A machine vision-based approach to measuring the size distribution of rocks on a conveyor belt

Simphiwe Mkwelo

Submitted to the Department of Electrical Engineering,
University of Cape Town, in fulfilment of the requirements
for the degree of Master of Science in Engineering.

May 2004

Declaration

I declare that this dissertation is my own work. It is being submitted for the degree of Master of Science in Engineering at the University of Cape Town. It has not been submitted before for any degree or examination at this or any other institution.

.....

Simphiwe Mkwelo

(Candidate's Signature)

Acknowledgements

I am grateful to the following people for their contribution towards this thesis and for being my personal inspiration:

- Professor Gerhard de Jager, the project supervisor, for his enthusiasm and guidance.
- Dr F. Nicolls, the project co-supervisor, for his guidance and assistance.
- Dr J.Francis for his useful comments and discussions.
- MPRU, UCT and the NRF for their financial support.
- Aubrey, Reuben, Musa and the rest of the team for their assistance in collecting sieved data at the Waterval mineral processing plant.
- The members of the Digital Image Processing group for their support and contributions.
- My family for their love and support.
- Last but not least, all my friends for their love and support.

Abstract

This work involves the development of a vision-based system for measuring the size distribution of rocks on a conveyor belt. The system has applications in automatic control and optimization of milling machines, and the selection of optimal blasting methods in the mining industry. Rock size is initially assumed to be the projected rock surface area due to the constraint imposed by the 2D nature of images. This measurement is facilitated by locating connected rock-edge pixels.

Rock edge detection is achieved using a watershed-based segmentation process. This process involves image pre-filtering with edge preserving filters at various degrees of filtering. The output of each filtering stage is retained and marker-driven watersheds are applied on each output resulting to traces of detected rock boundaries. Watershed boundary selection is then applied to select boundaries which are most likely to be rock edges based on rock features. Finally, rock recognition using feature classification is applied to remove non-rock watershed boundaries.

The projected rock area distribution of a test-set is measured and compared to corresponding projected areas of manually segmented images. The obtained distributions are found to be similar with an RMS error of 2.37% on the test-set. Finally, sieve data is collected in the form of actual rock size distributions and a quantitative comparison between the actual and machine measured distributions is performed. The overall quantitative result is that the two rock size distributions are significantly different. However, after incorporating a stereology-based correction, hypothesis tests on a 3m belt-cut test-set show that the obtained distributions are similar.

List of Figures

List of Tables

Table of Contents

Chapter 1

Introduction

1.1 Problem statement

There is a need for an instrument that automatically measures the size distribution of rocks on a conveyor belt in the mineral processing and blast fragmentation industries. In mineral processing, the output of such an instrument can be used to adjust milling parameters for optimum crushing performance, while in blast fragmentation it can be used to evaluate the effectiveness of blasting procedures [50].

In the mineral processing industry, sieving methods are perceived as the most accurate means of measuring rock-size distributions because each particle in the sample is measured. However, for an accurate and representative measurement of the overall plant ore size distribution, long belt-cuts [1] have to be taken. This is an undesirable situation because it affects production. As a result, plant managers are very reluctant to perform these long belt-cuts and instead opt for short belt-cuts. The problem with short belt-cuts is that they do not capture the overall ore-size distribution variability of the plant, and as a consequence the milling parameters thus obtained may not improve milling performance significantly, if at all. Therefore an automated solution is highly needed in this industry. In blast fragmentation analysis, sieving methods are not viable as extremely large rock fragments are produced. Thus there is no alternative solution in this application.

[1] Belt-cut is a term used in the comminution work which refers to the amount of the material to be sieved in terms of conveyor-belt length.

An automated rock-size measurement system should meet the following requirements:

- The measurement should be non-intrusive so that there is no wear and tear of mechanical parts and no degradation of materials under measurement.
- Measurements should be performed in near real time to facilitate quick adjustments of process parameters for control purposes.
- The instrument should be able to withstand the harsh conditions of a mineral processing plant.
- Production should not be affected by the operation or failure of the instrument.

1.2 Background

In the mineral processing industry, valuable minerals such as gold and platinum are extracted using high electricity consumption machines during comminution [11]. In the comminution stage, milling machines crush ore to a desired particle size for the liberation of the valuable mineral in the subsequent stages of the mineral extraction process [65]. In the absence of any form of mill control, the product size particles will either be ground to an unnecessary degree or the mill discharge will be too coarse due to inadequate residence-time [2] in the mill [65]. Both conditions are undesirable: the former is expensive due to excessively high electrical energy consumption, while the latter has too low a degree of liberation for separation and results in poor recovery in the concentration stage. Therefore some form of mill control is needed to keep the mill discharge particle size at a desired set-point.

There are two common types of mills in the mineral processing industry namely autogenous (AG) and semi-autogenous (SAG) mills [65]. The AG mill grinds ore using the ore itself while the SAG uses steel rods or balls as grinding media in addition to the ore itself. It is understood that both types of mills require certain rock size distribution profiles for proper operation [65]. After the detection of an improper size distribution profile, rock selection procedures can be

[2]Residence-time is a term used in comminution which refers to the time interval from the entrance of the material into the mill until it is discharged.

used for correction purposes [38]. The instrument can also be used for monitoring the input and output size distributions of systems such as crushers. Alternatively, the instrument can be incorporated as part of the mill control system, where the rock size distribution is a control variable to compensate the milling action for various input rock size distributions. This forms the main motivation for undertaking this research.

This dissertation proposes a machine vision-based solution to the rock size distribution measurement problem. This approach uses tools from machine vision to analyze scenes of rocks on a conveyor belt in order to facilitate the measurement of rock-size. This approach has been adopted by other researchers such as Crida[11] and Lange[38]. However, due to the limitations of computer hardware at the time, the complexity of the image analysis procedures was highly restricted. Most of the engineering was based on striking a compromise between speed and accuracy, with the accuracy component suffering most of the time. The complexity of the solutions were highly limited resulting in poor accuracies.

Nowadays recent developments in computer hardware have resulted in huge improvements in computing power and speed. This has led to the development of commercial machine vision-based rock-sizing instruments such as Split-engineering and Wipfrag [62, 63]. These systems have attracted great attention in the mineral processing and blasting industries where manual sizing procedures are not viable options. The long-term objective is to optimize the system implemented in this work to form a robust commercial rock-sizer product that can compete with these commercial rock-sizers.

1.3 Thesis objectives

Based on the above problem statement and background, this dissertation aims at developing a machine vision-based instrument for measuring the size distribution of rocks on conveyor-belt. The objectives of this work are to:

- Review machine vision in general, and specifically review previous work on the development of rock-sizing instruments using machine vision.
- Investigate image processing techniques that can be used as building blocks.

- Investigate the use of pattern recognition tools for rock recognition.
- Outline the implementation of the adopted segmentation approach.
- Implement various pattern classification approaches and compare them on the collected data.
- Empirically convert the 2D rock size distribution of sections to a 3D rock size distribution.
- Test the system on "real world" plant data and quantify its performance.
- Finally, draw conclusions and make recommendations

1.4 Thesis Format

The following is the format of this thesis document.

Chapter Two: Presents a literature review of Machine vision by introducing its main building blocks namely illumination control, digital image processing and object recognition. Finally, previous work on measuring rock size distributions using machine vision is presented.

Chapter three: Presents the theory behind a selected set of traditional digital image processing procedures and investigates their possible application to rock scene segmentation. In particular, the techniques under investigation can be broadly categorized into gray scale image thresholding and edge detection.

Chapter four: Investigates image de-noising methods. The selected set of approaches is divided into linear and nonlinear filtering methods. A qualitative comparison is carried out and the suitable approaches for rock-scene filtering are selected.

Chapter five: Investigates the use of mathematical morphology tools for the segmentation of rock scenes. The theory underlying the basic and advanced tools is presented. Finally, the most powerful tools in the form of the reconstruction operator and the watershed transform are investigated.

Chapter six: Presents the implementation details of the adopted rock scene segmentation approach. It is divided into preprocessing for estimating rock locations and rock edge tracing with a watershed-based segmentation approach. The implemented algorithm is tested for robustness to varying lighting conditions.

Chapter seven: Presents the background theory of a selected set of pattern recognition techniques. The set of classifiers under investigation includes k-nearest neighbor (KNN), the probabilistic neural network (PNN), regularized least squares classification (RLSC) and support vector machines (SVM) in the form of the kernel adatron. Feature vector dimensionality reduction methods are investigated for improving class separability in the input feature space. Data acquisition, feature measurement and classifier training for rock recognition are carried out. Finally, the classifiers are tested for generalization and the test results are presented.

Chapter eight: Presents the overall results of the dissertation. The rock size distributions are measured and compared to sieved data. A stereology-based "unfolding" method for obtaining a 3D rock size distribution from 2D size distributions of sections is used to correct for obvious errors. The performance of the modified system is evaluated on a 3m belt-cut test-data.

Chapter nine: Draws conclusions based on the results of the dissertation and makes recommendations for future developments.

Chapter 2

Relevant literature review

2.1 Introduction

Machine Vision is probably the most suitable solution for the the rock-size distribution measurement problem simply because it has many attractive advantages. In [11], the following advantages are stated:

- A non-intrusive measurement is performed on-line without stopping the conveyor belt and disturbing the feed to the mill.
- The instrument takes measurements continuously for 24 hours a day.
- The obtained results are consistent given the same scene with the same lighting conditions.
- There are no moving mechanical parts and thus reduced maintenance.

Even though these advantages are valid enough to justify the use of this approach, two major disadvantages are revealed in the literature. The first one is that only a two dimensional (2D) view of the scene is available. Thus the actual size distribution of the material cannot be measured, due to "fragment overlap" and occlusion of material underneath the surface of the stream. The second is that due to the camera's limited resolution, rock sizes below a certain threshold cannot be measured. The details of other "inherent sources of error" are provided in [43]. However there is a consensus in the field of machine vision for rock-size measurement that a useful measure of the actual distribution can be obtained [38].

In this chapter, a review of the literature on machine vision in general is given in section 2.2. This is followed by previous work on the measurement of rock size distribution using machine vision in section 2.3.

2.2 Machine Vision

Shapiro and Haralik define machine vision as a science that develops algorithms for automatically extracting and analyzing useful information from observed images using a computer [26, 27]. A typical machine vision system consists of a controlled illumination arrangement, a video camera and a personal computer with a grabber card. The camera captures the scene and transforms it to an analog video signal which is then fed to a frame grabber card for the construction of the final digital image [18]. The overall physical set up is shown in figure 1. It has been identified in [11] that a typical machine vision system executes the following processes in the specified order:

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

Except for image capture, most of the processes are executed in software.This implies that the design of suitable lighting conditions is not incorporated into this approach. However from intuition, and from previous work on the effect of lighting on machine vision systems[19, 17, 25, 18], it is clear that the design of a suitable lighting arrangement should be incorporated as part of the system. In general, the lighting arrangement should emphasize features which are important to the segmentation and subsequent recognition of objects of interest.

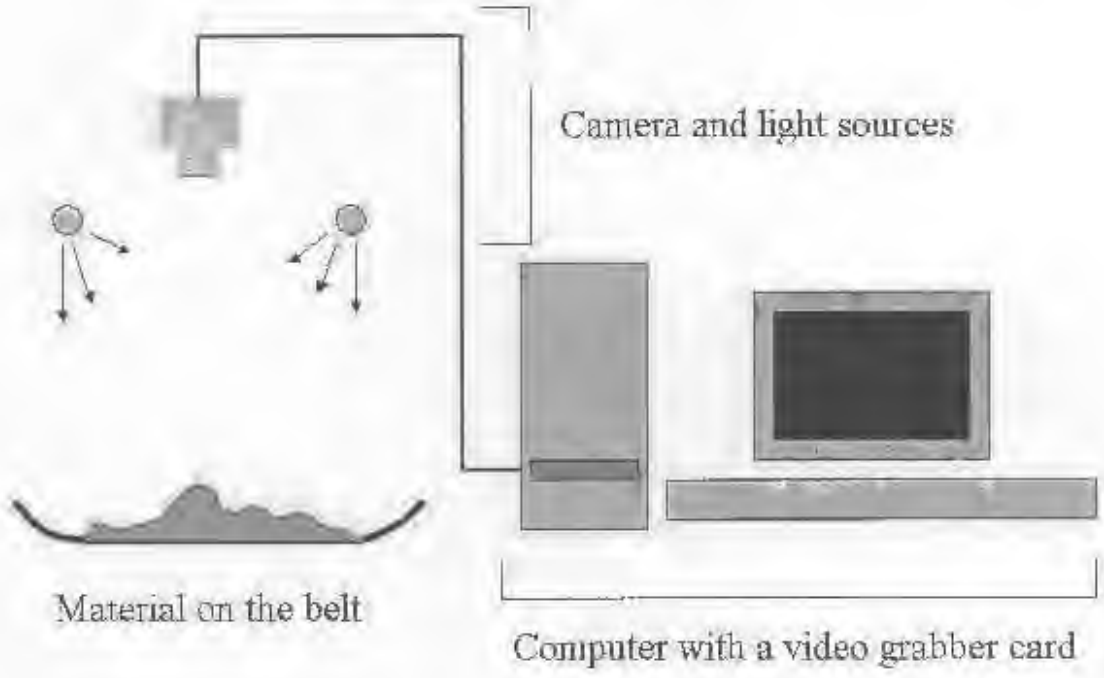


Figure 2.1: A Typical Machine vision system for rock sizing

In the next sections an introduction to the major building blocks of a machine vision system is given in the form of controlled illumination, digital image processing and pattern classification.

2.2.1 Controlled illumination

The appearance of objects in images is due to the reflectance properties of objects under imaging. For there to be reflection there should incident illumination. During the image formation process, the reflection of light on the object's surface (radiance) is captured by the light sensors of the camera. In a charge-coupled device (CCD) camera, the light sensors form a rectangular grid of electron collection sites [18]. Each site has three layers: the thin silicon wafer, on top of which is the layer of silicon dioxide, and finally the conductive gate structure deposited on the dioxide [18]. The light energy falling on each collection site is measured. The measured light is then converted to an analog voltage which is sampled and digitized for computer representation.

The brightness intensity at a point in an image depends on the characteristics of object's reflectance and the geometry of the imaging process [1]. The gray-level is perceived as a quantized measurement of image irradiance and is proportional to scene radiance. It is further stated that the gray-level f at an image point, is proportional to the scene irradiance E (the amount of light incident on the objects) and the reflectance r at the corresponding world point x. It is given by:

$$f(\mathbf{x}) = E(\mathbf{x})r(\mathbf{x}). \tag{2.1}$$

The irradiance E at x is the sum of the contributions from all the light sources, while the reflectance r is the portion of the irradiance that is reflected toward the camera [1]. The scene irradiance E has a low spatial variation across the image, while the reflectance component varies with surface orientation and thus changes quickly at the object edges [1]. It is apparent that the scene irradiance is the so called "Background illumination". An attempt to remove it is usually performed during digital image processing by a filtering operation, which models the structure of the background illumination and subtracts it from the acquired image.

In summary, there is a strong link between the captured image and the illumination conditions under which it was taken. This relationship is exploited whenever possible in machine vision applications [19, 25, 17, 18]. As far as the the application of measuring rock size distributions is concerned, the illumination arrangement should enforce the appearance of shadows around each rock in the scene to facilitate rock edge detection.

2.2.2 Digital image processing

Once the image is acquired in digital form, it is processed using digital image processing routines with the final goal of image segmentation. The initial processing is usually a filtering operation to reduce the noise which is introduced by the image formation process. This process is noisy due to sampling, quantization and random disturbances in the capture hardware. Filtering has its own associated degradations such as blurring the edges of objects. At this point the filtered image is then segmented into disjoint regions using a possible combination of edge detectors, thresholding techniques, morphological operations and other image processing transforms. This form of processing can be classified as low-level processing where the objects of interest are revealed. There is no object recognition or feature-based classification of regions.

The reader is referred to [31, 8, 58, 21] for digital image processing concepts, algorithms and applications.

Digital image processing of rock-scenes aims at revealing objects of interest which are rocks in this application. The key assumption is that of a controlled illumination set up, where shadows form rings around rocks at this stage.

2.2.3 Pattern recognition

The purpose of pattern recognition in this work is to distinguish between the segmented regions that represent the object of interest and those that do not. The classification is based on region properties or features, such as shape, texture, edge and regional gray value characteristics. Each region is viewed as a point in an N-dimensional feature vector space where N is the number of properties. Neighboring points are expected to be similar and thus form a cluster with the same class label.

In supervised learning, the training data is labelled and a functional mapping is learned which maps the measured properties to the labels. In unsupervised learning the training data is not labelled. Instead clustering algorithms are used to partition the feature vector space into a number of clusters and a new test point is assigned to a cluster closest to it. In the case of supervised learning, the system is tested for generalization so that it responds well to unseen data (data which is not in the training set). There are many ways of obtaining this mapping and this distinguishes the various classifiers. The reader is referred to [10, 55, 12, 53] for concepts, algorithms and applications of pattern recognition techniques.

The next section presents previous work on measuring the size distribution of rocks.

2.3 Previous work on Machine vision for rock-sizing

The quantification of rock fragmentation is not a new problem and this is evidenced by the amount of publication on the matter [62, 63, 38, 11, 4, 16, 46, 36, 54, 64, 32, 14, 37, 22]. The work on particle sizing using machine vision ranges in time from 1976 to 2004. It began in the mid seventies with the ARMCO Autometrics MSD-95 material size distribution transmitter

[32]. This instrument was developed at the Julius Kruttschnitt Institute of Australia and later manufactured by ARMCO. Its operation involves using 1-D signal processing methods to determine rock chord lengths. This is achieved by scanning the conveyor belt along a straight line parallel to and at the center of the belt using an optical sensor. Shadows and gaps between rocks are emphasized by a low incident lighting arrangement. This results in the output of the optical sensor being a 1-D signal which assumes high values in the presence of a rock and low values in a shadow region. This output is then compared to a threshold such that higher outputs are deemed to be rocks. The threshold comparator output is a series of pulses whose durations are the times it took the optical signal to stay above the threshold. These durations are measured and knowing the speed of the conveyor belt, the rock chord lengths are determined. Even though this system is fast enough to operate on-line in real-time, it has obvious limitations. The list of limitations includes: Ignoring rocks which are outside the probe line resulting in a bias in the measurement, the measured chord-lengths are not representative of the measured rocks resulting in a bias to smaller particles, the use of absolute intensity for rock detection may not be appropriate for different colors of ore and may therefore result to inaccurate chord lengths, and finally low intensity values do not necessarily represent gaps or shadows between rocks.

In this section machine vision approaches to rock size measurement from selected academic and commercial institutions are presented. These approaches were developed between the years 1990 and 2004.

2.3.1 The University of Witwatersrand

In this institution, Lange [38] developed a machine vision system for the measurement of rock size distributions in 1990 as part of his PhD research. His system involves a sequence of operations which commences with image low pass filtering using a neighborhood averaging filter to remove impulse noise. At this point the filtered image is processed in parallel by 2 processes: the first is a morphological gradient operation followed by thresholding for rock edge detection, the second is a process which computes 2 thresholds from the histogram of the image. The higher threshold highlights bright intensities and the lower highlights darker areas.

The information contained in the resultant set of images consists of detected edges, bright and dark areas. These images are combined to correct the edge detection image. The output of

this correction stage is the final segmented image. Measured cord-lengths of the closed regions are then used to estimate the sieve size distribution of the material on the conveyor belt. The conclusion is that the accuracy of the system is limited by the inaccuracy of the thresholding and edge detection processes, in addition to the conversion from 1-D to 3-D distributions of the chord-length measurements.

2.3.2 The University of Cape Town

Crida [11] developed a machine vision system for rock fragmentation in 1995 as part of his PhD research. He adopted a diffuse lighting arrangement where the response to specular reflection is low, and shadows separate rocks lying close together. The image processing is based on the Human Visual System (HVS) in that it has a pre-attentive stage and incorporates an attention focus stage.

The processing begins with the generation of a multi-scale pyramid, where the image is filtered with varying degrees to produce n-filtered images. The top image of the stack is the least filtered and has the largest size, while the highly filtered image at the bottom has the smallest size. Rock edges are highlighted using adaptive thresholding to obtain rock outlines. Elliptically shaped outlines are detected on each image in the scale-space using the Hough transform.

Spurious ellipses are detected by comparing the shape of each ellipse to a corresponding outline on the thresholded image and are subsequently eliminated. These ellipses encode expected size, position and shape of rocks. This is followed by attention focusing, comprised of rock edge tracing followed by region classification for removing false alarms. The rock edge tracing step is guided by the knowledge encoded in the ellipses to determine rock-edge pixels. Finally, the potential rock regions on each scale-space image are combined using a hierarchical procedure to select regions which are most likely to represent rocks.

No attempt is made to estimate a 3D rock-size distribution. Instead, the system performance is quantified by comparison to manually segmented data. The amount of fines [1] in the image

[1] Fines are very small particles which are typically less than a millimeter in diameter. However, in this work, fines are regarded as the particles that cannot be measured by the system, due to limited resolution.

are not estimated. In addition, the system tends to split large non-elliptical rocks due to its bias towards detecting elliptical rocks.

2.3.3 The Wipfrag commercial system

Wipfrag is a commercially available machine vision system for measuring rock fragmentation for the milling and blast fragmentation industries [42]. It was developed by Franklin and Maerz in collaboration with the University of Waterloo and industrial groups from the blasting, mining and mineral processing industries. The inner workings of the image analysis have not been revealed in their publications.

The following publications address issues related to the workings of their image-based granulometry system: mapping from 2-D distributions of rock sections to 3-D rock-size distributions using stereology and geometric probability principles [40]; inherent sources of error of optical digital fragmentation measuring systems [43]; case studies using the Wipfrag system [51]; image sampling techniques [39]; system calibration [44]; online fragmentation analysis [42]; and aggregate sizing and shape determination [41].

In [40], it is discovered that many of the underlying assumptions of stereology and apriori knowledge of geometry probabilities are violated. In response to this, a new method of unfolding a 3D distribution from a 2D size distribution of sections is proposed. The transform function involves an empirical calibration to compensate for missing fines and overlapping hidden particles. The main drawback of this approach is the assumption that rocks are spherical particles.

2.3.4 The commercial Split-Online system

The Split-Online machine vision system was made available for commercial purposes in 1997 after eight years of research [62]. It was initially developed at the Department of Mining and Geological Engineering, University of Arizona. This is currently the state-of-art in the field of rock fragmentation quantification [62]. Their image analysis processes are not revealed in detail. From [62] it appears that some form of pre-processing is followed by an automatic

thresholding procedure to prepare the image for delineation. This consists of 4 steps: gradient filter, shadow convexity analysis, split algorithm and watershed algorithm.

The output of their watershed algorithm is a binary image where black areas denote fines and the white areas represent rocks. At this point, user intervention is required to correct for the inaccuracies of the system due to poor lighting, quality of the image, etc. Since the information on the inner workings is not supplied, its successes and failures cannot be provided.

2.4 Summary

This chapter presented a literature review of machine vision in general, and in particular provided descriptions of its fundamental components. Images of an illuminated scene are captured, processed by a desktop computer using image processing tools to reveal objects and further processed to recognize objects of interest.

Previous work on the development of a vision-based rock-sizing machine is presented. It must be stated that the list is not exhaustive but rather presents systems from both academic and commercial perspectives. However, commercial institutions tend to reveal little information about the inner workings of their systems and as a result a proper review of their methods is not possible.

In closing, the unifying theme is that machine vision in general involves illumination control for enhancing particular object features, digital image processing for extracting such objects, and finally some form of feature-based classification for object recognition.

Chapter 3

Traditional Image Processing

3.1 Introduction

Digital image processing involves the development of computer algorithms for image analysis and the subsequent extraction of the objects of interest [8]. The core processes are usually image enhancement and segmentation. In this work, digital image processing is divided into traditional and morphological image processing. Traditional image processing techniques are regarded as the very old techniques which are commonly used in the image processing field. This chapter is concerned with the provision of theoretical background of traditional image processing techniques that can be used as the building blocks of the rock-scene segmentation algorithm.

An image is considered as a continuous 2D function $f(x,y)$ where x and y are the horizontal and vertical spatial displacements. A digital image is a sampled and quantized version of the continuous $f(x,y)$. This digital image is then stored on computer memory as a 2D array of picture elements or pixels. Figure 3.1 shows an example of a digital image in the form of a scene of rocks on a moving conveyor belt.

The scene consists of fines and rocks of various sizes, brightness values and surface textures. The smaller rocks appear to have the same brightness as that of fines. These were pre-classified as fines by a grizzle [1] for storage purposes [38] before being loaded on the conveyor. The image is taken from a video sequence of frames taken during a sieving test-work exercise at the Waterfall plant in Rustenberg.

[1] A grizzle is a form of screen or sieve which is used to allow rocks of certain size to pass through.

Figure 3.1: A scene of rocks on a conveyor belt

The main objective of the image processing in this context is to extract rock edges to facilitate the measurement of enclosed areas as rock size. As means of achieving this objective, two main approaches in the form of edge detection and grey-scale thresholding are investigated. This chapter is mainly divided into two sections, a background theory on gray level thresholding techniques is given in 3.2 and edge detection techniques in 3.3.

3.2 Gray level Thresholding

Thresholding is the simplest form of image segmentation where gray value thresholds are selected for discriminating between two or more classes of gray-levels in an image [11]. The simplest case is bi-level thresholding where a single threshold is selected for discriminating between two classes of gray-levels. The threshold is usually chosen from the gray value distribution or histogram of the image. These methods are called image histogram dependent threshold selection methods in the literature.

Most of the techniques, such as Tsai's moment preserving and Otsu's discriminant analysis methods assume that the histogram of the image is bimodal and attempt to extract a gray value with the minimum frequency count between the two peaks as the threshold [34]. The underlying assumption is that the background and the foreground gray values are generated from two different gaussian distributions resulting in a clear distinction between foreground and background[34]. However, this assumption does not always hold, particularly for complex

scenes such as a scene of rock assemblages which tend to have rocks with similar gray values as the fines.

The rock-scene image of figure 3.1 is a particular example where global thresholding is not expected to be successful, due to most of the rocks appearing to have approximately the same brightness as the fines. As a consequence the resultant rock-scene segmentation is inaccurate as shown in figure 3.2. In this figure, the threshold is manually tuned to be the gray value at the local minimum between the two peaks. An ideal case is the situation where one peak represents the shadow regions and the other represents the rocks; a threshold in the middle would then result in a good segmentation output where rocks are revealed.

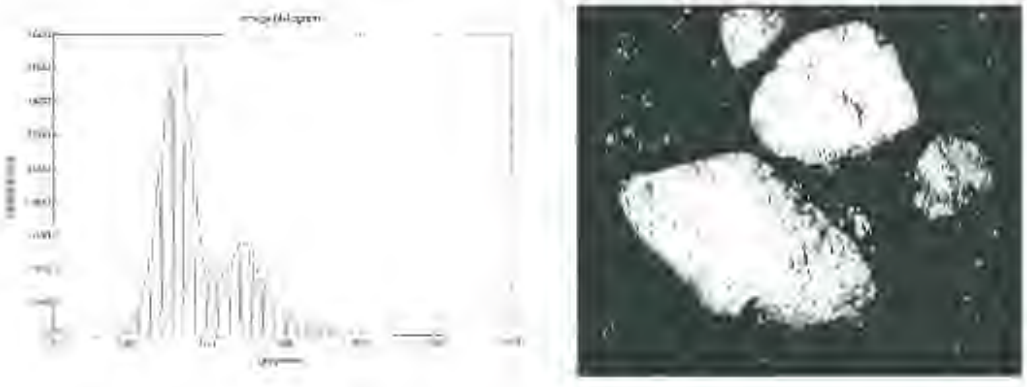


Figure 3.2: A scene of rocks on a conveyor belt, its corresponding histogram with a manually selected threshold of 107 and the resultant binary image. It can be seen that selecting the optimal value of T as the minimum gray-level between the 2 peaks does not reveal all the rocks in the image.

From here onwards, background theory on various thresholding techniques is presented.

3.2.1 Moment Preserving technique

The moment preserving threshold technique was developed by Tsai [11] for selecting an optimal threshold given an input image. It uses the first four moments of a gray-level probability distribution namely the area, mean, variance and skewness. For Bi-level thresholding it assumes that the gray-level probability distribution of an image is bimodal where the two mean gray-levels of the foreground and background, and the threshold T are estimated using the following procedure: A gray-level probability distribution is computed from the image histogram using

$$P_j = \frac{n_j}{N} \tag{3.1}$$

where n_j is the frequency count of pixels with gray-level j, N is the total number of image pixels and j ranges from 0 to 255 gray-levels. The first 4 moments are then computed using

$$m_i = \sum_{j=0}^{255} j^i P_j \tag{3.2}$$

where m_i is the i^{th} moment with i ranging from 0 to 3. This technique requires that these moments be preserved after thresholding. In a bi-level thresholding case, a single threshold T is used to classify two classes with mean gray-levels w_0 and w_1. The first four moments of the resultant binary image are given by

$$m'_i = \sum_{k=0}^{1} w_k^i p_k \tag{3.3}$$

Here p_0 is the fraction of gray-levels below T, p_1 is the fraction above T and i runs from 0 to 3. Equating (3.2) and (3.3) produces four equations which are called moment preserving equations for bi-level thresholding[11]:

$$p_0 + p_1 = 1 \tag{3.4}$$

$$p_0 w_0 + p_1 w_1 = m_1 \tag{3.5}$$

$$p_0 w_0^2 + p_1 w_1^2 = m_2 \tag{3.6}$$

$$p_0 w_0^3 + p_1 w_1^3 = m_3. \tag{3.7}$$

The fraction of gray-levels less than the threshold T is given by

$$p_0 = \sum_{j=0}^{T} P_j. \tag{3.8}$$

The solution of the moment preserving equations determines the value of p_0, which is then used in equation (3.8) to determine the threshold T. The details of the procedure for solving the moment preserving equations is outlined in [11]. Applying this technique to the image under analysis we get the following variables of interest from the solution of the moment preserving equations.

Table 3.1: Variables of interest of the moment preserving technique

variable	value
T	112
w_0	78.2
w_1	143.3
p_0	0.74
m_1	95

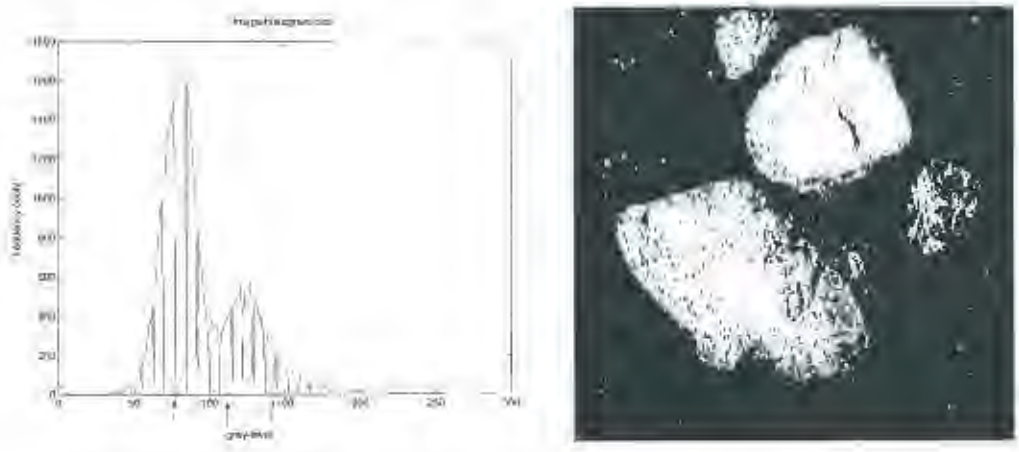


Figure 3.3: Moment preserving threshold output. The threshold is determined to be 112. The background and foreground mean gray-levels are found to be 78 and 143 respectively.

The results show that a global threshold determined from the moment preserving technique does not reveal most of the rocks. A global threshold of 112 is higher than the manually selected value of 107 and thus reveals rocks even less than before. The image under analysis has rocks which appear to be covered by fines, and these are the rocks which are not detected.

3.2.2 Discriminant Analysis technique

Otsu proposed a threshold selection method which determines the gray-level threshold using discriminant analysis theory [34]. The following mathematical formulation of this technique is obtained from [34].

As in the moment preserving technique discussed in the subsection 3.1, the gray-level histogram is normalized and regarded as a gray-level probability distribution using equation 3.2.1. For Bi-level thresholding the image pixels are classified into two classes C_0 and C_1, where the probabilities of class occurrences ω_0 and ω_1 are determined from

$$\omega_0 = \sum_{i=0}^{I} p_i. \tag{3.9}$$

$$\omega_1 = \sum_{i=T+1}^{255} p_i \tag{3.10}$$

The class means μ_0 and μ_1 are defined as:

$$\mu_0 = \sum_{i=0}^{T} \frac{ip_i}{\omega_0} = \frac{\mu(T)}{\omega(T)} \tag{3.11}$$

and

$$\mu_1 = \sum_{i=T+1}^{255} \frac{ip_i}{\omega_1} = \frac{\mu_{tot} - \mu(T)}{1 - \omega(T)} \tag{3.12}$$

where

$$\omega(T) = \sum_{i=0}^{T} p_i \tag{3.13}$$

$$\mu(T) = \sum_{i=0}^{T} ip_i \tag{3.14}$$

and

$$\mu_{tot} = \sum_{i=0}^{255} ip_i = \omega_0\mu_0 + \omega_1\mu_1 \tag{3.15}$$

The between-class variation or class separation is then computed as:

$$\sigma_B^2 = \omega_0(\mu_0 - \mu_{tot})^2 + \omega_1(\mu_1 - \mu_{tot})^2 \tag{3.16}$$

which is then written in terms of T-values as:

$$\sigma_B^2 = \frac{[\mu_{tot}\omega(T) - \mu(T)]^2}{\omega(T)[1 - \omega(T)]} \tag{3.17}$$

An optimal threshold value is determined by maximizing the between-class variance σ_B^2 to obtain T_{opt}:

$$T_{opt} = max_{0 \leq T \leq 255}[\sigma_B^2(T)] \tag{3.18}$$

The result of applying this technique is shown in figure 3.4.

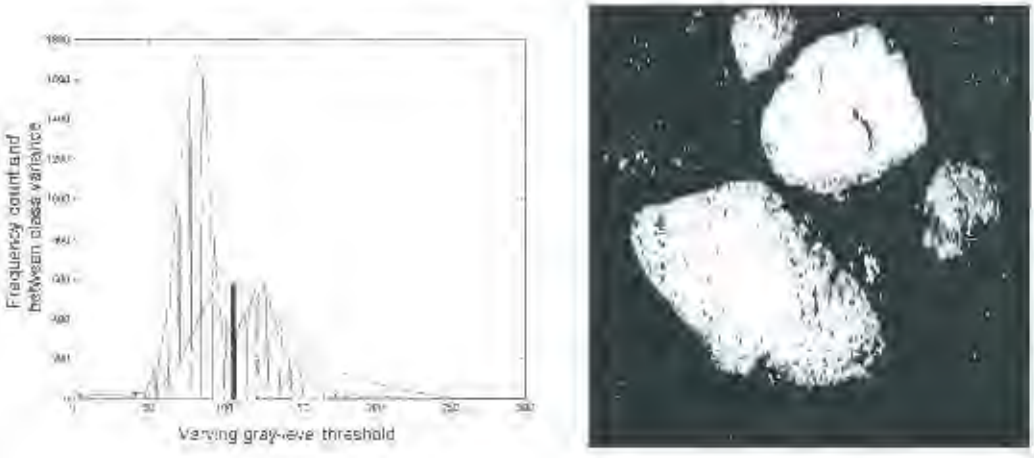


Figure 3.4: A Binary rock scene obtained using Otsu's method with an optimal threshold value of 106.5 and its corresponding histogram plot shown with the between-class variance on the same set of axes

The results show that a global threshold value determined from the discriminant analysis technique does not reveal all of the rocks. The global threshold obtained is 106.5 and thus the same rocks are revealed as for the manual procedure.

3.2.3 Adaptive thresholding

In order to extract regions of varying brightness and contrast, an adaptive gray level threshold which varies depending on pixel neighborhood statistics is desirable. The simplest and most

widely used statistic is the mean or the moving average. Other variants include the adaptive gaussian thresholding, where the weights decay as a gaussian function of the distance from the kernel center [30].

