TEXTURE ANALYSIS
OF
BONE MINERALISATION SURFACES

by

CAROL ANNE REID

A dissertation submitted to the
UNIVERSITY OF GLASGOW
for the degree of
DOCTOR OF PHILOSOPHY

Department of Statistics
August 1995

Acknowledgements

I wish to thank my supervisors Hugh Elder and the late Victor Moss for their help with the physiological aspects of the project, and Jim Kay, for his guidance on the statistical side and for his constant encouragement and enthusiasm.

I am grateful to the SERC and Joyce Loebl for financial support in the first two years of research.

I am indebted to all my colleagues and friends at the University of Paisley and the University of Glasgow for their support and friendship over the years.

My thanks are also due to my family and friends for their patience, understanding and moral support while I completed this research.

Finally, special thanks are due to my parents for their continual encouragement and financial support throughout my many years of study.

Summary

A better understanding of the process of bone mineralisation is needed before effective treatments for bone diseases such as osteoporosis can be found. Bone growth is essentially a surface process, and because of this, scanning electron microscopy has proved of great value in imaging large areas of growth surface. Previous studies have identified six mineral surfaces, each associated with a known metabolic growth state, and each characterised by a particular surface texture. It is desirable to have quick and accurate methods for recognising these surface types and evaluating them quantitatively.

The objective of this study was to develop algorithms to segment and classify, automatically, the surface textures found in scanning electron microscopy images of rats' bones. This thesis describes the techniques used, their success and their limitations.

Chapter 1 is a brief introduction to the physiology of bone and the use of the scanning electron microscope to study bone surfaces. It explains the background to the study and describes the texture types identified in bone surfaces. It also describes the preparation of specimens for the scanning electron microscope, how the images are captured by the SEM and the transfer of the images to disk for use on a Sun SPARCstation.

Chapter 2 introduces the idea of texture, and reviews methods previously used to classify texture. The techniques are divided into 3 groups - statistical, structural and modelling. The statistical techniques reviewed include variations based on co-occurrence matrices, grey-level difference statistics and run-length matrices. Structural methods mentioned include a linguistic approach to extracting primitives, an approach based on grey-level thresholding and an edge-based approach. Modelling techniques include random mosaic models, Markov and Gibbs random field models, and random walk models. Each method is briefly described and results, where relevant, are quoted and discussed.

In chapter 3 the classification techniques described in chapter 2 are tested on the scanning electron microscopy images of rats' bones. To accelerate the investigation

64x64 windows are extracted from the larger 512x512 images. Image standardisation techniques are also discussed and investigated. None of the statistical classification methods works as well as stated in published results, where, on the whole, images from Brodatz' (1966) texture album were used. The best classification technique for the images in this study is a maximum likelihood classifier based on the co-occurrence matrix without feature extraction. High rates of classification (88%-100%) are obtained on training sets, although results are poorer when applied to test sets.

Chapter 4 deals with the problem of segmenting the images. The idea of edge-detection is introduced and texture segmentation techniques are described and investigated. Initially the results in this study are worse than those in published literature but by combining and modifying some of the techniques the segmentation is greatly improved. The chapter also examines automatically thresholding the images to highlight edges. Results of the fully automatic segmentation process are shown.

Chapter 5 brings together the work of chapters 3 and 4. It deals with the larger 256x256 images, the drawing of boundaries between texture types using the methods of chapter 4 and subsequent classification of each region into one of 5 types. This chapter also discusses the detection of vascular channels and bone cells and illustrates some of the difficulties in segmenting and classifying the images. Examples are given which show that, to a certain extent, the objective has been achieved. Boundaries can be found between formative and resorptive texture types and around vascular channels. Windows can then be classified, with reasonable accuracy, into one of the 5 texture types.

Chapter 6 discusses the limitations of the methods used and the problems of image capture. It also discusses possible ways of improving the results. These include treating the texture types as a continuous scale rather than 5 (or 6) discrete types, and using back-scattered electron images instead of secondary electron images. The latter would result in very different images and may require a completely new set of techniques to the ones used here. This chapter also summarises several other techniques which have been used by authors for either texture classification or segmentation, and could be tested on the images in this study. These include relative extrema techniques, fractal-based approaches, and the use of neural networks.

Contents Page

CHAPTER 1 A Physiological Background 1

1.1 The Structure of Bone 1
1.2 Bone Growth and Bone Resorption 3
1.3 The Study of Bone Surfaces 4
1.3.1 The Scanning Electron Microscope 4
1.3.2 Methods 8
1.3.3 Classification of Bone Surfaces 9
1.4 Current Objectives 15
1.5 Specimen Preparation 15

CHAPTER 2 Texture Classification : A Review 20

2.1 Texture 20
2.2 Previous Approaches to Texture Classification 21
2.3 Statistical Approaches 22
2.3.1 Grey-Level Co-occurrence Matrices 23
2.3.2 Grey-Level Difference Statistics 36
2.3.3 Run-Length Statistics 39
2.3.4 Fourier Power Spectrum 43
2.3.5 Neighbouring Grey-Level Dependence Matrices 48
2.3.6 Relative Extrema Measures 51
2.3.7 Filter Masks 52
2.4 Structural Approaches 54
2.4.1 Pure Structural Approaches 54
2.4.2 Structural-Statistical Approaches 55
2.4.3 Conclusions 57
2.5 Modelling Approaches 57
2.5.1 Gibbs and Markov Random Field models 58

2.5.2 Random Mosaic Models 66
2.5.3 Random Walk Models 68
2.6 Conclusions 71

CHAPTER 3 Classification of Texture in Bone 76

3.1 Subjective Impressions 76
3.2 Classification Algorithm 78
3.3 Co-occurrence Matrices 82
3.4 Difference Statistics 85
3.5 Fourier Power Spectrum 89
3.6 Run-Length Statistics 91
3.7 Image Standardisation 93
3.7.1 Histogram Equalisation 95
3.7.2 Histogram Normalisation 98
3.7.3 Conclusions 102
3.8 Training and Test Data 102
3.9 Co-occurrence Matrices and Difference Statistics Revisited 106
3.10 Maximum Likelihood 122
3.11 Conclusions 132

CHAPTER 4 Segmentation 137

4.1 Introduction 137
4.2 What is an Edge? 137
4.3 Texture Edge-Detection 139
4.3.1 Notes on Rosenfeld & Thurston (1971) algorithm 145
4.4 Application of Rosenfeld & Thurston (1971) algorithm 147
4.5 Smoothing 149
4.6 Relaxation Labelling 155

4.7 Raster Tracking 163
4.8 Thresholding 168
4.9 Conclusions 174

CHAPTER 5 Combining Segmentation and Classification 177

5.1 Introduction 177
5.2 Haversian Channels 177
5.3 Bone Cells 180
5.4 The Synthesis 191
5.5 Conclusions 200

CHAPTER 6 Discussion 202

6.1 Success 202
6.2 Limitations 203
6.3 Further Work 205
6.3.1 The Quad-Tree Approach 206
6.3.2 Relative Extrema 206
6.3.3 Markov and Gibbs Random Fields 207
6.3.4 Fractals 208
6.3.5 Spatial Filtering 210
6.3.6 Neural Networks 210
6.3.7 Continuous Texture Types 211
6.3.8 Back-Scattered Electron Images 212
6.4 Conclusions 213

Appendix 214

References 219

Chapter 1 A Physiological Background

1.1 The Structure of Bone

Bone is a form of connective tissue which forms most of the skeleton of higher vertebrates. It consists of two main components - 1) bone cells and 2) an intercellular matrix.

The matrix contains an organic component and an inorganic component. The organic component is mainly collagen fibres which contribute greatly to the strength and resilience of the bone, whereas the inorganic component, which accounts for approximately two-thirds of the weight of the bone, gives it its hardness and rigidity.

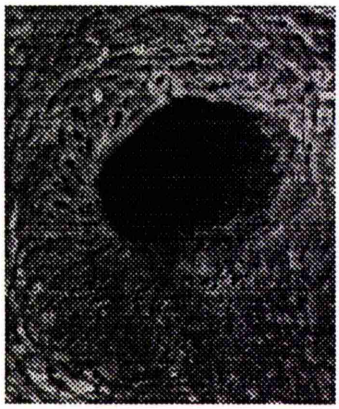
Examination by eye reveals two types of bone - spongy bone and compact bone. Spongy bone consists of thin, irregular branches or trabeculae which connect with one another to form a mesh, the spaces of which are filled with bone marrow. Compact bone appears solid except for microscopic spaces. Both spongy and compact types are present in almost every bone, although the amount and distribution of each vary considerably. In long bones such as the tibia the shaft is mainly compact bone surrounding a bone marrow cavity. Each end consists of spongy bone covered by a thin shell of compact bone. In flat bones, e.g. the skull, two plates of compact bone enclose a middle layer of spongy bone.

Each bone is enveloped by a specialised tissue coat called the periosteum. A similar, but less well-developed connective tissue layer called the endosteum lines the marrow cavity and marrow spaces.

Microscopically, the most characteristic feature of mature bone is its lamellar structure. The calcified intercellular substance, or bone matrix, is organised into lamellae or layers arranged in various ways. Within the intercellular substance there are small cavities called lacunae which contain the bone cells. Radiating from each lacuna are numerous narrow channels, termed canaliculi, which penetrate adjacent lamellae to join with canaliculi of neighbouring lacunae. Thus all lacunae are connected by a system of these narrow channels. In addition, compact bone is traversed by vascular channels called Haversian channels. These contain the blood vessels and nerves of the bone and provide a means of transporting nourishment to the bone cells. In cross-

section, Haversian channels appear as round openings surrounded by ring-shaped concentric lamellae. (Fig. 1.1)

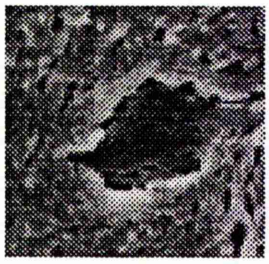
Fig. 1.1
A scanning electron microscope image of bone showing a Haversian channel.



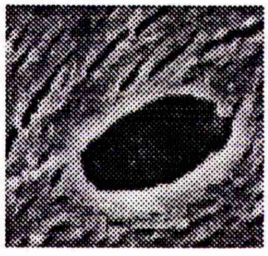
There are three cell types peculiar to bone - osteoblasts, osteocytes and osteoclasts (Fig. 1.2).

Osteoblasts are associated with bone formation. They vary in shape, have a large nucleus and several canaliculi extending from them. Osteoblasts secrete the intercellular substance of bone both around their cell bodies and around the canaliculi. When mature, osteoblasts are termed osteocytes.

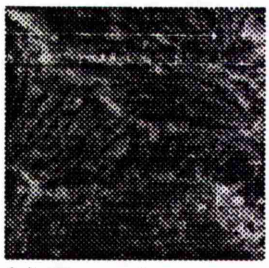
Osteoclasts, the third major type of bone cell, contrast with the other two types in being very large with multiple nuclei and in being involved in bone resorption rather than bone production. At one time they were thought to have formed from the fusion of osteocytes but recent research has suggested this is not the case. Osteoclasts are now believed to be derived from separate cell lines from osteoblasts and these fuse together to form the multinucleate osteoclasts. Osteoclasts are however known to be associated with bone resorption and are often found in shallow excavations known as Howship's lacunae.



(a) Osteoblast



(b) Osteocyte



(c) Osteoclast

Fig. 1.2 The bone cells.

1.2 Bone Growth and Bone Resorption

Bone is a living tissue and its composition and structure are constantly changing. These changes are the result of two processes - the formation of new bone and the resorption of older bone.

In a normal mammal, the formative and resorptive processes are finely balanced to maintain a constant bone mass in the adult or, in the child, a controlled rate of bone growth. Thus as new bone is added to the outside of a shaft of long bone, at the same time, older bone must be resorbed from the inside of the shaft. If for some reason the two processes of growth and resorption get out of balance with one another the bone becomes abnormal. One example of this is the fairly well-known disorder, common in post-menopausal women, called osteoporosis. Here, bone resorption is greater than bone formation resulting in "thinning of the bones".

1.3 The Study of Bone Surfaces

Both bone growth and bone resorption occur on the surface of bone so that all alterations in the shape of bones that occur through their growth and development are the result of bone being added to surfaces and resorbed from surfaces. Examination of bone surfaces can therefore provide valuable information about the bone such as its age and whether there is any history of bone disease. For this reason several studies have involved examining various bone surfaces. Many of these studies have made use of an instrument known as a scanning electron microscope.

1.3.1 The Scanning Electron Microscope

Electron microscopy has grown from man's efforts to see, more clearly and in increasing detail, the material world in which he lives. Regardless of how good a light (or optical) microscope is, the resolution, which is defined as the closest spacing of two points which can clearly be seen through the microscope, is limited. As a result, the light microscope has, in recent years, been complemented by its electron-optical counterparts, such as the scanning electron microscope (SEM), the transmission electron microscope (TEM) and more recently the transmission scanning electron microscope (STEM). All three electron microscopes project a beam of electrons (as opposed to a beam of light in a light microscope) at the specimen. At the simplest level, a scanning electron microscope provides images of external morphology, similar in appearance to those formed by eye, whereas a transmission electron microscope probes the internal structure of solids and provides microstructural detail not familiar to the human eye. A scanning transmission electron microscope combines some of the advantages of the SEM and the TEM.

One of the main differences between electron optics and light optics is that electrons are very much more strongly scattered by gases than light is. This means that in order to use electrons in a microscope all the optical paths must be evacuated to a pressure of about 10^{-7} of atmospheric pressure. Another main difference is that electrons carry a charge. This opens up the possibility of easily scanning a beam of electrons back and forth, as happens in a cathode-ray tube or a television tube. The application

of this approach has led to the development of the scanning electron microscope which has revolutionised attitudes to the study of surfaces.

Figure 1.3 shows the main components of a simple scanning electron microscope.

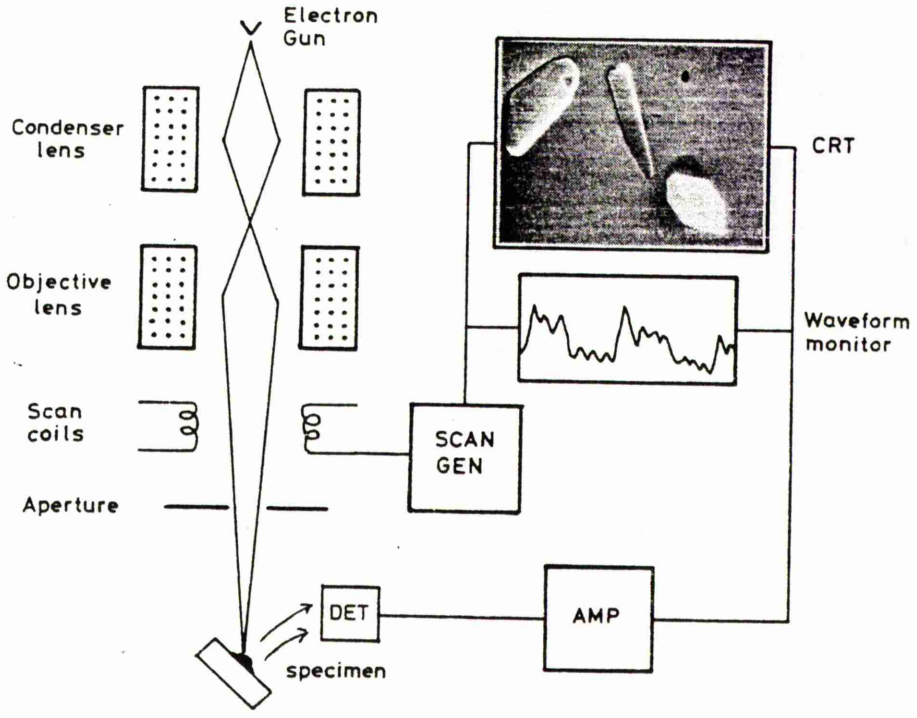


Fig. 1.3 The Scanning Electron Microscope

An electron gun, which is usually a heated tungsten hairpin filament, emits an abundance of (primary) electrons which generate an electron beam. A series of lenses then demagnifies this electron beam until it hits the specimen, which is held in the specimen chamber. The fine beam of electrons is then scanned across the specimen by the scan coils while a detector records the number of low energy secondary electrons, or other radiation, given off from each point on the surface. At the same time, the spot of the cathode ray tube (CRT) is scanned across the screen, while the brightness of the spot is modulated by the amplified current from the detector. The electron beam and the CRT spot are both scanned in a similar way to a television receiver, i.e. in a rectangular set of lines known as a raster. By suitable choice of the number of turns of the scan coils or by moving the coils which deflect the electron beam in the microscope, the size of the raster scanned in the specimen can be arranged to be much smaller than

on the surface of the cathode ray tube. The magnification is the ratio of the length of one side of the raster on the display cathode ray tube to the length of the corresponding side of the raster scanned by the electron probe on the specimen. Thus the final picture will be a magnified image of the object and within limits the magnification can be made as large as wished. In practice, the magnification depends on the resolution, and the fact that for small spot sizes, the number of electrons in the beam, and hence the signals derived from their interaction with the sample, become very low.

One of the main features of the SEM is that, in principle, any radiation from the specimen or any measurable change in the specimen may be used to provide the signal to modulate the CRT and thus provide contrast in the image. Each signal is the result of some particular interaction between the electrons and the specimen and may provide us with different information about the specimen. Figure 1.4 shows some of the signals which may be used in the SEM.

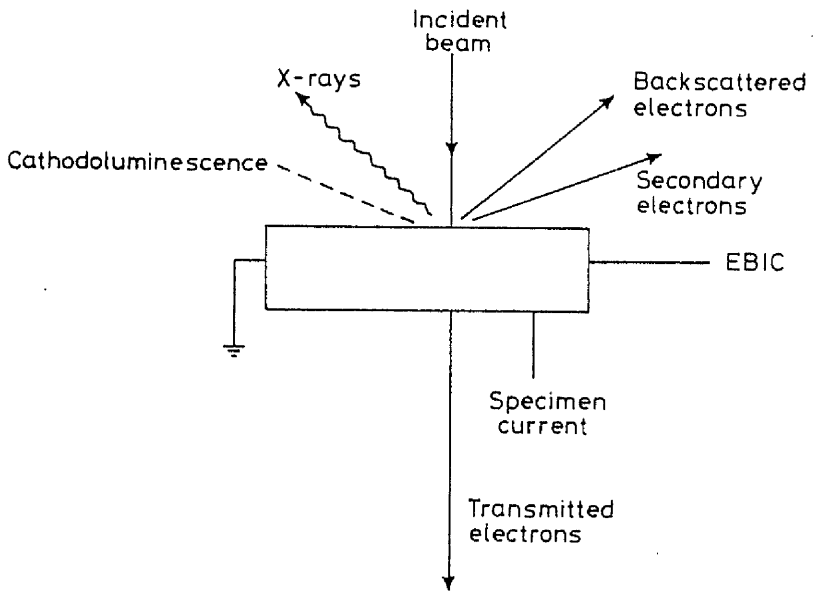


Fig. 1.4 Signals which may be used in the SEM

Secondary electrons are electrons which escape from the specimen with energies below about 50eV. The yield of secondary electrons, that is the number emitted per primary electron, can be as high as 1. Secondary electrons are therefore

abundant and are the most commonly used imaging signal in scanning electron microscopy. Secondary electrons are detected by a scintillator-photomultiplier system known as the Everhart-Thornley detector. Electrons are attracted from the specimen towards a metal cup with a metal grid over its opening. The energy of the secondary electrons is too low to excite the scintillator and so they are accelerated by applying a voltage to a thin aluminium film covering the scintillator. The light created in the scintillator then passes down a light pipe to a photomultiplier where it is converted to an electric current ready for subsequent amplification.

One of the principal uses of the SEM is to study the topography of a sample. Topographic images obtained with secondary electrons look remarkably like images of solid objects viewed by light and the specimen in the SEM appears as if it were being looked at from above.

The preparation of specimens for the SEM is far simpler than for other types of microscope where thin specimens are used. Since there are no lenses below the specimen in the SEM there is a great deal of space available to accommodate the specimen, and therefore fairly large specimens can be viewed. For effective viewing of the specimen in the SEM it is necessary for the surface of the specimen to be electrically conducting. This is because, during normal operation, a surplus of electrons builds up on the specimen surface. If these were not conducted away to earth, the specimen surface would become negatively charged and a distorted image would be formed. In biological materials it is common to coat the specimen with a thin conducting layer of gold or carbon using sputter coating.

In summary, scanning electron microscopy is preferable to other microscopy methods for various reasons. Probably the main advantage is the excellent resolution compared with optical microscopes. In addition, specimens can be examined over a wide range of magnifications and the SEM provides a greater depth of field than that provided by light microscopy methods. Also specimen preparation is easier because it is not necessary to prepare thin, flat sections, and the scanning electron microscope provides a clear view of much larger specimens than, for example, a transmission electron microscope.

1.3.2 Methods

Using a variety of methods of preparation, several different bone surfaces have been exposed and examined using the scanning electron microscope (Fig. 1.5).

There have been several studies in which the morphological appearance of the cellular layer lining bone surfaces has been investigated with the scanning electron microscope (Jones (1974), Jones, Boyde & Pawley (1975), Davis et al. (1975), Jones & Boyde (1976a & b), Jones et al. (1977)).

Other studies involve examining the surface of the bone matrix directly underlying the cells (Boyde & Hobdell (1969)).

A third method is to remove the organic component of the bone matrix to expose the mineral component for examination by the SEM. This method was first applied to mammalian bone and dental hard tissues in 1968 (Boyde (1968), Boyde, Jones & Hobdell (1968), Jones & Boyde (1968)). The organic component was dissolved using sodium hypochlorite which exposes the surface of the mineral component without affecting it to any great extent.

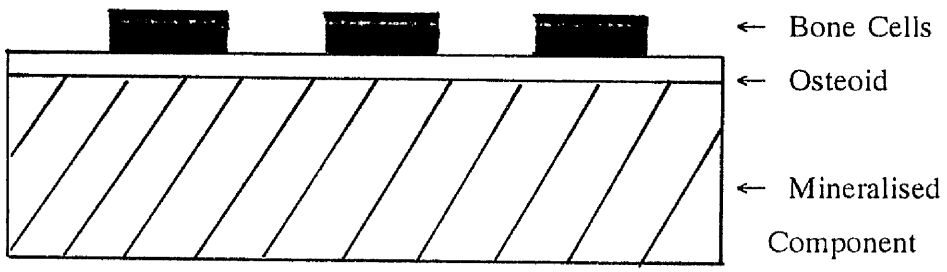


Fig. 1.5 Surfaces examined by SEM

Using this third method, in 1969 Boyde & Hobdell (1969a & b) studied the appearance of free surfaces of anorganic bone and demonstrated that, based on this appearance, it was possible to determine whether a surface was undergoing or had undergone formation or resorption or if it was in a resting phase of activity.

In 1970, again using anorganic preparations, Jones & Boyde investigated the

changes induced in the surface topography of rat bones by hormone administration. They found that they could now differentiate further the states of activity. Actively resorbing zones could now be distinguished from previously resorbed surfaces which were currently in a resting phase of activity and surfaces which had only recently entered a resting phase could be distinguished from prolonged resting surfaces.

1.3.3 Classification of Bone Surfaces

Six different types of anorganic surface could now be distinguished on the basis of their morphological appearance. These six surfaces represent different stages in the bone turnover cycle.

Type 1 : Surfaces with a granular appearance composed of small fusiform mineral particles. In some areas the particles display a random orientation, in others they are aligned in rows. This type of surface is taken to represent an early phase of mineralisation of the matrix (Figs. 1.6I, 1.7).

Type 2 : Surfaces consisting of rows of larger mineral nodules resembling pebbles or strings of beads. These areas are taken to represent a later stage in the calcification of the collagen matrix (Figs. 1.6II, 1.8).

Surface types 1 and 2 together represent incompletely mineralised matrix.

Type 3 : Smooth surfaces consisting of collagen fibre skeletons. These areas are taken to represent completely mineralised, resting bone matrix. The rims of forming osteocyte lacunae are smooth, representing a high level of mineralisation (Figs. 1.6III, 1.9).

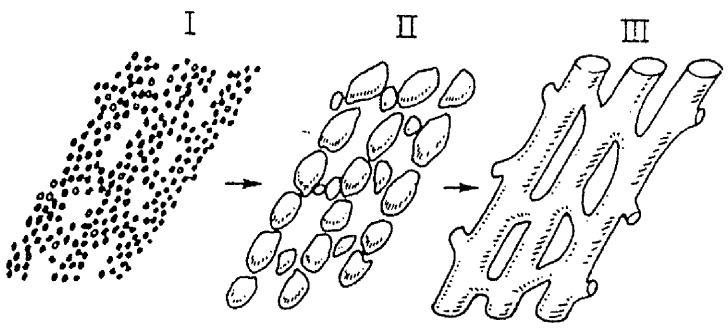


Fig. 1.6 Stages in the mineralisation sequence.

There are also surfaces displaying numerous depressions or pits with bright scalloped edges. These surfaces are always at a lower level than the adjacent bone matrix and their borders are marked by fairly well-defined edges. These surfaces correspond to bone which has undergone osteoclastic resorption. Two types of resorbed surface can be distinguished.

Type 4 : Surfaces where individual collagen fibre skeletons are visible and their orientation can be distinguished. The orientation varies between successive exposed lamellae. These surfaces represent bone matrix which was undergoing erosion at, or close to, the time the rat was killed (Fig. 1.9).

Type 5 : In other resorbed areas the surface has a much smoother texture and individual collagen fibre skeletons cannot be distinguished. This surface type corresponds to bone which was resorbed some time before the animal was killed and has subsequently undergone some degree of mineralisation. This accounts for its smooth appearance. Such surfaces are termed resting resorbed surfaces (Fig. 1.10).

Type 6 : This surface type occurs when resorbed surfaces are overlain by a mineralisation front. The circle is now complete and new bone, i.e. type 1 is formed on the resorbed surface (type 5). (Fig. 1.10)

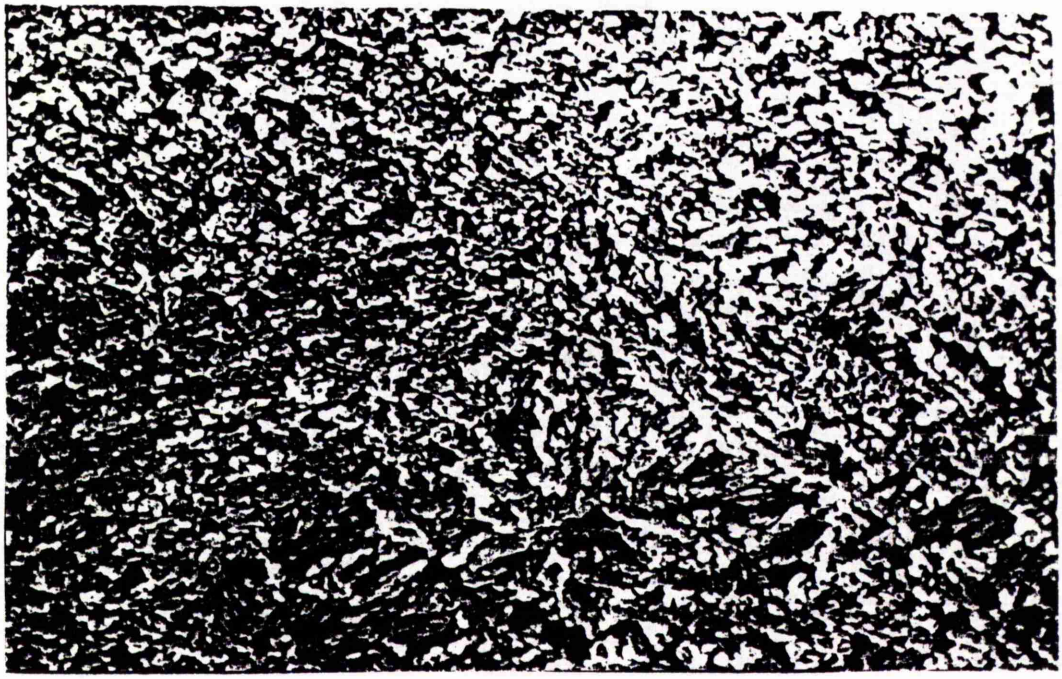


Fig. 1.7 The granular appearance of texture type 1.

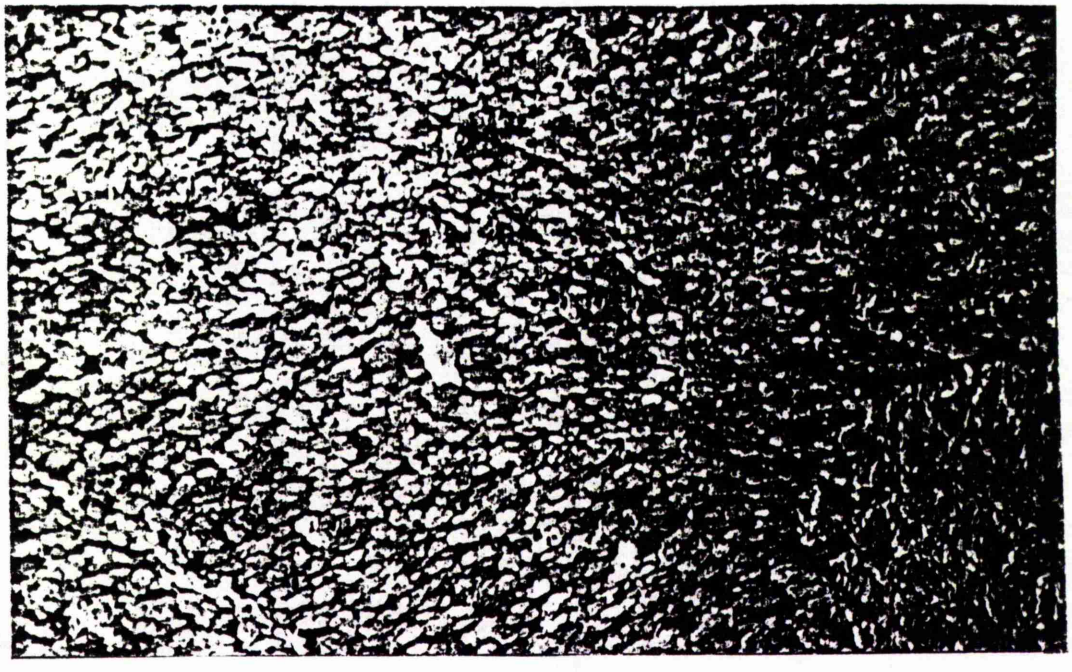


Fig. 1.8 Texture type 2, characterised by the larger mineral nodules resembling 'strings of beads'.

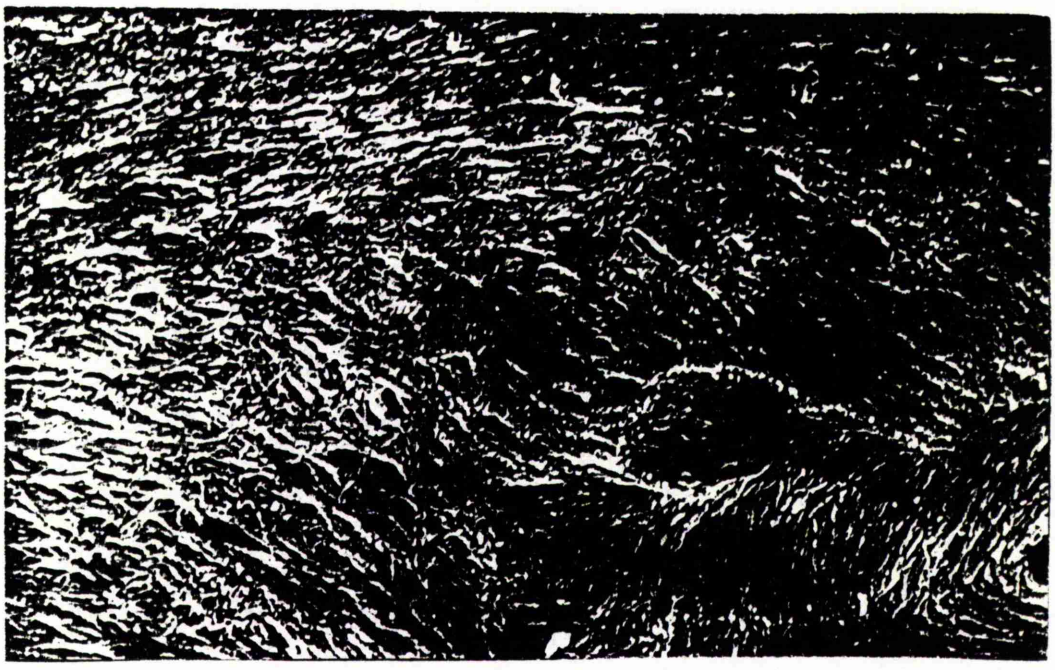


Fig. 1.9 Texture type 3, with its smooth collagen fibres, can be seen in the top left. Texture type 4 is in the bottom right. The collagen fibres are clearly visible.

