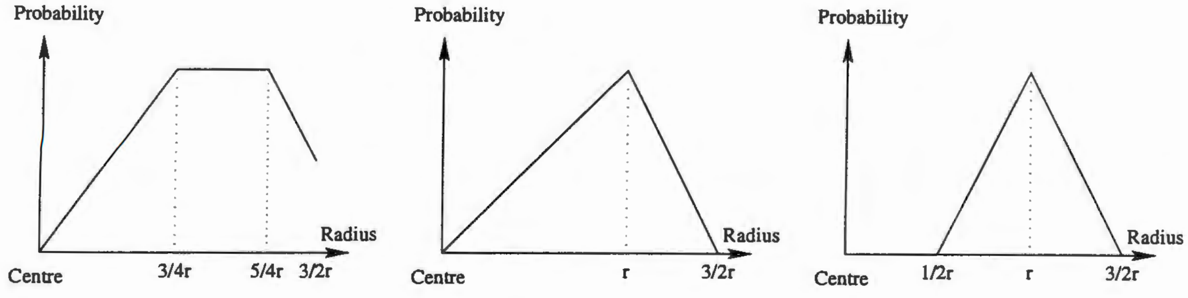
detected image (when edges are detected perpendicularly to radial lines) thus simplifying the segmentation process dramatically.

In order to adequately find blob edges while observing the first two rules implied by the rock model, the following two techniques have been combined.

- The expected shape and size of the rock is encoded in a probability mask. This mask image will be used to bias the detection of edge points to those of an optimally sized and shaped rock which best matches the underlying shape in the probability mask.

- An edge direction mask corresponding to the the expected shape is created. The direction mask encodes the direction in which edges should occur at each point within the region of interest if a correctly centred rock is present. Since rocks will be surrounded by a dark band, intensity will decrease most rapidly perpendicular to the direction of the rock edge.

The probability mask is described in Section 7.3.1 and the generation of the direction mask in Section 7.3.2. Detecting blob edge points involves combining edge probability and direction information to determine the most likely edge points. Grey level edge detection based on the algorithms described in Section 3.2.1 has been applied and its use is described in Section 7.3.3.



(a) Poor edge localisation (b) Medium edge localisation (c) Good edge localisation

Figure 7.4: Probability profiles designed to detect edges at the optimal radius, r = $r_{expected}$, from the centre of the region of interest.

7.3.1 Edge Probability Mask

The edge probability mask is used to encode size and shape information to facilitate detection of the expected rock. A separate mask is used for each possible shape (this includes each orientation and eccentricity for ellipses). The use of a shape mask increases the flexibility of the segmentation algorithm. This is achieved in two ways: the probability mask encodes the shape of the object to be detected, and it specifies a region in which the edge points may be found.

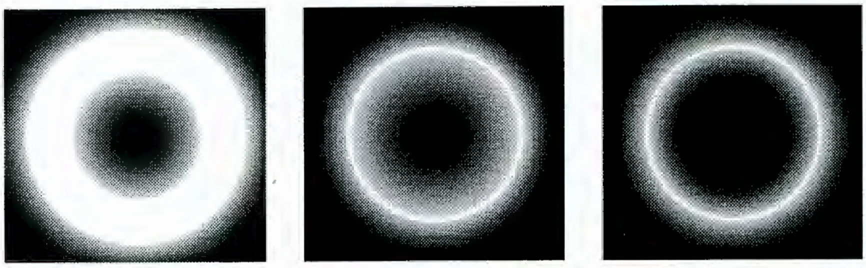
The probability mask is generated using a probability profile, $p(r)$. Profiles are designed to start at the centre of the mask and scan outwards along radial lines. The profile must be scaled so that the point of maximum probability (r) falls on the boundary of the shape of the mask. Several attempts have been made at designing probability profiles. The probability should have a maximum value at the expected edge point and decrease inwards and outwards from there. Three simple variations of this scheme have been implemented and are shown in Figure 7.4[1]. Any possible scheme could be implemented here such as a Gaussian profile, but considering the low resolution the difference would be minimal and these three capture the various stages of edge localisation adequately.

Initially, when rock shape was modelled with a circle, the probability mask was easily created by simply rotating the profile around the origin of the mask. Any point (x, y) is a distance $d = \sqrt{(x - x_0)^2 + (y - y_0)^2}$ from the centre (x_0, y_0) of the mask and is thus assigned a value $p(d)$. Figure 7.5 shows the three circular probability masks created using each of the probability profiles.

With the development of elliptical shapes came an increased complexity in the creation of the probability mask. The following equation which is derived in Appendix C.2 can be used

---

[1] Note the similarity between these probability masks and fuzzy logic membership sets. This could indicate a possible path to take for further developing the segmentation algorithm.

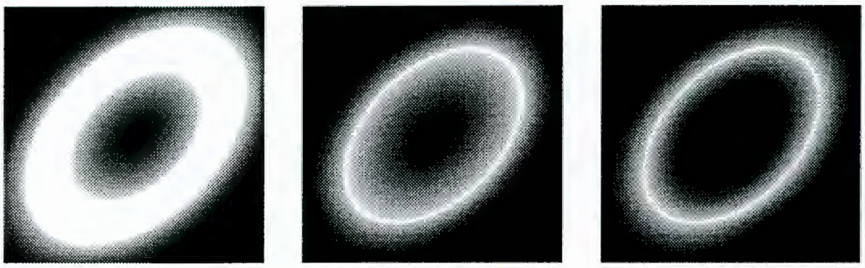


(a) Poor edge localisation

(b) Medium edge localisation

(c) Good edge localisation

Figure 7.5: Circular probability masks.



(a) Poor edge localisation

(b) Medium edge localisation

(c) Good edge localisation

Figure 7.6: Elliptical probability masks.

to calculate the distance from the origin along a line of orientation ψ to the ellipse as

$$r_\phi = \sqrt{(a\cos\theta\cos\phi + b\sin\theta\sin\phi)^2 + (b\cos\theta\sin\phi - a\sin\theta\cos\phi)^2} \tag{7.1}$$

where ϕ is calculated using

$$\phi = \arctan\left(\frac{a\sin\theta\cos\psi + a\cos\theta\sin\psi}{b\cos\theta\cos\psi - b\sin\theta\sin\psi}\right) \tag{7.2}$$

which is derived in Appendix C.4.

For each point in the probability mask, the orientation can be calculated as $\psi = \arctan(\frac{y-y_0}{x-x_0})$. The distance from the origin to the point is d as before. In order to scale the profile, the probability at the point (x, y) must be set at $p(r\frac{d}{r_\phi})$. Figure 7.6 shows elliptical probability masks created using the three profiles.

With only circular shapes allowed, it was necessary for the profile to allow for a reasonable amount of deviation from the expected shape in order to correctly segment non-circular rocks.

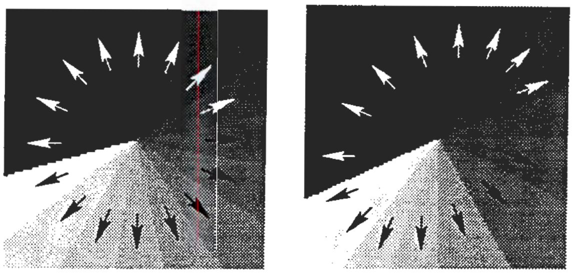

(a) Circular mask (b) Elliptical mask

Figure 7.7: Direction masks for circular and elliptical shapes.

For this reason, the first profile was most successful with circular masks. An ellipse can better model the shape of a rock and it is no longer desirable to allow a large degree of freedom in the profile as the edge is expected to remain close to the ellipse in all directions. The second and third profiles should be used when the actual rock shape can be modelled more accurately due to better localisation of blob edges.

The following section explains how the edge direction mask which was introduced in Section 3.2.1 is generated given an elliptical shape model.

7.3.2 Edge Direction Mask

The role of an edge direction mask is to encode the orientation at which edges must be detected for each point in an image using a directional edge detector. This is achieved by encoding the normal direction to the boundary of the shape that is encoded by the probability mask. Note that the direction will be the same for all points on a radial line. The following equation which determines the normal direction to the boundary of an ellipse of parameters a, b and θ given a parameter value of ϕ is derived in Appendix C.3:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} b\cos\theta\cos\phi + a\sin\theta\sin\phi \\ -b\sin\theta\cos\phi + a\cos\theta\sin\phi \end{pmatrix} \tag{7.3}$$

where x and y are the horizontal and vertical components of a vector normal to the ellipse. Given that the orientation of the radial line is ψ, ϕ can be calculated using (7.2).

Figure 7.7 shows the direction masks for the circular and elliptical shapes used for the probability masks. The direction masks are stored using direction numbers which increase in a clockwise direction. Sixteen possible edge directions are currently supported and are indicated with arrows on the direction mask. Note that the direction mask for the circular shape has equal quantities for each direction while the elliptical shape has flatter sides and therefore larger regions in those directions. The value at each point (x, y) in the direction

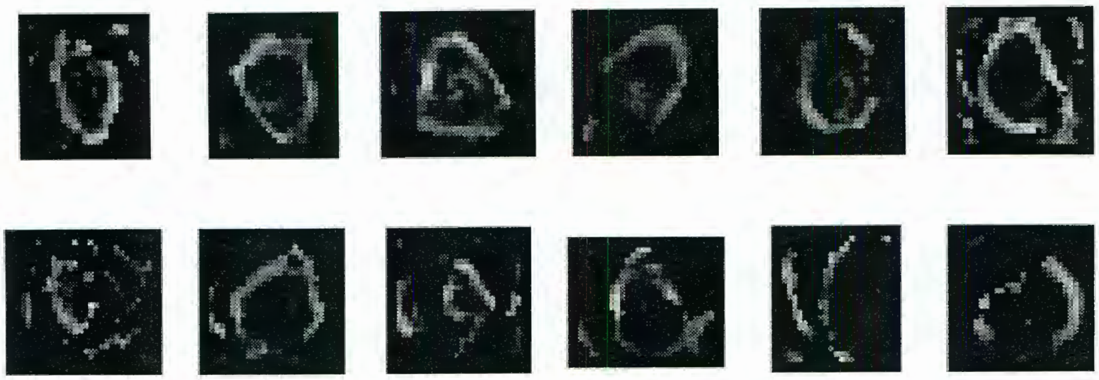


Figure 7.8: Directional gradient detection based on the expected rock shape direction mask.

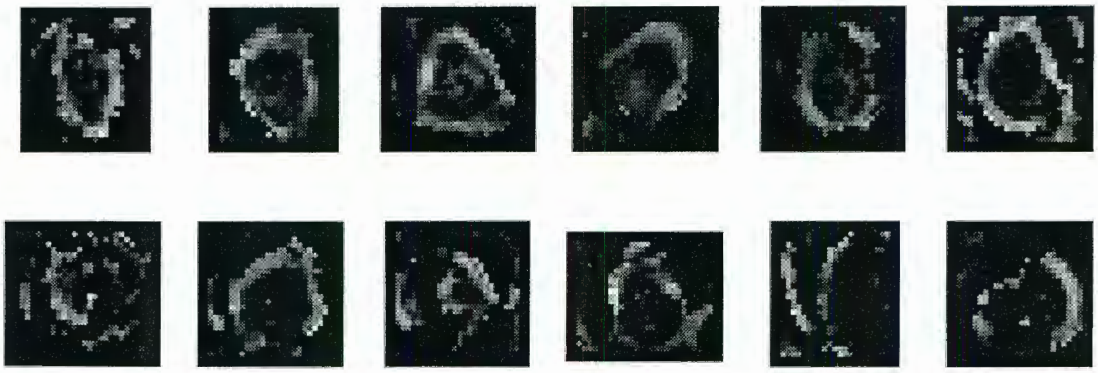


Figure 7.9: Directional valley detection superimposed on gradient images in Figure 7.8.

mask is calculated by determining the angle $\psi = \arctan\left(\frac{y}{x}\right)$, calculating ϕ, and substituting that into (7.3). The angle of the normal is then calculated and quantised to the nearest allowable value for edge direction.

The result of applying the directional gradient detection operator described in Section 3.2.1 to each of the sample regions of interest using a direction mask derived from the expected rock shape is shown in Figure 7.8. It was found that when the valley detection operator was applied using the same direction mask, very few valley points were found. From these results, it was concluded that the condition requiring three adjacent valley points is seldom satisfied within the rock images and is therefore too stringent a condition for this type of image. Relaxing the condition resulted in a proliferation of edge points corresponding to surface texture markings. It was therefore felt that in either situation, the valley operator on its own did not provide suitable detection performance and so if it is used, it could only provide reinforcement for the gradient operator. This can be achieved by summing the two components and is shown in Figure 7.9.

The next section describes a method based on directional greyscale edge detection for finding blob edge points in the region of interest given a probability mask.

7.3.3 Detecting Blob Edge Points in the Region of Interest

The greyscale edge detection stage is used to create a circular signal[2] S_{edge} representing the distances from the centre of the region of interest to blob edge points for all radial directions, and their associated edge strength. The number of points in S_{edge} is equal to the circumference of the expected shape at the given size to provide a sample space of approximately 1 pixel for the edge profile. The direction for each point is found by incrementing the parameter ϕ from 0 to 2π, calculating the respective point on the ellipse using

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a\cos\theta\cos\phi + b\sin\theta\sin\phi \\ b\cos\theta\sin\phi - a\sin\theta\cos\phi \end{pmatrix}$$

which is derived in Appendix C.1, and determining its angle to the origin. S_{edge} is determined by performing directional edge detection (using either the gradient or gradient and valley detector described in Section 3.2.1) and then selecting the best edge in each direction based on the edge strength and probability values.

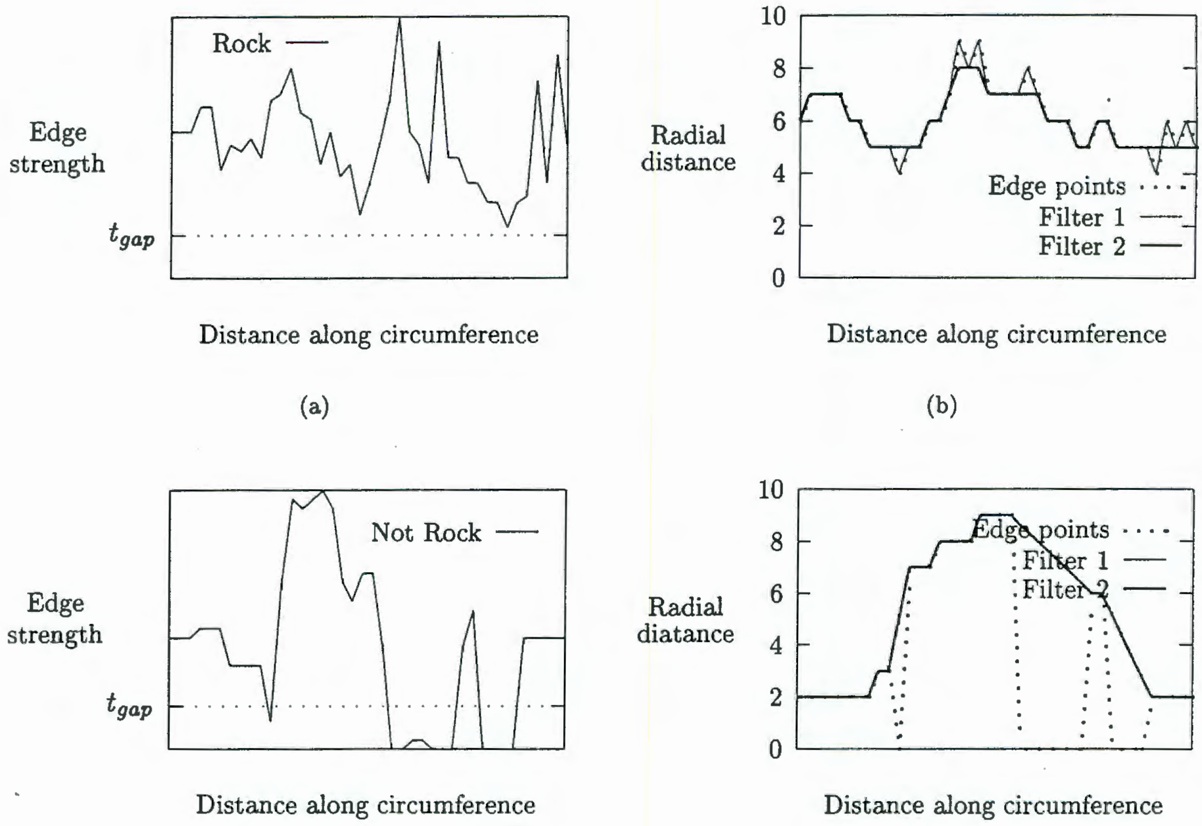
For any given direction, a profile of edge strength $e(a)$ can be obtained by scanning along a radial line in that direction in the edge detected image. Likewise, a profile of edge probability $p(d)$ can be obtained from the probability mask. The entry in the edge signal for the distance to the edge point is then determined as $S_{edge}(\theta) = d : e(d)p(d) > e(a)p(a), \forall a \neq d, 0 < a < \frac{3}{2}r, 0 < d < \frac{3}{2}r$ where r is the expected distance to the edge for the given direction. The associated edge strength is $e(d)$. Gaps can occur in the edge signal. This may occur if there is no point a where $e(a)p(a) > 0$. A gap in the edge signal is indicated by a zero value for $S_{edge}(\theta)$.

A self-normalising edge-strength filter is then used to remove very weak edge points from the edge signal using the following method: The maximum edge strength detected in the radial scans is recorded and if any point in the signal has an edge strength less than t_{gap} of that value, it is removed from the signal to result in a gap. This method can adapt to the degree of contrast in the image and inspection of edge strength profiles in Figure 7.10(a) and (c), shows that a value of $t_{gap} = \frac{1}{6}$ would not remove any significant edges. It is desirable to convert weak edges to gaps since it provides an easy-to-measure feature which is useful for discriminating between regions of interest containing a suitable rock and those that do not.

The process for generating the list of edge points is described in algorithmic form in Figure 7.11.

---

[2] A circular signal is defined to be a signal which wraps around at each end.



(a) (b)

(c) (d)

Figure 7.10: Graphs of edge and radius profiles from S_{edge} and $S_{boundary}$. (a) and (c) show the edge strength profiles for the last region of interest in Figures 7.1(a) and (b) respectively. A rock is present in (a) and not in (b). (b) and (d) show the corresponding distance profiles after edge detection, and boundary closing with both filtering processes.

From the results of this process (See Figure 7.12) it appears that the use of directional greyscale edge detection and the probability mask can produce rock edges with good localisation under diffuse lighting conditions. Some simplification of this algorithm could be achieved by using binary edge detection whereby the region of interest is thresholded at a value based on a fraction of the measured maximum intensity of the rock. This method could be applicable if surface intensity fell off quickly near rock edges as predicted by the lighting model in Chapter 5, but due to the complexity of the situation and observation of performance, a greyscale edge method is more generally applicable and robust because edges are found close to the point of steepest gradient and not where the intensity passes through a particular value. It is, however, possible to use an approximation like this to improve the preattentive vision stage and this is discussed in Section 7.6.

Given a region of interest (ROI), $R = (I, G, P, S, \mathbf{c})$ where

$$\begin{aligned} I &= \text{Grey level image of ROI} \\ G &= \text{Edge detected image of ROI} \\ P &= \text{Probability mask encoding expected shape} \\ S = (S_1, \ldots, S_n) &= \text{Points on the perimeter of expected shape} \\ \mathbf{c} &= \text{Centre point of ROI} \end{aligned}$$

then a list of the radial distances to the edge points and their strengths is denoted as $S_{edge} = (r, e)$ where

$$\begin{aligned} r = (r_1, \ldots, r_n) &= \text{List of radial distances} \\ e = (e_1, \ldots, e_n) &= \text{List of edge strengths} \end{aligned}$$

and are calculated using the following function:

Function: Find_Edges(R)

```
for($i = 1, 2, \ldots, n$)
    $\phi = \arg(S_i)$
    $radialmax = 0$
    for($j = 2, 3, \ldots, \frac{3}{2}|S_i|$)
        $x = \mathbf{c}_x + j \cos \phi$
        $y = \mathbf{c}_y + j \sin \phi$
        if($G(x, y) \times P(x, y) > radialmax$)
            $radialmax = G(x, y) \times P(x, y)$
            $r_i = j$
            $e_i = G(x, y)$
$t_{gap} = \frac{1}{6} \max\{e_1, \ldots, e_n\}$
for($i = 1, 2, \ldots, n$)
    if($e_i < t_{gap}$)
        $r_i = 0$
return $S_{edge}$
```

Figure 7.11: The blob edge finding algorithm.

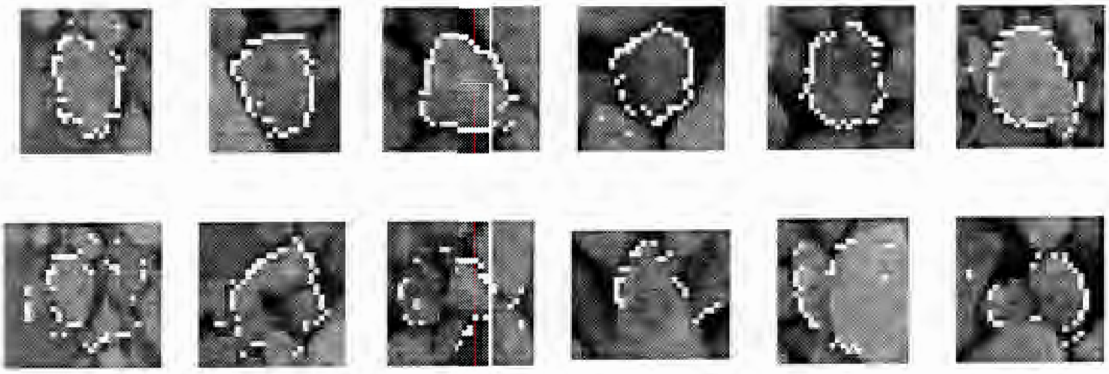


Figure 7.12: Labelled edge points in the regions of interest.

The resultant list of edge points is stored as a circular signal S_{edge} which must be processed to produce a complete blob boundary. This processing stage is described in the following section.

7.4 Closing Blob Boundaries

The aim of the second stage in the segmentation algorithm is to produce a continuous boundary by filtering the edge signal S_{edge} found in Section 7.3.3. The edge signal stored in S_{edge} contains two potential problem features: (1) gaps may be present where no edge point was detected and (2) some edge points may be detected where there are gaps in the correct edge and another edge is present within the region of non-zero probability.

Two procedures are proposed to tackle these problems. Firstly, an outlier removal algorithm is applied to the signal in an attempt to remove invalid edge points and secondly, gaps in the signal (where no edge points were detected) are filled by means of interpolation.

A signal $S_{boundary}$ which represents a continuous boundary is produced by filtering the edge signal to remove outliers using one of two filters, both of which operate on a three point moving window where the three points are i, j and k. The operation of the two filters are described in the following two points:

1. The first filter attempts to remove outliers using a thresholding approach:

$$S_{boundary}(j) = \begin{cases} 0 & |S_{edge}(i) - S_{edge}(j)| > t_{outlier}, |S_{edge}(j) - S_{edge}(k)| > t_{outlier} \\ S_{edge}(j) & \text{otherwise} \end{cases}$$

   The threshold value that is used is $t_{outlier} = \frac{1}{3}r$ where r is the expected radius of a circular rock. This value has been chosen since smaller variations could correspond to rock boundary features.

2. The second filter assigns the median value to the boundary:

$$S_{boundary}(j) = \text{median}\{S_{edge}(i), S_{edge}(j), S_{edge}(k)\}$$

Note that two aspects effect the operation of the filters and the determination of i, j and k:

- The filter only produces an output for boundary point $S_{boundary}(j)$ when $S_{edge}(j) > 0$.
- The closest values of i and k to j are used jumping across gaps if necessary, i.e. $i = \max\{a : S_{edge}(a) > 0, a < j\}$ and $k = \min\{a : S_{edge}(a) > 0, a > j\}$.

After filtering, the boundary signal may contain gaps corresponding to removed points or gaps in the edge signal. The gaps are filled by means of linear interpolation to produce a continuous boundary signal representing the radial distances to all predicted points of the blob boundary.

Figures 7.10(b) and (d) show two examples of the effect of the outlier removal and interpolation processes on the radial distance profiles of regions of interest with and without a rock. The algorithm for performing outlier removal and gap interpolation to produce a blob boundary $S_{boundary}$ is given in Figure 7.13. Figure 7.14 shows the example regions after boundary closing has been performed.

Although the first outlier removal algorithm is only capable of removing individual outlier points (or two adjacent points for the median filter), with a typical expected rock circumference of around 45, if a rock is present, there are very few outliers which will not be detected.

The closed boundary signal $S_{boundary}$ is used by the following section to produce a labelled region representing the blob in the region of interest.

7.5 Finding the Extent of a Blob

The third stage of the algorithm is required to produce a simply connected labelled region representing the extent of the blob in the region of interest. Since it is necessary to construct the region from the boundary signal which consists of radial distances, the following strategy was adopted:

1. Draw radial lines in the region of interest from the centre to the specified radius in the boundary signal for all directions.
2. Apply a hole filling operator to remove any small holes resulting from quantisation of the lines. The operator fills a hole if more than a certain number of its 8-connected neighbours are set. In this case, the number of neighbours was chosen to be 5 which results in small interior holes being filled while not enlarging the labelled region at its edges because edge points only have 5 neighbours or less (See Figure 7.15).

Figure 7.16 contains the algorithm used to produce the simply connected region of labelled points in the image A_{label} from the boundary signal $S_{boundary}$. Figure 7.17 contains example regions that have been labelled after using the median filter version of the boundary closing algorithm. Note that the boundary of the labelled region has been highlighted here. Although similar, it is not quite the same as the closed boundary signal $S_{boundary}$ since all points are equally spaced.

7.6 Refining Preattentive Vision

The first level of the preattentive vision stage which was aimed at detecting ellipses in a binary image was described in Section 6.3. A problem that occurs with the preattentive vision process

Given a detected edge signal S_{edge} determined using Find_Edges(R). It is desired to determine a smooth boundary in which outliers have been removed and gaps have been filled. This will be denoted by $S_{boundary} = (r, e, g)$ where

$$\begin{aligned} r = (r_1, \ldots, r_n) &= \text{List of radial distances} \\ e = (e_1, \ldots, e_n) &= \text{List of edge strengths} \\ g = (g_1, \ldots, g_n) &= \text{Edge intensity profile} \end{aligned}$$

which is calculated using the following algorithm:

```
Function: Find_Boundary( $S_{edge}, R$)
    $threshold = \frac{1}{3} r_{opt}$
    for($i = 1, 2, \ldots, n$)
        if($r_i > 0$)
            $j = \text{next}\{a : r_a > 0, a > i\}^a$
            $k = \text{next}\{a : r_a > 0, a > j\}$
            switch(outlier_filter)
            case difference:
                if($|r_i - r_j| > threshold$ and $|r_j - r_k| > threshold$)
                    $r_j = 0$
            case median:
                $r_j = \text{median}\{r_i, r_j, r_k\}$
    for($i = 1, 2, \ldots, n$)
        if($r_i > 0$)
            $j = \text{next}\{a : r_a > 0, a > i\}$
            for($k = i + 1, \ldots, j$)
                $r_k = r_i + (r_j - r_i)\frac{k-i}{j-i}$
            $i = k - 1$
    for($i = 1, 2, \ldots, n$)
        $\phi = \arg(S_i)$
        $x = c_x + r_i \cos\phi$
        $y = c_y + r_i \sin\phi$
        $e_i = G(x, y)$
        $g_i = I(x, y)$
    return $S_{boundary}$
```

---

[a] Note that the list r is circular so that if j or k become greater than n then the value j or k modulus n should be used.

Figure 7.13: Algorithm for determining a continuous boundary.

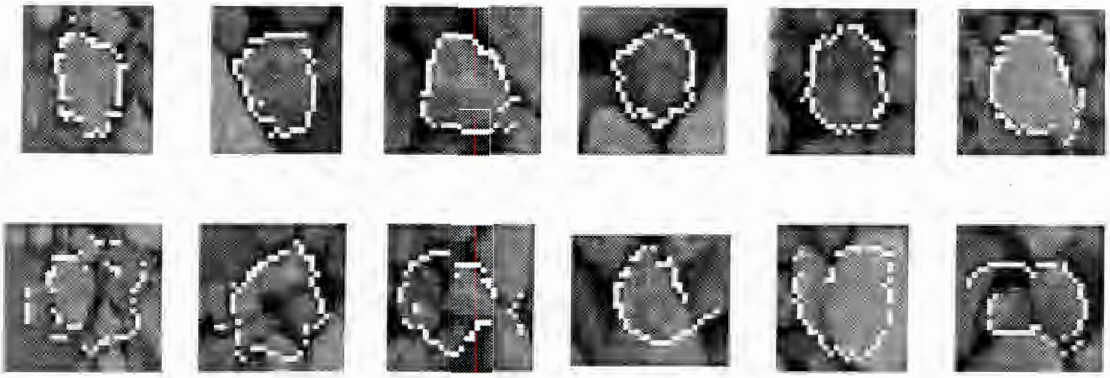


Figure 7.14: Detected boundaries in the regions of interest using the threshold filtering method.

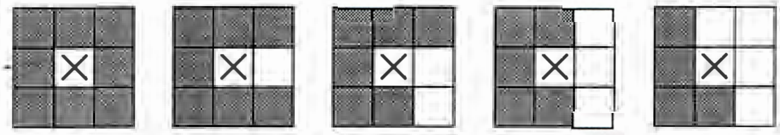


Figure 7.15: 8-connected neighbours used to determine a threshold for a spot filling algorithm. If more than 5 neighbours are set (left four figures), then it can be seen that the point itself can be considered an interior or edge point and can therefore be set too.

Given a smooth boundary signal $S_{boundary}$ calculated using Find_Boundary(S_{edge}, R), it is desired to produce a labelled image A_{label} which indicates the extent of the outlined blob in the region of interest. This is achieved using the following algorithm:

```
Function: Find_Area($S_{boundary}, R$)
        $A(\mathbf{c}_x, \mathbf{c}_y) = 1$
        for($i = 1, 2, \ldots, n$)
            $\phi = \arg(S_i)$
            for($j = 1, 2, \ldots, r_i$)
                $x = \mathbf{c}_x + j\cos\phi$
                $y = \mathbf{c}_y + j\sin\phi$
                $A_{label}(x, y) = 1$
        for($x = 1, 2, \ldots, \mathrm{width}\{I\}$)
            for($y = 1, 2, \ldots, \mathrm{height}\{I\}$)
                if($A_{label}(x, y) = 0$)
                    $neighbours = \mathrm{count}\{A_{label}(x_k, y_k) > 0 : \sqrt{(x - x_k)^2 + (y - y_k)^2} < 2\}$
                    if($neighbours \geq 5$)
                        $A_{label}(x, y) = 1$
        return $A_{label}$
```

Figure 7.16: Algorithm for determining a simply connected rock extent.

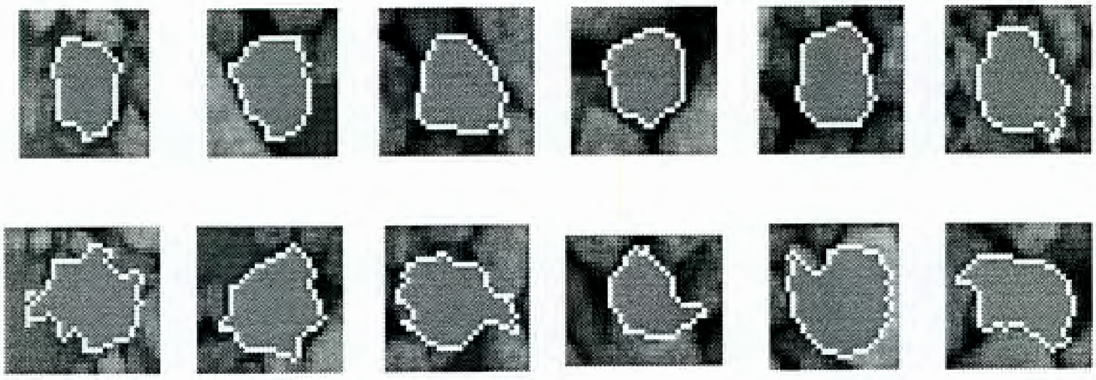


Figure 7.17: Labelled images of the detected extents in the regions of interest. Boundary outlier removal was performed using the median filter.

is that the Hough transform responds strongly to regions in which there is a large shadow which should not actually be considered as a candidate rock position. What is actually required from the preattentive vision stage is an indication of positions where approximately elliptical objects may be found which are not filled in in their centres in the binary image. With slight modification, the first stage of the segmentation algorithm described in Section 7.3 provides a very useful tool for achieving this goal.