Adaptive mean thresholding

This form of thresholding is implemented by firstly highly smoothing the image with a large uniform or moving average filter, and using the blurred image for thresholding the original. The sliding neighborhood kernel has equal weights that sum to unity and its shape is square. Each pixel intensity in the smoothed image is computed as the weighted average gray level of the pixels within the sliding neighborhood on the input image $I(x,y)$ using the convolution operation

$$I_s(x,y) = I * K = \sum_i \sum_j I(i,j)K(x-i,y-j). \tag{3.19}$$

Here $I_s(x,y)$ is the smoothed image and $K(x,y)$ is a p by p normalized mask with p being an odd integer. The original image is thresholded with I_s as the adaptive threshold, and the output is the binary image B where for all indices i and j:

$$B(i,j) = 1 \text{ if } I(i,j) > I_s(i,j) \tag{3.20}$$

and

$$B(i,j) = 0 \text{ if } I(i,j) \leq I_s(i,j) \tag{3.21}$$

Results after applying this scheme to the rock scene image under analysis are shown in figure 3.5.

It appears that small kernel window sizes detect small structures, such as small rocks and lines at the expense of large rocks being split and noise detected. On the other hand, larger window sizes reveal big rocks with high success at the expense of smaller structures being missed.

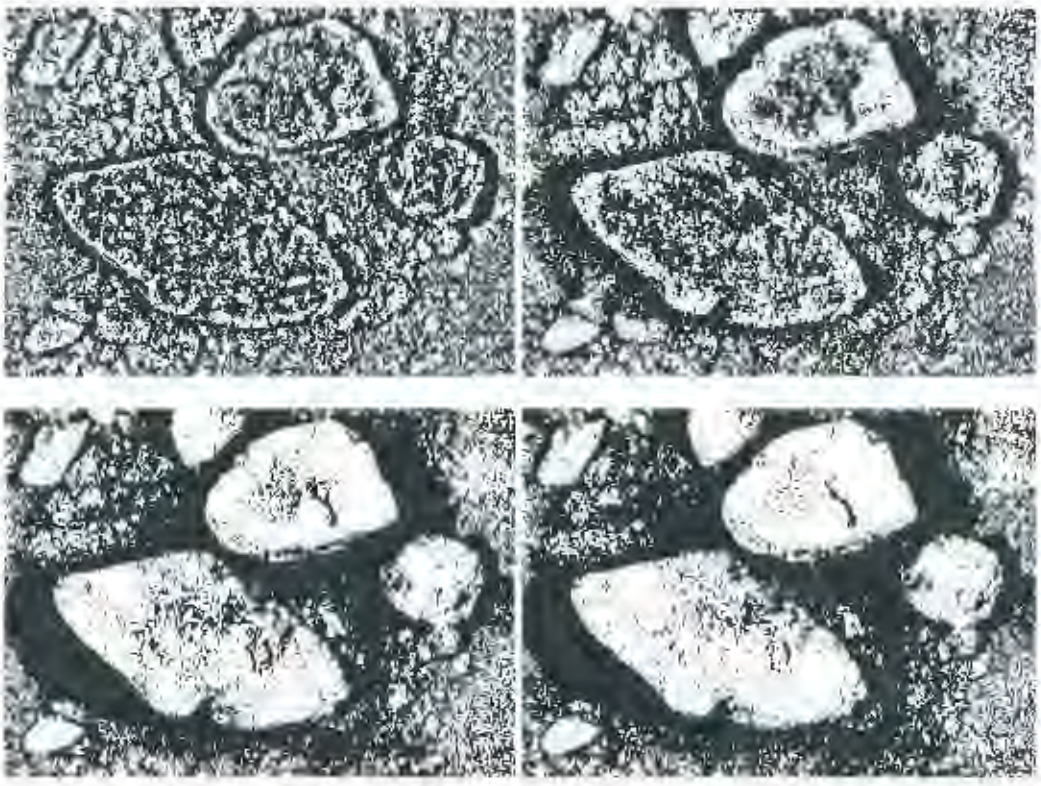


Figure 3.5: Adaptive mean thresholding outputs with kernel window sizes of 15x15 and 31x31 at the top from left to right and kernel window sizes of 75x75 and 91x91 at the bottom from left to right.

3.2.4 The two-window adaptive thresholding technique

It is of utmost importance that all the rock sizes within the image are detected. In this work, a two window adaptive thresholding technique is implemented based on this requirement. It uses the observation pointed in previous subsection, namely small kernel window sizes tend to detect small rocks while larger window sizes detect larger rock sizes.

Depending on the size of the image under analysis, two windows of sizes w_1 and w_2 are selected. Two successive adaptive thresholding processes are applied on the input resulting to two binary images. A logical OR operator is then used to combine the two binary outputs to improve the connectivity of detected pixels. The final output of applying this technique to the image under analysis is shown in figure 3.6. The window sizes of 25x25 and 95x95 were used.

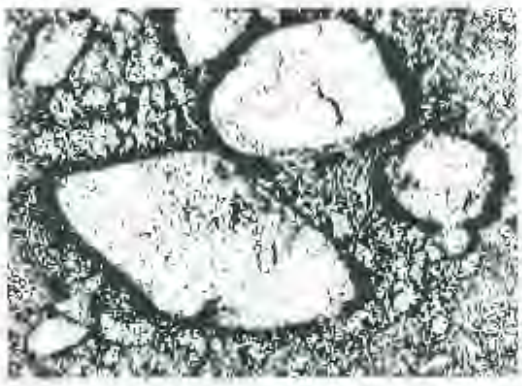


Figure 3.6: The output of the two-window adaptive thresholding technique

3.3 Edge detection

Edge detection is required in this application to trace the edge of each rock in the image so that the areas of the enclosed regions can be accurately measured. It is one of the oldest segmentation techniques in computer vision, together with greyscale thresholding[13]. However, it is expected to produce more accurate results than thresholding in this application due to the variations in gray levels across surfaces inhibiting the success the thresholding approach. Edge detection involves detecting gray level discontinuities in an image with the aim of revealing objects of interest.

There are two types of edge detectors and these are template matching and differential gradient or Laplacian techniques [1]. Differential gradient operators use two masks to detect edges in the vertical and horizontal directions. On the other hand, the template matching approach can use up to 12 masks for detecting edges in various directions [13]. This work only investigates the differential gradient techniques due their simplicity and easy implementation. The original image $I(x,y)$ is convolved with masks $K_i(x,y)$ which detect edges in the horizontal and vertical directions. The outputs $O_i(x,y)$ are combined via some function G to obtain the edge detection map.

$$O_i(x,y) = I(x,y) * K_i(x,y) \tag{3.22}$$

$$O = G(O_1, O_2) \tag{3.23}$$

A commonly used combining operator is the rms (root mean square) operator defined as:

$$O = \sqrt{O_1^2 + O_2^2}. \tag{3.24}$$

It is common practice to threshold the edge strength map so that the strongest edges are detected and the noise is suppressed. The noise is a result of the noise enhancing property of a derivative giving rise to the detected edges being accompanied by spots. Edge detector outputs are not thresholded in this work since the actual edge strength is required by the watershed transform which is described in chapter 5. This section provides descriptions of different edge detectors and investigates their application to rock fragmentation images.

3.3.1 The simple gradient operator

Horizontally and vertically adjacent pixels are subtracted and divided by the pixel spacing to approximate the gradient. The two outputs are combined using the rms operator. Let the original image be $I(x,y)$ then the vertical component of the gradient is given by:

$$G_v(i,j) = I(i,j) - I(i,j-1) \tag{3.25}$$

and the horizontal component is:

$$G_h(i,j) = I(i,j) - I(i-1,j) \tag{3.26}$$

Then the rms operator produces

$$G = \sqrt{G_v^2 + G_h^2} \tag{3.27}$$

Figure 3.7 shows the output after applying the gradient operator on the image under analysis. The output appears to be noisy particularly on rock surfaces. Ideally, high responses are desired at the boundary and not on the surfaces. A suitable pre-filter should reduce the effect of noise and rock texture.

Figure 3.7: The simple gradient operator output

3.3.2 The Sobel operator

The Sobel operator computes the derivative of the image using the sobel approximation to the derivative as shown in the two masks below. Partial derivatives in the horizontal and vertical directions are approximated by weighted pixel differences in the respective directions. Neighborhood pixels which are vertically and horizontally aligned with the center have higher weights than the rest. The differences are divided by the number of neighborhood pixels excluding the center pixel. The rms operator is used to combine the two outputs.

Table 3.2: The horizontal direction detection mask

-1/8	-2/8	-1/8
0	0	0
1/8	2/8	1/8

Table 3.3: The vertical direction detection mask

-1/8	0	1/8
-2/8	0	2/8
-1/8	0	1/8

Figure 3.8: The Sobel operator output obtained using a neighborhood size of 3x3

Figure 3.8 shows the result of applying this operator on the image under analysis. This output is slightly better than the simple gradient output in that it is less responsive to rock-texture but also suffers from noise sensitivity. It also requires pre-filtering with a low pass filter.

3.3.3 The Prewitt operator

The Prewitt operator computes the derivative of the image using the Prewitt approximation to the derivative as shown in the 2 masks below. This operator is similar to the Sobel except that the nonzero elements of the mask add up to unity. The masks of Prewitt are apparently optimal while those of Sobel are not [13]. This is because each of the non-zero elements in mask has equal contributions to the final response while the Sobel operator response is dominated by the elements which are aligned with the center element.

Table 3.4: The horizontal direction detection mask

-1/6	-1/6	-1/6
0	0	0
1/6	1/6	1/6

Figure 3.9 shows the output after applying Prewitt's operator on the image under analysis. It is similar to the gradient and Sobel output in that it requires pre-filtering to reduce noise sensitivity.

Table 3.5: The vertical direction detection mask

-1/6	0	1/6
-1/6	0	1/6
-1/6	0	1/6

Figure 3.9: The Prewitt operator output obtained using a neighborhood size of 3x3

3.3.4 The Laplacian of Gaussian

This operator locates edges by convolving the image with a Laplacian of Gaussian mask. A continuous 2D Gaussian function $K(x,y)$ is given by:

$$K(x,y) = e^{-(\frac{x^2+y^2}{2\sigma^2})}. \tag{3.28}$$

The Laplacian operator ∇^2 on the gaussian function can be written as:

$$\nabla^2 K(x,y) = \frac{\partial^2 K(x,y)}{\partial x^2} + \frac{\partial^2 K(x,y)}{\partial y^2}. \tag{3.29}$$

After substitution and taking derivatives the final Laplacian of gaussian operator works out to be:

$$\nabla^2 K(x,y) = (\frac{1}{\sigma^2})^2 K(x,y)(x^2+y^2-2\sigma^2). \tag{3.30}$$

This operator effectively detects local edges around the pixel in questions due the dip in weights around the center. The space parameter σ controls the width of the dip and thus the width of the detected edge traces. The larger it is, the wider the edge tracks become. Figure 3.10 shows 25x25 Laplacian of Gaussian mask with a σ value of 2.5.

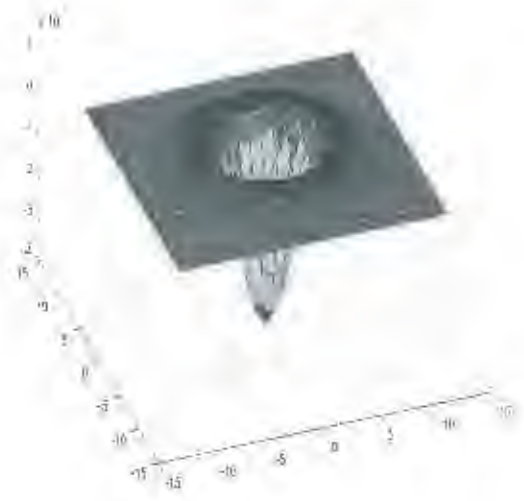


Figure 3.10: An example of 25x25 Laplacian of gaussian operator with the sigma value of 2.5

Figure 3.11 shows the result of applying the Laplacian of Gaussian on the image under analysis. The edge strength at the boundary is very high all around the rock boundary. The response is high even for the weak edges and thus it is highly sensitive rock-surface cracks and texture.

Figure 3.11: The Laplacian of Gaussian operator output with a neighborhood size of 41x41 and a sigma value of 2.5

3.3.5 The Canny operator

This operator filters each column of the image with a 1-D Gaussian and then filters the outputs with a 1-D derivative of the Gaussian to obtain gradients in the vertical direction. This operation is repeated on the transposed image to obtain gradients in the horizontal direction. The two gradients are finally combined using the rms operator. A simpler implementation convolves the 1-D gaussian mask and its derivatives first and uses the resultant 1D mask to convolve the image along each column in each of the directions. In formal terms:

$$K(x) = e^{-(\frac{x^2}{2\sigma^2})} \tag{3.31}$$

$$K'(x) = (\frac{-x}{\sigma^2})e^{-(\frac{x^2}{2\sigma^2})} \tag{3.32}$$

The 1D operator is computed as the convolution:

$$C(x) = K(x) * K'(x) \tag{3.33}$$

The image $I(x,y)$ is convolved with the column vector $C(x)$ along each column to obtain edges in the vertical direction. An example of a 1-D Canny operator $C(x)$ with a σ value of 2.5 is shown in figure 3.12. The image $I(x,y)$ is then transposed and convolved to obtain horizontal edges. The rms operator is used to combine the two outputs.

The parameter σ controls the width of the traces and the rejection of noisy responses. A smaller value reduces the width of the traces and results in a noisy response due to rock-surface cracks

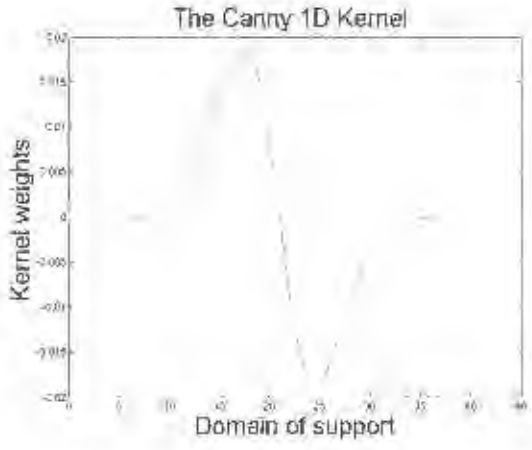


Figure 3.12: The 1-D canny operator with a sigma value of 2.5

and texture. A larger value increases the width of the traces and results in a less noisy image. Figure 3.12 shows the output of the Canny operator on the image under analysis. It has lower edge strength response on rock surfaces and high responses on the boundary. It appears to be less sensitive to noise, rock-surface cracks and texture than the other gradient operators which are under investigation.

Figure 3.13: The Canny operator output with a sigma value of 2.5

3.4 Summary

This chapter introduced traditional image processing procedures that can be used in the development of the segmentation algorithm. It is apparent that thresholding techniques can be used for locating possible rock locations in the image, while rock-edge detection can be used as a guide for tracing out edges of rocks in an image.

The investigation shows that the adaptive thresholding method is more suitable than the histogram-dependant global thresholding approaches on rock-scene images. This is because of its ability to adapt the threshold according to gray-value changes within the image. However the window size of the filter affects the range of sizes of the particles that are detected. As a solution to this problem a two-window adaptive thresholding method in introduced. It is found to be the best method for thresholding images of rock-scenes.

Finally, rock-edge detection methods are also investigated. It is apparent from the investigation that, with the exception of the Canny operator, most of these methods require a pre-filtering stage. The canny operator output appears to be the best of all the operators' outputs since it has weak responses from noise and texture variations but strong responses at the actual rock edges. The Laplacian of Gaussian appears to have the worst response on the rock-scene image under analysis. However, with the introduction of pre-filtering methods the rankings of the operators may change.

Chapter 4

Filtering methods

4.1 Introduction

In computer vision applications, the noise that is imposed on digital images comes from digital imaging processes such as sampling, quantization and random disturbances in the capture hardware. Noise is undesirable and should be reduced before further processing can take place. A common property of noise is the wide spectrum of frequencies that it occupies. As a consequence, the general objective of smoothing techniques is to remove the high frequency components without severely distorting the signal of interest in the base-band. This can be achieved using low pass filters such as the neighborhood averaging filter.

The type of noise that corrupted the signal of interest is usually unknown. The general assumption in the signal processing community is that of a zero-mean white gaussian noise with small variance, while the signal of interest is a smooth function. The objective is to recover the underlying signal of interest. However, a filtering operation may be required for more than just noise suppression but also to enhance objects' appearances by making gray level discontinuities more prominent[1]. In this case linear approaches fail as they blur edges and surfaces equally. Nonlinear filtering approaches, which are tailored for edge preservation, are better alternatives and they have recently found great attention within the research communities of signal processing.

This chapter is concerned with noise suppression and image enhancement techniques, some of which will be used in the subsequent chapters of this dissertation. A background theory of linear and nonlinear approaches is given. A qualitative comparative evaluation of these methods on the rock fragmentation image under analysis is carried-out simultaneously.

4.2 Linear filtering

Linear filters smooth the image by convolving it with a mask of weights, such that each pixel's gray level in the output image is a linear combination of gray levels in a predefined neighborhood. The simplest of these is the box filter or the uniform filter where the normalized mask consists of equal weights. However, pixels which are spatially close together within a predefined neighborhood tend to have approximately the same gray value, and the box filter does not exploit this. On the other hand the Gaussian filter exploits this by assigning a higher weighting to the pixels within the mask which are closer to the mask-center. In this section, theory of linear filters such as the box and the gaussian filters is discussed and their applicability to the problem of smoothing rock scenes is investigated.

4.2.1 The uniform or box filter

The image is convolved with a mask of equal weights whose total sum is unity for preserving the gray value of a flat region. The underlying assumption here is that the image is a smooth low-frequency signal corrupted by white gaussian noise with a small variance. The assigning of the mean gray value of the neighboring pixels to each pixel as the estimate is optimal inside the almost flat regions since the least squares error is minimal. However images have discontinuities which characterize object shapes where the estimate is affected by the huge deviation in gray values. The gray values which differ significantly from the rest in the neighborhood should be treated as outliers and not be included in the estimation process, so that the estimate lies on one of the flat surfaces.

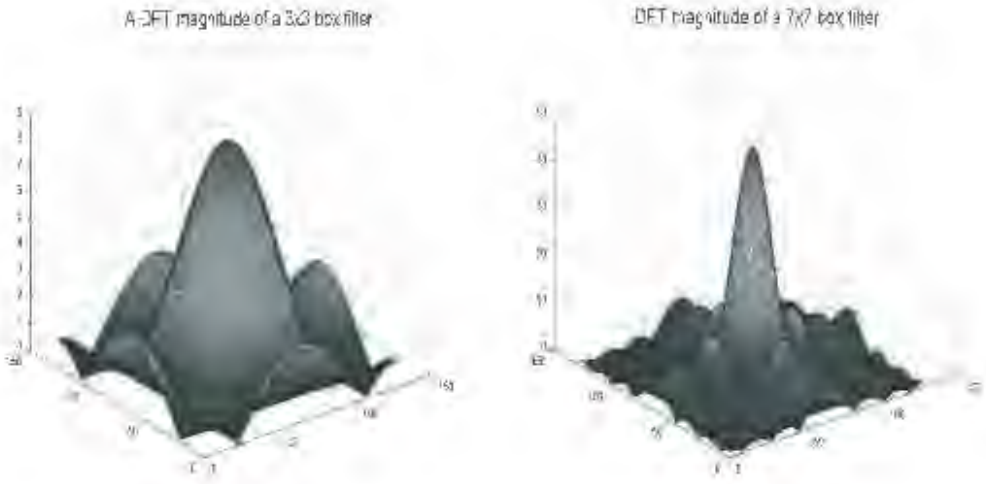


Figure 4.1: The Fourier transform of 3x3 and 7x7 box filters

The box filter has a discrete Fourier transform (DFT) in the form of a sinc function as shown in figure 4.1. This is a poor low pass filter because the attenuation does not increase monotonously with increasing spatial frequency instead it oscillates [31]. The significance of this is that high frequency noise is not removed entirely by the box filter due to the huge side-lobes as shown in figure 4.1. As an example, figure 4.2 shows the effect of increasing the box filter size from 3x3 to 7x7 in the spatial domain. As can be seen, local detail in the form of small sized rocks and texture across surfaces is being lost. This is due to the main-lobe being narrower as the window size is increased. Texture reduction is desired but the destruction of small sized rocks is not a desired effect. In addition to the destruction of small structures, rock edges are blurred as well.

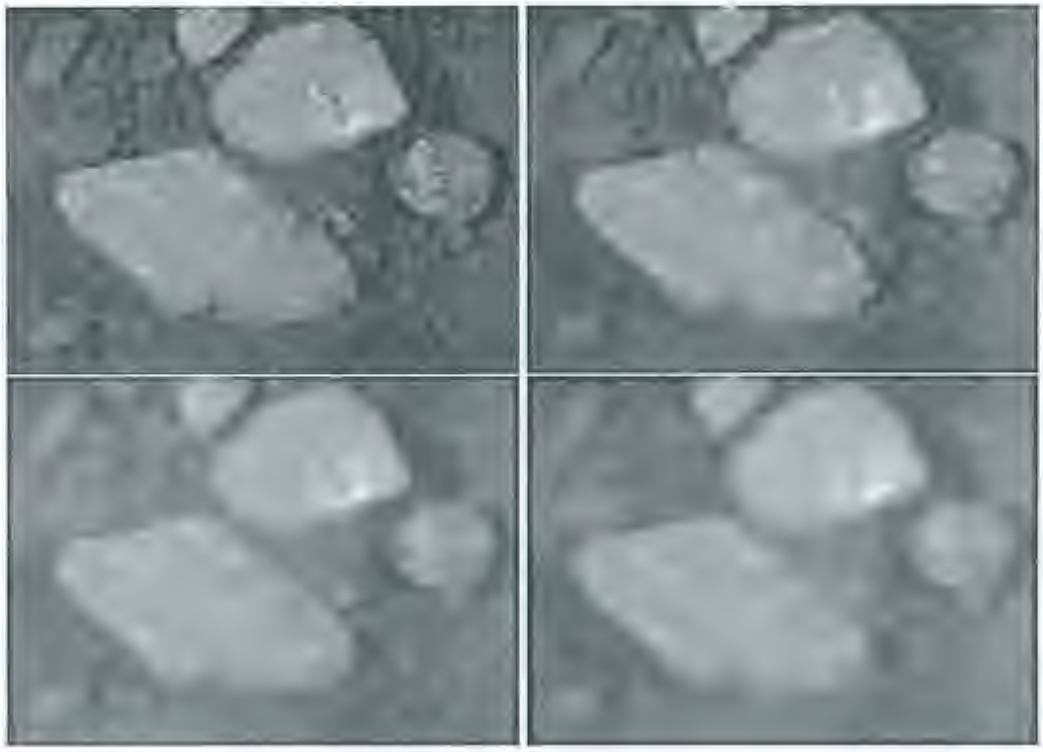


Figure 4.2: The effect of varying the mask size of a uniform filter from 3x3, 5x5 to 7x7

4.2.2 The Gaussian Filter

The Gaussian filter is a special kind of a moving averaging filter where the mask weights decrease radially from the center to the margins of the mask. It is expected to smooth better than the uniform filter as the DFT of a gaussian with space parameter σ is another Gaussian with the space parameter being the reciprocal of σ. Thus the frequency decreases monotonously with increasing spatial frequency. It has a number of desirable mathematical properties which are used to increase computational speed. One such property is its separability as a result of:

$$G(x,y) = e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} = e^{-\left(\frac{x^2}{2\sigma^2}\right)} e^{-\left(\frac{y^2}{2\sigma^2}\right)}. \tag{4.1}$$

The space parameter σ controls the degree of smoothing. A larger σ value reduces the bandwidth of the filter, thus destroying the high frequency content such as noise, rock texture, fines and small rocks. On the other hand a lower value increases the bandwidth and thus reduces the degree of filtering. The space parameter is exploited in many applications for the detection of a wide range of object sizes in an image by filtering the image with various σ values. This result

to a scale-space paradigm introduced by Witkin[66] defined as

$$I(x,y,t) = I_0(x,y) * G(x,y,t) \tag{4.2}$$

where $I(x,y,t)$ is a population of Gaussian filtered images, $I_0(x,y)$ is the original image and t is the scale space parameter which varies from 0 (the unfiltered original) to the maximum value (the coarsest scale). It is shown by Koenderink[35] and Hummel[28] that this population of images can be viewed as the solution to the heat diffusion equation:

$$I_{t+1} = \frac{\partial^2}{\partial x^2}(I_t) + \frac{\partial^2}{\partial y^2}(I_t) \tag{4.3}$$

where I_0 sets the initial condition. As an example, figure 4.3 shows a scale-space representation of rock scenes. The increase in the scale parameter transforms the images from a fine to coarser scales. The drawback of this approach, which has been pointed out in [52] is the loss of "semantically meaningful" object descriptors, namely edges. In their work Perona and Malik go further by proposing a new scale space paradigm using anisotropic diffusion as opposed isotropic diffusion offered by the linear Gaussian filtering. Anisotropic diffusion is described in detail in section 4.3.2.

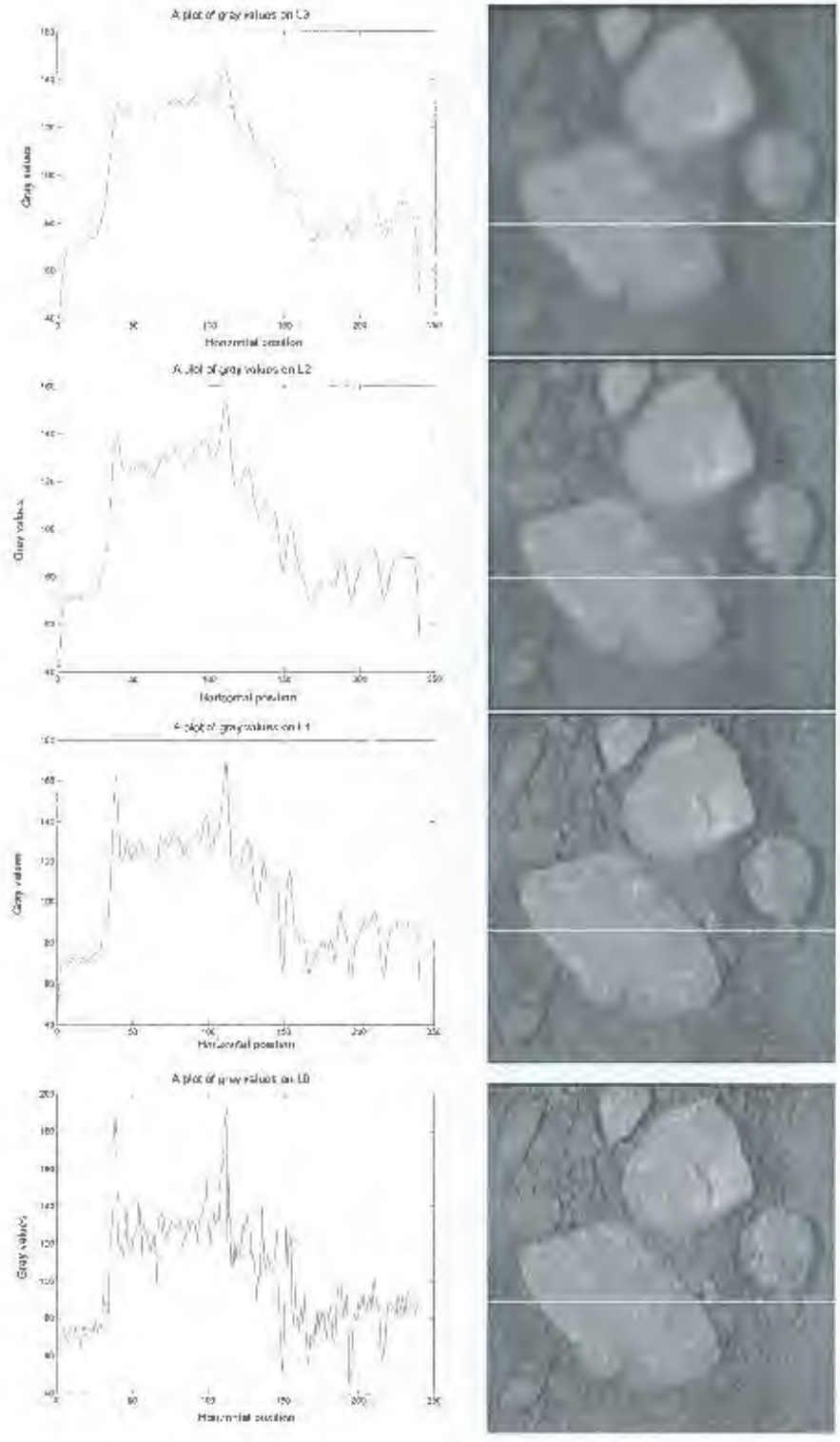


Figure 4.3: A scale space set of derived images with the space parameter t on a 25x25 neighborhood increasing from 0 to 3 where t is zero corresponds to the original image. L3 denotes the low-pass filtered image at t is 3.

4.3 Non-linear filtering

As pointed out in the previous section, linear filtering methods do not preserve the edges of objects in an image, and thus cannot be used as a pre-filtering scheme for rock edge detection. In this section non-linear filtering methods which preserve edges during filtering are investigated. The set of filters under investigation consists of rank value, anisotropic diffusion and bilateral filters.

4.3.1 Rank value filters

The filters that have been considered so far use the convolution operation to smooth the image, hence the linearity property. The convolution operation multiplies the signal values which fall within a window and sums the products. This operation blurs edges due its sensitivity to outliers in gray values. On the other hand, rank value filtering is a non-linear operation where the gray values of pixels within the mask are compiled into a list, sorted in ascending order and the gray value in a particular position of the list selected as the gray value of the center pixel. A median filter is a common type of a rank value filter where the medium (50^{th} percentile) value in the list is selected. Figure 4.4 illustrates the operation of a 3x3 median filter.

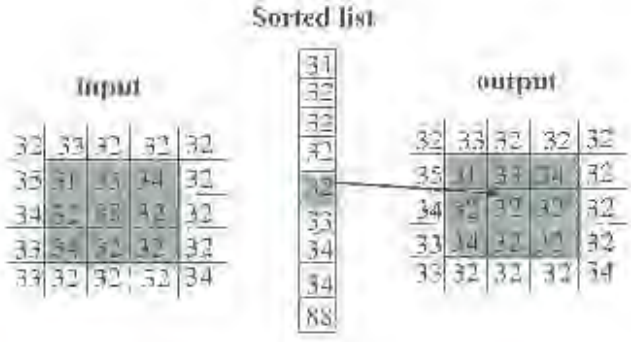


Figure 4.4: Illustration of rank value filtering in a 3x3 neighborhood with the median value being selected from the sorted list

The median filter is robust to outliers as it eliminates gray-value or color outliers from the image. It preserves edges to some degree due the selection of the median value which ignores outliers in the mask. It is expected to remove impulse noise without severely affecting the neighboring pixel gray values. It can be used for pre-filtering to reduce the texture variation across rock surfaces without significantly blurring the edges. Figure 4.5 shows the filtering performance of 3x3, 5x5 and 9x9 median filters on the analysis image. It is apparent that spots and gray-value spikes are progressively removed as the filter support is increased. In addition, the edges of smaller structures also tend to diminish with increasing filter-support size.

Based on these results it is apparent that the median filter is not a powerful edge preserving filter for this application and thus cannot be used for this purpose. Instead it can complement a powerful edge preserving filter by pre-filtering to remove impulse noise and gray-value spikes prior to applying an edge preserving filter.

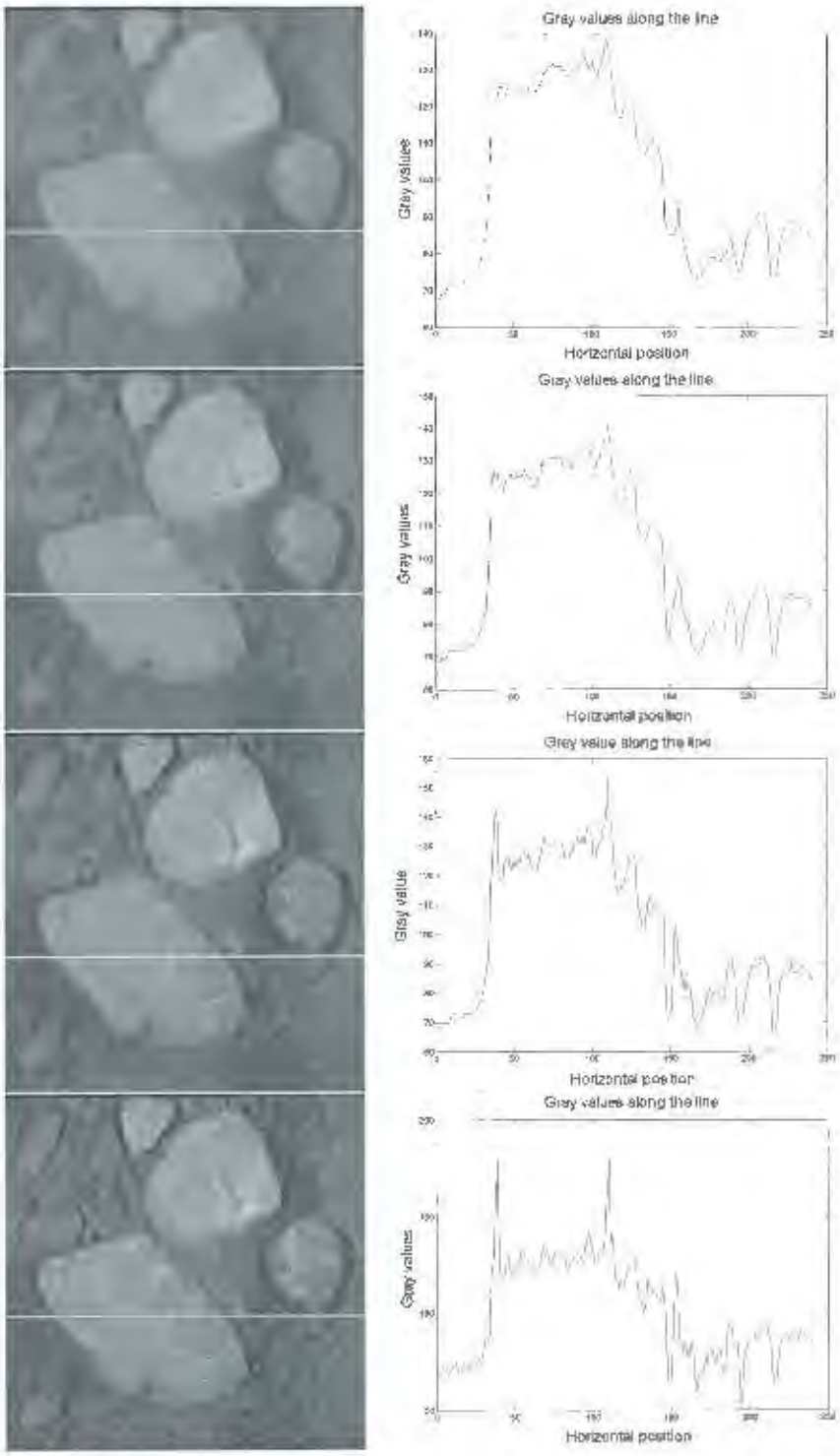


Figure 4.5: The effect of varying the neighborhood size of a median filter. The bottom output results from filtering with a 3x3 mask, the second from the bottom with a 5x5, the third from the bottom with a 7x7 and the top with 9x9 mask.

4.3.2 Anisotropic Diffusion

Anisotropic diffusion is presented by Perona and Malik in [52] where they improve the standard scale space paradigm to preserve edges. The linear scale space paradigm introduced by Witkin[66] can be viewed as isotropic diffusion which is defined by the partial differential equation:

$$I_t = \mathrm{div}(c(x,y,t)\nabla I) = c\triangle I \tag{4.4}$$

where

$$c(x,y,t) = 1. \tag{4.5}$$

The *div* operator denotes the divergence operator, ∇I is the gradient of the image I, $\triangle I$ is the Laplacian operator on I and $c(x,y,t)$ is the diffusion coefficient. As pointed out by Koenderik[66], this form does not permit space variant smoothing due to the constant diffusion coefficient c. As a consequence surfaces and edges are blurred equally. This is the fundamental flaw of this paradigm.

In the anisotropic diffusion framework, Perona and Malik allow spatial varying smoothing by selecting a suitable nonconstant function for the diffusion coefficient resulting in the partial differential equation

$$I_t = \mathrm{div}(c(x,y,t)\nabla I) = c(x,y,t)\triangle I \quad \nabla c \cdot \nabla I. \tag{4.6}$$

Ideally the diffusion coefficient should be unity on flat surfaces for isotropic diffusion and zero at the boundaries for edge preservation. A gradient image ∇I is a good indicator of region boundaries and can be used to adapt the diffusion coefficient $c(x,y,t)$ using:

$$c(x,y,t) = g(||\nabla I(x,y,t)||), \tag{4.7}$$

where $||\nabla I(x,y,t)||$ is the gradient magnitude and $g(||\nabla I(x,y,t)||)$ is the so called "edge stopping" function. The desirable properties of this function are that it should approach unity for small gradient magnitudes and zero for large gradient magnitudes. The success of the filtering depends highly on the choice of the edge stopping function used. In the next section, the discrete formulation is presented.