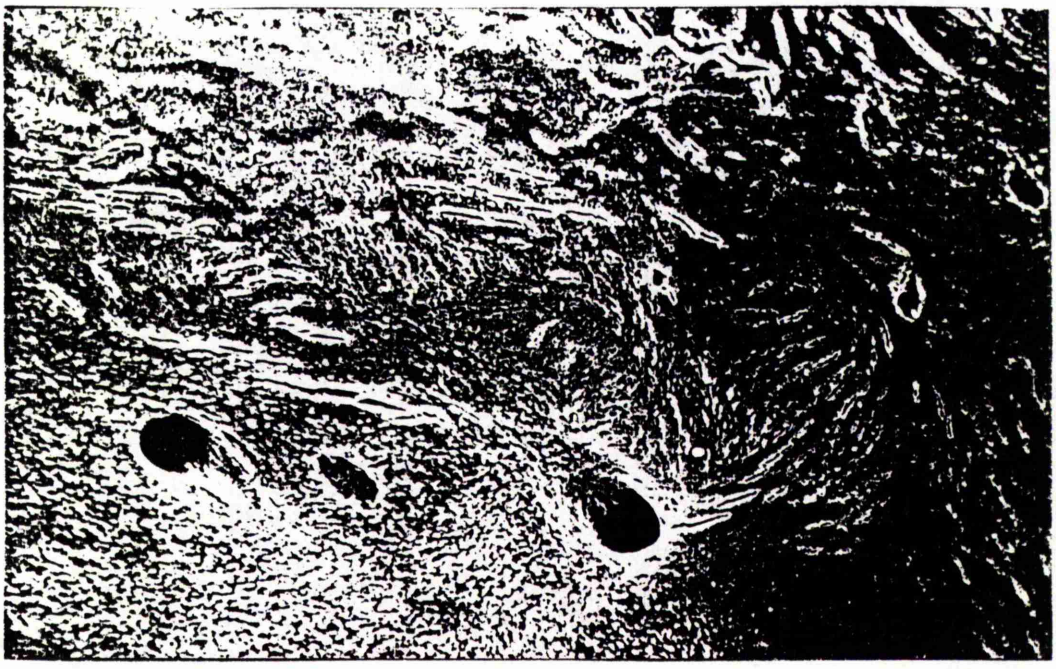


Fig. 1.10 The smooth appearance of texture type 5 can be seen at the top of the picture with the mineralisation front of type 6 in the bottom left.

The proportion of each of these six surface types present in bone varies along the length of a bone and with age. Reid (1986) found that the amount of active surface decreases with age, and depends on whether the bone is diseased or healthy. Dempster et al. (1979) measured the percentage area of each of the six different types of surface described above on tibiae from three rats in each of three experimental groups.

1) Rachitic (R) - rats fed on a Vitamin D deficient and low phosphorous diet.
2) Vitamin D-Treated (D) - rats on a similar diet to group R but with a Vitamin D supplement.
3) Normal (N) - rats fed on a "normal" diet.

More detailed descriptions of the diets can be found in Dempster et al. (1979).

The proportions were measured by taking a series of twenty-five micrographs along the middle of each bone segment and analysing the micrographs using a computer-linked planimeter (Biddlecombe et al. (1977)).

The results are illustrated overleaf and show that in rachitic bones a significantly greater proportion of the surface was occupied by incompletely mineralised bone, i.e. surface types 1 and 2, than was present in bones from Vitamin D-treated and Normal rats.

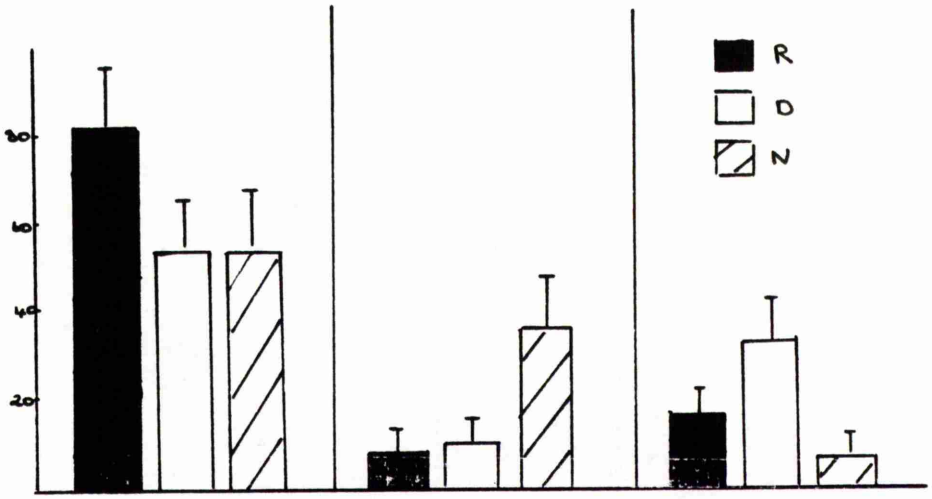


Fig. 1.11 Percentage area occupied by total incompletely mineralised, resorbed and resting surfaces on the tibial endosteum of rachitic (R), D-treated (D) and Normal (N) rats. Values are means +/- S.E.M. (n=3).

These results suggest that in rachitic and Vitamin D-treated rats the rate of mineralisation between stages 2 and 3 is much slower and a lower percentage of bone surface becomes completely mineralised.

Similar studies have been carried out on other types of diseased bone, e.g. examination of osteogenesis imperfecta samples revealed all surfaces were either mineralising or resorbed, with no resting, fully mineralised surfaces being found (Ornoy & Kim (1977), Reid & Boyde (1983), Teitelbaum et al. (1974)). Boyde et al. (1986) mention studies of various types of diseased bone, e.g. osteomalacia, renal bone disease & secondary hyperparathyroidism, fluorosis, paget's disease etc., in which differences from normal can be seen in the morphology of the bone surface. Thus measurement of certain features and surface types can provide a great deal of useful information about the effects of such diseases. It is therefore important that we are able to compare the proportions of the different surfaces which are present. Dempster's study in 1979 did obtain useful quantitative data but the method used was time-consuming and expensive

and the location of the boundaries between the surface types was decided by the physiologist.

1.4 Current Objectives

The ultimate objective of this study is to make the analysis of bone structure completely automatic. Thus the recognition of the different surface types and the drawing of the boundaries between them would be carried out by computer. This would make the procedure faster, less expensive and less dependent on the opinion of the physiologist. Whether this task is possible, remains to be seen. With this in mind, however, this study assesses various image processing techniques to investigate how automatic the procedure can become.

1.5 Specimen Preparation

The bones examined in this study were tibiae of rats. Right and left tibiae were extracted from several rats and a length of shaft was cut from the bone using a rotary saw with a very thin (0.3mm) sintered diamond blade. Distilled water was used as a coolant and a lubricant. The length of the shaft ran from the point of attachment of the fibula to the point where the crest of the tibia meets the shaft.

This shaft segment was then cut longitudinally into medial and lateral halves (Fig. 1.12).

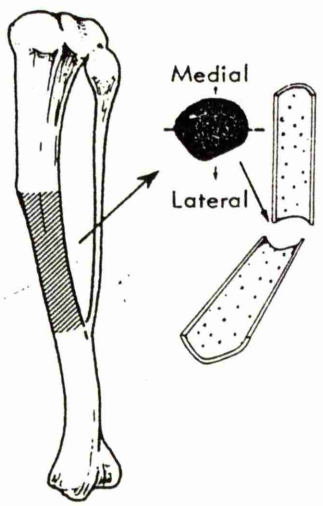


Fig. 1.12 The lateral segment of tibial shaft used for SEM.

The lateral half was soaked in 5-7% sodium hypochlorite solution overnight (Boyde & Jones (1974)). This removes all of the organic material in the bone, leaving only the mineral component of the bone matrix. The specimens were then washed in distilled water for one hour (4 changes) and dehydrated in ascending concentrations of ethanol as follows:

70% EtOH - 3 hours (3 changes)

90% EtOH - 1 hour (4 changes)

They were then soaked in absolute alcohol for one hour (2 changes) and were air-dried for 3 hours. The specimens were then glued (endosteum upwards) on to aluminium stubs and given a 50 nanometer coating of gold.

The specimens were viewed on a JEOL T100 scanning electron microscope using an accelerating voltage of 15kV. Digital images were saved on disk using a Joyce Loebl Magiscan computer interfaced to the scan output of the scanning electron microscope. The SEM magnification used was 350 but because the Magiscan only uses the central part of each scan, the final digital image magnification is in fact twice this.

Using Kermit, the images were then transferred across for use on a Sun SPARCstation, on which all the following analysis was carried out. Figures 1.13, 1.14 and 1.15 show three examples of the images obtained. Figure 1.13 contains texture types 1 and 5, figure 1.14 contains types 2 and 3 and figure 1.15 contains types 1,3 and 5. Embedded within each of the texture types, in all of the images, are vascular channels and bone cells. The images also all contain, to varying degrees, a number of artefacts. This is probably due, to a certain extent, to the age of some of the specimens - although some specimens were prepared especially for this study, many of the ones used in this initial phase of the study were prepared for Dempster's work in 1979. It should also be noted that the boundaries between texture types are far from clear. Figure 1.15 contains a reasonably clear, though not continuous, boundary between types 3 and 5, but in all other cases one texture type tends to merge into another. In addition, there is not always a natural progression through the 5 types. This is illustrated in the bottom right-hand corner of figure 1.15 where texture type 1 lies beside texture type 3. It can therefore not be assumed that anything surrounding, for example a type one texture, will be one of either texture type 5 or 2.

All the facts mentioned above make the task at hand a difficult one, and the

problems they create will be discussed in more detail throughout the thesis.

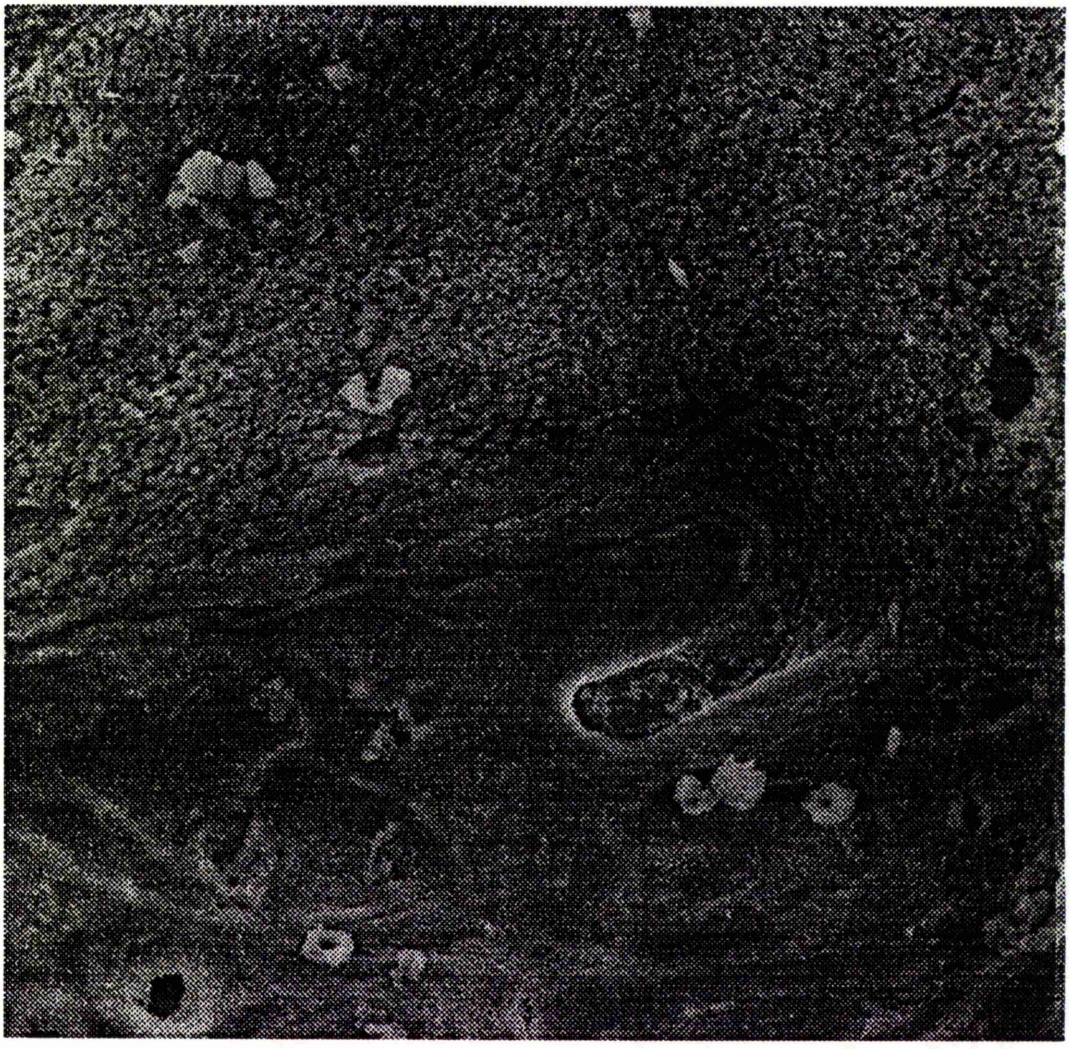


Fig. 1.13 A 512×512 image containing texture type 5 in the lower half and texture type 1 above.

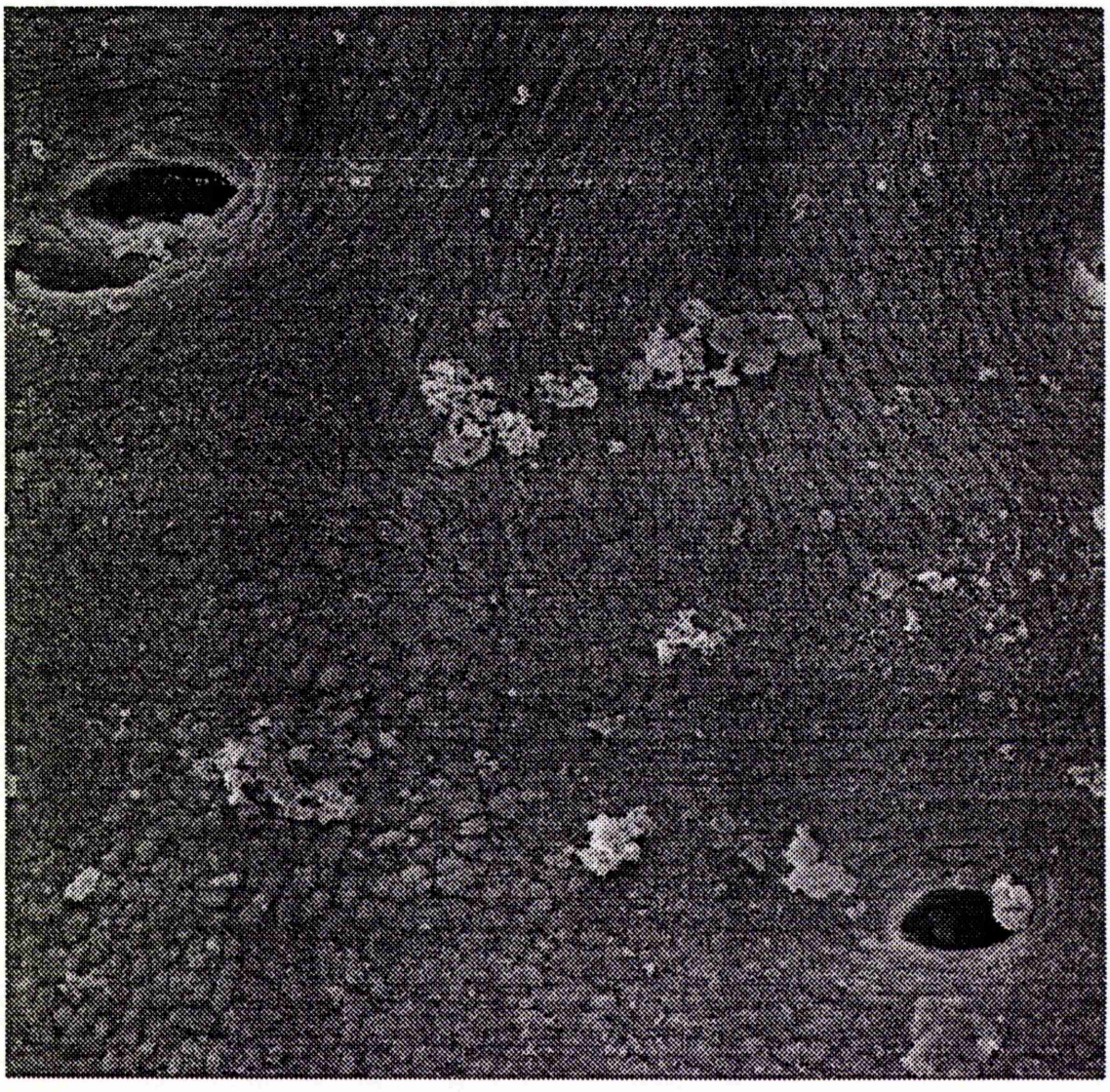


Fig. 1.14 A 512×512 image containing texture type 2 in the lower half, with texture type 3 above.

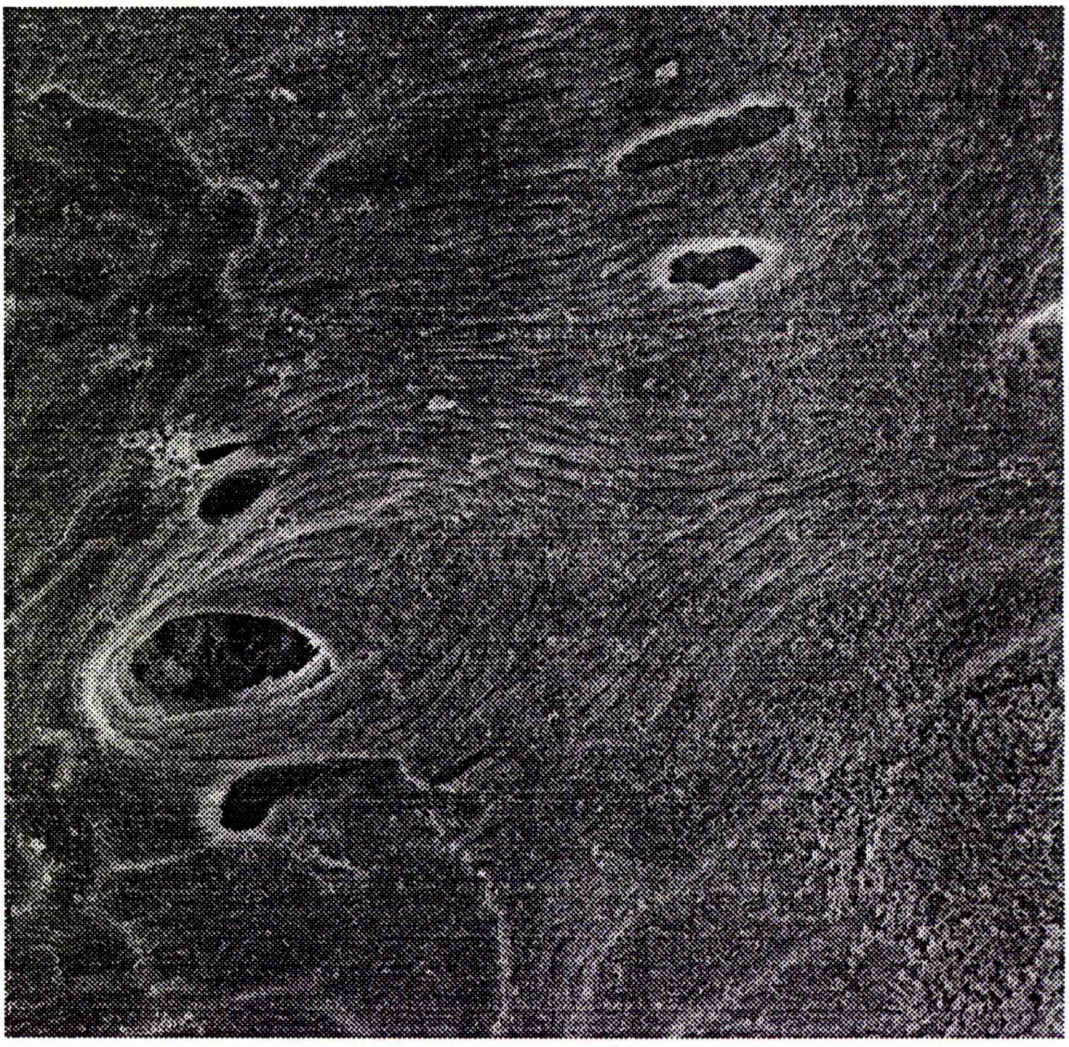
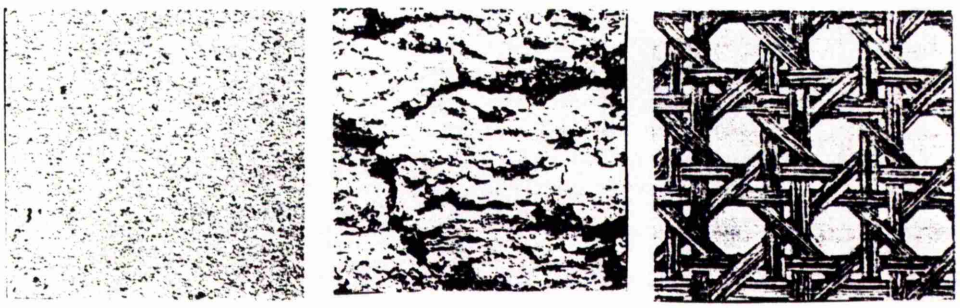


Fig. 1.15 A 512×512 image containing texture type 5 on the left, texture type 3 to the right and above, with a patch of type 1 in the bottom right.

Chapter 2 Texture Classification. A Review

2.1 Texture

Although a property of almost all surfaces, no formal definition of texture exists. It is generally viewed as a measure of properties such as smoothness, coarseness and regularity. Figure 2.1 shows examples of a) smooth b) coarse and c) regular textures. Part of the difficulty in defining texture is the large number of attributes required in the definition. In 1978, Tamura et al. identified 6 dimensions of texture - 1) coarseness (coarse v fine), 2) contrast (high contrast v low contrast), 3) directionality (directional v non-directional), 4) line-likeness (line-like v blob-like), 5) regularity (regular v irregular) and 6) roughness (rough v smooth).



(a) (b) (c)

Fig. 2.1 Examples of (a) smooth, (b) coarse and (c) regular textures.

Texture is a property of areas, the texture of a point being undefined. It is invariant to changes in orientation (i.e. rotating an image by 90° does not change its texture), brightness of the image and the size of the area. It is, however, dependent on resolution and magnification. An image magnified 50 times may appear to have a fine texture. If it is magnified 1000 times however, the texture will appear coarser.

For these reasons, standardisation of images is an important element in texture classification and will be discussed in chapter three.

2.2 Previous Approaches to Texture Classification

There are essentially 3 main approaches used in image analysis to classify the texture of a region. These are a) the statistical approach, b) the structural approach and c) the modelling approach. Modelling approaches are sometimes incorporated into the first two sections depending on whether the models are considered to be statistical or structural.

In statistical texture analysis procedures, a textured image is represented as a set of measurements called a feature vector. Many statistical methods are based on variations of grey-level co-occurrence matrices. Other statistical approaches include the use of grey-level difference statistics and grey-level run-length matrices, methods based on texture edges and filter masks, and procedures using the power spectrum of the image.

Structural approaches try to characterise the pattern which is repeated in texture. They are therefore more appropriate for textures with fairly regular structures. They involve the definition of primitives, i.e. objects or subpatterns which are repeated in a given area. These primitives are extracted from an image and the texture is characterised by means of some rule that limits the number of possible arrangements of the primitive.

With the modelling approach, texture is assumed to be a realisation of a stochastic process which is governed by some parameters. Texture analysis is viewed as a parameter estimation problem: given a textured image the problem is to estimate the parameters of the assumed random process. The estimated parameters can then serve as features for texture classification and segmentation problems. Although several models have been used to generate and represent textures, the most common approaches are based on Markov random field models.

The modelling approach has received much attention in recent literature. A difficulty with texture modelling, however, is that many natural textures do not conform to the restrictions of a particular model. Similarly, structural approaches tend not to be applicable to natural, irregular textures and have therefore been less popular than statistical methods. Due to the fairly irregular textures in the data sets used in this study the approach taken has been a statistical one. Structural and modelling approaches,

although reviewed, have not been tested on the images in this study.

2.3 Statistical Approaches

Statistical approaches compute local features at each pixel in an image and derive a set of statistics from the distributions of these features. The simplest features are the grey levels and the simplest statistics are the moments of the grey-level histogram of an image or region. Let z be a random variable denoting the grey levels and $p(z_i)$, i=1,2...,LVL be the corresponding histogram, where LVL is the number of distinct grey levels. The n^{th} central moment of z is given by:

$$\mu_n(z) = \sum_i (z_i - m)^n \, p(z_i)$$

$$where \; m = \sum_i z_i \, p(z_i)$$

is the mean value of z and is the first moment.

The central second moment

$$\sum_i (z_i - m)^2 \, p(z_i)$$

is the variance and is a measure of the spread of grey levels. The third central moment measures the skewness of the histogram: a value of 0 indicates a symmetric distribution, whereas positive or negative values indicate distributions skewed to the right or left respectively. The fourth central moment measures the flatness of the histogram.

When the problem is to classify images based on texture, the grey levels are usually standardised to be either uniformly or normally distributed with the same mean and variance. In such cases the moments could not be used to discriminate the textures.

In addition, these first-order statistics suffer from the limitation that they carry no information regarding the relative position of pixels with respect to each other. One way to bring this type of information into the process is to consider not only the distribution of grey levels, but also the position of pixels with equal or nearly equal grey levels. Thus statistics based on relative position as well as grey level are more frequently used in texture classification problems.

2.3.1 Grey-Level Co-occurrence Matrices

In 1973, Haralick et al. computed a set of matrices, called grey-level co-occurrence matrices, for a given image and suggested a set of 14 textural features which could be extracted from these matrices. He proposed that these features contain information about image textural characteristics such as homogeneity, linearity and contrast, and could therefore be used to classify images into texture types.

The first step is to construct a co-occurrence or grey-tone spatial-dependence matrix. Suppose we have an n×n image consisting of 256 (0 to 255) grey levels. A co-occurrence matrix is a 256×256 matrix computed by counting the number of times two pixels separated by distance d at a specified angle θ° (displacement (d,θ)) occur in the image, one with grey level i and the other with grey level j. From this co-occurrence matrix a normalised co-occurrence matrix can be computed by dividing each element of the matrix by the number of possible pairs of neighbouring pixels. For example, when the displacement is (1,0) i.e. the relationship is nearest horizontal neighbour, in an n×n image there will be 2(n-1) neighbouring pairs in each row and there are n rows, giving 2n(n-1) nearest horizontal pairs. Similarly there will be 2n(n-1) nearest vertical pairs and 2(n-1)(n-1) nearest diagonal pairs.

N.B. Co-occurrence matrices are usually taken to be symmetric with pairs of grey levels at angle θ and θ+180° being counted.

As a simple example, given two 16×16 grey level images IP(i,j):

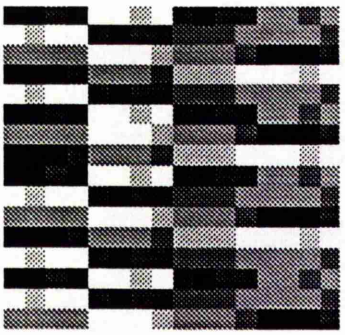


Image 1

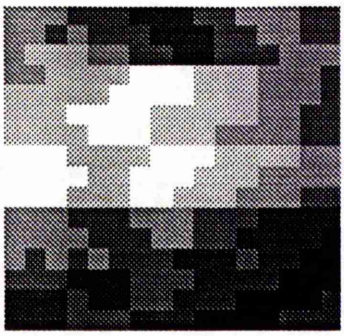


Image 2

with displacement (1,0) i.e. a horizontal displacement of 1 pixel, then the normalised co-occurrence matrices would be, respectively:

Image 1:

Grey Level	0	1	2	3	4	5
0	0.075	0.0688	0.0083	0	0	0.0083
1	0.0688	0.0375	0.0104	0.01455	0	0.0188
2	0.0083	0.0104	0.0417	0.0521	0.0063	0
3	0	0.01455	0.0521	0.1875	0.0083	0.0083
4	0	0	0.0063	0.0083	0.025	0.0646
5	0.0083	0.0188	0	0.0083	0.0646	0.0958

Image 2:

Grey Level	0	1	2	3	4	5
0	0.0375	0.01458	0	0	0	0
1	0.01458	0.1250	0.04167	0	0	0
2	0	0.04167	0.1250	0.0542	0	0
3	0	0	0.0542	0.1333	0.0208	0
4	0	0	0	0.0208	0.1542	0.0354
5	0	0	0	0	0.0354	0.0917

If a texture is coarse, and d is small compared to the size of the texture elements, the pairs of points at separation d should tend to have similar grey levels. This means that the high values in the co-occurrence matrix will be concentrated on or near its main diagonal. This is illustrated in the normalised co-occurrence matrix of image 2 above. Conversely, for a fine texture, if d is comparable to the texture element size, then the grey levels of points separated by d will be quite different, so that values in the co-occurrence matrix will be spread out relatively uniformly. Image 1 above was an example of a fairly fine texture. Thus a good way to analyse texture coarseness would be to compute some measure of the scatter of the co-occurrence matrix values around the main diagonal. Similarly, if a texture is directional then the degree of spread of the values about the main diagonal in the co-occurrence matrix should vary with the direction θ. Thus texture directionality can be analysed by comparing spread measures of the co-occurrence matrix for various values of θ.

Details of the 14 measures obtainable from these normalised co-occurrence matrices can be found in the appendix. Four of the more commonly used features are described here.

In the following, P(i,j) is the (i,j)th element of the given normalised co-occurrence matrix for a given displacement. LVL is the number of grey levels.

1) Angular Second Moment

$$\sum_{i=0}^{LVL-1} \sum_{j=0}^{LVL-1} P(i, j)^2$$

2) Contrast

$$\sum_{k=0}^{LVL-1} k^2 \sum_{|i-j|=k} \sum P(i, j)$$

3) Correlation

$$\frac{\sum_{i=0}^{LVL-1}\sum_{j=0}^{LVL-1}[i\ j\ P(i,j)] - \mu_x\mu_y}{\sigma_x\sigma_y}$$

where μ_x,μ_y,σ_x & σ_y are the means and standard deviations of the marginal distributions associated with P(i,j).

4) Entropy

$$-\sum_{i=0}^{LVL-1}\sum_{j=0}^{LVL-1} P(i,j)\ \ln P(i,j)$$

Angular second moment (ASM) is a measure of homogeneity. A homogeneous image will result in a co-occurrence matrix with a combination of high and low P(i,j)'s. In particular, where the range of grey levels is small, the P(i,j)'s will tend to be clustered around the main diagonal resulting in a high value of ASM. A non-homogeneous image will result in an even spread of P(i,j)'s and hence a low ASM.

Contrast is a measure of the local variations present in an image. If there is a large amount of variation in an image the P(i,j)'s will be concentrated away from the main diagonal and the contrast feature will be high.

Correlation is a measure of grey-tone linear dependencies in the image and will be high if an image contains a considerable amount of linear structure.

Entropy is another measure of the homogeneity of an image. It is large when all the P(i,j)'s are of similar magnitude and small when the P(i,j)'s are unequal.