Figure 7.18 shows the result of using the adaptive thresholding process described in Section 6.3.2 on each of the regions of interest with a value of $t_{edge} = 0.75$. It can be seen from these results that the shape of the rocks are clear with very little noise in the interiors due to surface texture. It is therefore possible to use a binary version of the blob edge detection algorithm to find whether or not the ellipse specified by the Hough transform does actually correspond to a similarly shaped object in the image.

Only the first stage of the segmentation algorithm is required, that is, the blob edge point detection process. As before, the same probability mask can be used given the detected ellipse parameters. The edge direction mask is also used to determine where edges are present in the region of interest, but in this case, since the image is binary, a simplified form of the edge detection process is used which considers only a point and its neighbour specified by the direction mask. If the point is foreground and the neighbour is background, then there is an edge point of unit strength, otherwise there is not. For each radial line, the point is kept where there is an edge with the highest probability. Figure 7.19 shows the result of this procedure superimposed on the binary image.

The average probability of all the selected edge points gives an indication of the fit of the ellipse to the object image in that region of interest. Setting a threshold t_{prob} on the average probability then allows a proportion of the spuriously detected ellipses to be discarded. Note that although some spurious points have been detected in the regions of interest containing

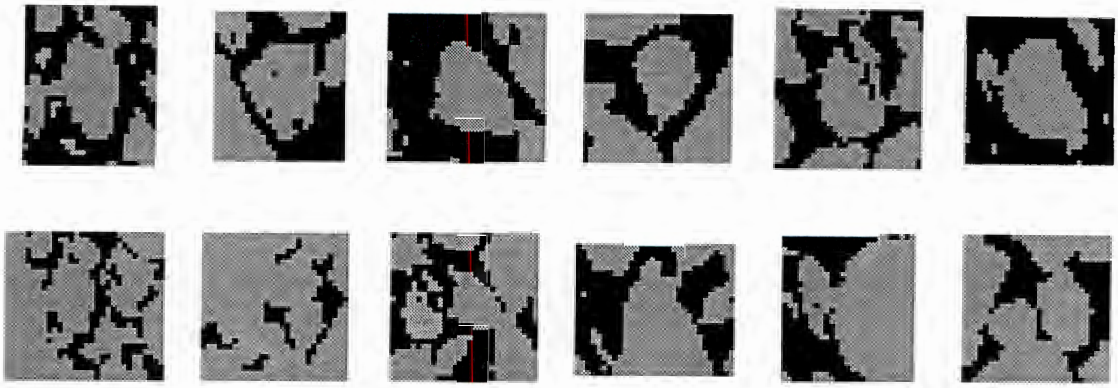


Figure 7.18: Regions of interest thresholded using a proportion of the measured intensity of the object in the centre.

rocks, there will still be a relatively high probability since most of the edge points occur near to the position of the predicted ellipse.

A refined preattentive vision output can now be obtained by cascading the two levels: first detect ellipses using the Hough transform then discard ellipses which do not match the shape in the region of interest by performing the binary version of the edge point detection algorithm[3].

The output of this procedure will be a new list $L_{highlight}$ of all the highlighted target points in the image for further investigation by attention focusing. Each element of the list consists of the following information: $L_j = (\mathbf{p}, e, \theta, h, p)$ where

$\mathbf{p}$ = The coordinates of the centre of the ellipse
e = The aspect ratio of the ellipse equal to $\frac{a}{b}$.
θ = Ellipse orientation
h = The grey level peak value of the Hough transform normalised by the circumference
p = The average probability of the detected edge points

Figure 7.20 shows ROC curves of the overall performance of the preattentive vision stage obtained by adjusting the probability threshold t_{prob} and cascading the two preattentive vision levels. A graph is shown for each case of σ_d which determines the spatial frequency content of the images. It can be seen by comparing these graphs with those in the previous chapter for the Hough transform method alone, that a significant proportion of the false hits have been removed without affecting the true hit percentage thereby justifying the use of this procedure.

---

[3] Note that it is necessary to use both of these processes to implement preattentive vision efficiently. The segmentation stage could operate successfully on its own but the computational cost of examining every possible ellipse in the image is prohibitive and the Hough transform therefore acts as a relatively fast screening process.

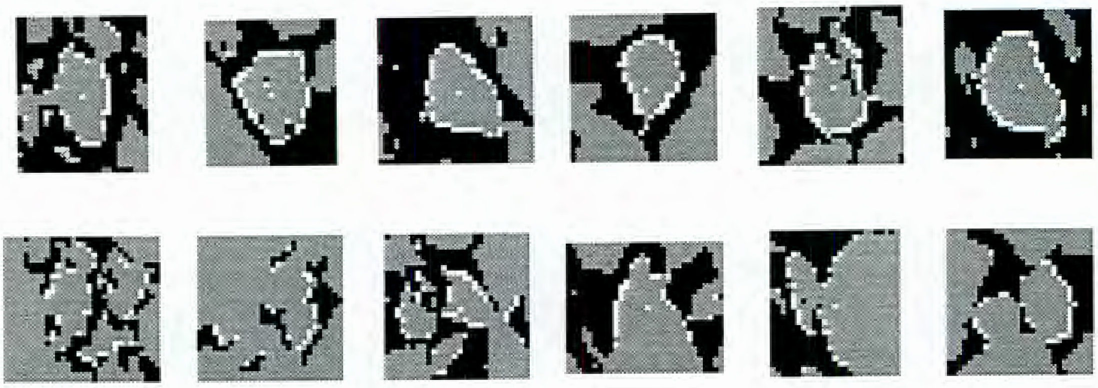


Figure 7.19: Edge points detected in the binary thresholded image to improve preattentive vision. The average probability of each of the edge points gives an indication of the match of the ellipse to the shape in the region of interest.

Figure 7.21 shows a comparison of the average ROC curves for the two values of σ_d. It can be seen from this graph that the performance when $\sigma_d = \frac{1}{3\pi}$ is considerably better than when $\sigma_d = 1$. This indicates that the extra detail which is visible in the un-bandlimited version is required for more accurate results.

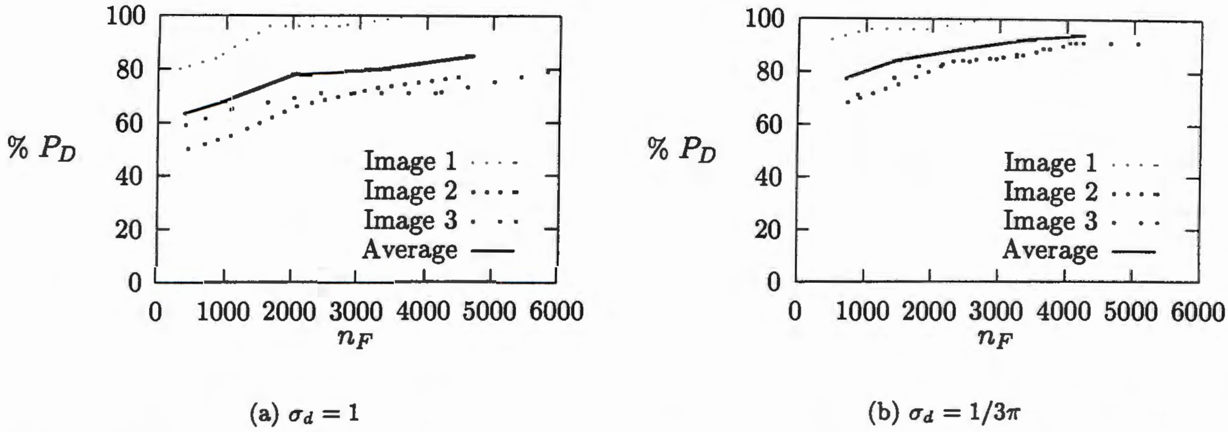
7.7 Conclusion

In this chapter, a second level of the preattentive vision stage is used to filter the list $L_{ellipse}$ by removing some of the false alarms to produce a list $L_{highlight}$ of target points of further investigation by attention focusing.

Each element of the list $L_{highlight}$ indicated a region of interest which could contain a rock which was expected to have a certain elliptical shape and orientation. Using this knowledge and general characteristics of the appearance of rocks, a segmentation algorithm was developed consisting of three stages.

Firstly, knowledge of the size and shape of the object was used to bias the selection of edge points in the region of interest which had appropriate orientation and position for a rock. Next, a boundary was created by removing extremities from the edge points and closing gaps. Finally, a labelled image A_{label} was produced by filling the interior of the detected blob boundary.

$S_{boundary}$, the blob boundary, and the labelled region A_{label} are used in the next chapter for measuring features from the grey level and edge images which will then be used to discriminate between regions of interest to determine which do and which do not contain the predicted rock.



(a) $\sigma_d = 1$

(b) $\sigma_d = 1/3\pi$

Figure 7.20: ROC obtained when varying t_{prob} in steps of 0.05 from 0.6-0.8 with $t_{edge} = 0.75$ and $t_{clip} = 0.55$. A high value of t_{prob} corresponds to a low % P_D.

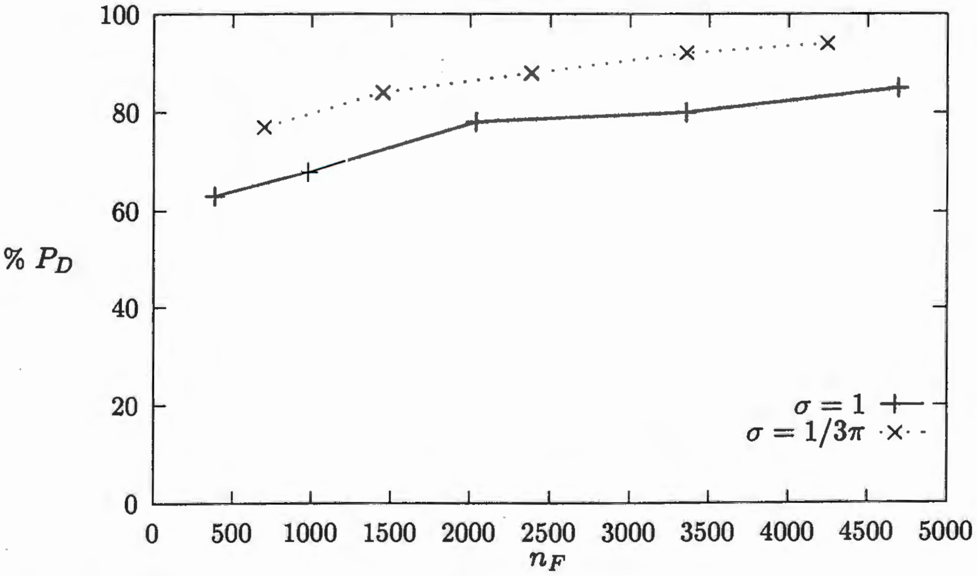


Figure 7.21: Comparison of the mean ROC curves obtained for the two cases of σ_d clearly show that this extension to the preattentive vision stage works better when $\sigma_d = 1/3\pi$.

Chapter 8

Blob Classification for Rock Recognition

8.1 Introduction

Classification theory was introduced in Section 3.4 where it was found that recognition is a basic attribute of living organisms [89]. The field of pattern recognition involves discriminating between input data to determine which population they belong to. There are two basic problems which must be tackled to achieve this: (1) Suitable features must be identified which will allow discrimination between the different populations while reducing the dimensionality of the classification problem. (2) The feature space must be partitioned into regions which correspond to the different populations.

Once this has been achieved, a pattern recognition procedure commonly consists of the following three stages [7]: (1) Data sensing involves measuring the object to be recognised. This was achieved in Chapter 7. (2) The features must be measured from the measured input data. (3) Feature vector classification must then be performed by identifying the partition in the feature space which contains the vector and labelling the current input data with the appropriate population. Since it is not usually possible to do this with complete accuracy, it is necessary to be able to measure the classification error and perform optimisation to reduce it to a minimum.

The goal of this chapter is to solve the two problems associated with a pattern recognition system and then perform optimisation to minimise the classification error. An alternative view of this process is that it is an initial analysis of the labelled region corresponding to each target that was selected by the preattentive stage in Chapter 6. The analysis is aimed at identifying which targets could actually represent a rock in the image. This corresponds to the second stage in the human visual system which compares detected objects to an internal representation and performs identification [90]. In practice, this is achieved by analysing characteristics of the segmented region created in Chapter 7 to determine whether or not it could correspond to a rock which matches the predicted ellipse model.

Information concerning the context of the rock is not considered at this stage. The context at this stage refers to the actual arrangement of rocks in the image. It is therefore possible that there may be several hits on the same rock (because of multiple close hits during preattentive vision) or in contradictory positions in the image (due to classification errors). A later stage of hierarchical analysis is used to extract the most likely set of rocks from the list of detected positions by considering their relative positions and the evidence for their existence.

A statistical pattern recognition approach has been taken to perform classification. Bartlett *et al* [3] describe a visual inspection system which uses statistical pattern recognition for analysing solder joints. They highlight the necessity of selecting good features to ensure the success of any classification algorithm. They describe good features as having the following characteristics:

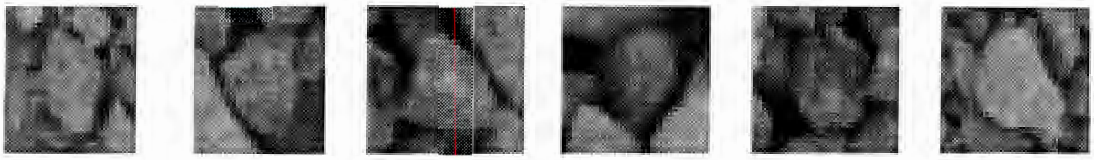
- They can be calculated reasonably quickly.
- Values for the features should cluster together for each of the different classes.
- The clusters of values for different classes should be spaced far apart.

Methods for measuring fifteen features have been developed for use in this system. The features were designed using a process of identifying common classification errors and then producing a new feature which should improve performance under that particular condition. A process of interaction was then used to refine the feature measurement process using real data. The characteristics of good features were kept in consideration and all the features have been normalised to the range 0–1.

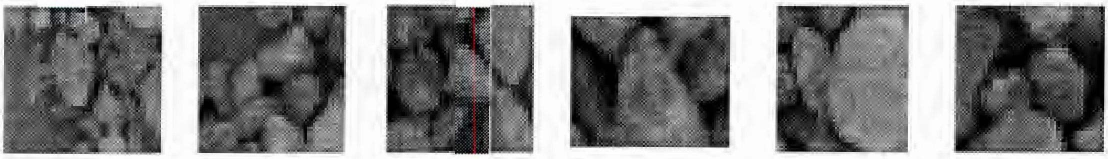
Section 8.2 describes several sources of information made available by the segmentation process, from which the features are measured. Section 8.3 describes how the features were developed in four groups to facilitate discrimination between the sub-classes of situations. In order to perform classification and to optimise the performance, it is necessary to determine representative samples to create training and test sets. This problem is analysed in Section 8.4. Section 8.5 describes the classification algorithm that has been implemented and finally Section 8.6 discusses the selection of an optimal feature set for best classification performance.

8.2 Information Bands

The use of multiple bands for an image is common. Colour images are usually described using three bands, one each representing red, green and blue. Zuerndorfer and Wakefield [97] also state that many machine-sensing systems are required to process signals from multiple *sensors* which represent the same region using different sensing technologies such as optical, sonic, nuclear and electromagnetic. They also state that the situation may be further complicated



(a) Regions of interest containing suitable rocks.



(b) Regions of interest not containing suitable rocks.

Figure 8.1: Sample regions of interest.

when the different sensors have different spatial transfer functions which result in a *multi-resolution* system.

In this application, the region of interest image can be processed in a number of ways in order to highlight different aspects of the image such as the position of rock edges, dark narrow gaps caused by deep shadows between rocks, and thresholding to show foreground and background. Combining information from each of these sources should result in a more robust classification since it considers more than one aspect of the object. Each of the image bands which can be created using a different processing technique will be referred to here as information bands.

Each of the information bands is a preprocessed version of the region of interest grey level image. They therefore all have spatially equivalent coordinate systems and so measurements may be made easily from each band.

Figure 8.1 shows the sample regions of interest that were used in the previous chapter for developing the segmentation algorithm. They are used again here to illustrate the effect of the various forms of preprocessing which are described in the next four sections.

Expected Intensity

The expected intensity is not actually a complete information band but is rather a value which has been measured from the grey level image of the region of interest and is important for normalising many of the other features.

If a rock is present in the region of interest, the centre of R should be the middle of the rock which is very likely to be horizontally orientated i.e. $\psi = 0$ (See Section 5.2[1]). The perceived intensity should therefore correspond to $\pi I_i \sin^2 \mu$ which is the maximum value under the diffuse lighting configuration. Knowledge of the slope versus intensity graph for the given lighting arrangement can then be used to detect rock surface and shadow points.

The expected intensity I_e of the rock is determined by measuring the average of the centre values in R in a small region with a footprint of $\frac{2}{3}$ times the expected rock diameter (See Section 6.3.2 where an averaging convolution kernel K of this size is used to reduce the effect of surface texture on the surface of a rock).

Highlighting Narrow Gaps

A major problem that occurs in segmentation of rock scenes is that of merging. The object of this information band is to enable the measurement of a feature which gives a strong indication of the presence of rock merging. This is achieved by observing that often when rocks are close together, there is a dark narrow shadow between them. Clearly if the shadow appears in the middle of a labelled region, then merging has probably occurred.

The morphological tophat function $HAT(f, g)$ which was described in Section 3.2.3 is used here as a valley detector for the purpose of highlighting the shadows in the image. Given that the grey level image of the region of interest is I, then the gap detected band is determined using the following formula:

$$A_{gap} = HAT(I^{-1}, g)$$

where I^{-1} is the photometrically inverted grey level image. g is a small disc structuring element with a radius of $\frac{1}{4}r$ where r is the expected rock radius. As with several other parameters, it is necessary to have a value which is proportional to the expected rock size in case it changes.

The operation of the tophat operator is to remove all structure from the image which is larger than the structuring element. It is therefore necessary to have g slightly larger than the width of the gaps between adjacent rocks. $\frac{1}{4}r$ allows for gaps up to $\frac{1}{8}$ of the width of the predicted rock size to be detected. With the actual resolution of $r \approx 8$, this corresponds to gaps of less than 3–5 pixels wide.

Figure 8.2 shows the result of applying the valley detector to each of the regions of interest. Note that for each region of interest representing a rock, the response within the boundary of the rock is very low.

---

[1] Note that in this case, the intensity is used instead of the bi-directional reflectivity ρ_d. Intensity is proportional to ρ_d multiplied by the radiance of incident light I_i which is constant for the scene.

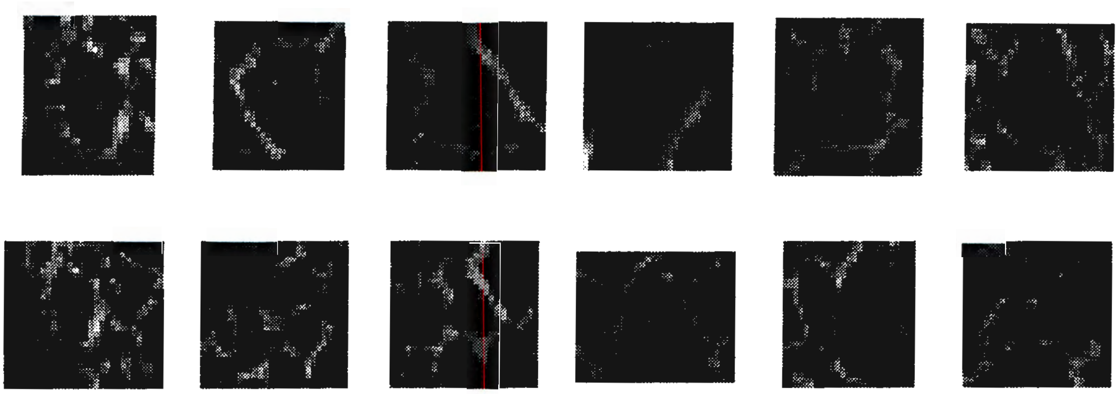


Figure 8.2: Highlighted narrow gaps using the tophat operator in the regions of interest.

Image Foreground

The aim of thresholding the image is to provide a band containing foreground and background information. This provides an alternative estimate of which points represent the object foreground in the region of interest which can be compared with the labelled region.

Thresholding can be achieved using either the moment-preserving threshold (See Section 3.2.2) or by using the expected intensity and the lighting model to determine a threshold.

If the moment-preserving threshold value is determined to be t_{moment} then each point in the thresholded image is calculated using:

$$A_{M-P-threshold}(x,y) = \begin{cases} 1 & \text{if } I(x,y) \geq t_{moment} \\ 0 & \text{otherwise} \end{cases}$$

The measured intensity of a horizontal surface was used to determine a threshold t_{edge} in Section 6.3.2 for highlighting the edges of rocks by finding shadows. In this case, the thresholded image is obtained using the following expression:

$$A_{I-threshold}(x,y) = \begin{cases} 1 & \text{if } I(x,y) \geq t_{edge} I_e \\ 0 & \text{otherwise} \end{cases}$$

Figures 8.3 and 8.4 show the results of thresholding the regions of interest using the two methods just described. It can be seen that the model based threshold generally provides a more accurate segmentation of the foreground when a rock is present. However, when the assumption that the centre of R is the centre (and hence top) of a rock is false, then the model based threshold is not as accurate as the moment-preserving threshold.

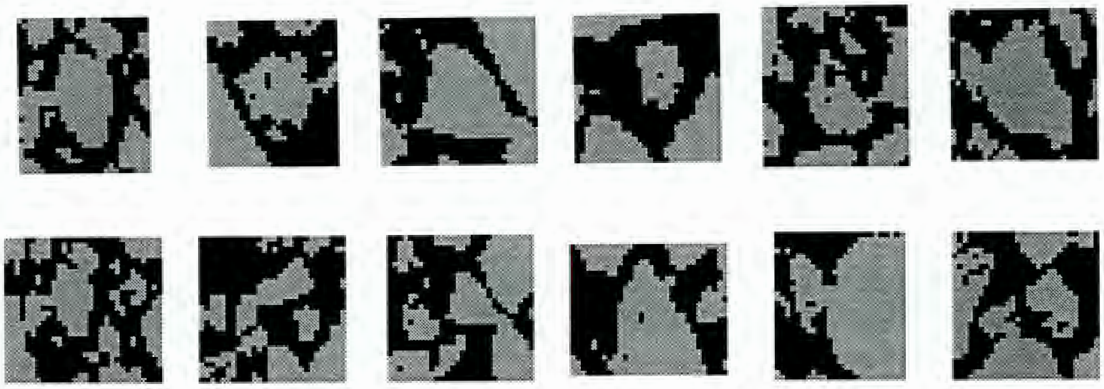


Figure 8.3: Regions of interest thresholded using moment preserving threshold.

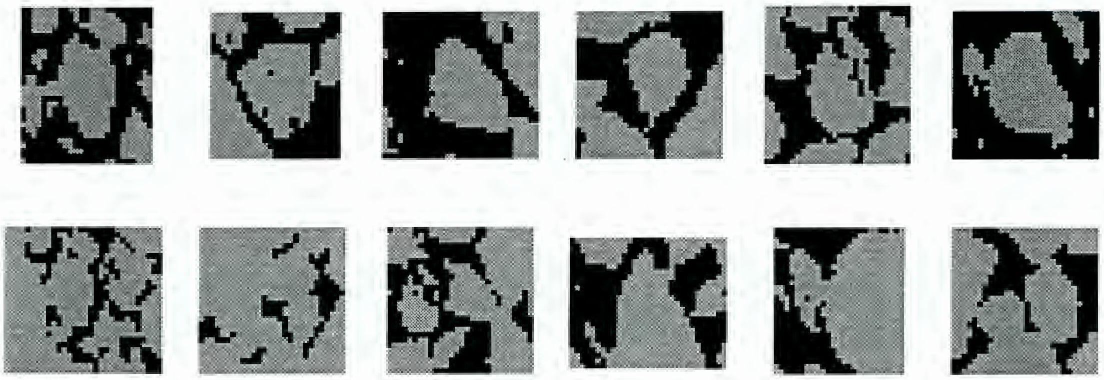


Figure 8.4: Regions of interest thresholded using $t_{edge} = 0.75$ times the expected rock intensity.

Edge Strength

Image edge strength can be used as an indication of the presence of rapid intensity changes in the image which could correspond to physical features such as the edge of a rock. The directional edge detector was shown in the previous chapter to be particularly effective when a rock of known shape and position is in the region of interest. Unfortunately, this is not always the case, and it is necessary to be able to detect when other situations have occurred.

Although a rock may have surface texture, the edge strength at the edge of the rock should be consistently stronger than that caused by texture, otherwise the shape of the rock would be effectively camouflaged and it would be very difficult to see it at all. It is therefore necessary to be able to measure edge strength associated with texture and other interior features which will not be found using the directional edge detector. For this reason, an alternative, the morphological gradient operator $MGRAD(f, g)$ (See Section 3.2.3) was used since it is easy

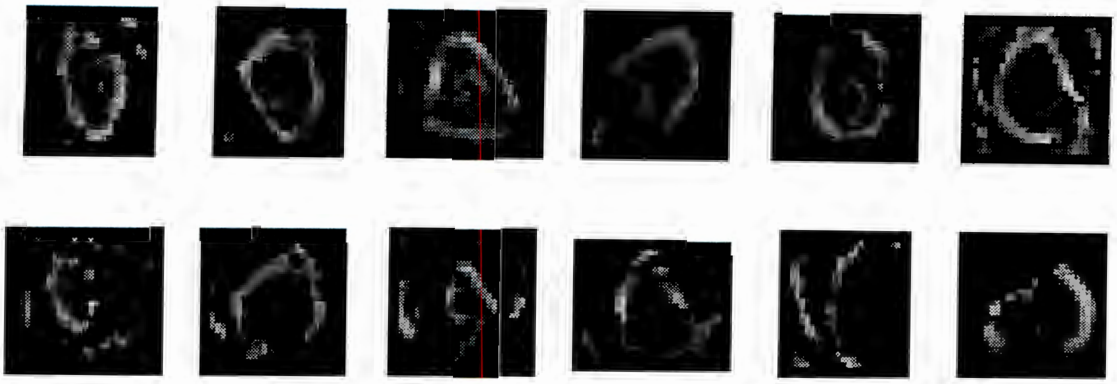


Figure 8.5: Directionally edge detected regions of interest.

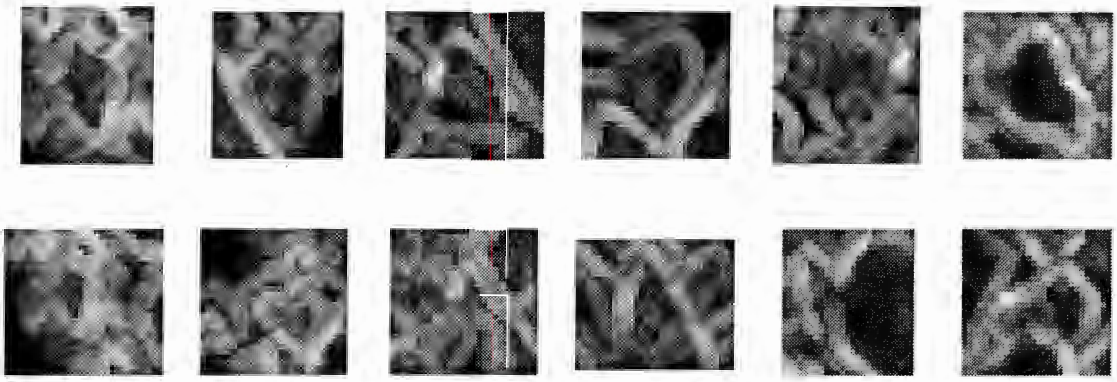


Figure 8.6: Morphologically edge detected regions of interest.

to implement and provides isotropic response. Using this band, the edge strength at the edge of the rock can be compared with that in the interior.

Figures 8.5 and 8.6 show the effect of performing edge detection in the regions of interest using the directional and morphological gradient methods to result in the two image bands; A_{dir} and A_{grad} respectively.

8.3 Feature Selection

As stated earlier, the goal of feature selection is to provide a list of features which are easy to calculate and produce different values for different classes. In this case, it is necessary to distinguish between regions of interest in which a rock is actually present and has been properly segmented, and those in which there is not.

Because there are actually several different reasons why a labelled region should not be classified as being a rock (it has been incorrectly segmented due to size or position offset, or error; or because it is part of a larger rock or contains a group of smaller rocks) the non-rock class can actually be considered as several sub-classes. It was therefore necessary to identify sub-classes by observing situations in which classification was incorrect, and then attempting to design a feature which would be effective for discriminating between the rock class and the particular non-rock sub-class.

This process resulted in the identification of four basic groups of features. Each group is designed to respond to a different non-rock sub-class. Essentially, the four groups identify correctness of the shape of the segmented region, characteristics of the boundary, and the problems of rock splitting and merging.

It is important to consider how a feature is to be measured to ensure its effectiveness. It was found in Section 3.4.1 that measurement error reduces the usefulness of a feature to the point where it may actually become detrimental to the overall classification performance.

The discrete nature of the image representation means that special attention must be paid to the measurement of perimeter length. A list L_{edge} of edge points of the segmented region is found by tracking around the boundary of the labelled region in A_{label}. The perimeter of the blob is measured from L_{edge} by assigning unit length to horizontal and vertical steps and $\sqrt{2}$ to diagonal steps. A list $L_{interior}$ of the interior points of the segmented region contains the remaining points in A_{label} after L_{edge} has been removed. L_{edge} and $L_{interior}$ can be seen in Figure 7.17. L_{edge} is shown in white and $L_{interior}$ is shown in grey superimposed on the background.

The remainder of this section is split into four sections which give an explanation of the role of each feature group. All the features that have been developed in each group are also identified and described.

8.3.1 Optimal Rock Shape

Shape is commonly used in computer vision systems where non-information preserving features are used [47, Section 10.2.2]. These are defined as measurements which do not permit the unique reconstruction of the original object but are very useful as indicators of shape. Typical examples of such shape features are perimeter, area and two-dimensional extent.

Since it is not yet certain whether or not the blob actually represents a rock or not, shape features are designed to determine whether or not the labelled extent of the blob could correspond to the predicted rock in terms of its visual footprint. The features in this group are measured from both the boundary of the blob and the labelled extent.

The following list provides a description of each of the features in this group:

Surface area of blob is calculated in the following way: Given a_p is the area of the predicted ellipse model and a_b is the area of the blob extent in A_{label} then this feature value is $\frac{a_b}{a_p} - 0.5$. A value near 0.5 is desirable as it indicates a blob with an area close to the predicted value.

Circularity. The circularity [47] is defined as $1 - \frac{4\pi a_b}{Perimeter^2}$. This value is zero for a circle and increases to 1 as the boundary increases in complexity. A low value is desired because rocks are expected to be roughly circular in shape.

Centre of mass offset. The offset is defined as $\sqrt{\overline{x}^2 + \overline{y}^2}$ where $\overline{x}$ and $\overline{y}$ are the zero order moments of the labelled image A_{label}. This feature measures the distance from the centre of the ROI to the centre of mass of the blob. Clearly, if a rock is actually present and has been correctly segmented, this value should be small.

Edge position probability. This is the average value of the probability mask for each of the boundary points $S_{boundary}$ of the blob. It gives a measure of how close the boundary is to the predicted size and shape and should have a value close to 1 if the predicted rock is present.

8.3.2 Boundary

These features ensure that the detected boundary obeys the restraints imposed by a physical object like a rock i.e. that the boundary is continuous (see the fourth rock characteristic in Section 7.2). As a result of this, all of these features involve inspection of the measured boundary of the blob.

The following features have been developed for this group:

Proportion of good points. This is the proportion of defined points in S_{edge} after outliers have been removed and before gaps are filled. A high value for this feature indicates the presence of a complete edge in the region of interest.