Perona and Malik's discrete formulation

Perona and Malik approximated the continuous anisotropic diffusion equation with the following discrete version [47]:

$$I_s^{t+1} = I_s^t + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} g(\nabla I_{s,p}) \nabla I_{s,p} \tag{4.8}$$

where I_s^t is the discrete image after the t iterations in a 2D grid, s denotes the current pixel position, λ determines the rate of diffusion, η_s represents the spatial neighborhood of pixel s, and $|\eta_s|$ is the number of nearest neighbors of pixel s as shown figure 4.6. The image gradient in a particular direction is approximated by,

$$\nabla I_{s,p} = I_p - I_s^t \tag{4.9}$$

where p is a member of the 4 nearest neighbors of s.

Figure 4.6: The neighborhood relations between the pixels of the discrete anisotropic diffusion equation shown at the boundary

The edge stopping function

The issue of selecting the appropriate edge stopping function requires further consideration. Black and others [47] investigate various edge stopping functions, where they establish a relationship between anisotropic diffusion and robust estimation. The set of edge stopping functions include the Lorentzian, Huber's minimax and Tukey's biweight which are derived from relevant error norms. In their work, anisotropic diffusion is viewed as a robust estimation procedure for estimating a piecewise smooth image from a noisy image. The assumption is made that the

noise is Gaussian distributed with zero mean and a small variance, and it is pointed out that this assumption is not valid at object's boundaries as some of the pixel's neighborhood gray values are outliers. Some of the highlights of the work in [47] are outlined in the remainder of this section.

In robust statistics, the problem of estimating a piecewise smooth image from a noisy one can be posed as an optimization problem, where an image is sought which satisfies the criterion:

$$\min_{I} \sum_{s \in I} \sum_{p \in \eta_s} \rho(I_p - I_s, \sigma). \tag{4.10}$$

Here $\rho(\cdot)$ is a robust error norm and σ is the scale parameter which controls its width. The design of the robust error norm should ensure that the effect of outliers at the boundaries is reduced. Applying gradient descent to the problem results in:

$$I_s^{t+1} = I_s^t + \frac{\lambda}{|\eta_s|} \sum_{p \in \eta_s} \psi(I_p - I_s^t, \sigma), \tag{4.11}$$
$$\text{where } \psi(I_p - I_s, \sigma) = \rho'(I_p - I_s, \sigma)$$

The function ψ is termed the influence function in robust statistics literature [47]. Comparing the robust estimate equation 4.11 to the discrete formulation by Perona and Malik in 4.8, it is clear that the relationship between the 2 formulations in terms of an arbitrary input x is:

$$g(x)x = \psi(x, \sigma). \tag{4.12}$$

This relationship can be exploited by viewing anisotropic diffusion as robust estimation and determining how large a magnitude gradient should be for it to be classified as an outlier ("edge") using σ. Figure 4.7 shows the relationship between the 2 formulations graphically using the least-squares, Lorenzian and Tukey's biweight functions.

A qualitative comparison of the functions reveals that the Lorentzian is more robust to outliers than the least squares since its edge stopping function $g(x)$ tends to zero as x tends to infinity. Ideally, the edge stopping function should be zero above a certain threshold. Tukey's bi-weight function achieves this by stopping diffusion completely above a threshold σ. In [47], the value of sigma is determined as the "robust scale", σ_r, of the image found using

$$\sigma_r = 1.4826\text{MAD}(\|\nabla I\|) = 1.4826\text{median}_I[\|\nabla I - \text{median}_I(\|\nabla I\|)\|]. \tag{4.13}$$

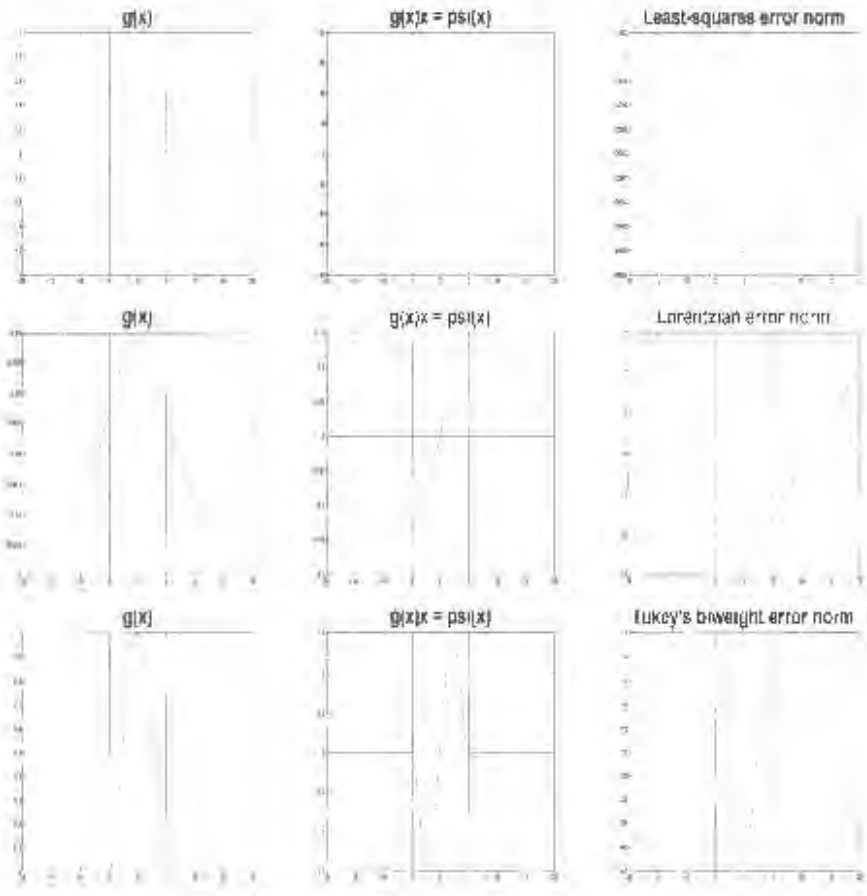


Figure 4.7: Various edge stopping functions, influence functions and associated error norms

The parameter σ_e is a robust estimate of the variation in gradient magnitude of image I. The MAD denotes the median absolute deviation and the constant 1.4826 is selected such that σ_e is unity for a zero mean normal distribution with a variance of unity. This occurs when the MAD is 0.6745 [7]. The results of applying the defined anisotropic diffusion filter with a σ value of 0.0196 on the image under analysis, for a varying number of iterations t are shown in figure 4.8. Observing the cross sectional profiles, it can be seen that where huge deviations occur in gray value, diffusion does not take place, while on regions with small changes in gray value diffusion increases with the number of iterations. Inspecting the actual filtered images, it can be seen that texture on rock surfaces, cracks and edges are preserved. This is due to the low σ value. This is the scale space paradigm proposed by Malik and Perona as a replacement of the linear standard paradigm.

An interesting scenario is where the image is anisotropic diffused with various σ values while keeping the number of iterations t and the diffusion rate constant. It is expected that the set of filtered images will range from images containing only strong edges with high σ values, to images preserving most edges at low σ values. However, the anisotropic diffusion is already computationally intensive due to the high number iterations required for effective smoothing. Therefore such a scheme would increase the computational burden on the system.

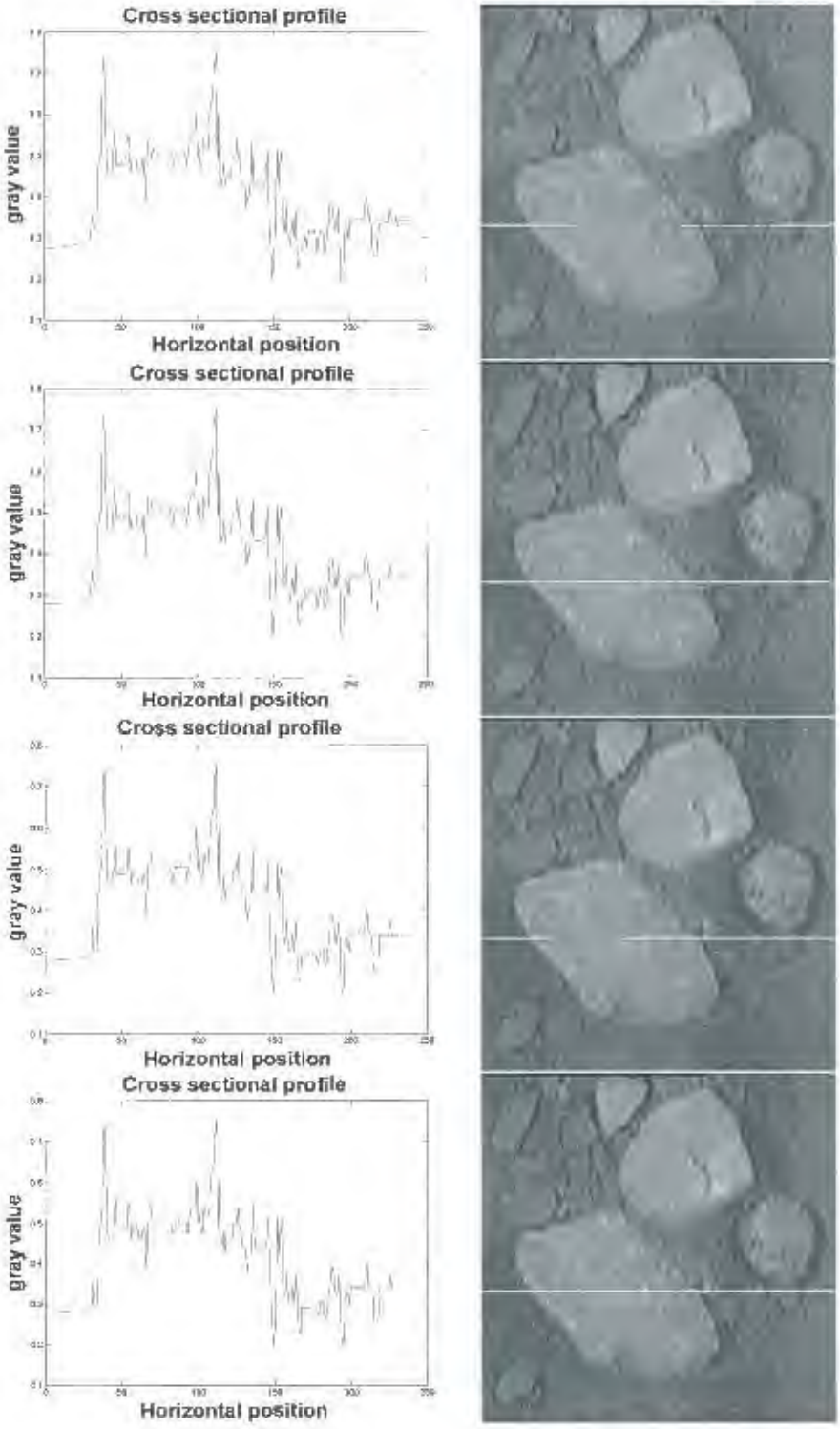


Figure 4.8: Results of setting the number of iterations to 50 in the top image, 100 in the second image from the top, 250 in the third image and 300 iterations in the bottom image while keeping the diffusion rate and sigma value constant at 1 and 0.0196 respectively.

4.3.3 The Bilateral Filter

The bilateral filter is a non-iterative edge preserving filtering technique introduced by Tomasi and Manduchi[59]. The filtering is spatially varying as in the anisotropic diffusion method. The usual Gaussian kernel has weights which decrease with increasing distance from the center of the kernel. On the other hand, the kernel weights of the bilateral filter are a combination of spatial distance from the center of the kernel and the similarity in gray-value or color between the pixel in question and the kernel center pixel. The bilateral filter as introduced in [59] has the form

$$\mathbf{h}(\mathbf{x}) = k^{-1}(\mathbf{x}) \int \int \mathbf{f}(\xi) c(\xi, \mathbf{x}) s(\mathbf{f}(\xi), \mathbf{f}(\mathbf{x})) d\xi \tag{4.14}$$

where the normalization term is

$$k(\mathbf{x}) = \int \int c(\xi, \mathbf{x}) s(\mathbf{f}(\xi), \mathbf{f}(\mathbf{x})) d\xi \tag{4.15}$$

The terms $c(\xi, \mathbf{x})$ and $s(\mathbf{f}(\xi, \mathbf{x}))$ are the closeness and tonal weight functions of the Euclidean distance between their arguments respectively. The gray value $\mathbf{f}(\xi)$ in the neighborhood of the central point x is compared to the gray value $\mathbf{f}(\mathbf{x})$ to obtain the spatial tonal weight $s(\mathbf{f}(\xi), \mathbf{f}(\mathbf{x}))$. In [60], the generalized version of the Bilateral filter is introduced. It is argued that the central value $\mathbf{f}(x)$ is noisy and should not be used as a good estimate for the underlying true signal value [60]. Instead a less noisy estimate should be used by introducing a second image g which provides a better estimate $\mathbf{g}(\mathbf{x})$. The new formulation of the Bilateral filter becomes

$$\mathbf{h}(\mathbf{x}) = k^{-1}(\mathbf{x}) \int \int \mathbf{f}(\xi) c(\xi, \mathbf{x}) s(\mathbf{f}(\xi), \mathbf{g}(\mathbf{x})) d\xi, \tag{4.16}$$

where the normalization term is

$$k(\mathbf{x}) = \int \int c(\xi, \mathbf{x}) s(\mathbf{f}(\xi), \mathbf{g}(\mathbf{x})) d\xi. \tag{4.17}$$

This is called Spatial-Tonal Normalized Convolution. The problem with this formulation is that the procedure for obtaining the less noisy image g is not specified. If non-edge preserving smoothing is used to obtain the image g, then $\mathbf{g}(x)$ will not be an accurate estimate at the edges. On the other hand, an edge preserving scheme will use the noisy $\mathbf{f}(x)$ value as the true estimate to obtain the image g. Based on this dilemma, the original formulation of Tomasi and Manduchi as expressed in equations 4.14 and 4.15, is therefore used in the rest of this work.

Gaussian Bilateral filtering

In this version, the closeness and tonal-weight functions are defined as Gaussian functions of their arguments, where

$$c(\xi,\mathbf{x}) = e^{-1/2(d(\xi,\mathbf{x})/\sigma_d)^2} \tag{4.18}$$

and

$$d(\xi,\mathbf{x}) = \|\xi - \mathbf{x}\| \tag{4.19}$$

is the Euclidean distance between ξ and $\mathbf{x}$. The tonal-weight is given by

$$s(\mathbf{f}(\xi),\mathbf{f}(\mathbf{x})) = e^{-1/2(\delta(\mathbf{f}(\xi),\mathbf{f}(\mathbf{x}))/\sigma_r)^2} \tag{4.20}$$

where

$$\delta(\mathbf{f}(\xi),\mathbf{f}(\mathbf{x})) = \|\mathbf{f}(\xi) - \mathbf{f}(\mathbf{x})\| \tag{4.21}$$

is the absolute difference of the intensity values $\mathbf{f}(\xi)$ and $\mathbf{f}(\mathbf{x})$. This implementation requires prior values of the space parameter σ_d and the similarity or range scale parameter σ_r. Based on tests, the space parameter has little effect on the preservation of edges: it only imposes a closeness constraint so that pixels far away from the kernel center have little influence on the kernel weighted mean. On the other hand, the selection of the similarity parameter σ_r is critical, and can be estimated using the robust scale parameter proposed for the robust anisotropic diffusion framework given by equation 4.13. As shown in figure 4.9, a single iteration of the bilateral filter does not remove noise effectively because the central gray-value $\mathbf{f}(x)$ is noisy.

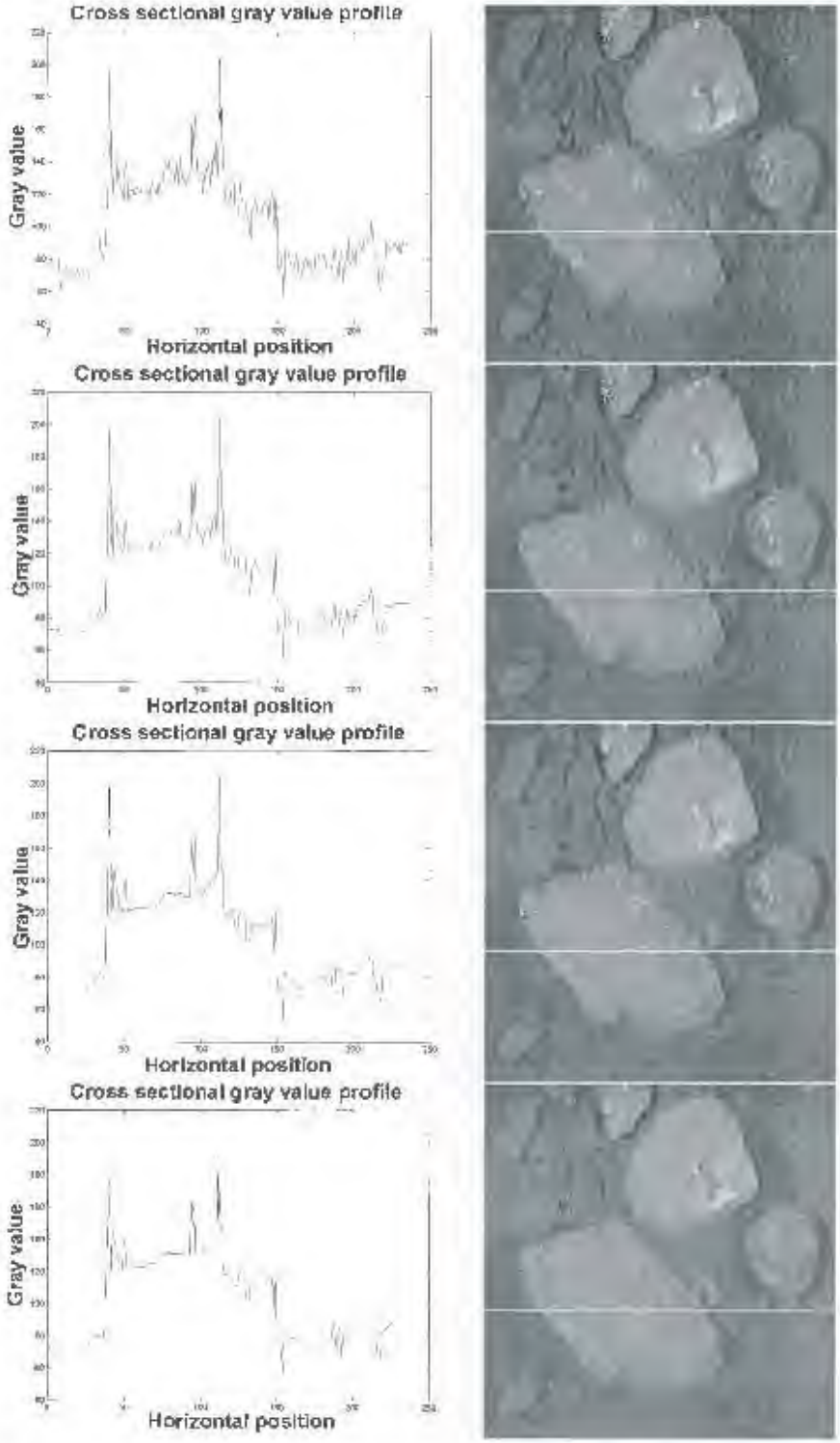


Figure 4.9: Results of setting the number of iterations to 1, 5, 10 and 15 from the top to the bottom image respectively, while keeping range and space sigma constant at 5 and 3 respectively.

Iterative Bilateral filtering is expected to remove noise and rock surface texture effectively as the iteration index progresses, due to better estimates of the kernel center pixel gray values. The required number of iterations is low compared to those required for anisotropic diffusion. After experimentation and tests it was determined that 15 iterations are suitable for rock scenes. This number can be reduced further by increasing the σ_r value with the increasing iterations as shown in figure 4.10. This is expected to smooth rock surface cracks in addition to surface texture and noise. Preceding this scheme with a 5x5 median filter should remove gray-value spikes due to cracks and white spots on rock surfaces.

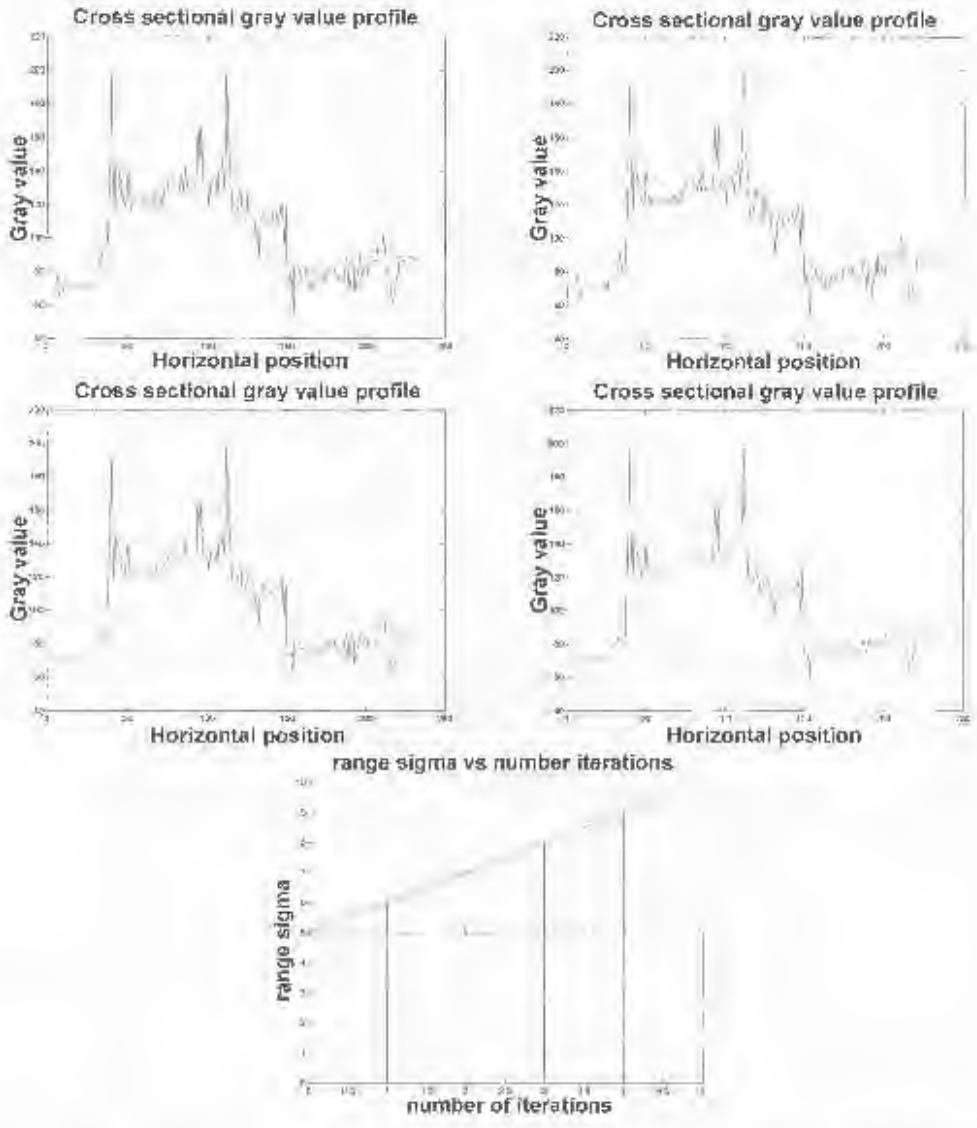


Figure 4.10: A comparison between two iterative bilateral filters where one has a constant range sigma value and the other has a varying sigma value with the number of iterations. The number of iterations is set to 1 and 5. It is expected that they should have the same output at iteration 1 as shown on the top two profiles. The two profiles on the right show the output of the bilateral filter when the range sigma value is varied with iterations as shown on the lower graph, while the sigma value is kept constant at 5 for the two profiles on the left.

4.4 Summary

In this chapter, various linear and non-linear filtering methods were investigated. In particular, linear methods in the form of the uniform and Gaussian filters were examined. These filters are not suitable for smoothing rock scenes since they blur edges and flat surfaces equally. However non-linear filters in the form of median filters, anisotropic diffusion and the bilateral filter are found to be suitable.

It was found that impulse noise in the form of rock-surface texture and gray-value spikes can be removed using the median filter, while rock-edges can be emphasized using either the anisotropic diffusion or the bilateral filter approach. However, anisotropic diffusion is found to be more computationally intensive than the bilateral filter. The bilateral filter is the adopted method for smoothing rock-scenes.

The standard bilateral filter has a flaw in that it uses a noisy estimate of the true signal against which the image gray-values are compared and as a result the filtering is not effective. An iterative version is expected to perform better since less noisy estimates of the true signal are achieved as the iteration index progresses. This is the version of the bilateral that is used in the rest of this work.

Chapter 5

Morphological Image Processing

5.1 Introduction

It is apparent that it would be a cumbersome task to use only traditional image processing techniques with pre-filtering methods in the development of a segmentation algorithm. In [38], edge detection with correctional procedures in the form of histogram-dependent image thresholding is adopted. However, it is reported that the desired accuracy is not achieved. Thus an integrative approach which involves combining tools from both traditional and morphological image processing with pre-filtering methods is a more viable route to follow.

Mathematical Morphology is an image processing paradigm whose tools modify the image content using geometrical structures of particular shapes, with the image being viewed as a landscape [11]. The structural elements vary from square to ball-shaped structures. Its initial ideas originated from the work of Serra and Matheron in the 1960's [58] and now have grown into a field that provides many useful image processing tools. Its primitive tools such as erosion and dilation, and the advanced tools in the form of closing and opening, modify the image content non-linearly by subjecting them to mathematical set operations [38] that either contract or expand the regions in the image. In addition, modern morphology in the form of the watershed transform is recently perceived as the method of choice for many image segmentation tasks [56]. In this work, the watershed transform forms the kernel or the core of the adopted segmentation approach.

This chapter is broken down as follows: firstly, formal definitions of the basic and advanced morphological techniques are given in sections 5.2 and 5.3 respectively. The reconstruction

operator is then introduced in section 5.4. In section 5.5, the watershed transform is introduced in terms of its background theory and definitions. Finally marker-based watershed is introduced in section 5.6.

5.2 Basic transformations

This section presents background theory of basic morphological transformations and their application to rock scene images.

5.2.1 Erosion and dilation

Erosion and dilation are the fundamental transformations of mathematical morphology from which the advanced transformations are derived.

Binary Dilation can be regarded as the expansion of a set by an amount limited by the properties of the structural element. These properties are the shape, size and the reference or the origin of the structural element. The formal definition for binary images in terms of set theory is

$$\delta_g(f) = f \oplus g = \{p \mid g_p \bigcap f \neq \emptyset\}, \tag{5.1}$$

where p is the current pixel position, g_p is the translated structuring element with its origin at p and $\delta_g(f)$ denotes the dilation operation on the binary image f. In a nutshell, the dilation result is the set of image pixels where the intersection between the translating structuring element and the foreground region is not empty [9].

Binary Erosion can be regarded as the contraction of a set by an amount limited by the properties of the structuring element. The formal definition of erosion is

$$\varepsilon_g(f) = f \ominus g = \{p \mid g_p \subset f\}. \tag{5.2}$$

The result of image erosion is the set of all the pixels of f where the translated structuring element g_p is completely contained within the foreground area [9]. Binary morphology can be regarded as a particular example of grey-scale morphology where the gray values are limited to one and zero.

The generalized gray-scale morphological operations are realized using the concept of an Umbra function $U(f)$ [11]. The Umbra of a d-dimensional function $f(\mathbf{x})$, where $\mathbf{x}$ is a d-dimensional vector, is given by

$$U(f) = \{(\mathbf{x}, b) : b \leq f(\mathbf{x})\}. \tag{5.3}$$

The Umbra is the set of all points on and below the surface function $f(\mathbf{x})$ as shown in figure 5.1. The original function $f(\mathbf{x})$ can be recovered using the top operation T defined as

$$f(\mathbf{x}) = T(U(f)) = max\{b : (\mathbf{x}, b) \in U(f)\}. \tag{5.4}$$

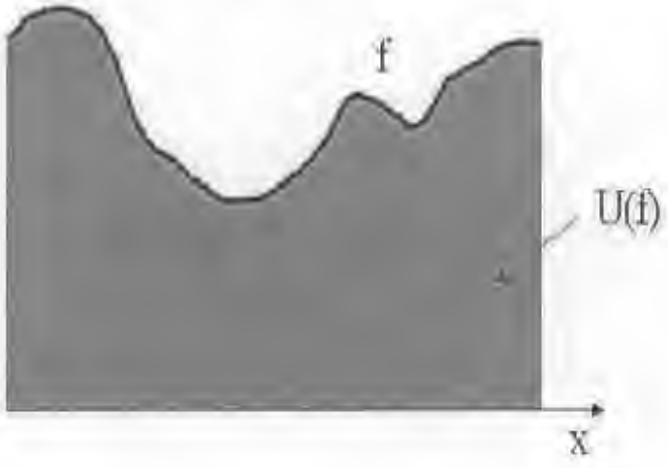


Figure 5.1: The umbra U(f) of a function f(x)

Gray scale erosion and dilation are defined in terms of the Umbra of an image f as

$$f \ominus g = T[U(f) \ominus U(g)] = \min\{f(\mathbf{x}+\mathbf{y}) - g(\mathbf{y})\} \tag{5.5}$$

and

$$f \oplus g = T[U(f) \oplus U(g)] = \max\{f(\mathbf{x}-\mathbf{y}) + g(\mathbf{y})\} \tag{5.6}$$

respectively, where g is the translating structural element, the positions $\mathbf{x} \pm \mathbf{y}$ fall within the domain of the function f and $\mathbf{y}$ is within the domain of g. Figure 5.2 shows the result of applying binary image erosion and dilation on a binary rock scene. It appears that erosion reduces the foreground and if holes are present, they are enlarged. Foreground areas smaller than the structuring element are removed. Dilation enlarges the foreground areas and closes holes whose size is smaller than the structuring element size.

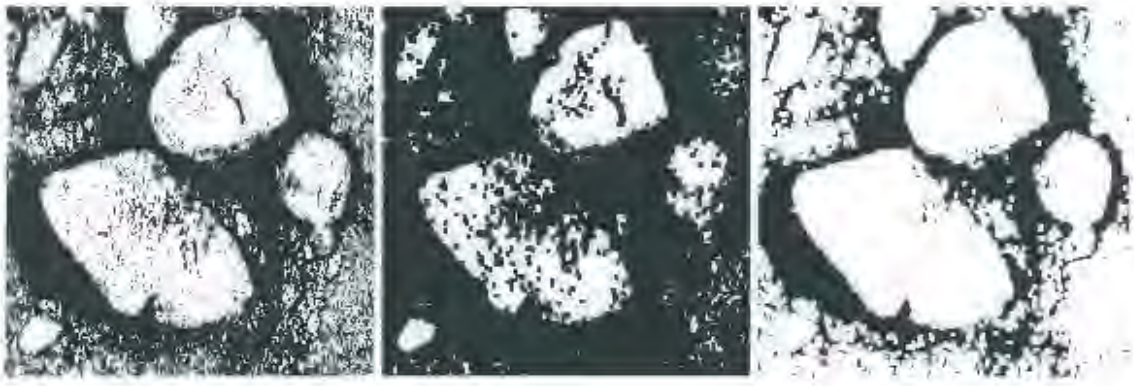


Figure 5.2: The original binary image, binary eroded and dilated images with a square structuring element of size 3x3. The original sample image is adaptively thresholded with a window size of 91x91

Figure 5.3: The sample image, greyscale eroded and dilated images with a disk structuring element of radius 5.

Figure 5.3 shows the results of applying gray-scale erosion and dilation on the sample image. The erosion output has dark disk shaped patches whose gray values are the minima values within the disk shaped neighborhood. The dilation output has bright disk shaped patches whose gray values are the maxima values within the disk shaped neighborhood.

5.3 Advanced transformations

This section presents the background theory of advanced morphological operations [6] and their application to the rock scene sample image.

5.3.1 Opening and closing

These operations are derived from the combinations of erosion and dilation. The opening operator is defined as

$$f \circ g = (f \ominus g) \oplus g \tag{5.7}$$

Opening can be thought of intuitively by considering the image as a landscape where the gray value at the point x is the height. Sliding the structuring element g underneath the surface and determining how high it can be pushed up defines the opening operation.

The closing operator is defined as

$$f \bullet g = (f \oplus g) \ominus g \tag{5.8}$$

Intuitively, the structuring element is slid on top of the image instead of underneath. The extent to which it can be pushed downwards is determined. The result is a morphologically closed image.

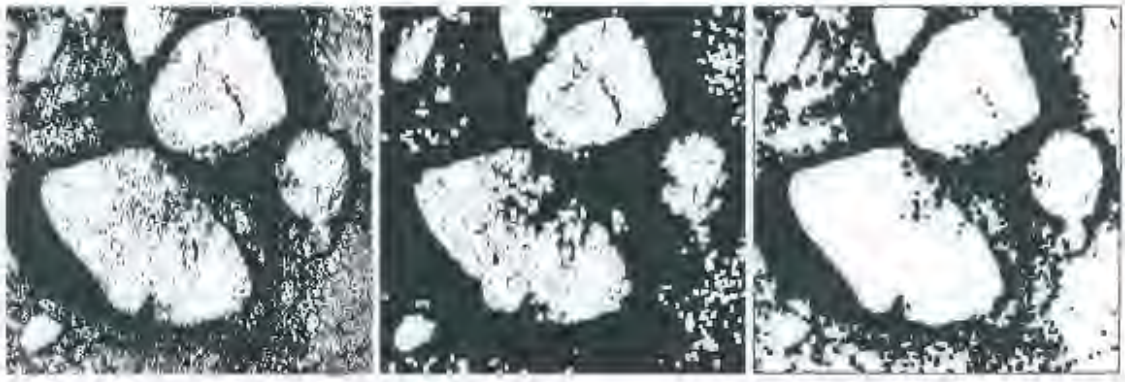


Figure 5.4: The original binary scene, binary opened and closed images with a square structuring element of size 3x3. The original sample image is adaptively thresholded with a window size of 91x91

Figure 5.4 shows the result of applying binary opening and closing operations on a binary rock scene image. It can be observed that the opening operation removes noisy spots without severely distorting larger structures. Closing fills holes without severely distorting the shape of larger structures.

Figure 5.5 shows the result of applying gray-scale opening and closing operations on the sample image. It can be observed that the opening operation's output is a non-linearly low pass filtered

Figure 5.5: The sample image, greyscale opened and closed images with disk structuring element of radius 5.

image. Bright and noisy structures whose widths are less than the diameter of structuring element are removed. On the other hand, the closing operation fills holes whose radii are smaller than the radius of the structuring element.

5.3.2 Morphological gradient

Edge detection is presented in chapter 3, where various differential edge detectors are investigated. The morphological gradient operator achieves edge detection using morphological operations. A simple subtraction of the eroded image from the original image gives the required quantity. This is written as

$$grad(f) = f - (f \ominus g) \tag{5.9}$$

Figure 5.6: Morphological gradient with a disk structuring element of unity radius

5.3.3 The Distance Function

The distance function assigns to each pixel in a binary image a value equivalent to its distance from the background. In formal terms, let Y be a set of Z^2 then for every element y of Y the distance $d(y)$ to the complementary set Y^c is defined as:

$$d(y) = dist(y, Y^c), \text{ for all } y \in Y \tag{5.10}$$

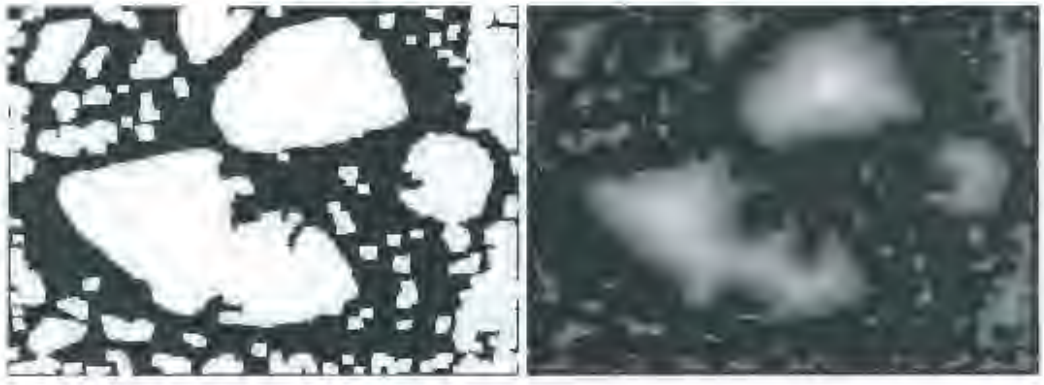


Figure 5.7: A binary rock scene and its distance transform version

Figure 5.7 shows a binary rock scene image as an input to the distance function and the distance function output.