For the two images illustrated above, the values of the 4 features are as follows:

	Image 1	Image 2
ASM	0.0789	0.0957
Contrast	1.7250	0.3333
Correlation	0.0218	0.0210
Entropy	2.8606	2.5209

Thus, for example, contrast is much higher for image 1 where there is much more variation. Image 2 is more homogeneous than image 1 resulting in a higher ASM.

Haralick et al. (1973) tested their 14 textural features based on co-occurrence matrices on three data sets. These sets were taken from a) photomicrographs of different rocks, b) aerial photographs of man-made and natural scenes and c) high-altitude satellite pictures of the earth. By computing the mean and variance over four directions, 1 pixel apart, of several of the features (4 features for the micrograph data and the satellite images and 11 for the aerial photographic data) between 80% and 90% of the textures in each of the sets were classified correctly.

In a study by Conners and Harlow in 1980 it was claimed that the 14 textural features proposed by Haralick et al. (1973) do not contain all the important information about the texture and more recent studies have sought to address this problem. This has resulted in the proposal of two further measures, called cluster shade and cluster prominence, which can be computed from co-occurrence matrices (Conners et al. (1984)). These measures are believed to guage the perceptual concepts of uniformity and proximity (Julesz (1962)) and are given by:

$$\textit{Cluster Shade} = \sum_{i=0}^{LVL-1} \sum_{j=0}^{LVL-1} (i+j-\mu_i-\mu_j)^3 \, P(i, j)$$

$$\textit{and} \quad \textit{Cluster Prominence} = \sum_{i=0}^{LVL-1} \sum_{j=0}^{LVL-1} (i+j-\mu_i-\mu_j)^4 \, P(i, j)$$

$$\textit{where } \mu_i = \sum_{i=0}^{LVL-1} i \sum_{j=0}^{LVL-1} P(i, j) \textit{ and } \mu_j = \sum_{i=0}^{LVL-1} \sum_{j=0}^{LVL-1} j \, P(i, j).$$

In 1984, Conners et al. included cluster shade and cluster prominence in the set of co-occurrence matrix features used to segment a high-resolution black-and-white image of Sunnyvale, California. Regions were classified as belonging to one of 9 classes, e.g. residential, mobile home, water, runway etc.. Training data consisted of 1135 samples of size 145×145. Altogether 48 displacements were considered - d= 1,2,4,6,8,12,16,20 and θ = 0°,19°,75°,90°,109°,161°. The values of θ were selected according to the orientation of the streets. The features computed were ASM, contrast, entropy, IDM, cluster shade and cluster prominence. Images were first classified as Uniform (i.e. containing only one of the 9 classes), Boundary (2 or more of the 9 classes) or Unspecified (one or more unknown classes). Boundary and Unspecified regions were then broken down into smaller regions and subjected to the same procedure until Uniform classes were obtained. These were then classified into one of the nine types. Testing the method on the training set, 90% of the images were correctly classified. When the procedure was tested on a test set consisting of 1156 samples, the number correctly classified was 83.4%. The results did not specify which features were most useful in classifying the data.

Yogesan et al. (1993) included cluster shade and cluster prominence in their investigation to discriminate between electron microscopy images of normal, proliferating, pre-cancerous and cancerous mouse liver cell nuclei, but did not find them to be particularly useful features for classifying livers of mice. They found the best features extracted from co-occurrence matrices to be variance, the kappa statistic (Parkkinen & Selkainaho (1990)), which is described later, and a new measure called sum of homogeneity: SH = $\sum$ P(i,i) which measures homogeneous areas in an image.

Carstensen (1993) also investigated cluster shade and cluster prominence in classifying 15 textures from Brodatz' (1966) texture album. As with Yogesan et al. (1993) he did not find them to be particularly useful. He found the most useful features to be correlation, contrast, sum variance, and a feature proposed by Laws in 1980 called diagonal moment defined by:

$$\sum_{i=0}^{LVL-1} \sum_{j=0}^{LVL-1} |i-j| \; |i+j-2\mu| \; P(i, j).$$

The fact that the co-occurrence matrix depends on the displacement means that different displacements reveal different information about the texture, and several co-occurrence matrices with varying displacements could be combined in classification. This, however, raises the problem of which displacement is best. This problem was addressed by Zucker and Terzopoulos in 1980 and Terzopoulos in 1985. Based on the theory of contingency tables, Zucker and Terzopoulos (1980) suggested a χ^2 statistic to measure the association between the grey levels of two pixels situated (dk,dl) apart. The test statistic is given by:

$$\chi^2 = \sum_{i=1}^{LVL} \sum_{j=1}^{LVL} \frac{[P(i, j)-P(i, .)P(., j)]^2}{P(i, .)P(., j)}$$

where P(i,.) and P(.,j) are row and column sums of the normalised co-occurrence matrix.

For a periodic texture, if the displacement is in the orientation of the periodicity, and if the distance between the pixels is equal to one period, then there will be perfect association and the χ^2 statistic will have a high value. Thus the χ^2 statistic can be used to measure the periodicity of the texture, and can be used to find the 'best' displacement, or simply as a feature for classification. In 1990, Parkkinen & Selkainaho pointed out that there exists another statistic, called kappa, also based on the co-occurrence matrix, which measures the periodicity of a texture, and the authors claim

that the kappa statistic is preferable because it is of order LVL, whereas χ^2 is of order LVL^2.

The normalised kappa statistic (Cohen (1960)) is given by:

$$Kappa\ (\kappa) = \frac{P_o - P_c}{P_m - P_c}$$

$$where\ P_o = \sum_{i=0}^{LVL-1} P(i,i),$$

$$P_c = \sum_{i=0}^{LVL-1} P(i,.)P(.,i)$$

$$and \quad P_m = \sum_{i=0}^{LVL-1} \min\ [P(i,.),P(.,i)].$$

Parkkinen & Selkainaho (1990) compared the kappa statistic for measuring texture periodicity with the χ^2 statistic used by Zucker and Terzopoulus (1980) and Terzopoulos (1985). To do this, they used images taken from Brodatz' (1966) texture album. No attempt was made to classify the images, but Parkkinen & Selkainaho (1990) proposed that if texture periodicity is appropriate (i.e. if the texture has a regular structure) the kappa statistic is a better measure than the χ^2 statistic.

In 1982 Vickers and Modestino classified textures using co-occurrence matrices without further feature extraction. Their method was based on a maximum likelihood approach, allocating an image to its most likely texture type based on the whole co-

occurrence matrix. The grey-level co-occurrence matrix C(d,θ) is an LVL×LVL matrix whose (i,j)th element is given by the number of times the grey levels i and j occur d pixels apart at an angle θ°. The matrix is then made rotation invariant by averaging over θ in all directions and the result is denoted C(d). A co-occurrence matrix C(d), and hence an image IP is classified as texture type k_0 if

$$L_{k_0}\,[C(d)] = \max_{0<k<K-1} L_k[C(d)]$$

(K possible texture types)

where

$$L_k[C(d)] = \ln \frac{P(C(d)\,|k)}{P_0[C(d)]}$$

and $P_0[C(d)]$ is a normalising function.

Assuming a multinomial distribution,

$$P[C(d)\,|k] = \frac{\sum_{i,j} C(i,j;d)!}{\prod_{i,j} C(i,j;d)!} \prod_{i,j} Q(i,j;d\,|k)^{C(i,j;d)}$$

where Q(i,j;d|k) = Probability of observing grey levels i and j at a pair of points separated by distance d if an image belongs to class k.
By taking the normalisation function $P_0[C(d)]$ as

$$\frac{\sum_{i,j} C(i, j;d)!}{\prod_{i,j} C(i, j;d)!}$$

$L_k[C(d)]$ simplifies to

$$\sum_i \sum_j C(i, j;d) \ln Q(i, j;d|k).$$

To complete the description of the classifier it is necessary to estimate Q(i,j;d|k). The maximum likelihood estimator of Q can be shown to be

$$\hat{Q}(i, j;d|k) = \frac{C(i, j;d|k)}{C}$$

$$= P(i, j;d|k)$$

$$\textit{where} \quad C = \sum_i \sum_j C(i, j;d)$$

and so the maximum likelihood rule becomes:
allocate to k_0 if:

$$\sum_i \sum_j C(i, j;d) \ln P(i, j;d|k_0) > \sum_i \sum_j C(i, j;d) \ln P(i, j;d|k)$$

for all k.

Vickers and Modestino (1982) used their maximum likelihood method to classify nine textures from Brodatz' (1966) texture album. For each texture, 16 64×64

windows were extracted from a 512x512 image and divided into training and test sets. Results are published with displacements d=1, d=3 and d=5. Displacement d=5 produced the best results on test images, with 98% of the images being classified correctly. This method therefore achieved considerably better results than other studies where only a subset of textural features was used, e.g. those of Haralick et al. (1973) and Weszka et al. (1976).

A generalisation of co-occurrence matrices was suggested by Davis et al. in 1979. Instead of constructing **grey-level** co-occurrence matrices, the authors suggested constructing a co-occurrence of, for example, edges. An edge-detector is applied to an image (the authors used a Kirsch operator) to find edge-maxima and their corresponding directions, e.g.

.	.	.	.	.	.	V	.
.	H	H	H	.	.	V	.
.	V	.	V	.	.	.	L
.	H	H	H	.	.	.	.
.	.	.	.	.	R	.	.
.	R	.	.	R	.	L	.
R	.	V	.	V	.	R	.
H	H	.	.	H	H	.	.

V=Vertical Edge L=Left Diagonal Edge
H=Horizontal Edge R=Right Diagonal Edge

A generalised co-occurrence matrix can then be constructed by looking at the relationship between these edge-directions. For the above example, counting pixels distance one apart in any direction would produce the following matrix:

	H	V	L	R
H	12	5	0	1
V	5	2	0	1
L	0	0	0	1
R	1	1	1	0

Various relationships can be considered, and in fact Davis et al. (1979) counted the occurrences of edges in cone-shaped regions as shown in Figure 2.2.

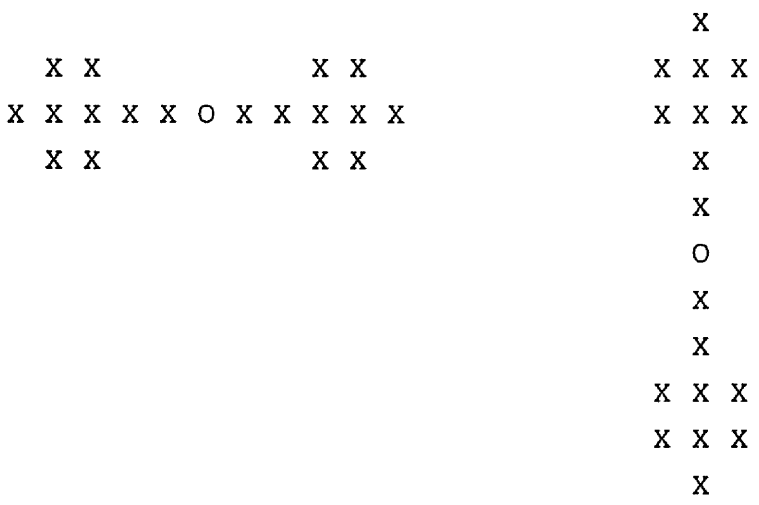


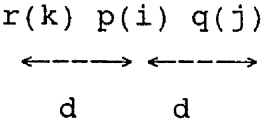
Figure 2.2 Cone-shaped regions used by Davis et al. (1979).

The authors suggested several features which could be extracted from these generalised co-occurrence matrices. These are similar to those suggested by Haralick et al. in 1973, i.e. angular second moment, contrast, entropy and correlation, although the interpretation depends on the operator used to obtain the co-occurrence matrix, e.g. edge-detector, spot-detector etc., and also the displacement used. The authors also suggested that ratios of features could provide useful texture information.

In their experiment, Davis et al. (1979) classified 6 samples of each of 5 texture

types from Brodatz' (1966) texture album using 4 generalised and 2 grey-level co-occurrence matrices. They computed four features - angular second moment, contrast, correlation and entropy, and, using a leave-one-out method, classified a maximum of 87% correctly, based on contrast and entropy from a generalised co-occurrence matrix using a 5×5 edge operator. They compared this method with grey-level co-occurrence matrices where they obtained a maximum correct classification of 57%. In 1981, Davis, Clearman and Aggarwal conducted a similar study using a larger database. Again they claimed to have obtained better classification results with generalised co-occurrence matrices than with grey-level co-occurrence matrices, (61% versus 52%).

In 1980, Dyer et al. suggested a modification of the generalised co-occurrence matrices above. They also applied an edge-detector to an image, this time taking the absolute difference in grey level in 2×2 neighbourhoods in 4 directions. They suggested 6 different co-occurrence matrices that could be constructed from this edge map. One example was 'most similar neighbour along an edge'. Let p be a point having nonzero edge magnitude, and edge orientation θ, and let d be a given distance. Let q be the point at distance d from p in the θ direction and let r be the point at distance d from p in direction $\theta+\pi$. Let i,j,k be the grey levels of points p,q,r respectively.



Let P(i,j) be the (i,j)th element of the co-occurrence matrix. If $|i-j| < |i-k|$, increment P(i,j). Otherwise increment P(i,k). Thus both the edge information and grey-level information are used. Other co-occurrence matrices can be constructed by computing e.g. most similar neighbours across an edge, least similar neighbours across an edge etc. etc.. For details, see Dyer et al. (1980). Once the co-occurrence matrix is computed, features such as ASM, contrast, entropy and correlation can be extracted.

Although no quantitative results were published by Dyer et al. in 1980, the authors reported an improvement in classification of images from Brodatz' (1966) texture album over fixed separation grey-level co-occurrence matrices. Entropy, from any of the 6 co-occurrence matrices produced good separation. For a sample of

LANDSAT images, however, separation was poor in all cases.

2.3.2 Grey-Level Difference Statistics

Grey-Level Difference Statistics are based on absolute differences between pairs of grey levels. As with co-occurrence matrices the starting point is an n×n image consisting of 256 grey levels. The grey-level difference statistics are contained in a 256-dimensional vector and are computed by taking the absolute difference of all possible pairs of grey levels distance d apart at angle θ° and counting the number of times the difference is 0,1,...,255. The difference statistics are then normalised by dividing each element of the vector by the number of possible pairs of pixels.

As examples, consider the two 16×16 grey-level images illustrated in 2.3.1, i.e.

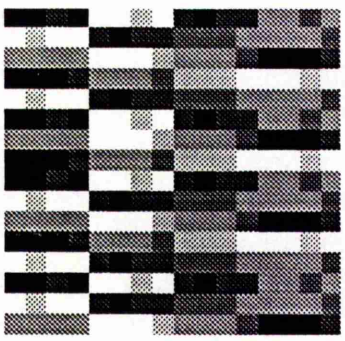


Image 1

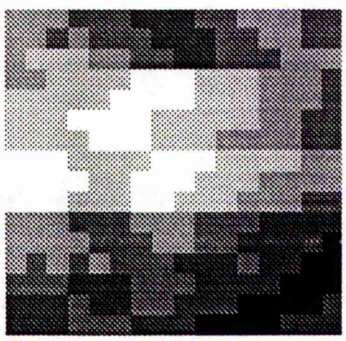


Image 2

and displacement (1,0).

The vectors of grey-level difference statistics would be:

$$Image\ 1 = \begin{bmatrix} 0.4625 \\ 0.4083 \\ 0.0750 \\ 0 \\ 0.0375 \\ 0.0167 \end{bmatrix}$$

$$Image\ 2 = \begin{bmatrix} 0.6667 \\ 0.3333 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

In image 2 only the top two values in the vector are non-zero. This indicates that neighbouring pixels have the same or similar (a difference of 1) grey levels, i.e. the texture is fairly coarse. In image 1 the values in the vector of difference statistics are more evenly spread indicating a finer texture.

Weszka et al. (1976) suggested 4 measures which could be computed from these vectors which may be useful in classifying texture.

In the following, the P(i)'s are the i^{th} elements of the vectors of normalised grey-level differences.

1) Angular Second Moment

$$\sum_{i=0}^{LVL-1} P(i)^2$$

2) Contrast

$$\sum_{i=0}^{LVL-1} i^2 \, P(i)$$

3) Entropy

$$- \sum_{i=0}^{LVL-1} P(i) \ln P(i)$$

4) Mean

$$\frac{1}{LVL} \sum_{i=0}^{LVL-1} i \, P(i)$$

Angular second moment, contrast and entropy provide similar textural information to those with the same name based on co-occurrence matrices. Mean is essentially another measure of contrast. It will be small when neighbouring pixels have similar grey levels and the P(i) are concentrated near the top of the vector.

The corresponding features for the images above are as follows:

	Image 1	Image 2
ASM	0.3880	0.5556
Contrast	1.725	0.3333
Entropy	0.1080	0.6365
Mean	0.7917	0.3333

Comparing the two images, entropy is larger for image 2, indicating a more homogeneous image. Contrast and mean are higher for image 1 indicating more variation.

2.3.3 Run-Length Statistics

Another set of measures, suggested by Galloway in 1975 are based on grey-level run-length statistics. Again with an n×n image consisting of 256 grey levels and direction θ, run-length statistics are calculated by counting the number of runs of a given length (from one to n) for each grey level. In a coarse texture it is expected that long runs will occur relatively often whereas a fine texture will contain a higher proportion of short runs.

Again, using the simple grey-level images in 2.3.1, i.e.

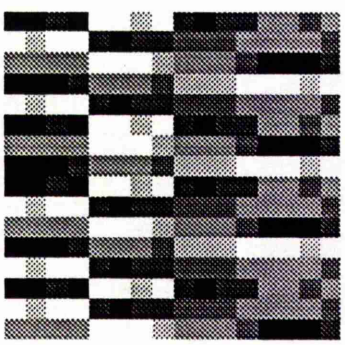


Image 1

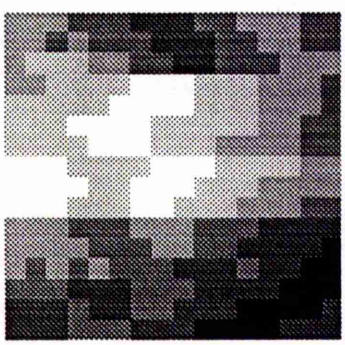


Image 2

and direction $\theta = 0°$,

the run-length matrices would be:

Image 1:

		RUN-LENGTH			
		1	2	3	4
GREY LEVEL	0	13	4	7	0
	1	20	9	0	0
	2	16	0	5	0
	3	4	4	7	9
	4	16	0	3	0
	5	12	9	7	0

Image 2:

		RUN-LENGTH				
		1	2	3	4	5
GREY LEVEL	0	1	1	1	2	0
	1	1	3	9	3	0
	2	7	10	10	0	0
	3	5	7	8	3	0
	4	0	1	8	4	2
	5	0	1	6	3	0

Thus, for example, image 1 contains 9 runs of length 4 of grey level 3.

Galloway (1975) proposed five measures which can be computed from run-length matrices, where P(i,j) is the (i,j)th entry of the run-length matrix, N_g = the number of grey levels and N_r = the number of different run lengths.

1) Long Run Emphasis

$$LRE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} j^2 P(i, j)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} P(i, j)}$$

2) Grey-Level Distribution

$$GLD = \frac{\sum_{i=1}^{N_g}[\sum_{j=1}^{N_r} P(i, j)]^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} P(i, j)}$$

3) Run-Length Distribution

$$RLD = \frac{\sum_{j=1}^{N_r}[\sum_{i=1}^{N_g} P(i, j)]^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r} P(i, j)}$$

4) Run Percentage

$$RP = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i, j)}{n^2}$$

(where n^2 = no. of pixels)

5) Short Run Emphasis

$$SRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{P(i, j)}{j^2}}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} P(i, j)}$$

Long run emphasis will be large when there are lots of long runs of the same grey level, whereas short run emphasis and run percentage will be large when there are lots of short runs. If there is a mixture of long and short runs, i.e. the run lengths are not evenly distributed, run-length distribution will be large. If run lengths are not evenly distributed over the different grey levels, grey-level distribution will be large.

The five features for the two images above are:

	Image 1	Image 2
SRE	0.6295	0.2649
LRE	4.0690	8.0625
GLD	24.6828	19.4167
RLD	56.2690	28.3125
RP	145.0000	96.0000

Run percentage and short run emphasis are higher for Image 1 which is as expected given the variability of the image. Grey-level distribution is higher for Image where it can be seen from the run-length matrix that the run lengths are not evenly distributed over the grey levels.

Galloway (1975) used run-length matrices to classify 54 64×64 images into one of nine terrain categories such as orchard, wood, urban, suburb etc.. Before the run-length matrices were computed, grey levels were grouped into eight sets. The runs were therefore computed for eight grey-level groups: 0-7, 8-15, 16-23, 24-31, 32-39, 40-47, 48-55 and 56-63. The run-lengths were also grouped into the ranges 1, 2-3, 4-7, 8-15, 16-31 and 32-64. Thus, the run-length matrices were 8×6 arrays containing 8 grey-level groups and 6 run-length groups. Using these matrices Galloway, 1975 achieved a classification accuracy of around 83%.

Yogesan et al. (1993) examined various features taken from run-length matrices as well as co-occurrence matrices in their study to classify livers of mice. Of the run-length matrix features they found the most useful to be short-run emphasis.

2.3.4 Fourier Power Spectrum

Bajcsy (1973) computed the power spectrum, which gives the magnitude of the frequency components in the Fourier transform of an image.

The Fourier transform of an n×n digital picture is defined by:

$$F(u, v) = \frac{1}{n^2} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} f(i, j) \exp\left[-2\pi \sqrt{-1}(iu+jv)\right]$$

$$0 \le u, v \le n-1$$

and the Fourier power spectrum is $|F|^2 = FF^*$ (where * denotes the complex conjugate).

A coarse texture will have high values of $|F|^2$ concentrated near the origin,

while in a fine texture the values of $|F|^2$ will be more spread out. Texture directionality can also be identified by examining the values of $|F|^2$. If a texture contains features orientated in one direction, the high values of $|F|^2$ will also tend to lie in a single direction. This direction will be perpendicular to the direction in the image. For example, horizontal streaks in the image will result in vertical streaks in the power spectrum. If the texture has no obvious direction, no directional tendencies in the power spectrum would be expected.

In order to simplify interpretation, the two-dimensional power-spectral function is usually condensed to a one-dimensional function. To do this, the spatial frequency domain is broken up into either rings or wedges centred at the origin.

Firstly considering the rings, the average of $|F|^2$ within a circle of radius r centred at the origin is computed, i.e.

$$\phi_r = \int_0^{2\pi} |F(r,\theta)|^2 \, d\theta.$$

for various values of r, the ring radius.
Thus the rings are given by:

$$\int_{r_1}^{r_2} \int_0^{2\pi} |F(r,\theta)|^2 \, d\theta dr$$

where r_1 and r_2 are the inner and outer radii respectively.

The equivalent discrete version for an n×n image is given by:

$$\sum_{u=0}^{n-1} \sum_{v=0}^{n-1} |F(u,v)|^2$$

where $r_1^2 \leq u^2+v^2 \leq r_2^2$

$0 \leq u,v \leq n-1.$

Coarse textures will result in high values at low spatial frequencies whereas fine textures will tend to result in values at higher frequencies.

The wedges are of the form:

$$\phi_\theta = \int_0^x \int_{\phi_1}^{\phi_2} |F(r,\theta)|^2 d\phi dr$$

where the angles ϕ_1 and ϕ_2 delineate the wedge. The equivalent discrete equation is given by:

$$\sum_{u=0}^{n-1} \sum_{v=0}^{n-1} |F(u,v)|^2$$

where $\phi_1 \leq \arctan(u/v) < \phi_2$

and $0 < u,v < n-1$.

In this case a flat distribution implies a nondirectional texture. On the other hand, peaks in the function suggest a specific orientation of the texture.

As a pilot study Weszka et al. (1976) tested co-occurrence matrices, difference statistics, run-length matrices and Fourier features on Haralick's aerial photographic data set. This consisted of 6 samples of each of nine classes of land i.e. lake, woods, marsh, orchard, urban, suburb, scrub, railroad and swamp.

Each sample was a 64×64 pixel image with 64 grey levels. Weszka et al. (1976) tested the four methods using various displacements (4 directions 0°,45°,90° & 135°, and distances 1,2,4 & 8 pixels apart). The only features examined were contrast taken from co-occurrence matrices, mean from difference statistics and wedges and rings for the Fourier features. Four run-length features were tested in a preliminary study but performed so poorly compared with the other methods that they were not tested further.

Results using individual features were poor, with between 7 and 25 samples out

of a possible 54 being correctly classified. No one method performed consistently better than any other. Combining pairs of features produced much better results with features from difference statistics classifying the largest percentage of images correctly. The best result was 43 out of 54, i.e. nearly 80%. This result was obtained using the two displacements (2,135°) and (4,135°).

In this study co-occurrence matrices performed better than Fourier features but less well than difference statistics.

Weszka et al. (1976) also tested what they called composite features, i.e. taking the mean and standard deviation of the various features over 4 directions and also over 4 sizes. The authors concluded that composite features performed no better than raw features.

In their main study Weszka et al. (1976) classified LANDSAT images of three geological terrain types: Mississippian limestone and shale, Lower Pennsylvanian shale and Pennsylvanian sandstone and shale. A set of sixty 64×64 windows was selected from each of the three regions. The displacements used were the same as in the pilot study, as was the Fourier feature set. Contrast was again measured from co-occurrence matrices where the co-occurrence matrices were computed based on both pairs of average grey levels and pairs of grey levels of single points. Where averaging was used, the averaging neighbourhoods were square and were of the same size as the displacement (i.e. 1,2,4 or 8). Contrast and mean were computed from difference statistics again using pairs of average grey levels and single points.

The authors concluded that the best features always involved small sizes. Features based on averages did better than single points for large sizes but less well than single point features using small sizes. Thus, from the results published there appears to be no advantage in looking at averages since the best overall results were obtained from single points using small sizes. Large sizes gave poor results with both averages and single points. The best result (167 out of 180) was obtained from difference statistics taking the mean of single points and displacements ($\sqrt{2}$,45°) and (2,0°).

In a supplementary study the authors investigated three other features proposed by Haralick et al. (1973). These were ASM, entropy and inverse difference moment (IDM), both from co-occurrence matrices and difference statistics. As in the main study, these were computed based on both average grey levels and single points. IDM

performed worst, entropy from co-occurrence matrices did slightly better than entropy from difference statistics, while ASM performed best, with the measurements from difference statistics doing marginally better than those from co-occurrence matrices. Similarly, when pairs of features were investigated, the best results were obtained using ASM from difference statistics based on single points with displacements ($\sqrt{2}$,45°) and (2,0°).

In summary, the computationally cheapest of the features, namely the means of the single-point difference statistics seemed to do as well as any of the other features. It should be noted, however, that in this study only a small subset of the possible feature set was examined. An improvement in the classification results may have been achieved by using some of the other 14 features or indeed a combination of two or more features.

In 1980, Conners and Harlow did a theoretical comparison of the four methods - Fourier power spectrum, co-occurrence matrices, difference statistics and run-length matrices. In this study the procedure did not depend on the displacements used but sought to measure the amount of texture information contained in the matrices from each of the four methods.

The data set consisted of texture pairs, where a texture pair is a set containing two textures $X_1(n,m)$ and $X_2(n,m)$. One texture is represented by the random field $X_1(n,m)$ and the other is represented by the random field $X_2(n,m)$. Co-occurrence matrices, difference statistics etc. can be computed for each texture pair.

If the matrices are equal the textures cannot be discriminated. If, however, two of the corresponding matrices are not equal, the algorithm can use the inequality to discriminate the textures.

On the whole the results of this study agreed with those of Weszka et al. (1976). The study showed that both co-occurrence matrices and difference statistics perform better than Fourier power spectrum, and that run-length matrices are poor at discriminating texture. The only difference in conclusions from the two studies is that Conners and Harlow (1980b) found that co-occurrence matrices performed better than difference statistics. They proposed that the reason for the different results may be that the difference statistics do not necessarily contain more textural information than co-occurrence matrices but simply that the most commonly measured features from co-occurrence matrices i.e. ASM, contrast, correlation, entropy and IDM do not contain all

the important textural information.

One problem with all the methods discussed so far is that they depend on the orientation of the image. The solution used by many researchers has been to compute the relevant feature over 4 (or more) directions and calculate the mean value.

2.3.5 Neighbouring Grey-Level Dependence Matrices

In 1983 Sun and Wee introduced the neighbouring grey-level dependence matrix (NGLDM) where the relationship between an element and all its neighbouring elements is considered at one time instead of one direction at a time. They proposed that this method eliminates the angular dependency and reduces both computational time and storage requirements. In addition, because the method only depends on the **relationship** between a pixel and its neighbours, the authors claim the method is invariant to grey-level transformation.

For an image IP(i,j), the neighbouring grey-level dependence matrix is computed for given d and a by counting the number of times the difference between each element in IP(i,j) and its neighbours is equal to or less than a at a certain distance d. As an example, if the image is

$$IP(i, j) = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 2 & 2 & 2 \\ 2 & 2 & 3 & 3 \end{bmatrix}$$

and d = 1, a = 0, (i.e. 1 pixel apart), P(i,j,1,0) for the four pixels in the centre of the image is given by:

(0,4) (1,3)

(2,3) (2,3)

i.e P(2,2) has grey level 0 and has 4 neighbours also with grey level 0. The number of neighbours is called the NGLDM number and the matrix Q(i,j), where Q(i,j) is the total number of entries with grey level i and NGLDM number j is as follows:

		NGLDM Number								
		0	1	2	3	4	5	6	7	8
	0	0	0	0	0	1	0	0	0	0
G	1	0	0	0	1	0	0	0	0	0
L	2	0	0	0	2	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0

Sun and Wee (1983) suggested five measures that can be taken from the matrix Q(i,j) to describe the texture.

1) Small Number Emphasis

$$SNE = \sum_i \sum_j \frac{\frac{Q(i,j)}{j^2}}{R}$$

2) Large Number Emphasis

$$LNE = \sum_i \sum_j \frac{[j^2\ Q(i,j)]}{R}$$

3) Number Nonuniformity

$$NNU = \frac{\sum_i [\sum_j Q(i, j)]^2}{R}$$

4) Second Moment

$$SM = \frac{\sum_i \sum_j [Q(i, j)]^2}{R}$$

5) Entropy

$$ENT = \frac{-\sum_i \sum_j Q(i, j) \ln Q(i, j)}{R}$$

where R is the normalising factor and is equal to the total number of entries in the Q(i,j) matrix.

SNE measures the fineness of an image. An image consisting of fine texture will have a reasonable range of grey levels in a neighbourhood and hence NGLDM numbers concentrated in the small NGLDM number columns, making $Q(i,j)/j^2$ large for small j. Thus, the larger the SNE of an image is, the finer the image is. For a coarse texture picture, a neighbourhood will contain similar grey levels and the large NGLDM numbers will be concentrated in the large j columns, making j^2 Q(i,j) larger for large j. Thus if LNE of one image is larger than the other, then the former will be coarser than the

latter. The measurement of the homogeneity of the Q matrix is given by SM. In a homogeneous image, there are only a few large entries in the Q matrix. Then SM will be large because it is the sum of the squares of all entries in the Q matrix. SM is equivalent to ASM in co-occurrence matrices. NNU and ENT are related to the coarseness of an image. These are similar measures to contrast based on co-occurrence matrices.

Sun and Wee (1983) used a data set similar to that of Weszka et al. (1976) to compare their neighbouring grey-level dependence matrices with other methods. 64×64 images of 3 geological terrain types, histogram flattened to contain 64 grey levels, were classified using a nearest-neighbour rule. Using one feature, 79.44% were correctly classified, compared with Weszka et al.'s (1976) 75.5%. For pairs of features, Weszka et al.'s (1976) results were slightly better - 93% compared with Sun and Wee's (1983) 85%.

Berry and Goutsias (1991) extended the maximum likelihood classifier of Vickers & Modestino (1982) to investigate neighbouring grey-level dependence matrices (NGLDM) and neighbouring spatial grey-level dependence matrices (NSGLDM), as well as spatial grey-level dependence matrices (SGLDM or co-occurrence matrices). They tested the methods on 9 Brodatz' (1966) textures - a similar data set to that used by Vickers and Modestino (1982). They achieved 100% correct classification with both neighbouring grey-level dependence matrices and co-occurrence matrices with a displacement of 1 pixel and 64×64 images, although for smaller images - 32×32 and 16×16 - the authors found that co-occurrence matrices performed better than neighbouring grey-level dependence matrices and neighbouring spatial grey-level dependence matrices. For 64×64 images, results were comparable for all 3 methods but the authors found that computational time was considerably reduced by using NGLDM or, to a greater extent, by NSGLDM rather than co-occurrence matrices.