Largest boundary gap size. If l_{gap} is the length of the longest gap in the edge signal S_{edge} after outliers have been removed and before interpolation, then this feature is measured as $\frac{l_{gap}}{Perimeter}$. A low value is desirable because a rock should not have large gaps in its boundary. This feature can play a dual role in preventing splitting by detecting when the corners of large rocks appear in the region of interest and are segmented as small rocks because there is likely to be a gap in the edge where the corner joins the rest of the rock.

Edge strength. This is the average value of edge strength for all points in $S_{boundary}$. The directional edge strength operator is used for this feature since incorrectly (i.e. perpendicular to the expected direction) oriented edges will have a low response. When the predicted rock is present, there will be correctly (i.e. tangential to the expected direction) oriented edges in all directions so this feature should produce a high value.

Edge discontinuities is the number of boundary discontinuities in $S_{boundary}$ as a proportion of the boundary length. A discontinuity is defined as a radial step change greater than $t_{outlier}$ which was defined in Section 7.4 and is greater than allowable variations in the boundary.

8.3.3 Splitting

Splitting occurs when the region of interest only contains part of a large rock. Accepting regions of interest in which splitting has occurred could result in a bias in the size distribution towards smaller rock sizes.

This group of features is designed to detect artifacts which are associated with a blob in which splitting has occurred. Since the labelled region represents part of something larger, it is necessary to analyse the boundary of the blob to identify splitting. Rock splitting can be identified by detecting large sections of the boundary where there is high intensity and very low edge strength.

The following two features are designed to identify splitting artifacts:

Minimum edge strength. Given a signal E_{edge} representing the edge strength at each point on the boundary $S_{boundary}$, this feature is calculated as:

$$\mathbf{x}_j = \min((E_{edge} \ominus s_1) \oplus s_2)$$

where s_1 and s_2 are flat disc structuring elements.

The role of eroding with s_1 is to remove narrow peaks from the edge signal and so s_1 is a short structuring element of width $\frac{1}{3}r$, where r is the expected radius of a circular rock. All peaks which are narrower than $\frac{1}{3}r$ will be removed from the signal.

Dilation with s_2 is responsible for filling in narrow regions of the boundary with low edge strength up to the length of s_2 which is set to $0.9r$. Any regions of low edge strength which are wider than this structuring element will survive the morphological processes and be present in the resulting signal. The minimum value of the resulting signal is taken as the feature value. If this feature does have a low value, it indicates that there is a wide region of low edge strength which could indicate splitting.

Edge intensity too high. It was observed in Section 5.2.5 that given the intensity I_e of a horizontal surface of a rock, it would be possible to determine an intensity threshold above which image points would probably belong to same rock. This relationship is used here to find interior points which are at the boundary of the rock. The feature is measured as the proportion of boundary points which are brighter than a threshold of $0.9I_e$. The value of 0.9 is used since such bright points will have a very high probability of being interior points of a rock if I_e is correct.

8.3.4 Merging

Merging is the condition in which the blob actually represents several rocks in close proximity in the region of interest. If this situation is not detected, it could result in a bias of the size distribution in favour of larger sizes.

The role of this group of features is to detect artifacts which indicate that merging has occurred within the current region of interest. Merging can be identified by looking for gaps between rocks in the interior of the labelled region.

Several features have been developed for this purpose and they are listed below:

Edge strength ratio. The average edge strength of the points on the boundary L_{edge} and the interior $L_{interior}$ of the labelled region are measured as e_{edge} and $e_{interior}$. The morphological gradient band is used for this feature since it is isotropic and can therefore respond to internal intensity features in $L_{interior}$, therefore allowing comparison of the boundary to the interior. The feature value is calculated as $\frac{e_{edge}}{e_{interior}}$. A high value is desirable as the gradient should be stronger at the edge of a rock than in its interior.

Proportion of background. This is calculated as the proportion of the labelled pixels in A_{label} that are above the threshold in $A_{M-P-threshold}$. A high value is desirable because the shape of the rock should be closely preserved in $A_{M-P-threshold}$. If there is actually a group of rocks in the region of interest, there is likely to be a relatively large proportion of background points in the labelled region.

Maximum gap strength. This feature is calculated as the maximum value of the highlighted gap band A_{gap} for all points $L_{interior}$ in the interior of the blob. A high value for this feature indicates a gap within the blob which gives strong evidence for rock merging.

Interior intensity below threshold. It was observed in Section 5.2.5 that an intensity threshold could be obtained below which image points are likely to represent shadows or edges points. This information is used here to find non-rock points within the labelled region. The feature is measured as the proportion of the area of the interior of the blob $L_{interior}$ which is darker than $0.5I_e$. The lighting model shows that under no circumstance can a point on the rock be that dark unless it is in shadow.

Darkest interior point. If the darkest point in the the list $L_{interior}$ has intensity I_d, then this feature is evaluated as $\frac{I_d}{I_e}$. Since interior points should not be very dark, this should be a reasonably large fraction.

8.4 Generating Representative Training Data

The aim of this section is to describe a procedure that can be followed to create representative training sets since these will result in more accurate and reliable classification. The following three considerations must be overcome in order to create a good training set.

- Is it more important to correctly classify rock or non-rock regions of interest? Since any ROI containing a rock that is misclassified cannot be detected later — while non-rock ROI's may still be deleted — it would appear that classification should be biased in favour of the rock class.
- What relative weighting is required between the two classes? Note that the training data typically contains more non-rock samples since there are clearly more permutations of the region of interest which don't contain a suitable rock. This would tend to emphasise non-rocks in classification by having a higher proportion of the training set representing non-rocks. This is partially offset by rock features being clustered closely while non-rock features are more spread around the feature space.
- How can the training set be extracted so that it is representative of the feature vectors that must be classified? This situation occurs because the regions of interest are not randomly selected due to the preattentive vision stage which highlighted target positions.

A solution has been proposed and implemented to overcome these problems. The basis for this solution is to use the same procedure for selecting the training set samples as that used during operation. The actual population of each sample can then be determined using *a priori* information. The samples that are selected this way will be representative of the types of samples that must be classified during operation and should also give an indication of the relative occurrence of each class or population.

A priori information must then be used to determine the actual population of each sample for the training set. This is achieved by using a list of the size, position, orientation and aspect ratio of a best fitting ellipse for each rock in the image. The ellipses are fitted manually for this purpose.

The distance measure described in Appendix D can then be used to determine the distance $d_{feature}$ between any ellipse selected by the preattentive vision stage and the ellipse representing the closest actual rock in the image. Since the distance measure is normalised to have a value of 1 for the threshold between rocks and non-rocks, it can be used with the following discriminant function to determine the appropriate population for the given sample and its corresponding feature vector:

$$\hat{g}(\mathbf{x}) = \begin{cases} 0 & d_{feature} \leq 1 \\ 1 & \text{otherwise} \end{cases} \tag{8.1}$$

It is sometimes desirable to associate a weight or cost of error with each sample for the purpose of calculating classification error. This can also be achieved using the distance measure and the following function, which returns a weight of unity when the classification is certain (very close to or far from a rock) and is reduced for marginal samples where the distance is close to 1.

$$f(\mathbf{x}) = \begin{cases} 1 & d_{feature} \leq \frac{1}{2} \\ 2(1 - d_{feature}) & \frac{1}{2} < d_{feature} \leq 1 \\ 2(d_{feature} - 1) & 1 < d_{feature} \leq \frac{3}{2} \\ 1 & \text{otherwise} \end{cases} \tag{8.2}$$

The information encoded in the resulting training set can then be used with a suitable discriminant function to classify or determine the population of any input feature vector. The weight value can then be used with a pattern error function for determining a measure of the probability of misclassification using the methods described in Section 3.4.1.

The discriminant function that will be used is described in the next section, after which classification error is used with a genetic algorithm to maximise classification performance by finding the optimal feature set.

8.5 A Combined Discriminant Function

The role of a discriminant function was introduced in Section 3.4.1 where it was found that it is responsible for classifying a feature vector by partitioning the multidimensional feature space into regions corresponding to the different populations.

In this case there is a 15-dimensional feature space which must be partitioned into two regions; a region for vectors representing rocks and the remainder for vectors which don't represent rocks. In the ideal situation, the regions should be such that there are no misclassifications. Typically, this is not possible and one should thus try to minimise the probability or error, or the average cost of errors if some are more expensive than others.

The k-nearest neighbour classifier was described in Section 3.4.1 and is used here in conjunction with a rule based approach to provide rapid execution without compromising accuracy.

A knowledge based approach to identifying features and rules for classifying the resulting feature vectors has been used by Bartlett *et al* [3]. Rules provide fast classification since it is unnecessary to scan through a training set of preclassified samples. For this reason, the feature vector is initially classified using two threshold values for each feature. The threshold values divide the feature space into the interior and exterior of a hyper-cube. All feature vectors representing a rock should lie inside the hyper-cube along with some representing non-rocks. Feature vectors outside the hyper-cube can be rapidly classified as non-rocks at this stage. The classification of the remaining vectors is then verified using a k-nearest-neighbour classifier which requires a training set of sample feature vectors with their true classification and operates more slowly.

The training set used for the k-nearest-neighbour is also used to determine the threshold values for the hyper-cube based on a proportion limit $t_{feature}$. The thresholds are determined so that the given proportion $t_{feature}$ of samples representing rocks will be correctly classified by a rule based discriminant function based on that feature alone. If the proportion is unity, all rock feature vectors in the training set will be correctly classified.

This classification rule is very quick to compute but not very accurate for the non-rock class. Accuracy is improved by using the k-nearest-neighbour rule to check the classification when the feature vector lies within the hyper-cube and is hence classified as a rock. In this way obvious non-rock vectors are discarded using the thresholds and the k-nearest-neighbour rule determines the final result. Note that most classifications are expected to return a non-rock result due to the low proportion of possible rock to non-rock regions of interest so that this method can provide a substantial computational saving.

There are therefore two motivations for using this combined method of classification:

1. Knowledge of rocks can be integrated into the classification decision in an expert system approach by expressing fundamental limits on features.

2. The threshold based classification stage can perform rapid discrimination. Since the k-nearest-neighbour method is only required for vectors lying inside the hyper-cube, a dramatic decrease in execution time can be expected for the combined method compared to k-nearest-neighbour method alone.

Figures 8.7, 8.8 and 8.9 show the results of testing the k-nearest neighbour, rule based and combinational discriminant functions using cross-validation, and the hold-out method with unweighted error counting (See Section 3.4.1). Justification for the combinational method is found by comparing error counts achieved by the k-nearest-neighbour method and the combinational method. This shows that there is no perceptible difference in classification accuracy between the two methods when $t_{feature} = 1$. Both methods achieve optimum performance at an error count of 0.29 at k=27[2]. Threshold only classification is also shown for comparison. It can be seen that the best performance of 0.36 can be achieved when $t_{feature} = 0.96$ but is considerably worse than the other two methods.

The training set containing 2233 samples was generated using the procedure described in Section 8.4 for determining representative samples from two images. An additional test set containing 938 samples was obtained from a third image using the same procedure.

---

[2] Note that the probability of misclassification for each population is measured separately and the error count is determined to be the average value. This compensates for the relative imbalance between the numbers of rock and non-rock samples.

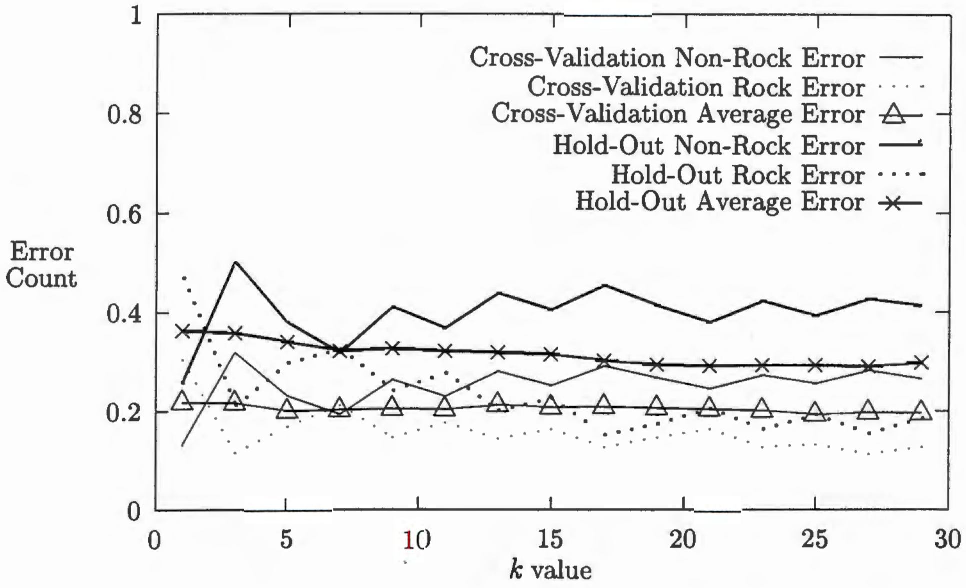


Figure 8.7: Graph of Error Count versus k value for k-nearest-neighbour method using Hold-Out and Cross-Validation methods. The Error Count is shown for each class as well as the average Error Count for each method. It can be seen that Cross-Validation produces better results than the Hold-Out method. This is expected since the former method uses samples from the training set to measure error. Note that with the Hold-Out method, performance improves as k increases up until 27.

8.6 Determining the Optimal Feature Set

As described in Section 3.4.1, classification error occurs because perfect classification cannot be achieved resulting in a non-zero probability of misclassification (PMC). The aim of this section is to minimise the PMC by means of feature sensitivity analysis. Feature sensitivity analysis was described in Section 3.4.1 as a process of ranking the effectiveness of features and discarding those that contribute to an increase in classification error or PMC. If a feature provides only marginal benefit for classification, it was found that noise on the measurement of the feature could be more significant than the possible gain resulting in an overall increase of PMC.

PMC will be minimised by selecting an optimal feature set p_{opt} using a genetic algorithm (GA) as the optimisation mechanism (See Section 3.4.2).

In order to perform optimisation, it is necessary to be able to measure the classification error or PMC. Given a pattern error function, h^{EC} or h^{WE} which assigns a cost to each error, an estimate of the cost of error $\hat{P}$ is obtained for the rock and non-rock populations independently using the training and test data sets obtained in the previous section. An overall estimate of error is then obtained by calculating a weighted average of the two values. This approach compensates for a relatively small proportion of the data sets comprised of rock

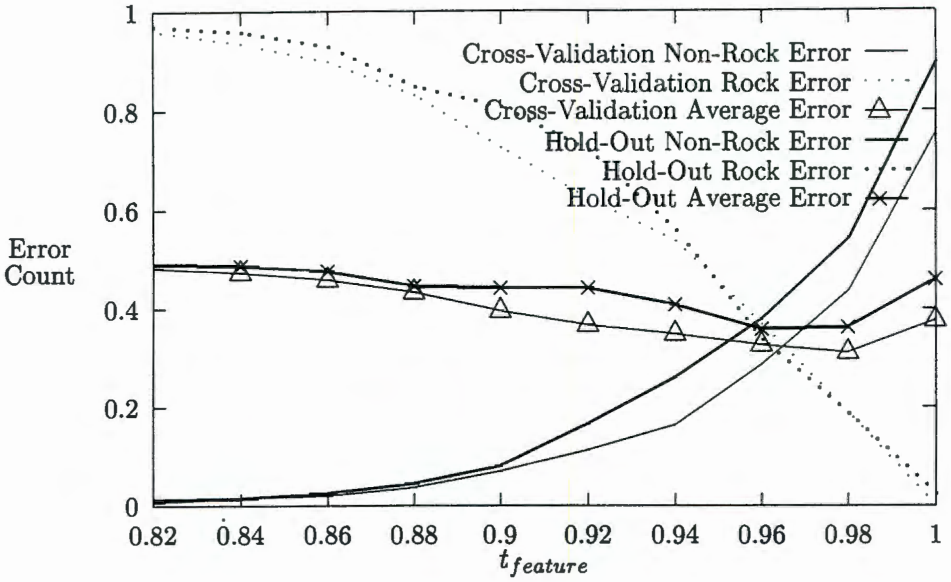


Figure 8.8: Graph of Error Count versus Threshold level for threshold classification using Hold-Out and Cross-Validation methods. The Error Count is shown for each class as well as the average Error Count for each method. It can be seen that the best Hold-Out rule based classification occurs when $t_{feature} = 0.96$ but is not as good as k-nearest neighbour performance. Note that the rock class error increases as $t_{feature}$ decreases as fewer rock samples are contained in the hyper-cube. The smaller hyper-cube also results in more non-rock samples being correctly classified. It can be seen from the very high non-rock class error when $t_{feature} = 1$ that there is considerable overlap between the classes in the feature space.

samples. Adjusting the weighting allows for emphasising the correct classification of rock or non-rock samples. Since it is important not to lose samples which correspond to rocks, a weighting in favour of the rock class would be indicated. For this reason, optimisation is performed using the relative weightings of 50:50, 40:60 and 30:70; the second class corresponding to rocks. A relative weighting of 60:40 has also been included for comparison purposes.

The chromosome string that is manipulated by the GA is used to determine which features are enabled and the value of k for the k-nearest-neighbour algorithm. The first 15 bits each correspond to one of the features. The final four bits encode k starting with the least significant bit to give a value n in the range 0 to 15. k is calculated as $2n + 1$ since it should be odd.

The genetic algorithm was used with a 19 bit chromosome string, a population of 21, $p_{crossover} = 0.6$ and $p_{mutation} = 0.01$. Elitism was used to ensure a monotonic improvement in performance. Given these settings for the GA, and an evaluation function which returned the classification error for each chromosome string, 30 generations were allowed for optimisation.

Figure 8.10 shows the result of using the genetic algorithm in the described manner with 8 different evaluation functions. Weighted error counting results in a lower error count due to the lower cost of some errors and so these results cannot be compared directly with those

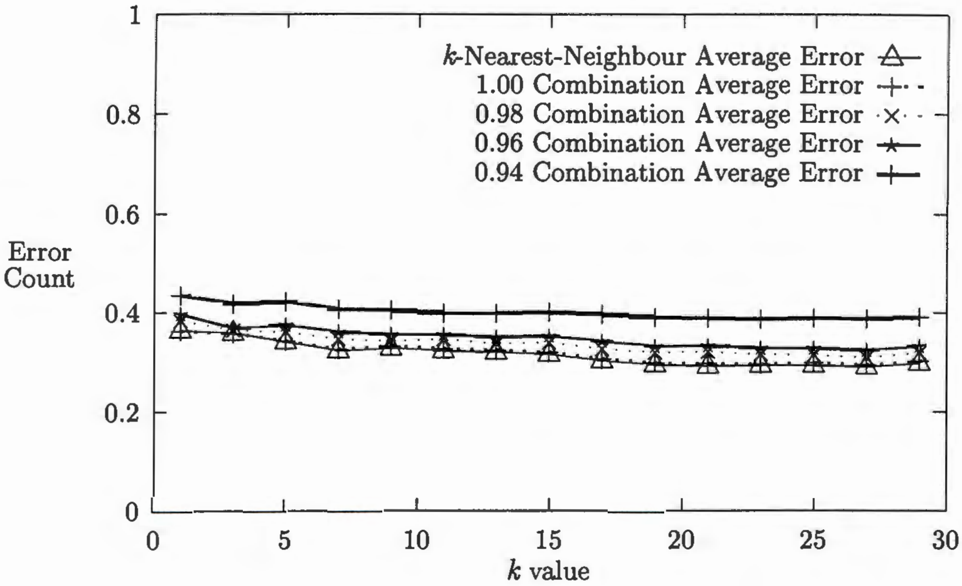


Figure 8.9: Graph of Error Count versus k value for the combination classification rule using Hold Out method. The average Error Count is shown for each threshold level. It can be seen that when $t_{feature} = 1$ there is no compromise on the classification accuracy of the k-nearest neighbour classifier alone since the difference in performance is imperceptible.

obtained from unweighted error counting. It can be seen that the genetic algorithm is able to improve the error count by about 10% for each of the evaluation functions. Note that in each case, this corresponds to approximately 2 hours of computation.

Table 8.1 shows the final result obtained from each of the evaluation functions. In each case, the features are shown in groups in the same order as their description in Section 8.3. For each feature, a 1 indicates that the feature should be enabled and a 0 shows that the feature should not be enabled. The best value of k is shown at the end of the row.

It can be seen from Table 8.1 that the most consistently useful features are the third shape feature (offset), the second and fourth boundary features (gap size and discontinuities) and the third merging feature (highlighted gap strength). On the other hand, both the splitting features and the second and last merging features (background and darkest point) should be avoided. It appears from this result that the success of the second stage of preattentive vision described in Chapter 7 has reduced the number of cases in which splitting and merging can occur therefore making those features less useful for classification.

Apart from one case, there is also reasonably strong evidence that the value of k should be fairly high.

Figure 8.11 shows ROC curves which shows the combined performance of the preattentive vision and classification processes at selecting rocks in the image. Results from three images are averaged to determine the curves. The same three images were used as in Section 6.3.4.

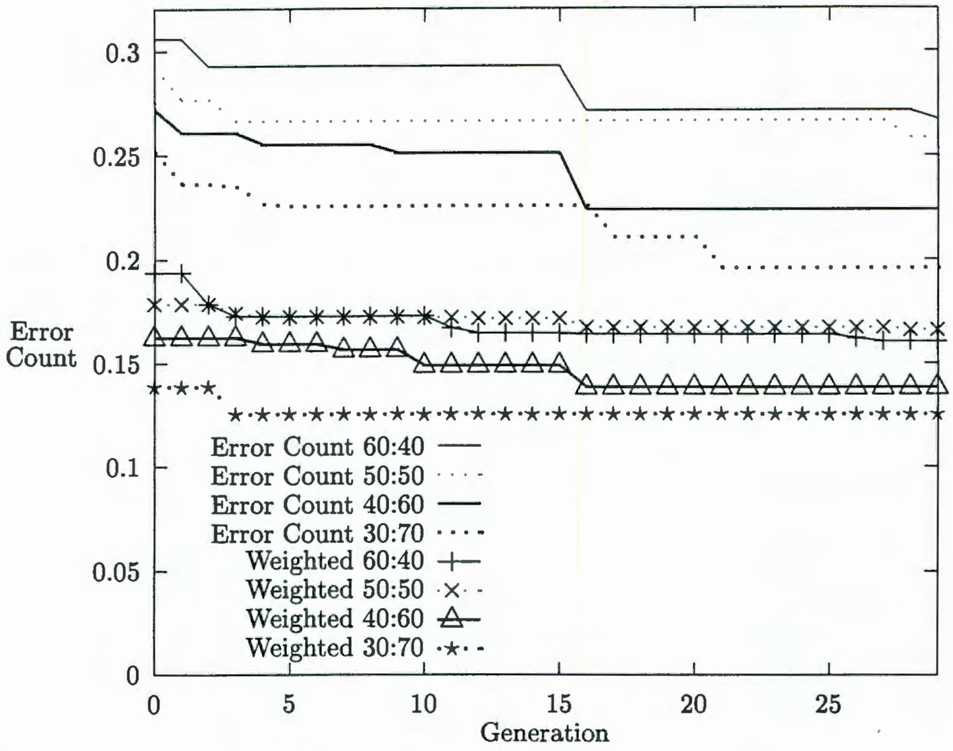


Figure 8.10: Graph of classification performance improvement during optimisation using a Genetic Algorithm. Two sets of curves are obtained by using the two pattern error functions h^{EC} and h^{WE} for error counting and weighted error counting respectively. In each case, the relative weight shown is given to errors in the non-rock and rock classes respectively. It can be seen that increasing the weight in favour of correct rock class classification results in a lower cost of classification error which indicates that the classification process is biased towards correctly classifying the rock class at the expense of the non-rock class. Note that due to the different pattern error functions, the two sets of curves cannot directly be compared.

Type		Shape				Boundary				Split		Merge					k
Counted	60:40	1	1	1	1	0	1	1	1	0	0	0	0	1	1	0	15
	50:50	0	0	1	1	1	1	1	1	0	0	1	0	1	0	0	29
	40:60	1	1	1	0	0	1	1	1	0	0	0	0	1	1	0	23
	30:70	1	1	1	0	0	1	1	1	0	0	0	0	1	1	0	31
Weighted	60:40	0	0	0	1	1	1	0	1	1	0	1	1	0	0	0	7
	50:50	1	0	1	0	1	1	0	1	0	0	1	0	1	0	0	25
	40:60	1	1	1	0	0	1	1	1	0	0	0	0	1	0	0	31
	30:70	1	0	1	0	1	0	1	0	0	0	1	0	1	0	0	27

Table 8.1: Optimisation results for each of the evaluation functions. The optimum selection of features is shown for each method with the best value for k.

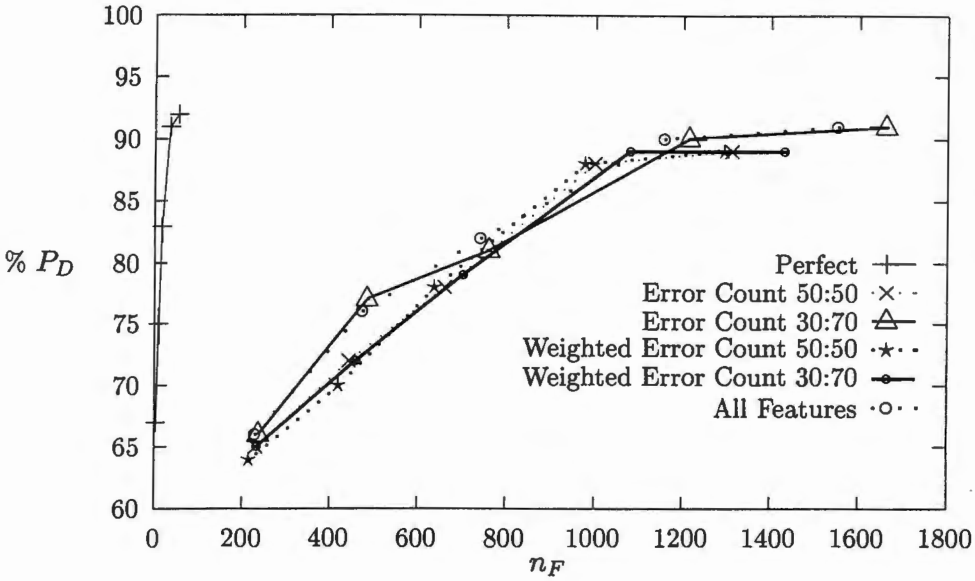


Figure 8.11: ROC curves of classification performance while varying t_{clip} from 0.5 to 0.7. Perfect classification is achieved using a list of rocks in the image and shows the limit that is imposed by the preattentive vision stage. It can be seen that using Weighted Error Counting for optimisation provides the best results. Note that varying the relative cost of errors does not have a particularly strong effect on the overall performance of the system.

The variable is t_{clip} which determines the clipping level after the Hough transform in the preattentive stage. A curve is shown for perfect classification. This is achieved using a list of rocks in the image and using the distance measure between the current ellipse and the closest rock to classify the blob. This shows the limit that is placed on the system by the preattentive stage given the particular value of t_{clip}. Note that some false alarms have occurred during perfect classification as a result of quantisation errors. These errors all correspond to marginal cases and are not important.

Four curves are also shown for optimal feature sets that were selected using four of the evaluation functions as indicated. Finally, a curve is shown in which no optimisation has taken place and all the features are used. It can be seen that for none of the optimised cases is there a significant improvement over the performance of the unoptimised classification.

Although there is clearly still a considerable number of false alarms in the list of possible rock positions, it can be seen that classification is achieving the goal of reducing the number of false alarms dramatically without adversely affecting the true detection proportion. In order to improve on these results, it is necessary to reduce the number of misclassified non-rock samples while maintaining the current level of rock class acceptance.

The use of a genetic algorithm does not guarantee that the absolute minima will be found which is usually the goal, but it does ensure a good result in a reasonably short period of time. In fact, finding the global maximum does not necessarily correspond to the best result, since the optimisation is only applicable for a particular pair of training and test data. In addition, even if overall classification accuracy could be optimised, what is actually required is optimum performance of the system as a whole which may operate more successfully under some error conditions than others. The results that are achieved here can therefore not be assumed to represent the best possible settings for these parameters, but should only provide a guideline for the selection of features.

A possible explanation for the failure of optimisation to improve performance significantly and the relatively high error in the non-rock class is that the training and test data are not correctly classified. More specifically, it is possible that the distance threshold of 1 in (8.1) and (8.2) is too large. The result of this could be that the rock class forms a cluster in the feature space which is correctly labelled in the training set. Furthermore, a proportion of the non-rock samples are misclassified with the result that the cluster is enlarged and the non-rock error is increased without compromising the rock class error.

8.7 Summary

A feature vector classification process is described which is used to produce a list of probable targets L_{class} by filtering the list $L_{highlight}$ to remove false alarms, i.e. those targets at which a rock is not actually present.

Fifteen features are measured and feature sensitivity analysis has been used to determine which subset p_{opt} provides optimum performance with respect to a test set which is representative of the actual samples which must be classified.

It is found that the number of false alarms can be dramatically reduced with minimal impact on the number of true detections; however, there is still a large proportion of false alarms and it is necessary to analyse the remaining lists of targets from each level of the pyramid in a hierarchical procedure. This problem in tackled in the next chapter.

Chapter 9

Consolidation of Multiscalar Results to Separate Rocks

9.1 Introduction

The aim of the hierarchical analysis which is described in this chapter is to determine the final list of rocks in the image from the results of the classification stage for each level of the image pyramid. This stage is necessary because the preattentive vision process typically highlights more positions in the image where there are not actually rocks than where there are rocks. The classification process which was described in the previous chapter was aimed at identifying which highlighted points in the image do correspond to rocks. Unfortunately, although the classification stage removes a large proportion of the false alarms from the list of possible rocks, there still remains the possibility of multiple hits on each rock and further false alarms.

False alarms occur most frequently in the situations where a group of rocks is merged together, or a single rock is split at relatively high resolution such that its corners appear as individual rocks. The overall result of these factors is that there are many extraneous detections of rocks in the image pyramid which must be removed before a size distribution can be calculated. Figure 9.1 gives an example rock scene in which all the detected rock positions have been highlighted using ellipses. The proliferation of ellipses gives an indication of the extent of the problem.

The use of tree structures, which have been used for decision making and hierarchical representation, is proposed as a solution to this problem. The following few examples show a wide range of applications in which a tree structure of some form has been successfully used.

Breiman, *et al* [8] use decision trees for classification and regression problems. Cheng and Lu [12] have used a tree representation for matching waveforms. This application is typical of hierarchical representation schemes where the main branches of the tree represent coarse structure and the leaves represent the finest detail.

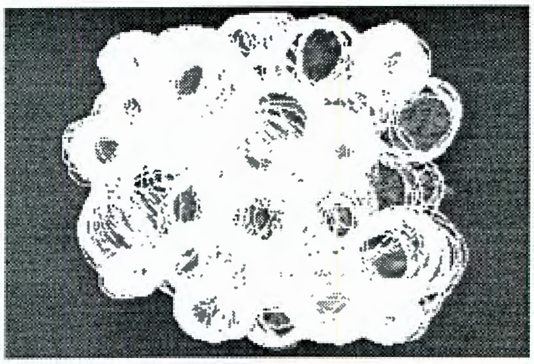


Figure 9.1: Rock image with all detected positions highlighted using ellipses. This result was obtained using the following values: $t_{edge} = 0.75$, $t_{clip} = 0.55$, $t_{prob} = 0.65$. All features were used for classification.

Shapiro and Haralick [83] define structural descriptions of objects and use a manipulation of a tree representation to perform inexact matching. The form of descriptions used in this application is a tree structure of object components, for example: a chair consists of six parts; a back, seat and four legs. The various parts can themselves be described in successively greater detail.

Trees have also been used in image processing applications. Beaulieu and Goldberg [4] define a possible form of hierarchical structure in an image. In their definition, an image can be divided into parts which could represent different objects. These parts can in turn be subdivided into subparts which represent components of the objects, and so on.

Multiresolution image processing techniques often attempt to achieve a global view of a scene by viewing it at several scales. This process is usually accompanied by the use of a tree structure for representing the information in the multiresolution hierarchy. For example, Lifshitz and Pizer [48] generate a tree representation of an image for the purpose of an interactive display. The image is described in terms of nested light and dark regions in a multiresolution stack of images formed by successively smoothing the starting image. The level of the tree that is selected determines how much detail will be used when segmenting that portion of the image in a display.

Crowley and Parker [16] and Crowley and Sanderson [17] generate a tree representation for two-dimensional grey-scale shapes in images. The representation is constructed from peaks and ridges in an array of bandpass images created using the difference of low-pass transform. The descriptions of shapes which are thus encoded are invariant to size, orientation and position and can be used for matching different shapes.