5.4 Reconstruction

Reconstruction is a very useful tool of mathematical morphology. Some of the applications where it is used are in characterizing froth flotation performance in [20] and [67]. Reconstruction operators can be divided into binary and greyscale reconstruction operators.

5.4.1 Binary reconstruction

Binary reconstruction can be defined in terms of connected components and geodesic dilations[67].

Connected components

Let I and J be two images defined over the same domain D with $J \subseteq I$. This means that J is the marker image and I the mask image, according to the literature. Let the connected components of I be $I_1, I_2, I_3 \ldots\ldots I_n$. Then the binary reconstruction $R_I(J)$ of the mask I from marker J is the union of the connected components of I which contain a pixel from J. In formal terms

$$R_I(J) = \bigcup_{J \cap I_k \neq \emptyset} I_k \tag{5.11}$$

Geodesic dilations

Geodesic dilations (erosions) are defined in terms of geodesic distance. Geodesic distance is defined in [6] as follows: suppose that a set X is a subset of the Z^2 space, and x and y are two points within X. Then the geodesic distance $d_X(x, y)$ between x and y is the length of the shortest path included in X that links x and y, as shown in figure 5.8.

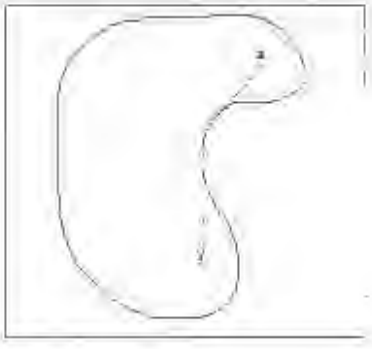


Figure 5.8: Geodesic distance between points x and y

The concept of Geodesic dilations is defined in terms of geodesic distance as follows in [67]: suppose that the set X is a subset of Z^2 and Y is a subset of X. Then the geodesic dilation of size n of the set Y (where n is a positive integer) is the set of all points within X whose geodesic distance to Y is less than or equal to n. In formal terms

$$\delta_X^{(n)}(Y) = \{p \in X | d_X(p, Y) \leq n\}. \tag{5.12}$$

An elementary geodesic dilation of the set Y is obtained using the following equation

$$\delta_X^{(1)}(Y) = (Y \oplus B) \bigcap X, \tag{5.13}$$

where B is a unity sized structuring element. A geodesic dilation of size n (where n is a positive integer) is obtained by iterating elementary geodesic dilations n times. Based on these definitions, the binary reconstruction $R_X(Y)$ of the set X from marker Y is given as

$$R_X(Y) = \bigcup_{n \geq 1} \delta_X^{(n)}(Y) \tag{5.14}$$

This performs iterations of geodesic dilations of marker Y until stability is reached. Figure 5.9 shows the binary reconstruction of the white areas using the black areas within the white blobs as the set of markers.

Figure 5.9: Binary reconstruction of the white areas with the black areas within the white blobs as the set of markers.

5.4.2 Greyscale reconstruction

Greyscale reconstruction is defined here in terms of geodesic dilations as obtained from [67]. An elementary geodesic dilation of greyscale image Y under X is defined as:

$$\delta_X^{(1)}(Y) = (Y \oplus B) \wedge X \tag{5.15}$$

where $Y \oplus B$ is the greyscale dilation of Y by a structuring element of size unity, and the $\wedge$ is the point-wise minimum operator analogous to the intersection operator.

A geodesic dilation of size n is obtained by iterating elementary geodesic dilations n times as in the binary case. The Greyscale reconstruction $R_X(Y)$ of X from Y is obtained by iterating geodesic dilations of marker Y until stability is reached. In formal terms

$$R_X(Y) = \bigvee_{n \geq 1} \delta_X^{(n)}(Y), \tag{5.16}$$

where '$\vee$' is the supremum operator which is analogous to the union operator of the binary case.

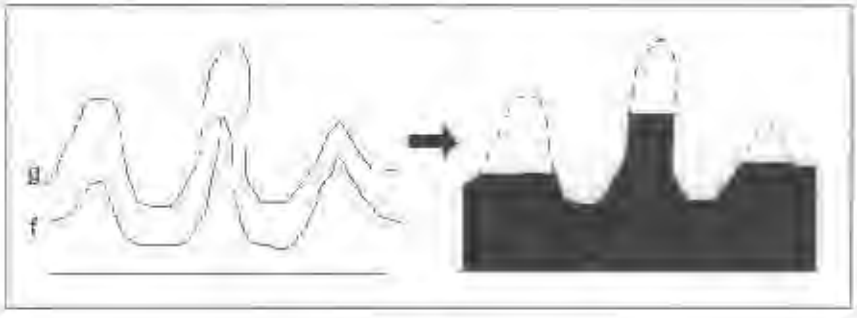


Figure 5.10: Greyscale reconstruction

Figure 5.10 shows the greyscale reconstruction of the mask function g using function f as the marker. It appears that greyscale reconstruction extracts or removes the peaks of the mask function. The dual reconstruction can be defined in terms of iterations of elementary geodesic erosions ε_X^1 defined as

$$\varepsilon_X^{(1)}(Y) = (Y \ominus B) \bigvee X. \tag{5.17}$$

Here $\bigvee$ is the point-wise maximum operator, $\ominus$ is the erosion operator and B is a unity sized structuring element. Based on this information, the dual reconstruction $R_X^*(Y)$ of mask X from marker Y is obtained by iterating geodesic erosions of marker Y until stability is reached. In formal terms

$$R_X^*(Y) = \bigwedge_{n \geq 1} \varepsilon_X^{(n)}(Y). \tag{5.18}$$

The relationship between the reconstruction $R_Y(Y)$ and its dual reconstruction $R_X^*(Y)$ is given as

$$R_X^*(Y) = (R_{X^c}(Y^c))^c \tag{5.19}$$

This shows that the dual reconstruction can be implemented using the reconstruction operator by exploiting the complementary feature of the reconstruction and its dual reconstruction.

5.5 The Watershed transformation

The watershed transform is the method of choice for many segmentation tasks in the field of mathematical morphology [6]. This is mainly due to the attractive features which it possesses in the form of continuous contour detection, and the efficiency in computation which lends itself well to real time implementation.

The segmentation is based on the natural phenomenon of flooding a topographical landscape. The image is seen as a topographical landscape where each point's gray value is equal to the height of the landscape at that point. Holes are pierced at the minima of the landscape and it is lowered into some fluid at a constant speed. The liquid is expected to penetrate the holes and fill up the catchment basins until fluids from various basins begin to merge. At the occurrence of this event, dams are built at the points of merging to separate flows from different basins. These dams are the watersheds and the barriers are the watershed lines.

This forms the immersion definition, and it is argued to be better suited to practical implementation than the definition by catchment basins [61]. The catchment basin definition is based on the fact that points will only be assigned to the same catchment basin if imaginary drops of water falling on them descend to the same minimum [5].

This section presents the formal definition of the watershed transformation.

5.5.1 Preliminary definitions

The concepts of geodesic zone of influence and skeleton by influence zones are introduced, as obtained from [6].

Geodesic zone of influence $iz_X(Y_i)$ of connected component Y_i in X is defined as the set of points in X at a finite geodesic distance, and closer to Y_i than any other Y_j. In formal terms:

$$iz_X(Y_i) = \{x \in X, d_X(x, Y_i) \text{ finite }, \forall j \neq i, d_X(x, Y_i) < d_X(x, Y_j)\}. \tag{5.20}$$

The boundaries between the various geodesic zones of influence are called the geodesic skeleton by zones of influence $SKIZ_X(Y)$ of Y in X. This is formally written as:

$$SKIZ_X(Y) = X / IZ_X(Y). \tag{5.21}$$

The symbol "/" denotes the set difference, and $IZ_X(Y)$ is the union set of all the geodesic zones of influence $iz_X(Y_i)$. Figure 5.11 shows the visual descriptions of the various concepts.

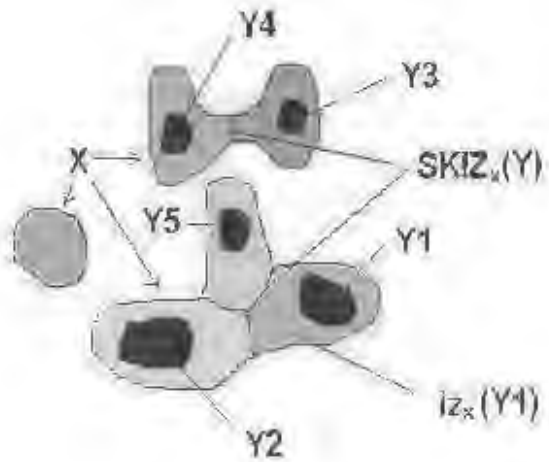


Figure 5.11: Geodesic zones of influence and skeleton by zones of influence

5.5.2 The definition of the watershed by immersion

The formal definition of immersion that is given here is obtained from [6] and [67]. The greyscale image f is assumed to be a 2D function of the image space grid. The absolute minimum of f on the entire domain is denoted as h_{min} and h_{max} is the absolute maximum. A section of f at level h is denoted by $Z_h(f)$. More formally:

$$Z_h(f) = \{x \in Z^2 : f(x) \leq h\}. \tag{5.22}$$

The immersion procedure begins at the minima $Z_{h_{min}}$ where the water penetrates the pierced holes. If the h_{min} level is incremented by one, the section $Z_{h_{min}+1}$ is obtained. Obviously the relationship

$$Z_{h_{min}} \subseteq Z_{h_{min}+1} \tag{5.23}$$

holds. Let Y be one of the connected components of the section $Z_{h_{min}+1}$. In [67] it is identified that there are three possible relationships between the connected components of the sections $Z_{h_{min}}$ and $Z_{h_{min}+1}$, namely:

- $Y \bigcap Z_{h_{min}} = \emptyset$: a new minimum is encountered at connected component Y where its surrounding pixels have gray values greater than $h_{min} + 1$. This minimum should be pierced so that water enters to fill up its associated catchment basin.
- $Y \bigcap Z_{h_{min}} \neq \emptyset$ and is connected: Y is still within the catchment basin associated with the minimum $Z_{h_{min}}$. The grey value of the pixels of Y are less than or equal to $h_{min} + 1$.
- $Y \bigcap Z_{h_{min}} \neq \emptyset$ and is not connected: other minima are encountered and the concept of geodesic zones of influence is used to create the watershed lines.

Combining all the possibilities for the second set, and denoting by $W_h(f)$ the catchment basins of f at section level h, the following relationship is obtained:

$$W_{h_{min}+1}(f) = min_{h_{min}+1}(f) \bigcup IZ_{Z_{h_{min}+1}}(W_{h_{min}}). \tag{5.24}$$

Here $min_{h_{min}+1}(f)$ denotes the set of points at altitude $h_{min} + 1$ belonging to the minimum of f. This relationship is generalized for any value of h as follows:

$$W_{h+1}(f) = \{min_{h+1}(f) \bigcup [IZ_{Z_{h+1}(f)}(W_h(f))], \forall h \in [h_{min}, h_{max}]\}. \tag{5.25}$$

The minima $min_{h+1}(f)$ of at level $h + 1$ are obtained from:

$$min_{h+1}(f) = Z_{h+1}(f)/R_{Z_{h+1}(f)}(Z_h(f)), \tag{5.26}$$

where the "/" denotes the set difference operator and R is the reconstruction operator defined in section 5.4. The algorithm is initiated at $W_{-1} = \emptyset$. The final watershed line set $WL(f)$ is

$$WL(f) = \{W_N^c(f), with\ N = max(f)\}. \tag{5.27}$$

Figure 5.12 shows the construction of the watershed by the process of immersion as defined in equation 5.24.

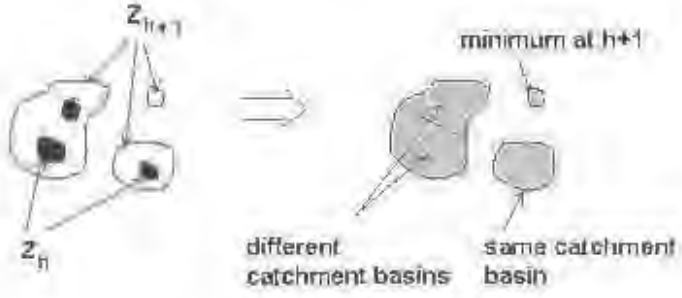


Figure 5.12: Watershed construction based on the three possible relationships

5.5.3 The over-segmentation problem

The watershed algorithm as presented over-segments the image undergoing segmentation. The reason for this is that in real world images catchment basins are not well defined due to image noise and texture. This result in the image having too many catchment basins which are identified by the immersion process.

Pre-smoothing

One approach to reducing over-segmentation is to apply a pre-smoothing filter before applying the watershed transform. Obviously the noise and texture level will be reduced so that there are fewer catchment basins than before. However, low-pass filters which are not sensitive to edges will destroy some of the boundary gradients, and as a result the shape of objects outlines will be distorted. Assuming that an edge sensitive pre-smoothing filter is used, such as the bilateral or the anisotropic diffusion filter, the over-segmentation will be reduced. However, this does not remove the problem completely.

5.6 Marker-based watershed segmentation

Another approach to solving the over-segmentation, which can be used in tandem with pre-smoothing is enforcing a pre-determined set of markers to be the local minima of the gradient[6]. Most of the time in practice the centers of the objects can be determined. En-

forcing these centers to be the global minima with respect to the objects will lead to a single catchment for each object in the image. These markers will be the new minima of the gradient function and the real minima will not be taken into account. As a result the behavior of the watershed transform will change. The following is the modified watershed definition as obtained from [6].

Let g be the gradient image, $W_h(g)$ be the section of g at level h, and M be the marker set. Then

$$W_{h+1}(g) = \{IZ_{(Z_{h+1} \cup M)}(W_h)], \forall h \in [h_{min}, h_{max}]\} \tag{5.28}$$

where the initial flooding sources W_{-1} are the predetermined set of markers:

$$W_{-1}(g) = M. \tag{5.29}$$

This new definition of the watershed transform appears simpler than the original version simply because the real minima of g are not determined.

5.6.1 Homotopy modification

The process of modifying the gradient function g as described above is termed homotopy modification. In other words, the homotopy of the function has been changed by piercing holes and thus creating new minima as shown in figure 5.13. This can be achieved by reconstructing the sections of g with the marker set M as follows [6]:

$$Z_h(g') = \{R_{Z_h(g) \cup M}(M), \forall h\}. \tag{5.30}$$

The function g' can be recovered from the sections using

$$g'(x) = min(h : x \in Z_h(g')) \tag{5.31}$$

Once the homotopy of the gradient function g is applied to obtain g' then the final step is performing the marker-based watershed segmentation of g'.

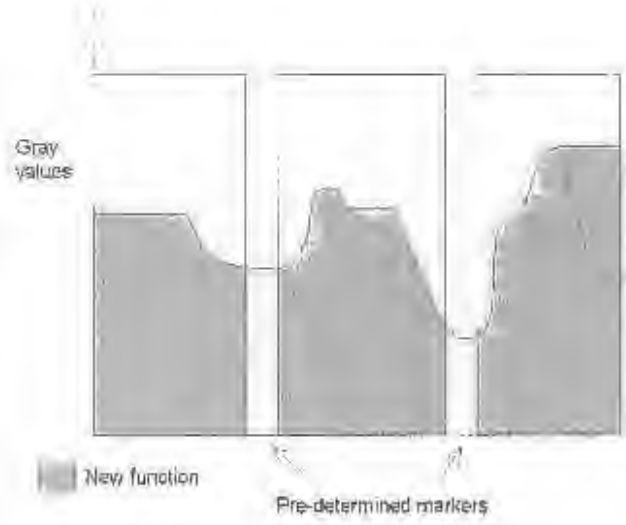


Figure 5.13: Watershed construction based on the three possible relationships

5.6.2 The watershed transform of a rock scene

The watershed transform is applied on the rock scene image under analysis and a few remarks are made. Firstly, the successful application of the watershed transform requires a gradient transform of the original image to enforce catchment basins. Secondly, a set of markers of local minima of the gradient image has to be pre-determined. These markers form the initial positions of the immersion process.

The immersion process requires well defined boundaries of the catchment basins in order to stop flooding at the correct points. Unfiltered images tend to have inaccurate watershed lines due to the noise and cracks across rock surfaces. Figure 5.13 shows the predefined set of markers overlayed on the original image, the noisy and inaccurate watershed lines, and the more accurate watershed lines due to pre-filtering with a low pass filter before applying the gradient operator.

(a) Markers

(b) Noisy and inaccurate watershed lines of the original image

(c) Accurate watershed lines of a smoothed image

Figure 5.14: Applying the watershed transform to a rock-scene image

5.7 Summary

In this chapter, the underlying theory behind the most useful tools of mathematical morphology is presented. In particular, the basic morphological tools in the form of erosion and dilation are handled. This is followed by the description of more advanced tools which are derived from the basic ones in the form of opening, closing and the morphological gradient. The distance function is another advanced tool but it is not based on the basic tools. All of these tools can form useful building blocks of the rock scene segmentation process.

The theory behind the most powerful tools in the form of the reconstruction operator and the watershed transform is finally presented. These tools form the core of the work presented in this dissertation as far as the segmentation of rock scenes is concerned. Reconstruction is used for determining the positions of the local minima of the gradient image. The watershed transform is a very powerful image segmentation tool.

In addition, a combination of the watershed transform and greyscale reconstruction result in a marker-based watershed transform, where a pre-selected set of minima is used to significantly reduce the over-segmentation problem. However, when this is applied on a rock-scene without a pre-filter the resultant watershed lines do not accurately trace the edges of the objects of interest. An edge preserving pre-filter should be used to increase the accuracy of these lines

Chapter 6

Rock scene segmentation

6.1 Introduction

Rock scene segmentation involves partitioning a scene of rocks into disjoint regions, such that each region is a closed outline of each rock in the image. These outlines should resemble as closely as possible the perceived edges of rocks. In this chapter, a rock scene segmentation procedure that is expected to be capable of segmenting any scene of rocks is constructed. The complete system is then tested on ore images from two different mineral processing plants, which are captured under varying lighting conditions. The test-set is expected to test the system for robustness under the specified conditions.

Generally, the implemented segmenter has two main components: the rock locator for determining markers of rocks, the actual segmentation process which is based on marker-based watershed segmentation and a scheme for locating fines. This chapter is broken down as follows: in section 6.2, an approach for pre-selecting a set of minima or markers is implemented. This is followed by the development of a watershed-based segmentation algorithm in section 6.3. The implemented segmentation algorithm is then tested for robustness to varying lighting conditions in section 6.4. Finally, the procedure for locating fine particles is described in section 6.5.

6.2 Automatic rock location for marker extraction

The immersion process of the watershed transform requires initial positions of the gradient local minima within the image. It is obvious that in the case of rock scenes there should be one marker for each rock to avoid splitting or merging. Splitting is erroneous when a two or more markers are assigned to a single rock, while merging is erroneous when a single marker is assigned to two or more rocks. Thus an effective procedure for identifying and marking of rocks, should determine markers without resulting in splitting or merging of regions.

In ideal conditions where a rock is smooth without any texture and cracks across the surface, the gradient of its grey values will have a maximum at the boundary and a minimum at the approximate rock center. These approximates centers are the regions that should be located. Two methods in the form of the centroid method and the reconstruction method are devised and investigated for determining these approximate centers. In what follows it is assumed that a binary rock image is obtained by thresholding a grayscale rock image using the two-window method.

6.2.1 The centroid method

The centroid method determines the centroids of the white areas in a binary rock scene. The centroid co-ordinates of a connected foreground region in a binary image are determined by finding its center of mass co-ordinates, as shown figure 6.1. For each connected region, the center of mass co-ordinates are computed as

$$(C_x, C_y) = (mean(x_i), mean(y_i)), \tag{6.1}$$

where i is the counter for each pixel co-ordinate position (x_i, y_i) in the foreground area. This method locates a center for each connected foreground region. However, it is not robust to situations where adjacent rocks are in contact with each other. This is because the bi-level thresholding process introduces errors in situations where there is no shadow separating two rocks in close proximity. This results in erroneous merging. Some of these events are shown in figure 6.1.

Figure 6.1: A Binary rock scene with marked foreground areas

6.2.2 The reconstruction method

The reconstruction method attempts to remedy the drawback of the centroid method by extracting the positions of the local maxima of the distance transform as rock locations. The distance transform (DT) is described in chapter 5 and there it is shown that the DT local maxima positions can be extracted using the greyscale reconstruction operator. The signal under reconstruction is the distance transform of the binary rock scene. This signal is termed the mask in reconstruction terminology, and the marker image is the mask signal negatively offset by a constant h:

$$marker = DT - h. \tag{6.2}$$

Here DT denotes the distance transformed image. The reconstruction is then performed as

$$reconstructed = R_{DT}(DT \quad h). \tag{6.3}$$

The reconstructed signal is the mask signal with local maxima chopped off as shown in figure 5.10. The local maxima are recovered by subtracting the reconstructed from the original DT image:

$$LMP = DT - R_{DT}(DT - h), \tag{6.4}$$

where LMP denotes Local Maxima Positions. Figure 6.2 shows the positions of the DT local maxima overlayed over the original rock scene image. As can be seen, the reconstruction method does acknowledge connected rocks by assigning multiple markers in such situations. However, due to the imperfections of the thesholding method giving rise to irregularly shaped

foreground areas, a single rock can be assigned many markers resulting to erroneous splitting. This occurrence can often be reduced by selecting a suitable value for the h-parameter of the grey-scale reconstruction method.

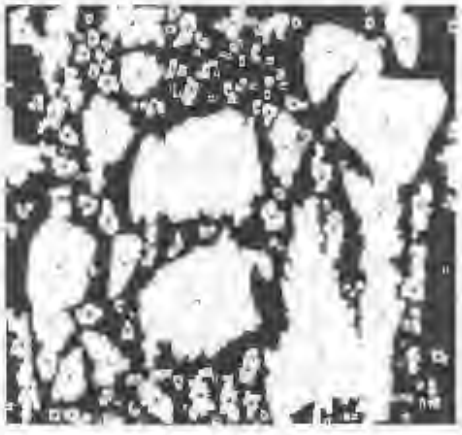


Figure 6.2: Binary rock scene with markers determined from the reconstruction method

6.2.3 The combinational approach

This approach aims at correcting for the errors caused by the reconstruction method by using the centroid method. For each foreground region in the binary rock scene, a decision needs to be made of whether to select a marker from either the centroid output or markers from the reconstruction output, as shown in figure 6.3. This selection is expected to correct for either erroneous splitting or merging. It requires a suitable criterion for making the decision. However, it is difficult to devise a criterion which does not involve the segmentation output.

A more realizable approach involves the case where the two methods are used with the segmentation process as shown in figure 6.4. Regions can be selected which resemble rocks in a sense that they are roughly circular in shape. For each white area of the binary rock scene a comparison between the corresponding segmented regions derived from both the centroid and the reconstruction method is made. The watershed region features that are used are the circularity of the region

$$C = \frac{4\pi A_{region}}{P^2_{region}} \tag{6.5}$$

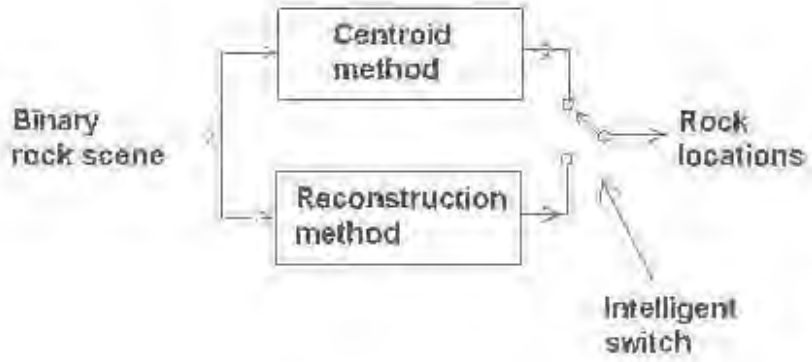


Figure 6.3: A block diagram representation of the combinational approach

and the area proportion between the watershed region in question and the corresponding foreground region in the binary scene

$$A_p = \frac{A_{region}}{A_{binary}} \tag{6.6}$$

Here A_{region} and P_{region} are the area and the perimeter of the region in question, and A_{binary} is the corresponding foreground area in the binary rock scene. The outputs of the method that have higher feature values than the other are selected as possible rock candidates. In a sense, a

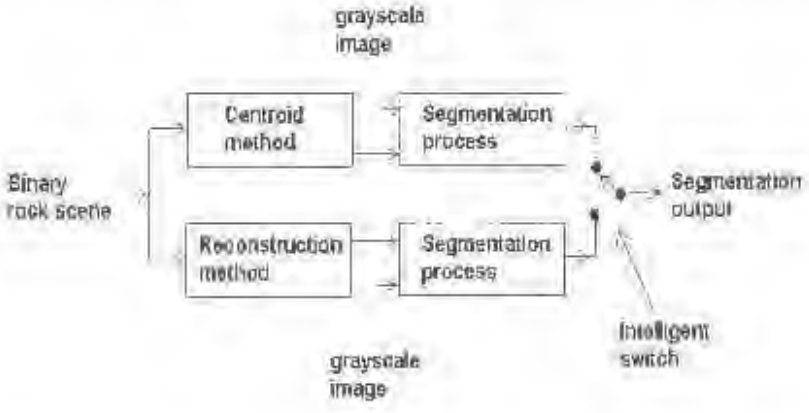


Figure 6.4: A block diagram representation of the realizable combinational approach

decision is being made of whether to split or merge a region given the features or characteristics of segmented regions derived from the two methods.

6.3 Watershed-based segmentation

The watershed transform requires a set of markers of the local gradient minima and the actual gradient image as inputs. Given these two items, a watershed segmentation of the gradient image is executed and the outlines of the rocks are determined. However, real images have noise and texture imposed on them and as a result the watershed lines will be inaccurate. Therefore a suitable pre-smoothing technique for rock scene segmentation is required. In chapter 4, the iterative bilateral filter with an increasing photometric similarity parameter σ_r is proposed. This technique is expected to be superior to the single iteration bilateral filter in that various edge preservation levels are explored and edges are enhanced. A filtered image is tapped at each output stage resulting in a population of images filtered at various degrees of edge preservation and edge enhancement is achieved via the iterative property.

6.3.1 The basic framework

The structure of the iterative bilateral filter is shown in figure 6.5. It is clear that this filter has one input and n outputs. Provided that the range σ_r values increase monotonically, the filtered images are expected to vary in edge preservation from preserving most of the variation (edges due to texture, some noise and cracks) at small σ_r values, to preserving only highly varying regions (strong rock edges) at higher σ_r values.

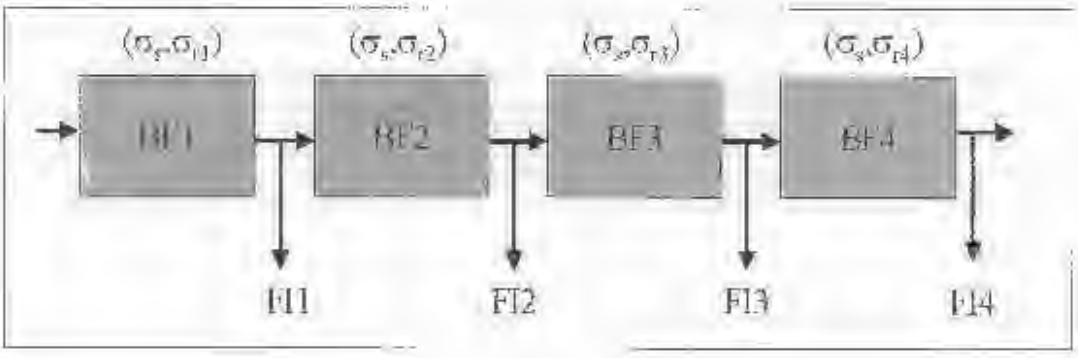


Figure 6.5: The structure of the iterative bilateral filter

In the context of the watershed transform, at small σ_r values the watershed is expected to fol-

low the gradient lines of small structures such as texture, small rocks, and cracks across rock surfaces. However, this is driven by the selected set of markers in that the number of detected small structures is determined by the number of markers. On the other hand, at higher σ_r values most of the small variation due to texture, small rocks and cracks is diminished. In this case, the watershed lines are expected to follow the outlines of larger structures such as big rocks. However, these outlines might not be accurate as most of the detail has diminished at this stage.

The basic structure which includes the watershed processes is shown in figure 6.6. A gradient operator (GO) is applied at the output of each bilateral filter (BF) stage. Any of the available gradient operators of chapter 3 can be used. In this work, the morphological gradient operator is used. Marker-based watershed segmentation processes (W) are then executed on each gradient image. The information from the n-watershed images is then combined using a multiple watersheds analysis process to form a single final output image.

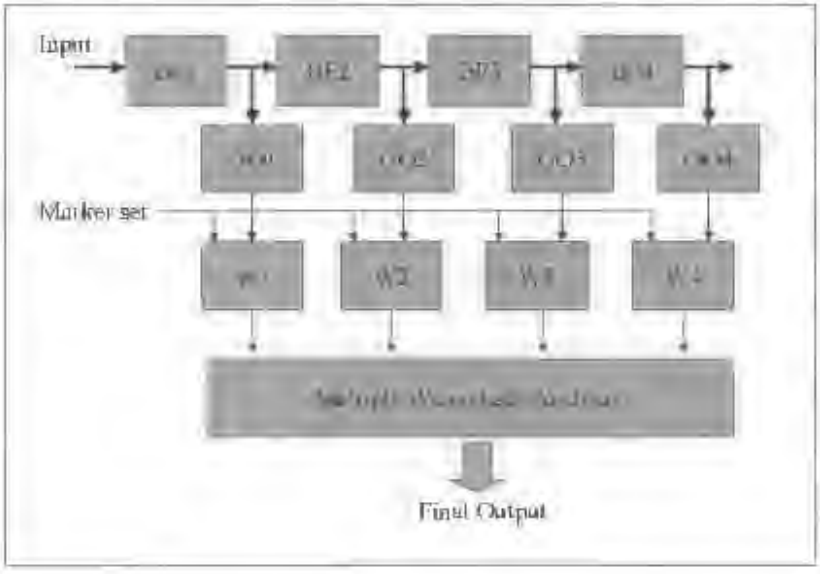


Figure 6.6: The structure of the watershed-based rock-scene segmentation framework

6.3.2 Incorporating the combinational approach

The structure of the basic framework as shown in figure 6.6 shows a single marker set input. However, as pointed out in section 6.2 there are two marker sets. Recall that the combinational approach, where both marker sets from the centroid and the reconstruction methods are used

in the segmentation, is proposed in section 6.2. It is clear that incorporating this approach as shown in figure 6.4 would require twice the number of executions of watershed processes that the basic framework currently holds. This would be a significant burden as far as computational speed is concerned.

A less burdensome solution is shown in figure 6.7. As can be seen, the marker-set from the reconstruction method (RM) is applied in more stages than the centroid method (CM) marker-set. This is because it is expected to be more accurate than the centroid method on scenes which contain predominantly small rocks. However, it tends to split larger rocks and therefore requires a correcting signal. The centroid method is expected to provide this signal by providing highly merged watersheds in such situations. Based on this reasoning, the CM marker-set can be injected at a position closer to the lower σ_r values, where the accuracy at tracing rock edges is expected to be higher, as shown in figure 6.7. This ensures accurate segmentation of larger rocks in the absence of cracks. There is a wide scope of options consisting of heuristic position assignments of the marker sets within the basic framework. Once the watershed images are obtained, the goal is to determine those regions which are most likely to be rocks using the watershed region features defined in section 6.2.3. This is achieved via multiple watersheds analysis.

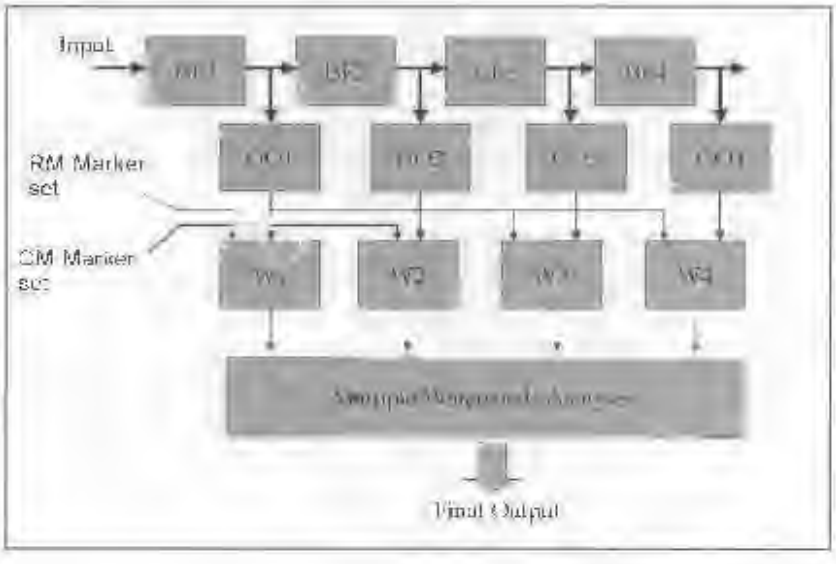


Figure 6.7: The structure of the modified watershed-based rock scene segmentation framework.

6.3.3 Multiple Watersheds Analysis

Multiple watersheds analysis is a procedure that combines the information from the various watershed output images to form a single segmented output image. This approach is required to extract from the population of n-watershed images those watershed lines which are most likely to be the actual rock edges. After much experimentation with various n values, the suitable value for n was found to be 8. Therefore 8 stages of the iterative bilateral filter are used resulting in 8-watershed outputs as shown in figure 6.8. The CM marker-set was injected at the 3rd watershed process from the highly merged watershed. As shown in figure 6.8, the images are arranged so that the 8^{th} watershed image was obtained using the centroid method marker-set.

Figure 6.8: The outputs of the watershed processes

The execution of the watershed-based segmentation algorithm outputs the n-binary watershed images so far. The proposed structure of the multiple watersheds analysis solution is shown in figure 6.9. In this figure, $n = 8$ and the binary watershed images (BW) are arranged so that the output from the CM marker-set is the 8^{th} image on the highly merged side. The procedure begins by adding the n-binary watershed outputs to obtain a single image. The motivation for doing this is that traces of rock edges on the n-images are expected to overlap. Therefore adding these binary watershed images is expected to result in high response values at the rock edges of the resultant image. It is also expected that the overlap of traces will not be perfect and this will result to disconnected boundaries when a high threshold is used for detection. On the other hand, a lower threshold will result to spurious traces being also detected. This serves as a motivation for adopting a multi-level thresholding approach. The image is thresholded at $(n-1)$-levels ranging from 1 to $n-1$, resulting in a further $(n-1)$-set of binary images (BT). As shown in figure 6.9, the binary watershed output from the CM marker-set is then appended as one of the resultant images, to give n-outputs.

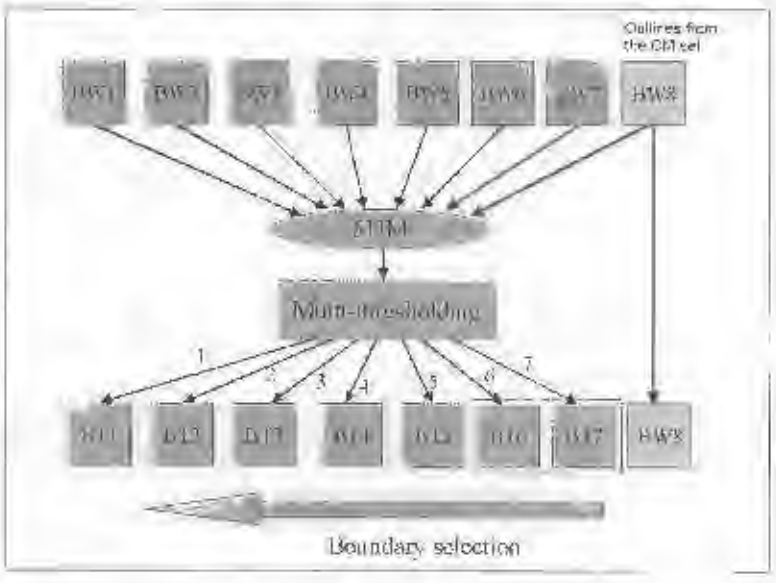


Figure 6.9: The proposed structure for multiple watershed analysis

In practice, the outputs of the highest two thresholds tend to have poor pixel connectivity and as a result they are removed, as shown by the bounding box of BT6 and BT7 in figure 6.9. An example of outputs after multi-thresholding the summation of the images of figure 6.8 is shown in figure 6.10. As can be seen, at low threshold values, the outputs tend to be highly split. While at high thresholds including the appended image, there is merging.