2.3.6 Relative Extrema Measures

Mitchell et al. (1977) suggested looking at the number of grey level extrema along a one-dimensional scan line. The images are first smoothed using a threshold T, so that only principal extrema are retained, and then the number of extrema outside

various thresholds are counted.

e.g. 

Thus for threshold T_1 the number of extrema would be 4. At threshold T_2 the number of extrema would be 6. The vector (4,6) could then be used as a texture feature. The authors claimed that by taking logarithms of the grey levels the method is made invariant to changes in lighting, while looking at ratios of extrema makes the method invariant to the size of the texture. In addition, although the authors only tested the method using a one-dimensional scan-line, they proposed that the method could be made invariant to rotation by counting extrema in several directions.

Mitchell et al. (1977) tested their max-min method on 8 textures from Brodatz' (1966) texture album. Forty-nine, 64×64 samples of each texture were taken, 36 of which made up the training set, with the remaining 13 used as a test set. Six features were computed for each of 2 methods of selecting the thresholds. 79.8% of the test images were classified correctly using a 3-nearest neighbour decision rule. They compared this with 66.3% correct classification using 6 features from grey-level co-occurrence matrices.

2.3.7 Filter Masks

Harwood et al. (1985) introduced a method of texture classification using local rank correlation. Most of the techniques described in this section so far are sensitive to noise, changes in lighting conditions and other monotonic shifts in grey level. The authors claimed that by using the rank of grey levels, a texture classification technique which is invariant to such grey-level shifts, could be derived. The method of Harwood

et al. (1985) is based on the masks used by Laws (1979) although the masks, as well as the grey levels are ranked.

e.g.

$$\textit{Original Laws Mask} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

$$\textit{Ranked Mask} = \begin{bmatrix} 2.5 & 5.0 & 7.5 \\ 1.0 & 5.0 & 9.0 \\ 2.5 & 5.0 & 7.5 \end{bmatrix}$$

Spearman's rank correlation coefficient is used to measure the correlation between the ranked masks and the ranked neighbourhoods for all neighbourhoods in an image. Then, to compare two samples, the distributions of the correlations of the samples are compared using a likelihood ratio test. The classification of a sample is based on computing, for each texture class, the median likelihood ratio test statistic of comparisons of the sample with all samples in the class. The sample is then assigned to that class for which the test statistic is a minimum.

Harwood et al. (1985) tested their local rank correlation method on 6 texture classes taken from Brodatz' (1966) texture album and achieved a maximum 100% success rate with 120×120 images, and 89% with 60×60 images using 3×3 rank masks. They compared this with Laws' texture energy method (1979) with histogram flattened images where the maximum correct classification rate was 93% for 120×120 images and 90% for 60×60 images. The authors suggested that an extension to their method might be to use ranked co-occurrence matrices as well as ranked masks.

2.4 Structural Approaches

Structural approaches assume that textures are made up of regions or primitives which appear in regular repetitive arrangements. They therefore tend to be more appropriate for man-made textures, although several authors, e.g. Wang et al. (1981), Zucker et al. (1975) have used structural techniques to classify Brodatz' (1966) texture images and terrain samples from LANDSAT images.

The first step in describing a texture is to identify the primitives. A primitive is a connected set of resolution cells characterised by a list of attributes. The simplest primitive is a single pixel with its grey level as its attribute. More commonly, primitives are sets of pixels all with the same or similar grey level or edge direction. Attributes may simply be the grey level of the region although other possibilities are size, shape and direction. Once the primitives and their attributes have been identified, the next step is to determine the placement rule, i.e. the spatial relationship between the primitives.

The division between structural and statistical methods is not always clear cut. Haralick (1979) divided structural methods into purely structural and structural-statistical. His definition of structural-statistical approaches was those where the method is structural in the sense that the primitives are explicitly defined, but statistical in that the spatial relationship is measured by probabilities. Haralick (1979) also divided textures into weak textures and strong textures. Weak textures are those which have only weak spatial relationships between the primitives, whereas strong textures have nonrandom spatial relationships. Several of the methods Haralick (1979) classified as structural-statistical are included within the statistical approaches section in this study. For example, weak texture methods include run-length approaches, and strong texture methods include the relative extrema approaches of Mitchell et al. (1977) and generalised co-occurrence matrices (Davis et al. (1979)).

2.4.1 Pure Structural Approaches

Haralick (1979) mentioned only 3 purely structural approaches. These were those of Carlucci (1972), Zucker (1976) and Lu and Fu (1978).

Carlucci (1972) examined textures from a linguistic point of view where languages, whose sentences are descriptions of textures, were defined. In Carlucci's (1972) model of texture the primitives considered were line segments and open and closed polygons, and the placement rules were based on a tree-like structure.

Zucker (1976) proposed the idea of texture modelling in terms of an ideal texture and its transformations. According to Zucker (1976) a real texture is a distortion of an ideal texture. The underlying ideal texture has a nice representation as a regular graph in which each node is connected to its neighbours in an identical way. Each node corresponds to a cell in a tessellation of the plane. The ideal texture is transformed by distorting the primitive at each node to make a realistic texture.

In Lu and Fu's model (1978) the primitives are pixels and a texture is divided into 9×9 windows. The spatial structure of the resolution cells in the window is expressed as a tree and the arrangement of grey levels to the resolution is given by the rules of a stochastic tree grammar. Lu and Fu (1978) used this tree grammar for both texture synthesis and texture discrimination. They classified 400 9×9 windows of 4 Brodatz' (1966) textures. Of these 400, 30 were misclassified.

2.4.2 Structural-Statistical Approaches

Other approaches which may perhaps be described as structural-statistical are those where the spatial relationship between pixels is not defined at all. For texture classification, extracted primitives can be used as features, and standard classification techniques can then be used. Wang et al. (1981) extracted primitives by thresholding, resulting in a binary image, using one of three methods: a) fixed percentage (25%) above threshold, b) histogram peak sharpening and c) superslice.

Once the primitives were selected, various properties such as area, perimeter, compactness, eccentricity, direction and average grey level, were measured. The means and standard deviations of these properties were computed, resulting in first-order statistics. Second-order statistics were also computed by identifying neighbours of the primitives and constructing co-occurrence matrices of attributes. From these co-occurrence matrices, features such as angular second moment, entropy, inverse difference moment and contrast were computed.

Wang et al. (1981) tested their methods on 128×128 Brodatz' (1966) and LANDSAT images. They used 4 samples of each of 4 Brodatz' (1966) textures and 3 terrain types. No quantitative results were given, but the authors claimed that the features could discriminate between texture types. Two features were needed for the Brodatz' (1966) images, and only one was needed for the terrain samples. The authors found that the simplest method of thresholding (fixed percentage) produced the best results, with the second-order statistics giving slightly better results than the first-order statistics.

A similar technique was used by Hong, Dyer and Rosenfeld in 1980, although they used an edge-based approach to extract primitives. An edge-detection operator was applied to the image, followed by thresholding to eliminate weak edges, and nonmaximum suppression to thin edges. Antiparallel edges were then paired and region interiors were filled. Features, similar to those used by Wang et al. (1981) were then computed for the resulting primitives.

Hong et al. (1980) used similar data sets to Wang et al. (1981) to test their methods. They also used 128×128 images of 4 Brodatz' (1966) textures and 3 terrain types. They also measured 6 features from their primitives and computed the means and standard deviations of each. Again no quantitative results were published but the authors claimed that mean of area, mean of perimeter and standard deviation of average grey level could separate Brodatz' (1966) images, and standard deviation of area could discriminate terrain samples.

Tsjui and Tomita (1973) extracted primitives to segment a picture consisting of regions of different textures. The primitives ('atomic regions') were defined to be sets of points with almost the same grey level. Once the atomic regions were extracted, properties such as size, shape, position, colour and average grey level were measured. Histograms of the property values were constructed and the textures identified as peaks in these histograms. The picture was then segmented by labelling the primitives with the names of the textures to which they belonged. Zucker et al. (1975) extended this idea to more realistic scenes, extracting their primitives using spot detectors.

Ehrich and Foith (1976) extracted primitives using a one-dimensional scan-line. A relational tree representation recursively partitions the function at the relative minima into finer partitions. Textural features, such as segment contrast, i.e. the mean of the

differences between relative maxima and minima over the partitions at a given level of the tree, can be extracted at any level of the tree.

Conners and Harlow (1980a) formulated a structural approach to texture analysis based on the co-occurrence matrix. They used the contrast feature from co-occurrence matrices to characterise the placement rules and the primitives of periodic textures, and proposed a model based on mathematical tiling theory. They concluded that a structural approach to texture could be formulated based on co-occurrence matrices although the contrast feature did not provide all the textural information contained in the matrix (Conners and Harlow (1980b)).

2.4.3 Conclusions

Structural techniques for texture analysis have not been widely investigated. Of the approaches that have been used, most fall into the category of what Haralick (1979) defined as weak texture measures, and what other authors have classified as statistical techniques. Few purely structural approaches have been proposed, especially in recent years. The main reason for this is that natural textures tend not to have the nice, regular patterns required for such methods. More recent research has concentrated on developing structural models (Ahuja & Rosenfeld (1981a&b), Wechsler & Citron (1980)). These modelling techniques will be discussed in the next section.

2.5 Modelling Approaches

Texture modelling techniques involve constructing models to specify textures. The object is to capture the intrinsic character of the texture in a few parameters. The models can then be used to generate synthetic textures or to describe an observed texture. Texture models can be broadly classified into stochastic models such as those based on Markov and Gibbs random field models, and structural models, which specify the manner in which an image is generated but have no probabilistic description. Structural models which have been proposed include random walk models (Wechsler and Citron (1980)) and random mosaic models (Ahuja and Rosenfeld (1981a&b),

Schachter, Rosenfeld & Davis (1978)).

Early literature in texture modelling concentrated on methods for texture synthesis, but recently there has been more interest in using texture models to segment texture. Relatively little has been published on texture classification using modelling techniques.

2.5.1 Gibbs and Markov Random Field Models

Markov random field and Gibbs models seem to be the most popular models used to describe texture. The origins of Gibbs distributions lie in physics and statistical mechanics literature. Ising (1925) used a special Gibbs distribution, now known as the Ising model, to describe the magnetic properties of ferromagnetic materials. The source of the revived interest in Gibbs distributions, especially in the context of image modelling and processing is an important result known as the Hammersley-Clifford theorem. This result, proved in the 1970's by several researchers, establishes a one-to-one correspondence between Markov and Gibbs random fields.

An image is assumed to consist of a square array of pixels or sites (i,j), $1 \leq i \leq n$, $1 \leq j \leq n$. These sites are numbered row by row from 1 to n^2 starting in the upper left. Each site in an image is assigned a 'colour', where colour may be grey level in the texture modelling context, or texture type in classification and segmentation contexts. The idea of a neighbourhood is central to Gibbs and Markov models and the probability of colour i at site t will depend on the neighbours of t. In a first-order model, t will have 4 neighbours (horizontal and vertical), in a second-order model the four diagonal sites will also be neighbours. Figure 2.3 shows 1st to 5th order neighbours of site t. Figure 2.3 (a) is a relative numbering system which defines sites relative to a site t. Figure 2.3 (b) marks all pixels which are distance k away, for k from 1 to 5.

t:-11	t:-7	t:-6	t:+8	t:+12
t:-9	t:-3	t:-2	t:+4	t:+10
t:-5	t:-1	t	t:+1	t:+5
t:-10	t:-4	t:+2	t:+3	t:+9
t:-12	t:-8	t:+6	t:+7	t:+11

(a)

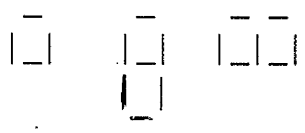
5	4	3	4	5
4	2	1	2	4
3	1	t	1	3
4	2	1	2	4
5	4	3	4	5

(b)

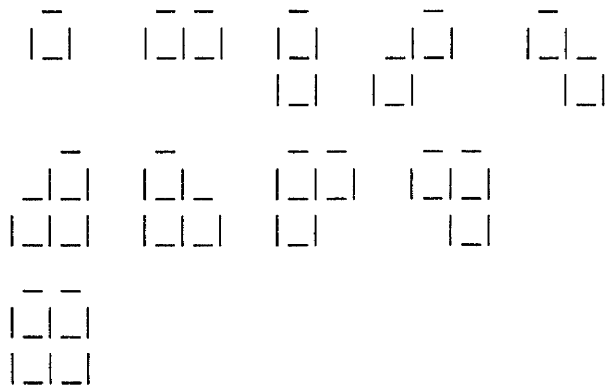
Figure 2.3 Neighbours of site t.

A clique is defined to be a set of sites where every pair of sites are neighbours, e.g.

1st-order cliques



2nd-order cliques



(N.B. individual sites are defined to be cliques.)

A Gibbs model then defines the probability of colour x at a particular site to be:

$$P(X=x) = \frac{\exp[-U(x)]}{Z}$$

where U(x) is called the energy function and is given by:

$$U(x) = \sum_{c} V_c(x).$$

$V_c(x)$ is a clique function and depends only on the colours in clique c.
Z is a normalising constant called the partition function and is given by:

$$Z = \sum_{x} \exp[-U(x)].$$

Various clique functions have been proposed. Derin and Elliott (1987) proposed taking

$V_c(x) = -\theta$ if all colours in a clique are the same

$= +\theta$ otherwise.

For single pixel cliques, they defined $V_c(x) = \alpha_k$.

If only cliques of size 2 are considered the energy function can be written in the form:

$$U(x) = \sum_{t=1}^{M} F(x_t) + \sum_{t=1}^{M} \sum_{r=1}^{K} H(x_t, x_{t+r})$$

where H(a,b) = H(b,a), H(a,a) = 0, and K depends on the size of the neighbourhood around each site.

Writing the Derin-Elliott model in this form,

$$F(x_t) = \alpha x_t \quad \textit{and} \quad H(x_t, x_{t+r}) = \theta_r I(x_t, x_{t+r})$$

where I(a,b) = -1 if a=b

= +1 if a≠b.

For 2nd-order models there are 4 parameters - θ_1, θ_2, θ_3, θ_4 where θ_1 influences pixels in the horizontal direction, θ_2 influences pixels in the vertical direction and θ_3 and θ_4 influence pixels in the diagonal directions.

Other possible models are the auto-binomial model where

$$F(x_t) = \alpha x_t - \ln[\frac{(G-1)!}{x_t!(G-1-X_t)!}]$$

$$\textit{and } H(x_t, x_{t:+r}) = \theta_r x_t x_{t:+r}$$

and the Gaussian Markov random field model where

$$F(x_t) = \frac{(x_t-\mu_t)}{2\sigma^2}$$

$$\textit{and } H(x_t, x_{t:+r}) = \frac{-\theta_r(x_t-\mu_t)(x_{t:+r} - \mu_{t:+r})}{\sigma^2}$$

One approach to the classification problem would be to fit a Markov random field model to each texture type using sample images to estimate parameters, then apply standard maximum likelihood methods to make a decision. The unknown normalising constant Z, however, makes this unworkable. In 1988 Chen suggested a more practical approach which simply uses the parameters estimated from samples of each texture type as feature vectors. For example, fitting a second-order model would mean that 4 parameters would be estimated and four-dimensional feature vectors would be used in decision-making. Any of the standard classifiers such as Fisher's canonical variate analysis or maximum likelihood can be used. One advantage of this method is that the models need not necessarily fit the data well, all that is required is that the models for each class are different.

The main problem with Gibbs and Markov random field models is how to estimate these parameters. Several methods have been suggested for doing this. These include Besag's coding method (1974), least square error method (Derin and Elliott (1987)), logit model fit method (Chen (1988)) and maximum pseudolikelihood (Besag (1974)). Of these, only the coding method and the maximum pseudolikelihood

method can easily be used in images with more than 2 grey levels.

The coding method is based on Besag's (1974) coding scheme. For a second-order model, the sites of an image are partitioned into 4 codings, as shown below, where the rows and columns of an array represent an image.

1 2 1 2 ... 1 2 1 2
3 4 3 4 ... 3 4 3 4
.
.
.
1 2 1 2 ... 1 2 1 2
3 4 3 4 ... 3 4 3 4

All pixels labelled j are used to estimate the jth parameter, j=1,2,3,4.
Let S_j be the set of sites labelled j in the above diagram. The coding method defines the estimate θ as follows:
First, find parameter θ_j which maximises $L_j(\theta)$, the loge-likelihood in coding j.

$$L_j(\theta) = \sum_t \ln [P(X_t=x_t \mid X_{\delta t}=x_{\delta t}]$$

where δt refers to all sites in the neighbourhood of site t, excluding t.

This can be written:

$$\sum_t \ln \frac{\exp(-[w(x_t, x_{\delta t})]^T)\theta}{\sum \exp(-[w(x_t, x_{\delta t})]^T)\theta}$$

where $w(x_t,x_{\delta t}) = [w_1(x_t,x_{\delta t})\ w_2(x_t,x_{\delta t})\ w_3(x_t,x_{\delta t})\ w_4(x_t,x_{\delta t})]^T$
and $\theta = [\theta_1\ \theta_2\ \theta_3\ \theta_4]^T$.

For the Derin-Elliott model the vector w is

$w_r(x_t,x_{\delta t}) = I(x_t,x_{t-r}) + I(x_t,x_{t+r})\quad 1 \le r \le 4$

$L_j(\theta)$ is then optimised using Newton-Raphson and the coding method estimator is defined as:

$$\theta = \frac{1}{4}\sum_{j=1}^{4} \theta_j$$

The coding method is basically a maximum likelihood estimation method that yields parameter estimates that maximise the conditional joint distribution of a subset of the random variables in the field, conditioned on the rest of the field. It requires the solution of a set of nonlinear equations and is therefore cumbersome and difficult to use reliably. Furthermore, depending on the order of the neighbourhood system a number of different estimates are obtained from a single realisation and an established method to combine these estimates into one does not exist. An alternative method suggested by Besag (1974) is the maximum pseudolikelihood method:

$$P(X=x;\theta) = \frac{\exp[-U(x,\theta)]}{Z(\theta)}$$

$$\textit{where } U(x,\theta) = \sum_t w(x_t, x_{\delta t})^T \theta$$

$$\textit{and } Z(\theta) = \sum \exp[-U(x,\theta)]$$

Suppose we are given a single sample x. The standard approach to estimate θ is to maximise the likelihood, i.e. choose θ to maximise P(x,θ). The problems of computing Z(θ) make this is infeasible. Besag's pseudolikelihood function is given by:

$$PL(x,\theta) = \prod_{s \in S} \prod_{|\delta S} (X_s = x_s | X_r = x_r,\ r \neq s; \theta)$$

where δ_s is the boundary of S under the neighbourhood system determined by the energy U and $S|\delta_s$ is the complement of δ_s relative to S. The pseudolikelihood estimator is the θ that maximises PL(x,θ).

Thus rewriting $P(X_s=x_s|X_{\delta t},\theta)$, this becomes:

$$\frac{\exp[-w(x_t, x_{\delta t})^T]\theta}{\sum \exp[-w(x_t, x_{\delta t})^T]\theta}$$

where $\sum$ is summed over the grey levels.

Chen and Huang (1993) conducted two experiments to classify texture based on 2nd-order Markov random field models. In the first, their data set consisted of 64×64 images of 4 types of texture (16 of each) taken from Brodatz' (1966) texture album. They investigated the Ising model, the auto-binomial model and the Gaussian random field model. For the first two models the images were grey-level equalised to contain 4 grey levels. Besag's pseudolikelihood method was used to estimate the parameters

of the models. Three different classifiers were tested - nearest neighbour, quadratic and Fisher's. The best results (62/64) were obtained using Fisher's classifier with the auto-binomial model. In the second experiment the data consisted of four types of sandpaper. Models and classifiers tested were the same as the first experiment. Again Fisher's classifier produced the best results but using the Gaussian Markov random field model resulted in one less misclassification than the auto-binomial model (7/64 and 8/64). These results are better than most of those obtained using co-occurrence matrices and the authors claim that the features are easier to extract because they involve only one step. They suggest that both feature extraction and classifier design should be considered simultaneously in designing an optimal classification scheme.

Ohanian and Dubes (1992) compared 4 types of textural features for classifying texture. The features investigated were Markov random field parameters, Gabor multi-channel filtering features, fractal based features and co-occurrence features. The authors used these features to classify 4 types of images - fractal images, Gaussian Markov random field images, natural images of pieces of different types of leather and natural images of painted surfaces. All images were of size 32×32. The numbers of grey levels ranged from 4 for the MRF features to 256 for the Gabor features. The classifier used was a leave-one-out nearest-neighbour classifier. The authors concluded that co-occurrence features performed best and suggested they should be tried first when working with small images. They advised that MRF models should only be used for large images, and the Gabor features used by the authors are not recommended.

2.5.2 Random Mosaic Models

Random mosaic models use random mosaics to represent a textured region. A region A is first tessellated into cells A_j, j=1...m, and one of m colours is independently assigned to each cell according to a fixed set of probabilities, $p_1...p_m$, $\sum p_i=1$. The set of colours may correspond to values of any property, not necessarily grey level. Many types of process that generate random mosaics have been proposed. In the Checkerboard model, the origin and orientation of the axes are chosen randomly and the plane is tessellated into squares. The hexagonal and triangular model are similar but with hexagonal and triangular regions.

In the Poisson Line model, a Poisson process chooses points in the strip $0 \leq \theta \leq \pi$, $-\infty < \rho < \infty$. Each of these points defines a line of the form $x\cos\theta + y\sin\theta = \rho$, and these lines define a tessellation of the plane.

In the occupancy model, a Poisson process chooses points (nuclei) on a plane. These nuclei spread out to form 'Dirichlet cells' or 'Voronoi' polygons' consisting of all points in the plane nearer to it than any other nucleus.

In Bombing models, points are also dropped onto the plane by a Poisson process. Each point then expands into a region of specific size and shape (e.g. a circle of radius r with the point at the centre). The process divides the plane into a foreground (the bombed out region) and a background.

Various statistical properties of the grey levels in a random mosaic can be computed. The variogram of a random mosaic is the mean-squared grey-level difference between a pair of points distance d apart:

$$V(d) = E[(g(s) - g(s'))^2]$$

where s and s' are 2 points distance d apart and g(s) is the grey level at point s.

For cellular textures, letting z_i be a random variable describing the grey levels in subregion A_i with mean μ_i and standard deviation σ_i, $1 \leq i \leq m$, the variogram can be written in the form:

$$V(d) = E[(z_i - z_j)^2], \qquad i \neq j.$$

For i=j, $E(z_i - z_i')^2$ depends on the correlation between pixels in the ith subregion. It is commonly assumed that this correlation is of the form $e^{-\alpha d}$, where $\alpha \geq 0$.

Writing p_i for the probability that a point lies in the ith region, and P_{ij} for the probability that two points distance d apart lie in the ith and jth subregions, $1 \leq i \leq m$, then

$$V(d) = \sum_{i>j} E[(z_i - z_j)^2]\ [p_j\ P_{ij}(d) + p_i\ P_{ji}(d)] + \sum_i E[(z_i - z'_i)^2]\ p_i\ P_{ii}(d)$$

$$= \sum_{i>j} 2p_i\, p_j(\sigma_i^2 + \sigma_j^2 + (\mu_i-\mu_j)^2(1-\omega(d))$$
$$+\sum_i 2p_i\sigma_i^2(1-\exp[-\alpha_i]\ d)\ (p_i(1-\omega(d))+\omega(d)).$$

where ω(d) is the probability that two points distance d apart are in the same cell. For Poisson Line and Checkerboard models ω(d) can be calculated and theoretical variograms can therefore be computed.

Schachter et al. (1978) compared theoretical and actual variograms using Poisson Line and Checkerboard models for various samples taken from Brodatz' (1966) texture album. They found that the Poisson Line model gave better fits for more random textures. In this study no attempt was made to classify textures but the value of V(d) at different distances could possibly be used as texture features. In 1981 Ahuja and Rosenfeld (1981a) also studied random mosaic models. They mentioned various properties such as autocorrelation, edge density and variogram that can be computed for mosaic models. No results were presented in this paper but in another publication (Ahuja & Rosenfeld (1981b)) they computed variograms from samples taken from Brodatz' (1966) texture album using six types of mosaic models. Based on their results, they claim that random mosaic models are powerful tools for texture analysis and synthesis. They claim the models are easier to specify than random field or time series models and because of the large number of possible models, they provide a means of controlling or matching many different texture features.

2.5.3 Random Walk Models

Wechsler and Citron (1980) suggested a procedure for texture classification based on random walks. A planar random walk is characterised by a particle moving in unit steps in one of four directions parallel to the x and y axes, assuming four-neighbour connectivity. The random walk is fully specified by defining the probabilities that a particle will leave a given pixel for any of its 4 neighbours. If the random walk is performed over some array A, the absorbing barrier Γ of A is defined as those pixels in A which have at least one neighbour outside A. The pixels belonging to Γ are called

boundary points B, while all the other points in A are mesh points M (Figure 2.4).

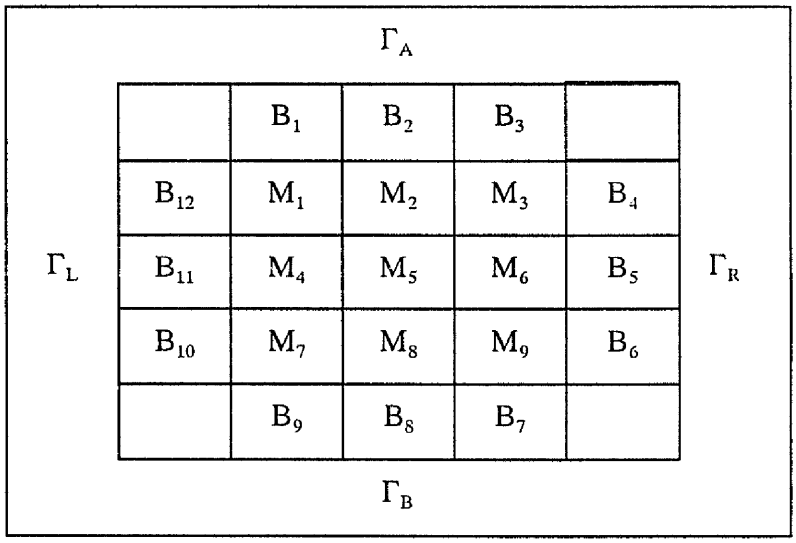


Fig. 2.4 Window A within which the random walk is performed.

Once a particle hits the absorbing barrier Γ it is absorbed and its random walk stops. The probabilities related to the random walks are as follows:

1) The probability of leaving a pixel belonging to a homogeneous array for any of its 4 neighbours is 0.25.

2) The probability of leaving Γ is zero.

3) The probabilities of leaving any mesh points are strictly positive, and for a non-homogeneous array the probabilities are defined such that particles are more likely to be shifted in the direction of smallest ascent or descent in the picture array. The probabilities are therefore based on absolute grey-level differences.

The probabilities of moving up, right, down and left from a mesh point M to its neighbours $N_k(M)$ are denoted $P_k(M)$, k=1,2,3,4 and are calculated by first computing:

$$DIF(k) = |F(M)-F[N_k(M)]|$$

$$AVER = \frac{\sum_{k-1}^{4} DIF(k)}{4}$$

$$DIF(k) = |DIF(k) - AVER| + 1$$

where F(M) and $F[N_k(M)]$ stand for the picture value at M and its neighbours $N_k(M)$ respectively.

The probabilities $P_k(M)$ are then defined as:

$$P_k(M) = \frac{[DIF\ (k)]^{-1}}{\sum_{k-1}^{4} [DIF\ (k)]^{-1}}.$$

The hitting distributions are defined as those probability functions which specify the likelihood that a random walk started at M_i will end at Γ_j. The authors estimated these hitting distributions by simulating random walks for the point M_5 and the absorption distributions were obtained by multiplying the hitting distributions by the number of times the random walk was simulated for that window. For texture classification, the authors took 5×5 windows and calculated gradients with magnitude m defined by:

$$m = [(\Delta x)^2 + (\Delta y)^2]^{\frac{1}{2}}$$

where $\Delta x = S_R - S_L$ and $\Delta y = S_A - S_B$

and S_A, S_R, S_B and S_L are the absorption probabilities corresponding to Γ_A, Γ_R, Γ_B and Γ_L respectively.

Histograms of magnitudes for the whole image are then constructed and compared with histograms of training images. An image is classified into the class with the highest correlation between histograms.

Wechsler and Citron (1980) tested their random walk model on 128 images of 32 texture types from Brodatz' (1966) texture album. Using a nearest-neighbour classification approach, they achieved 89.7% correctly classified for the leave-one-out test procedure.

A similar technique was used by Wechsler and Kidode in 1979 but they estimated absorption probabilities for all nine mesh points (Figure 2.4). They tested their model on 56 64×64images of 14 texture types taken from Brodatz' (1966) texture album. Again using a nearest-neighbour classification approach they achieved a maximum of 86% correctly classified for the leave-one-out test procedure. The authors compared their performance with that obtained using co-occurrence matrices and difference statistics. Using a nearest-neighbour classification and a leave-one-out approach they classified 78% and 79% respectively with the two techniques.

2.6 Conclusions

It is difficult to compare all the methods discussed above objectively because the results published are only applicable to the data sets on which they were tested. Also, for several of the approaches, no quantitative results are given. On the whole, two types of data sets have been used - Brodatz' (1966) texture images and/or satellite images, with the majority of researchers using Brodatz' (1966) images. Even there, comparison of the techniques is often not possible because of the different subsets of textures extracted from the album.

Almost all the studies have involved 64×64 images, most of which have had their grey-level histograms standardised by varying methods to contain 64 grey levels. Different studies have attempted to discriminate different numbers of categories and clearly the number of texture classes used will affect the results. Other important factors are the sizes of the training and test sets, when used, and the classification criteria applied. As an

example, in their pilot study, Weszka et al. (1976) had six samples of each of nine texture types, whereas in their main study they had 60 samples of each of only three types. Obviously fewer classes improve the chances of correct classification, as should larger numbers of samples in the training set. In addition, some studies such as those by Weszka et al. (1976) have obtained biased results by using the training set to test the method. Other studies, on the other hand, have used either the 'leave-one-out' technique, (e.g. Ohanian & Dubes (1992)) or have tested their methods on separate test sets (e.g. Haralick et al. (1973)).

Many authors have used the co-occurrence matrices of Haralick et al. (1973) as a base and have compared their results with those obtained using co-occurrence matrix features. For example, Davis et al. (1979) compared generalised co-occurrence matrices with grey-level co-occurrence matrices in classifying 5 texture classes taken from Brodatz' (1966) texture album. They claimed a success rate of 80% correctly classified with generalised co-occurrence matrices compared with only 50% with grey-level co-occurrence matrices. Similarly, Mitchell et al. (1977) used their relative extrema method to classify 8 classes of Brodatz' (1966) images and obtained 79.8% correct compared with 66.3% with co-occurrence matrices. Although they have obtained reasonably high success rates with their proposed techniques, perhaps these authors should have questioned why their results with co-occurrence matrices were so bad when other authors such as Haralick et al. (1973) have obtained much better results with, what would appear to be, data sets that are much harder to classify.

Sun and Wee (1983) tested their neighbouring grey-level dependence matrices on images of 3 types of terrain, similar to the images used by Weszka et al. (1976) in their main study, and obtained success rates of 79.4% and 85% with single features and pairs of features respectively. They compared their results with those of Weszka et al. (1976) - 76% for single features and 93% for pairs of features. Although the results for pairs of features are good, one criticism of these 2 studies is that the methods were tested on the training data. This is certain to lead to better results than if test images were used.

Wechsler and Citron (1980) obtained a correct classification rate of 89.7% when they classified Brodatz' (1966) texture images into 32 classes. Considering such a large number of classes, this result was encouraging, but the size of images used was 128×128, four times the area of images used by most authors. Harwood et al. (1985) also quoted

high rates of classification for large images using their local rank correlation method - 100% for 120×120 images of 6 classes from Brodatz' (1966) texture album, but the rate reduced to 89% when 60×60 images were used.