A tree representation provides advantages in the multiresolution framework as it forms a compact representation which reduces computational complexity because it only presents significant detail at any stage of processing.

Due to the suitability of tree representations for multiresolution systems, their use has been adopted for solving the multiple detection problem. This is achieved by analysing all the detections to create tree representations. The trees are then compared and analysed to determine the most probable collection of rocks in the image. Discrimination is based on a method of representation normalisation and a weighting function.

The classification stage described in the previous chapter resulted in a list L_{class} of detected points for each level of the pyramid. Each point in the list is stored with the following information: $L_j = (\mathbf{p}, e, \theta, h, p)$ where

$\mathbf{p}$ = The coordinates of the centre of the ellipse
e = The aspect ratio of the ellipse equal to $\frac{a}{b}$
θ = The ellipse orientation
h = The Hough transform peak value
p = The average probability of edge points in the binary image

The lists of points are used to generate the tree structure which represents the appearance and disappearance of rocks in the scale structure of the image. The tree structure may be complicated since there can be more than one point corresponding to each rock which will lead to multiple branches at that level of the tree. Each tree represents a possible rock in the image. All the trees are processed to form a normalised representation before ranking and deleting of overlapping trees to result in a final list of rocks contained in the image.

The remainder of this chapter describes the various stages of tree growing, processing and final rock size distribution calculation. Section 9.2 explains how the tree structures are composed from the lists of detected points. Section 9.3 describes an algorithm for calculating a weighting factor for a tree to facilitate the comparison of trees. The processes of normalising the tree representation by pruning and determining the parameters of the rock represented by the tree are given in Sections 9.4 and 9.5 respectively. Tree filtering is used to delete overlapping trees based on their scores and is described in Section 9.6. Section 9.7 explains how a final size distribution is calculated and presented, Section 9.8 examines the performance of this stage and finally, Section 9.9 describes an efficient approach for implementing these tree algorithms.

9.2 Growing Trees

The first stage of the multiscalar analysis process involves analysing all the detected points in each of the L_{class} lists corresponding to each level of the image pyramid to grow trees. Each list L_{class} of points becomes a list of tree nodes L_{node} where each node N contains the following information: $N = (\mathbf{p}, e, \theta, h, p, D)$ where D is the size of the object.

Section 9.2: Growing Trees

In order to facilitate comparison of points and nodes from different levels of the image pyramid, it is necessary to convert them from an image pyramid to a scale-space coordinate system. This is illustrated in Figure 9.2. Note that all detected points are represented by ellipses which are all the same size (with respect to area) in the image pyramid; their size D is a function of the expected rock size $r_{expected}$. The ellipses correspond to progressively larger rocks in the higher levels of the pyramid and this is shown in the scale-space stack by increasingly large ellipses. The ellipses have been labelled to facilitate identification. All the points or nodes are converted to a normalised spatial representation in the scale-space stack. This results in a modification of their position $\mathbf{p}$ and their size D. Given the current level of the image pyramid is i and the scale factor between adjacent pyramid levels is f_{scale} (see Section 6.2), then $\mathbf{p}$ and D must be multiplied by f_{scale}^{-i} to convert to the scale-space coordinate system.

Each node will become the root of its own tree and may then grow downwards through scale-space. It may also only have branches on the next level of scale-space. Furthermore, only nodes on the next level which are less than a distance t_{branch} from the root node will be accepted as branches. The distance is measured using the distance measure described in Appendix D which finds the distance between two nodes based on their relative values of $(\mathbf{p}, e, \theta, D)$. The aim of t_{branch} is to adjust the sensitivity of the tree growing routine to result in the most representative trees possible. Figure 9.3 shows each node in scale-space as a grey spot. On the level beneath each node is a possible region of acceptance for branch nodes based on the distance measure. Note that given the distance measure used, the size of the ring would vary for different combinations of ellipse shape.

Because the distance measure is defined to have a value less than 1 if the two nodes are close enough to represent the same rock, the value of t_{branch} should be less than 1. A ROC curve is used in Section 9.8 to determine a suitable value of t_{branch}.

A limit is placed on the maximum depth to which a tree may grow. This is in the form of a limit of the range of scales that may be occupied by the tree. A scale change of f_{limit}^2 is used in Appendix D as a limit on the possible size range over which any rock could be detected in the same place as the same object. The number of levels which this corresponds to can be calculated by solving the exponential pyramid level size equation, $f_{limit}^2 = 1/f_{scale}^n$, for n to give $n = -2 \ln f_{limit} / \ln f_{scale}$.

As a result of applying the rules described above, the trees which are grown will be expected to have the following characteristics:

- There are as many trees as there are nodes.
- The root of a tree occurs at the highest pyramid level and it grows downward toward the base of the pyramid.
- Each node may have multiple branches to the level below.

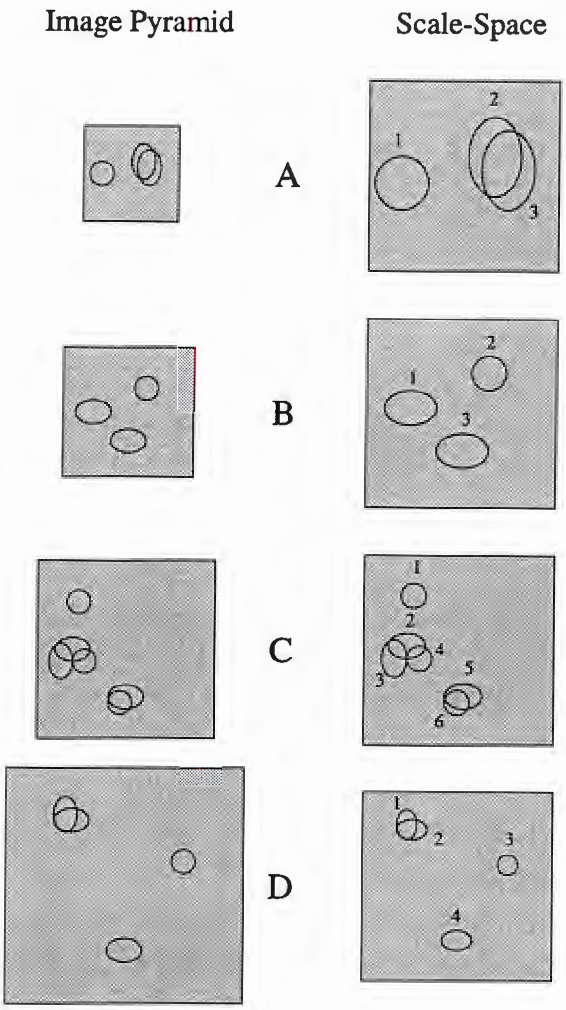


Figure 9.2: Rock positions are detected in an image pyramid and the results are converted to a scale-space stack for the generation of a hierarchical representation. Each detected rock position is represented by an ellipse which is labelled on each level in the scale-space stack.

- The depth of any tree is limited to a scale change of f_{limit}^2.

A raw tree is shown in Figure 9.3 with all the possible branches that can occur in the generation of trees for that example. Note that each node also represents the root of its own tree. This is necessary to allow for positions where several detections have been made at the same position $\mathbf{p}$ at various scales. If rocks have been merged, only considering a tree rooted at that node and none below it may result in incorrectly detecting a group of merged rocks as a single rock. With the current system, the tree with the best score will be kept. Tree score should be higher when a rock is actually present in the image since it will be better matched to the blob detection process. The method used for calculating the score for a tree is described in the next section.

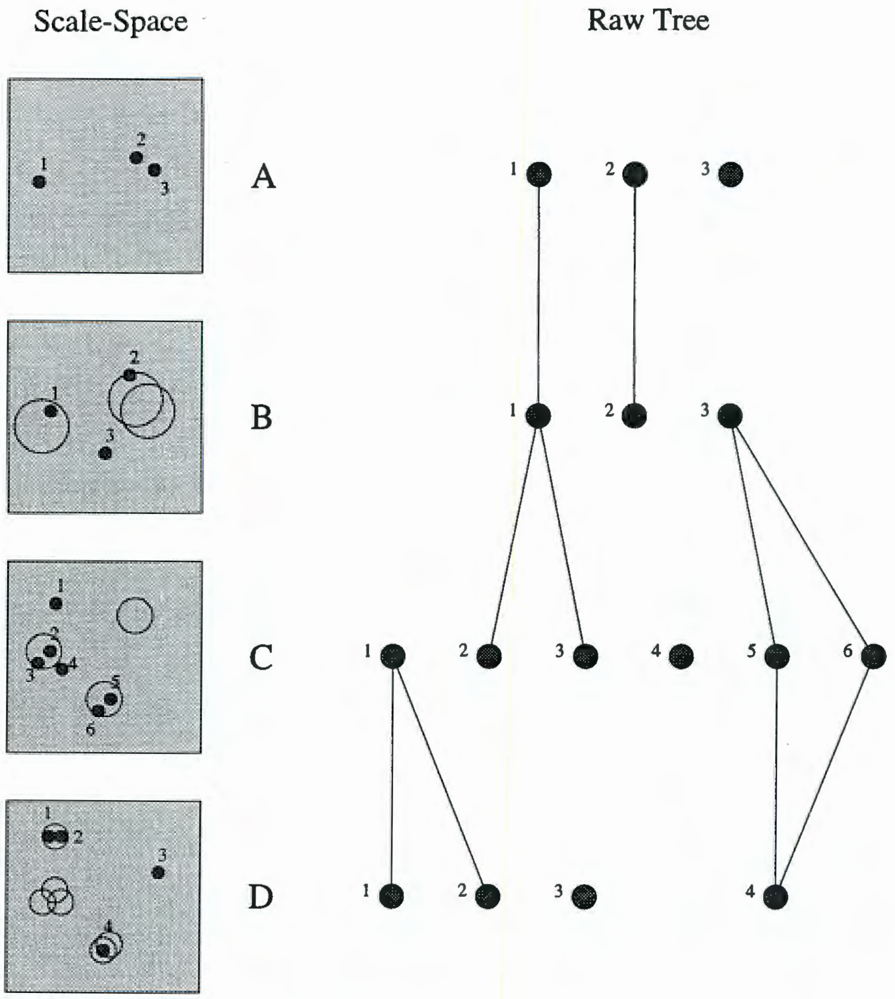


Figure 9.3: A region of acceptance for each node is shown on the level beneath it in scale-space. The tree links are made where a node on a level lies less than a distance t_{branch} from a node on the level above it.

9.3 Scoring Trees

The aim of measuring a tree score is to provide a means by which trees and subtrees can be compared. This is achieved by determining desirable and undesirable features which a tree may have and weighting them appropriately.

Ideally, a perfect tree should be present where there actually is a rock and should have the following characteristics: detected points or nodes in several adjacent scale levels which are spatially close and have similar ellipses, and the ellipse should be well matched to the object in the image.

The actual trees which were grown in the previous section do not necessarily have these properties and can exhibit some other behaviour which is undesirable such as: tracing a devious route through scale-space, having multiple branches on any level and having different ellipse parameters between adjacent levels. The tree score is required to provide a means for comparing different trees which occupy the same position to determine which best represents a rock.

The Hough transform value h and the average edge probability value p which were obtained during the two levels of preattentive vision give an indication of the match of the ellipse to the image content at that point. High values for both h and p are desirable since that indicates that the ellipse is well matched in the image. The tree score is increased by adding these two values.

The distance measure gives an indication of the differences in size, position, aspect ratio and orientation of two ellipses and this can be used to bias against bad continuity through scale-space. Subtracting the distance from the tree score will decrease the score more where the ellipses are not similar. Because it is possible to have multiple branches, the cost of the differences can be normalised by determining the average distance from the root node to each of the branch nodes.

The score of a tree with multiple layers can be increased by adding the scores of each branch node to the root node score. Again, the average should be calculated for normalisation so that a tree with many branches can be compared to one with fewer branches. Adding the branch score should increase the overall score since all branch distances, and therefore the average distance, must be less than t_{branch} which should not be much more than 1. Also, both h and p have lower bounds of t_{clip} and t_{prob} respectively which are both greater than 0.5. Recall that t_{clip} is the lower bound on acceptable peak values after the Hough transform in Section 6.3.3 and t_{prob} is the lower bound on the average probability of detected edge points in the second stage of preattentive vision (Section 7.6). It should therefore not be possible for the score of any node to become negative, and adding the branch scores to their root node should increase its score.

A recursive procedure for calculating a tree score is presented in algorithmic form in Figure 9.4. An example calculation for a small tree is shown in Figure 9.5.

It is also necessary to calculate the score for each branch of a node for the purpose of determining the best normalised core representation of the tree. This is achieved by calculating a tree score but assuming that the root node has only the one branch node. An algorithm for calculating branch score is also shown in Figure 9.4.

9.4 Normalising Trees by Pruning

As stated previously, trees can be created with undesirable properties. The aim of this section is to provide a procedure for tackling the problem of trees containing multiple branches from any of its nodes. The result of this will be a normalised tree representation in which branches

Given a root node of a tree, $N = (\mathbf{p}, e, \theta, h, p, D, B)$ where

$\mathbf{p}$ = Spatial position of the node
e = Ellipse aspect ratio
θ = Ellipse orientation
h = Hough transform weight value of the node
p = Average edge point probability
D = Size of the current node
$B = (B_1, \ldots, B_n)$ = Branch nodes of the current node

then the tree score of a node is calculated using:

```
Function: Tree_Score($N$)
        score = $h + p$
        if($n > 0$)
            score -= $\frac{1}{n}\sum_{i=1}^{n}$ Distance($N, B_i$)[a]
            score += $\frac{1}{n}\sum_{i=1}^{n}$ Tree_Score($B_i$)
        return score
```

and the branch score for branch i of node N is calculated using:

```
Function: Branch_Score($N$,$i$)
        score = $h + p$
        if($n > 0$)
            score -= Distance($N, B_i$)
            score += Tree_Score($B_i$)
        return score
```

---

[a] Distance() is a function which takes the ellipse parameters of two nodes and calculates the distance between them based on the procedure described in Appendix D

Figure 9.4: Recursive algorithm for calculating tree score.

have been pruned to leave a central branch which has the highest strength in terms of the tree score determined in Section 9.3. Tree pruning is a standard tree manipulation procedure whereby a branch is removed from the tree [8].

The branch score algorithm which was described in the previous section is used at each node of the tree to find the best branch from that level and all other branches are deleted or pruned. The following simple recursive algorithm, which starts at the root node, describes a computationally cheap approximation to this procedure.

1. Determine which branch of the current node has the highest branch score and delete all other branches.
2. Apply the pruning algorithm to the remaining branch until there are no branches.

In the ideal situation described earlier, it was stated that nodes should be detected at the position of the rock over a limited range of scales. The result of the pruning algorithm is that the best branch is kept at each level of the tree to give a final list of nodes, one for each

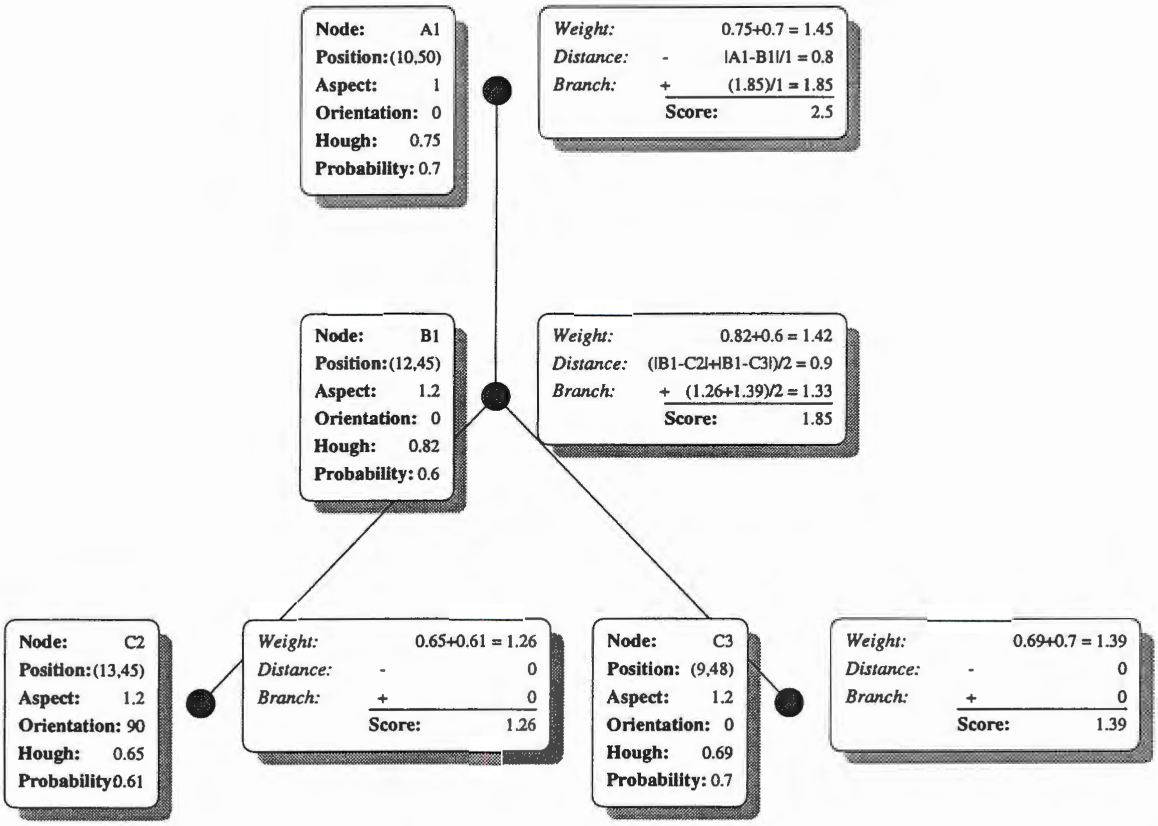


Figure 9.5: Example tree score calculation for the four node tree shown in Figure 9.3. Note that values for h and p have been fabricated for each node; however, the position, orientation and aspect ratio are based on the ellipses shown in Figure 9.2.

level. These nodes provide the straightest tree with the least possible deviation in position and ellipse parameters and thus allow for a more accurate qualitative comparison of tree scores. Figure 9.6 shows all the pruned trees that would be obtained from the tree structure in Figure 9.3. The score is shown at the base of each tree. Note that since each node is the root of a tree, none of the nodes have been irrevocably removed from the lists yet.

A more exhaustive algorithm which could be used instead is:

1. Apply the pruning algorithm to all the branches of the current node.
2. Determine which branch of the current node has the highest branch score and delete all other branches.

9.5 Extracting Rocks from Trees

The aim of this section is to determine the parameters of the rock that is being represented by each of the normalised tree to produce a list of rocks L_{rock}. It is assumed that the depth

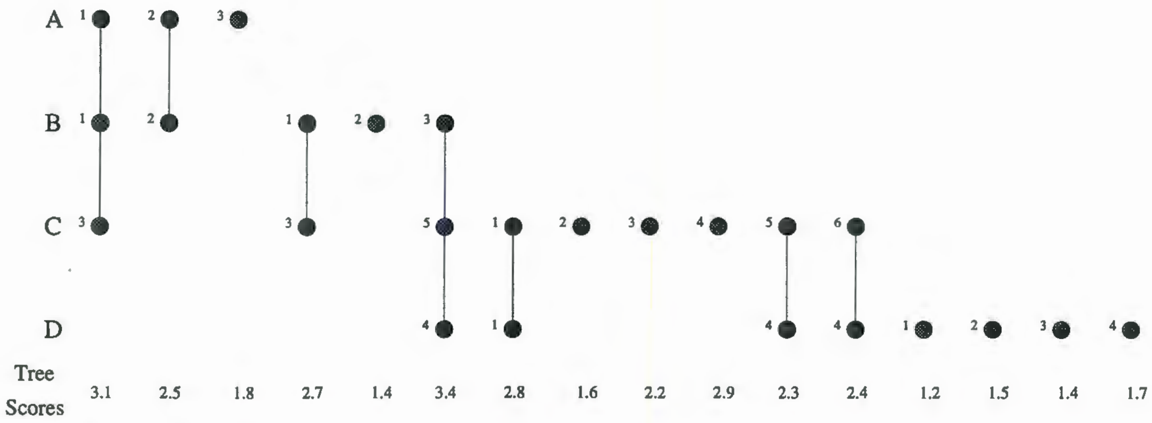


Figure 9.6: All remaining linearised trees after the pruning algorithm. The example score for each tree is given beneath it. Note that the score for the root node will usually increase when the tree is pruned.

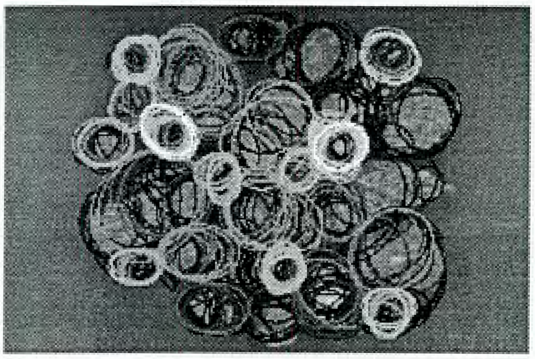


Figure 9.7: Rock image showing all possible rock positions encoded in the normalised trees. The intensity of each ellipse gives an indication of the relative strength of each rock position determined using the tree score. Trees were grown using $t_{branch} = 0.7$.

of a tree has been limited and then assigns the parameters of the middle node in the list to the rock. This allows compensation for detecting rocks at larger and smaller scales than the correct one.

Each rock in the list will have the following information: $L_j = (\mathbf{p}, e, \theta, s, D)$ where s is the tree score value. This is illustrated in Figure 9.7 which shows the rocks associated with each of the normalised trees by means of a shaded ellipse. The ellipses are shaded according to the score of each rock where white corresponds to the highest score.

9.6 Filtering to Find the Final Tree List

Filtering is the final stage of hierarchical analysis that is required to produce a list of rocks in the image. It uses knowledge in the form of rules and the tree score associated with each rock to remove rocks from the list L_{rock}. The aim of the process is to result in a list of rocks which is physically realisable and for which a size distribution can be calculated.

Knowledge is encoded in the form of the rule that two rocks in the image cannot overlap. If this rule is broken for any pair of rocks, one of the rocks must be deleted. The tree score associated with each rock is used to decide which rock should be deleted. Since the higher tree score should correspond to a tree which is more likely to be a rock, the rock with the lower score is deleted.

The amount of overlap is measured using the ellipse parameters which are associated with each rock. Given ellipse aspect ratio e_1 and size D_1, the major and minor axes can be calculated as: $a_1 = D_1\sqrt{e_1}$ and $b_1 = \frac{D_1}{\sqrt{e_1}}$. The angle from the centre of the first ellipse to the centre of the second can be calculated as $\psi_1 = \arg(\mathbf{p}_2 - \mathbf{p}_1)$. The ellipse parameter value that corresponds to ψ_1 is calculated as ϕ_1 using

$$\phi_1 = \arctan\left(\frac{a_1\sin\theta_1\cos\psi_1 + a_1\cos\theta_1\sin\psi_1}{b_1\cos\theta_1\cos\psi_1 - b_1\sin\theta_1\sin\psi_1}\right) \tag{9.1}$$

which is derived in Appendix C.4.

The distance from the centre of the first ellipse to its boundary in the direction of the second ellipse is equal to

$$r_1 = \sqrt{(a_1\cos\theta_1\cos\phi_1 + b_1\sin\theta_1\sin\phi_1)^2 + (b_1\cos\theta_1\sin\phi_1 - a_1\sin\theta_1\cos\phi_1)^2} \tag{9.2}$$

which is derived in Appendix C.2. The value of r_2 for the second ellipse can be calculated similarly.

The first rock is defined as being too close to the second rock if the ratio of the amount of overlap to r_1 is greater than the fraction $f_{overlap}$. This can be expressed by the following equation which is satisfied when there is too much overlap:

$$\frac{r_1 - (|P_2 - P_1| - r_2)}{r_1} > f_{overlap} \tag{9.3}$$

where $f_{overlap}$ is a filter overlap factor which specifies the proportion of rock overlap which is allowed. Some degree of overlap must be allowed to account for close proximity between rocks which are not accurately modelled by an ellipse. Figure 9.8 illustrates the relationship between the two ellipses and the amount of overlap. It is not immediately apparent what value should be used for $f_{overlap}$ and so ROC curves are used in Section 9.8 to determine a suitable operating point.

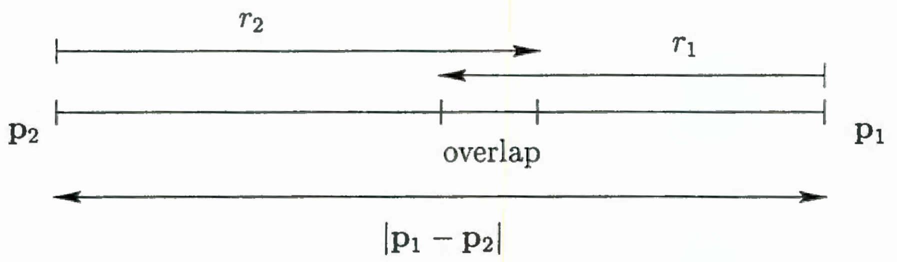


Figure 9.8: Diagram of the amount of overlap between two ellipses centred at $\mathbf{p}_1$ and $\mathbf{p}_2$.

```
Given a list of rocks, $L_{rock} = (\mathbf{p}, e, \theta, s, D)$ where

        $\mathbf{p} = (\mathbf{p}_1, \ldots, \mathbf{p}_n)$  =  The position of each rock
        $e = (e_1, \ldots, e_n)$  =  The ellipse elongation for each rock
        $\theta = (\theta_1, \ldots, \theta_n)$  =  The ellipse orientation for each rock
        $s = (s_1, \ldots, s_n)$  =  The tree score for each rock
        $D = (D_1, \ldots, D_n)$  =  The size of each rock

Function: Filter_Trees( $L_{rock}$)
    Sort $L_{rock}$ into descending order of $s_i$
    for( $i = 2, 3, \ldots, n$)
        for( $j = 1, 2, \ldots, i - 1$)
            Calculate $r_1$ and $r_2$ for $(\mathbf{p}_i, e_i, \theta_i, D_i)$ and $(\mathbf{p}_j, e_j, \theta_j, D_j)$
            if( $(r_1 - (|P_1 - P_2| - r_2))/r_1 > f_{overlap}$)
                Delete rock $j$ from $L_{rock}$.
                Break out of inner loop.
```

Figure 9.9: Algorithm for filtering tree list.

An algorithm for filtering the list of rocks is given in Figure 9.9.

Figure 9.10 shows the result of filtering on the input image of Figure 9.1. Although the final result is not perfect, the majority of the rocks have been correctly identified and the measured size distribution should therefore be representative of the actual distribution of rocks in the image.

9.7 Calculating Size Distributions

It was found in Section 2.1.3 that the desired measure of rock size distribution for this system is a rock volume distribution since it is simple to infer a mass distribution from this given the density of the material. Since only the projected area of rocks that are actually visible at the surface of the pile on the conveyor belt can be measured using machine vision, the volume distribution must be estimated based on the rock area measurements.

In general, much of the material on the conveyor belt will actually be buried in a pile

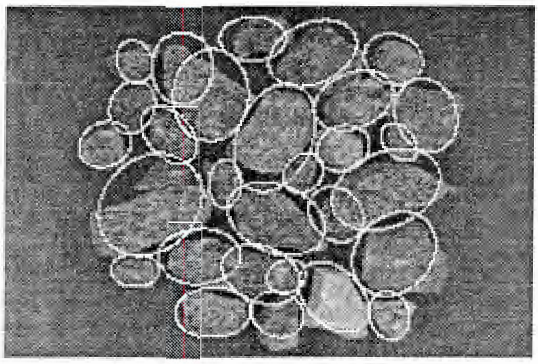


Figure 9.10: Rock image with detected rock positions highlighted using ellipses after tree filtering. An overlap factor of $f_{overlap} = 0.9$ was used to obtain this result.

which makes measurement impossible for the hidden material, and makes volume estimation difficult. Lange [46] has investigated this problem and found that while a stereological sizing technique based on an idealised particle shape (i.e. spherical or cubic) can be developed, when the particles are of arbitrary shape, modelling the transformation from two- to three-dimensions becomes extremely difficult. *Stereological* methods are concerned with the determination of three-dimensional parameters from measurements taken in two dimensions [56].

Although such work is beyond the scope of this thesis, a simpler approach has been taken for calculating size distribution which will give representative results under the circumstances where the material is not packed in a pile and all the rocks are sufficiently separated to be virtually unobstructed. The approach is based on the measurement of rock area and extends the elliptical rock model to an ellipsoid.

A method for estimating the cumulative mass curve was given by (2.3) in Section 2.1.4. That method was found to be particularly error prone since it relied on raising an equivalent spherical diameter D to the third power in order to estimate volume. In order to improve this situation, another method for estimating volume or mass is given here. This relies on measuring the area of the object and then extrapolating from this two-dimensional value to the three-dimensional volume estimate using the projected shape of the object.

The topics of area and volume distribution calculation are explained in the following two sections.

9.7.1 Area Distributions

An estimate of rock area used to calculate d_a (diameter of an equivalent sphere based on projected area) is stored with each rock in L_{rock} in the form of the ellipse representation. All ellipses have the same area $A_{ellipse} = \pi r_{expected}^2 = \pi ab$ at the base of the pyramid. Given this information, the level, l, of the pyramid where the rock was found and the pyramid level

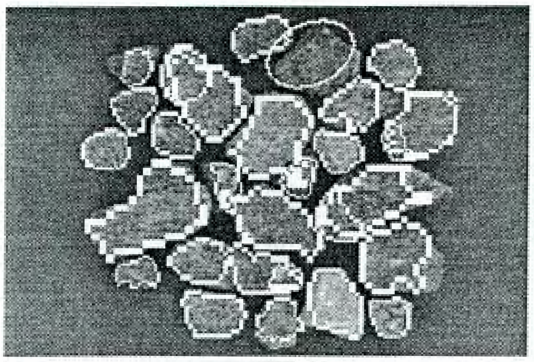


Figure 9.11: Detected rock positions highlighted with their respective boundaries determined using the blob detection algorithm. Note that a rock at the top of the scene is outlined with a ellipse. This is due to quantisation errors which resulted in the system being unable to extract the appropriate region of interest of that rock because it is too close to the edge of the image. In that case, the elliptical approximation to rock size is used.

reduction factor, f_{scale}, then the elliptical approximation to rock area is

$$\text{Elliptical Area} = A_{ellipse}(f^{l}_{scale})^2 \tag{9.4}$$

This area measurement has units pixel^2. In order to convert this to a measurement in cm^2 for representation purposes, it is necessary to know the image magnification factor, m, described in Section 3.3.1 and the size of a pixel, T_β, defined in Section 5.3.3. The modified form of (9.4) with units cm^2 is

$$\text{Elliptical Area} = A_{ellipse}\left(\frac{100 f^{l}_{scale} T_\beta}{\text{m}}\right)^2 \tag{9.5}$$

where the ratio T_β / m is the size of a pixel (in metres) on the object plane and the factor of 100 is used to convert the units from metres to centimetres.