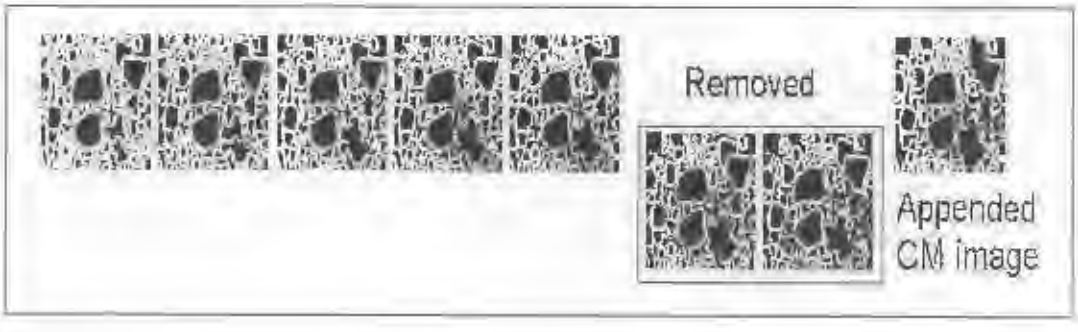
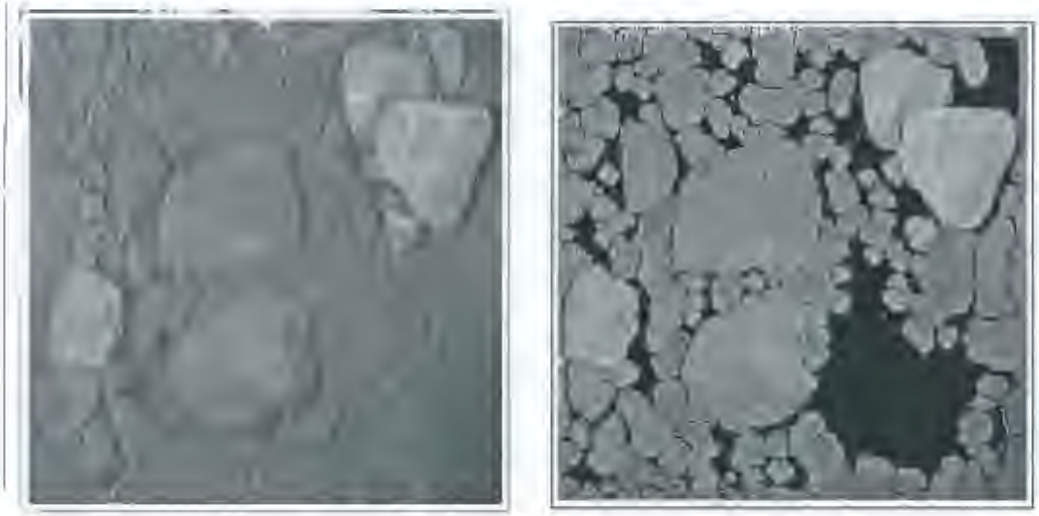


Figure 6.10: The multi-thresholding outputs

The multiple watersheds analysis is finalized by a boundary selection scheme. In this scheme, the multi-thresholding outputs are traversed from the highly merged to the highly split image as shown by the arrow of figure 6.9. At each image, boundary outlines which are roughly circular

in shape and whose proportional area with the corresponding area in the binary rock scene is greater than some threshold are collected. This scheme favors merged regions in a sense that if a particular outline meets the requirements in the current image, then in the next image the corresponding outline will not be considered. The final output of the multiple watershed analysis is shown in figure 6.11.



(a) Original rock-scene image (b) Segmentation output

Figure 6.11: Final segmentation results

The black patches are a result of the circularity C and the area proportion A_p parameters not being set to zero on BT1, as shown in figure 6.9. As a result, regions of BT1 which do not match the criteria are not selected by the boundary selection procedure. Provided that the BT1 parameters assume the lowest values as compared to the other BTs, then the ignored regions are the worst regions, and it is justifiable to leave them out.

6.4 Testing for robustness to varying lighting conditions

In this section, the watershed-based rock-scene segmentation is tested for robustness to noise, texture and surface cracks under varying lighting conditions. Six different images of ore from two mineral processing plants are used as the test-set. The images of the test-set consisting of

image1 to image6 are shown on the left in figures 6.12 to 6.17 respectively.

The first three scenes of figures 6.12, 6.13 and 6.14 are captured from the Waterval mineral processing plant. The ore under imaging is on a conveyor belt under a hood. A single light source at about 2m vertically from the conveyor belt to the hood is used. The camera is installed next to the light source. On the other hand, the ore images of figures 6.15, 6.16 and 6.17 are captured from another mineral processing plant under different lighting conditions. The images from the latter appear to have been captured under brighter lighting conditions than those from the former.

6.4.1 Parameter settings

The rock-scene segmentation algorithm has the following user-tunable parameters:

- Two windows w_1 and w_2 of the two-window adaptive thresholding method.
- The h-parameter of the reconstruction method.
- The number of stages N of the iterative bilateral filter and range σ_r of each stage is computed by determining the robust scale σ_s of each stage defined by equation 4.13. The weights $(0.25/1.4826, 0.4/1.4826, 0.55/1.4826, 0.7/1.4826, 0.85/1.4826, 1/1.4826)$ are used to weigh each σ_s value so that each weight from left to right weighs the corresponding σ_s value from the highly split to the highly merged end respectively. The resultant values are the selected σ_r values.
- The area proportion $(A_p)_l$ and circularity C_l at each level of the stack of watershed images. At $l = 1$ is the highly merged watershed image and at $l = 5$ is the highly split watershed image.

The actual settings are shown in table 6.1.

6.4.2 Algorithm evaluation

The aim here is to judge the output of the algorithm visually and numerically. Presented below are the outputs after applying the implemented segmentation algorithm on the test-set. The

Table 6.1: Parameter settings of the rock-scene segmentation algorithm.

Parameter	Regions
w_1	25x25
w_2	95x95
h	25
M	7
$(A_p)_1, C_1$	0.75, 0.80
$(A_p)_2, C_2$	0.70, 0.70
$(A_p)_3, C_3$	0.60, 0.60
$(A_p)_4, C_4$	0.25, 0.50
$(A_p)_5, C_5$	0.0, 0.50

visual results are shown on the right-hand side of figures 6.12 to 6.17. It can be seen that the algorithm performs well on average on all the images of the test-set. At this stage the average processing time for a 236 by 250 image is 36.032 seconds. The algorithm is implemented in matlab with C-mex files.

Based on the assumption that the actual rocks in the image are accurately segmented, the total negatives (fraction of non-rock) for each image are computed and presented in table 6.1. The actual number of rocks was determined manually from each image. The fraction of non-rocks N is defined as the fraction of the total number of regions (TNR) which are non-rocks. In formal terms

$$N = \frac{TNR - ANR}{TNR}, \tag{6.7}$$

where ANR is the actual number of rocks. Clearly, the fraction of rocks or total positives P is

$$P = \frac{ANR}{TNR}. \tag{6.8}$$

The fractions of non-rock regions are above the 50% mark for all the analyzed images, which is undesirable. These regions should be removed before measuring the rock size distributions, otherwise the errors will be large. Ideally, the fraction of non-rock regions should be reduced without significantly reducing the fraction of detected rock regions. In the next chapter, a set of tools from pattern recognition are applied to address this issue.

Table 6.2: False alarm rates of the segmentation algorithm.

image	TNR	ANR	N	P
image1	106	32	69.80%	30.20%
image2	106	27	74.52%	25.48%
image3	102	36	64.70%	35.30%
image4	145	39	73.10%	26.90%
image5	150	40	73.33%	26.67%
image6	138	57	58.70%	41.30%

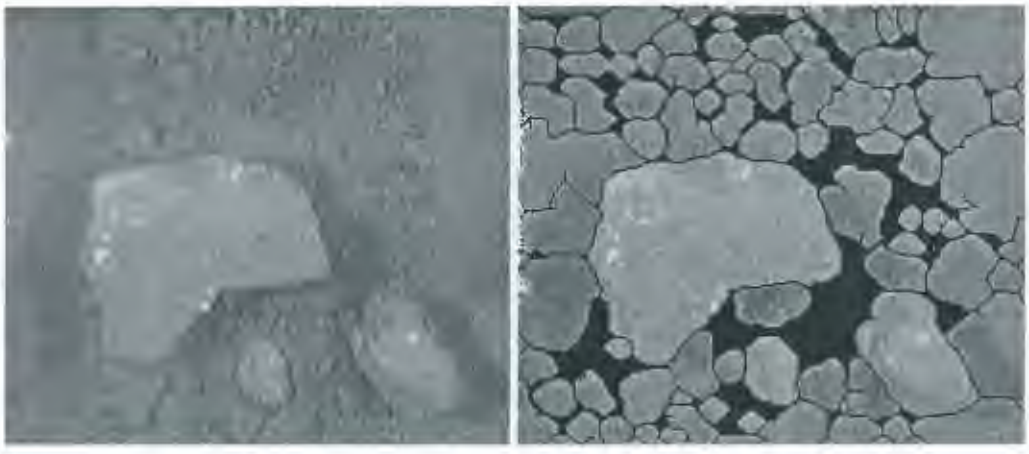


Figure 6.12: Test-image1 and the corresponding output after segmentation.

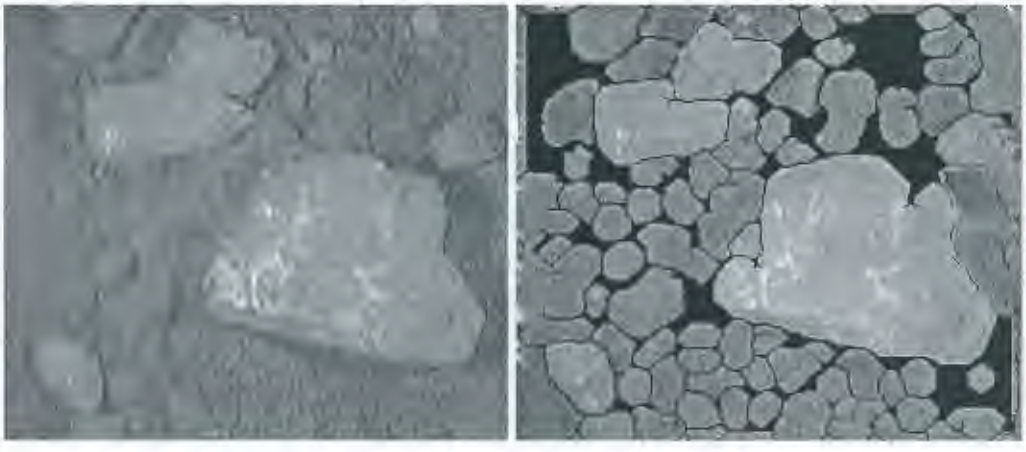


Figure 6.13: Test-image2 and the corresponding output after segmentation

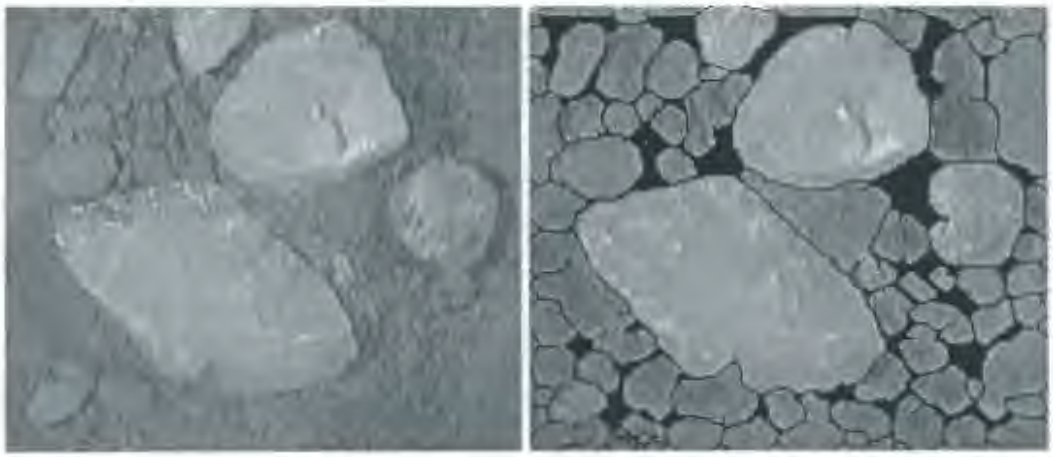


Figure 6.14: Test-image3 and the corresponding output after segmentation

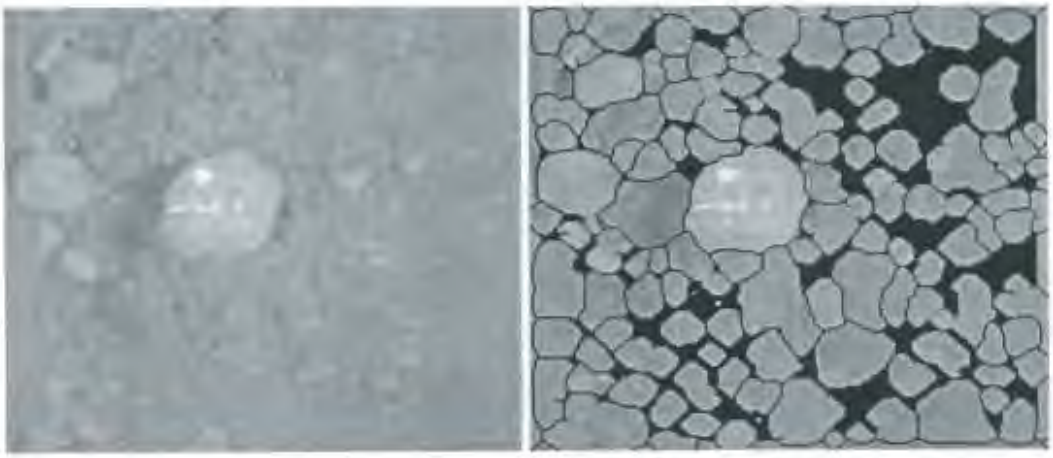


Figure 6.15: Test-image4 and the corresponding output after segmentation

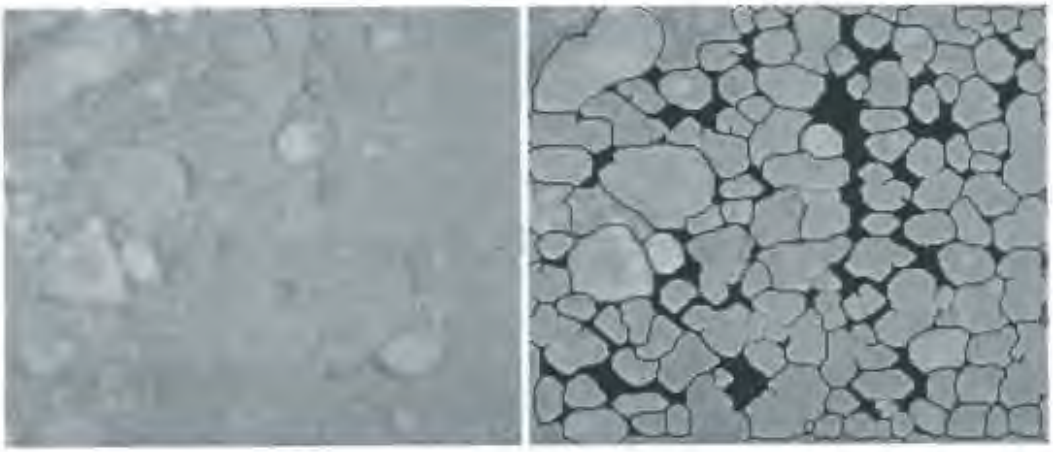


Figure 6.16: Test-image5 and the corresponding output after segmentation

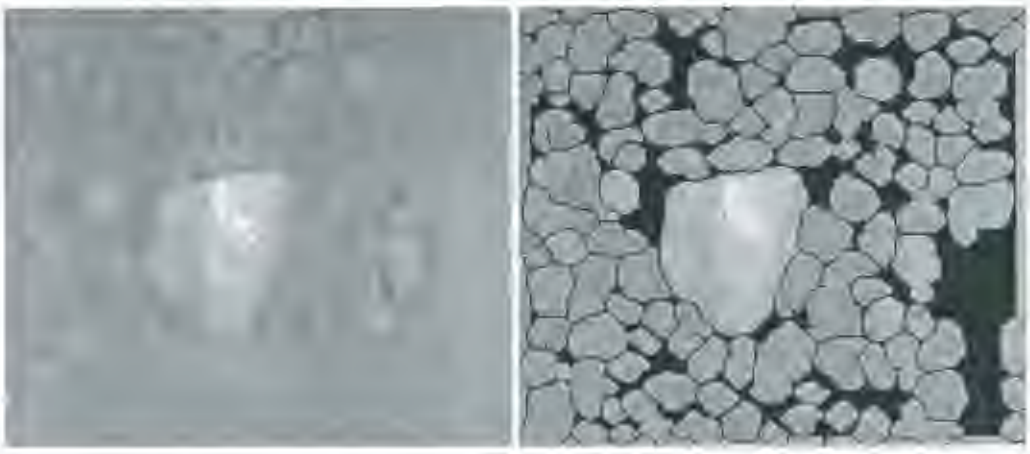


Figure 6.17: Test-image6 and the corresponding output after segmentation

6.5 Locating fines

The segmentation algorithm as it stands does not facilitate the measurement of fine particles. These are the small particles which range in size from 1 pixel to about 50 pixels in area. These particles are ignored by the algorithm since it is biased to selecting large circular particles as rocks. Most of the fine particles are rejected by the rock locator process to limit computational costs. These particles should be recovered via a less computationally expensive scheme. In this section, the moment preserving threshold is proposed as a method for detecting fine particles.

6.5.1 Using the moment preserving threshold technique

The moment preserving technique of chapter 3 has three intensity values as its output namely, the moment preserving threshold T, the background mean gray value w_0 and the foreground mean gray value w_1. The proposed method of locating fine particles using the moment preserving technique assumes that rock recognition has been applied prior to fines location. This allows negative masking of the coarse rocks of the original gray-scale image as shown in figure 6.18. The moment preserving threshold technique is then applied and using the foreground mean gray value w_1 as a threshold, the fine particles are detected, as shown in figure 6.19. A logical OR operator is the used to combine the detected fines and coarse particles and the final output is also shown in figure 6.19 (see also the flow of processing diagram in figure C.1). As can be seen, this technique does not recover all the fine particles in the image. This is mainly

due to the bias in detecting bright particles and the limited camera resolution.

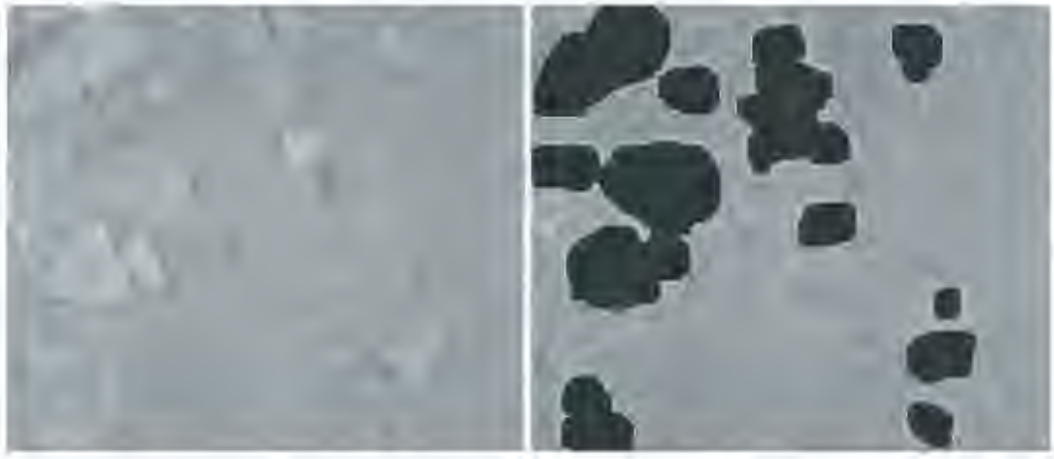


Figure 6.18: Original image and the corresponding negatively masked image

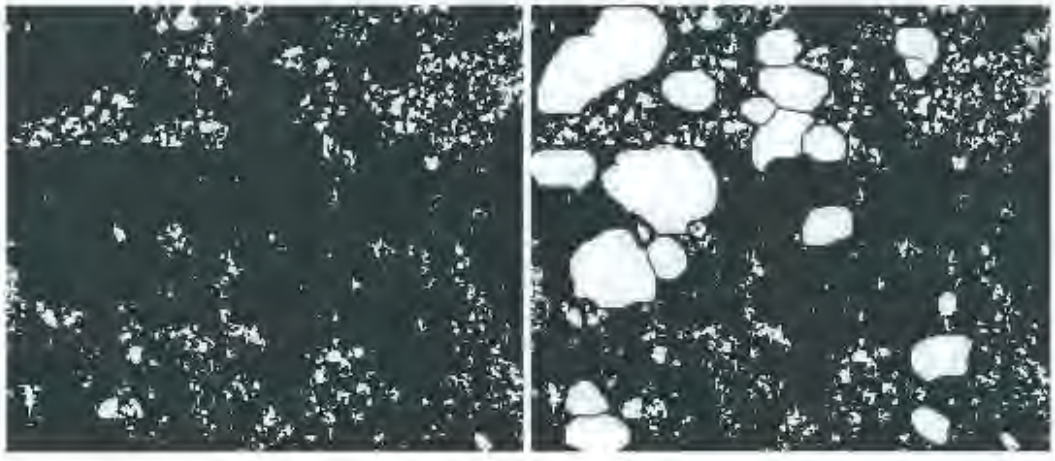


Figure 6.19: Located fines and the overall fine and coarse particles

6.6 Summary

In this chapter, the implementation of the rock scene segmentation algorithm is described. The main components of the algorithm are the automatic rock locator, the iterative bilateral filter, the marker-based watershed transform including marker extraction, the multiple watersheds analysis scheme and the fines locator component.

A rock locator for marker extraction is firstly applied on the image to provide indications of rock

locations. Two sets of markers are derived from a binary rock-scene using two methods in the form of the centroid and reconstruction methods. The centroid marker method tends to merge regions while the reconstruction method tends to split them. These properties are exploited in the watershed-based segmentation framework for segmenting rock-scenes.

The watershed-based segmentation of rock-scenes firstly filters the image at various degrees of edge preservation using the iterative bilateral filter. Watershed processes are then applied on each filtered output to obtain a set of *n*-watershed images. A multiple watershed analysis scheme that selects the watershed boundaries which are most likely to represent rock boundaries is then executed.

The final algorithm is applied to various ore images to test for its robustness to noise, texture and surface cracks under varying lighting conditions. The visual results are encouraging. However, the numerical results show high fractions of non-rock regions of greater than 50% on the test-set. The final component of the rock scene segmentation algorithm is fines locator. This component is expected to facilitate the measurement of fine particles.

Chapter 7

Rock feature classification for rock recognition

7.1 Introduction

The rock-scene segmentation algorithm that has been described thus far has very high fractions of non-rock regions, which is undesirable. This problem will cause large errors in the rock size distribution measurement by biasing the measurement towards larger rock sizes. It is a major requirement to reduce the non-rock fraction to a tolerable level, ideally without significantly reducing the fraction of detected rocks.

In this chapter, feature classification methods are investigated for the purpose of recognizing both rock and non-rock regions. Four feature classification methods, in the form of k-nearest neighbor (KNN), probabilistic neural network (PNN), kernel adatron support vector machine (SVM) and regularized least squares classification (RLSC), are selected for the investigation.

Feature subset selection methods are then investigated for the purposes of removing redundant and irrelevant features [23, 24]. These are the features that impair the performances of feature-based classification methods. Two commonly used feature subset selection methods in the form of separability index (SI) optimization and principal component analysis are investigated.

Finally, data is collected in the form of watershed boundary properties to form feature vectors. Feature subset selection methods are applied to remove redundant and irrelevant features. The feature classification methods are then trained and tested on the acquired data-set.

7.2 Preliminary definitions

A feature vector $[x_1, x_2, ..., x_n]$ of length n is a vector of n measurements whose elements are the measurements of each object's properties. Each object is then represented in an n-dimensional feature vector space as a point [53]. The matrix of feature vectors denoted by $\mathbf{X}$ can be constructed by arranging all the feature vectors into a block of numbers. It has size $p \times n$, where p is the number feature vectors and n is the number of measurements.

In supervised learning, each feature vector of $\mathbf{X}$ is assigned a target label t which takes on $\{+1, -1\}$ for binary classification and real numbers for regression problems. The labels for each measurement can then be appended to form a vector $\mathbf{t}$, leading to the training set $[\mathbf{X}, \mathbf{t}]$, as a set of examples from which a feature classification method can learn. In problems where there are many classes, a set of sequential binary classifiers can be used [23]. A functional mapping $f : \mathbf{X} \rightarrow \mathbf{t}$ is then learned and is expected to generalize well to new and unseen test feature vectors. This is the common property of all the feature classification methods that are investigated in this work.

7.3 Feature classification methods

This section presents the theory behind the workings of the selected set of feature classification methods.

7.3.1 The k-nearest neighbor (KNN)

The KNN is the simplest form of feature classification, where a test feature vector is assigned a majority label of the k-closest training feature vectors. For each test feature vector an indicator

function value y_i is computed from

$$y_i = \sum_{j=1}^{j=k} t_j/k. \tag{7.1}$$

Here t_j is a member of the k-nearest training features. A threshold at zero is then used to decide on the appropriate label. The constant k is an odd and positive integer. For the simplest case where the k-value is unity, the classifier is known as the simple nearest neighbor classifier. In this case, a test feature vector is assigned the label of the training vector that is closest to it. The commonly used distance metric is the Euclidian distance, which is also used in this work.

The underlying assumption for all the proximity-based classifiers is that feature measurements are relevant, independent and properly scaled [23]. If these assumptions hold then feature vectors with similar labels are expected to form distinct clusters in the feature vector space. In such situations, the performance of the KNN is state-of-the-art.

7.3.2 The probabilistic neural network (PNN)

The PNN is similar to the KNN method in that some measure of proximity is employed to quantify the similarity between two feature vectors. The underlying assumption that a feature vector should be surrounded by feature vectors of the same label also applies here, otherwise the feature measurements are irrelevant. The significant difference from the KNN is that a non-linear measure of similarity is obtained by non-linearly transforming the Euclidian distance via a kernel matrix K. The elements of this matrix are given by

$$K(i,j) = e^{(\frac{-D(x_i,x_j)^2}{2\sigma^2})}, \tag{7.2}$$

where D is the function for computing the Euclidean distance between feature vector x_i and x_j, and σ controls the degree of smoothness of the decision boundary and therefore the generalization to new data[53]. The choice of a kernel function is usually a Gaussian because it is widely used and is found to be successful for a wide range of problems[53].

The PNN decision function for each test point has the form

$$y_i = \sum_{j=1}^{j=p} t_j K(x_i,x_j), \tag{7.3}$$

where p is the number of points in the training set. This is a Gaussian weighted mean of the training targets t_j to obtain the estimate y_i of the test target for each test point x_i. A zero threshold value is then used to assign the appropriate label. This is similar to the KNN in that, a mean of the training targets is computed to estimate a label for each test point. However in the PNN, the whole training set (as opposed to k-training points) is used in the determination of each test label. Each training point label is assigned a weight which decreases with the distance from the test feature vector.

For small σ values, the indicator function of the PNN has a spiky appearance where each local gaussian function has a purely local influence. In this case, its behavior is similar to that of the simple nearest neighbor algorithm. For moderately large σ values, each local gaussian function has a slightly wider "sphere of influence", with the indicator function being smooth and therefore expected to generalize well. Extremely large σ values result in linear regressor where a hyperplane is fitted through the training data.

This classifier has been found to work well for problems with a moderate feature vector dimension and a large number of training feature vectors that cover the input space[24].

7.3.3 The kernel adatron Support Vector Machine (SVM)

The SVM is similar to the PNN in that it is also a kernel-based classifier. However, in the SVM the data are non-linearly projected to a higher dimensional feature vector space, where a linear separator or hyperplane can be used to classify the data more effectively[45]. This non-linear projection is achieved by using a kernel function in the form of the gaussian at each data point, as in the PNN case. However, for the SVM, each gaussian is weighted by the α_j weights as shown in equation 7.4:

$$y_i = \sum_{j=1}^{j=p} \alpha_j t_j K(x_i, x_j) \tag{7.4}$$

This equation shows that the target label of each test feature vector is a weighted sum of the Gaussian functions at that test point in the input space. This decision function is then thresholded at zero, in order to assign the appropriate label for each test point. The α weights are updated via an optimization process where the empirical risk of misclassification is minimized by maximizing the margin between the decision boundary and the support vectors [45]. This is

shown in figure 7.1.

An additional regularization parameter C is used to reduce the local distortion of the decision boundary, caused by the mixing of classes where class separation cannot be achieved by a reasonably smooth boundary [45]. The parameter C smooths the decision boundary by constraining the upper-bound on the magnitude of the largest α weight to C. This has the benefit of improving the generalization of the classifier to new data.

The SVM has been found to be state-of-art for many problems from a wide range of disciplines [45], [57], [33]. Its advantages are that it does not require a lot of data as the PNN because of its data compression feature, where only a subset of about 20% of the original training data set is retained as the support vectors [24].

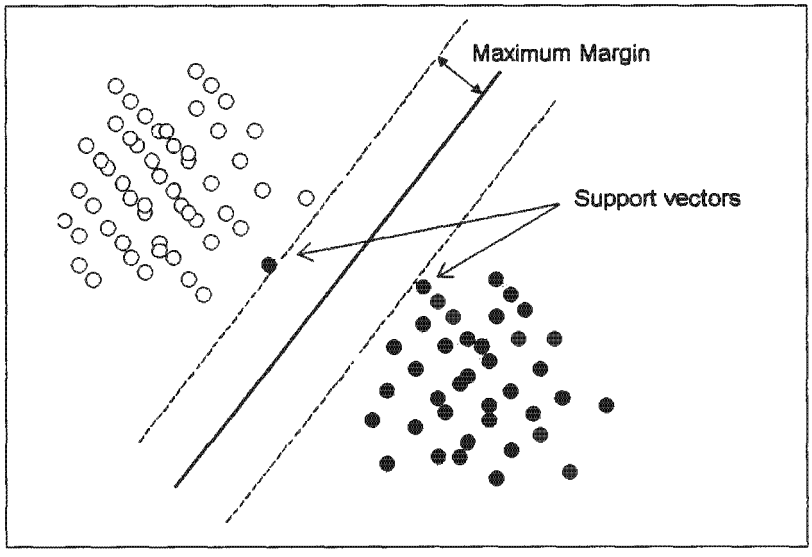


Figure 7.1: A support vector illustrating a case where a linear class separator is sufficient.

7.3.4 The regularized least squares classification (RLSC)

The RLSC or simple "regularization" algorithm is similar to the PNN and the SVM in that the indicator or decision function is also of the form:

$$f(x) = \sum_{i=1}^{i=p} c_i t_i K(x_i, x) \tag{7.5}$$

where the weights c_i are determined by least squares regression to minimize the empirical error on the training data set [53]. In mathematical terms, the following nonhomogeneous set of p linear equations has to be solved:

$$(p\gamma\mathbf{I}+\mathbf{K})\mathbf{c}=\mathbf{t} \tag{7.6}$$

$\mathbf{I}$ is the identity matrix, $\mathbf{K}$ is the kernel matrix where the diagonal elements are unity and γ is a user-tunable parameter. Let $p\gamma\mathbf{I}+\mathbf{K}$ be represented by the matrix $\mathbf{P}$. Then equation 7.6 becomes:

$$\mathbf{Pc}=\mathbf{t} \tag{7.7}$$

The matrix $\mathbf{P}$ is square and positive definite and therefore the inverse $\mathbf{P}^{-1}$ exists. It is also well conditioned provided that the $p\gamma$ of equation 7.6 is sufficiently large to strengthen the leading diagonal [53]. In practice, the inverse is replaced by pseudo-inversion where numerical ill-conditioning problems are anticipated. The solution of equation 7.7 is given by:

$$\mathbf{c}=\mathbf{P}^{-1}\mathbf{t} \tag{7.8}$$

where each element of vector $\mathbf{c}$ can be interpreted as the weight of importance for each Gaussian function of each training point. The smaller it is the less important the training feature vector is to the classification, and vise-versa. The decision function is then constructed and thresholded at zero, so that one of the +1,-1 is assigned to each test point.

It is apparent in [53] that the "regularization" algorithm has been applied successfully in many different applications. It has been used in both regression and binary classification type problems. One of the many mentioned modern regression applications is in the synthesis of images using computer graphics. In modern binary classification type applications, the problem of face recognition and classification of people by sex from digital images is mentioned. As a result, it may also be expected to perform well in this rock recognition problem.

7.4 Feature vector dimensionality reduction techniques

Feature classification methods usually perform well in a particular application provided that the following assumptions are valid:

- The feature measurements are independent.

- The feature measurements are relevant to the problem.
- There is enough data for the corresponding number of feature measurements.

Feature vector dimensionality reduction techniques are concerned with ensuring that the first two assumptions are valid. Given a particular data set, these techniques remove redundant and irrelevant features [23]. The techniques under investigation are the separability index (SI) optimization and principal component analysis (PCA).

7.4.1 Optimizing the Separability index (SI)

The SI is a figure-of-merit that gives an indication of the degree of class separation for a particular data set [24]. It was introduced by Thornton as the proportion of the feature vectors whose target label is the same as that of their nearest neighbors [23]. In formal terms this can be written as:

$$SI = \frac{\sum_{i=1}^{i=p} (\|t_i + t_i'\| + 1) mod 2}{p} \tag{7.9}$$

where t_i' is the target label of the nearest neighbor x_i' of feature vector x_i with target label t_i. For well-separated clusters of oppositely labelled points, the SI assumes a value close to unity. However as the clusters begin to merge, the SI falls to a point where the probability of each point having the same as its neighbor is 50% and the SI will be close to 0.50. A grid of points in the form of a chess-board will have an SI value close to zero.

It appears that subset selection can be achieved by optimizing the SI of a particular data-set. This can be achieved by evaluating the SI for each subset of features in the set and selecting the subset that has the maximum SI. This involves exhaustive search where $2^n - 1$ feature sets are evaluated with n being the dimensionality of the feature vectors. For moderate dimensions of sizes less than 10, this is a viable option [24].

However for larger dimensions, other search methods in the form of evolutionary algorithms are more suitable [24]. Evolutionary algorithms are biologically inspired processes which are based on Darwinian theory of evolution [24]. These methods do not require that all the feature subsets to be evaluated. Instead a form of guided random search is performed where only

the promising areas of the search space are explored. However, cautionary measures are also taken so that premature convergence does not occur[48], where the algorithm is trapped on a local maximum. An evolutionary algorithm in the form of Baluja's Probabilistic Incremental Learning (PBIL) algorithm [2, 3] is used in this work.