Ohanian and Dubes (1992) compared 4 different methods - co-occurrence matrices, Markov random field models, Gabor models and fractal models and found that for small images co-occurrence matrices were best, followed by fractal models. The authors advised that Markov random field models should only be used for large images. The authors did not state how large images should be before using Markov random field models but certainly for models of size 32×32 or less, they proposed that co-occurrence matrices are a better option for classification.

The best overall results have been obtained by Vickers & Modestino (1982) and Berry & Goutsias (1991). Vickers & Modestino (1982) obtained 98% correct classification on a test set of Brodatz' (1966) images using the whole co-occurrence matrix and a maximum likelihood classifier. Berry & Goutsias (1991) classified 100% of images, similar to those used by Vickers & Modestino (1982) by again using a maximum likelihood classification technique, either with the whole co-occurrence matrix or the whole neighbouring grey-level dependence matrix. Although neighbouring grey-level dependence matrices were as good as co-occurrence matrices for classifying 64×64 images, Berry & Goutsias (1991) found that when smaller images, i.e. 32×32 and 16×16 were to be classified, co-occurrence matrices gave better classification results. The results in these two studies are certainly impressive, one criticism of them, however, is perhaps in the data sets used. For each texture type Vickers & Modestino (1982) took one 512×512 image from Brodatz' (1966) texture album and from this extracted 16 64×64 images, 8 of which made up the training set and the remaining 8 formed the test set. Berry & Goutsias (1991) selected training and test sets in a similar way. By taking all their samples from one image they have eliminated any problems which might have occurred due to changes in lighting, focusing, orientation etc.. It is doubtful that such high success rates would have been achieved if training and test sets had been selected from a variety of images. Nevertheless, using the whole co-occurrence matrix, rather than extracting a subset of features, appears to give better rates of classification and is certainly worthy of further investigation.

The overall conclusion, therefore, seems to be that co-occurrence matrices do possibly contain more textural information than any of the other techniques investigated.

The problem is that most authors do not select enough information from these matrices. This agrees with the conclusion of Conners & Harlow (1980b) that the most commonly selected features of angular second moment, contrast, correlation and entropy do not contain all the textural information. Perhaps future studies should concentrate on some of the other textural features available from co-occurrence matrices. It is unlikely, however, that one subset could be found for all types of texture and, therefore, the best solution may in fact be to use the whole co-occurrence matrix.

There are two main criticisms of co-occurrence matrices. The first is that they are computationally relatively expensive. If this is a problem, then neighbouring grey-level dependence matrices may be an alternative option, but again without further feature extraction.

The second criticism is the difficulty in finding the best displacement to use. Currently there is no solution to this and most authors have either used a trial and error technique or have taken pixels a small distance (typically one pixel) apart and have averaged over four directions. Although some method is needed to make the techniques rotation invariant, whether averaging over 4 directions is the best solution, is questionable, especially since some research, e.g. Weszka et al. (1976) has found that using raw features (i.e. individual directions) gives better results than composite features (i.e averaging over directions).

In conclusion, although many techniques have been proposed in the past two decades, very few seem to be able to match the results of the co-occurrence matrices suggested by Haralick et al. in 1973. Many researchers have used angular second moment, contrast, entropy and correlation to classify images, and although other features such as cluster shade and cluster prominence have been proposed, very little work appears to have made use of the various other features. Considering the suggestion made by Conners and Harlow in 1980 that these four features do not contain all the important textural information, and more recent research using the whole co-occurrence matrix, this is perhaps surprising.

This chapter is not an exhaustive study of every method that has ever been considered for texture classification. Texture analysis is an active area of research, and modifications of existing methods, and new methods, are continually being suggested. To describe each of these would be a never-ending task. Instead, this chapter is an attempt to

outline some of the 'standard' methods for analysing texture, which have prevailed through time, such as those based on co-occurrence matrices, and to mention some more recent techniques which have successfully classified certain types of texture. How these techniques perform when applied to the rats' bones data used in this study, however, will be examined in the next chapter.

Chapter 3 Classification of Texture in Bone

3.1 Subjective Impressions

As stated in chapter 2, the main approaches to texture classification are the statistical approach, the structural approach and the modelling approach. Due to the irregularity of the images used in this study it was thought that structural approaches would not be appropriate. Similarly, it was thought that modelling such irregular images would not be feasible. The approach taken, therefore, was a statistical one. Probably one of the best known and most widely investigated statistical techniques for classifying texture is the co-occurrence matrix. Haralick et al. (1973) suggested 14 texture features that could be extracted from these matrices and in recent years this list has been extended. In the past two decades many authors have successfully classified various textures using co-occurrence matrix features. Other authors have taken classification rates using co-occurrence matrix features as a base and have compared rates obtained using their proposed methods with those obtained using co-occurrence matrices. Most authors have extracted a small subset of the available features from the matrices and, although it has been suggested (Conners and Harlow (1980b)) that certain of the features do not contain all the textural information available, in most cases the use of three or four texture features does produce fairly high rates of success. At the outset of the present study this seemed a sensible place to start - would features from co-occurrence matrices be of any use in classifying the textures found in the rats' bones?

The initial data set consisted of 25 64×64 pixel images with 256 (0....255) possible grey levels. The 25 images were made up of 5 of each of the first 5 texture types which were manually sampled from 512×512 pixel images such as those shown in figures 1.13, 1.14 & 1.15. At this stage all samples of the same texture type were taken from the same 512×512 square image. The problem of standardising the images is dealt with in Section 3.7. The images are shown in Figure 3.1 overleaf.

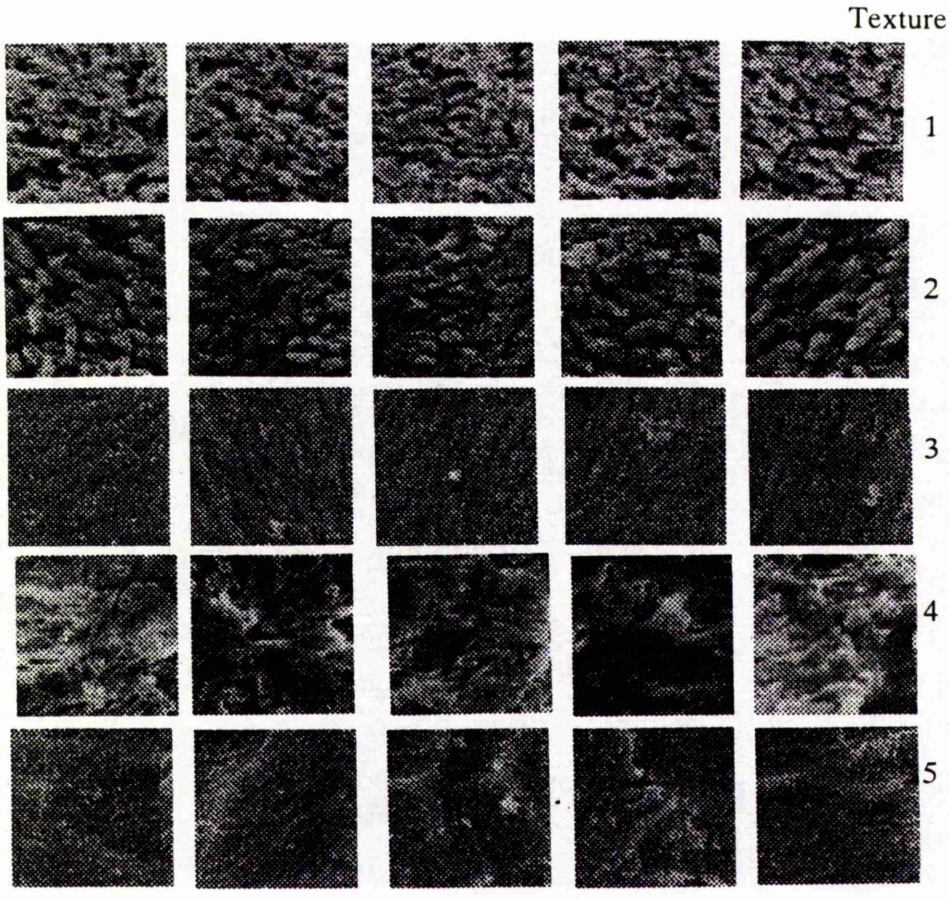


Fig. 3.1 64×64 images used in the initial study.

All the computation was carried out on a Sun SPARCstation using a package called Spider (Subroutine Package for Image Data Enhancement and Recognition). This is a Japanese government software product which contains a number of algorithms for texture analysis.

The first 13 of Haralick's features from co-occurrence matrices were computed in 4 different directions (0°,45°,90° and 135°) and 4 different distances (1,3,5,10 pixels) apart.

A subjective impression suggested that many of the measures would be of little use in distinguishing between textures. For most of the measures there seemed to be a wide variation of values amongst texture types and a lot of overlap between types, although angular second moment, contrast and correlation looked as if they might be useful in distinguishing between certain of the texture types. The differences were more noticeable with small displacements, either 1 pixel apart or 3 pixels apart. Table 3.1

below shows means and standard deviations of various features averaged over the 5 images of each type for displacement (1,0), i.e. horizontally adjacent pixels.

Type	ASM		Cont.		Corr.	
	Mean	St.dev.	Mean	St.dev.	Mean	St.dev.
1	0.00020	$6.29e^{-6}$	946.17	29.78	$3.87e^{-6}$	$5.75e^{-7}$
2	0.00021	$1.10e^{-5}$	806.25	80.51	$6.73e^{-6}$	$3.05e^{-7}$
3	0.00039	$4.34e^{-4}$	529.51	55.83	$6.50e^{-7}$	$1.85e^{-8}$
4	0.00044	$1.24e^{-4}$	382.30	160.72	$2.05e^{-5}$	$7.04e^{-7}$
5	0.00047	$6.79e^{-5}$	352.85	99.21	$2.19e^{-6}$	$5.73e^{-7}$

Table 3.1 Mean & standard deviation of various features taken from co-occurrence matrices.

Although this table only shows features measured from one displacement, and is based on a very small sample of images, it does appear that ASM is slightly higher for type 5 than both types 1 and 2, with type 1 having the lowest ASM and type 5 the highest. Contrast, on the other hand is considerably higher for types 1 and 2 than it is for types 3,4 and 5, although it should be noted that the standard deviation of type 4 images is high. Correlation is highest for type 2 but will depend very much on the displacement used.

In conclusion, at this stage, although there is a lot of variability in the texture measures, it looks as if it may be possible to discriminate between at least some of the texture types using a combination of ASM, contrast and correlation. A first impression, however, suggests that these measures may not be able to distinguish between texture types 4 and 5, and some other method may be needed to separate these two types.

3.2 Classification Algorithm

As an initial experiment, angular second moment, contrast and correlation based on co-occurrence matrices were examined more formally. This time, ten 64×64

windows of each of the 5 texture types, similar to those used above, were extracted from the 512×512 images and the above features were calculated for each window to form training sets. The displacements used were 1 pixel apart (horizontally & vertically) and 3 pixels apart (horizontally & vertically).

Test sets were constructed in a similar manner by extracting a further 5 windows of each texture type and combining them in a pairwise fashion to give 10 test images, i.e. a test set consisted of 5 images of one texture type and 5 images of a second type.

Although the aim of the project is to discriminate between all five texture types simultaneously, at this early stage of research only two texture types were considered at a time. The idea was to investigate whether any of the texture features could discriminate between **any** of the types and, if so, to build on this information to discriminate between all types. A routine was therefore written to classify each of the windows forming the test set as one of two types. Since the covariance matrices for each texture type seemed to vary considerably, quadratic discrimination was used (Lachenbruch (1975)).

The set of texture features measured on a training set image is treated as a feature vector $\underline{x}$. Initially we have two groups G_1 and G_2 and a set of feature vectors $x_{11},x_{12},...x_{1k}$ and $x_{21},x_{22},...x_{2k}$ which are subsets of the 14 co-occurrence matrix texture features computed from images belonging to each of the two types, G_1 and G_2 respectively.

An image with feature vector $\underline{x}$ is assigned to G_1 if

$$P(G_1|\underline{x}) > P(G_2|\underline{x})$$

and to G_2 if $P(G_1|\underline{x}) \leq P(G_2|\underline{x})$

i.e. the more likely texture, based on the computed texture features, is chosen.

(Since $P(G_1|\underline{x}) + P(G_2|\underline{x}) = 1$, the above simply reduces to whether $P(G_1|\underline{x})$ is greater than or less than 1/2.)

Images where $P(G_1|\underline{x})$ is equal to, or nearly equal to $P(G_2|\underline{x})$ should probably be considered to be grey areas, but this was not allowed for at this stage and all images were allocated to one of the two possible groups.

Then, assuming $\underline{x}|G_i \sim N_p(\underline{\mu}_i, \sum_i)$, using Bayes Theorem

$$P(\underline{x}|G_i) = \frac{1}{(2\pi)^{\frac{p}{2}} |\sum_i|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}D_i^2(\underline{x})\right]$$

$$\text{where } D_i^2(\underline{x}) = (\underline{x}-\underline{\mu}_i)^T \sum_i^{-1} (\underline{x}-\underline{\mu}_i)$$

and the above rule then simplifies to: decide G_1 if

$$\frac{1}{2}D_1^2\,\underline{x} - \frac{1}{2}D_2^2\,\underline{x} > -\frac{1}{2}\ln\frac{|\sum_1|}{|\sum_2|}\frac{P(G_1)}{P(G_2)}$$

In practice μ_i and $\sum_i$ are estimated from the training set images.

Then, if it is assumed that $P(G_1)=P(G_2)$, i.e. the prior probabilities for each texture are the same, the rule becomes:

decide G_1 if

$$\frac{1}{2}D_1^2(\underline{x}) - \frac{1}{2}D_2^2(\underline{x}) > -\frac{1}{2}\ln\frac{|\sum_1|}{|\sum_2|}$$

$$or \quad if \quad \frac{1}{2}\ln \frac{|\sum_2|}{|\sum_1|} - \frac{1}{2}D_1^2 + \frac{1}{2}D_2^2 > 0.$$

Otherwise assign to G_2.

This rule can easily be extended to cases where there are (K>2) groups. Assuming $x|G_i \sim N_p(\mu_i, \sum_i)$, (i=1,2,..,K), and assuming that the K groups are a-priori equally likely, the discriminant rule becomes:

assign to group i if:

$$-\frac{1}{2}\ln|\sum_i| - \frac{1}{2}(\underline{x}-\underline{\mu}_i)^T \sum_i^{-1} (\underline{x}-\underline{\mu}_i) > -\frac{1}{2}\ln|\sum_j| - \frac{1}{2}(\underline{x}-\underline{\mu}_j)^T \sum_j^{-1}(\underline{x}-\underline{\mu}_j)$$

$$j=1,2...K,$$
$$i\neq j.$$

$$i.e. \quad if \quad -\frac{1}{2}\ln|\sum_i| - \frac{1}{2}D_i^2(\underline{x}) > -\frac{1}{2}\ln|\sum_j| - \frac{1}{2}D_j^2(\underline{x})$$

$$j=1,2,...K,$$
$$i\neq j.$$

The assumption that all texture types have the same probability of occurring in a bone may not be valid. For example, younger rats are more likely to have a higher

percentage of the formative types 1, 2 and 3, whereas older rats' bones may consist more of types 4 and 5. The probabilities of each type are unknown, however, and are assumed to be equal here.

The results below show the number of images out of 10 which were classified correctly when applying this procedure. (1) indicates that displacements of 1 pixel apart were used, both horizontally and vertically; (3) indicates displacements of 3 pixels apart. Thus the feature vectors $\underline{x}$ in the following results are 2-dimensional.

3.3 Co-occurrence Matrices

	ASM (1)	ASM (3)	Contr. (1)	Contr. (3)	Correl. (1)	Correl. (3)
1v2	10	8	10	10	10	9
1v3	10	10	10	10	10	10
1v4	10	10	10	10	10	9
1v5	10	10	10	10	10	10
2v3	10	10	10	10	10	10
2v4	10	10	10	10	10	9
2v5	10	10	10	10	10	10
3v4	2	6	6	5	9	5
3v5	6	5	7	8	6	7
4v5	6	5	7	8	6	7
	84%	84%	90%	91%	91%	86%

Table 3.2 Number of images correctly classified using 64×64 co-occurrence matrices.

The results were quite good considering the small sample sizes used. This is especially true for images containing either texture type 1 or texture type 2, with nearly all the test images being correctly classified. These results confirm the subjective impression stated earlier that the differences in ASM and contrast between types 1 and 2 and the three other texture types are quite marked. The differences between texture types 3,4 and 5 are less obvious, with contrast and correlation being the best features

for separating them.

Ideally, we would like to be able to classify windows of size smaller than 64×64. To see whether the textural differences could be picked up in smaller windows, a similar experiment was carried out using 16×16 windows. This gave the following results:

	ASM (1)	ASM (3)	Contr. (1)	Contr. (3)	Correl. (1)	Correl. (3)
1v2	2	2	8	10	8	5
1v3	8	9	10	10	10	10
1v4	8	10	10	10	10	7
1v5	9	9	10	10	9	7
2v3	10	9	10	10	10	10
2v4	9	9	9	10	10	9
2v5	10	8	10	10	9	8
3v4	2	5	5	8	8	8
3v5	6	7	7	8	7	8
4v5	7	5	7	4	7	8
	71%	73%	86%	90%	88%	80%

Table 3.3 Number of images correctly classified using 16×16 co-occurrence matrices.

Not surprisingly the results are not as good with 16×16 windows as with 64×64. This is most noticeable from the results based on ASM, with only two images in the test set containing types 1 and 2 being correctly classified. The difference in contrast and correlation between texture types is still picked up using smaller windows. Contrast seems to be the best feature for discriminating between most texture types although correlation is better for discriminating between types 4 and 5.

Since the eventual aim is to discriminate between all texture types, the routine was adapted to include 2 textural features. Thus using the same method as previously but now with essentially 4 features (2 features at 2 different displacements -horizontal and vertical) all pairwise combinations of features were investigated. The best results were obtained with pairs which included contrast. The results are shown overleaf.

	ASM+Contr. (1)	ASM+Contr. (3)	Contr.+Correl. (1)	Contr.+Correl. (3)
1v2	6	9	10	5
1v3	10	10	10	8
1v4	10	10	10	10
1v5	10	10	10	10
2v3	9	9	10	8
2v4	10	10	9	9
2v5	10	8	10	8
3v4	4	4	8	4
3v5	8	4	9	5
4v5	7	6	7	4
	84%	80%	93%	71%

Table 3.4 Number of images correctly classified by 16×16 co-occurrence matrices using 2 features.

Using contrast and correlation with a displacement of 1 pixel apart has classified 93% of the images correctly. The main difficulty is in discriminating between types 4 and 5, with only 7 out of 10 images being correctly classified.

Next, even smaller windows were examined to see whether measuring the contrast and correlation could pick up a difference in images as small as 8×8. For each texture type 20 images of size 8×8 were extracted from the same 512×512 images as previously and their contrast and correlation computed. These sets of features were used as training sets. Ten test images were also extracted for each type and their contrast and correlation calculated. Test sets were constructed as previously by combining features from sets of 10 test images in a pairwise fashion to give 20 test images in each run of the programme. The following results were obtained:

	Contr.+Correl. (1)	Contr.+Correl. (3)
1v2	15	14
1v3	20	20
1v4	16	18
1v5	20	20
2v3	20	20
2v4	16	17
2v5	20	20
3v4	10	19
3v5	20	11
4v5	10	13
	83.5%	86%

Table 3.5 Number of images correctly classified using 8×8 co-occurrence matrices and 2 features.

Although the classification rate has dropped to 86%, the results are surprisingly good for such small windows. The difference in coarseness between types 1 and 2 is not picked up as well in such small images and there are still problems in discriminating between types 4 and 5. Otherwise, however, the discrimination routine works well even on windows as small as 8×8.

3.4 Difference Statistics

Since contrast has a corresponding feature based on difference statistics, rather than continuing the investigation of co-occurrence matrix features at this stage, it was decided to move on and examine difference statistics. The main advantage of difference statistics is that they are much faster to compute. In addition, although Conners and Harlow (1980b) found that co-occurrence matrices contain more textural information than difference statistics, Weszka et al. (1976) obtained better classification results with difference statistics than with co-occurrence matrices.

Contrast and mean were calculated for each of the 16×16 training and test set images, again using displacements of 1 and 3 pixels apart, horizontally and vertically. The following results were obtained:

	Contr.+Mean (1)	Contr.+Mean (3)
1v2	9	8
1v3	10	10
1v4	10	10
1v5	10	10
2v3	8	10
2v4	4	10
2v5	10	10
3v4	7	3
3v4	10	6
4v5	5	5
	83%	82%

Table 3.6 Number of correctly classified images out of 10 using 16×16 difference statistics.

For test sets containing type 2 images a displacement of 3 gives better discrimination. For all other texture types a displacement of 1 pixel is at least as good as a displacement of 3.

The discrimination is worse than that obtained from contrast and correlation based on co-occurrence matrices. Since contrast measured from difference statistics and co-occurrence matrices is essentially measuring the same thing the difference probably arises in using mean instead of correlation. Unfortunately there is no feature from difference statistics which corresponds to the co-occurrence correlation feature.

An examination of the grey levels of the images indicated that the mean grey level and the standard deviation of the grey levels vary between certain types. For example, the table below shows that the standard deviation tends to be lower for types 3, 4 and 5 than for types 1 and 2. Also, the mean grey level of type 1 images tends to

be higher than that of types 2, 3 and 4, and type 5 images tend to have a slightly lower mean grey level.

Type	Mean	St. Dev
1	122.9	42.4
2	106.5	35.7
3	106.8	17.1
4	105.6	18.1
5	103.3	17.2

Table 3.7 For each 16×16 image belonging to the training set the mean grey level and corresponding standard deviation has been computed. The table shows the average values for the 10 images of each type.

It was therefore thought that incorporating the mean and standard deviation of the grey levels into the discrimination routine may reduce the number of images being misclassified. The training and test sets still consisted of 4 features - contrast, 3 pixels apart, horizontally and vertically, mean and standard deviation. The results were as follows:

16×16 Difference Statistics

	Contr. (3) +Mean+St. Dev.
1v2	10
1v3	10
1v4	10
1v5	10
2v3	10
2v4	9
2v5	10
3v4	10
3v5	10
4v5	5
	94%

Table 3.8 Number of images correctly classified using difference statistics and grey-level mean and standard deviation.

Thus, by including mean and standard deviation in the discrimination nearly all the images are classified correctly. There is, however, still a problem in discriminating between types 4 and 5.

Looking at 8×8 images with training sets consisting of 20 images and test sets consisting of 20 images gave the following results:

8×8 Difference Statistics

	Contr. (3) +Mean+St. Dev.
1v2	18
1v3	20
1v4	18
1v5	20
2v3	19
2v4	18
2v5	20
3v4	15
3v5	12
4v5	10
	85%

Table 3.9 Number of images correctly classified by 8×8 difference statistics and grey-level mean and standard deviation.

Thus, again the differences are only really picked up between types 1 and 2 and the other 3 types. Discrimination between types 3,4 and 5 is fairly poor, with between 25% and 50% of images being misclassified.

3.5 Fourier Power Spectrum

Although Weszka et al. (1976) concluded that features based on co-occurrence matrices and difference statistics perform better than Fourier features and features from run-length matrices, for completeness these features were investigated at this stage.

In the case of the Fourier power spectrum, averages over wedge-shaped and over ring-shaped regions were examined.

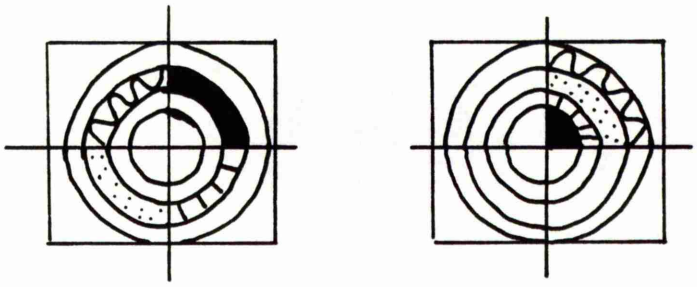


Fig. 3.2 The regions used in the Fourier power spectrum analysis.

The Fourier power spectrum was calculated in 16 regions and then averaged over size or direction, to give a feature set consisting of 4 values in each case.

	Radius (pixels)			
	(0-2)	(2-4)	(4-6)	(6-8)
A	0-π/2			Angle 1
n	π/2-π			Angle 2
g	π-3/2π			Angle 3
l	3/2π-2π			Angle 4
e				
	Size 1	Size 2	Size 3	Size 4

Fig. 3.3 The features used in the Fourier power spectrum analysis.

The results were as follows:

16×16 Fourier Spectrum

	Average over angles	Average over size
1v2	7	6
1v3	10	10
1v4	9	9
1v5	6	8
2v3	7	9
2v4	7	9
2v5	5	6
3v4	6	5
3v5	8	7
4v5	7	8
	72%	77%

Table 3.10 Number of images correctly classified.

As expected, in agreement with Weszka et al. (1976), Fourier features seem to perform less successfully than difference statistics and co-occurrence matrix features. Averaging over size gives marginally better results than averaging over direction but neither method discriminates well between types 1 and 2, and 3 and 4.

3.6 Run-length Matrices

In the examination of run-length matrices, once again 4 displacements were considered and the average and standard deviation of the features over the four directions computed. The best pair of features was grey-level distribution and run-length distribution. As the results below show, run-length matrix features discriminate well on test sets containing types 1 and 2, but are unable to discriminate between types 3,4 and 5.

A subjective look at the values of the features suggested that the standard deviations played a small part in the discrimination - the major differences appeared to

be in the means. The next step was therefore to include only the means of the features in the discrimination. The first four features, i.e. short run emphasis, long run emphasis, grey-level distribution and run-length distribution produced the results shown in Table 3.11 below. The improvement over using the mean and standard deviation of two features is, however, very slight. Again the discrimination is fairly good for sets containing types 1 and 2 although the features seem to be unable to discriminate between types 1 and 2. The features do however appear to be able to discriminate between types 3 and 4.

16×16 Run-length Matrices

	GLD+RLD (Mean+St. Dev.)	SRE,LRE,GLD & RLD (Mean)
1v2	8	4
1v3	10	10
1v4	10	10
1v5	10	10
2v3	10	9
2v4	10	9
2v5	10	9
3v4	6	9
3v5	5	7
4v5	1	4
	80%	81%

Table 3.11 Number of images correctly classified by run-length matrices.

The results obtained so far agree with Weszka et al. (1976) to the extent that difference statistics and co-occurrence matrices perform considerably better than Fourier power spectrum and run-length matrices. Computing contrast and correlation from co-occurrence matrices gives better results than any combination of features from difference statistics, but the best combination of features investigated is contrast based on difference statistics plus mean and standard deviation. (Corresponding features taken

from co-occurrence matrices would give similar results.)

There are several criticisms of this initial study that can be made. Firstly, discrimination between only two texture types at a time was considered. This will be of little use in 'real' situations where it is desirable to know to which of 5 or 6 classes of texture an image belongs. This criticism will be addressed in the next section. Secondly, images consisted of 256 grey levels. Co-occurrence matrices of sizes as small as 8×8 were computed and thus 64 pixels were allocated to a matrix which contained 65536 cells. Inevitably, the co-occurrence matrices were fairly sparse and it is perhaps surprising that the classification results were so good. In the remainder of work in this chapter the images were transformed to contain 64 grey levels and the majority of images considered were 64×64, giving 4096 pixels to be allocated to 4096 cells.

Only a small subset of texture features was considered, particularly in the case of co-occurrence matrices. Various studies (Conners & Harlow (1980b)) have suggested that ASM, contrast, entropy and correlation do not contain all the relevant textural information, and some authors (Conners et al. (1984)) have found some of the more recent features such as cluster shade and cluster prominence to be useful. In addition, only vertical and horizontal displacements were considered here, whereas in order to make the methods rotation invariant, many authors have computed averages over 4 (or more) directions (Haralick et al. (1973), Weszka et al. (1976)).

Finally, the images of each type are taken from the same bone. Although several authors such as Weszka et al. (1976) and Vickers & Modestino (1982) have constructed both training and test sets from one image, to make the results in this study applicable to any bone in the future, and therefore of any use medically, the data set has to be extended to include several bones, and indeed several rats.

Thus, although this initial analysis was useful to get a subjective impression of some methods which might be useful in future, no firm conclusions can be drawn at this stage. Each of the criticisms made above will be addressed in the analysis that follows.

3.7 Image Standardisation

In the previous section, mean grey level and standard deviation were found to be useful in discriminating between texture types. Although these features may provide

useful textural information - for example the standard deviation of a smooth, fairly uniformly grey, type 5 image is likely to be lower than for a type 1 texture which is speckled with white granules -the differences may occur simply because of the variation in conditions under which the images were saved. The manner in which the images are saved as regards brightness, focusing etc. is very subjective and may result in consecutive captures of the same image having a different distribution of grey levels. The problem becomes worse when images are saved from different bones on different days, etc., etc..

As an illustration of this the table below shows the difference in mean, standard deviation and contrast of 3 type 2 images, 2 of which are from the same bone.

	Mean	St. Dev.	Contr. (1)	Contr. (3)
Bone 1 image a	106.5	35.7	2033	2616
Bone 1 image b	119.0	20.8	789	757
Bone 2 image c	119.2	28.1	1661	1272

Table 3.12 Features computed from 3 images illustrating the differences in similar texture types.

Where training sets and test sets have been extracted from just one image of each texture type, this may in fact aid the discrimination - e.g. a type 1 texture may appear to have higher grey levels than a type 5 texture, but this may be due more to the conditions under which the images were captured rather than actual differences in the texture types. If the methods are to be useful in future it is obviously necessary to choose training and test sets which are representative of each texture type over all images and all bones. The subjective manner in which the images are saved, made necessary by the limitations of the scanning electron microscope, makes it impossible to save all images at exactly the same brightness without losing sharpness, and until a SEM with automatic focus becomes available, because the texture features are based on the grey levels of an image, it is imperative to find some way of standardising the grey levels.

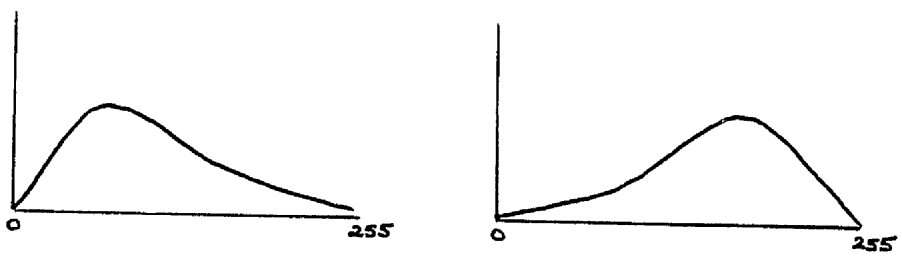
3.7.1 Histogram Equalisation

The most commonly used method of standardisation is histogram equalisation or histogram flattening (Hall et al. (1971), Haralick et al. (1973), Weszka et al. (1976), Vickers & Modestino (1982)).

Histogram equalisation transforms the grey scale histogram to a uniform distribution so that each grey level within the range desired occurs the same number of times.

Thus histogram equalisation defines a mapping of p grey levels in the original image into q grey levels in the new image such that the distribution of grey levels in the new image is uniform.

The number of grey levels, p, in the original image will normally be less than 256. Most images are either predominantly dark or predominantly bright, i.e. the histogram of grey levels is skewed to the right or left respectively.



Dark Bright

The mapping can expand the range of grey levels to 256 or more commonly, reduce the number of grey levels to 64.

There are various methods of performing histogram equalisation. Some methods do not divide grey levels, i.e. all pixels with the same grey level in the original image will have the same grey level in the transformed image, and therefore do not result in a perfectly uniform distribution.

Consider an 8×8 image with 6 grey levels which has to be transformed to contain 4 grey levels. Let p_i represent the frequency of grey level i in the original image, q_i

the desired frequency of grey level i in the transformed image and r_i the actual frequency of grey level i in the transformed image.

e.g.

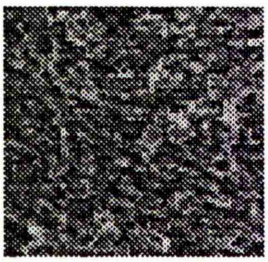
GL	0	1	2	3	4	5
p_i	12	16	13	10	8	5
q_i	16	16	16	16		
r_i	12	16	13	23		

As no division of grey levels is allowed, all that has happened in this case is that pixels with grey levels 4 and 5 in the original image have been assigned grey level 3 in the transformed images and, as can be seen, the resulting histogram is not in fact uniformly distributed.