A more accurate area measurement is obtained by using the blob segmentation algorithm described in Chapter 7 to identify the actual boundary shape and therefore measure the area, A_{rock}, for each of the rocks individually. Figure 9.11 shows the result of finding rock boundaries and superimposing them on the image. Note that because all rocks are detected at the same size, the boundaries of the larger rocks have to be enlarged to the correct scale which results in their blockiness. The actual rock size is then calculated using the same procedure that was used for the elliptical estimate resulting in the following formula:

$$\text{Rock Area} = A_{rock}\left(\frac{100 f^{l}_{scale} T_\beta}{\text{m}}\right)^2 \tag{9.6}$$

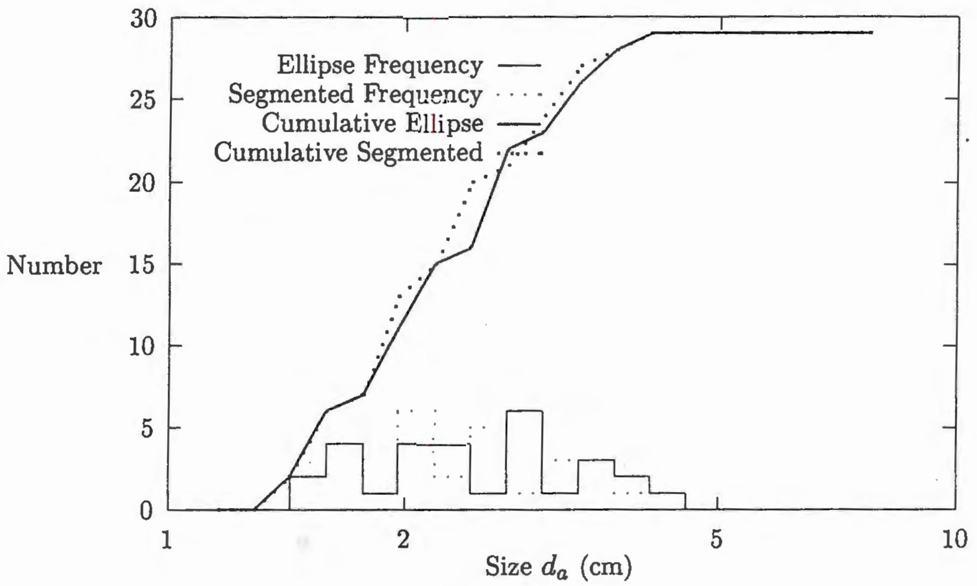


Figure 9.12: Graph of the number frequency histogram for the sample image. A histogram is shown for both the ellipse size measurement and the segmented size measurement.

Rock size number distributions which were calculated using the elliptical approximation from Figure 9.10 and the more accurate area calculation from Figure 9.11 are shown in Figure 9.12 for comparison. For each case two curves are drawn. The first curve shows the number frequency of rocks in each size bin and the second shows the number frequency cumulative distribution.

Figure 9.13 shows rock size percentage area distributions for the two methods of calculating rock area. The first curve shows the percentage of material by area in a particular size bin and the second curve shows the cumulative percentage area of material less than or equal to the particular size. The graph is plotted as percentage versus the diameter of a circle which has the same area as the current bin size.

Note that the pairs of curves generated using the two methods for area calculation do correspond reasonably closely. The cumulative curves essentially represent the integral of the histogram and this form of logarithmic cumulative distribution with its characteristic S shape is a standard form of representation for size distribution results in mineral processing and quarrying [38, p24]. This graph represents the maximum detail that can be obtained from the image analysis process. It is not necessary to display all the detail and the graph could be modified to show only material in user specified ranges.

9.7.2 Volume Distributions

A volume distribution is of interest to the mining industry because it is related to a mass distribution by a material density factor (See Section 2.1.4) and the mass distribution influences

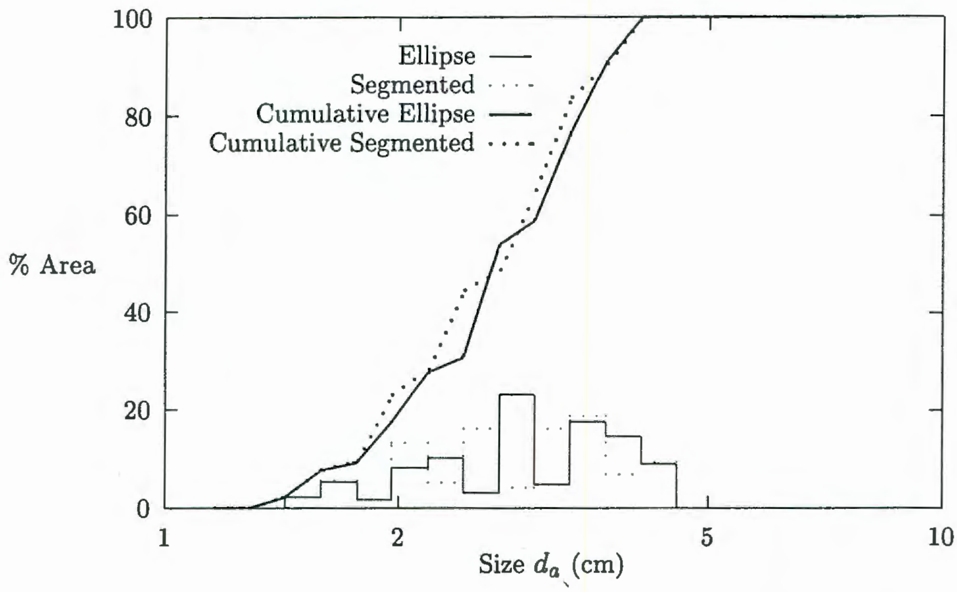


Figure 9.13: Graph of area size distributions for the sample image. Both a % area histogram and a cumulative % area curve are shown for the elliptical area approximation and blob segmentation area.

the performance of an autogenous mill. As stated in Chapter 2, using the measure of object shape and area instead of a characteristic diameter will result in considerably less error in the volume because extrapolation is being performed from two- to three-dimensions instead of one- to three-dimensions.

If the rock material is sparsely spread out on the conveyor belt, it is possible to find a volume distribution by estimating the volume of each of the rocks individually. A comparison of Figures 9.10 and 9.11 give reasonable justification for modelling rock shape using ellipses since they correspond quite closely to the detected boundaries. It should therefore be possible to extend the elliptical model to three-dimensions by modelling rocks as ellipsoids for the purpose of calculating volume. This could easily be done by extending the ellipse to an ellipsoid with parameters a, b and c where a and b are known and c must be determined. In general, material on the belt has settled in a most stable position and it is therefore not very unreasonable to assume that c will be a fraction of the minor radius, b, such that $c = f_{depth}b$ where f_{depth} is a depth proportion factor which could be adjusted to give acceptable results. The individual rock volume could be expressed as

$$\text{Ellipse Volume} = \frac{4}{3}\pi abc\left(\frac{100 f^{l}_{scale} T_{\beta}}{\text{m}}\right)^{3} \tag{9.7}$$

In a similar manner, the rock boundary could be used to calculate a volume of revolution around the major axis of the ellipse to result in a volume $V_{revolution}$. The depth proportion

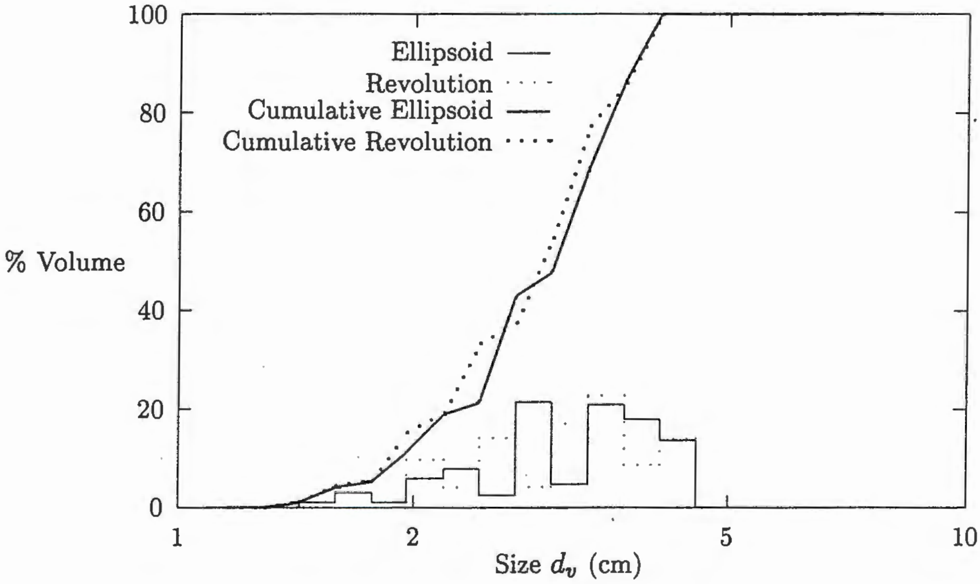


Figure 9.14: Graph of volume size distributions for the sample image. Both a % volume histogram and a cumulative % volume curve are shown for the ellipsoidal and blob segmentation volume approximation.

factor could be incorporated to give a rock volume estimate $V_{rock} = f_{depth} V_{revolution}$ and scaled volume of

$$\text{Rock Volume} = V_{rock} \left(\frac{100 f^{l}_{scale} T_{\beta}}{\text{m}} \right)^{3} \tag{9.8}$$

Because it is doubtful how accurate this method would be and there are several complexities involved in the calculation, the following simple method has been used to calculate volume given the measured area A_{rock} after segmentation. The equivalent diameter d_a is determined as $d_a = 2\sqrt{\frac{A_{rock}}{\pi}}$ then the volume is calculated as

$$\text{Rock Volume} = f_{depth} \frac{1}{6} \pi d_a^3 \left(\frac{100 f^{l}_{scale} T_{\beta}}{\text{m}} \right)^{3} \tag{9.9}$$

An equivalent graph to Figure 9.13 for rock volume measurements is shown in Figure 9.14. Since it is not generally possible to assume that rock material is spread out on the conveyor belt and due to the inaccurate nature of the volume estimation procedure, the remaining results in this thesis will be given in terms of area measurements since that is the actual result that is obtained.

9.8 Performance Evaluation

Figure 9.15 shows ROC curves of the performance of the complete system as the value of

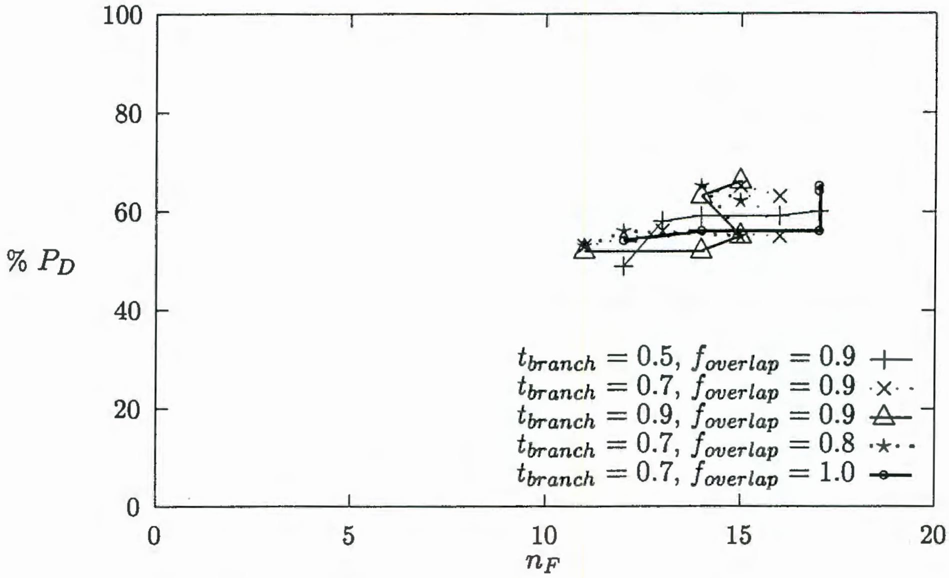


Figure 9.15: ROC curves for the performance of the complete system for five combinations of t_{branch} and $f_{overlap}$ based on results from three images. The following parameters have been used: $t_{edge} = 0.75$, $t_{prob} = 0.65$, all features for classification and t_{clip} in the range 0.5–0.7. Due to the nature of the hierarchical analysis procedure, the usual structure of an ROC curve is not apparent. It can, however, be seen that the number of false alarms has been reduced from around 1000 after classification to a low value around 15. This has occurred at the cost of a decrease in the percentage of true detections. This curve does not differentiate between the misses of various degrees and is therefore not particularly informative about the actual performance of the system.

t_{clip} after the Hough transform is adjusted. The five curves are obtained for five combinations of the variables t_{branch} and $f_{overlap}$ as shown. It can be seen that at this stage, the ROC curve does not provide a particularly useful indication of the performance of the system when measured in the same manner as previously. This is because no information regarding the relative "badness" or cost of the false alarms is considered.

An indication of the cost of the false alarms can be obtained by measuring their distances from the nearest rock and compiling a distribution. Figure 9.16 shows this result for the sample image results shown in Figure 9.10. True detections correspond to a distance of less than 1 while the severity of a false hit increases with increasing distance. It can be seen from the distribution that the measured size distribution should be accurate since a very large proportion of the hits are either true or very close. Recall that t_{branch} is the region of acceptance of a node during tree growing (Section 9.2) and $f_{overlap}$ is the amount of overlap that is allowed for adjacent ellipses to compensate for inaccuracies in the fit of the model to the rock (Section 9.6).

The distribution of the distance or cost of false alarms can be used to modify the ROC curve to give a more realistic indication of the performance of the system. Given the distri-

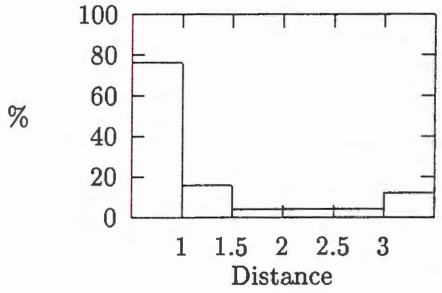


Figure 9.16: Graph of the distribution of the distance of each detected ellipse from the nearest ellipse for the result shown in the sample image. Ideally, the first bin with a distance of less than 1 should contain 100% with no hits in any other bin. In reality, detected points range in distance from near misses when distance is less than 1.5 up to complete misses for distance greater than 3. It can be seen from the sample image that the majority of the rocks have been accurately detected. Given the distribution among the misses, it can be seen that the result should predominantly represent the actual distribution in the image.

bution of costs is represented by $C = (c_1, c_2, c_3, c_4, c_5, c_6)$ where c_1 are true detections and c_6 are the worst false alarms which are completely wrong, then the following new values for the percentage true detections and the number of false alarms are:

$$\begin{aligned}\text{True Detections} &= c_1 + 0.5c_2 \\ \text{False Alarms} &= c_6 + 0.8c_5 + 0.6c_4 + 0.4c_3 + 0.2c_2\end{aligned}$$

Note that this is a simple estimate of the relative weights of each bin of the histogram. It is known that the cost of each bin should increase with its distance from the nearest rock and this relationship is approximated here as a linear one; true detections have weight zero and complete misses have weight 1 when calculating the number of false hits. In addition, since near misses are close to actual rocks, they are able to contribute a proportion to the number of true detections.

The result of using these values to replot the ROC curves is shown in Figure 9.17. Clearly, the percentage true detections increases slightly while the number of false alarms decreases. The curves also become more similar to the normal appearance of an ROC curve although this does not provide any new insight into the selection of operating parameters.

9.9 Fast Tree Implementation

The method used for generating trees which is described in this chapter is both time consuming and expensive in memory due to much redundant generation and storage of trees prior to pruning. This was found by observation of program performance. A non-recursive version of an algorithm will generally be more efficient with respect to time and space than a recursive

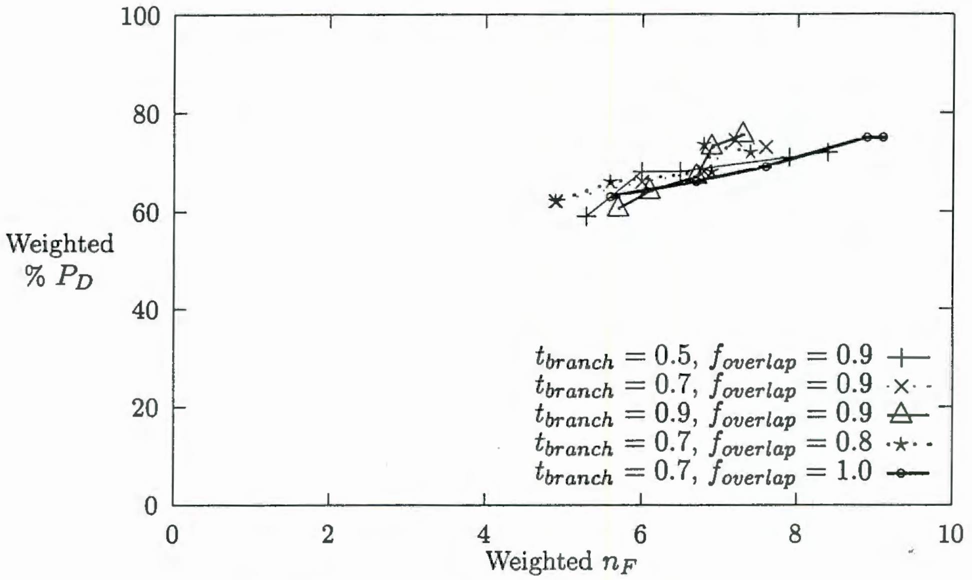


Figure 9.17: A modified form of the ROC curves in Figure 9.15 in which the false alarms are weighted to result in a more representative indication of the system performance. It can be seen that results are generally poorer when $t_{branch} = 0.5$ and $f_{overlap} = 1$. The combination of $t_{branch} = 0.7$ and $f_{overlap} = 0.8$ gives generally better results than the remaining two curves, but the combination of $t_{branch} = 0.9$ and $f_{overlap} = 0.9$ could also be useful for low values of t_{clip}.

version [88]. The improvement is achieved because the entry and exit overheads of a function call can be largely avoided in addition to a reduction in the amount of redundant processing.

Based on that observation, a more efficient approach has been implemented in which a list of nodes is created for each level of the pyramid. A list of branches from each node to the nodes on the level below is created. This results in one global tree structure with multiple roots. Each of the trees generated previously can now be accessed by performing a tree walk starting at the root node of the desired tree. The only precaution that must be taken is to ensure that the depth of the walk is limited to the depth of the original tree.

The list of pruned trees is generated using each node as the root of a depth limited tree. The same branch scoring technique is used as previously to determine which branch would produce the highest scoring tree. That branch is then added to the new tree and the branch scoring routine goes one step lower. This is repeated until the maximum allowable depth is attained.

A drastic improvement may also be achieved by remembering scores at each node. The only complication that occurs is that a different score will be returned by a node depending on the current depth of the tree. Since the depth limit is usually quite small (less than 10) it is possible to store an array of scores corresponding to each of the possible depths. In practice, if a score is not already known, it must be calculated using a recursive routine which then stores the result in the relevant position and returns it. Regardless of the number of nodes or tree depth limit, this improvement reduces the cost of tree scoring to an almost negligible amount.

Finally, for each of the pruned trees, the score, footprint and rock represented by the tree are calculated for use in the filtering routine which remains the same as before.

The modifications to tree analysis described here resulted in a speed improvement of several orders of magnitude with no compromise to results since the same algorithm has just been recoded more efficiently. If it were necessary to modify the tree pruning method to the more exhaustive method, it would be considerably more difficult to modify the iterative code and the pruning routine would have to be implemented using a stack in order to achieve fast results. Based on the performance of the less exhaustive method that is actually used, it does not appear to be necessary to change the algorithm.

9.10 Summary

Lists of possible rock positions from each level of the image pyramid were produced by the classification process. Hierarchical analysis was designed to determine the structure of the positions to find a final list of rocks in the image for measuring the rock size distribution. This was achieved by generating tree structures to represent the possible positions of rocks in scale-space which were then filtered using a weighting function to delete weaker trees where there was overlap. The result of this process was a physically realisable list of rocks which could occupy the image. Given the list of rocks, calculation of a size distribution is a relatively trivial matter.

Evaluation of the performance of this stage of the system indicated false alarms were removed from the lists only at the cost of some true hits. When one considers that the number of false alarms was reduced by around two orders of magnitude and the number of true detections only decreased by about 20%, the result is quite favourable and gives a strong indication that this form of analysis is suitable for the task.

Chapter 10

Results

10.1 Experimental Procedure

The aim of this research was to produce an instrument for measuring rock size distribution that is resistant to, or tolerant of, the following conditions in that it should be able to achieve the correct results in their presence. (1) It should correctly identify all rocks in the image regardless of their individual intensity. Specifically, it should not be biased against dark rocks or rocks in the shadow of larger, light rocks. (2) It should also find the same result regardless of rotation, reflection, magnification and translation of the image or rocks in the image. (3) The instrument should also not confuse areas of fines in the image for larger rocks.

The instrument is capable of measuring the projected area of each particle that is visible at the surface of the rock-stream. Stereological methods for converting this two-dimensional measurement to a volume distribution have not been investigated in this research and so it is not possible to estimate the volume distribution of the rock-stream accurately from the segmented image. Since a volume distribution is not going to be calculated, it is unnecessary to know the actual volume distribution of the material in order to verify the results.

The results that are given here are aimed at showing that the instrument is capable of accurately identifying rocks in the image and can therefore provide an accurate measurement of the projected area distribution. In order to validate these results, it is necessary to know the actual projected size and position of each rock in the image, as well as the actual projected area distribution.

The mechanisms that are used here to validate the performance of the instrument are as follows:

Area distribution The measured distribution and cumulative distribution can be compared to the actual, manually determined, distributions.

Error histogram The histogram of the minimum distance from each detected rock to the nearest actual rock (See Section 9.8) gives an indication of how many rocks have been accurately detected, how many mistakes the instrument made and how serious they were.

Images Three images are shown for each result. The first image shows the actual rock scene that is to be segmented at the full resolution captured by the camera. The next two images show the segmentation results superimposed on the smoothed version of the image that was analysed. One image shows the detected ellipses and the other shows the measured boundaries of the detected rocks.

Eight images have been selected for testing the instrument to provide the results just described. In each case, the image has been analysed manually to fit an ellipse to each of the rocks that are visible in the image. This information is required to determine the histogram and is also used to estimate the actual projected area distribution by calculating the areas of each of the ellipses.

The images were selected to illustrate the tolerance of the instrument to rotation and translation of the rocks in the image as well as ensure that it is capable of operating with any size distribution of rocks in the image. For this reason, the first five images are of the same set of rocks which have been rearranged. This ensures that there is the same distribution in each image while testing for rotation and translation.

The three remaining images are the ones that have been used during previous chapters for generating ROC curves and each contains a different distribution of rocks. The first image contains only large and medium rocks which are positioned to be touching in the image but not overlapping[1]. The second contains large, medium and small rocks, also not overlapping. The third image contains only medium and small rocks which are sparsely scattered around the image.

In each of the eight images, the rocks have been arranged so as not to touch the edge of the image. This is because the instrument cannot see rocks at the edge of the image. It would also not be possible to fit an ellipse to a rock that cannot be completely seen thus making the measured distribution inaccurate.

The next consideration is that of selecting the operating parameters for the instrument in order to achieve optimum performance. The parameters that are used in this chapter were found by examining the ROC curves that were created in Chapters 6, 7, 8 and 9. It was found after the two stages of preattentive vision that the performance was significantly better for the case where $\sigma_d = \frac{1}{3\pi}$ which corresponds to minimal smoothing of the image to match the input from the camera. Subsequently, for Chapters 8 and 9, all results were shown for the case $\sigma_d = \frac{1}{3\pi}$.

Before determining the final set of operating parameters, it is necessary to identify the relationship between the performance of the complete instrument and the image smoothing

---

[1] Note that laboratory rocks have been pre-sieved into three size categories: large 4cm, medium 3cm and small 2cm.

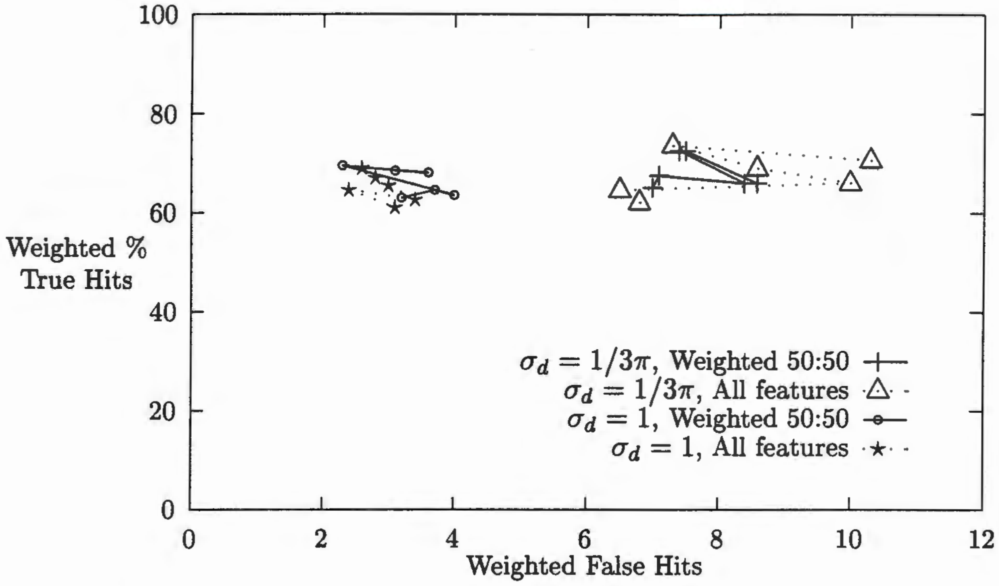


Figure 10.1: Comparison of results obtained for the two values of σ_d used to determine the amount of smoothing when creating the image pyramid. The ROC curves were obtained by varying the pyramid scaling factor f_{scale} from 0.7–0.95 in steps of 0.05 while keeping $t_{edge} = 0.75$, $t_{clip} = 0.55$, $t_{prob} = 0.65$, $t_{branch} = 0.7$ and $f_{overlap} = 0.8$. It can be seen that when $\sigma_d = 1$ there is an average reduction by a factor of 3 in the number of false hits with only a marginal decrease in the percentage of true hits. Although results after early stages of processing indicated better performance with $\sigma_d = \frac{1}{3\pi}$, it is clear from these results that smoothing is desirable for improved performance. This could be attributed to smoothing decreasing the amount of surface texture.

σ_d and pyramid scaling factor f_{scale}. Figure 10.1 shows the ROC curves that are obtained as the pyramid scaling factor f_{scale} is changed. Four curves are shown to illustrate how the performance changes with the selection of features and as the value of σ_d is changed from 1 to $\frac{1}{3\pi}$. Exactly the same operating parameters have been used for each case of σ_d except for the value of $r_{expected}$ which is changed from 7 to 9 (See Section 6.2.3) and a new set of training data was obtained for the $\sigma_d = 1$ case using the method described in Section 8.4. Note that feature selection was not performed again so that the same selection of features is used for both the Weighted 50:50 curves to allow direct comparison.

Despite the relatively poor performance with $\sigma_d = 1$ during the preattentive vision stages, it is clear that the overall performance is considerably better than the $\sigma_d = \frac{1}{3\pi}$ since there is virtually the same level of rock detection with about one third the number of false alarms. It can also be seen that the performance is relatively insensitive to the selection of features for classification although the optimised feature set using weighted error counting with equal contributions does give slightly better results for each value of σ_d when $f_{scale} = 0.85$.

Surprisingly, the instrument performance does not increase as f_{scale} increases. Because

Parameter	Description	Value(s)
σ_d	The degree of pyramid smoothing	1
$r_{expected}$	Expected radius of a rock	9
f_{scale}	The scale reduction factor for the pyramid	0.85
n_θ	Number of ellipse orientations	8
n_e	Number of ellipse eccentricities	3
f_{aspect}	Aspect ratio factor to calculate ellipses	1.1
t_{edge}	Rock edge intensity as a proportion of rock intensity	0.75
t_{clip}	Hough transform detection threshold	0.55
t_{prob}	Edge probability threshold for 2nd preattentive stage	0.65
$t_{feature}$	Hypercube rule threshold	1
Features	Which features are enabled for classification	Weighted 50:50
k	k-nearest-neighbour value	25
t_{branch}	Distance limit for tree branches	0.7
$f_{overlap}$	Maximum proportion of overlap relative to radius	0.8

Table 10.1: Parameter values that are used for all results shown in the remainder of this chapter.

the value of f_{scale} cannot affect the performance of preattentive vision or classification, the dependence must be introduced in the hierarchical analysis stage and this should be further investigated.

From the result shown in Figure 10.1, the values of $f_{scale} = 0.85$, $\sigma_d = 1$ and the Weighted 50:50 optimised feature set will be used for the remaining results in this chapter. Table 10.1 gives a list of all the parameter values that are used for all the remaining results in this chapter.

10.2 Output Distributions

This section contains two sets of results. The first set consists of the eight images which were described in the previous section and will be used to provide a quantitative analysis of the instrument's performance. This is achieved by comparing its results to values which have been calculated from a manually segmented version of the image.

The distribution is calculated by summing the area of each rock in each of the bins rather than by multiplying the number of rocks in the bin by the area represented by the bin. This gives a slightly more accurate result since the area is no longer being estimated from a one-dimensional measure of size (equivalent diameter). Note that the manually estimated area distribution is found by calculating the areas of each of the ellipses that have been manually fitted to the rocks. Figure 10.2 includes an area distribution that has been measured by manually delineating each of the rocks in that image. It can be seen that there is close agreement between the two manually estimated distributions which indicates that the use of ellipses for calculating the manual distribution is adequate.

The remaining nine images serve to give a qualitative analysis of instrument performance under a wide variety of operating conditions. The important thing to determine is whether or not it is possible to be certain that the instrument is reasonably robust and can provide consistent results. Clearly, since calculation of the size distribution is trivial once the segmentation is performed, the crux lies with the segmentation of the image; a good segmentation will give an accurate size distribution and vice versa. For this reason, the remaining images are shown for a variety of conditions to illustrate the consistency of the instrument under adverse conditions. Only the segmentation is shown for these cases since if that is correct and convincing, it follows that the size distribution will be too.

Due to the nature of many of the images, it is not practical to determine the size distribution manually and it is necessary to rely on a qualitative analysis of the segmentation. The reasons for this are:

- The material extends off the edge of the image and so it would not be possible to accurately determine its size distribution manually or using the instrument. See Section 11.2 for a possible solution to this problem.
- When images contain fines it would not be practical to perform manual segmentation since the smaller particles become more and more indistinct.

Despite these problems, the segmentations shown in Figures 10.10–10.18 are consistently accurate under various conditions and will therefore result in a measured size distributions which closely match the actual distribution.

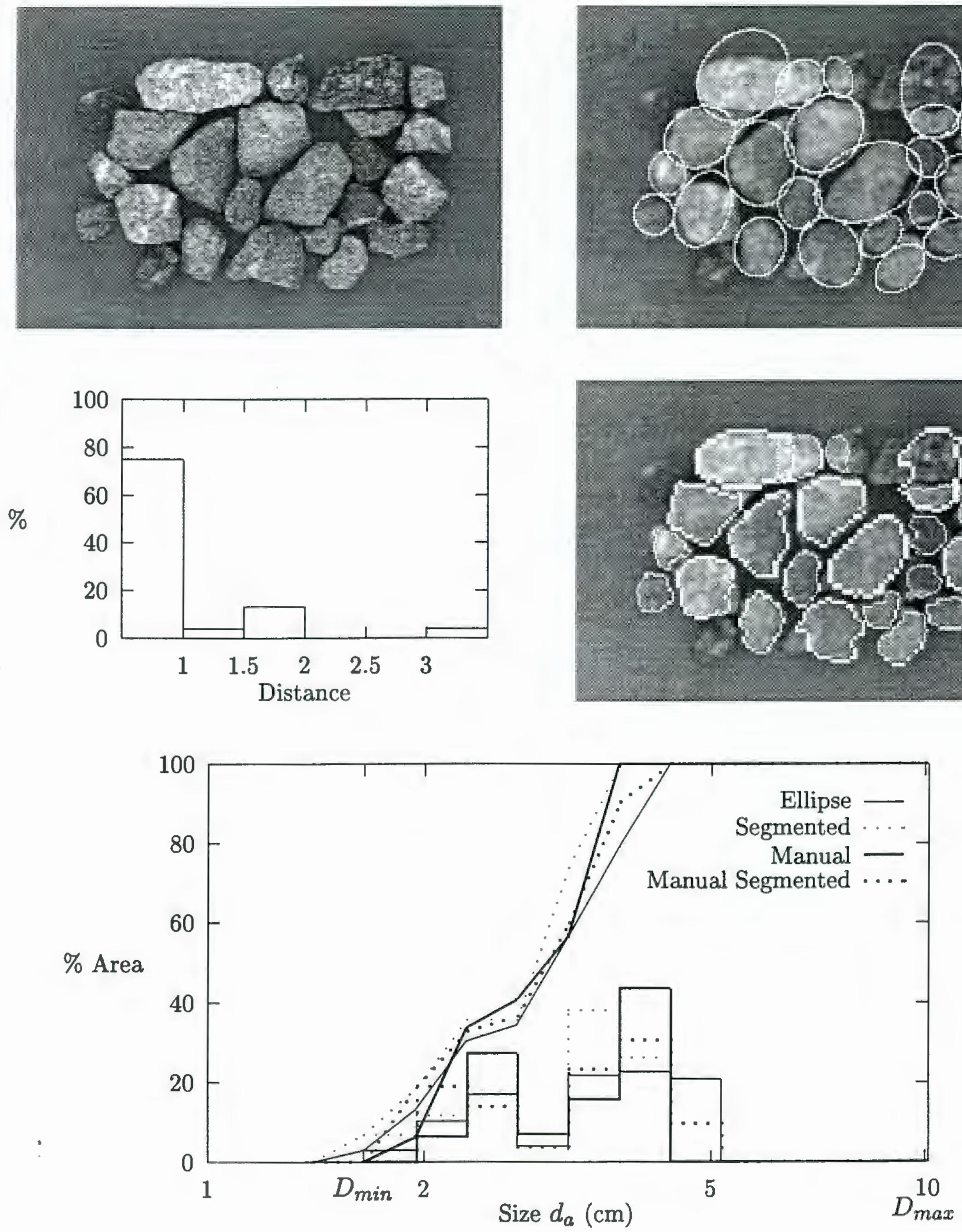


Figure 10.2: 1st image of sample rocks. Execution time 9:23min. Note that the true hit rate is 75% with a very low proportion of false hits. Two rocks have been missed. The long bright rock has been split. Two rocks have been incorrectly segmented; one has identified a bright mark and the other has resulted in a merging.