7.4.2 Principal component analysis (PCA) via singular value decomposition (SVD)

Principal component analysis (PCA) is a procedure for extracting the principal directions of a body of data where there is significant variability [24]. The principal directions are based solely on the feature vectors and therefore ignore the target labels. As a feature vector dimensionality reduction technique, PCA eliminates the principal components with insignificant variability. This can be achieved using singular value decomposition [24] as follows:

$$\mathbf{X}_{normalized} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{V}^T \tag{7.10}$$

where $\mathbf{X}_{normalized}$ is the normalized matrix of feature vectors. The normalization is performed by subtracting the mean of each column from the respective column and dividing by the respective standard deviation. The matrix $\mathbf{U}$ has columns which are mutually orthogonal unit vectors. These are the principal directions arranged such that their significance decreases with increasing index. The amount of variation for each principal direction is captured in the diagonal matrix $\mathbf{S}$. The matrix $\mathbf{V}$ has its columns as eigenvectors. It can be shown that the diagonal matrix $\mathbf{P}$ of corresponding eigenvalues can be determined from $\mathbf{S}$ using

$$\mathbf{P} = \frac{\mathbf{S}^2}{p-1}. \tag{7.11}$$

The closer to zero an element on the diagonal of $\mathbf{S}$, the more it is necessary to remove its corresponding principal direction. If the k columns of $\mathbf{S}$ with values less than some threshold are deleted, the following approximation to $\mathbf{X}_{normalized}$ is obtained:

$$\mathbf{Y} = \mathbf{U}_k \cdot \mathbf{S}_k \cdot \mathbf{V}_k^T \tag{7.12}$$

This is called the best rank-k least squares approximation to $\mathbf{X}_{normalized}$ [24]. The matrix $\mathbf{Y}$ has fewer features than $\mathbf{X}_{normalized}$ and thus feature dimensionality reduction is achieved. However it must be stated that since PCA does not take into consideration the target labels, it can eliminate features which are essential to the classification.

7.5 Methodology

A data set a consisting of 20 ore images taken under varying lighting conditions was assembled. These images are segmented using the rock-scene segmentation algorithm described in the previous chapter. The complete set of segmented images has a total of 2415 watershed regions which should be classified into rock and non-rock regions. The data is then partitioned as shown in table 7.1. A supervised machine learning approach is adopted, where the true target label for each region in the data set is assumed known. The true targets are determined manually and therefore some human error is expected. The fraction of actual rocks is also shown in the same table. The test set has an 87.31% fraction of negatives and the main aim is to recognize these regions and subsequently remove them.

Table 7.1: Data partitioning .

Partition	Images	Regions	Fraction of actual rocks	Fraction of non-rocks
Training	10	1117	22.74%	77.26%
Validation	5	644	13.35%	86.65%
Testing	5	654	12.69%	87.31%
Total	20	2415	17.52%	82.48%

7.6 Rock feature extraction

Eleven features are measured and can be broadly divided into rock shape, gray value and gradient characteristics.

7.6.1 Rock shape

The only feature under the rock shape category is the regional centroid to boundary distance variance feature. It measures roughly the circularity of the watershed region. Its computation involves firstly finding the centroid of the region and its boundary pixel coordinates b_i. A vector of distances d_i from each b_i to the centroid is constructed. Finally, the feature is then computed

as

$$f_0 = \frac{\sum_{i=1}^{i=n}(d_i - avg(d_i))^2}{n-1}. \tag{7.13}$$

This feature is expected to assume a low value for roughly circular regions and higher values for irregular regions.

7.6.2 Rock gray value

Seven features have been extracted which depend on rock gray value characteristics. The relevant descriptions are provided below.

Proportion of dark interior pixels

This feature is based on the appearance of the rocks as compared to shadows and fines. In most ore images, rocks appear brighter than the rest of the material. If a cumulative distribution of a gray value histogram of the image is computed, it can be concluded with high confidence that the bottom 10 % constitutes the fines and shadows. The gray value at this level is used as the threshold for detecting dark areas, in the form of fines and shadows in the image. The feature is then computed as the ratio of the number of detected dark interior pixels of the region to the total area of the region. It is expected to be low for rock regions because rocks are expected to be bright. It is expected to be high for non-rock regions because these regions are expected to be dark.

Proportion of dark boundary pixels

The same procedure as above is used to obtain the binary image of detected dark areas. However, instead of measuring from the regional interior, the detected set of pixels is measured on the boundary of the region. The feature is computed as the ratio of the number of detected pixels on the boundary to the region perimeter. It is expected to be high for rocks because shadows are expected on the boundary of the rock. Ideally, it is expected to be low for non-rock regions because shadows are not expected only on the boundary.

Proportion of thresholded area to region area

Recall the binary image that was used to derive the centroid and reconstruction methods marker-sets. The feature calculated here is computed as the ratio of the area of a region in this binary image to the area of the corresponding watershed region. It is expected to be high for rock regions and low for non-rock regions.

Average interior gray level

Rocks tend to have brighter gray values as compared to fines and shadows. The feature is therefore computed as the average gray value on a small disk around the region centroid. It is expected to be high for a rock region and low for a non-rock region.

Average boundary gray level

This feature is based on the knowledge that the gray value profile of a rock tends to fall off at the edges. It is computed as the average boundary gray level on the perimeter of the region. It is expected to be low for rock regions and high for non-rocks.

Boundary and interior gray level absolute difference

This feature is computed as the absolute difference of the boundary and interior average gray value features. It is expected to be high for rocks and low for non-rocks.

Interior gray level variance

This feature is computed as the variance of the gray values inside the region. It is expected to be low for rocks and high for non-rocks.

7.6.3 Rock gradient

This category has three features which depend on rock gradients. Their descriptions are given below.

Average interior gray level gradient

This feature is based on the knowledge that gray value discontinuities tend to be located at the edges under ideal lighting conditions. One of the many gradient operators can be used to obtain the gradient image. The feature is then computed as the ratio of the average gradient value on a small disk around the regional centroid. It is expected to be low for rocks and high for non-rocks.

Average boundary gradient

The gradient image is obtained as for the previous feature. The measurement of the average gradient value is performed on the boundary. This feature is expected to be high for rocks and low for non-rocks.

Boundary and interior gradient absolute difference

This feature is computed as the absolute difference of the boundary and interior average gradient feature values. It is expected to be high for rocks and low for non-rocks.

7.6.4 Summary of features

The summary of the features that are extracted is given in table 7.2. In addition each feature is assigned a symbol f_n where n is the feature number.

Table 7.2: Summary of extracted features.

feature	Symbol	rock	non-rock
Centroid to boundary distance variance	f_0	high	low
Proportion of dark interior pixels	f_1	low	high
Proportion of dark boundary pixels	f_2	high	low
Proportion of thresholded area to region area	f_3	high	low
Average interior gray-value	f_4	high	low
Average boundary gray-value	f_5	low	high
Average interior gradient value	f_6	low	high
Average boundary gradient value	f_7	high	low
Boundary and interior gray value absolute difference	f_8	high	low
Boundary and interior gradient value absolute difference	f_9	high	low
Interior gray value variance	f_{10}	low	high

7.7 Performance evaluation

The eleven features are measured for each region in the data-set summarized in table 7.1 to form data matrices $[\mathbf{X}_{train}, \mathbf{t}_{train}]$, $[\mathbf{X}_{validation}, \mathbf{t}_{validation}]$ and $[\mathbf{X}_{test}, \mathbf{t}_{test}]$. Before evaluating the performances of the various classifiers on the data-set, it is necessary to firstly evaluate the separability of the classes in the feature vector space. The SI is a quick and easy technique for doing this. The SI of the training set is found to be 78.25 %. This figure can be improved using feature subset selection methods.

7.7.1 Rock feature dimensionality reduction

Optimizing the SI using the PBIL

The feature vector dimension is eleven and thus exhaustive search is expected to be computationally expensive since 2047 evaluations will be performed. Therefore evolutionary search in

the form of Baluja's PBIL is a viable option. The settings as well as the amount of processing time it took for a pentium three 1.2GHz processor with 256 MB of RAM to complete a 1000 evaluations are shown in table 7.2.

Table 7.3: PBIL settings.

no. of evaluations	Learning rate	forgetting factor	processing time
1000	0.1	0.01	36 minutes

Applying this algorithm to the training data-set result in the plot of SI versus the number of evaluations shown in figure 7.2. As can be seen, the SI improves until 600 evaluations after which it remains at 82.45%. The selected subset of features consists of features f_0, f_1, f_4, f_5, f_8 and f_9.

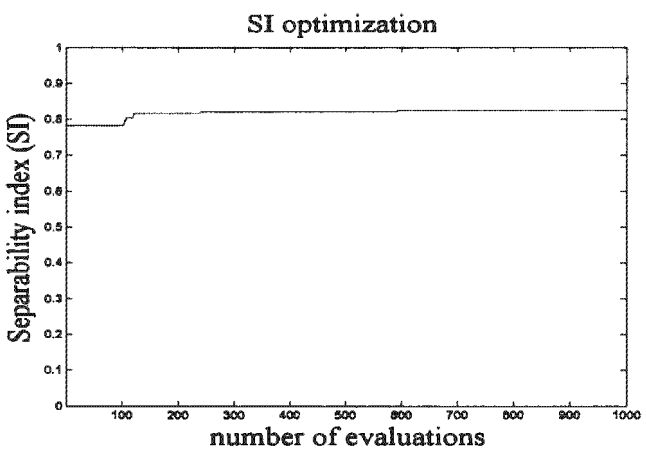


Figure 7.2: Optimizing the separability index.

In search of reasons as to why the features f_2, f_3, f_6, f_7 and f_{11} are not selected, the correlation matrix presented in table of figure 7.2 is analyzed. The selected subset of features is shown in vertically shaded columns. The non-shaded columns are the unselected features and the target correlation values under the label t. It is clear that features f_2,f_6 and f_7 should be eliminated as they are either highly correlated with one of the other features or are weakly correlated with the target. The elimination of f_3 and f_{10} might be justified by checking whether one feature is a linear combination of the other features. However, this is time-consuming and thus was not attempted.

	f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	t
f_0	1.0000	0.1536	0.0953	0.0359	-0.0703	-0.1148	0.2092	0.1117	-0.1188	0.0802	-0.0690	-0.0826
f_1	0.1536	1.0000	0.9170	-0.1351	-0.6909	-0.7573	0.2277	-0.1007	-0.5225	-0.1939	-0.1223	-0.1513
f_2	0.0953	0.9170	1.0000	-0.3327	-0.5872	-0.6620	0.1096	-0.0258	-0.4725	-0.0737	-0.0749	0.0016
f_3	0.0359	-0.1351	-0.3327	1.0000	0.5408	0.3724	-0.0389	0.3634	0.0956	0.4681	0.0299	0.3351
f_4	-0.0703	-0.6909	-0.5872	0.5408	1.0000	0.7476	-0.0398	0.4563	0.2625	0.5154	0.1721	0.3464
f_5	-0.1148	-0.7573	-0.6620	0.3724	0.7476	1.0000	-0.2799	0.1911	0.8215	0.3068	0.1695	0.1624
f_6	0.2092	0.2277	0.1096	-0.0389	-0.0398	-0.2799	1.0000	0.3801	-0.3861	0.1023	0.0412	-0.0004
f_7	0.1117	-0.1007	-0.0258	0.3634	0.4563	0.1911	0.3801	1.0000	-0.1161	0.9439	0.2161	0.4053
f_8	-0.1188	-0.5225	-0.4725	0.0956	0.2625	0.8215	-0.3861	-0.1161	1.0000	0.0051	0.0920	-0.0627
f_9	0.0802	-0.1939	-0.0737	0.4681	0.5154	0.3068	0.1023	0.9439	0.0051	1.0000	0.2259	0.4409
f_{10}	-0.0690	-0.1223	-0.0749	0.0299	0.1721	0.1695	0.0412	0.2161	0.0920	0.2259	1.0000	0.2035
t	-0.0826	-0.1513	0.0016	0.3351	0.3464	0.1624	-0.0004	0.4053	-0.0627	0.4409	0.2035	1.0000

Figure 7.3: The table of correlation coefficients

Removing insignificant principal components

The principal component analysis via singular value decomposition is applied on the normalized training data set. A scree plot of $\sqrt{\frac{eigenvalues}{p-1}}$ values where p is the sample size is shown in figure 7.4. It is apparent that the first eight principal components should be retained, based on the fact that the eigenvalues fall off significantly after the 8^{th} principal component.

Table 7.4: SVD-based data transformations and their merits.

Input feature transformation	Notation	SI
normalized training data	$\mathbf{X}_{normalized}$	78.25%
best rank-8 approx. to whitened data	$\mathbf{U}_8$	78.69%
best rank-8 approx to data rotated into principal components	$\mathbf{U}_8 \cdot \mathbf{S}_8$	78.25%
best rank-8 approx. to original data	$\mathbf{U}_8 \cdot \mathbf{S}_8 \cdot \mathbf{V}_8^T$	78.33%

Table 7.4 shows various attempts at improving the SI by reducing the feature vector dimensionality to eight and transforming using PCA via SVD. The highest improvement in SI of 0.44%

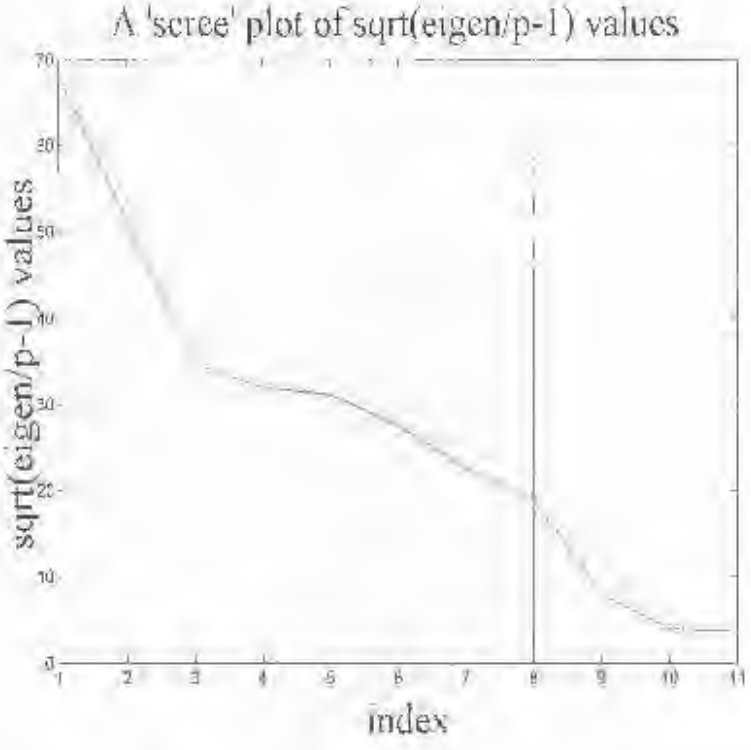


Figure 7.4: A scree plot of eigen values showing where the cut-off threshold should be taken

is achieved when the best rank-8 approximation to data rotated into the principal component transformation is used. Based on these results it is apparent that PCA is not significantly useful at improving the separability of the classes in the feature vector space.

7.7.2 Training and testing

The training of classifiers entails finding the parameters that will result in acceptable performance levels. The set of classifiers under investigation, consisting of KNN, PNN, SVM and RLSC, were trained separately on the validation set and the parameter settings were determined. The settings and the training accuracy on the validation set, are shown in table 7.5.

Testing the classifiers for generalization on the test set results in the generalization results shown in table 7.6. It appears that the RLSC or "regularization" algorithm has outperformed the rest of the feature classification methods. This is followed by the PNN with generalization accuracies of 91.59%, which outperforms the SVM. Finally, the KNN has the least generalization accuracy of about 90%.

Table 7.5: Classifier training results.

Feature classification method	Parameter1	Parameter2	Training accuracy
KNN	$k = 13$	none	90.00%
PNN	$\sigma = 0.51$	none	90.99%
SVM	$\sigma = 0.51$	$C = 100$	90.22%
RLSC	$\sigma = 0.51$	$\gamma = 0.09$	90.37%

Table 7.6: "Accuracy" and "Precision"

Classifier	Time	accuracy	FPR	TPR	precision
KNN	1.5 sec	89.91%	6.13%	63.04%	91.13%
PNN	1.5 sec	91.59%	1.22%	42.16%	97.20%
SVM	6.39 sec	90.67%	1.41%	36.17%	96.25%
RLSC	12.03 sec	92.35%	2.28%	55.40%	96.05%

In addition to the accuracies of the classifiers, the overall false positive (FP) and true positive (TP) rates are also shown in table 7.6. The fraction of rocks in the test set is 12.69% before rock recognition is applied. This constitutes the total positives (P), while the fraction of non-rocks of 87.31%, constitutes the total negatives (N). For each classifier the false positive, true positive rate and precision are computed from the following[15]:

$$FPR = \frac{FP}{N}, \tag{7.14}$$

$$FPR = \frac{TP}{P}, \tag{7.15}$$

and

$$\text{precision} = \frac{TP}{TP + FP}. \tag{7.16}$$

As can be seen in table 7.6, the best classifier in terms of a good compromise between TPR, FPR and processing speed is the simple KNN classifier. This is because it has an acceptable FPR and precision, and the highest TPR on the test-set, and its computation is almost instantaneous.

However, the TPR and FPR values of a particular classifier determine the operating point of that classifier on an ROC (Receiver Operating Characteristic) plane [15]. The operating point

is usually changed by varying some threshold that distinguishes between the two classes (in binary classification). This results in a curve that specifies the performance of the classifier for each threshold value. The TPR and FPR values that are reported here, form a single point on the ROC for each classifier, as shown in figure 7.5. The diagonal line is the ROC curve of a purely random classifier. For each threshold, this classifier is expected to achieve equal true and false positive rates. It is random because the likelihood that a positive is true or false is the same for all the thresholds. Good classifiers should have ROC curves which reside within the top triangle or above this random line.

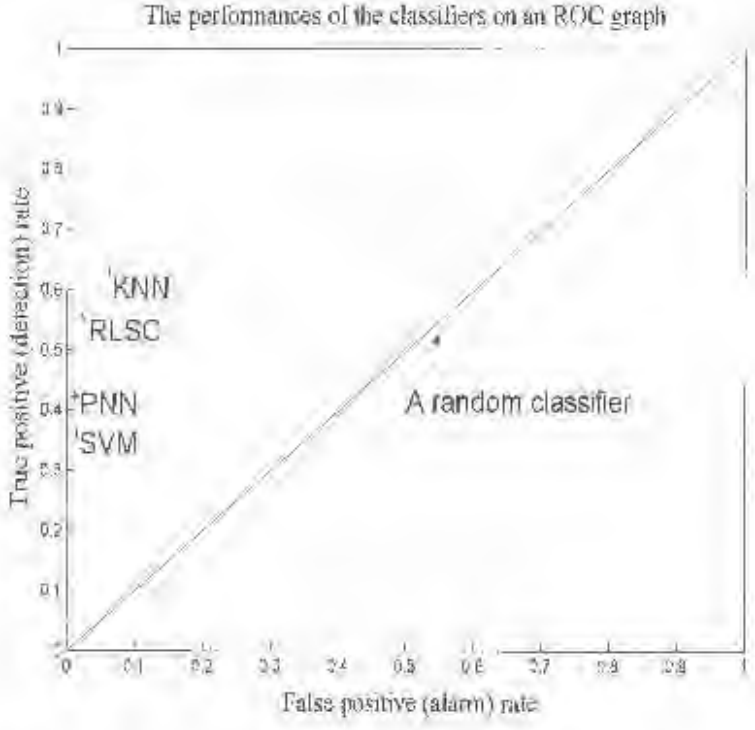


Figure 7.5: The ROC space, showing the performances of the four classifiers

In [15], it is argued that classifiers should not be compared based on a single point on an ROC plane, because the result of the comparison might not be valid at other points. It is also stated that an ROC curve of a classifier can be generated by varying its decision threshold. In this work, the decision thresholds were set to zero for all the classifiers. In other words, the decision functions of equations 7.1, 7.3, 7.4 and 7.5 were compared to a zero threshold, in order to decide on the most probable label. Clearly, varying this threshold changes the decision boundary of the classifier, and therefore the TPR and FPR. The threshold is varied between $-\infty$ and $+\infty$.

The ROC curve starts at (0,0), where the threshold is $+\infty$. As the threshold is lowered to the point that the decision function is encountered, more true positives and less false positives are progressively encountered. This results in the slope of the ROC curve being steeper. As the global minimum of the decision function is approached, more false positives and less true positives are progressively encountered, as a result the slope of the ROC curve decreases. At $-\infty$, both the true positive and false positives are fully detected and therefore both TPR and FPR assume unity values. The ROC curves of the four classifiers are shown in figure 7.6.

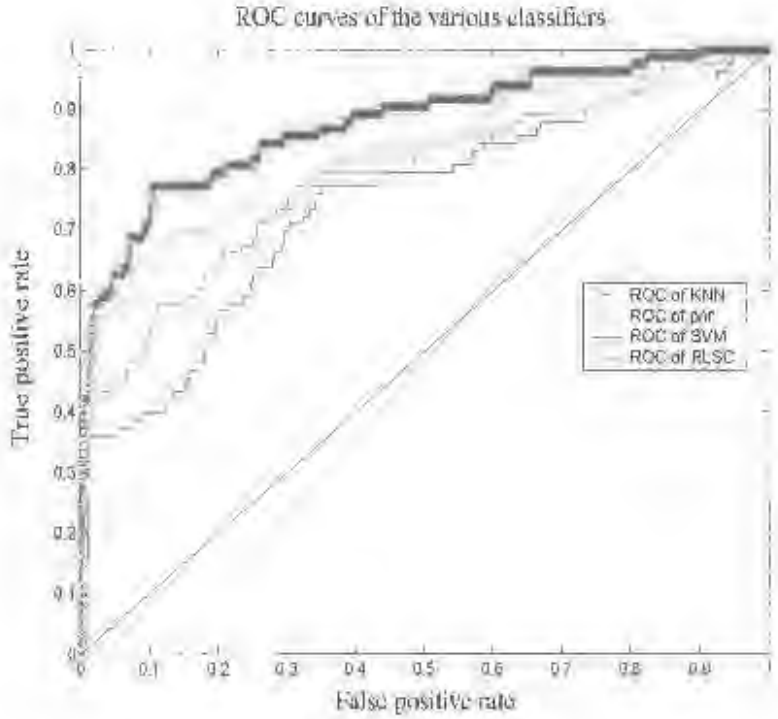


Figure 7.6: The ROC curves of the four classifiers

A perfect classifier should have an ROC curve that passes through the point (0,1). In other words, there should be a threshold where its false positive and true positive rates are zero and unity respectively. Any classifier that is closest to this perfect ROC curve is the the best classifier on the test set. As can be seen in figure 7.6, the KNN is the best classifier on this test-set. The KNN is further tested on a subset of the images that were used to test the segmentation for robustness to variations in lighting conditions. The obtained visual results are shown in appendix A.

The visual results on the test-set of the KNN and the RLSC are shown in figures 7.5 and 7.6.

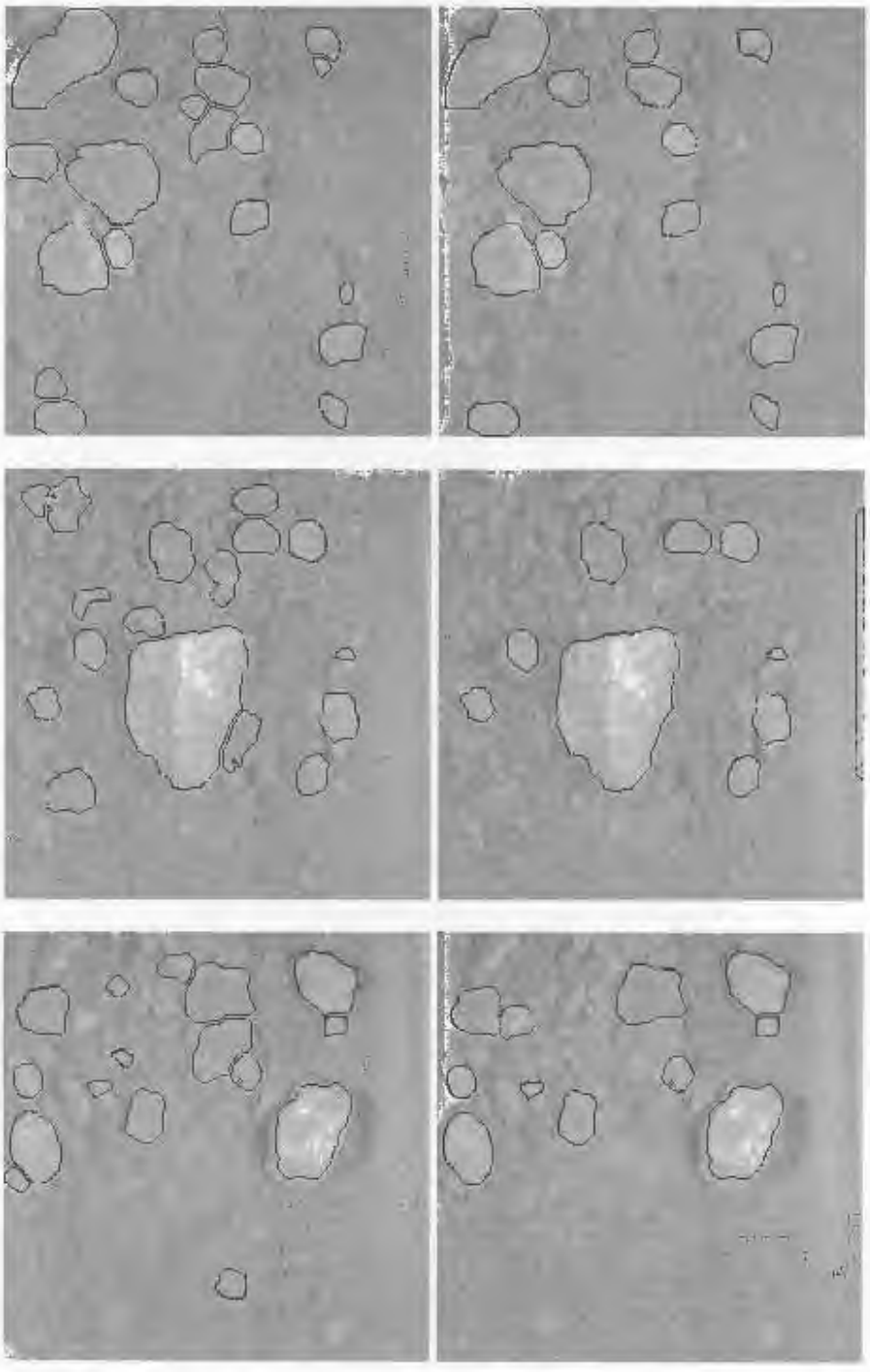


Figure 7.7: Test results on test-set images 1-3, showing KNN classified regions on the left and RLSC classified regions on the right.

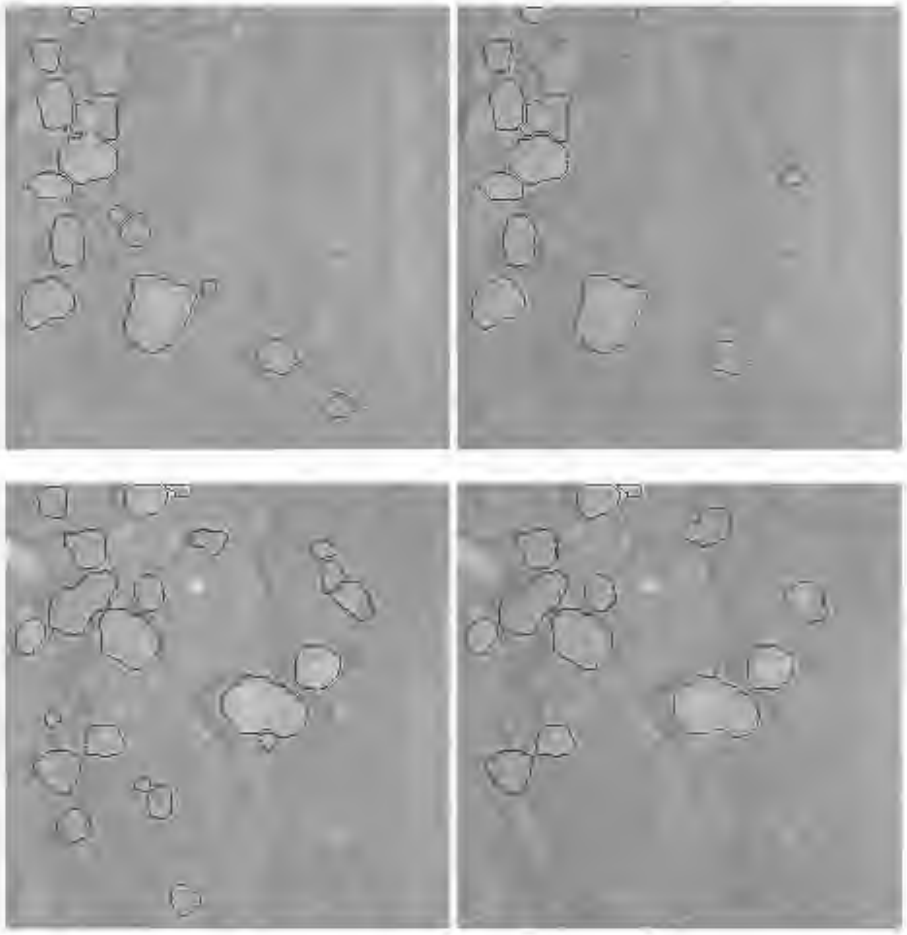


Figure 7.8: Test results on test-set images 4-5, showing KNN classified regions on the left and RLSC classified regions on the right.

7.8 Summary

The main aim of this chapter was to investigate, implement and apply feature classification methods to reduce the high false alarm rates of the segmentation algorithm. Four classification methods in the form of KNN, PNN, SVM and RLSC are selected. All of these classifiers are data-driven.

Data was collected and feature vector reduction methods in the form of feature subset selection using SI and PCA via SVD were applied. Optimizing the SI using the PBIL algorithm improved the SI from 78.25% to 82.45%. The PCA via SVD method was not successful at improving the separability of the classes in the feature space. This was probably due to its ignorance of the target classes.

The classifiers were then trained to determine their optimal parameter settings. Testing for generalization followed, and the results showed that the RLSC method outperformed all the other methods in terms of generalization, but that the KNN has the best compromise between the false alarm rates, true detection rates and speed of computation. The emphasis is placed on the achievement of the main objective of reducing the FAR on the test-set from 87.31% to 5.35% with the simple KNN classifier.

Chapter 8

Results

8.1 Introduction

In this chapter, the overall results of this dissertation are presented. Firstly, projected rock area distributions at pixel level on test-set images are measured. These measurements are then compared to projected area distributions of manually segmented images. The primary objective of this experiment is to evaluate the ability of the system to accurately measure projected rock area distributions. The secondary objective is to ascertain whether the system has reasonable outputs by using the hand-segmented data as ground truth.

Secondly, real plant data in the form of sieved size distributions with corresponding video sequences are collected. The video sequence is analyzed by the system, and the projected rock-area distribution is measured and compared to the actual mass fraction distribution of the sieved data. A difference in the two distributions is expected since the methods are not measuring the same thing.

Empirical corrections are then made to enable a fair comparison between the two size distributions of the belt-cuts. Both visual and statistical results of the comparisons are presented. Finally, a summary of the overall results is presented.

8.2 Measuring projected rock area distributions

In this section the results are presented in terms of two images for visualization purposes, and the corresponding distributions for qualitative comparisons. One image is the system's output and the other is the corresponding manually segmented image. The Human Visual System (HVS) is capable of locating rocks at a global level, where it is easy to distinguish between rock and non-rock patches. As a consequence the time consuming process of location of fines is not attempted. Consequently, the system's fines locating capability is also switched off to ensure a fair comparison between automatically and manually determined rock size distributions. As far as the plotting of the distribution is concerned, a method for creating suitable bin sizes is required. In this work, bin widths are determined through the use of the Freedman-Diaconnis rule[29]. The bin width h is given by:

$$h = 2(IQ)n^{-\frac{1}{3}} \tag{8.1}$$

where IQ is the interquartile range and n is the size of the data-set. The projected rock area distributions are plotted on the same set of axes for comparison purposes. The test-set has five images, and thus five area distributions of each category (manually or automatically determined) can be averaged for each bin to obtain a single distribution. Finally, a quantitative comparison between the average distributions from the manual and automatic method is made.

At this point, the system can measure projected area distributions in terms of the number of pixels as shown in figures 8.1 and 8.2. As can be seen on this image, the automatically determined projected area distribution closely matches that of the manually determined distribution. However, it must be stated clearly that the manually determined distribution is not the actual distribution of the material in the picture. It is only the approximate projected area distribution of the rocks which are visible to the human eye.

The projected rock area distributions of the rest of test-set images are shown in figures 8.3 to 8.10. Figure 8.11 shows the average distributions for each category. It can be deduced by qualitative comparisons that the automatically determined distributions closely match that of the manually determined projected area distributions. However a quantitative measure of this match is required.

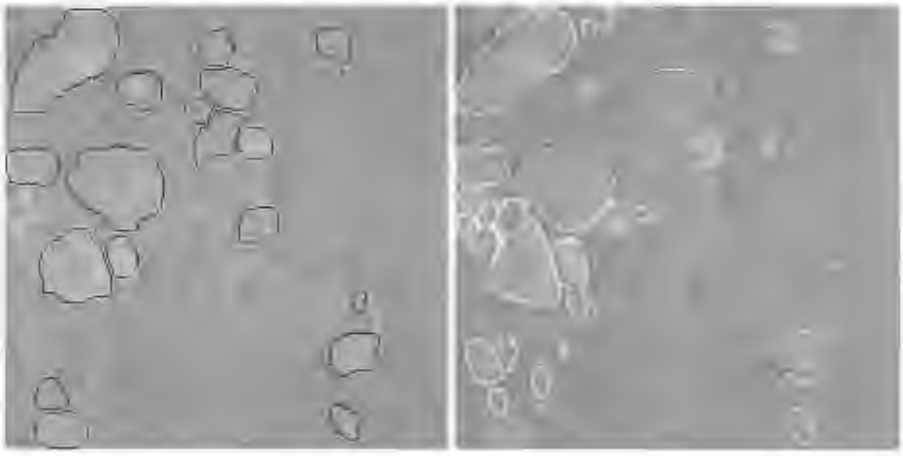


Figure 8.1: Test image1 segmented using the KNN-based segmentor is shown on the left and on the right is the manually segmented version.

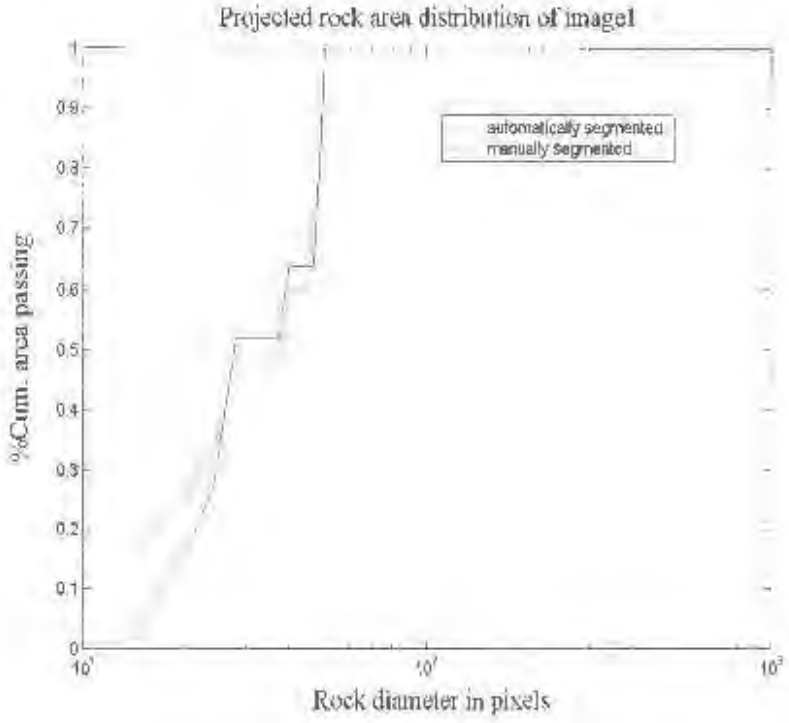


Figure 8.2: The corresponding projected area distributions of the automatically and manually segmented images of test-set image1.

One quantitative measure of difference is obtained by computing the RMS (root mean square) error between the two distributions. This error measure is found to be 2.37% on this data-set. Another approach of quantitatively comparing two cumulative distributions is outlined in [49].

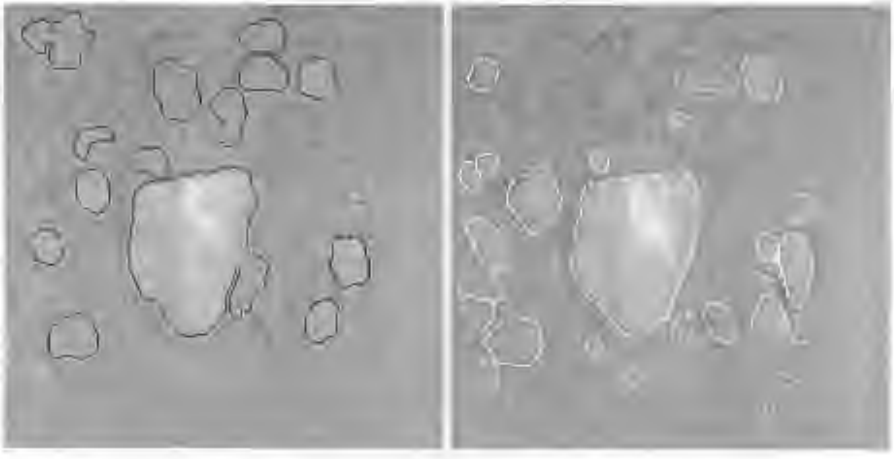


Figure 8.3: Test image1 segmented using the KNN-based segmentor is shown on the left and on the right is the manually segmented version.