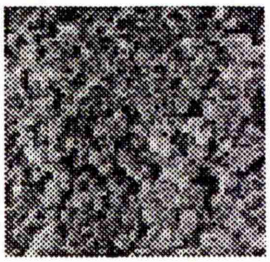
A second method which results in the frequencies q_i and hence a uniform distribution would randomly select 4 pixels with grey level 1 in the original image to have grey level 0 in the transformed image. It would then randomly select 4 pixels with grey level 2 to have grey level 1 in the transformed image, and so on.

In practice, usually there is very little difference between the two methods (Fig. 3.4) though care has to be taken when using either method. The first method cannot be used to expand the range of grey levels because no division of grey levels is allowed. The second method can produce strange results, e.g. if an image is half black and half white.

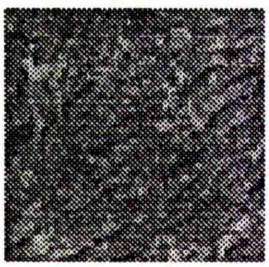
Some images of rats' bones and their histogram equalised transformations are shown in figure 3.4 below.



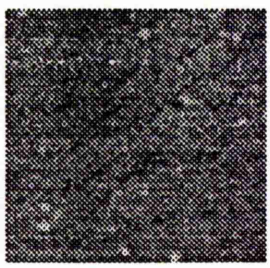
(a)



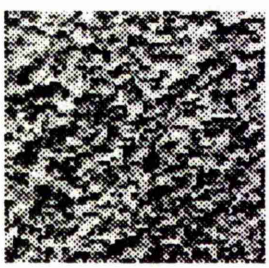
(b)



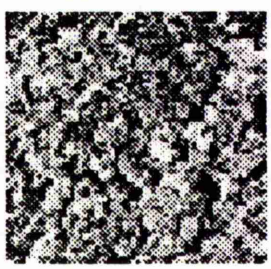
(c)



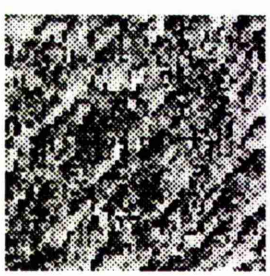
(d)



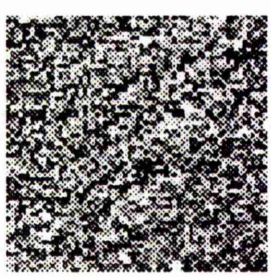
(e)



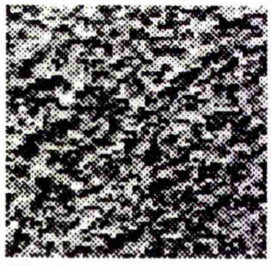
(f)



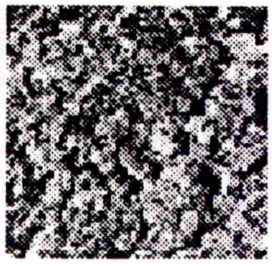
(g)



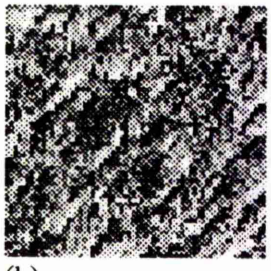
(h)



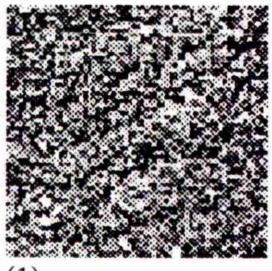
(i)



(j)



(k)



(1)

Fig. 3.4 Original images (a),(b),(c) and (d) and histogram equalised transformations method 1, (e)-(h), and method 2, (i)-(l).

Both the transformations have had the effect of standardising the images to some extent, so that they all consist of roughly the same number of pixels of each of the 64 grey levels. Using the second transformation, all images contain exactly the same grey levels, i.e. a 64×64 image contains 64 pixels with grey level 0, 64 with grey level 1 ... 64 with grey level 63. The texture in the images is still visible but there is now no variation in first order information such as mean and standard deviation. Differences in second order statistics such as co-occurrence matrices and difference statistics will still exist, but there is a loss of useful information.

3.7.2. Histogram Normalisation

Another possible method of standardising the images is by histogram normalisation, i.e. transforming the grey-level histogram to be normally distributed. Again there are at least two different ways of achieving this. One way is to specify the

desired mean and standard deviation, e.g. mean 32, standard deviation 12.5 (Vickers & Modestino (1982)) and transform IP(i,j) to JP(i,j) using the following algorithm:

$$JP(i, j) = (IP(i, j) - \mu_{old}) \times \frac{\sigma_{new}}{\sigma_{old}} + \mu_{new}.$$

An alternative method would be to specify the range of grey levels [0,d] in the transformed image, and use the algorithm:

$$JP(i, j) = (IP(i, j) - \mu_{old} + \sigma_{old}) \times \frac{d}{2\sigma_{old}}$$

With the first method all images of all texture types will again have the same mean and standard deviation resulting in loss of information. The second method has the advantage of not specifying the new mean and standard deviation but has the disadvantage of producing values of JP(i,j) outside the desired range, [0,d], e.g. < 0, where $IP(i,j) < \mu_{old} - \sigma_{old}$.

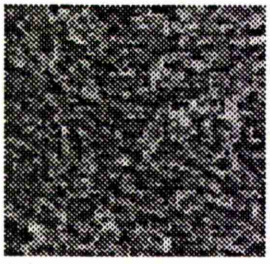
A possible modification of method 2, resulting in all pixels being transformed to within the desired range could be obtained by computing

$$KP(i, j) = \frac{(IP(i, j) - \mu_{old})}{\sigma_{old}}$$

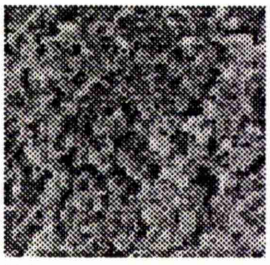
and then stretching the grey levels to cover the range [0,d] by computing

$$JP(i, j) = \left[KP(i, j) - \min\ [KP(i, j)]\right] \times \frac{d}{\max\ [KP(i, j)]\ -\ \min\ [KP(i, j]}$$

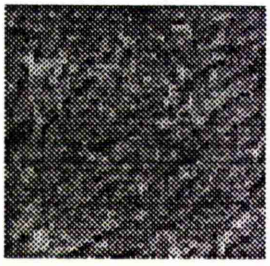
Figure 3.5 shows the resulting images from each of these three transformations.



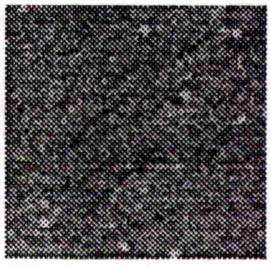
(a)



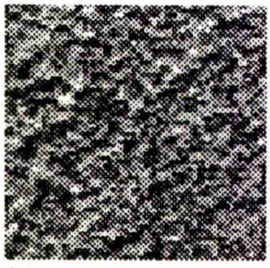
(b)



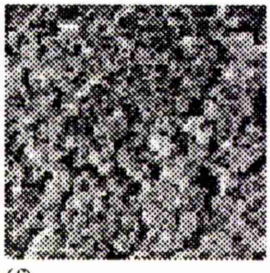
(c)



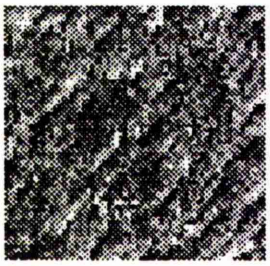
(d)



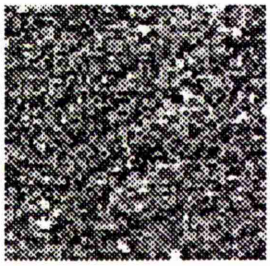
(e)



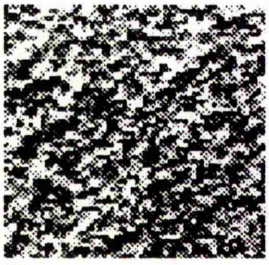
(f)



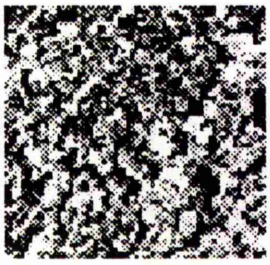
(g)



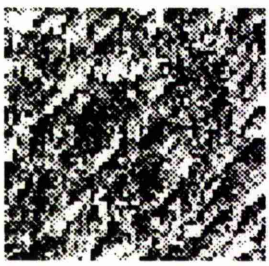
(h)



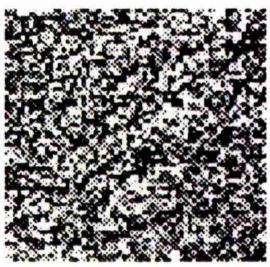
(i)



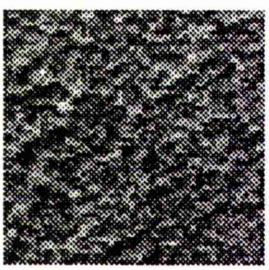
(j)



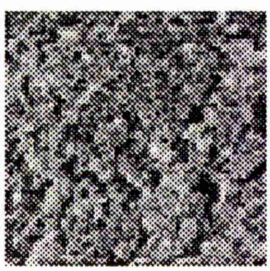
(k)



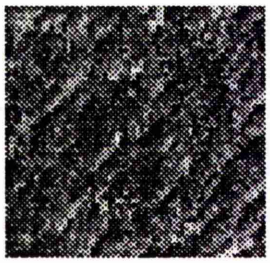
(l)



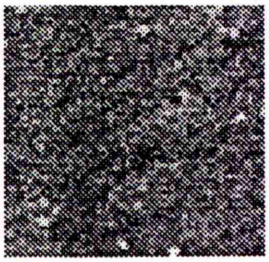
(m)



(n)



(o)



(p)

Fig 3.5 Original images (a)-(d) and transformed images (e)-(p), method 1, (e)-(h), method 2, (i)-(l), method 3, (m)-(p).

3.7.3 Conclusions

There appears to be very little to choose between the methods. Visually there is very little difference between the two types of uniform distribution standardisation although the histograms of grey levels are not the same. The second type of histogram normalisation results in fairly black and white images. This is because if the transformation results in a pixel with a grey level outwith the desired range (for example 0-63) it is assigned either grey level 0 or grey level 63. This occurs fairly regularly.

The method used by most authors seems to be to transform to a uniform distribution and therefore all the analysis that follows has been performed on histogram equalised images with division of the grey levels (i.e. method 2).

3.8. Training and Test Data

The aim is to have a database for each texture type which is representative of all images from all rats. Towards this aim new training sets were constructed. Training data for each texture type now consist of 10 64×64 images containing images from several bones. As stated above, all the images have now been histogram equalised to contain 64 grey levels. The original images and the corresponding histogram-equalised images are shown in figures 3.6 and 3.7 respectively. This new data set consists only of images of the three formative types, 1,2 and 3, plus the resorptive type 5. Type 4

was not included in this study because of the lack of clear images containing this texture type available at the time. Texture type 4 is included in future analysis, the results of which appear later in this chapter.

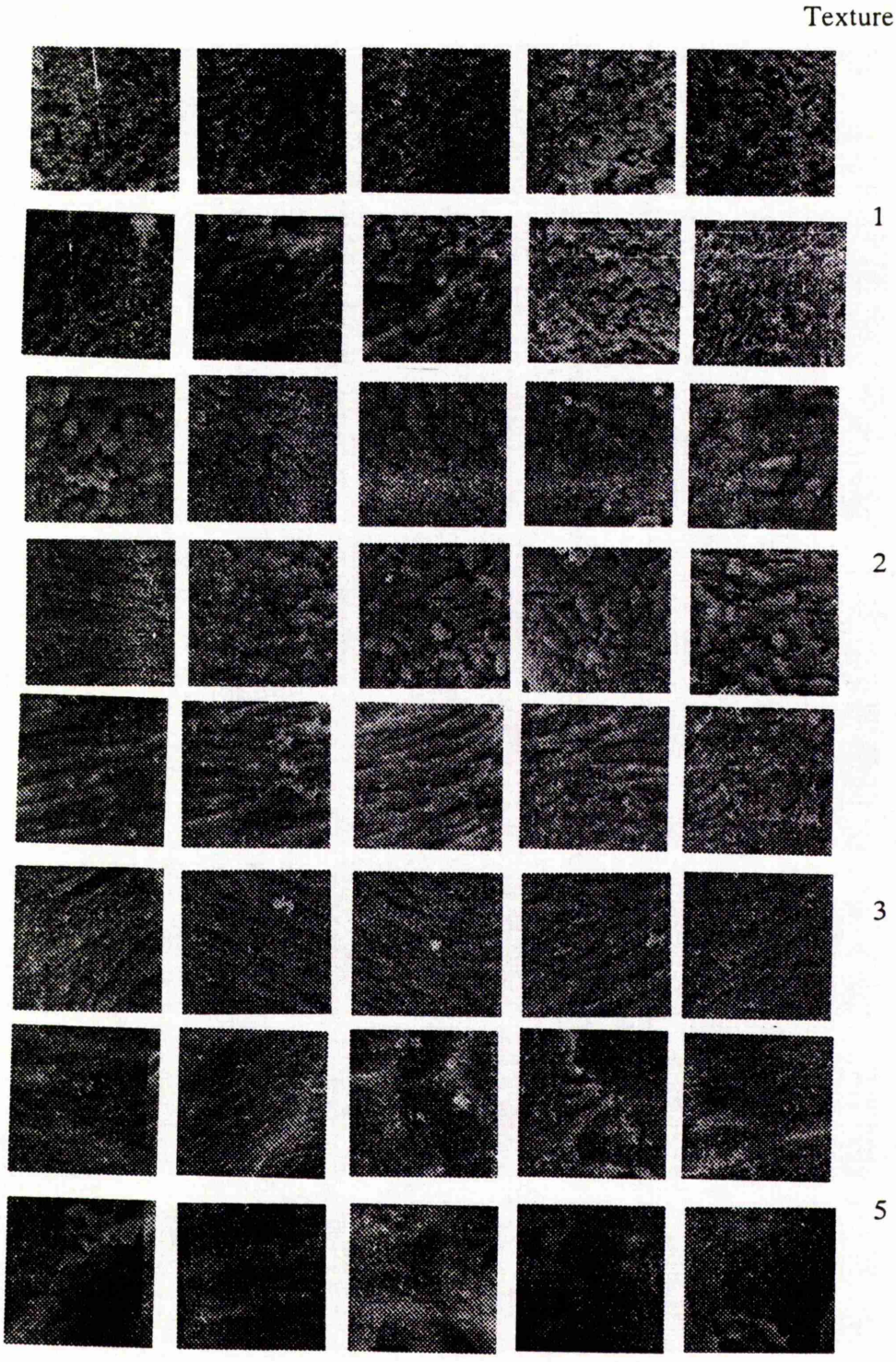


Fig. 3.6 New training set images before histogram equalisation.

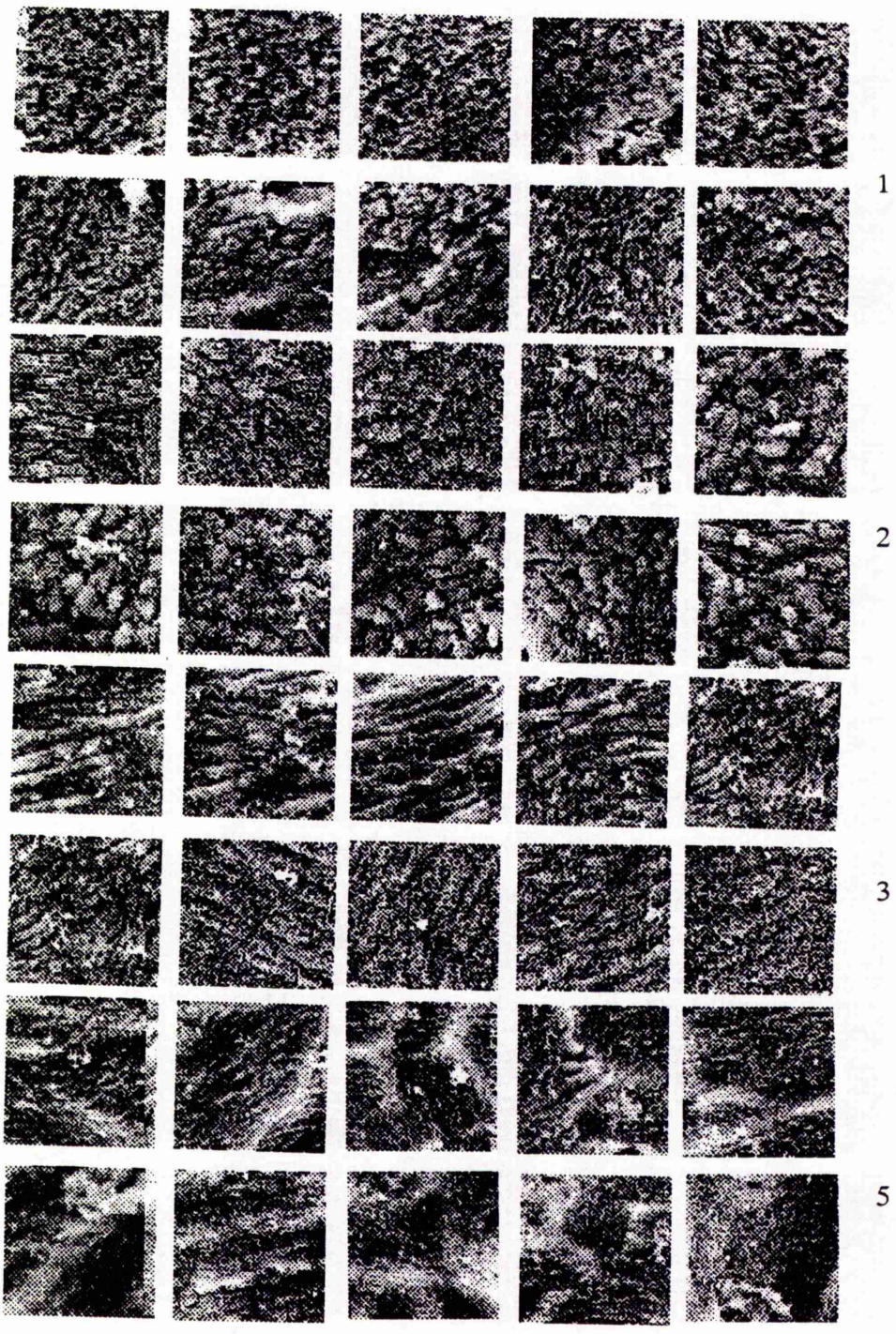
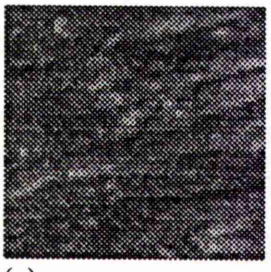


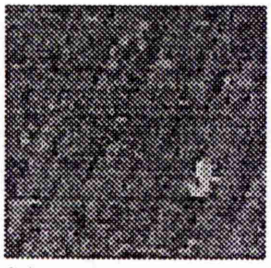
Fig. 3.7 New training images after histogram equalisation.

Due to the histogram equalisation it was no longer possible to use first-order statistics such as mean grey level and standard deviation to discriminate between types because these measures were now the same for all texture types (mean approximately 31.5, standard deviation approximately 18.475). Although fairly good classification results were obtained on 16×16 and even 8×8 images in the initial study without using the mean and standard deviation of the grey levels, the first order statistics certainly helped to improve the results. Other features therefore had to be re-examined to try to improve the classification. A finer edge than 64×64 windows will allow is desired, but, considering that most other authors who have classified Brodatz' (1966) images have only classified 64×64 images, it is perhaps unlikely that accurate discrimination will be possible on smaller windows of the natural images used in this study.

Also, because training sets now contain images from several bones with different orientations, features which are orientation independent are required. For example consider the two type 3 images in Figure 3.8.



(a)



(b)

Fig. 3.8 Type 3 textures with different orientations.

The human eye can see that these are the same texture type, only orientated differently. Textural features are required so that the computer will also recognise these as the same texture.

3.9. Co-occurrence Matrices and Difference Statistics re-visited

Returning to the paper by Haralick et al. in 1973, 64×64 windows were classified by measuring different subsets of the 14 features taken from co-occurrence matrices with a displacement of 1 pixel apart and in each of 4 directions (0°,45°,90°,135°). The mean and standard deviation of each feature over the 4 angles were then computed. Between 80% and 90% of images were classified correctly.

This technique was used to classify each of the 40 images in the new training sets in this study into one of four types -1,2 3 or 5. Because of the small number of images available at this stage, instead of testing the methods on test sets, jackknifing (or the leave-one-out method) of Lachenbruch and Mickey (1968) was used. Using this method, the discriminant rule is obtained from n-1 observations and the rule is used to classify the remaining case. Efron (1983) claimed that, although the method can have high variability, especially for small data sets, the method is nearly unbiased. Further investigation of other methods can be made once a method for classifying the images is found but it was felt that at this stage jackknifing would provide an adequate indication of the error rate.

As well as investigating a displacement of 1 pixel apart, co-occurrence matrices were also computed for displacements of 2,3 and 4 pixels apart. The features investigated were the first 12 suggested by Haralick et al. in 1973, the kappa statistic (Parkinnen & Selkainaho (1990)) and cluster shade and cluster prominence (Conners et al. (1984)). The figures in Table 3.13 show the number of images out of 40 classified correctly when using the resubstitution method, i.e. no jackknifing. Table 3.14 shows corresponding numbers using jackknifing.

	Displacement			
	1	2	3	4
ASM	25	25	18	19
Con	29	29	25	26
Corr	24	24	21	22
SofS	23	23	23	25
IDM	17	24	23	23
SAv	14	14	15	16
SVa	26	25	24	25
SEn	26	24	28	26
Ent	27	27	27	21
DVa	29	28	25	26
DEn	27	27	25	24
MofC	25	23	24	22
Kap	18	20	21	16
ClSh	28	26	23	22
ClPr	28	23	21	20

Table 3.13 Number of training images out of 40 correctly classified by co-occurrence matrix features.

[ASM=Angular Second Moment
Con=Contrast
Corr=Correlation
SofS=Sum of Squares
IDM=Inverse Difference Moment
SAv=Sum Average
SVa=Sum Variance
SEn=Sum Entropy
Ent=Entropy
DVa=Difference Variance
DEn=Difference Entropy
MofC=Information measure of correlation
Kap=Kappa
ClSh=Cluster Shade
ClPr=Cluster Prominence]

	Displacement			
	1	2	3	4
ASM	19	17	18	19
Con	23	20	19	24
Corr	19	22	20	21
SofS	18	18	18	19
IDM	14	20	19	18
SAv	10	10	11	11
SVa	24	19	18	16
SEn	21	16	20	21
Ent	20	19	19	15
DVa	23	20	22	24
DEn	25	20	20	19
MofC	18	19	20	15
Kap	12	15	17	6
ClSh	24	19	14	16
ClPr	25	17	15	19

Table 3.14 Number of images out of 40 correctly classified by co-occurrence matrix features using jackknifing.

The results are quite poor, with only a maximum of 72% of images being correctly classified when the discrimination rule was tested on the training set, and a maximum of 62% correctly classified when jackknifing was used. The results are, however, comparable to those of Weszka et al. (1976). In their main study, they obtained a maximum of 75% correctly classified using single features from co-occurrence matrices, and testing their method on the training set. In their study, larger training sets were used, with only three categories, each containing 60 samples.

The considerable drop in the number of correct classifications when jackknifing was used in this study is almost certainly at least partly due to the small size of the training sets used. It was hoped that a larger training set would improve the results. What can be seen from these results is that, on the whole, small displacements seem to

be best, with a displacement of one pixel resulting in the highest number of correct classifications. This agrees with conclusions made by Weszka et al. (1976). Another interesting point is that, as suggested by Conners and Harlow in 1980 the most commonly used features of ASM, contrast, correlation and entropy are not necessarily the most informative texture measures and in this data set the best results were obtained from difference entropy and cluster prominence.

A similar study was carried out with features taken from difference statistics and run-length matrices. The results are shown in Tables 3.15, 3.16 and 3.17.

	Displacement			
	1	2	3	4
Con	29	28	25	26
ASM	24	22	25	22
Ent	26	24	25	24
Mean	27	26	26	26

Table 3.15 Number of correctly classified images out of 40 using features from difference statistics without jackknifing.

	Displacement			
	1	2	3	4
Con	23	20	22	24
ASM	20	18	20	19
Ent	22	19	20	19
Mean	22	20	23	21

Table 3.16 Number of correctly classified images out of 40 using features from difference statistics with jackknifing.

	No Jackknifing	Jackknifing
SRE	16	11
LRE	19	10
GLN	16	10
RLN	15	9
RP	18	12

Table 3.17 Number of correctly classified images out of 40 using features from run-length matrices. [SRE=short runs emphasis, LRE=long runs emphasis, GLN=grey-level nonuniformity, RLN=run-length nonuniformity and RP=run percentage.]

The results for features based on difference statistics are similar to the corresponding features computed from co-occurrence matrices. As found by Weszka et al. (1976) and Conners & Harlow (1980b) run-length matrices perform less well than both co-occurrence matrices and difference statistics.

In an attempt to improve the classification results, features were combined. All pairwise classifications using co-occurrence matrices and a displacement of 1 were investigated. The greatest number of correct classifications without jackknifing was 90% using sum average and cluster prominence. This dropped to 57% when jackknifing was used. With jackknifing the best results were obtained from sum variance and either cluster prominence or cluster shade, sum average and difference entropy and measure of correlation and cluster shade. All combinations classified 60% correctly. The best combinations are shown in Table 3.18 (no jackknifing) and Table 3.19 (jackknifing).

Features	No. correctly classified
SAv + ClPr	36
IDM + ClPr	34
Ent + DEnt	34
DVa + Kap	34
Con + Kap	34
ASM + Con	33
ASM + SVa	33
IDM + SVa	33
SAv + SVa	33
Ent + ClPr	33
SVa + ClSh	33
IDM + ClSh	33
Ent + ClSh	33

Table 3.18 Number of images out of 40 correctly classified by pairs of features. No jackknifing.

Features	No. correctly classified
SAv + DEnt	24
SVa +ClPr	24
SVa + ClSh	24
MofC + ClSh	24
Cor + DEnt	23
SofS + DEnt	23
SAv + ClPr	23
Ent + ClSh	23
DEnt + ClPr	23
MofC + ClPr	23
SEnt + Kap	23

Table 3.19 Number of images out of 40 correctly classified by pairs of features. Jackknifing.

Thus, although using two features has improved the results considerably when the method is tested on the training set, there is no improvement in results when jackknifing is used, and in fact the maximum percentage of correctly classified images has dropped from 62% to 60%. Again, this is probably mainly due to the lack of images in the training set. Introducing a third feature improved the classification results for the training set, e.g. IDM, sum average and cluster prominence gave 95% correctly classified. When jackknifing was used, however, this figure dropped to 52%. The best result using jackknifing was 55% with sum average, difference entropy and cluster prominence, but this was worse than the best result using just 2 features.

In an attempt to find the best combination of features for classification, SPSS (Statistical Package for the Social Sciences) was used to carry out a stepwise discrimination analysis. Two methods of selecting variables were investigated. The first method was based on minimising the overall Wilks' lambda. Wilks' lambda is given by: $|W|/|T|$, where

$$W = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)(x_{ij} - \bar{x}_j)^T$$

i.e. within-groups sum-of-squares

$$\text{and } T = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x})(x_{ij} - \bar{x})^T$$

i.e. total sum-of-squares.

(x_{ij} is the ith point from class j, n_j is the total number of sample points in class j, $\bar{x}$ is the overall sample mean and k is the number of classes.)

At each step the variable that results in the smallest Wilks' lambda for the discriminant function is selected for entry. Each entry or removal of a variable is considered a step and the maximum number of steps permitted is twice the number of

independent variables. The significance of the change in Wilks' lambda when a variable is entered or removed from the model is based on an F statistic.

The second method used was Mahalanobis' distance. Mahalanobis' Distance D^2 is a generalised measure of the distance between two groups. The distance between groups a and b is defined as:

$$D^2_{a\,b} = (n-g) \sum_{i=1}^{p} \sum_{j=1}^{p} w_{i\,j}^{*} \, (X_{i\,a} - X_{i\,b}) \, (X_{j\,a} - X_{j\,b})$$

where p is the number of variables in the model, X_{ia} is the mean for the ith variable in group a and w_{ij}^{*} is an element from the inverse of the within-groups covariance matrix. When Mahalanobis' distance is the criterion for variable selection, the Mahalanobis' distances between all pairs of groups are calculated first. The variable that has the largest D^2 for the two groups that are closest (have the smallest D^2 initially) is selected for inclusion.

Covariance matrices were assumed to be equal, although in reality this was unlikely to be the case, and the discriminant rules were only tested on the training set. Although this was likely to lead to slightly biased results, at this stage the aim was simply to get an idea of variables which would be useful for discrimination, rather than carrying out any formal analysis. Taking each displacement separately, all variables were included in the analysis (i.e. 12 features from Haralick et al. (1973), cluster shade and prominence, and kappa). Features consisted of the means and standard deviations computed over 4 directions. The results, showing the variables entered and removed, plus the final classification tables, are shown below.

Displacement 1			
Wilks' λ		Mahal. Dist.	
In	Out	In	Out
ClPr(S)		SEn(M)	
Con(M)		Ent(S)	
Ent(S)		SofS(S)	
IDM(M)		ClPr(S)	
SEn(M)		IDM(S)	
IDM(S)			SofS(S)
Kap(M)		SVa(M)	
SofS(M)		Corr(M)	
Kap(S)			SVa(M)
DEnt(S)		IDM(M)	
DEnt(M)		DEn(S)	
MofC(S)		Kap(M)	
	Kap(M)	Kap(S)	
	DEn(M)	SVa(M)	
ASM(M)		DEn(M)	
Ent(M)			
			DEn(S)

(M = Mean, S = Standard Deviation)

[In=Features in classifier

Out= Features not in classifier]

Classification Tables

Wilks' λ

		Predicted			
		1	2	3	5
T	1	9	0	1	0
r	2	0	9	1	0
u	3	1	2	7	0
e	5	0	0	0	10

87.5%

Mahal. Dist.

		Predicted			
		1	2	3	5
T	1	10	0	0	0
r	2	1	8	1	0
u	3	2	1	7	0
e	5	0	0	1	9

85%

Table 3.20 Number of images correctly classified (displacement 1).

[ASM=Angular Second Moment
Con=Contrast
Corr=Correlation
SofS=Sum of Squares

IDM=Inverse Difference Moment
SAv=Sum Average
SVa=Sum Variance
SEn=Sum Entropy

Ent=Entropy
DVa=Difference Variance
DEn=Difference Entropy
MofC=Information measure of correlation
Kap=Kappa
ClSh=Cluster Shade
ClPr=Cluster Prominence
]

Displacement 2			
Wilks' λ		Mahal. Dist.	
In	Out	In	Out
Cor(M)		Kap(M)	
DEn(S)		SAv(S)	
IDM(M)		IDM(M)	
ASM(M)		SoS(M)	
ClSh(S)		Kap(S)	
Ent(M)		DEn(M)	
ClSh(M)			SAv(S)
SofS(M)		SEn(S)	
Ent(S)		ClPr(S)	
SEn(M)		ASM(M)	
SAv(M)		ClSh(S)	
			ASM(M)
		SEn(M)	
		SAv(M)	
			DEn(M)
		SVa(S)	
			SEn(M)
		DEn(M)	
		IDM(S)	
		Cor(M)	
		ASM(M)	
		Ent(M)	
			DEn(M)
			SAv(M)
		SEn(S)	
			SVa(S)
		Ent(S)	
		DVa(S)	
		ClSh(M)	
			IDM(S)

Classification Tables

Wilks' λ

		Predicted			
		1	2	3	5
T	1	8	1	1	0
r	2	0	10	0	0
u	3	2	3	5	0
e	5	0	0	0	10

82.5%

Mahal. Dist.

		Predicted			
		1	2	3	5
T	1	10	0	0	0
r	2	0	9	1	0
u	3	1	1	8	0
e	5	0	0	0	10

92.5%

Table 3.21 Number of images correctly classified (displacement 2).