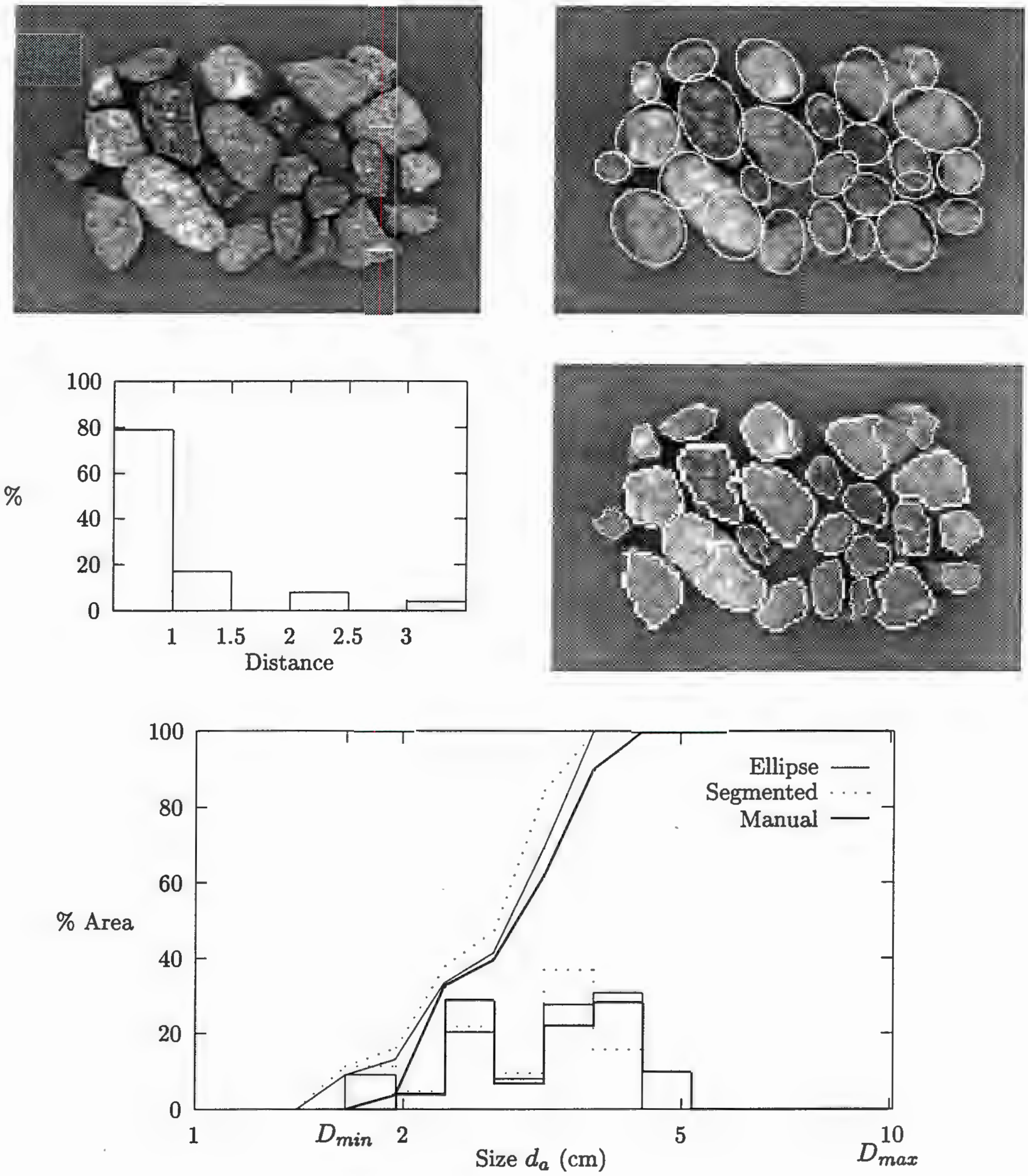


Figure 10.3: 2nd image of sample rocks. Execution time 10:44min. Again the long light rock is split. No rocks have been missed but a few rocks have been slightly inaccurately segmented. In two cases the ellipse does not appear to be correctly orientated and in one other case, a rock is detected once correctly and the corner is also detected as a small rock. There appears to be a large overlap here because the corner has been detected more strongly and so cannot be deleted by the correct ellipse, but is smaller and therefore sufficiently far from the correct ellipse to not delete it.

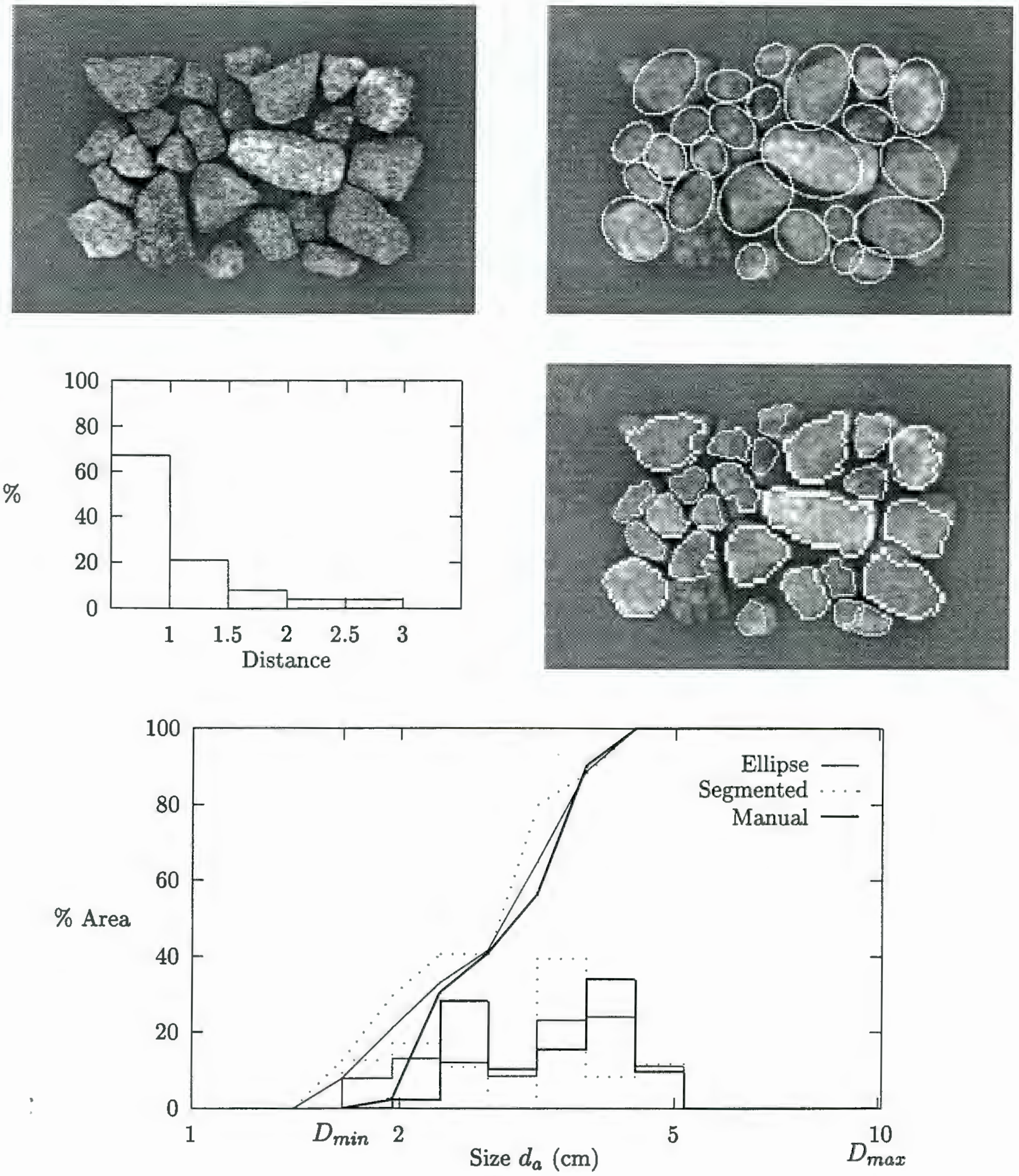


Figure 10.4: 3rd image of sample rocks. Execution time 9:47min. The light rock has been correctly segmented. A large rock near the lower left corner has been badly split probably because the lower edge of the rock is very indistinct. The rock at bottom right has been split into two parts and a few smaller rocks have not been very accurately matched with an ellipse.

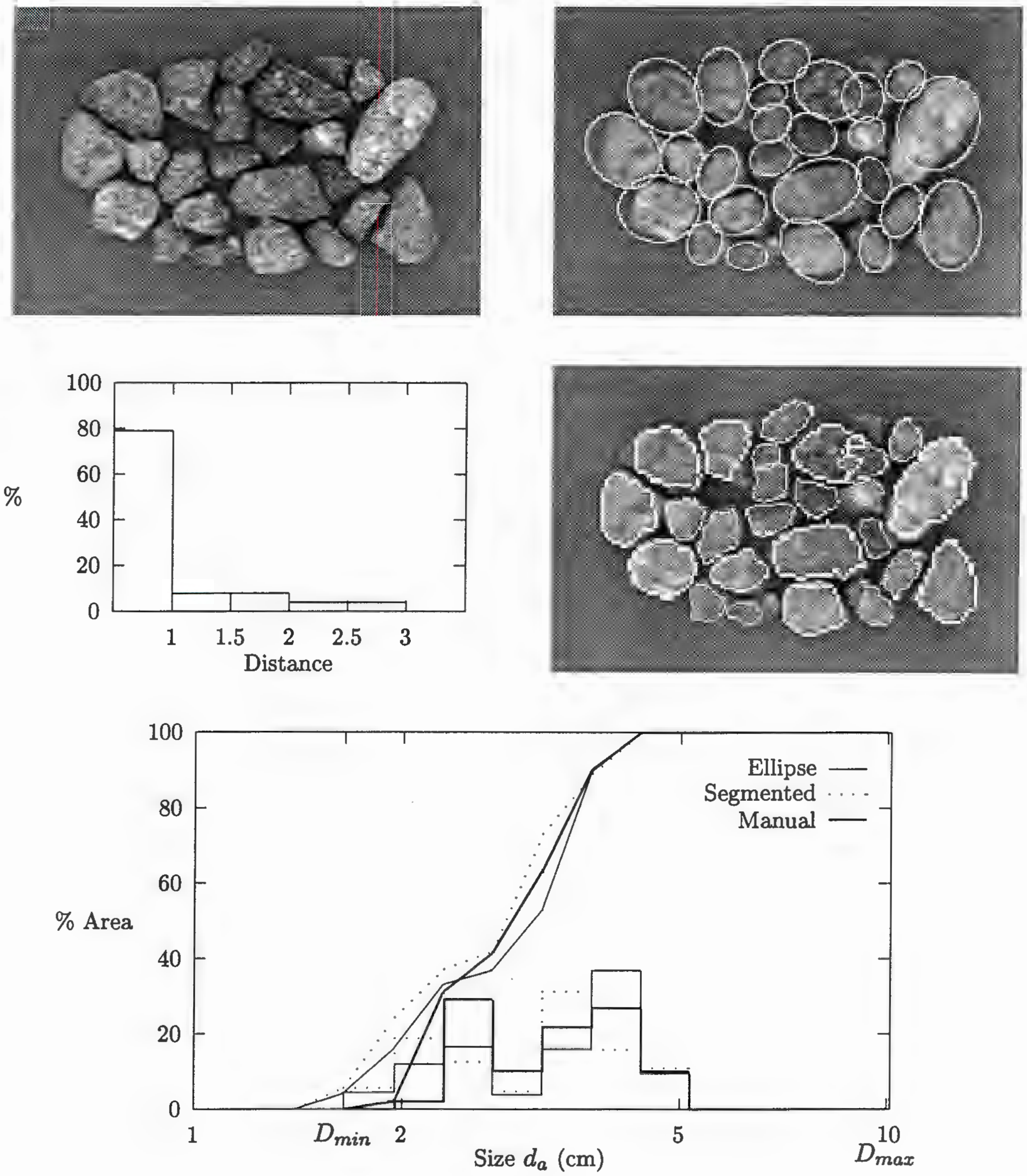


Figure 10.5: 4th image of sample rocks. Execution time 10:44min. Only one rock has been missed and a medium sized rock in the upper off-centre to the left has been split. A dark large rock has been split probably partly due to bright markings on its surface.

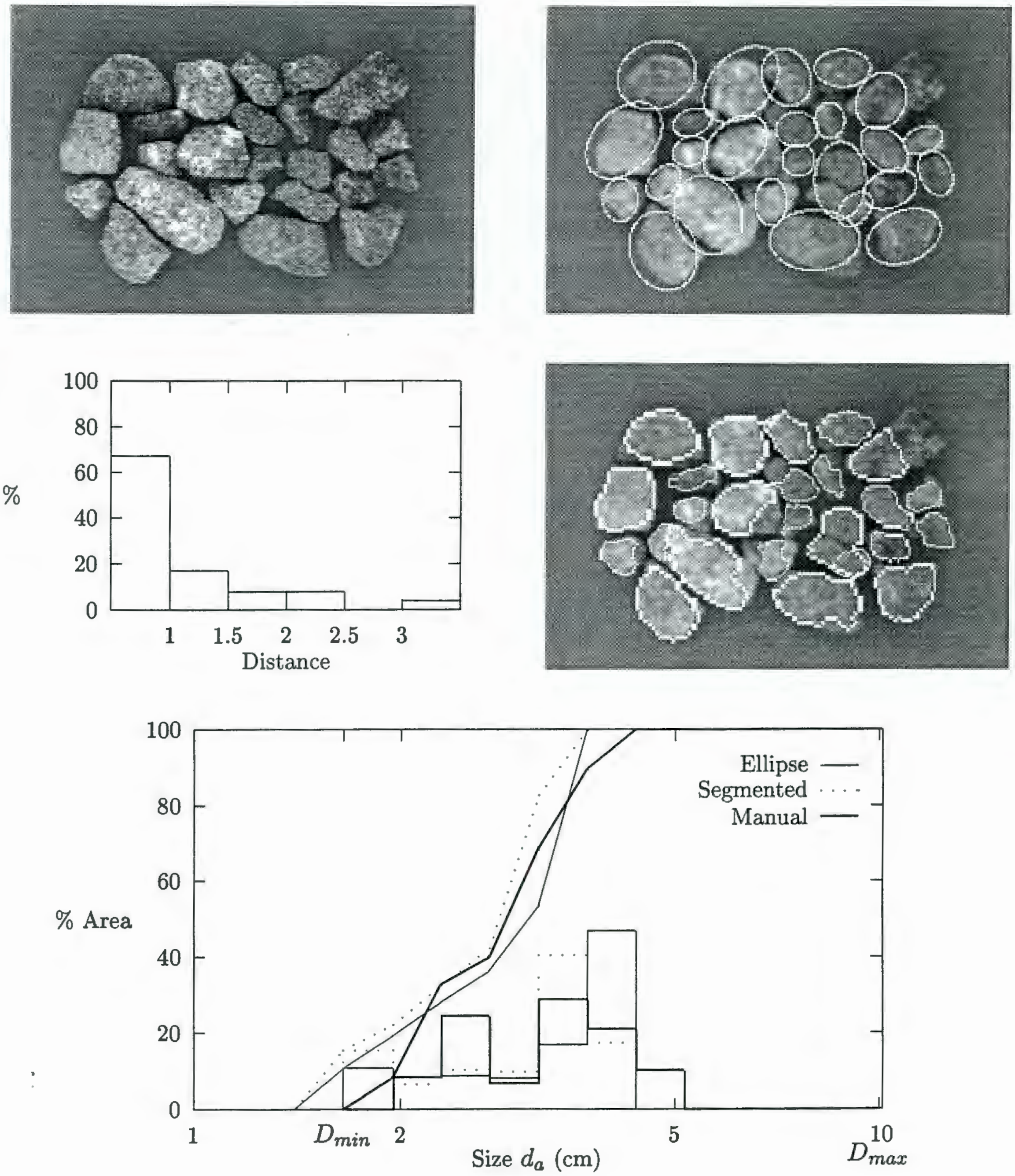


Figure 10.6: 5th image of sample rocks. Execution time 10:36min. The long light rock has been split again and two small rocks have been merged. The top right rock has been incorrectly modelled due to indistinct edge of rock beyond pile.

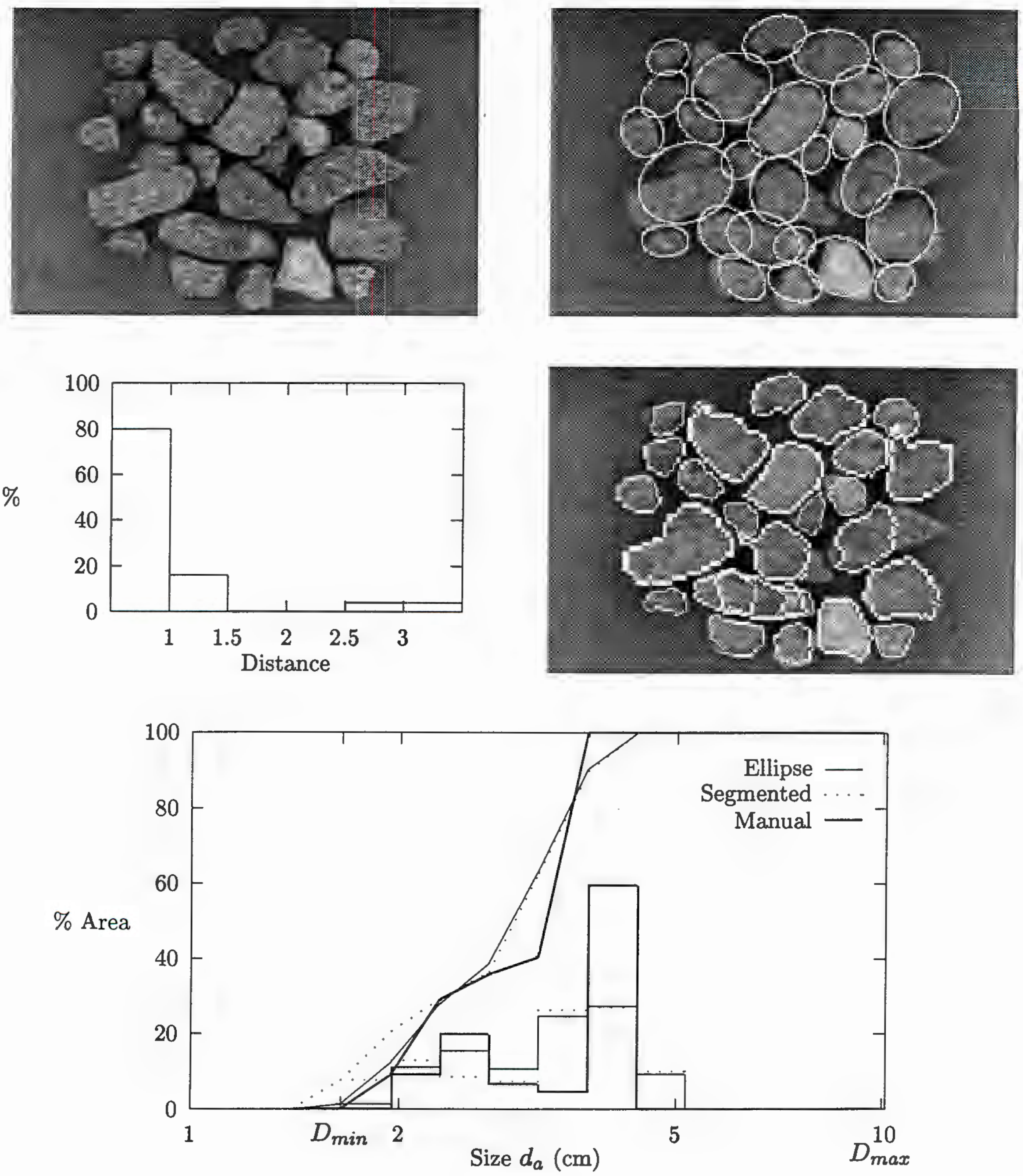


Figure 10.7: 1st development image. Execution time 5:32min. The large rock at the left has been merged with a small rock and the long rock below centre in the middle has been split into three parts although the central one results in a fairly good segmentation. The pointed rock at the right has not been accurately detected. This image contains only large and medium rocks. It can be seen that there is very little response in the distribution around the small rock size.

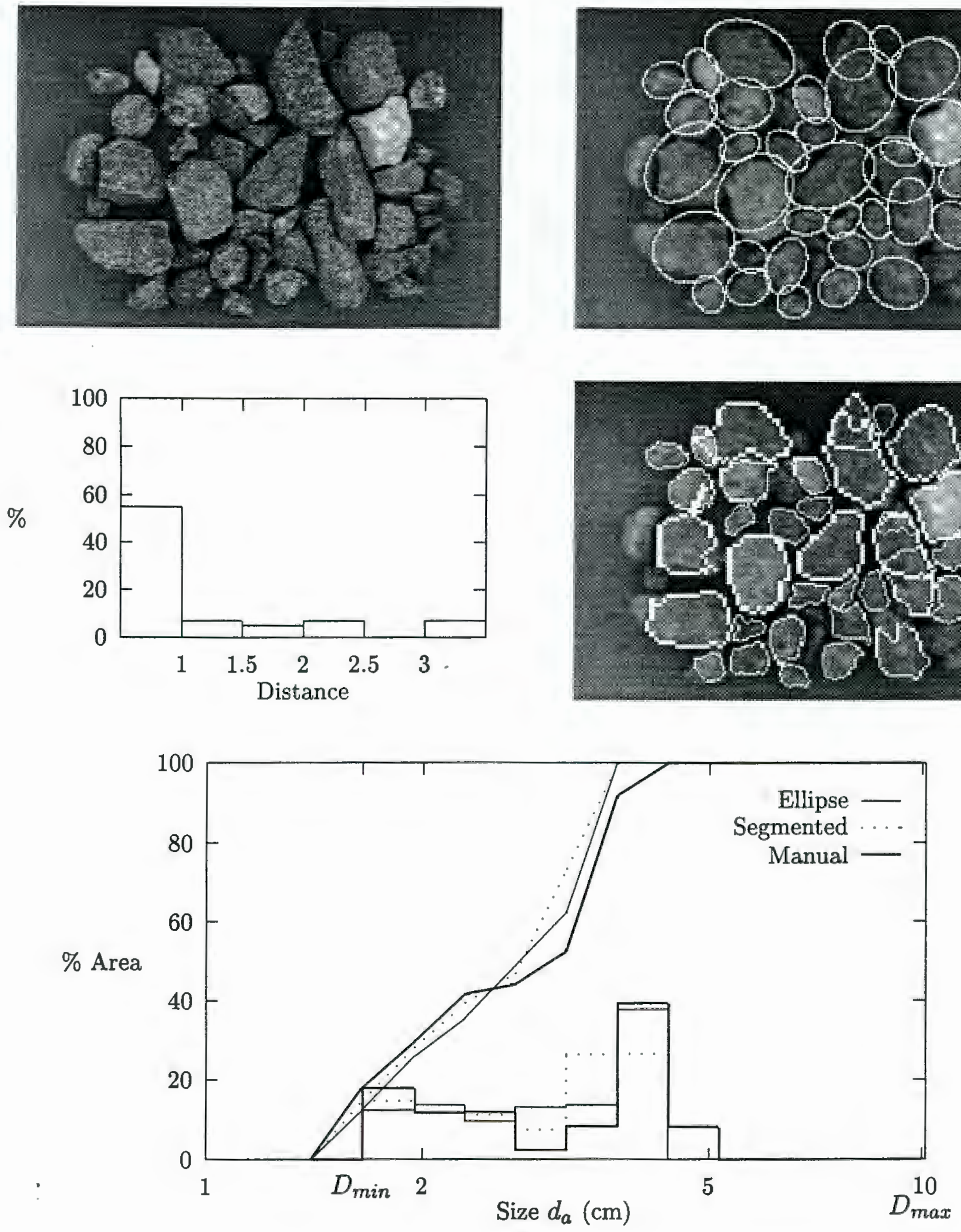


Figure 10.8: 2nd development image. Execution time 6:05min. A fairly low detection rate of only 55% is obtained for this image although the distribution matches reasonably well. Most of the inaccuracy results because the rocks are small and faint because they are near to the edge of the pile. There are also two long rocks which are not accurately detected. The distribution is not too inaccurate because it is normalised to a cumulative percentage and so the actual number of rocks detected is irrelevant as long as they are representative of size. This image contains large, medium and small rocks and this is apparent from the distribution.

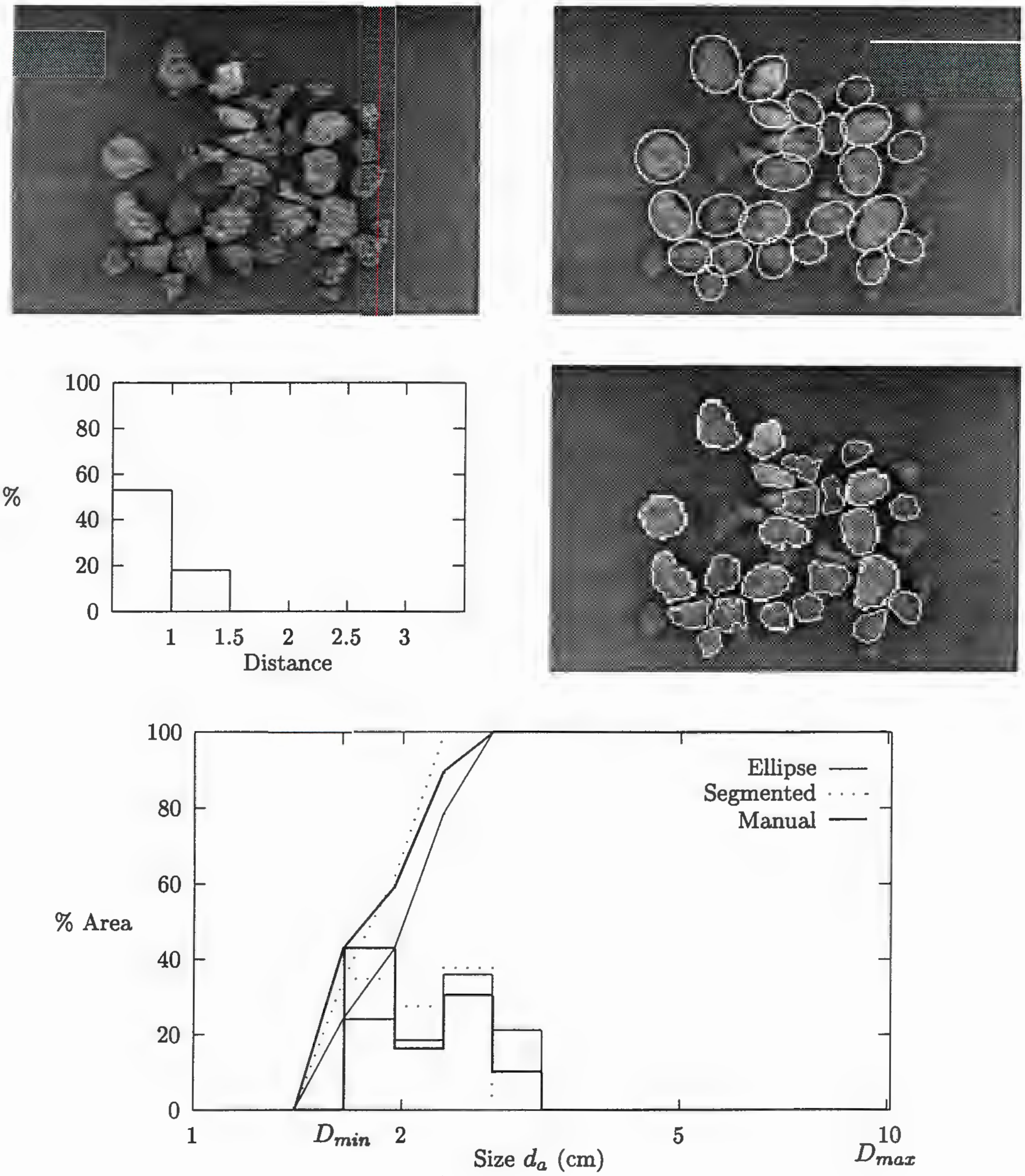


Figure 10.9: 3rd development image. Execution time 5:59min. As with the previous image there is a fairly low detection rate due to the small sizes of the rocks and because they are less distinct due to their sparseness resulting in reduced contrast at their edges from shadows. The rocks that are detected are done so correctly so that the size distribution is reasonable. Note from the histogram that there are only true hits and close misses. This image contains only medium and small rocks and it can be seen that this is found in the distribution.

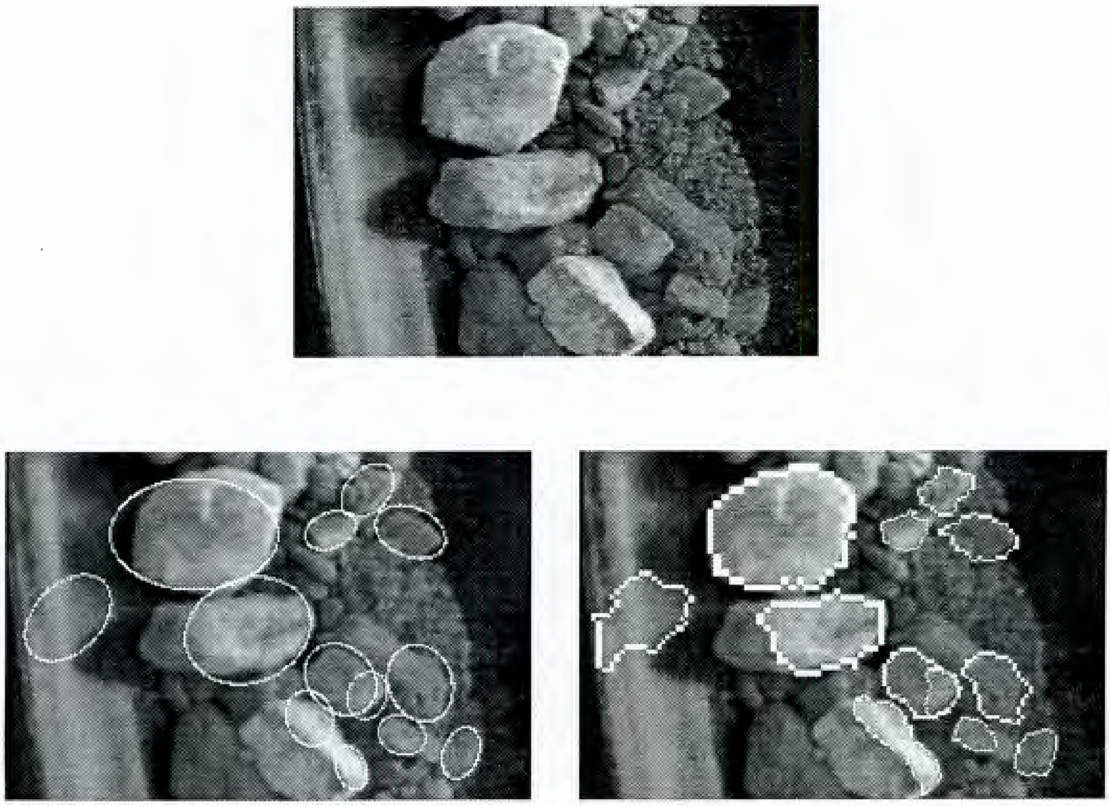


Figure 10.10: 1st real image under indirect light with few fines and spread out material. Execution time 9:37min. The system has responded to a section of the belt. This is probably due to the shadows caused by the two large rocks and the edge of the belt providing edges. The mistakes that have been made with segmenting the rocks have occurred because the light is directional with the result that facets of the surface are of different intensities. The instrument has responded to the lighter facet and discarded the darker one adjacent to it because it does not have a suitable edge at the transition from dark to light.