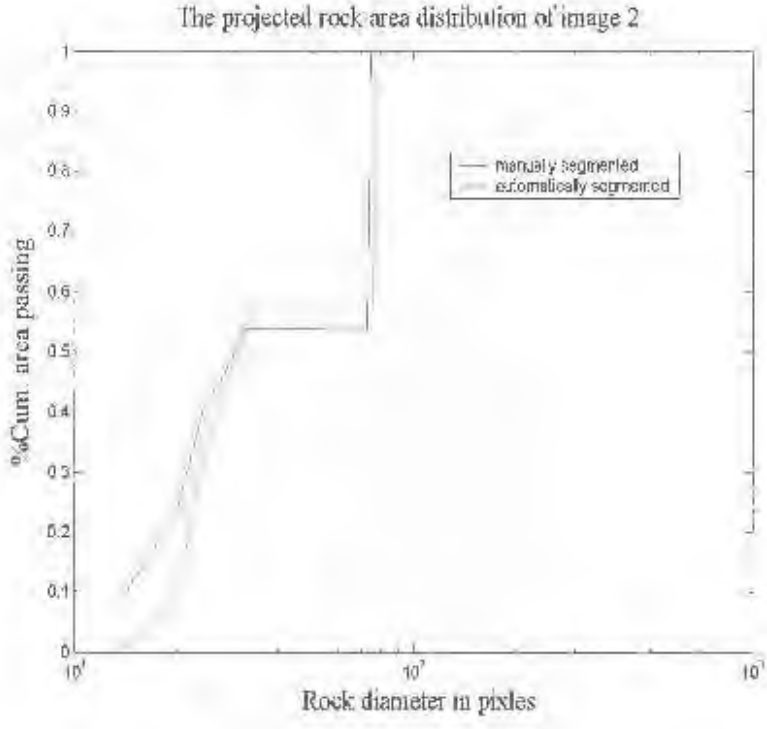


Figure 8.4: The corresponding projected area distributions of the automatically and manually segmented images of test-set image2.

The process begins by linearizing the two distributions using a Rosin-Rammler transformation, where the x-axis and the y-axis are transformed using:

$$x_{new} = \ln(x) \tag{8.2}$$

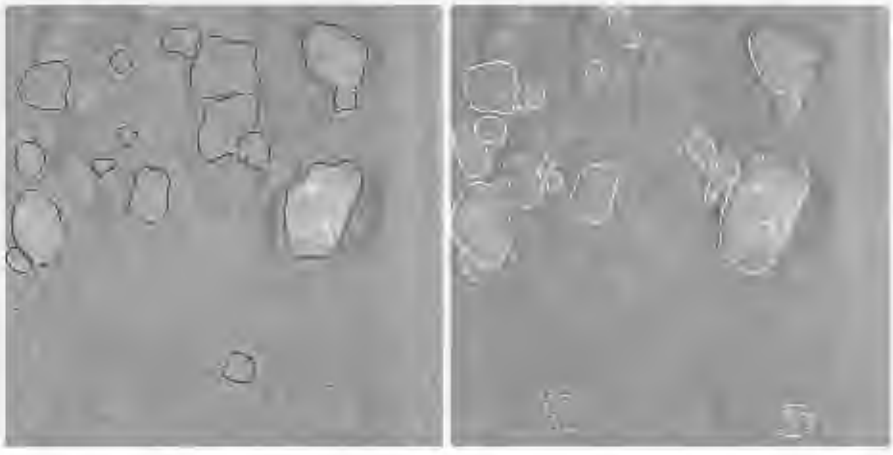


Figure 8.5: Test image1 segmented using the KNN-based segmentor is shown on the left and on the right is the manually segmented version.

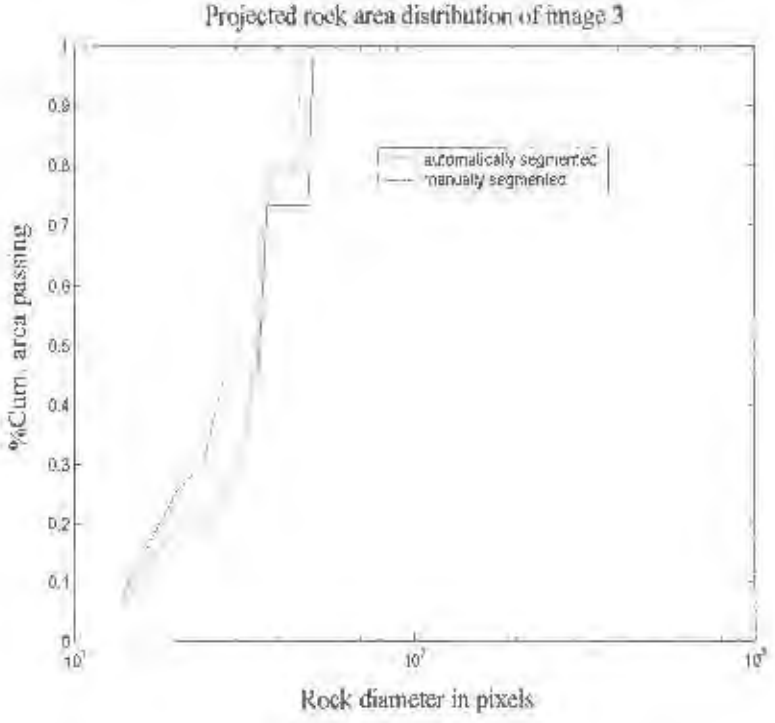


Figure 8.6: The corresponding projected area distributions of the automatically and manually segmented images of test-set image3.

and

$$y_{new} = \ln(\ln(\frac{100}{y})) \tag{8.3}$$

It is claimed in [49] that this should lead to a linear relationship between the transformed vari-

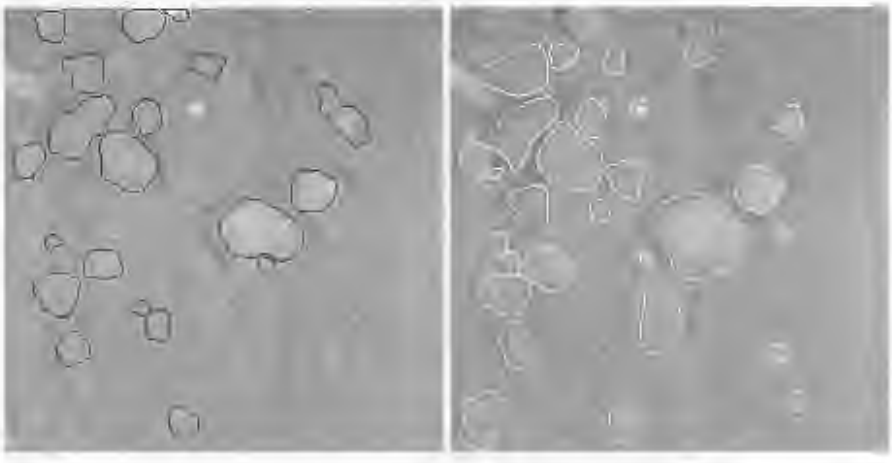


Figure 8.7: Test image1 segmented using the KNN-based segmentor is shown on the left and on the right is the manually segmented version.

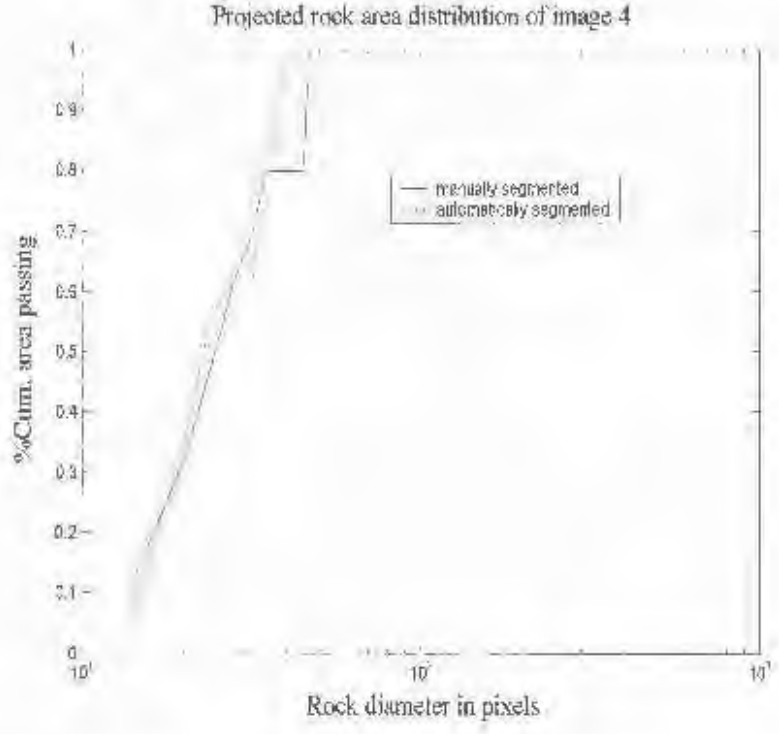


Figure 8.8: The corresponding projected area distributions of the automatically and manually segmented images of test-set image4.

ables. However, the variable y assumes values less than or equal to unity, and the reciprocation followed by the amplification results in large numbers for values close to zero. The attenuation provided by the two natural logarithms is insufficient to compensate for this distortion, and as

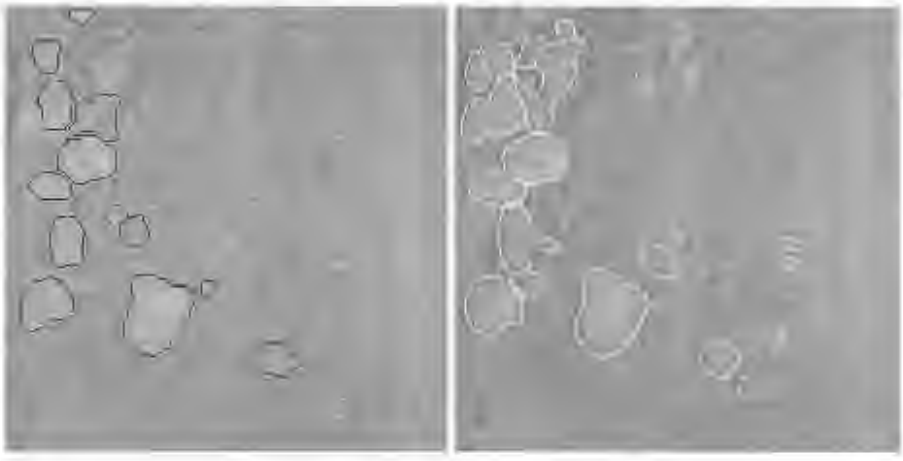


Figure 8.9: Test image1 segmented using the KNN-based segmentor is shown on the left and on the right is the manually segmented version.

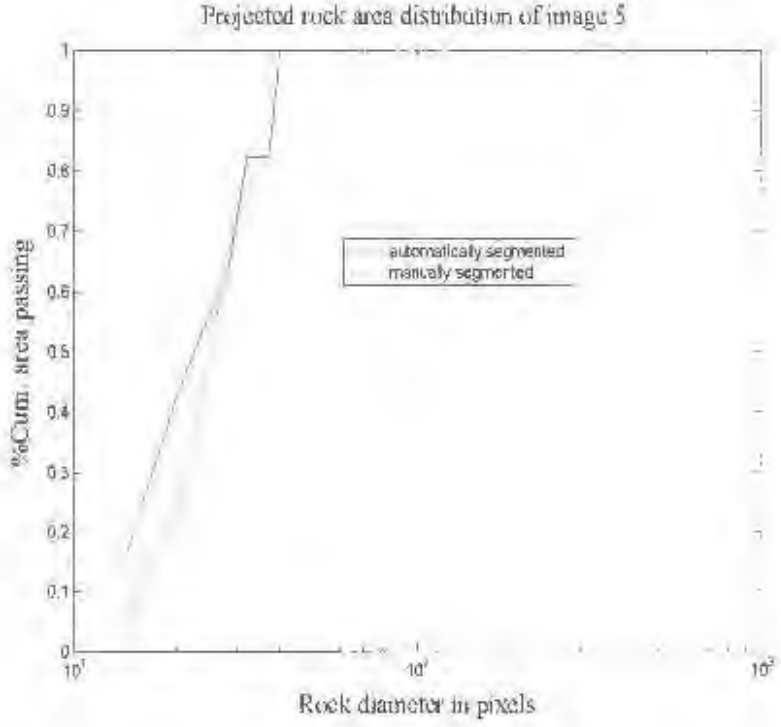


Figure 8.10: The corresponding projected area distributions of the automatically and manually segmented images of test-set image5.

a result the output is non-linear. Based on this fact, an alternative procedure for linearizing is adopted in this work. This procedure involves transforming only the x-axis as done in equation 8.2. Regression lines are then fitted on the resultant data and the results are shown in table 8.1.

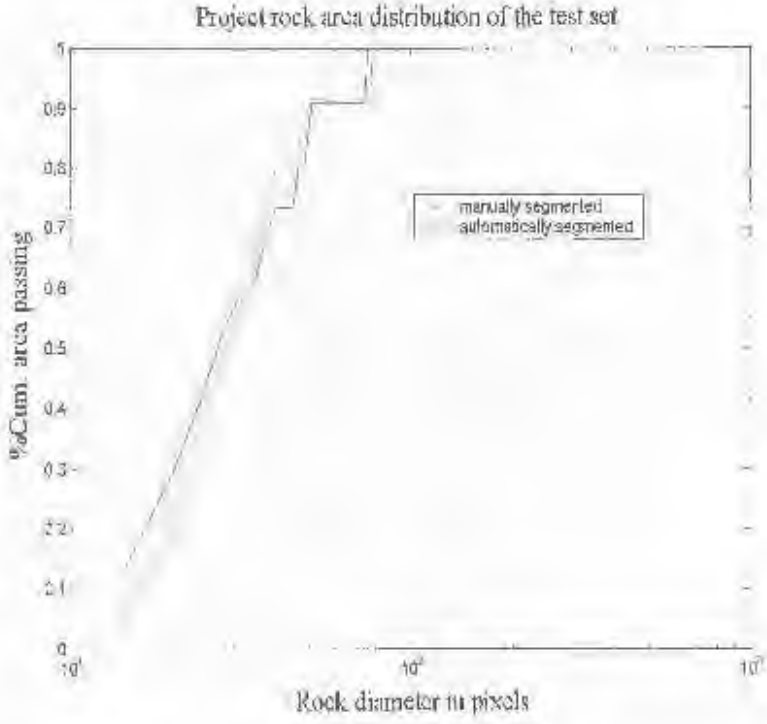


Figure 8.11: The overall projected rock area distributions on the test-set

It appears that a linear fit is significant for both lines since the F-statistics obtained from the regression outputs are significantly greater than the $F^{0.005}_{(1,28)}$ value of 9.28 from the F-tables. Therefore the linearity assumption is justified.

Table 8.1: Machine vs Manual measurements linear regression results

Parameter	Automatic Line	Manual Line
b_0	-1.2396	-1.1263
95% confidence interval on b_0	(-1.4867 to -0.9926)	(-1.2831 to -0.9695)
b_1	0.5214	0.4923
95% confidence interval on b_1	(0.4583 to 0.5846)	(0.4523 to 0.5324)
R^2	91.09 %	95.77%
Observed F-statistic	286	633.57

The two lines are shown in figure 8.12. Visually, they appear to be very close to each other. A hypothesis test for slopes, with H_0 being the null hypothesis stating that the two slopes are

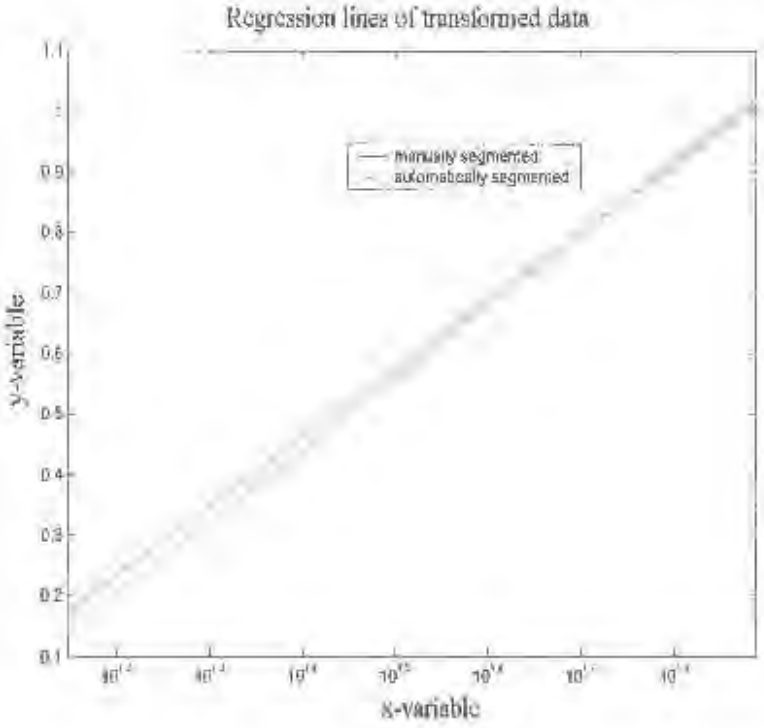


Figure 8.12: Regression lines of linearized projected area distributions of the test-set.

similar, and the alternative hypothesis H_1 being the two slopes are different is carried out to investigate this closeness. In formal terms,

$$H_0 : b_a = b_m \tag{8.4}$$

$$H_1 : b_a \neq b_m \tag{8.5}$$

where b_a is the slope of the automatically determined data and b_m is the slope of the manual measurements. The test is a two-sided t-test because the null hypothesis H_0 should be rejected when $b_a - b_m$ is either significantly negative or positive. The test statistic is given by

$$t = \left|\frac{b_a - b_m}{s_{b_a}}\right| \sim t_{n-k-1} \tag{8.6}$$

where t_{n-k-1} is the student's t-distribution with $n-k-1$ degrees of freedom and s_{b_1} is the standard error on b_1. The variables n and k are the number of samples (30) and independent variables (1) respectively. The standard error on the slope b_a is given by

$$s_{b_a} = \sqrt{\frac{MSE}{s_x^2(n-1)}} \tag{8.7}$$

where MSE is the mean square error of the automatic regression line on the data, and s_x is the standard deviation of the independent variable x. The quantities s_x and MSE are determined to be 0.4262 and 0.0047 respectively. Using these values, the test statistic value is determined to be 0.960. The 5% significance level for the two-sided test is $t_{28}^{0.025} = 2.048$ which is significantly greater than the test-static value. It can be concluded that there is insufficient evidence to reject H_0 at the 5% significance level, and therefore the two slopes are deemed to be similar.

A similar approach as the one used for testing for similarity in slopes is adopted for testing for similarity in constants. The null and alternative hypothesises are:

$$H_0 : c_a = c_m \tag{8.8}$$

$$H_1 : c_a \neq c_m. \tag{8.9}$$

The test statistic is given by

$$t = |\frac{c_a - c_m}{s_{c_a}}| \sim t_{n-k-1} \tag{8.10}$$

as before, where s_{c_a} is the standard error on c_a and is computed as

$$s_{c_a} = \sqrt{\frac{MSE}{n-1}}. \tag{8.11}$$

The test-statistic is determined to be 8.7735 and using a significance level of 5%, H_0 is rejected. Therefore based on statistical tests on the similarity of regression lines, the two distributions are not similar. The discrepancy is at the fine end where the machine generated size distribution appears to be coarser than the ground truth. The major sources of error are the segmentation algorithm and the KNN classifier. However, the encouraging outcome is that the visual results seem to suggest that the two distributions are similar. In addition, expanding the number of images in the test-set and repeating the statistical tests may confirm this suggestion.

For the purposes of representing distributions in terms of rock mass fraction against size in *mm* as is done in the mineral processing industry, a conversion from the representation of the projected rock area in terms of the number of pixels to squared-centimeter cm^2 is required. This conversion can be done using:

$$\text{Rock Area} = A_{rock} \cdot f(p_s, r_f, m) \tag{8.12}$$

where A_{rock} is the rock area in pixels and f is a function of pixel size p_s, image reduction factor r_f and a magnification factor m [11]. The conversion from area to diameter in cm is achieved using

$$\text{Rock diameter} = 2\sqrt{\frac{RockArea}{\pi}} \tag{8.13}$$

This is a diameter of a circle with the same area as the projected rock area $RockArea$.

8.3 Comparing machine measured size distributions to sieved data

The previous section tested the system on a test-set of images, and compared the distribution obtained to the distribution of corresponding manually segmented images. The visual results show that the two distributions are similar, but quantitative results are not convincing. In this section a further evaluation of the system on sieved data is carried out. In particular, qualitative and quantitative comparisons between the automatically measured and the actual sieved rock-size distribution are carried out.

8.3.1 Methodology

Sieved data was collected in the form of rock-size distributions of 3 and 5 metre belt-cuts with corresponding video at the Waterval mineral processing plant in Rustenburg. The two rock-size distributions as obtained from the test-work are shown in figure 8.13. It appears that the $5m$ belt-cut is finer than the 3m belt-cut at the fine end and coarser at the coarse end. This is because the first 5 meters of the total belt-cut consists of coarse and fine material, while the last 3 meters consists of only fine material.

Calibration objects in the form of discs with diameters of $64mm$ were included with the material under imaging. The scaling from pixels to cm^2 is determined to be one pixel area being approximately equal to one square-centimeter. This allows us to map the diameters from pixel values to millimeter values. The methodology for performing a comparison between the machine measured and sieved rock-size distributions involves using the sieved distribution of the $5m$ belt-cut

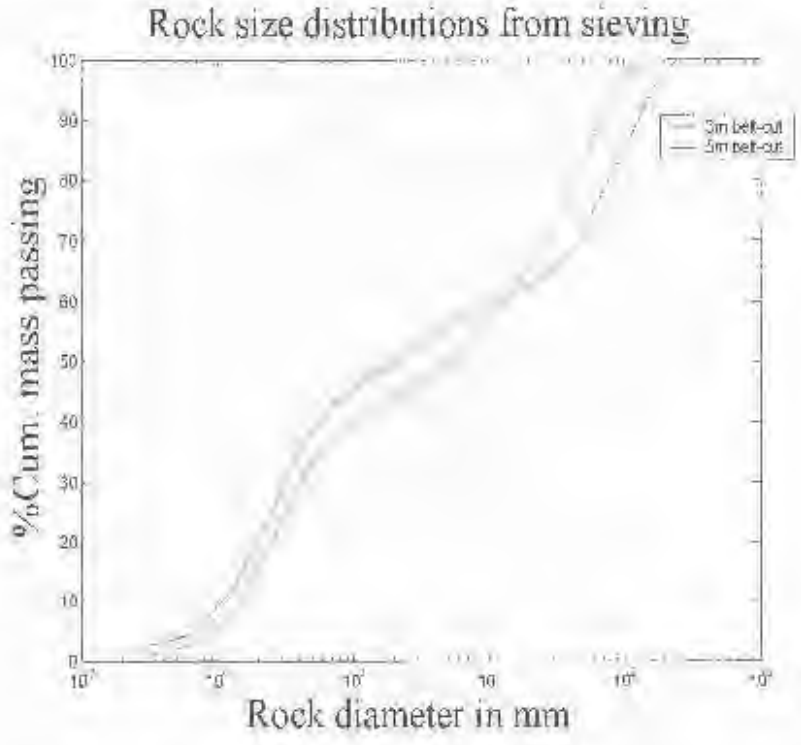


Figure 8.13: The 5m belt-cut sieve size distribution with the corrected projected rock area distribution of the machine vision system.

for empirical corrections, while retaining the $3m$ belt-cut for evaluating the effectiveness of the corrections.

The total belt-cut video sequence has 1084 frames. This is a 1 minute and 12 second long movie when played at 15 frames per second. The number of frames was reduced to 362 by sampling at 3 frame intervals to reduce redundancy. The first 227 of the 362 frames belong to the $5m$ belt-cut and the rest to the $3m$ belt-cut. The 5m belt-cut frames were analyzed by the system and the resultant distribution is shown in figure 8.14. It is apparent that there is a dead-band in the range between zero and $6mm$. This is due to mainly the limited resolution of the cameras and other factors outlined in [43], such as missing fines.

8.3.2 Dead-band correction

A slightly better approximation of the sieve distribution is obtained by fitting an s-curve at the fine end to correct for the dead-band problem and missing fines as shown in figure 8.14. Fitting

a standard "tail" or s-curve to estimate fines is standard procedure in the comminution industry [65]. This is done to estimate fine particles which are too small to be measured. The extent of dead-band correction is determined as the length of the dead-band and is found to run from the 1^{st} sample to the 19^{th} sample of the machine generated distribution curve. The procedure is as follows: The machine generated cumulative distribution is firstly converted to a probability distribution function (pdf); the first 19 samples of the pdf of the sieved 5*m* belt-cut are inserted into the first 19 slots (dead-band length) of the machine generated pdf. the modified machine generated distribution is then normalized to obtain a dead-band corrected pdf, and it is then converted to a cumulative distribution function(cdf).

This correction is tested on the 3m belt-cut to evaluate its usefulness. The visual results as shown in figure 8.15 are encouraging.

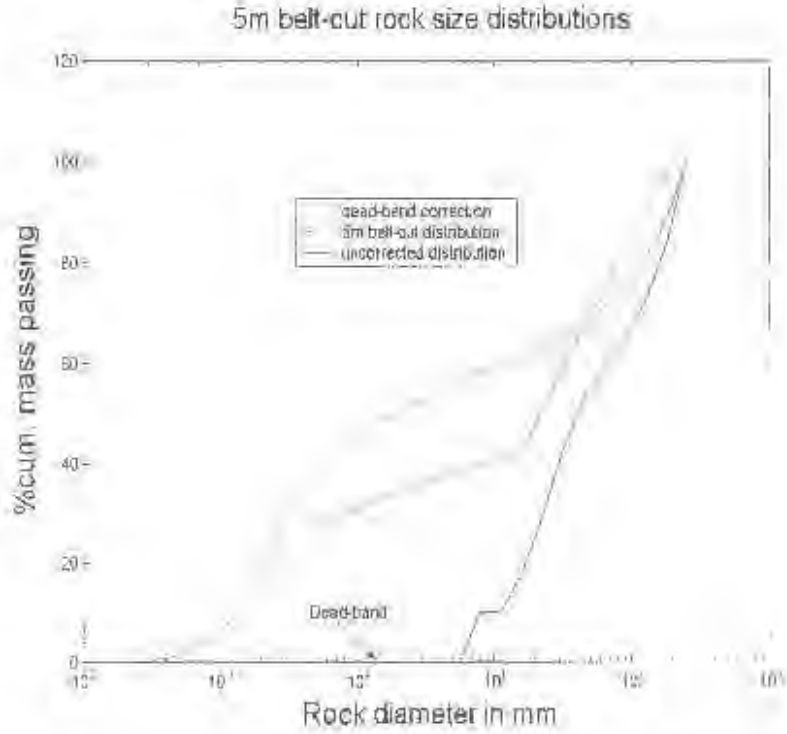


Figure 8.14: The 5m belt-cut sieve size distribution with the corrected projected rock area distribution of the machine vision system.

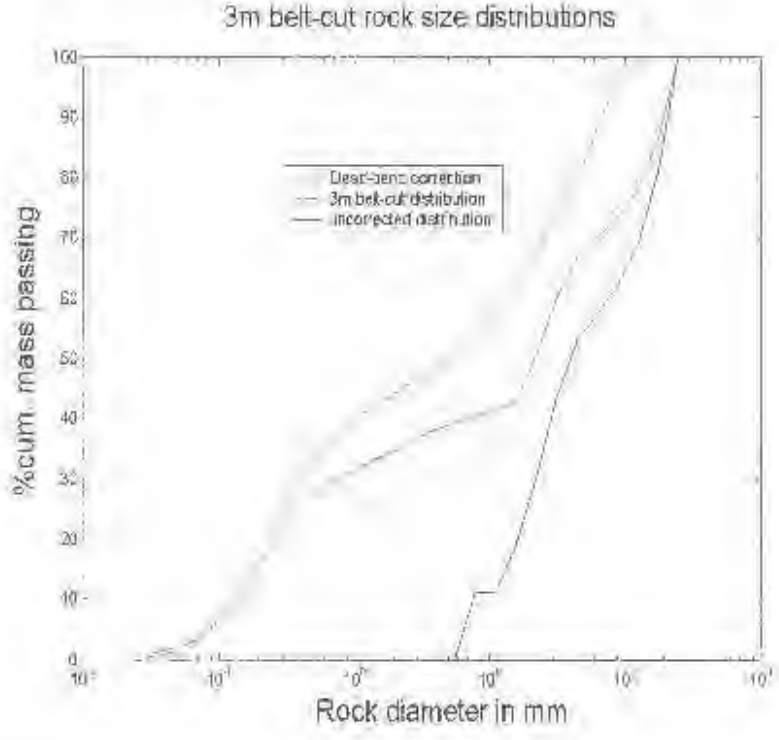


Figure 8.15: The 3m belt-cut sieve size distribution with the corresponding corrected projected rock area distribution of the machine vision system.

A quantitative comparison can be done using the same technique which is used to quantify the difference between manual and automatic distribution measurements. Results in table 8.2 show that the two linear fits are significant and thus the linearity assumption is justified.

Table 8.2: Machine vs sieve distributions linear regression results on the 3m belt-cut data

Parameter	Machine Line	Sieve Line
b_0	0.2917	0.3601
95% confidence interval on b_0	(0.2649 to 0.3185)	(0.3414 to 0.3788)
b_1	0.0959	0.1164
95% confidence interval on b_1	(0.0866 to 0.1052)	(0.1099 to 0.1229)
R^2	94.74 %	98.21%
observed F-statistic	450.5	1368.5

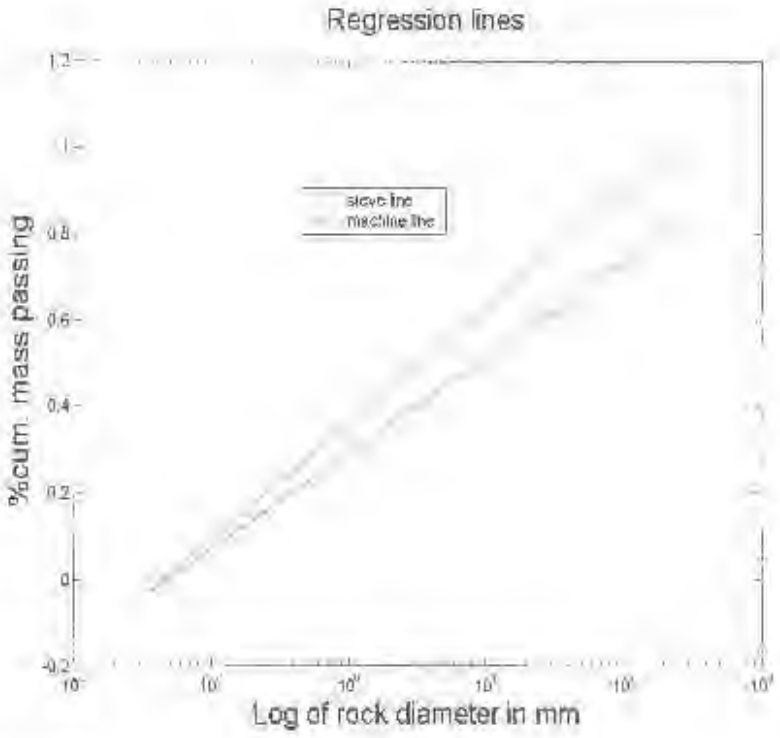


Figure 8.16: A visual comparison between the machine and sieve regression lines.

A visual comparison of the regression lines shown in figure 8.16, shows that the two lines appear to be different. But how significant is this difference? A similar procedure as that which was used to compare two regression in [49] is used here. The results of the t-test are summarized in table 8.3. It appears that the two slopes are significantly different since H_0 is rejected at the 5% significance level. There is no need to go further and test for the similarity of the constants. If lines have different slopes then they are different. Therefore the two rock-size distributions are significantly different.

Table 8.3: Hypothesis testing results for similarity between the machine and sieve measured size distributions.

Test	H_0	H_1	t-stat	rejection level
slopes	$b1_{machine} = b1_{sieve}$	$b1_{machine} \neq b1_{sieve}$	4.727	5%
constants	$b0_{machine} = b0_{sieve}$	$b0_{machine} \neq b0_{sieve}$	none	none

This result is expected because what is being measured by the two approaches is different. The sieving method measures the overall distribution of the material while the machine vision

system measures the projected areas of rocks as well as overlapping rocks on top of the pile. In [43], they go further and state that assuming that the rocks are anisotropic in shape and that they lie flat, the machine vision system will tend to measure the major and intermediate diameters. While the sieve methods will measure the minor and intermediate diameters. Therefore the machine measured distributions will tend to be coarser than those of the sieve methods as seen in figures 8.14 and 8.15.

Further corrections in the form of compensating for overestimation due to "fragment lay" and under-estimation due to overlapping fragments [43] should be made. The Wipfrag system of [43] uses statistical transformations which are based on stereology and geometric probabilities to reduce these errors. The same transformations of their work are attempted from here on.

8.3.3 Stereology-based correction

As seen in the previous section, that simply fitting an s-curve at the fine-end is not sufficient even though the visual results obtained are encouraging. In this section, other empirical corrections, based on stereology and geometric probabilities, are attempted. In particular, the selected transformation is based on the following stereological relationship from [43]:

$$N_A = \frac{M}{2\pi} N_V \tag{8.14}$$

where N_A is the number of rocks per unit area on the sectioning plane, N_V the number of particles per unit volume intersected by the sectioning plane and M is the mean curvature of the particles. Assuming rocks to be spheres where $M = 2\pi d$, the resultant relationship becomes

$$N_A = dN_V, \tag{8.15}$$

where d is the average diameter. Therefore the number of rocks per unit volume for each bin in the histogram can be determined using:

$$N_V(d_{ave}) = \frac{1}{d_{ave}} N_A(d_{ave}), \tag{8.16}$$

where d_{ave} is the average rock diameter of each bin. In [43], a calibration function f is included to account for overlapping fragments and the effect of missing fines. The new equation is

$$N_V(d_{ave}) = \frac{1}{d_{ave} f(d_{ave})} N_A(d_{ave}), \tag{8.17}$$

This appears to be a generalization of equation 8.15, since it is the specific case where $f(d_{ave})$ is a set of unity values.

The stereological solution of equation 8.15 is applied prior to applying dead-band correction. See also the visual results of applying the calibrated stereology correction 8.16 in appendix A. The extent of dead-band correction is determined on the 5*m* belt-cut data and its length is heuristically determined to be 19 samples. The two corrections are then applied in the specified order and the results on the two belt-cuts are shown in figures 8.17 and 8.18. The results appear to have improved tremendously.

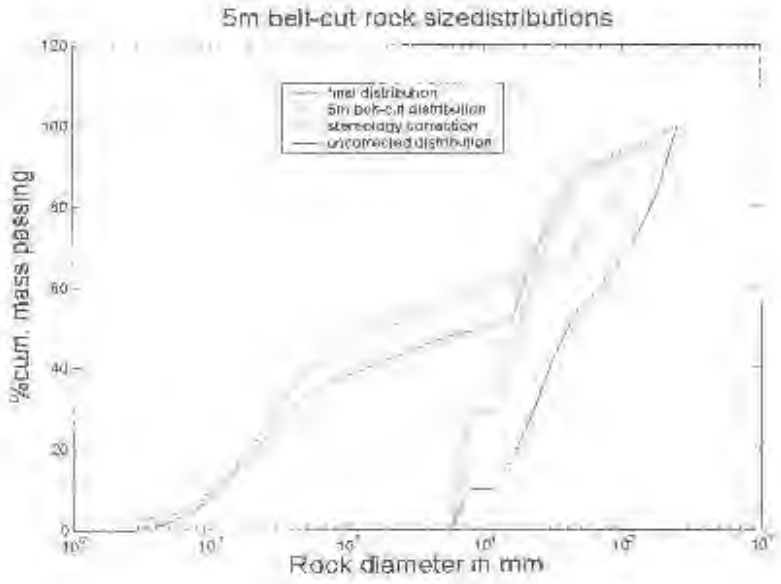


Figure 8.17: A visual comparison between the sieve and corrected machine size distributions on the 5m belt-cut data.