Displacement 3			
Wilks' λ		Mahal. Dist.	
In	Out	In	Out
Cor(M)		MofC(S)	
Con(S)		SAv(S)	
Kap(M)		DEn(S)	
Kap(S)		Ent(S)	
ClSh(S)		MofC(M)	
IDM(M)			SAv(S)
DEn(M)		Kap(M)	
MofC(S)		IDM(M)	
Ent(S)		Con(M)	
ClSh(M)		Kap(S)	
SofS(M)		ASM(M)	
SEn(M)			SAv(M)
IDM(S)		DEn(M)	
SAv(S)		ClSh(S)	
		SVa(M)	
		ClSh(M)	

Classification Tables

Wilks' λ

		Predicted			
		1	2	3	5
T	1	10	0	0	0
r	2	0	10	0	0
u	3	0	0	10	0
e	5	0	0	0	10

100%

Mahal. Dist.

		Predicted			
		1	2	3	5
T	1	9	0	1	0
r	2	0	10	0	0
u	3	0	0	10	0
e	5	0	0	0	10

97.5%

Table 3.22 Number of images correctly classified (displacement 3).

Displacement 4			
Wilks' λ		Mahal. Dist.	
In	Out	In	Out
SEn(M)		Con(S)	
IDM(S)		Kap(M)	
SofS(S)		IDM(S)	
Kap(M)		SEn(S)	
		SVa(S)	
		SofS(S)	
		SVa(M)	
			SofS(S)
		DEn(S)	
		MofC(S)	
		Ent(S)	
		DVa(S)	
		SEn(M)	
			SVa(M)
			Ent(S)
		IDM(M)	
		SVa(M)	
		ClSh(S)	

Classification Tables

Wilks' λ

		Predicted			
		1	2	3	5
T	1	6	3	0	1
r	2	2	7	1	0
u	3	2	3	5	0
e	5	0	0	0	10

70%

Mahal. Dist.

		Predicted			
		1	2	3	5
T	1	6	3	1	0
r	2	1	9	0	0
u	3	1	1	8	0
e	5	0	0	0	10

82.5%

Table 3.23 Number of images correctly classified (displacement 4).

There is very little pattern in these results, with different variables being included for each displacement used and for both methods of selecting variables. There are not even any variables which consistently appear at the top of the tables for all displacements. In fact, in each case, several variables are included, indicating that each successive variable entered is providing new information and a small subset of variables will not describe the texture adequately. One interesting point is that ASM, either mean or standard deviation is seldom included in the model. IDM, either mean or standard deviation is included in most models, as is sum entropy. The best results were obtained using a displacement of 3, with all the images being correctly classified by the Wilks' λ method.

A similar procedure was carried out using the 4 features from difference statistics. Again both mean and standard deviation over the 4 directions were calculated but this time features from displacements 1 and 2 and displacements 3 and 4 were considered together. The results were as follows:

Displacements 1 & 2			
Wilks' λ		Mahal. Dist.	
In	Out	In	Out
Con2(M)		Mean1(S)	
Mean1(M)		ASM2(S)	
ASM2(S)		Mean1(M)	
Ent1(S)		Mean2(S)	
Con1(M)		Ent2(M)	
Mean2(S)			ASM2(S)
	ASM2(S)	Mean2(M)	
ASM1(S)			Ent2(M)
Ent1(M)		Ent1(M)	
ASM2(M)		Con2(M)	
	Con1(M)	Con1(S)	
			Mean1(S)
		Ent2(M)	
			Mean1(M)

(Con2=Contrast, displacement 2 etc.)

Classification Tables

Wilks' λ

		Predicted			
		1	2	3	5
T	1	9	0	0	1
r	2	0	9	1	0
u	3	0	1	9	0
e	5	0	0	0	10

92.5%

Mahal. Dist.

		Predicted			
		1	2	3	5
T	1	9	0	0	1
r	2	0	9	1	0
u	3	1	0	9	0
e	5	0	0	0	10

92.5%

Table 3.24 Number of images correctly classified by difference statistics (displacements 1 and 2).

Displacements 3 & 4	
Wilks' λ	Mahal. Dist.
In / Out	In / Out
Con4(M)	Con4(S)
Con3(S)	Con3(S)
Con3(M)	Con4(M)
Mean3(M)	Con4(S)
Ent4(M)	Mean4(M)
	Ent3(S)
	Mean3(M)
	Con3(M)
	Ent4(S)
	Mean4(S)

Classification Tables

Wilks' λ

		Predicted			
		1	2	3	5
T	1	9	1	0	0
r	2	4	5	1	0
u	3	4	2	4	0
e	5	2	0	0	8

65%

Mahal. Dist.

		Predicted			
		1	2	3	5
T	1	9	1	0	0
r	2	4	5	1	0
u	3	1	3	6	0
e	5	1	0	0	9

72.5%

Table 3.25 Number of images correctly classified by difference statistics (displacements 3 and 4).

Displacements 1 and 2 seem to give better classification results, but again there is no real pattern. ASM seems to be the least useful feature for classification but contrast, mean and entropy are included in most cases. In some cases there is additional information in a feature already entered but with a different displacement, e.g. Con3(M) and Con4(M). The classification results were quite good using displacements 1 and 2, with 92.5% of images being correctly classified by both Wilks' λ and Mahalanobis'

distance. The number of variables included in the classification was large (9, out of a possible 12, with Wilks' λ and 10 with Mahalanobis' distance). It seems unlikely therefore that a smaller subset of variables could correctly classify future images and it appears that the best classification results may be obtained by using all the information in the difference statistics.

3.10 Maximum Likelihood

As mentioned above, because of the number of variables included in each model, a better method of classification may be to use all the information contained in either the co-occurrence matrices or difference statistics and use the maximum likelihood method suggested by Vickers and Modestino in 1982. They tested their method using displacements of 1, 3 and 5 and obtained the best results with a displacement of 5. The same displacements were investigated in this study. Testing the classifier on the training data from which it was constructed (i.e. no jackknifing) gave the following results:

Displacement 1

		Predicted			
		1	2	3	5
T	1	8	0	1	1
r	2	0	10	0	0
u	3	0	0	7	3
e	5	0	0	0	10

70%

Displacement 3

		Predicted			
		1	2	3	5
T	1	10	0	0	0
r	2	0	10	0	0
u	3	0	1	9	0
e	5	0	0	0	10

97.5%

Displacement 5

		Predicted			
		1	2	3	5
T	1	10	0	0	0
r	2	0	10	0	0
u	3	0	0	10	0
e	5	0	0	0	10

100%

Table 3.26 Number of images correctly classified using maximum likelihood.

Thus, using a displacement of 5, all the images are correctly classified. When the method was tested on a new test set containing 5 images of each of the 4 types, however, the results were:

		Predicted			
		1	2	3	5
T	1	4	0	1	0
r	2	1	2	2	0
u	3	2	3	0	0
e	5	0	0	0	5

55%

Table 3.27 Number of new images correctly classified by maximum likelihood.

Conners and Harlow (1980b) suggested that the reason co-occurrence matrices performed better than difference statistics in their study, but vice-versa in other studies (Weszka et al. (1976)) may be due to the fact that the most commonly used features from co-occurrence matrices do not contain all the textural information. It is, however, possible that, similarly, the features extracted from difference statistics do not contain all the textural information there is. Vickers and Modestino (1982) only tested their method using co-occurrence matrices. Using a similar method, again with displacements

1, 3 and 5, but taking difference statistics as the starting point, produced the following results:

Displacement 1

		Predicted			
		1	2	3	5
T	1	5	0	2	3
r	2	4	2	3	1
u	3	1	0	6	3
e	5	3	0	0	7

50%

Displacement 3

		Predicted			
		1	2	3	5
T	1	7	0	2	1
r	2	0	10	0	0
u	3	1	0	8	1
e	5	0	0	1	9

85%

Displacement 5

		Predicted			
		1	2	3	5
T	1	7	3	0	0
r	2	0	9	1	0
u	3	0	3	6	1
e	5	1	0	0	9

77.5%

Table 3.28 Number of images correctly classified by maximum likelihood based on difference statistics.

This time, when the classifier was tested on the training set, displacement 3 was best. As with co-occurrence matrices, when tested on a new test set, the results were poorer:

Displacement 3

		Predicted			
		1	2	3	5
T	1	1	1	3	0
r	2	1	3	1	0
u	3	2	2	1	0
e	5	0	0	1	4

45%

Table 3.29 Number of new images classified by maximum likelihood based on difference statistics.

Thus difference statistics do seem to contain less information than co-occurrence matrices.

Because such good results were achieved with co-occurrence matrices on the training set it is hoped that with a much larger training set, similar results will be obtained when tested on a test set.

In an attempt to increase the size of the database a whole new set of images was collected from different rats' bones. These images appeared much clearer on the screen and consisted of good examples of all 5 texture types. Training sets consisting of 30 images of each of the 5 types were constructed.

A selection of five images of each of the types is given in figures 3.9 and 3.10. The difficulty in distinguishing between the textures can be seen from these images. There is a very fine dividing line between type 4 and type 5 textures and many of the type 5 textures, when examined over such small windows, look remarkably like type 3 textures. With the formative types, type 1 and 2 textures can often look very similar, as can type 2 and type 3 textures.

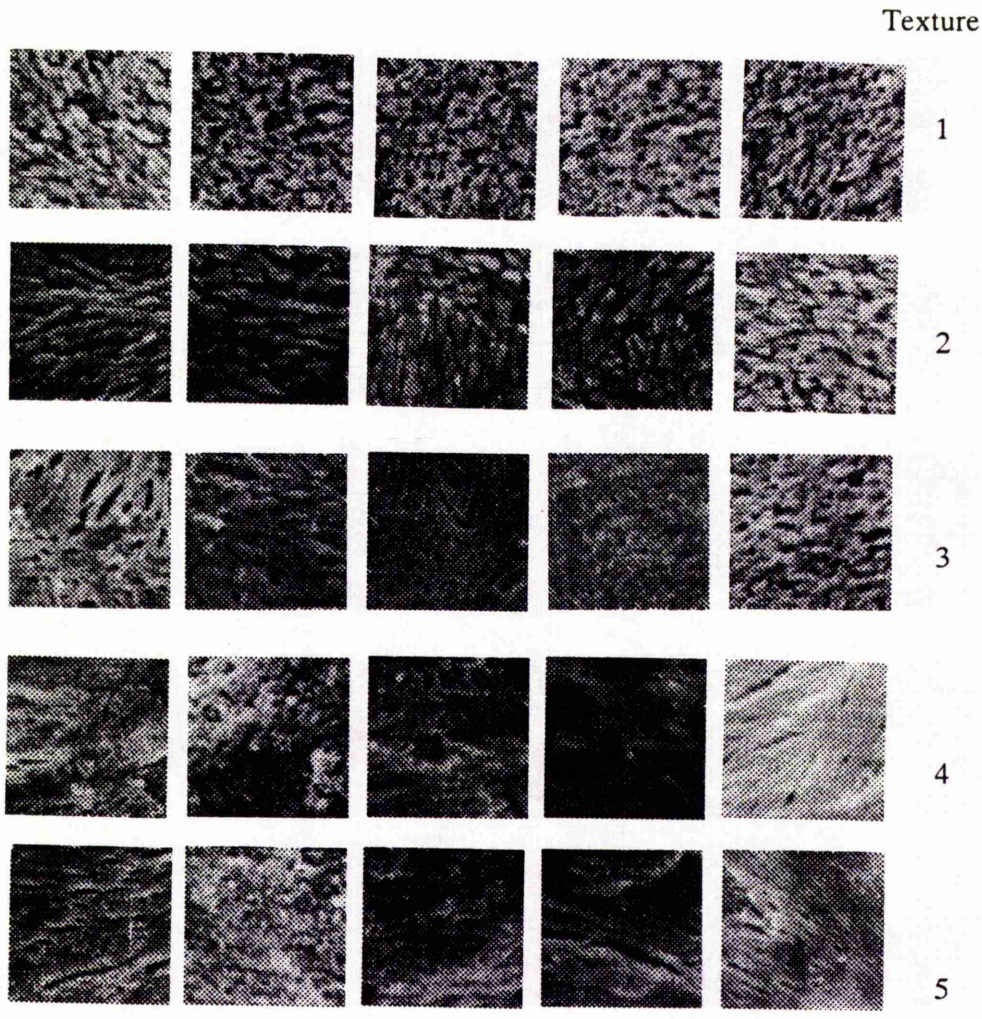
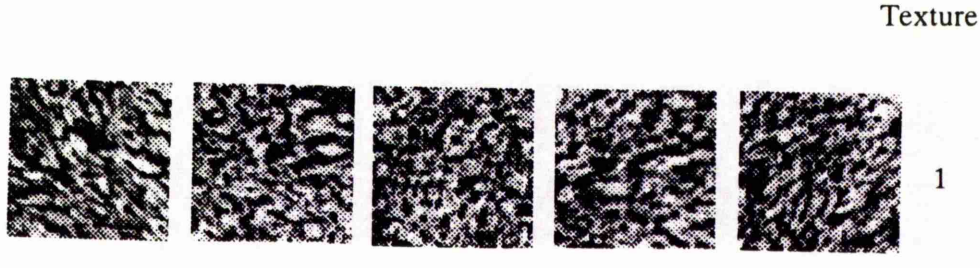


Fig. 3.9 A selection of the images that made up the new training set, before histogram equalisation.



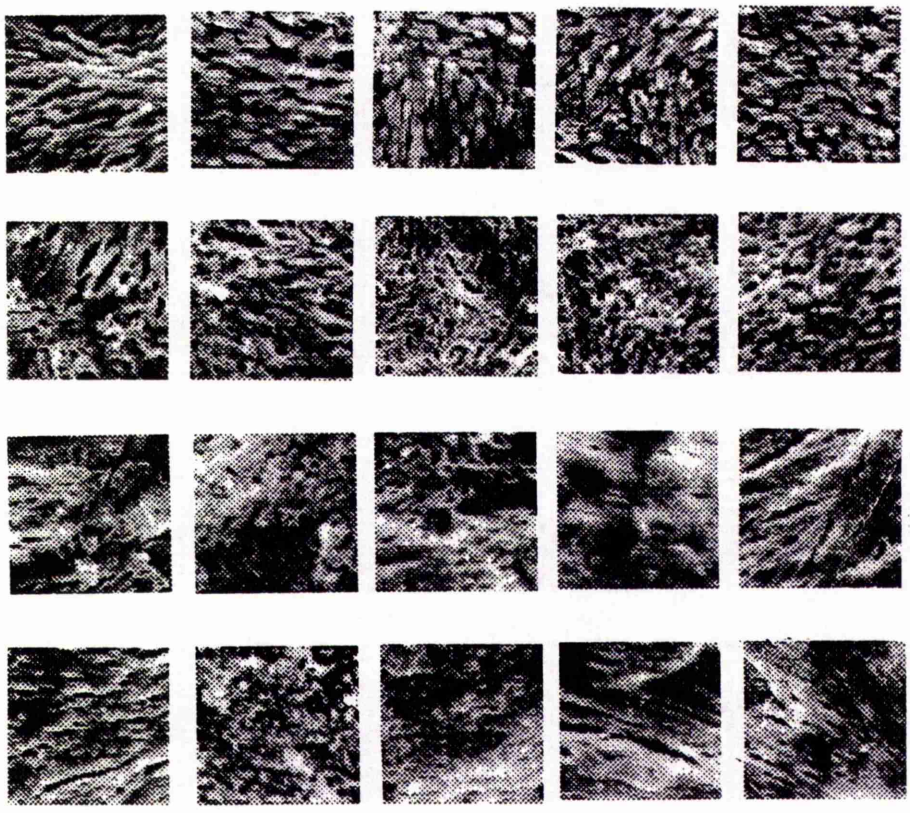


Fig. 3.10 The images in Fig. 3.9 after histogram equalisation.

The maximum likelihood method was then tested on this new data with the following results:

Displacement 1

		Predicted				
		1	2	3	4	5
T	1	29	1	0	0	0
r	2	2	28	0	0	0
u	3	1	2	26	0	1
e	4	0	2	1	22	5
	5	0	3	0	0	27

88%

Displacement 3

		Predicted				
		1	2	3	4	5
T	1	30	0	0	0	0
r	2	0	29	1	0	0
u	3	0	1	29	0	0
e	4	0	0	2	26	2
	5	0	0	3	2	25

92.6%

Displacement 5

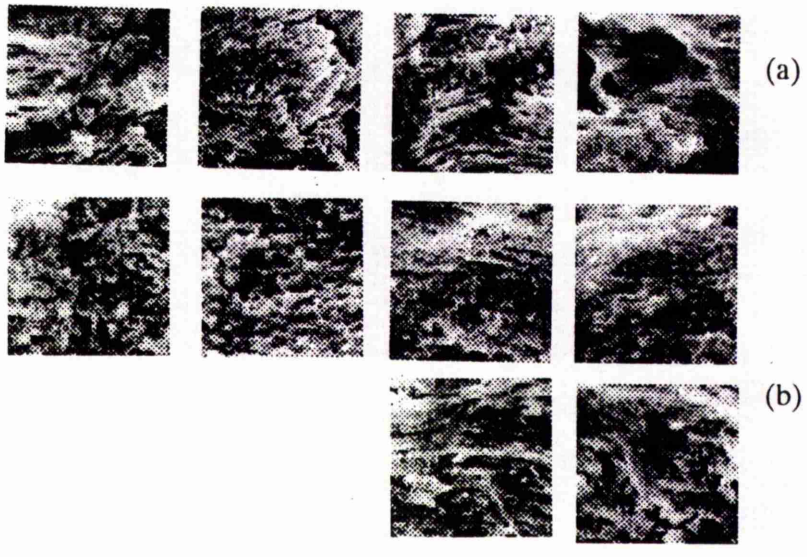
		Predicted				
		1	2	3	4	5
T	1	30	0	0	0	0
r	2	0	30	0	0	0
u	3	0	0	30	0	0
e	4	0	0	3	26	1
	5	0	0	4	2	24

93.3%

Table 3.30 Number of images correctly classified by maximum likelihood based on co-occurrence matrices.

The best displacement was 5, with 93.3% of images correctly classified. All of the type 1,2 and 3 images have been correctly classified with all the misclassifications being of type 4 and 5 images. A closer examination of the images that were misclassified, indicates to some extent, the reasons for the misclassifications. The images are shown in figure 3.11.

Fig. 3.11
Misclassified images.
a) Type 4
b) Type5



The first three images in figure 3.11(a) were classified as type 3 and the fourth image as type 5. Comparing these with the selection of images shown in figure 3.10, the misclassifications are not surprising. In figure 3.11(b) the first four images were classified as type 3 and the last two as type 4. Again, comparing with figure 3.10 this is not entirely unexpected. In addition, by comparing the $\log_e$-likelihood functions for each of the 5 types, in all the misclassified images, the correct texture type would have been the second choice and there is often very little difference between the values of the function for the first and second choices of texture. Table 3.31 below shows the $\log_e$-likelihood functions for the type 5 images that were misclassified.

Image	Type 1	Type 2	Type 3	Type 4	Type 5
a	-33252	-33219	**-33167**	-33191	-33175
b	-33278	-33263	**-33221**	-33275	-33237
c	-33233	-33187	-33045	**-32888**	-32891
d	-33215	-33150	-32960	**-32685**	-32708
e	-33261	-33232	**-33192**	-33233	-33212
f	-33249	-33223	**-33165**	-33200	-33179

Table 3.31 Log_e-likelihood functions for each of the 5 texture types for the misclassified type 5 images.

When the method was tested on a set of new test images the results were as follows:

Displacement 5

		Predicted				
		1	2	3	4	5
T	1	8	2	0	0	0
r	2	2	8	0	0	0
u	3	0	2	7	0	1
e	4	0	0	2	6	2
	5	0	1	6	0	3

64%

Once again, the results are worse than those obtained on the training sets, but are still reasonably good for the formative texture types. The problem lies with the two resorptive types, 4 and 5. What may be necessary is to first classify into one of types 1,2,3, and a resorptive type (either type 4 or 5) and then further classify the resorptive types into one of types 4 or 5. Investigating the first stage of this procedure, i.e. classifying the training images into one of the 4 classes above, gave the following results with displacement 5:

Displacement 5

		Predicted			
		1	2	3	4or5
T	1	30	0	0	0
r	2	0	30	0	0
u	3	0	0	30	0
e	4	0	0	3	27
	5	0	0	5	25

94.6%

This has not really led to any improvement in the classification. The type 4 and 5 textures are still being confused with texture type 3. The results with 5 texture types,

considering the wide variety of rats, bones, images, techniques for capturing the images etc. are, however, fairly good, and at least comparable to those obtained by other authors.

To see whether this new data set was compatible with the one used previously, and in fact the two could be combined to provide more data, the maximum likelihood routine was run using the new data set as the training set and the old data set was used as the test set. The results, using displacements 1,3 and 5 are given in Table 3.32.

Displacement 1

		Predicted				
		1	2	3	4	5
T	1	5	3	2	0	0
r	2	5	4	1	0	0
u	3	6	3	1	0	0
e	5	0	0	7	0	3

32.5%

Displacement 3

		Predicted				
		1	2	3	4	5
T	1	7	0	3	0	0
r	2	3	0	7	0	0
u	3	4	0	6	0	0
e	5	6	1	3	0	0

32.5%

Displacement 5

		Predicted				
		1	2	3	4	5
T	1	6	2	2	0	0
r	2	7	2	1	0	0
u	3	6	2	2	0	0
e	5	0	0	6	3	1

27.5%

Table 3.32 Number of images in old training set correctly classified from new training set images.

As can be seen, the results are very poor with only 32.5% of images being correctly classified with displacements 1 and 3 and 27.5% with displacement 5. Why the results are so bad is not immediately clear although by comparing the two sets of

images (figures 3.6 and 3.9 and 3.7 and 3.10) it is not surprising that a large percentage of images has been misclassified. Comparing either the original images or the images after histogram equalisation it can be seen that each of the types looks quite different in the two data sets. The images in figure 3.7 are quite blurred and the texture features in some images are barely visible. The images in figure 3.10, however, look much brighter and the texture features are more easily discernible. There are several possible reasons for the differences. Firstly, the images in figure 3.7 are taken from quite old specimens and it is possible that the gold covering has eroded away to some extent. The images in figure 3.10, on the other hand were freshly prepared for this study. The second major difference in the two sets of images is the method in which they were saved. The original images (figure 3.7) were saved on disk using a Magiscan computer connected to the SEM and then had to be transferred across, using Kermit, for use on a Sun SPARCstation. The later images, however, were photographed from the SEM and then the photographs were scanned onto disk. This method appeared to give clearer images. In addition, partly due to the different methods of image capture, the two sets of images have been magnified to slightly different degrees. As explained in Chapter 1, the Magiscan captures the central portion of the image and doubles the original magnification. The images were viewed in the SEM at a magnification of 350, thereby resulting in images with a magnification of 700. The later images, however, were magnified 750 times. Visually, this should not make a great difference but may, to a small extent, adversely affect the classification results.

The classification algorithm works reasonably well on both data sets and the solution is probably to use similar images for training and test sets, i.e. if test images have been captured using a Magiscan computer, use the old training set, whereas if test images are taken from photographs which have been scanned, use the new training set.

3.11 Conclusions

Many techniques for texture classification have been suggested over the years. Several of these are described in chapter 2. The majority of researchers have tested

their techniques on Brodatz' (1966) images, with varying rates of success. This study is one of few where classification techniques were tested on more natural images. Consequently the results in this study do not appear to be quite as good as some of those in published literature. Nevertheless, correct classification rates of over 80% are good considering the small data sets used and the wide variety of bones imaged to make up the sets. It is hoped that if a much larger database of images was used as a training set the classification could be improved further.

The approach taken was first of all to examine statistical techniques, particularly co-occurrence matrices, and to compare these with other techniques such as difference statistics, Fourier power spectrum methods and run-length matrices. The results obtained were similar to those of other authors such as Weszka et al. (1976) and Conners and Harlow (1980b), in that co-occurrence matrices and difference statistics were found to perform better than the other two methods. The results were encouraging, with a reasonably high percentage of images being correctly classified, particulary in the case of co-occurrence matrices, and so methods involving co-occurrence matrices were investigated further. Although most authors have tended to use only the features angular second moment, contrast, correlation and entropy, the use of a step-wise discrimination routine suggested that, in this study, including several textural features produced the best classification results. These features included, not only the original ones proposed by Haralick et al. in 1973, but also cluster shade and prominence (Conners et al. (1984)) and the kappa statistic (Parkinnen & Selkainaho (1990)). It was also noted that no subset of features consistently appeared in the list of those selected. As so many features seemed to be required in the model, it was thought that a quicker and simpler method may in fact be to use the whole co-occurrence matrix. Thus Vickers & Modestino's (1982) maximum likelihood method was employed. Vickers and Modestino (1982) claimed a correct classification rate of 100% when their method was applied to training sets and similar success rates were obtained in this study. If this research is to be of any use in the future, however, it is essential that the method also works well on test sets. In all the examples in this study it was found that despite very high rates of correct classification with training set images, the rates fell quite considerably when applied to test sets. This indicates the necessity of assessing the performance of classification techniques either using test sets or jackknifing. Although 100% correct

classification has not been achieved on test sets in this study it is unlikely that this will ever be achieved. A brief investigation of k-nearest neighbours classification technique using a set of 12 features taken from co-occurrence matrices also obtained 100% correct classification on the training set, but when tested on a test set the results dropped dramatically. Misclassification matrices for 1st, 3rd and 5th nearest neighbours are shown in table 3.33 below.

1st NN

		Predicted				
		1	2	3	4	5
T	1	3	4	3	0	0
r	2	2	4	1	2	1
u	3	2	1	6	0	1
e	4	0	1	1	3	5
	5	0	2	2	2	4

40%

3rd NN

		Predicted				
		1	2	3	4	5
T	1	5	2	2	0	1
r	2	1	7	1	1	0
u	3	4	0	5	0	1
e	4	0	2	1	5	2
	5	0	2	2	2	4

52%

5th NN

		Predicted				
		1	2	3	4	5
T	1	4	5	0	0	1
r	2	1	7	1	1	0
u	3	2	0	8	0	0
e	4	1	2	0	4	3
	5	1	2	0	3	4

54%

Table 3.33 Number of images correctly classified using k-nearest neighbour classification.

Neural networks were also examined briefly, but again did not perform well on

test sets. A feed-forward net with 13 inputs, 9 hidden units and 5 outputs trained using variable-metric learning produced the following misclassification matrix when applied to the test set.

		Predicted				
		1	2	3	4	5
T	1	2	5	1	1	1
r	2	0	5	1	1	1
u	3	1	0	5	1	3
e	4	0	1	1	5	3
	5	1	0	0	4	5

44%

Table 3.34 Misclassification matrix for a feed-forward neural network when applied to a test set.

Further investigation of these methods, especially with the neural network approach, where different architectures could be explored, may lead to an improvement in results. However, considering the variability in the images, the similarity in some of the texture types and the fact that physiologists may often not agree on the texture type in an image, it is unlikely that any method will ever achieve 100% correct classification on test sets. The results obtained here using the whole co-occurrence matrix and a maximum likelihood classifier have produced fairly high rates of success. One way of improving the results may be to increase the number of texture types and, for example, have a texture type 1.5 which would represent a late type 1 or an early type 2 texture. Similarly, 'in-between' classes could be created for types 2 and 3 and for types 4 and 5. For this to be feasible, and this is probably the main weakness of this study, much larger training and test sets would be required. More sophisticated techniques could perhaps produce comparable results, but it is unlikely that a significant improvement would be obtained using the same data sets. The greatest improvement could almost certainly be obtained by simply applying the maximum likelihood method to much larger training sets, where all images are saved in a similar way, with good examples

of each of the 5 texture types, and further examples of 'in-between' classes.

Chapter 4 Segmentation

4.1 Introduction

Segmentation is the process which divides an image into its constituent parts or regions. It is an important element in image processing because it is at this stage that regions of interest are extracted for further processing such as the classification discussed in chapters 2 and 3.

One of the simplest ways of extracting objects or regions from an image is by grey-level thresholding. Given a grey-level image IP(i,j) in the range 0 to 255 and a value of t between 0 and 255, by thresholding at t, we obtain the binary image:

$$IP_t(x,y) = 255 \quad \text{if } IP(i,j) \geq t$$
$$= 0 \quad \text{if } IP(i,j) < t.$$

If an image consists of objects that are either darker than their background or lighter than their background then thresholding is a natural way to extract the objects. Similarly, a 'grey' object can be extracted from a black and white speckled background by defining:

$$IP_{u,v}(i,j) = 255 \quad \text{if } u \leq IP(i,j) \leq v$$
$$= 0 \quad \text{otherwise.}$$

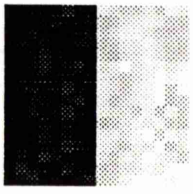
In many image processing situations, however, the problem is not quite as straightforward as this and segmenting a picture usually involves detecting edges between two regions which differ with respect to their distribution of grey levels or their texture.

4.2 What is an edge?

The classification techniques described in chapter 2 were designed to detect some sort of **uniformity** in grey levels or textures within each texture type. The problem of segmentation is based on detecting **discontinuities**, i.e. places where there is a change

in grey level or texture indicating the end of one region and the beginning of another. Such a discontinuity is called an edge. Ideally, an edge will be a continuous well-defined line but in practice an edge may be broken, where the change in regions is not abrupt enough, or in the extreme case where one region merges into another there will be no clearly defined edge at all.

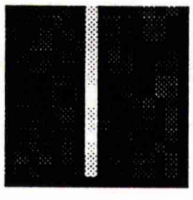
Various edge-detection operators have been developed. Kirsch (1971) and Robinson (1977) applied 3×3 masks to an image and took the maximum value to be the edge magnitude. Roberts (1965), Sobel (1971) and Prewitt (1970) devised differential operators where the difference between grey levels of two neighbouring pixels is taken. High absolute values are associated with areas where the grey level is changing rapidly. For example, given a grey-level image IP(i,j):



the Roberts' (1965) operator computes the sum of the squared differences of diagonally opposite pixel values. Thus

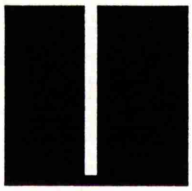
$$JP(i,j)= \sqrt{[\{IP(i,j)-IP(i+1,j+1)\}^2 + \{IP(i+1,j)-IP(i,j+1)\}^2]}$$

producing the image JP(i,j):



N.B. No values are obtained for the margin i=16 and j=16

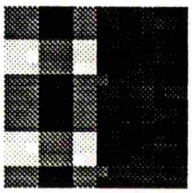
By thresholding at 20 this method would detect a continuous vertical edge:



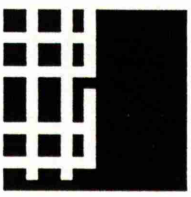
4.3 Texture Edge-detection

Segmentation of texture is slightly more complicated than grey-level edge-detection. Consider an image containing two textures, on the right a very smooth texture and on the left a texture with more contrast.

e.g. IP(i,j):



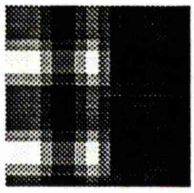
All the edge-detectors mentioned above would be able to detect the edge between the two textures. They would also, however, detect edges within the coarse texture on the left. For example, the resulting image using the Roberts' (1965) operator above is:



Thus although, on the whole, the edge between the textures has been picked up, so have the edges within the texture on the left. These abrupt changes in grey level are however part of the texture and have to be recognised as such.

One way of overcoming this problem is by taking averages over neighbourhoods and looking for discontinuities in the averages. Indeed, one of the most common ways of detecting texture edges is by locating regions where there is an abrupt change in **average** grey level.

In the example above, averaging over 2×2 neighbourhoods gives:



A differential operator, such as the Roberts' (1965) operator would now detect a vertical edge between the two textures. Averaging over neighbourhoods of size 3×3 or 4×4 would also detect the edge but a neighbourhood of size 8×8 for example would be too large to detect an edge between the two regions. The size of neighbourhood needed depends on the coarseness of the texture, a coarse texture requiring a larger neighbourhood than a fine texture, and also on the size of regions to be detected. If the smallest region size is known in advance the largest possible size should be chosen such that regions of this size can be detected. If this is not known, and it seldom is, a possible approach would be to consider a selection of different sizes. Several authors have measured image properties on windows of varying sizes using a split-and-merge algorithm (Chen & Pavlidis (1979), Burt et al. (1981), Spann & Wilson (1985) and Wilson & Spann (1988)).