10.3 Analysis of Results

The purpose of this section is to investigate the performance of the instrument as shown in Figures 10.2–10.18 and comment on the errors that are made. The constraints of the instrument will also be investigated.

Inspection of the first eight images shows that the instrument produces a projected area distribution that closely matches the manually estimated values. It can also be seen that there is generally a high success rate at correctly detecting the rocks in the image while producing very few false alarms. This is reflected in the histograms from which an average true detection rate of 69% can be calculated.

These results clearly show that the instrument is capable of accurately identifying and

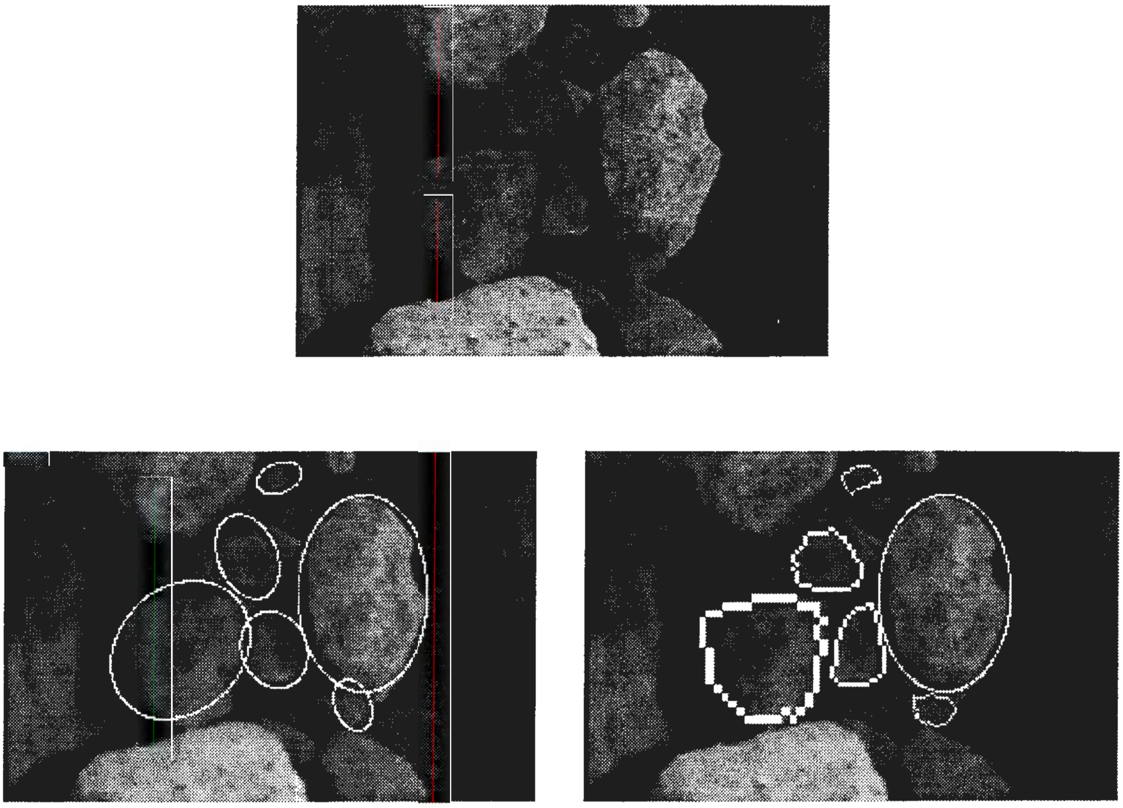


Figure 10.11: 2nd real image under indirect light with few fines and spread out material. Execution time 10:43min. This image has been very accurately segmented. Notice that the instrument has not responded to rocks which are only partially present since they do not have any part which could be mistaken for a smaller rock. The large rock is too close to the edge of the image (because of quantisation errors) to be segmented and so the ellipse is shown instead.

measuring rocks in the image, and that the measured projected area distribution will closely match the actual distribution.

Note that the size range of rocks that were available for creating laboratory images is only from around 2cm to 5cm which gives a dynamic range of not much greater than one octave. The instrument is actually capable of measuring rocks over a range of almost a decade (from D_{min} to D_{max}). It can be seen from the results that the system is very sensitive to size since it is able to distinguish between three size categories (in Figures 10.7, 10.8 and 10.9) over the small range. Although no quantitative results were obtained with a wider size range, there is no reason why the instrument should not be able to operate correctly, and there is some qualitative evidence from the images of real rock-streams that this is the case.

The results in Figures 10.10–10.18 indicate that the instrument is robust in that it can operate reasonably successfully over a wide range of input conditions varying from laboratory

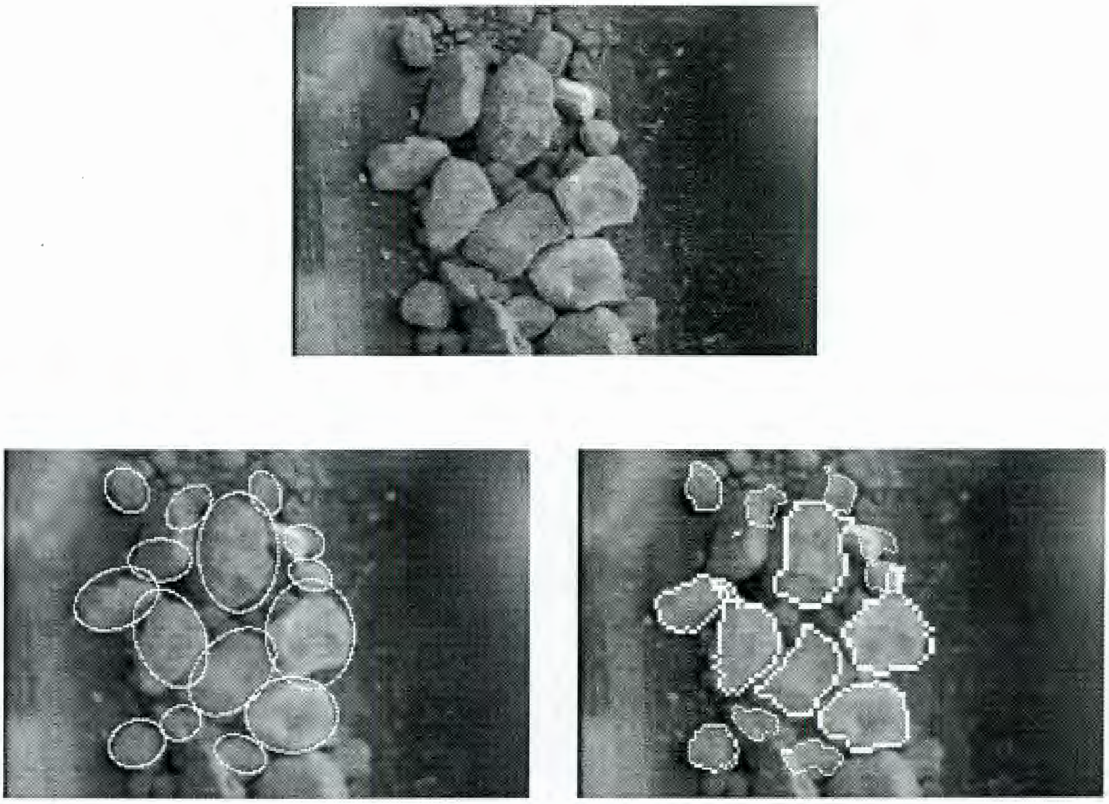


Figure 10.12: 3rd real image under indirect light with few fines and spread out material. Execution time 8:15min. This image has also been very accurately segmented. Notice, however, the segmentation error on the medium size rock near the upper left corner.

conditions with sparsely and closely packed material under controlled diffuse lighting, to a real rock-stream with uncontrolled lighting. Note that some of the images that were obtained from a real conveyor belt do contain a wider range of rock sizes than the laboratory images and they are generally accurately segmented unless the rocks lie too close to the edge of the image.

It can be seen from the results that the following conditions appear most likely to cause incorrect segmentation and hence incorrect measurement of a rock.

1. Small rocks are missed because they are below the resolution of the instrument (D_{min}) although the common error of merging clusters of small particles is generally not made by the instrument. This means that although the small rocks are not measured, they are not introducing a bias towards larger sizes in the distribution.
2. Indistinct rocks are missed. If there is not sufficient edge strength around a rock, it is not detected by the system. This condition is usually a result of the rock not having close neighbours and therefore only weak shadows around its edges. Alternatively, a

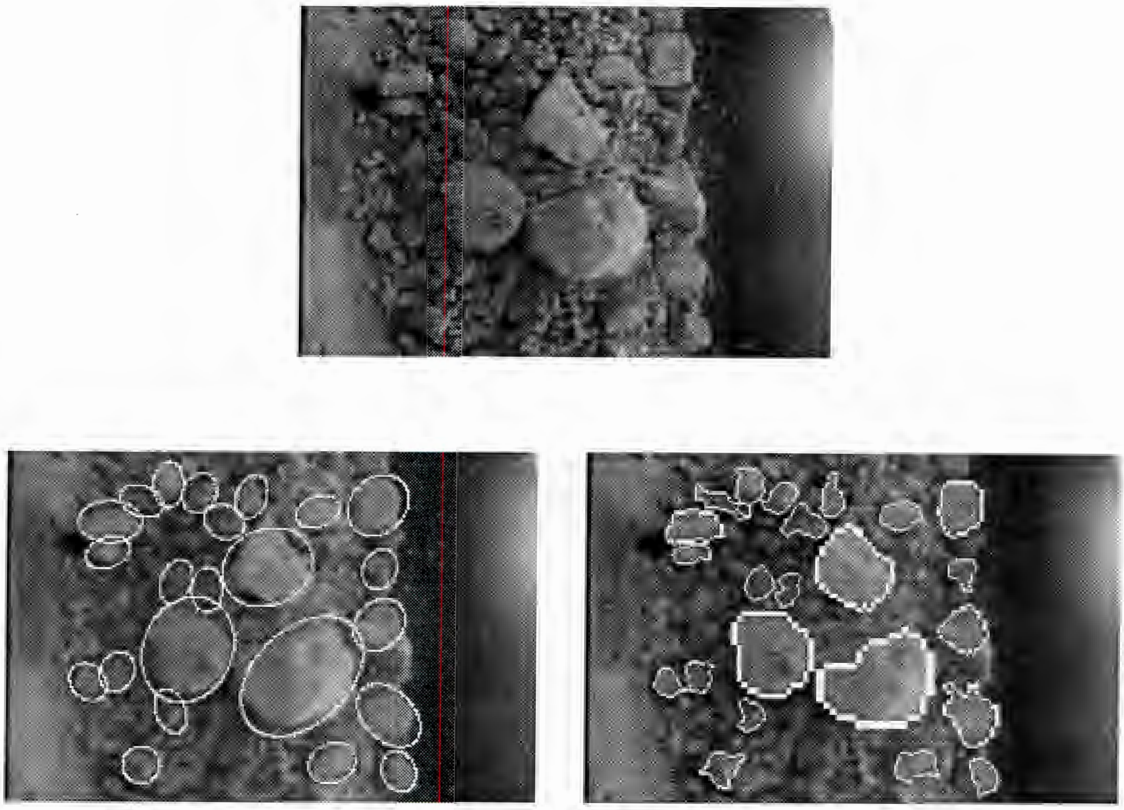


Figure 10.13: Real image under indirect light with many fines and some spread out rocks. Execution time 8:48min. Although there are many small clumps of fines in this image none of them appear to have been mistaken for a larger rock.

small dark rock next to much lighter rocks can also be missed because there is insufficient decrease in intensity at the edge of the small rock. This might only effect a size distribution if the indistinct rocks are generally of a particular size, otherwise it only results in a lower sample of the distribution.

3. Long rocks are difficult. Long rocks tend to be either split into parts or merged with neighbouring rocks of similar intensity because the instrument is slightly biased in favour of circular rocks despite the use of an elliptical shape model. This is partially due to square and circular kernels and structuring elements being used for convolution and morphological operations. Depending on the frequency of long rocks, it could result in a bias towards either smaller or larger rocks.

4. Bright surface markings or rough texture can result in incorrect segmentation where a bright part of a rock is identified as a rock itself. This error will bias the distribution in favour of smaller sizes.

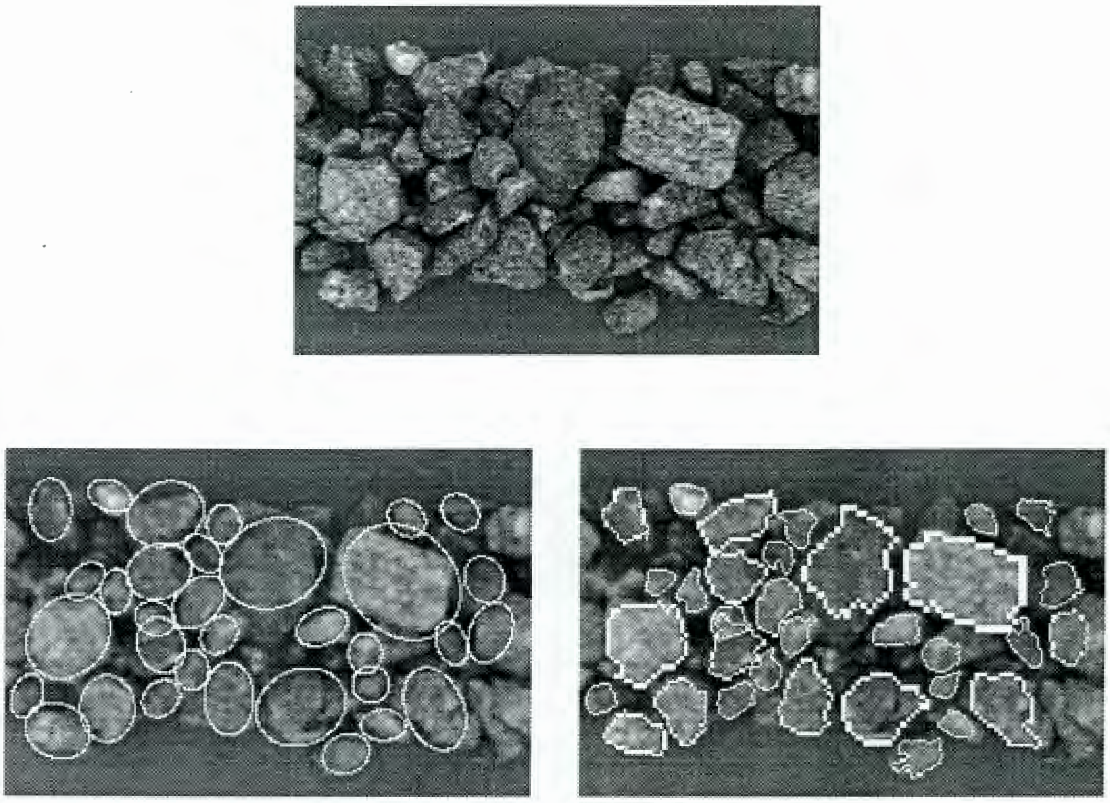


Figure 10.14: 1st image of packed rocks under laboratory conditions. Execution time 12:38min. This image contains rocks which have been placed in a pile and has therefore resulted in some degree of occlusion and a reasonably difficult image to segment since there is not much space between the particles. The results that are obtained appear very encouraging since there are very few particle mergings or splittings although some indistinct particles are missed.

It can be seen from all the distributions that although the measured amount of small rocks is generally greater than the actual value, it is not by a large degree. Note that in most cases, the correct answer is actually present after the classification stage but is not selected by hierarchical analysis. This could be due to inadequacies of the tree approach or the relative weights used for scoring. The situation could be improved by either improving the classification stage or the method for calculating tree scores.

The mistakes that are made by the instrument, especially in Figures 10.10 and 10.18, show that it does not actually recognise rocks, but rather it is a complicated way of finding light objects on a relatively dark background within certain shape constraints. It just happens that rocks generally produce this pattern and are hence correctly segmented. It is quite possible for other artifacts to produce the same pattern such as the combination of shadows and the edge of the conveyor belt in Figure 10.10. Under controlled conditions, however, the probability of

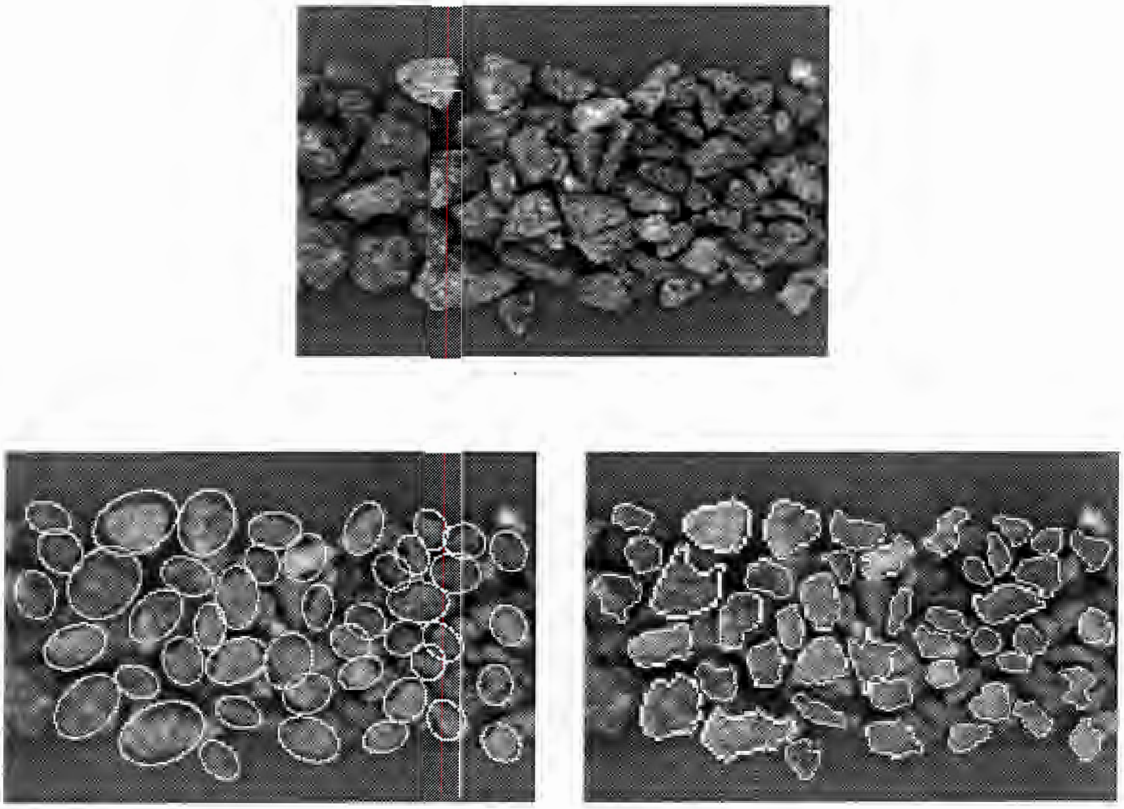


Figure 10.15: 2nd image of packed rocks under laboratory conditions. Execution time 17:52min. Another image with a considerable number of fairly small overlapping rocks. Again good results are achieved.

these situations can be decreased so that the assumptions that are encoded in the instrument actually correspond to good performance.

These results have been obtained using a Sun Sparcstation 20 and it can be seen that in this case, with the degree of image complexity that is present in images of real rock-streams, the execution time is less than ten minutes. Execution time is proportional to the number of rocks in the image which is why some of the laboratory images have taken considerably longer to process. The memory requirements of the process are reasonably high due to the complexity of the solution and the use of multiple image representations. However, it is possible to execute the program on an Intel 486 processor given 12 to 16 megabytes of RAM.

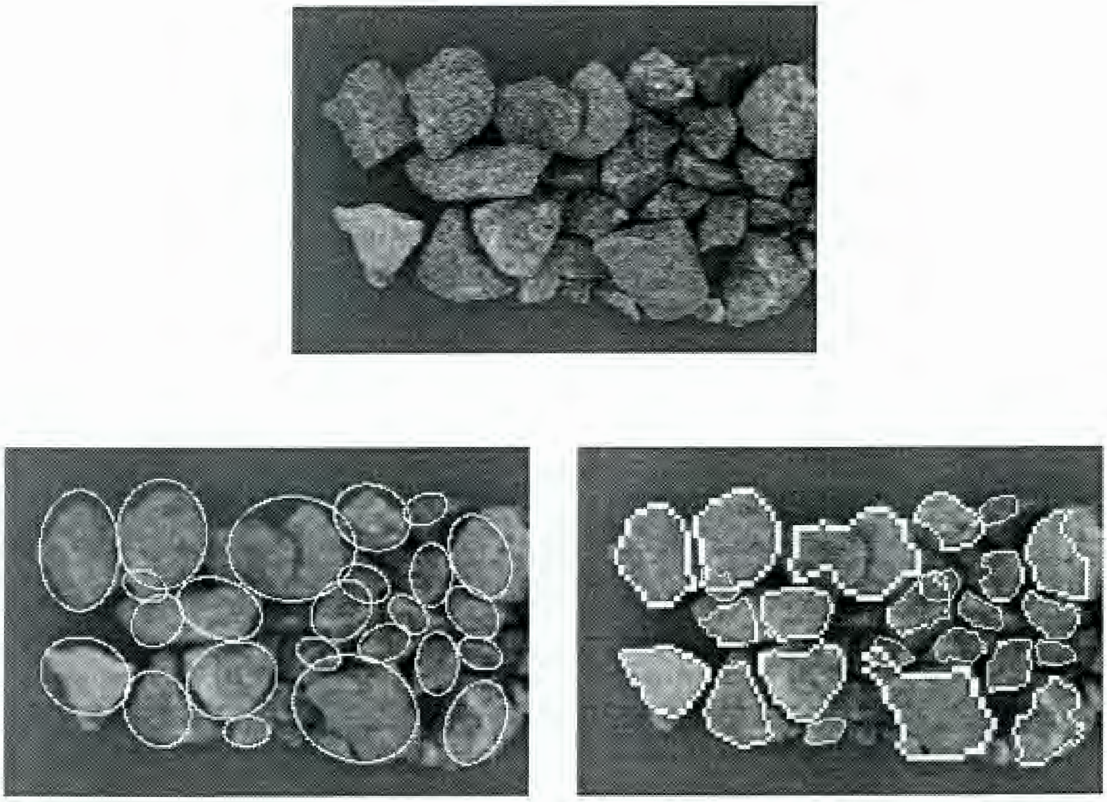


Figure 10.16: 3rd image of packed rocks under laboratory conditions. Execution time 9:35min. Two rocks have been merged at the middle top of the image and the long particle below them has been split. Some small dark rocks in the shadow of the larger ones have been missed.

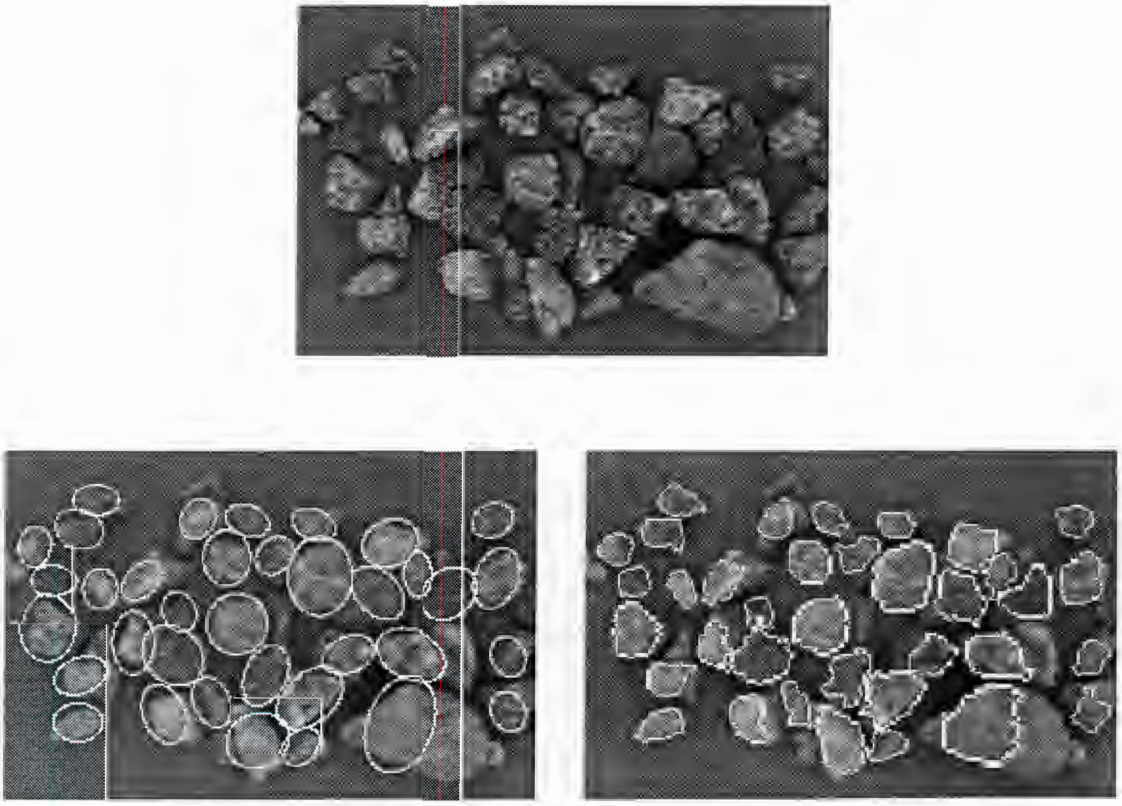


Figure 10.17: Image of spread rocks under laboratory conditions. Execution time 16:25min. The large particle at the lower right corner has been incorrectly segmented since it is too close to the edge of the image to be seen at the correct scale. This is because the segmentation algorithm requires a certain amount of contextual information.

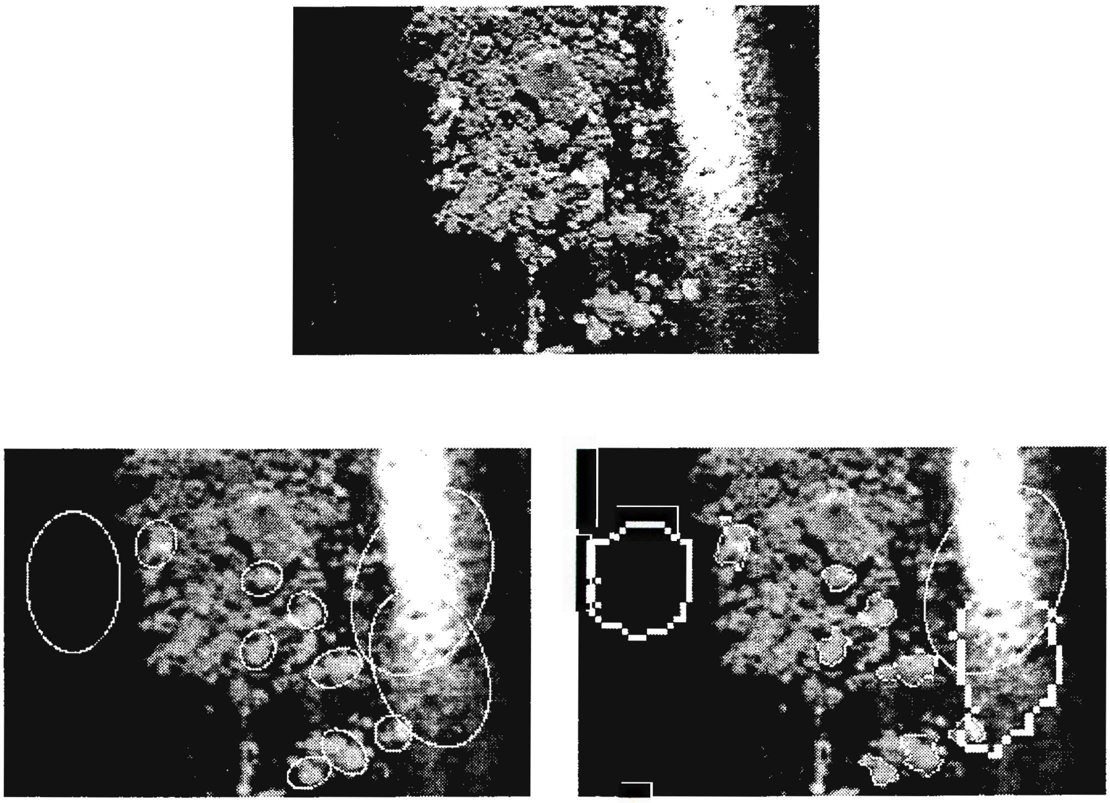


Figure 10.18: Image of fines and rock material on a wet belt under direct sunlight. Execution time 7:34min. This image contains a pathological case in which the conveyor belt is wet and there is direct sunlight from overhead which has resulted in a lot of specular reflection. The results for this image are poor due to the incorrect lighting which does not match the model which is implied in the segmentation algorithm and classification process.

Chapter 11

Conclusions

11.1 Summary and Conclusions

A need has been identified in the mining industry for an instrument for measuring rock fragmentation. A measure of rock fragmentation will be useful for increasing the efficiency of the gold extraction process by improving control over comminution using autogenous mills.

The work that has been performed in this thesis can be separated into the following categories: review of previous attempts at a solution to the problem; development of a computational framework based on the structure of the human visual system and multiscalar image processing techniques; investigation of the effects of lighting and the image acquisition procedure on the contents of an image; development of a rock detection algorithm consisting of three stages (preattentive vision, attention focusing and hierarchical analysis); and presentation of results.

It was found that although good performance has been reported previously for rock fragmentation analysis systems, certain common disadvantages could be identified in the methods used. The disadvantages were usually as a result of the use of low-level image processing techniques and resulted in poor performance in regions of low contrast.

As a result of an analysis of lighting, it was proposed that a diffuse lighting arrangement would be most suitable for an instrument measuring rock fragmentation for the following reasons: objects do not cast hard shadows; adjacent objects will be separated by a shadow regardless of orientation; different facets of an object will have similar intensities; and specular reflection from wet surfaces will not significantly change these results.

Preattentive vision has been accomplished using a two stage process. The Hough transform is used to detect elliptical shapes in the image after which the image content is compared to a simple model of rock shape. This process acts to highlight targets in the image for attention focusing and effectively reduces the search space from several million possible rock positions to thousands.

Attention focusing is implemented using a two stage process of segmentation and classi-

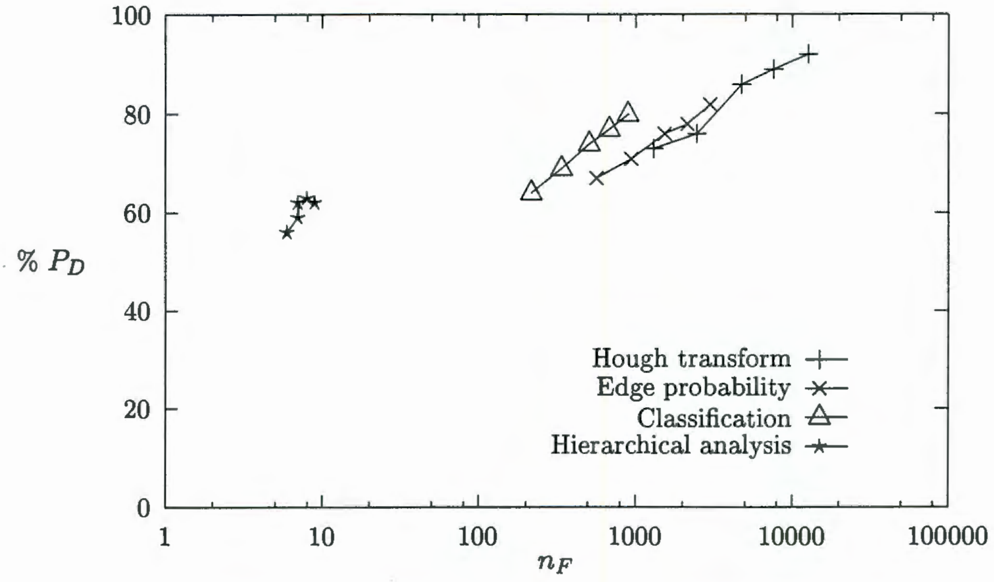


Figure 11.1: ROC curves showing the results after subsequent stages of processing. In each case the variable t_{clip} after the Hough transform is varied from 0.5–0.7. The values in Table 10.1 have been used for all other parameters.

fication. The segmentation algorithm incorporates knowledge of the shape of the rock from the Hough transform with other knowledge about the appearance of rocks to delineate the object in the image corresponding to a particular target. Features are then measured from the labelled region and classification is performed to determine whether or not it could correspond to a rock. This is similar to the HVS process of comparing objects to an internal representation for identification.