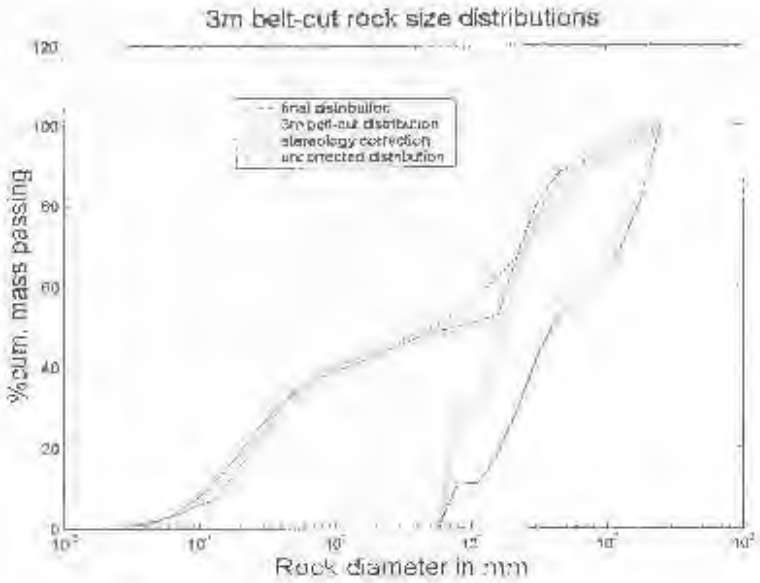


Figure 8.18: A visual comparison between the sieve and corrected machine size distributions on the 3m belt-cut data.

Treating the 3m belt-cut as the test-set, the next step is the quantification of the apparent improvement. The regression results are shown in table 8.4 and the obtained regression lines are shown in figure 8.19.

Table 8.4: Machine vs sieve distributions linear regression results on the 3m belt-cut data

Parameter	Machine Line	Sieve Line
b_0	0.3617	0.3601
95% confidence interval on b_0	(0.3366 to 0.3867)	(0.3414 to 0.3788)
b_1	0.1135	0.1164
95% confidence interval on b_1	(0.1049 to 0.1222)	(0.1099 to 0.1229)
R^2	96.66%	98.21%
Observed F-statistic	723.9	1368.5

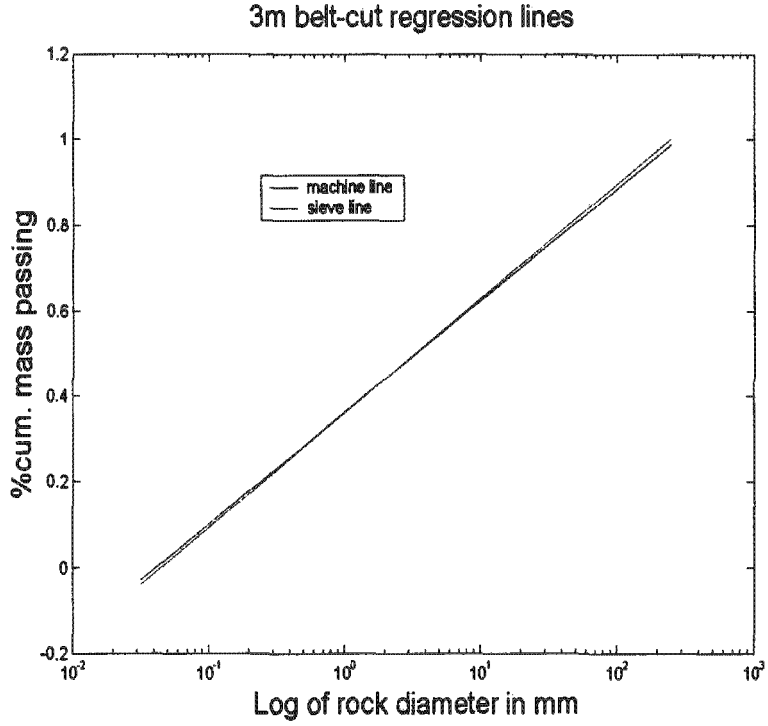


Figure 8.19: A visual comparison between the sieve and corrected machine regression lines.

Once again the linear fits are significant as shown in table 8.4. After carrying out the hypothesis tests of similarity between slopes and constants, the results show that at the 5% significance level the null hypothesis which states that the slopes are the same, is not rejected. Further tests on the similarity of constants reveal that at the 5% significance level, the null hypothesis which states that the constants are similar, is not rejected. Based on these results, it can be concluded that the lines are similar and therefore the sieve and machine measured rock-size distributions are similar on the 3m belt-cut test data.

Table 8.5: Hypothesis testing results for similarity between the machine and sieve measured size distributions.

Test	H_0	H_1	t-stat	significance level
slopes	$b1_{machine} = b1_{sieve}$	$b1_{machine} \neq b1_{sieve}$	0.7022	5%
constants	$b0_{machine} = b0_{sieve}$	$b0_{machine} \neq b0_{sieve}$	0.1411	5%

8.4 Summary

In this chapter, the major results of this dissertation are presented. In particular, the results of the comparison between the rock-size distributions of manually segmented test-set images and those measured by the system are presented. Secondly, the system's performance is evaluated on a conveyor-belt video sequence with corresponding sieved data.

The visual results of the comparison between the system output and the manual measurements reveals that the two distributions are similar with an RMS error of 2.37%. However statistical tests do not strongly agree with this conclusion. These tests highly depend on the availability of data. Based on the fact that only a handful of images were used as the test-set, test-set could be extended for a more convincing conclusion. Another source of error is the KNN classifier since is expected to have an error rate of approximately 10% with a true positive rate (TPR) of 63.04%. This means 36.96% of the actual rocks on the test-set are missed.

The final tests involved comparing the machine measured distributions of a belt video sequence to the corresponding sieve measured distributions. It was discovered that there are major differences between the two distributions as expected. The main differences are the apparent dead-band, size-distribution overestimation and underestimation. The major sources of error are the KNN, missing fines, "fragment lay", "overlapping fragments" and camera resolution.

The camera resolution is the primary contributor to the dead-band problem. The secondary contributor is the effect of missing fines. An s-curve from the 5m belt-cut is used to correct for the effect of missing fines and dead-band. The overestimation problem is due to "fragment lay" while underestimation is due to "overlapping fragments" [43]. These are corrected using an "unfolding" method for reconstructing 3D rock-size distributions from 2D distributions of rock-sections. After incorporating these correctional measures into the system, the performance evaluation of the updated system on the 3m belt-cut shows that the two distributions are similar.

Chapter 9

Conclusions and Future work

9.1 Conclusions

Based on the results of this work, the following conclusions can be drawn:

- A machine vision-based instrument for measuring the size distribution of rocks on a conveyor is developed. Its main components are a watershed-based segmentation algorithm, a rock feature classification method in the form of KNN, and a stereology-based correction scheme for unfolding a 3D rock-size distribution from a 2D size-distribution of rock sections.

- The visual results of watershed-based segmentation on a test-set of images taken under varying lighting conditions show that the algorithm is robust to these conditions. However, high false alarm rates due to lumps of fines appearing as rocks and imperfect lighting conditions were also obtained.

- Pattern recognition tools in the form of feature classification methods are investigated for rock recognition with the main objective of reducing false alarm rates. Overall, the investigated set of methods reduced the FAR significantly with moderate effect on the true detection rates (TDR). However the most suitable feature classification method in terms of a good compromise between FAR, TDR on the test-set, and high speed of computation is the simple KNN.

- The system is then tested for accuracy on a test-set of images by evaluating and comparing to size distributions of corresponding manually segmented images. Visual results show that the two distributions are similar with an RMS error of 2.37%. However statistical tests showed otherwise.

- The results of a comparison between the measured projected areas and the sieve size distribution on collected sieve data showed a significant difference between the two distributions. This difference is due to several sources of error, some of which can be compensated for. A combination of dead-band and stereology-based correction of the distributions was applied and tremendous reductions in the difference between the two distributions were achieved. A statistical decision test revealed that the two distributions are similar on a $3m$ belt-cut test data.

9.2 Recommendations for future developments

Based on the results and conclusions of this work, the following recommendations for future developments can be made:

- Operational lighting conditions are not intensively investigated in this work. A further investigation on the suitable types of lights as well the geometrical arrangements should be undertaken.

- A further investigation into the segmentation algorithm's parameter settings should be undertaken to optimize its performance. In particular, the parameter settings of the iterative bilateral filter should be given special attention.

- As far as rock feature classification for rock recognition is concerned, the simple KNN should be used since it performs reasonably well and has a high speed of computation.

- The final rock-size distributions are not as smooth as the sieve rock size distributions, smoothness can be enforced by using the modified version of the sterelogical transformation (see appendix B). Further investigations on stereological corrections should be undertaken.

- The system cannot run in real-time on a mineral processing plant as the implementation is not optimized. A more robust implementation of the system should be undertaken so that more tests can be performed on vast amounts of data.

- A Full 3D surface reconstruction of the rock scene and the subsequent image analysis thereof can improve the results achieved since accurate quantities such as rock volume can be computed. A feasibility study of 3D surface reconstruction for rock scenes should be undertaken.

Bibliography

[1] D. H. Ballard and C. M. Brown. *Computer Vision.* Prentice-Hall, inc., Englewood Cliffs, New Jersey 07632, Department of Computer science, University of Rochester, Rochester, New York, 1982.

[2] S. Baluja. *Genetic Algorithms and explicit search statistics*, pages 319–325. MIT Press, Cambridge, M.A., 1997.

[3] S. Baluja and R. Caruana. Removing the genetics from the standard genetic algorithm. *The International Conference on Machine Learning*, pages 38–46, 1995.

[4] Gunther Franz-Martin Berger. Software for a particle size analyser based on image analysis techniques. Master's thesis, University of Witwatersrand, 1985.

[5] S. Beucher and C. Lantuejoul. Use of watersheds in contour detection. *International workshop on image processing: Real-time edge and motion detection or estimation*, 1979.

[6] S. Beucher and F. Meyer. The morphological approach to segmentation: the watershed transformation. In E. Dougherty, editor, *Mathematical Morphology in Image Processing*, pages 433–481. Marcel Dekker, New York, 1992.

[7] M.J. Black and G. Sapiro. Edges as outliers: Anisotropic smoothing using local image statistics. Technical report, University of Minnesota, Department of Electrical and Computer Engineering, University of Minnesota, 200 Union street SE, Minmeapolis, 1998.

[8] S. Bow. *Pattern Recognition and Image Preprocessing.* Marcel Dekker, Inc., Department of Electrical Engineering, Northern Illinois University, Dekab, Illinois, 1992.

[9] L. Costa and R.M. Cesar Jr. *Shape analysis and classification: Theory and Practice.* CRC Press, 2001.

[10] L.F. Costa and R.M. Cesar Jr. *Shape Analysis and Classification Theory and Practice.* Image processing series. CRC Press, Pennsylvania Institute of Technology, 2001.

[11] R.C. Crida. *A Machine Vision Approach to rock fragmentation analysis.* PhD thesis, University of Cape Town, Department of Electrical Engineering, University of Cape Town, Cape Town, September 1995.

[12] R.C. Crida and G. De Jager. Rock recognition using feature classification.

[13] E.R. Davies. *Machine Vision: theory, Algorithms, Practicalities.* Academic Press, Department of Physics, Royal Holloway, University of London, Egham, Surrey, UK, second edition, 1997.

[14] O. Eriksen. In-stream measurements of grain size distributions by random intercepts. In *International Symposium on In-stream measurements of particulate solid properties*, pages 1–17, Berge, Norway, 1978.

[15] T. Fawcett. Roc graphs: Notes and practical considerations for data mining researchers. Technical report, Intelligent Enterprise Technologies Laboratory, January 2003.

[16] Xin Zhou Feng Lu 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.

[17] D. Forsyth and A. Zisserman. Mutual information. *IEEE Transactions on Pattern analysis and Machine Intelligence*, pages 466–473, 1989.

[18] D.A. Forsyth and J. Ponce. *Computer Vision- A modern approach.* Prentice-Hall, 2003.

[19] D.A. Forsyth and A. Zisserman. Reflections on shading. *IEEE Transactions on Pattern analysis and Machine Intelligence*, 13(7):671–679, July 1991.

[20] J.J. Francis. *Machine vision for froth flotation.* PhD thesis, University of Cape Town, Department of Electrical Engineering, Faculty of Engineering and Built Environment, University of Cape Town, Rondebosch 7700, South Africa, June 2001.

[21] J Liu G. Pok and A.S. Nair. Selective removal of impulse noise based on homogeneity level information. *IEEE Tranasactions on image processing*, 12(1):85–92, January 2003.

[22] S.G. Grannes. Determining the size distribution of moving pellets by computer image processing. In R.V. Ramani, editor, *19th Application of Computers and Operations Research in the Mineral Industry*, pages 545–552, 1986.

[23] J.R. Greene. Feature subset selection using thormton's separability index and its applicability to a number of sparse proximity-based classifiers. *Proceedings of the 12th Annual Symposium of the Pattern Recognition Association of South Africa*, 2001.

[24] J.R. Greene. Neural, fuzzy and evolving systems lecture notes of the course eee496s, 2003.

[25] J. Haddon and D. Forsyth. Shading primitives: Finding folds and shallow grooves. Technical report, University of California, Computer Science Division, University of California, Berkely CA 94720, 1998.

[26] R.M. Haralik and L.G. Shapiro. *Computer and Robot Vision*, volume 1. Addison-Wesley Publishing Company, 1992.

[27] R.M. Haralik and L.G. Shapiro. *Computer and Robot Vision*, volume 2. Addison-Wesley Publishing Company, 1993.

[28] A. Hummel. Representations based on zero crossing in scale-space. *Proc. IEEE Computer Vision and Pattern Recognition Conf.*, pages 204–209, June 1986.

[29] R. Hyndman. The problem with sturges' rule for constructing histogram. Technical report, Monash University, 1995.

[30] Intel Corporation, U.S.A. *Open Source Computer Vision Library, Reference manual*, 2001.

[31] B. Jahne. *Digital Image Processing Concepts, Algorithms and Scientific Applications*. Springer-Verlag, Scripps Institution of Oceanography, University of Carlifornia, San Diego, La Jolla, CA 92093-0230, USA, second edition, 1993.

[32] I. Elber J.H. Vignos and E. Gallagher. Coarse particle size transmitter. In *Meeting of SME of AIM*, Tucson, Arizona, October 1976.

[33] R. Jhumka and D. J. Mashao. Comparing svm and gmm on speaker identification. *Proceedings of the 13th Annual Symposium of the Pattern Recognition Association of South Afrca*, 2002.

[34] A. Kanfer. An investigation into the design of a low cost machine vision system for monitoring material flow in the mineral processing industry. Master's thesis, University of Cape Town, Department of Electrical Engineering, University of Cape Town, December 1993.

[35] J. Koenderik. The structure of images. *Biol. Cybern*, 50:363–370, 1984.

[36] O. Carlsson L. Nyberg 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.

[37] G. Von Borries L. Yachner, C. Gonzales and R. Nobile. Coarse particle size distribution analyzer. In *IFAC Symposium on Automation for Mineral Resource Development*, Brisbane Queensland, Australia, July 1985.

[38] T.B. Lange. *Measurement of the size distribution of rocks on a conveyor belt using machine vision*. PhD thesis, University of Witwatersrand, Faculty of Engineering, University of Witwatersrand, Johannesburg, 1990.

[39] N.H. Maerz. Image sampling techniques and requirements for automated image analysis of rock fragmentation. *Proceedings of the FRAGBLAST 5, Workshop on Measurement of blast fragmentation*, pages 115–120, August 1996.

[40] N.H. Maerz. Reconstructing 3-d block size distributions from 2-d measurements on sections. *Proceedings of the FRAGBLAST 5, Workshop on Measurement of Blast Fragmentation*, pages 39–43, August 1996.

[41] N.H. Maerz. Aggregate sizing and shape determination using digital image processing. *Center for Aggregates Research (ICAR) Sixth Annual Symposium Proceedings*, pages 195–203, 1998.

[42] N.H. Maerz and T.C. Palangio. Wipfrag system ii- online fragmentation analysis. *FRAGBLAST 6, Sixth Internation Symposium for rock fragmentation by blasting*, pages 111–115, August 1999.

[43] N.H. Maerz and W. Zhou. Optical digital fragmentation measuring systems-inherent sources of error. *FRAGBLAST, The international journal for blasting and fragmentation*, 2(4):415–431, 1998.

[44] N.H. Maerz and W. Zhou. Calibration of optical digital fragmentation measuring systems. *FRAGBLAST, The international journal for blasting and rock fragmentation*, 4(2):126–138, 2000.

[45] D.J. Mashao. Comparing svm and gmm on parametric feature-sets. *Proceedings of the 14th Annual Symposium of the Pattern Recognition Association of South Africa*, 2003.

[46] C. McDermott and N.J. Miles. The measurement of rock fragmentation using image analysis. *Departmental magazine, Department of Mining Engineering*, pages 49–61, 1988.

[47] D.H. Marimont M.J. Black, G. Sapiro and D. Heeger. Robust anisotropic diffusion. *IEEE Transactions on Image Processing*, 7(3):421–232, March 1998.

[48] S.G. Mkwelo. A comparative evaluation of evolutionary design methods in engineering. A 4th year undergraduate thesis in the department of electrical engineering, University of Cape Town, October 2001.

[49] T.J. Napier-Munn. *An introduction to comparative statistics and experimental design for Minerals Engineers*. Julius Kruttschnitt mineral research centre, The University of Queensland, 2 edition, 2001.

[50] T.C. Palangio N.H. Maerz and J.A. Franklin. Wipfrag image based granulometry system. *Proceedings of the FRAGBLAST 5, Workshop on measurement of Blast Fragmentation*, pages 91–99, August 1996.

[51] T.C. Palangio and N.H. Maerz. Case studies using the wipfrag image analysis system. *FRAGBLAST 6, Sixth Symposium for rock fragmentation by blasting*, pages 117–120, August 1999.

[52] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629 – 639, July 1990.

[53] T. Poggio and S. Smale. The mathematics of learning: Dealing with data. *Notices of the AMS*, 50(5), 2003.

[54] C. Pretorious and A.L. Nel. Rock size mintoring on conveyor belt systemsusing neural networks. *Proceedings of the second South African Workshop on Pattern Recognition*, pages 100–110, November 1991.

[55] W. Hue Q. Song and W. Xie. Robust support vector machine with bullet hole image classification. *IEEE Transactions on Systems, Man, and Cybernetics-Part C*, 32(4):440–448, November 2002.

[56] J.B.T.M. Roerdink and A. Meijster. The watershed transform: Definitions, algorithms and parallelization strategies. *Fundamenta Informaticae*, pages 187–228, 2000.

[57] G. de Jager S. Mkwelo and F. Nicolls. Watershed-based segmentation of rock-scenes and proximity-based classification of watershed regions under uncontrolled lighting conditions. *Proceedings of the 14th Annual Symposium of the Pattern Recognition Association of South Afrca*, 2003.

[58] J. Serra. *Image analysis and mathematical morphology*, volume 1. Academic Press, 1982.

[59] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. *Proceedings of the International Conference on Computer Vision*, 1998.

[60] R. van den Boomgard and J. van den Weijer. On the equivalence of local-mode finding, robust estimation and mean-shift analysis as used in early vision tasks. *Proceedings of the 16th International Conference on Pattern Recognition*, 3:927–930, 2002.

[61] L. Vincent and P. Soille. Watersheds in digital spaces: an effiecient algorithm based on immersion simulations. *PAMI*, 13(6):583–598, 1991.

[62] The Webmaster. The split-engineering website, February 2004. www.spliteng.com.

[63] The Webmaster. The wipfrag website, March 2004. www.wipware.com.

[64] David Wigeson. Fragmentation analysis using computer vision techniques. Master's thesis, University of Witwatersrand, 1987.

[65] B.A. Willis. *Mineral Processing technology*. Butterworth Heinenmann publishing (Ltd), 6 edition, 1997.

[66] A. Witkin. Scale-space filtering. *Int. Joint Conf. Artificial Intelligence*, pages 1019–1021, 1983.

[67] B.A Wright. The development of a vision based flotation froth analysis system. Master's thesis, University of Cape Town, Faculty of Engineering and Built Environment, University of Cape Town, Rondebosch 7700, South Africa, 1999.

Appendix A

Recognition results on experiment images

The following are visual rock recognition results on a subset of images used for testing the segmentation algorithm under varying lighting conditions. A KNN with a k-value of 13 is used for recognition.

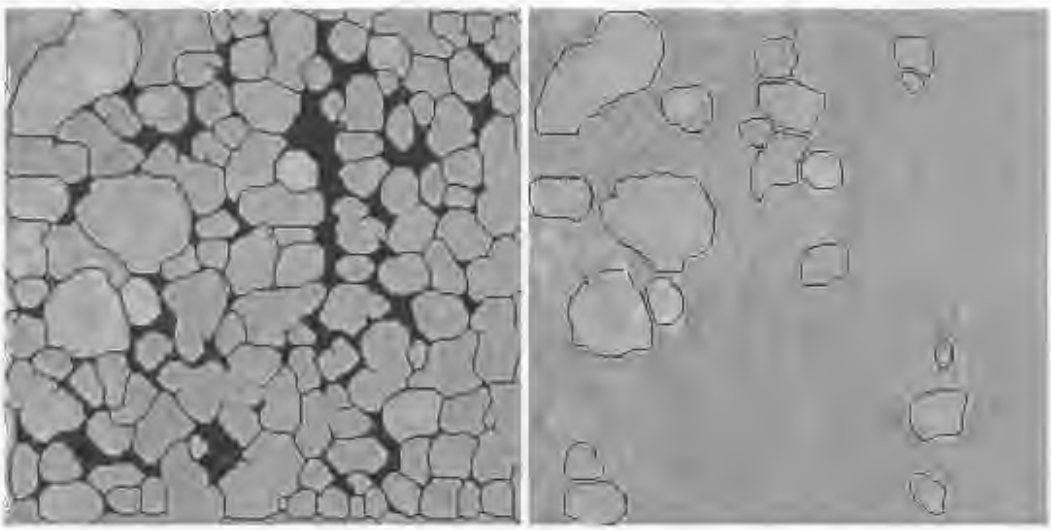


Figure A.1: Segmentation output of test-image5 and the corresponding rock recognition output.

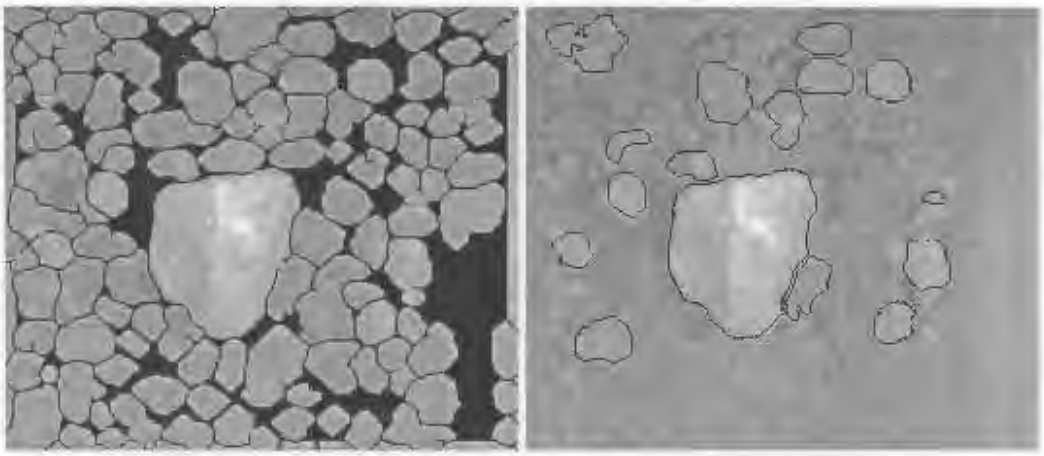


Figure A.2: Segmentation output of test-image6 and the corresponding rock recognition output.

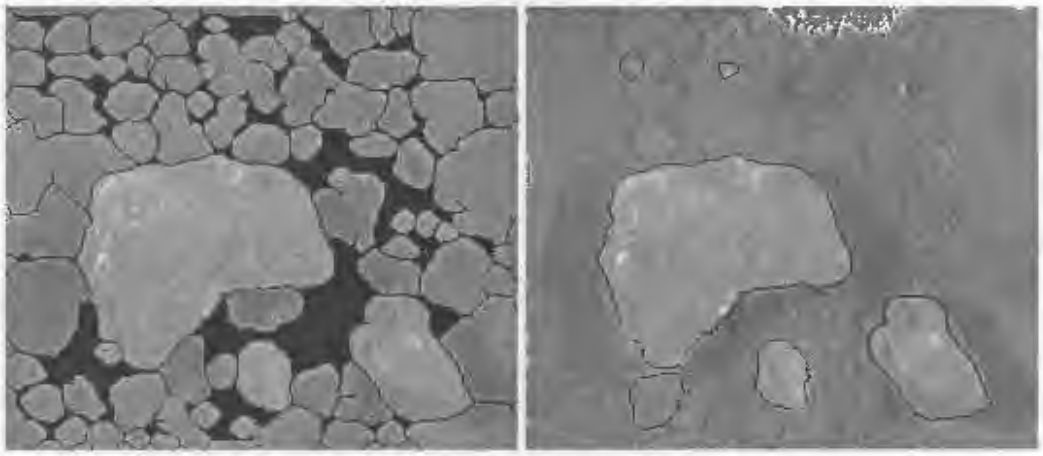


Figure A.3: Segmentation output of test-image1 and the corresponding rock recognition output.

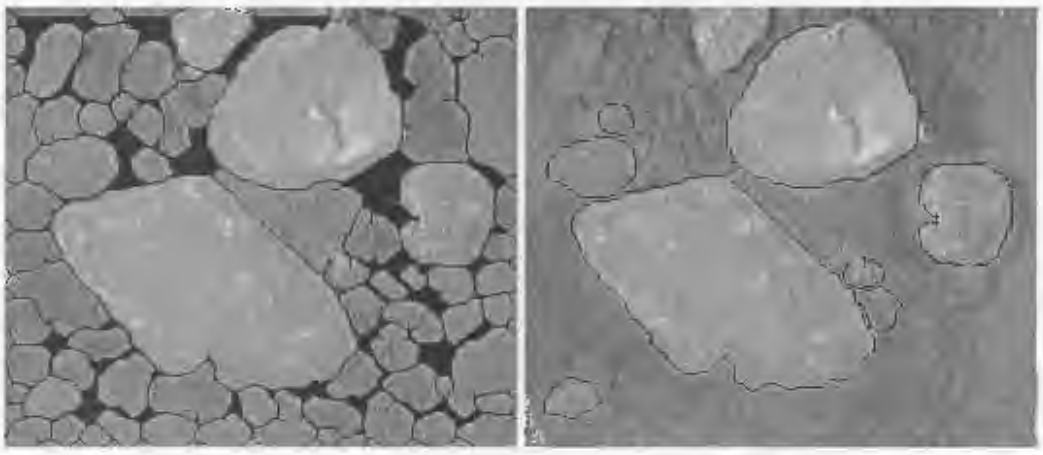


Figure A.4: Segmentation output of test-image4 and the corresponding rock recognition output.

Appendix B

Results of the calibrated stereology method

The calibrated stereology-based correction is applied and the obtained visual results are shown in figures B.1, B.2, B.3. The distributions are smoother than the distributions of the uncalibrated method, however there is a larger dissimilarity in the obtained regression lines. As a result, the statistical tests reveal that the two distributions are dissimilar, as shown in table B.2. Table B.1 shows that the linear fits are significant.

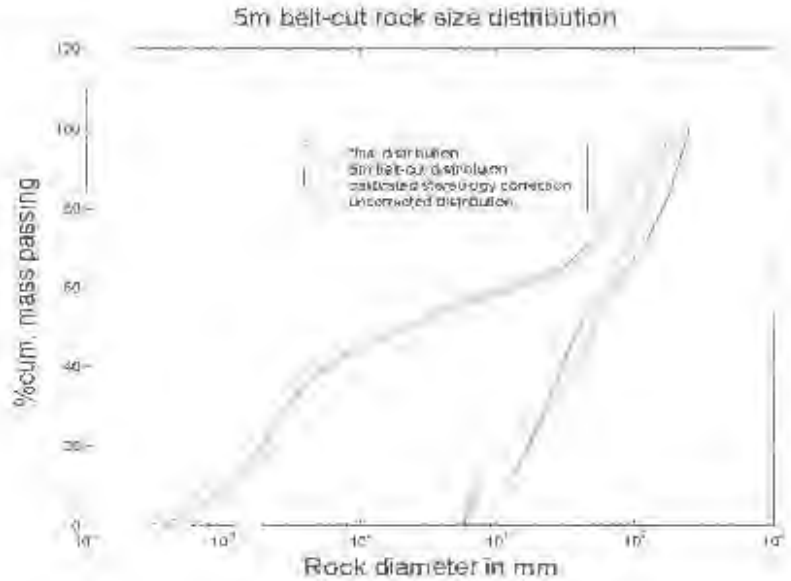


Figure B.1: The 5m belt-cut distribution after a calibrated stereology correction

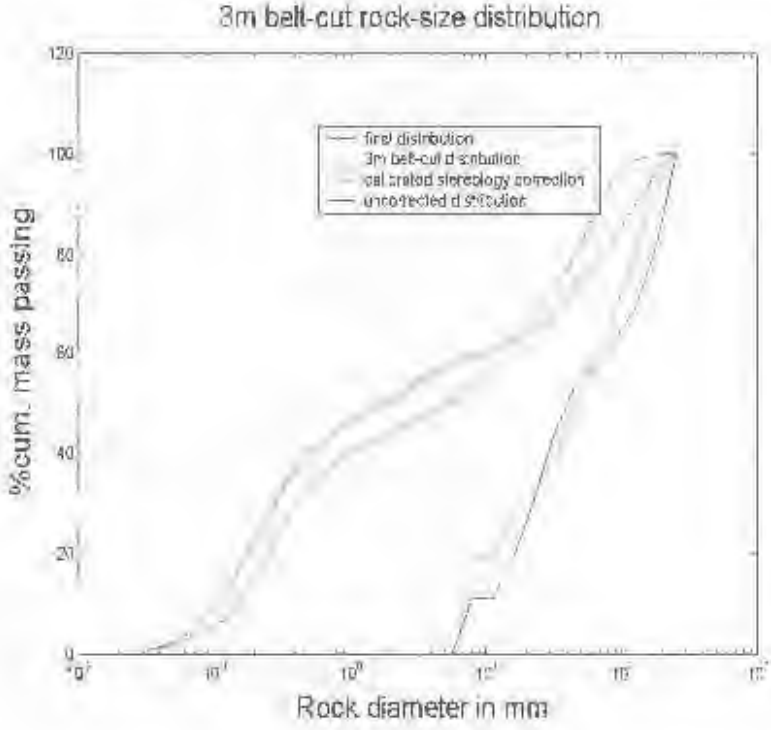


Figure B.2: The 3m belt-cut distribution after a calibrated stereology correction

Table B.1: Machine vs sieve distributions linear regression results on the 3m belt-cut data

Parameter	Machine Line	Sieve Line
b_0	0.3898	0.3601
95% confidence interval on b_0	(0.3679 to 0.4117)	(0.3414 to 0.3788)
b_1	0.1031	0.1164
95% confidence interval on b_1	(0.0955 to 0.1107)	(0.1099 to 0.1229)
R^2	96.9%	98.21%
Observed F-statistic	781.6	1368.5

Table B.2: Hypothesis testing results for similarity between the machine and sieve measured size distributions.

Test	H_0	H_1	t-stat	rejection level
slopes	$b1_{machine} = b1_{sieve}$	$b1_{machine} \neq b1_{sieve}$	3.7495	5%
constants	$b0_{machine} = b0_{sieve}$	$b0_{machine} \neq b0_{sieve}$	3.0549	5%

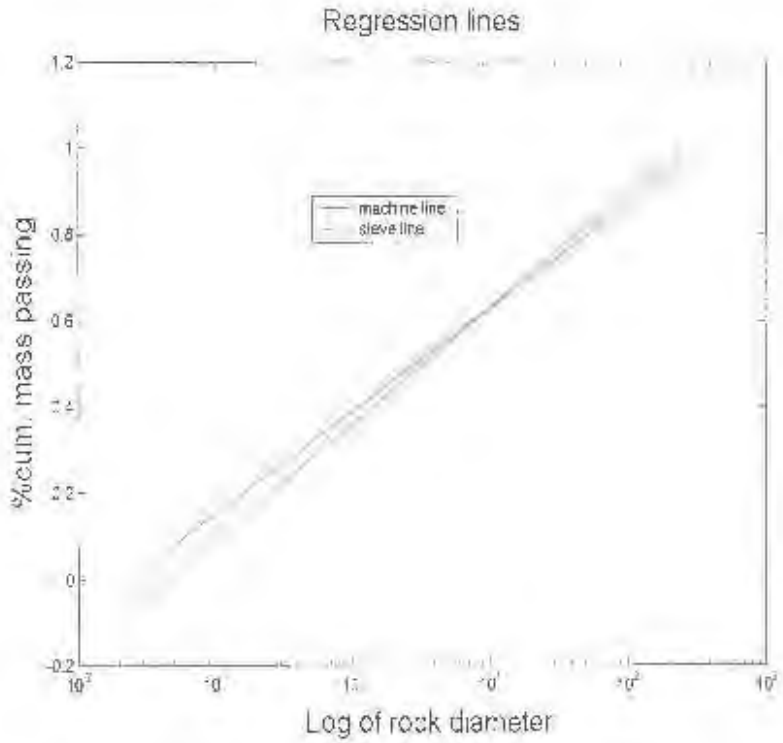


Figure B.3: The corresponding regression lines

Appendix C

The system's flow of processing

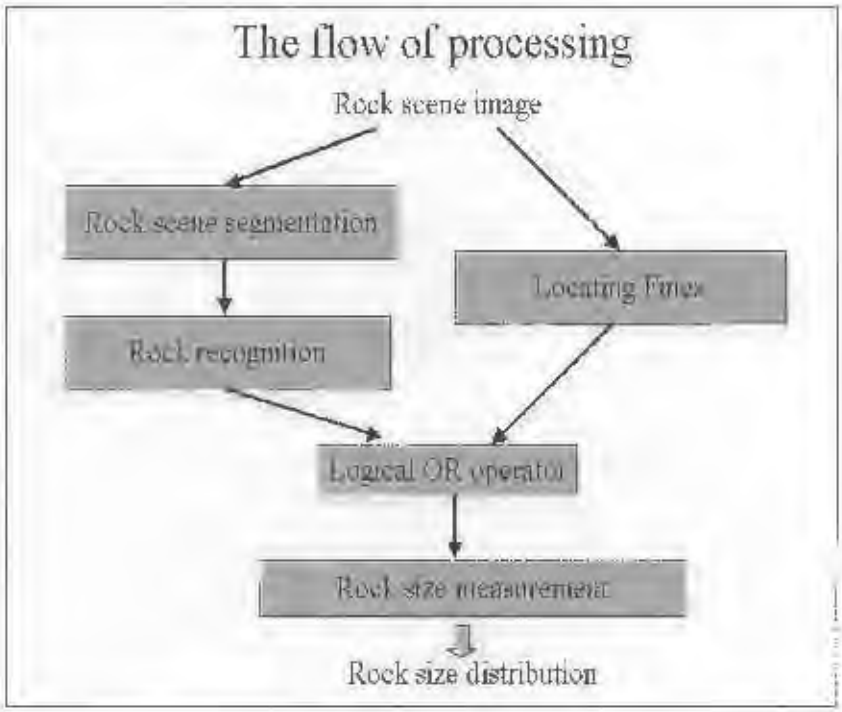


Figure C.1: The flow of processing diagram

Figure C.1 shows the flow of processing diagram, where a greyscale image of the rock scene is the input to the system. This is fed directly into the rock scene segmentation process to locate the boundaries of rocks in order to facilitate the measurement of rock size. However, some of the resultant boundaries are spurious and must therefore be removed. This is performed via rock recognition by using pattern recognition methods or classifiers. However, it was later

discovered that fines or smaller particles are missed after these processes are performed. A fines locator algorithm was then incorporated to solve this problem. The classifier and the fines locator outputs are then combined using a logical OR operator and the particles on this final image are measured for the final estimation of rock size distribution.