Finally, multiscalar results from each level of the image pyramid are combined in a stage of hierarchical analysis to determine a list of rocks for the image. This corresponds to the high-level component of the system which is responsible for coordinating the results from previous stages to resolve contradictions in the final representation which could result from multiple detections of the same object.

Figure 11.1 illustrates the effect of each of the stages required to obtain a list of rocks from an image. The result of each stage is shown using an ROC curve in which the variable is t_{clip} from the Hough transform. Note that before the preattentive vision is used, the position on the graph would have been $P_D = 100\%$ and $n_F \approx 4,000,000$.

From this graph it can be seen that the Hough transform which acts as the first stage of preattentive vision provides a very dramatic decrease in the number of false alarms (approximately two orders of magnitude). Although the usefulness of the second stage of preattentive vision may be brought into doubt by this result, it was shown in Section 7.6 that this stage performed considerably better when $\sigma_d = \frac{1}{3\pi}$.

The classification stage is shown to consistently improve the performance of the instrument by decreasing n_F without significantly decreasing P_D. Finally, hierarchical analysis has decreased n_F by another two orders of magnitude in order to arrive at a final result. Unfortunately, this has occurred at the cost of a decrease in P_D.

Despite the fact that a couple of the modules do not perform particularly well, the results shown in the previous chapter indicate that the instrument performs very well in the current configuration which is shown in Table 10.1. From this it can be concluded that the instrument has become to some degree distributed due to its examination of many targets in order to detect a final few rocks. It also indicates that further improvements in performance can be expected with further development to refine the modules.

The other conclusions which can be drawn from this research are the following:

- A diffuse lighting arrangement provides desirable conditions in which to identify rocks due to a low response to specular reflection and slow variation of intensity with changing surface orientation. It also ensures that a shadow will separate objects which lie close together thus facilitating their correct identification.
- The use of a multiscalar system has meant that a segmentation algorithm could be developed which gives good results for all rocks regardless of their size.
- Basing the computational structure on that of the human visual system has resulted in an instrument that is capable of producing accurate size distribution measurements in the presence of poor image capture conditions in a reasonable period of time.

11.2 Future Work

Although the instrument already produces accurate results, the performance could be improved in a number of ways as a result of its modular nature. Some possible improvements that could be made are suggested in this section.

Figure 11.1 and the results of Section 7.6 indicate that the second stage of preattentive vision does not perform very well when $\sigma_d = 1$, although the overall result is significantly better than the case where $\sigma_d = \frac{1}{3\pi}$. Early results indicated that better performance could be achieved for preattentive vision if $\sigma_d = \frac{1}{3\pi}$. A dual resolution system could be used whereby a high resolution image is used for preattentive vision and a low resolution image is used for attention focusing. This may sound like a contradictory situation since one would expect higher resolution for attention focusing, but all indications are that this would lead to an overall improvement in performance since the input to the classification stage would be better than it is at the moment.

Preattentive vision could also be improved in other ways. Firstly, a more sophisticated Hough transform algorithm could be used with thinned edges in I_{edge} to improve the detection rate and this might reduce the need for the second stage of preattentive vision. Another problem with preattentive vision is that the intensity based edge detection routine does not always respond to edges of low contrast at the edge of the rock-stream. Since missing a rock at this stage places an absolute limit on the performance, this problem should be investigated further.

The segmentation algorithm that is currently used benefits from the incorporation of knowledge about rock shape but does sometimes produce outliers which cannot be removed due to the simplicity of the outlier removal algorithm. It may therefore be worthwhile to investigate other segmentation schemes which incorporate shape information such as the radial contour model used by Brinkley [9] for segmenting human organs which have known shape. Although rock shape is not characteristic, certain restraints could be placed on it and with a process of relaxation to position edges, a more accurate segmentation could be obtained.

Classification has been demonstrated to play an important role in identifying rocks or removing non-rocks from the possible list of targets. Unfortunately, there still remain a large number of false alarms after classification and this situation could be improved in a number of ways: by identifying new features, using a more sophisticated discriminant function, determining a more appropriate measure of distance for creating representative training data, or identifying different classes of false alarms and discriminating between them. This information could be used by a high-level system to improve hierarchical analysis of the image in a manner which is examined next.

Essentially, the instrument currently operates in a bottom up fashion in that all low-level evidence from preattentive vision is examined using attention focusing before the high-level hierarchical analysis stage takes place. The instrument could be modified to use a high-level probabilistic framework for gathering evidence about the structure of the image such as that used by Sarkar and Boyer [82]. They use a perceptual inference network to manage a vision process by representing knowledge about the content of the image, performing reasoning about the structure of the image, and gathering further evidence to corroborate results.

This framework could accept initial results from preattentive vision, postulate an image structure or layout of rocks, gather further evidence by examining selected targets and incorporating classification results to modify the model of the image, until some degree of certainty is obtained. This approach could reduce processing time since it may not be necessary to examine all of the detected targets, and should improve the overall result because it could use classification results in a more global manner than the current tree approach.

One particular disadvantage of the way that the instrument is currently configured, is that it is possible to detect small rocks closer to the edge of the image than large ones. This can result in errors when a large rock is too close to the edge and is incorrectly identified as a number of smaller particles. A possible solution to this problem would be to determine a peripheral region around the image which is wide enough for half the size of the largest rock to be detected. Any rock whose centre lay in the periphery would not be detected. In this way, any large rock near the edge of the image would not be segmented at all so that there would be no chance of incorrectly biasing the distribution in favour of smaller sizes (although this does not appear to be a serious problem). The disadvantage of this approach is that it results in a smaller region in which rocks can be detected and as a result, a reduced resolution.

The computational structure that has been developed here has been shown to be capable of successfully identifying and measuring rock size from an image. Rocks are particularly difficult to recognise because they do not have a characteristic shape or appearance. For this reason, it should be possible to adapt this instrument to detect a variety of other types of objects. If the shapes of the objects are more constrained than that of a rock, some improvement in results should be expected.

Bibliography

[1] Fritz Ackermann and Michael Hahn. Image pyramids for digital photogrammetry. In *Proceedings of the 43rd Photogrammetric Weekly*, pages 43–58, Stuttgart, 1991.

[2] Terence Allen. *Particle Size Measurement.* Powder Technology Series. Chapman and Hall, London, 3rd edition, 1981.

[3] Sandra L. Bartlett, Paul J. Besl, Charles L. Cole, Ramesh Jain, Debashish Mukherjee, and Kurt D. Skifstad. Automatic solder joint inspection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(1):31–43, January 1988.

[4] Jean-Marie Beaulieu and Morris Goldberg. Hierarchy in picture segmentation: A stepwise optimization approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(2):150–163, February 1989.

[5] Günther Franz-Martin Berger. Software for a particle-size analyser based on image analysis techniques. Master's thesis, University of the Witwatersrand, 1985.

[6] Thomas O. Binford. Survey of model-based image analysis systems. *The International Journal of Robotics Research*, 1(1):18–64, 1982.

[7] Sing-Tze Bow. *Pattern Recognition, Applications to Large Data-Set Problems.* Marcel Dekker, Inc., 1984.

[8] Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. *Classification and Regression Trees.* The Wadsworth statistics/probability series. Wadsworth International Group, Belmont, California, 1984.

[9] James F. Brinkley. A flexible, generic model for anatomic shape: Application to interactive two-dimensional medical image segmentation and matching. *Computers and Biomedical Research*, 26:121–142, 1993.

[10] Peter J. Burt and Edward H. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, COM-31(4):532–540, April 1983.

[11] John Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, November 1986.

[12] Yao-Chou Cheng and Shin-Yee Lu. Waveform correlation by tree matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-7(3):299–305, May 1985.

[13] W. Cheung, F. P. Ferrie, R. Dimitrakopoulos, and G. Carayannis. Computer vision-based rock modelling. *Computing Systems in Engineering*, 3(5):601–608, 1992.

[14] Chung, H. Stephen, and Glenn Ludwig. Semi-automated fragmentation assessment. In *Proceedings of the 8th Annual Symposium on Explosives and Blasting Research*, pages 131–140, Orlando, Florida, USA, 1992. International Society of Explosives Engineers.

[15] Véronique Clément and Monique Thonnat. A knowledge-based approach to integration of image processing procedures. *CVGIP*, 57(2):166–184, March 1993.

[16] James L. Crowley and Alice C. Parker. A representation for shape based on peaks and ridges in the difference of low-pass transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(2):156–170, March 1984.

[17] James L. Crowley and Arthur C. Sanderson. Multiple resolution representation and probabilistic matching of 2-D gray-scale shape. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(1):113–121, January 1987.

[18] Luc Devroye. Automatic pattern recognition: A study of the probability of error. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):530–543, July 1988.

[19] Edward R. Dougherty. *An Introduction to Morphological Image Processing*, volume TT 9. SPIE Optical Engineering Press, 1992.

[20] Edward R. Dougherty and Jaakko Astola. *An Introduction to Nonlinear Image Processing*, volume TT 16. SPIE Optical Engineering Press, 1994.

[21] Richard O. Duda and Peter E. Hart. Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM: Graphics and Image Processing*, 15(1):11–15, January 1972.

[22] Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis*. John Wiley and Sons, 1973.

[23] Michael P. Ekstrom, editor. *Digital Image Processing Techniques*, volume 2 of *Computational Techniques*. Academic Press, Inc., 1984.

[24] John C. Feltz. Development of the modulation transfer function and contrast transfer function for discrete systems, particularly charge coupled devices. *Optical Engineering*, 28(8):893–904, August 1990.

[25] James D. Foley, Andries van Dam, Steven K. Feiner, and John F. Hughes. *Computer Graphics Principles and Practice*. The Systems Programming Series. Addison-Wesley Publishing Company, 2nd edition, 1990.

[26] John M. Gauch and Stephen M. Pizer. Multiresolution analysis of vertex curves and watershed boundaries. *SPIE Medical Imaging VI: Image Processing*, 1652:226–240, 1992.

[27] John M. Gauch and Stephen M. Pizer. Multiresolution analysis of ridges and valleys in grey-scale images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):635–646, June 1993.

[28] Michael A. Gennert, Norman Wittels, and Gary L. Leatherman. Uniform frontal illumination of planar surfaces: where to place the lamps. *Optical Engineering*, 32(6):1261–1271, June 1993.

[29] Rafael C. Gonzalez and Paul Wintz. *Digital Image Processing.* Addison-Wesley, 2nd edition.

[30] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing.* Addison-Wesley Publishing Company, 1992.

[31] William I. Grosky and Ramesh Jain. A pyramid-based approach to segmentation applied to region matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(5):639–650, September 1986.

[32] Robert M. Haralick and Linda G. Shapiro. *Computer and Robot Vision*, volume 1. Addison-Wesley Publishing Company, 1992.

[33] Robert M. Haralick and Linda G. Shapiro. *Computer and Robot Vision*, volume 2. Addison-Wesley Publishing Company, 1993.

[34] Robert M. Haralick, Stanley R. Sternberg, and Xinhua Zhuang. Image analysis using mathematical morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(4):532–550, July 1987.

[35] John H. Holland. *Adaption in Natural and Artificial Systems.* The MIT Press, Cambridge, Massachusetts, 1992 edition, 1992.

[36] B. K. P. Horn. *Robot Vision.* McGraw-Hill Book Company, 1984.

[37] Berthold K. P. Horn. Understanding image intensities. *Artificial Intelligence*, 8:201–231, 1977.

[38] G. C. Hunter, C. McDermott, N. J. Miles, A. Singh, and M. J. Scoble. A review of image analysis techniques for measuring blast fragmentation. *Mining Science and Technology*, 11(1):19–36, July 1990.

[39] E. Ivanov, D. Holm, W. Osten, and W. Jüptner. An image processing system for mineral flow control in coal mining industry. In Dmitry Chetverikov and Walter G. Kropatsch, editors, *Computer Analysis of Images and Patterns: 5th International Conference, CAIP'93*, number 719 in Lecture Notes in Computer Science, pages 716–720, Budapest, Hungary, September 1993. Springer-Verlag.

[40] Evgeni Ivanov, Wolfgang Osten, and Werner Jüptner. An image processing system for monitoring of conveyor belt systems. In *Machine Vision Applications in Industrial Inspections II*, IS&T/SPIE Symposiom on Electronic Imaging Science and Technology. SPIE, 1994.

[41] Rangachar Kasturi and Ramesh C. Jain, editors. *Computer Vision*, volume 1, Principles. IEEE Computer Society Press, Los Alamitos, California, 1991.

[42] Rangachar Kasturi and Ramesh C. Jain, editors. *Computer Vision*, volume 2, Advances and Applications. IEEE Computer Society Press, Los Alamitos, California, 1991.

[43] A. J. Katz and P. R. Thrift. Generating image filters for target recognition by genetic learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):906–910, September 1994.

[44] John M. Kemeny. Practical technique for determining the size distribution of blasted benches, waste dumps and heap leach sites. *Mining Engineering*, pages 1281–1284, November 1994.

[45] John M. Kemeny, Ashutosh Devgan, Roberta M. Hagaman, and Xingqiang Wu. Analysis of rock fragmentation using digital image processing. *Journal of Geotechnical Engineering*, 119(7):1144–1160, July 1993.

[46] T. B. Lange. *Measurement of the size distribution of rocks on a conveyor belt using machine vision.* PhD thesis, University of the Witwatersrand, July 1990.

[47] Martin D. Levine. *Vision in Man and Machine.* McGraw-Hill, Inc., 1985.

[48] Lawrence M. Lifshitz and Stephen M. Pizer. A multiresolution hierarchical approach to image segmentation based on intensity extrema. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(6):529–540, June 1990.

[49] Tony Lindeberg. Scale-space for discrete signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3):234–254, March 1990.

[50] Feng Lu, Xin Zhou, and Yong-Bao He. Image segmentation technique used in the estimation of the size distribution of rock fragments in mining. In *Proceedings of IAPR Workshop on Computer Vision: Special Hardware and Industrial Applications*, pages 351–354, Tokyo, Japan, October 1988.

[51] Petros Maragos. Optimal morphological approaches to image matching and object detection. *IEEE*, pages 695–699, 1988.

[52] Petros Maragos and Ronald W. Shafer. Morphological systems for multidimensional signal processing. *Proceedings of the IEEE*, 78(4):690–710, April 1990.

[53] D. Marr and E. Hildreth. Theory of edge detection. *Royal Society of London Proceedings B*, 207:187–217, 1980.

[54] David Marr. *Vision.* W.H. Freeman and Company, San Francisco, 1982.

[55] Takashi Matsuyama. Expert systems for image processing: Knowledge-based composition of image analysis processes. *Computer Vision, Graphics, and Image Processing*, 48:22–49, 1989.

[56] C. McDermott and N. J. Miles. The measurement of rock fragmentation using image analysis. *Departmental Magazine, Department of Mining Engineering*, pages 49–61, 1988.

[57] Peter Meer, Ernest S. Baugher, and Azriel Rosenfeld. Frequency domain analysis and synthesis of image pyramid generating kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(4):512–522, July 1987.

[58] N. J. Miles and S. T. Hall. Image processing: Diagnostic and control applications in mineral processing. Technical report, University of Nottingham, Nottingham, England.

[59] Irwin Miller and John E. Freund. *Probability and Statistics for Engineers.* Prentice-Hall International, Inc., 3rd edition, 1985.

[60] Peter Moon. An investigation into the design of a satellite based stereo imaging sensor and the use of automatic image matching in the production of digital elevation models. Master's thesis, University of Cape Town, 1993.

[61] Andrè Leon Nel. *Neural Network Edge Detection.* PhD thesis, Rand Afrikaans University, June 1993.

[62] W. Niblack. *An Introduction to DIGITAL IMAGE PROCESSING.* Prentice-Hall International (UK) Ltd, 1986.

[63] L. Nyberg, O. Carlsson, and B. Schmidtbauer. Estimation of the size distribution of fragmented rock in ore mining through automatic image processing. In *Technological and Methodological Advances in Measurement - Acta IMEKO 1982*, volume 3 of *Data Processing and System Aspects*, pages 293–302, Berlin, West Germany, May 1983. North Holland Publishing Company.

[64] A. Ord. Real-time image analysis of size and shape distributions of rock fragments. *The AusIMM Bulletin and Proceedings*, 294(1):28–31, February 1989.

[65] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-9(1):62–66, 1979.

[66] N. Ozaki, K. Harayama, W. Shinohara, and S Hayashi. Recovery boiler intelligent control. *3rd IFAC Workshop on AI in Real Time Control*, September 1991.

[67] N. Paley. Errors in sampling rock fragmentation. In *Symposium on Sampling Practices in the Minerals Industry*, Australasian Institute of Mining and Metallurgy, pages 81–87, Mount Isa, Australia, November 1992.

[68] Don E. Pearson and John A. Robinson. Visual communications at very low data rates. *Proceedings of the IEEE*, 73(4):795–812, April 1985.

[69] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, July 1990.

[70] William K. Pratt. *Digital Image Processing.* John Wiley and Sons, Inc., 2nd edition, 1991.

[71] C. Pretorius and A. L. Nel. Rock size monitoring on conveyor belt systems using neural networks. In *Proceedings of the Second South African Workshop on Pattern Recognition*, pages 100–110, November 1991.

[72] Sarunas J. Raudys and Anil K. Jain. Small sample size effects in statistical patterns recognition: Recommendations for practitioners. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(3):252–264, March 1991.

[73] RCA Corporation. *Electro-Optics Handbook.* Technical Series EOH-11. 1974.

[74] Stephen E. Reichenbach, Stephen K. Park, and Ramkumar Narayanswamy. Characterizing digital image acquisition devices. *Optical Engineering*, 30(2):170–177, February 1991.

[75] Irvin Rock and Stephen Palmer. The legacy of Gestalt psychology. *Scientific American*, 263(6):48–61, December 1990.

[76] Steven K. Rogers and Matthew Kabrisky. *An Introduction to Biological and Artificial Neural Networks for Pattern Recognition*, volume TT 4. SPIE Optical Engineering Press, 1991.

[77] Azriel Rosenfeld and Avinash C. Kak. *Digital Picture Processing*, volume 1. Academic Press, 2nd edition, 1982.

[78] Azriel Rosenfeld and Avinash C. Kak. *Digital Picture Processing*, volume 2. Academic Press, 2nd edition, 1982.

[79] Gerhard Roth and Martin D. Levine. Geometric primitive extraction using a genetic algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):901–905, September 1994.

[80] Sudeep Sarkar and Kim L. Boyer. Perceptual organisation in computer vision: A review and a proposal for a classificatory structure. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(2):382–399, March 1993.

[81] Sudeep Sarkar and Kim L. Boyer. A computational structure for preattentive perceptual organisation: Graphical enumeration and voting methods. *IEEE Transactions on Systems, Man, and Cybernetics*, 24(2):246–267, February 1994.

[82] Sudeep Sarkar and Kim L. Boyer. Using perceptual inference networks to manage vision processes. *Computer Vision and Image Understanding*, 62(1):27–46, July 1995.

[83] Linda G. Shapiro and Robert M. Haralick. Structural descriptions and inexact matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-3(5):504–519, September 1981.

[84] Martin L. Smith and John M. Kemeny. Blast fragmentation model evaluation via image analysis. In *Symposium on Emerging Computer Techniques for the Minerals Industry*, pages 229–238, Littleton, CO, USA, 1993. Society for Mining, Metallurgy and Exploration Inc.

[85] G. G. Stanley. Considerations in the application of autogenous milling. *Journal of the South African Institute of Mining and Metallurgy*, 76:53–55, October 1975.

[86] Stanley R. Sternberg. Grayscale morphology. *Computer Vision, Graphics, and Image Processing*, 35:333–355, 1986.

[87] Paul James Symonds. The investigation of the characterisation of flotation froths and design of a machine vision system for monitoring the operation of a flotation cell ore concentrator. Master's thesis, University of Cape Town, 1992.

[88] Aaron M. Tenenbaum, Yedydyah Langsam, and Moshe J. Augenstein. *Data Structures using C.* Prentice-Hall Internal, Inc., 1990.

[89] Julius T. Tou and Rafael C. Gonzalez. *Pattern Recognition Principles.* Number APPCM 7. Addison-Wesley Publishing Company, 1974.

[90] Anne Treisman. Features and objects in visual processing. *Scientific American*, 255:106–115, November 1986.

[91] Wen-Hsiang Tsai. Moment-preserving threshold: A new approach. *Computer Vision, Graphics, and Image Processing*, 29:377–393, 1985.

[92] Michael Arnoldus Heyns Venter. *Stereo Imaging in Low Bitrate Video Coding.* PhD thesis, University of Pretoria, December 1992.

[93] David Wigeson. Fragmentation analysis using computer vision techniques. Master's thesis, University of the Witwatersrand, 1987.

[94] Jeremy M. Wolfe. Hidden visual processes. *Scientific American*, 248(2):94–103, February 1983.

[95] Xingqiang Wu and John M. Kemeny. A segmentation method for multi-connected particle delineation. In *IEEE Workshop on Applications of Computer Vision*, number 92TH0446-5, pages 240–247, Palm Springs, CA, USA, November 1992.

[96] K. P. Yeo, C. C. Cheung, A. Ord, and W. A. Brown. Determination of rock fragment sizes using a transputer array. In Tariq S. Durrani, William A. Sandham, John J. Soraghan, and Sheila M. Forbes, editors, *Application of Transputers 3: Proceedings of the Third International Conference on Applications of Transputers*, Transputer and Occam Engineering Series, pages 142–147, Glasgow, UK, August 1991. Scottish Transputer Centre, University of Strathclyde, IOS Press.

[97] Brian Zuerndorfer and Gregory H. Wakefield. Extensions of scale-space filtering to machine-sensing systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(9):868–882, September 1990.

Appendix A

Definitions of Particle Size

Table A.1 gives a list of some different equivalent sphere diameters which may be used. A more complete table and analysis of this subject can be found in Allen [2, p104].

Symbol	Name	Definition	Formula
d_v	Volume diameter	Diameter of a sphere having the same volume as the particle	$V = \frac{\pi}{6} d_v^3$
d_s	Surface diameter	Diameter of a sphere having the same surface as the particle	$S = \pi d_s^2$
d_a	Projected area diameter	Diameter of a circle having the same area as the projected area of the particle resting in a stable position	$A = \frac{\pi}{4} d_a^2$
d_p	Projected area diameter	Diameter of a circle having the same area as the projected area of the particle in random orientation	Mean value for all possible orientations $d_p = d_s$ for convex particles
d_c	Perimeter diameter	Diameter of a circle having the same perimeter as the projected outline of the particle	
d_A	Sieve diameter	The width of the minimum square aperture through which the particle will pass	

Table A.1: Definitions of particle size.

Appendix B

Definitions of Particle Shape

Table B.1 gives a list of qualitative terms which have been used to give an indication of the nature of particle shape. This table is to be found in Allen [2, p107] who extracted it from the British Standard 2955: Glossary of Terms Relating to Powders.

Term	Definition
Acicular	needle-shaped
Angular	sharp-edges or having roughly polyhedral shape
Crystalline	freely developed in a fluid medium of geometric shape
Dentritic	having a branched crystalline shape
Fibrous	regularly or irregularly thread-like
Flaky	plate-like
Granular	having approximately an equidimensional irregular shape
Irregular	lacking any symmetry
Modular	having rounded, irregular shape
Spherical	global shape

Table B.1: Definitions of particle shape.

Appendix C

Ellipse Equations

C.1 Parametric Representation

A parametric representation of an ellipse is required given the major axis a, the minor axis b and the angle of rotation about the axis as θ. It is assumed that the ellipse will have no translation from the origin. The equation is expressed in terms of a parameter ϕ.

The basic equation for an ellipse is:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{C.1}$$

Rewriting (C.1) in parametric form results in:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a\cos\phi \\ b\sin\phi \end{pmatrix} \tag{C.2}$$

The final parametric form for the ellipse can be found by rotating (C.2) through an angle of θ using the standard two dimensional rotation transformation matrix.

$$\begin{aligned} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} a\cos\phi \\ b\sin\phi \end{pmatrix} \\ &= \begin{pmatrix} a\cos\theta\cos\phi + b\sin\theta\sin\phi \\ b\cos\theta\sin\phi - a\sin\theta\cos\phi \end{pmatrix} \end{aligned} \tag{C.3}$$

C.2 Distance to Ellipse

Given the parameter ϕ, it is desired to calculate the distance from the origin to the point on the ellipse specified by ϕ. The point on the ellipse can now be found using (C.3).

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a\cos\theta\cos\phi + b\sin\theta\sin\phi \\ b\cos\theta\sin\phi - a\sin\theta\cos\phi \end{pmatrix}$$

The euclidean distance from the origin is therefore calculated using the following expression.

$$\begin{aligned} r_\phi &= \sqrt{x^2+y^2} \\ &= \sqrt{(a\cos\theta\cos\phi + b\sin\theta\sin\phi)^2 + (b\cos\theta\sin\phi - a\sin\theta\cos\phi)^2} \end{aligned} \tag{C.4}$$

C.3 Ellipse Normal Direction

It is desired to find a parametric representation for direction of a normal to the ellipse at any point on its boundary. The ellipse has given parameters a, b and θ. The point on the boundary is determined using the parameter ϕ.

Given the parametric form of the ellipse from (C.3):

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a\cos\theta\cos\phi + b\sin\theta\sin\phi \\ b\cos\theta\sin\phi - a\sin\theta\cos\phi \end{pmatrix}$$

Next calculate the gradient (i.e. tangent) vector of the ellipse equation:

$$\begin{aligned} \begin{pmatrix} x \\ y \end{pmatrix} &= \frac{d}{d\phi}\begin{pmatrix} a\cos\theta\cos\phi + b\sin\theta\sin\phi \\ b\cos\theta\sin\phi - a\sin\theta\cos\phi \end{pmatrix} \\ &= \begin{pmatrix} -a\cos\theta\sin\phi + b\sin\theta\cos\phi \\ b\cos\theta\cos\phi + a\sin\theta\sin\phi \end{pmatrix} \end{aligned}$$

Finally, the normal vector is calculated as follows:

$$\begin{aligned} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} -a\cos\theta\sin\phi + b\sin\theta\cos\phi \\ b\cos\theta\cos\phi + a\sin\theta\sin\phi \end{pmatrix} \\ &= \begin{pmatrix} b\cos\theta\cos\phi + a\sin\theta\sin\phi \\ -b\sin\theta\cos\phi + a\cos\theta\sin\phi \end{pmatrix} \end{aligned} \tag{C.5}$$

C.4 Calculating ϕ Given Orientation

In order to use (C.4) and (C.5), it is necessary to be able to determine the appropriate parameter value for ϕ given the orientation of the point with respect to the origin is ψ.

A straight line through the origin with orientation ψ can be expressed using the following parametric representation:

$$\begin{pmatrix} x \\ y \end{pmatrix} = t \begin{pmatrix} \cos\psi \\ \sin\psi \end{pmatrix} \tag{C.6}$$

Equating (C.6) with (C.3) results in:

$$t \begin{pmatrix} \cos\psi \\ \sin\psi \end{pmatrix} = \begin{pmatrix} a\cos\theta\cos\phi + b\sin\theta\sin\phi \\ b\cos\theta\sin\phi - a\sin\theta\cos\phi \end{pmatrix}$$

Now ϕ can be found by equating the two equations for t and solving for ϕ.

$$\begin{aligned}
\frac{a\cos\theta\cos\phi + b\sin\theta\sin\phi}{\cos\psi} &= \frac{-a\sin\theta\cos\phi + b\cos\theta\sin\phi}{\sin\psi} \\
\sin\phi\left(\frac{b\sin\theta}{\cos\psi} - \frac{b\cos\theta}{\sin\psi}\right) &= -\cos\phi\left(\frac{a\sin\theta}{\sin\psi} + \frac{a\cos\theta}{\cos\psi}\right) \\
\text{Therefore: } \tan(\phi) &= -\left(\frac{a\sin\theta\cos\psi + a\cos\theta\sin\psi}{b\sin\theta\sin\psi - b\cos\theta\cos\psi}\right) \\
\text{And: } \phi &= \arctan\left(\frac{a\sin\theta\cos\psi + a\cos\theta\sin\psi}{b\cos\theta\cos\psi - b\sin\theta\sin\psi}\right) \qquad \text{(C.7)}
\end{aligned}$$

Appendix D

Distance Measure

It is necessary to be able to measure the distance between a pair of ellipses. This is important for the following reasons:

1. Measuring the "closeness" of fit of an ellipse to a rock by comparing it to a manually fitted ellipse.
2. For calculating the score of a tree by determining the similarity between ellipses at adjacent levels of the tree.
3. To facilitate the creation of a representative training set of samples of rock and non-rock feature vectors. The distance measure can be used to determine the actual classification of each feature vector.

The distance measure must provide an indication of the differences between two ellipses. Since the ellipses can differ with respect to several parameters, each one must be considered in the distance measure. Furthermore, the relative differences or costs caused by each parameter must be normalised to allow them to be compared. More specifically, it is desirable to normalise each of the costs to have a value of unity to correspond to the maximum difference for which the two ellipses could represent the same object. The following list shows each of the parameters that has been considered and how it has be normalised:

Position The scale-space coordinate of the centre of each ellipse is represented by $\mathbf{p}$. The cost of the difference in position can be estimated by means of the Euclidean distance metric given by $d_p = |\mathbf{p}_1 - \mathbf{p}_2|$. It has been found that if an ellipse is offset by more than one sixth of its size then it is too far to be considered as the same object and should have a cost greater than one. This can be achieved using the following modified form of the position difference cost $d_p = 6|\mathbf{p}_1 - \mathbf{p}_2|/size$ where $size$ is the size of the ellipse.

Size The size of each ellipse is $size = \sqrt{ab}$. In order to normalise the size cost, it is necessary to use a logarithmic measure so that the cost is proportional to the relationship between the sizes rather than the absolute values. For example, there should be an equal cost for a large ellipse and one half its size and for a small ellipse and one half its size. If the cost is normalised to have unit value when the ratio of size is equal to f_{limit} or $\frac{1}{f_{limit}}$ then the cost can be represented as $d_s = \ln \frac{size_2}{size_1} / \ln f_{limit}$. A value of $f_{limit} = \frac{4}{3}$ has been found to be suitable.

Aspect ratio The aspect ratio of an ellipse is measured as $e = \frac{a}{b}$ which is 1 for a circle and increases to infinity as the circle flattens to a line. Using this gives a measure of the cost of the difference between two ellipses as $d_e = |e_1 - e_2|$. The cost will be unity if one compares a circle to an ellipse in which $a = 2b$ and will be zero if the aspect ratios are equal.

Orientation The cost of an orientation difference between two ellipses is dependent on the aspect ratios of the ellipses. Clearly, if one ellipse is actually a circle, the orientation difference to any other ellipse should be zero since a circle can be any orientation. As the lower aspect ratio increases, so should the cost of an orientation difference. The following expression gives a possible cost of orientation difference $c_o = 1 - e^{1-\min(e_1,e_2)}$ for which there is no cost if either ellipse is a circle, and the cost increases exponentially to one as the lower aspect ratio value tends to infinity.

The actual cost of the difference in orientation is linearly proportional to the relative angles of the ellipses. Since the orientation of the ellipse is known to be θ, the normalised cost of orientation differences can be expressed as $d_o = c_o|\theta_1 - \theta_2|\frac{2}{\pi}$.

The overall distance is then calculated as the sum of the squares of these values.

$$d = \sqrt{d_p^2 + d_s^2 + d_e^2 + d_o^2} \tag{D.1}$$