The quad-tree approach is a particular case of a split-and-merge algorithm. The leaves of the tree correspond to single pixels and the root (level 0) represents the whole picture. If the entire picture is not considered to be homogeneous with respect to some feature it is split into four equally sized blocks and the feature measured again in each block. If any of the smaller blocks are still not homogeneous they are then in turn divided into four. At each stage, as well as splitting inhomogeneous regions, homogeneous regions may be merged. The procedure is repeated until the image consists of a number of homogeneous regions.

Chen & Pavlidis (1979) used co-occurrence matrices as features to segment textured images and a region R was said to have a uniform texture if:

$$\sum_{i=0}^{LVL-1} \sum_{j=0}^{LVL-1} [\max C_{ij}^1, C_{ij}^2, C_{ij}^3, C_{ij}^4) - \min (C_{ij}^1, C_{ij}^2, C_{ij}^3, C_{ij}^4)] < T_C$$

where C^1_{ij}, C^2_{ij}, C^3_{ij} & C^4_{ij} are the co-occurrence matrices of the 4 quadrants of R and T_C is a pre-specified threshold.

Spann & Wilson (1985) used a method which combined a quad-tree smoothing operation with statistical classification performed at the highest level of the tree, followed by a downward directed boundary estimation based on the segments obtained at the top of the tree. The advantage of this method is that it doesn't need any a priori information - even the number of classes. In Wilson & Spann (1988) the features are based on finite prolate spheroidal sequences derived from a tesselation of the frequency plane.

In 1971, Rosenfeld & Thurston proposed a method which computes averages for a range of sizes and picks a "best" size at each point of the picture by comparing these values. The procedure is summarised in the following steps.

1. For a 2^n by 2^n image calculate the average grey level over the neighbourhood of size 2^k by 2^k (k=1,2,...n-1) at each point (x,y)

$$A_k(x, y) = \frac{\sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} IP(i, j)}{2^k \times 2^k}$$

e.g. if

$$IP(i,j) = \begin{bmatrix} 10 & 2 & 9 & 3 & 6 & 5 \\ 9 & 3 & 10 & 2 & 5 & 6 \\ 10 & 2 & 11 & 3 & 5 & 5 \end{bmatrix}$$

$$A_1(3,4) = \{IP(2,3)+IP(3,3)+IP(2,4)+IP(3,4)\}/4$$

$$= 6.5$$

2. For each size, at each point, take differences between pairs of averages corresponding to pairs of nonoverlapping neighbourhoods in each of 4 directions (horizontal, vertical and two diagonals). These differences are measures of edge strength. Thus the difference between a pair of horizontally adjacent nonoverlapping averages of size 2^k by 2^k at (x,y) is

$$E_{k,h}(x,y) = |A_k(x+2^{k-1},y) - A_k(x-2^{k-1},y)|$$

e.g. $$E_{1,h}(3,5) = |A_1(3,6) - A_1(3,4)|$$

$$= |5.25 - 6.5|$$

$$= 1.25$$

The best orientation is the one which gives the highest absolute difference.

3. The next stage is to pick a best size. This is taken to be the largest size for which the next smaller size does not give a significantly higher absolute difference. Specifically, if $E_k(x,y)$ is the best of the E's of size $2^k \times 2^k$ in four directions at the point (x,y), and the sizes used are 1×1, 2×2, ... $2^L \times 2^L$, the best size is the largest k such that

$$E_{L-1} < \lambda E_{L-2} < \ldots < \lambda^{L-k-1} E_k, \qquad \text{but } E_k > \lambda E_{k-1}$$

where Rosenfeld & Thurston (1971) took λ =0.75 to define "significantly" higher.

The edge strength at a particular point will be high if the averages over the two neighbourhoods are very different, in particular, it will be high if the neighbourhoods lie just on opposite sides of an edge.

If k is large, E_k will be high, however, not only at an edge, but over a range of positions on either side of it. This is because the averages over the two touching neighbourhoods begin to differ as soon as one of the neighbourhoods overlaps both sides of an edge, while the other neighbourhood lies entirely on one side. E_k, however, should be a local maximum when the neighbourhoods touch just at an edge, and each neighbourhood lies entirely on one side, since the averages are then as different as possible. Thus, if we want to determine the location of an edge, we should ignore high values of E_k if there are higher ones nearby in the same direction, i.e. only keep the local maxima. This process is called thinning.

4. Erase the value at a point if there is a higher value at any point within a distance of half the best size in a direction perpendicular to the best orientation at the point. For example, if the best orientation is horizontal and the best size is $2^k \times 2^k$, $E_k(x,y)$ is suppressed if any $E_k(x',y)$ (where $x-2^{k-1} < x' < x+2^{k-1}$) is greater than it.

E.g. if the matrix of best edge-strengths is:

$$E(x, y) = \begin{bmatrix} 5 & 7 & 3 & 2 \\ 4 & 6 & 5 & 3 \\ 4 & 9 & 6 & 4 \\ 4 & 8 & 8 & 5 \end{bmatrix}$$

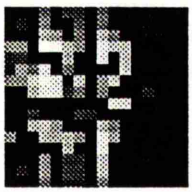
and the best size at E(2,2) is 2×2 corresponding to a horizontal edge, then the edge strength at E(2,2) would become zero because E(1,2) and E(3,2) are both greater than E(2,2).

$$E(x, y) = \begin{bmatrix} 5 & 7 & 3 & 2 \\ 4 & 0 & 5 & 3 \\ 5 & 9 & 6 & 4 \\ 4 & 8 & 8 & 5 \end{bmatrix}$$

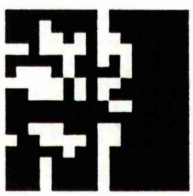
A similar procedure would be carried out for each point (x,y).

5. The resulting image is then thresholded to produce a binary image.

The result of applying the first 4 of these procedures to the textured image above, taking λ to be 0.75 and using a maximum edge size of 3, corresponding to 8×8 windows, produced the following image:



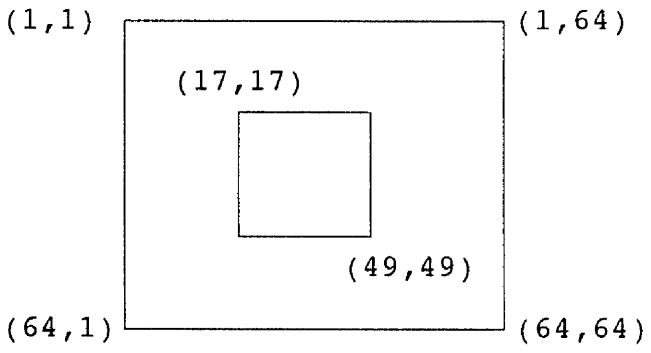
The resulting image after thresholding at 20 is as follows:



At this stage, the method does not look very promising. Not only are edges within the coarse texture still being highlighted, but also the main edge between the textures seems to be lost. There are various reasons for this, and modifications that can be made to improve the segmentation. These will be discussed throughout the rest of the chapter.

4.3.1 Notes on Rosenfeld & Thurston (1971) Algorithm

1) Rosenfeld & Thurston (1971) proposed that the largest neighbourhood used should have "size comparable to that of the entire picture". Thus, for a 64×64 image, L would be 6, giving neighbourhoods of size 64×64, 32×32, 16×16, 8×8, 4×4 and 2×2. For the 64×64 neighbourhood only one pixel could have its average grey level calculated and it would not be possible to compute any differences of 64×64 nonoverlapping windows. Thus all edge strengths would be zero. Moving down to L=5, i.e. 32×32 neighbourhoods, considering only the horizontal direction for simplicity, it is possible to calculate the average grey level of pixels in a square region from (17,17) to (49,49).



Edge values could then be computed for pixels (17,1)...(48,64). Note, however, that most of these edge values will not have been calculated from the difference of two averages, e.g. $E_{5,H}$ (17,1) = A_5 (33,1) - A_5 (1,1), where neither A_5 (1,1) nor A_5 (33,1) will be averages. If pixels where it is not possible to compute an average are left with their original grey level, what is being calculated is not the edge strength between two areas of size 32×32 but between single pixels. This problem will always arise at the boundaries of an image, and the greater the neighbourhood used, the larger the boundaries will be. To prevent these misleading results it is proposed here that edge strengths should only be computed where it is possible to compute the appropriate averages. Thus for a 64×64 image, the maximum edge size would be 5 (i.e. 32×32), but this would only be possible at certain pixels, i.e. (33,17)...(33,49) for horizontal

neighbourhoods.

2) Rosenfeld & Thurston (1971) took the value of λ to be 0.75. If 2×2, 4×4, 8×8, 16×16 and 32×32 neighbourhoods were investigated in a 64×64 image, k=2 i.e. a 4×4 neighbourhood would be 'best' if $E_4 < \lambda E_3 < \lambda^2 E_2$ and $E_2 \geq \lambda E_1$. Thus, with $\lambda = 0.75$, E_2 would have to be considerably bigger than E_3 and E_4, but could in fact be smaller than E_1. It is intuitively sensible to have λ less than one, since the aim is to detect edges between relatively large areas of texture and to ignore edges within the texture. Taking λ greater than one results in the larger edge sizes being discriminated against in favour of the smaller edge sizes and, hence, within-texture edges. Rosenfeld & Thurston (1971) gave no reason for choosing 0.75 in preference to any other value. In this study it was found that if the maximum edge size (L) was restricted to 3 then all values of λ between 0 and 1 produced the same results. Allowing neighbourhoods of sizes 4 and 5 led to spurious large edges. By increasing the value of λ these could be suppressed but when values of λ greater than 1 were used the result tended to be lots of extra smaller edges being highlighted. Thus it was found that for the 64×64 images used in this study, the best segmentation was obtained with L, the maximum edge size, equal to 3, i.e. only taking averages of 8×8, 4×4 and 2×2 neighbourhoods, and λ taken to be any value less than 1. From this point on, in fact, λ was always taken to be 0.75.

3) To thin the edges, Rosenfeld & Thurston (1971) suggested looking within a distance of half the best size in the direction perpendicular to the best orientation. For large neighbourhoods this may not be sufficient to thin the edges fully. A greater distance could be used but care would have to be taken if an image contains bands of texture, or objects of similar size to the neighbourhood used (Figs. 4.1 and 4.2).

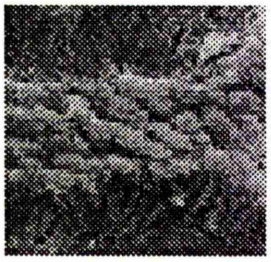


Fig. 4.1 A band of texture type 2.

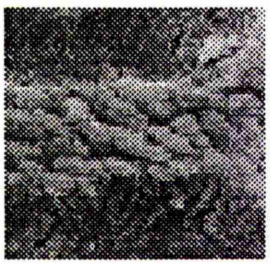
Fig. 4.2 An object (O) embedded in the texture.

In Fig. 4.1, if the band is of width 5 pixels and thinning is carried out to within a distance of 5 pixels, the weaker of the two edges of the band would be suppressed. A similar thing would happen with the object in Fig. 4.2. Rosenfeld et al. (1972) suggested that a way of getting round this problem would be to take signed differences, when computing edge values in step 2. Edges would then be suppressed if there were higher values of the same sign within a given distance and in a given direction. Since opposite edges of an object would have opposite signs they would not compete with each other. As illustrated in figures 4.1 and 4.2, the images in this study do contain objects such as vascular channels and bone cells, and may contain bands of texture, and these potential problems have to be considered when thinning is performed. In fact, thinning was carried out to within a distance of half the best size. Because the maximum edge size was limited to 8×8 pixels this did not result in many unwanted edges.

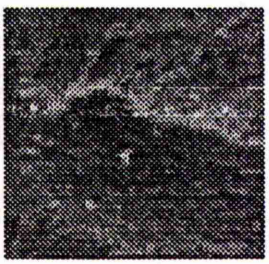
4.4 Application of Rosenfeld & Thurston (1971) algorithm

Some early results of applying the Rosenfeld & Thurston (1971) algorithm to images of rats' tibiae are given in Reid et al. (1990). The method was tested on several

64x64 images with varying success depending on the texture types they contained (Fig. 4.3). The method does, on the whole, highlight the boundaries between texture types but also highlights various other unwanted edges in the images (Fig. 4.4). This is particularly the case with texture type 2 where the edges round the mineral nodules are highlighted (Fig. 4.4a).



(a)



(b)

(c)

Fig. 4.3 Digital image from SEM showing

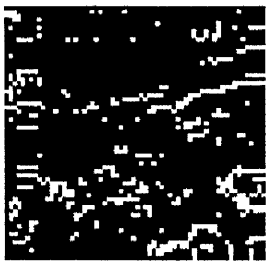
a) texture type 4 with a band of texture type 2

b) texture type 3 (upper half) & texture type 5 (lower half)

c) texture type 5 (upper half) & texture type 1 (lower half)

(a)

(b)



(c)

Fig 4.4 Application of the Rosenfeld & Thurston (1971) edge detection algorithm to locate boundaries, giving binary images (a), (b) & (c).

4.5 Smoothing

One possible reason for the segmentation algorithm performing so poorly is the fact that the images are very noisy. In an attempt to remove some of the noise from the raw images and improve the segmentation, two types of smoothing algorithms were investigated. The first was an edge-preserving smoothing routine proposed by Nagao and Matsuyama in 1979. The method looks at 9 masks - 4 pentagons, 4 hexagons and 1 square, and chooses the one with the minimum variance (Fig. 4.5). This reduces the likelihood of averaging over neighbourhoods containing more than one texture type and therefore preserves the jump in edge strength at an edge.

Fig.4.5 The masks used in the edge-preserving smoothing routine

The routine then averages over the grey levels and substitutes the average value

for the grey level at a point.

As can be seen the results for the images containing types 3 and 5 were very good (Fig. 4.6b). The edge is fairly continuous and the image contains very little noise. Results for images containing types 2 and 4 were poorer (Fig. 4.6a). As before, the main edge is detected but the edges around the type 2 pebbles are still highlighted. Lowering the threshold does not help but simply causes the edge between the texture types to become broken. Similarly, iteratively smoothing the images tends to make the edges more broken and does not lead to a reduction in the number of unwanted edges being highlighted (Fig. 4.7).

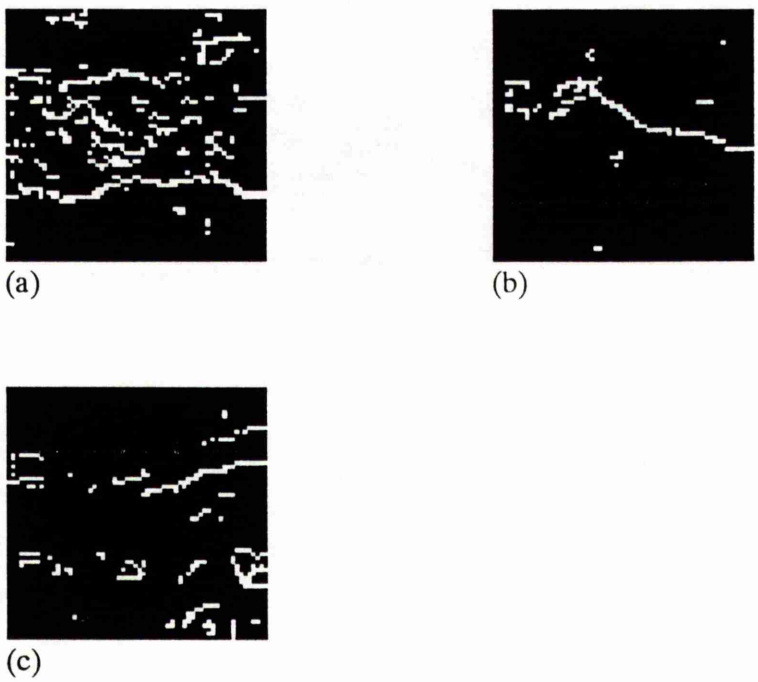
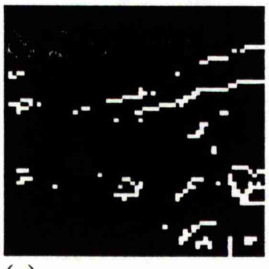


Fig.4.6 Application of the Nagao and Matsuyama (1979) smoothing routine before edge-detection improves the segmentation.

(a)

(b)



(c)

Fig. 4.7 Applying the Nagao and Matsuyama (1979) smoothing routine twice before edge-detection.

The second method of smoothing investigated was median filter smoothing, where the grey level of each pixel is replaced by the median of the grey levels in a neighbourhood of that pixel. Median filtering is often preferred to averaging when the preservation of edge sharpness is required. In order to perform median filtering in a neighbourhood of a pixel the grey levels of the pixel and its neighbours are ordered, the median is determined and this median value is assigned to the pixel. For example, in a 3×3 neighbourhood the median is the 5th largest value, in a 5×5 neighbourhood the 13th largest value and so on.

In this study, 3×3, 5×5, 7×7, 9×9 and 11×11 neighbourhoods were examined. The results for one image are shown in figure 4.8.

3×3

5×5

7×7

9×9

11×11

Fig. 4.8 Median filter smoothing using different neighbourhood sizes, followed by Rosenfeld & Thurston (1971) algorithm.

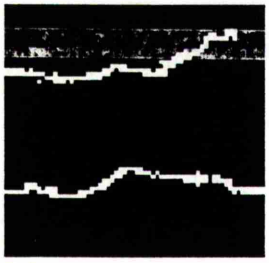
When 3×3 and 5×5 neighbourhoods are used, several edges within the texture are highlighted, suggesting a bigger neighbourhood is needed. 11×11 is too large resulting in the main edge being fairly broken. The best segmentation seems to be achieved by using either a 7×7 or a 9×9 neighbourhood, although the upper edges are fairly broken.

The segmentation results above could posssibly be improved by either:

a) using a larger neighbourhood size and lowering the threshold. This should remove noise and improve the edge (Fig. 4.9(a)),

or

b) using a smaller neighbourhood size and increasing the threshold. This should produce a less broken edge but may result in more noise (Fig. 4.9(b)).



(a)

(b)

11×11
lower threshold

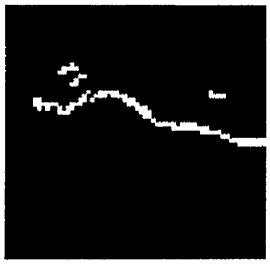
5×5
higher threshold

Fig. 4.9 Median filter smoothing

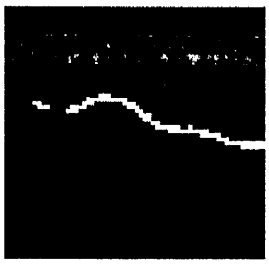
In fact, figure 4.9(a) illustrates that using a large neighbourhood size and lowering the threshold improves the segmentation considerably, giving almost perfect results. On the other hand, figure 4.9(b) shows that the smaller neighbourhood size with a higher threshold produces a broken main edge with some unwanted noise.

When median filtering was applied to images containing types 3 and 5 (Figs. 4.10) similar results were obtained.

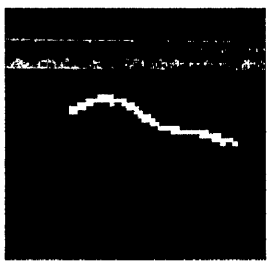
3×3



5×5



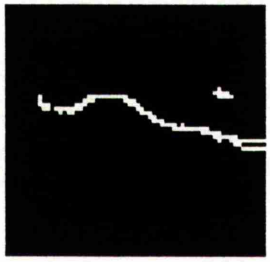
7×7



9×9

Fig. 4.10 median filter smoothing

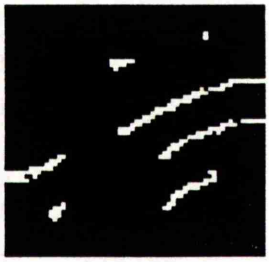
Neighbourhoods of size 7×7 and 9×9 produced fairly good results but, once again, using an 11×11 neighbourhood and lowering the threshold produced better results for the images containing types 3 and 5 (Fig. 4.11(a)). This was also true for images containing types 1 and 5 (Fig. 4.11(b)).



(a)

11×11

threshold=18



(b)

11×11

threshold=15

Fig. 4.11 median filter smoothing

Thus for all images, 11×11 neighbourhoods seem to produce the best results, with a good continuous edge between textures and few edges being detected within edges.

4.6 Relaxation Labelling

Although the methods used so far give fairly good segmentation results for most images, in some images where the boundary between textures is less clear, there is still a need for a method which will highlight the main edge between textures but suppress the smaller, weaker edges within the various texture types. Relaxation labelling is a method introduced by Rosenfeld, Hummel and Zucker in 1976 to reduce ambiguities in an image by using relationships among the objects of the image. Various applications of relaxation labelling have been proposed which include enhancing noisy images (Lev, Zucker & Rosenfeld (1977)) and enhancing lines and curves (Zucker, Hummel & Rosenfeld (1977)). In 1977, Schachter et al. proposed an application of relaxation labelling to reinforce edges.

Relaxation labelling is an iterative process which operates on objects (or pixels) which have labels attached to them. By examining relationships between labels, some labels are strengthened and some are weakened or eliminated.

Let:

$a=\{a_1,a_2,....,a_n\}$ = the pixels of an image

$\Lambda=\{\lambda_1,\lambda_2,...\lambda_m\}$ = the set of labels

and $p_i(\lambda)$ = Prob (λ is the correct label for a_i)

Initially line detection operators in several orientations are applied to the image, and their outputs are used to determine an initial probability that each point lies on a curve in a given orientation. These points are then iteratively reinforced. At each iteration, the new probability $p_i^{k+1}(\lambda)$ is a function of both the old probability $p_i^k(\lambda)$ of the given pixel and the old probabilities $p_j^k(\lambda')$ of neighbouring pixels.

There are 3 basic kinds of contributions that one label λ can make to another label λ on a neighbouring pixel.

1) If the two labels λ and λ' are compatible, i.e. are likely to occur together, then $p(\lambda')$ should contribute positively to $p(\lambda)$.

2) If the two labels λ and λ' are incompatible, i.e. they are unlikely to occur together then $p(\lambda')$ should contribute negatively to $p(\lambda)$.

3) If there is no relation between the labels, then one should not influence the other.

After a few iterations points that lie on smooth curves tend to have high probabilities of being curve points while other points' probabilities are low.

For edge enhancement, only two kinds of labels are considered - edge labels and non-edge labels. Initially a gradient edge detection operator is applied to the image, which gives an edge strength (or probability) and orientation at each point. These points are then iteratively reinforced, similar to the curve case, and new edge probabilities at a point (x,y) are defined in terms of the old probabilities at (x,y) and its neighbours.

Schachter et al. (1977) used a Prewitt operator to determine initial probabilities

and directions.

The Prewitt operator involves applying:

$$D_x = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

and

$$D_y = \begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

to an image.

The probability of an edge at a given point (x,y) is given by:

$$P(x, y) = \frac{mag\ (x, y)}{\max\limits_{u,v} mag\ (u, v)}$$

where max is taken over the entire image, and

$$mag = \sqrt{(D_x)^2 + (D_y)^2}.$$

The orientation is given by:

$$\theta = \tan^{-1} \frac{D_y}{D_x} .$$

The probability of a non-edge at (x,y) is defined by:

$$\bar{P}(x, y) = 1 - P(x, y).$$

Let α be the edge slope at (x,y), β the edge slope at (u,v), γ the slope of the line joining (x,y) to (u,v) and D the chessboard distance from (x,y) to (u,v) i.e. max (x-u, y-v). Then the edge/edge reinforcement process between the points (x,y) and (u,v) has strength given by:

$$R_{ee} = \frac{\cos(\alpha - \gamma)\cos(\beta - \gamma)}{2^D}.$$

The edge probability at (x,y) is weakened by the nonedge probability at (u,v) to the degree R_{en} defined by:

$$R_{en} = \frac{\min(0, -\cos(2\alpha - 2\gamma))}{2^D}.$$

The nonedge probability at (x,y) is affected by the edge probability at (u,v) to the degree R_{ne} defined by:

$$R_{ne} = \frac{1-\cos(2\beta - 2\gamma)}{2^{D+1}}.$$

The nonedge probabilities at (x,y) and (u,v) reinforce each other to the degree R_{nn} defined by:

$$R_{nn} = \frac{1}{2^{D}}.$$

For each point (x,y) the net effect of its neighbouring points on its edge probability P(x,y) and non-edge probability

$$\bar{P}(x, y) = 1 - P(x, y)$$

is computed as follows:

$$Q(x, y) = \sum_{u, v} C_1 P(u, v).\ R_{ee}((x, y), (u, v)) + C_2 \bar{P}(u, v).\ R_{en}((x, y), (u, v))$$

$$\bar{Q}(x, y) = \sum_{u,v} C_3 P(u, v).\ R_{ne}((x, y), (u,v)) + C_4 \bar{P}(u, v).\ R_{nn}((x, y), (u, v))$$

where C_1, C_2, C_3, C_4 are constants whose sum is taken to be 1.

$$Let \quad \hat{Q} = \frac{Q}{|Q| + |\bar{Q}|}$$

$$\bar{\hat{Q}} = \frac{\bar{Q}}{|Q| + |\bar{Q}|}$$

$$P' = P\,(1+\hat{Q})$$

$$\bar{P}' = \bar{P}\,(1+\bar{\hat{Q}})$$

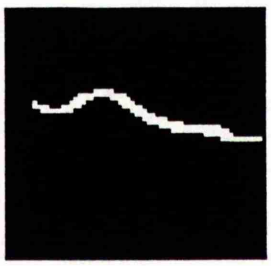
$$and \quad P^{new} = \frac{P'}{P + \bar{P}'} \quad .$$

Schachter et al. (1977) used a 5×5 square neighbourhood centred at (x,y). Thus 24 (u,v)'s influenced a given (x,y). The values of the constants used were as follows:

$$C_1 = 0.866 \quad C_2 = 0.124 \quad C_3 = C_4 = 0.005$$

Applying this method to the 3 images produced the output in Figure 4.12.

(a)



(b)

(c)

Fig. 4.12 Relaxation labelling after median filter smoothing.

Schachter et al. (1977) warned that the results are sensitive to the values of C_1-C_4. Large values of C_1 tend to thicken the edges and large values of C_4 lead to gaps appearing in the edges. Various values of C_1-C_4 were experimented with in this study but altering the values did not appear to make a great deal of difference and certainly did not improve the results. For example, taking $C_1 = C_4 = 0.4$ and $C_2 = C_3 = 0.1$ produced the image in Fig. 4.13.

Fig. 4.13
Relaxation labelling
with $C_1=C_4=0.4$ and
$C_2=C_3=0.1$.

Considerably reducing the value of C_1 has not made the edge any thinner and increasing the value of C_4 has not led to fewer gaps in the edge.

Considering the results in Figure 4.12, the segmentation is good in image 4.12(b) containing types 3 and 5 but there is no improvement in the images containing types 1 and 5 and 2 and 4. Here, the edges are broken and other edges within the texture types are highlighted. The main reason for this is that this application of relaxation labelling is not detecting texture edges. The initial probabilities are found using a simple differential edge detector and therefore detect the edges within the textures as well as those between textures. Even after several iterations the within-texture edges are highlighted. This is not surprising since they are not simply random noise, but are indeed edges with high edge strengths.

One solution to this would seem to be to compute texture edge strengths as the initial probabilities instead of grey-level edge strengths i.e. use the Rosenfeld & Thurston (1971) algorithm to compute the initial probabilities and then iteratively reinforce these texture edge probabilities. In order to do this different orientations from those used previously would have to be considered. Previously, possible orientations were 0°, 45°, 90° and 135°, i.e. →

↙ ↓ ↘ and all combinations of these could reasonably occur together, e.g. 0° and 45° →

↘.

Indeed, Schachter et al. (1977) state that orientations have to be specified modulo 360 rather than modulo 180, i.e. it is necessary to differentiate between 0° → and 180° ←.

A similar, but simpler technique could use edge strengths of neighbouring points to reinforce edges iteratively by firstly picking out points with high edge strengths, i.e. 'definite' edge points, and then looking at the neighbours of each of these points. If there is a neighbour with a 'relatively high' edge strength it is fairly likely this will also be an edge point. This procedure could be continued either until the boundary of the image is reached or there are no neighbouring pixels with the desired edge strength. This procedure is known as raster tracking.

4.7 Raster Tracking

Raster tracking is often used to extract objects by tracking them from row to row of the picture as the picture is scanned row by row in the manner of a T.V. raster. The same idea can be used to track an edge. Specifically, in each row a pixel is considered to lie on an edge if its edge strength exceeds some relatively high threshold d. In addition, once a point (h,k) is accepted as a point on an edge we accept any neighbour of (h,k) in seven directions: ↖↑↗

→

↙↓↘

provided that the neighbouring points have edge strength above some lower threshold t.

As a simple example consider the image in figure 4.14 (a)

Fig.4.14

$$\begin{bmatrix} 9 & 2 & 5 & 3 & 8 & 3 & 4 & 5 \\ 2 & 5 & 3 & 3 & 4 & 3 & 4 & 5 \\ 6 & 3 & 9 & 3 & 7 & 2 & 5 & 4 \\ 2 & 3 & 3 & 4 & 3 & 3 & 4 & 5 \end{bmatrix}$$

(a) Input image

$$\begin{bmatrix} 255 & 0 & 0 & 0 & 255 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 255 & 0 & 255 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(b) Result of thresholding (a) at 7.

$$\begin{bmatrix} 255 & 0 & 255 & 0 & 255 & 0 & 255 & 255 \\ 0 & 255 & 0 & 0 & 255 & 0 & 255 & 255 \\ 255 & 0 & 255 & 0 & 255 & 0 & 255 & 255 \\ 0 & 0 & 0 & 255 & 0 & 0 & 255 & 255 \end{bmatrix}$$

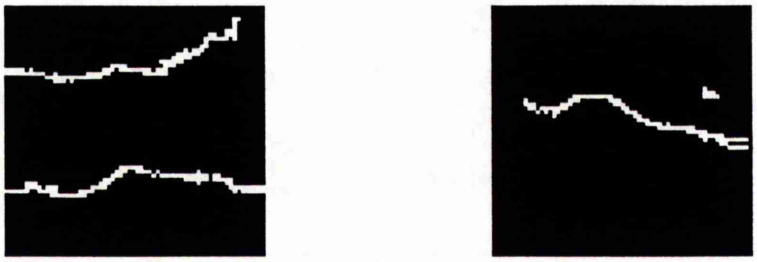
(c) Result of thresholding (a) at 4.

$$\begin{bmatrix} 255 & 0 & 0 & 0 & 255 & 0 & 0 & 0 \\ 0 & 255 & 0 & 0 & 255 & 0 & 0 & 0 \\ 0 & 0 & 255 & 0 & 255 & 0 & 0 & 0 \\ 0 & 0 & 0 & 255 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(d) Result of thresholding
with d=7, t=4.

If a high threshold is chosen, the edge has several gaps in it (Fig. 4.14(b)). If, on the other hand, the threshold is too low, too many pixels are highlighted (Fig. 4.14(c)). Tracking, using a combination of (b) and (c) gives a continuous edge with no added noise (Fig. 4.14(d)). The process could be applied iteratively to fill in gaps in edges.

Applying the algorithm to the rats' bone images, initially using a relatively high threshold to detect 'definite' edges, and then thresholding all neighbouring pixels of these points at a lower level produced the images in Figure 4.15.



(a) (b)

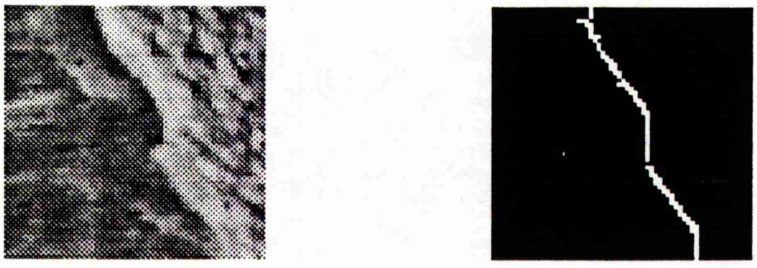
(c)

Fig.4.15 Result of tracking after applying Rosenfeld & Thurston (1971) edge-detection algorithm.

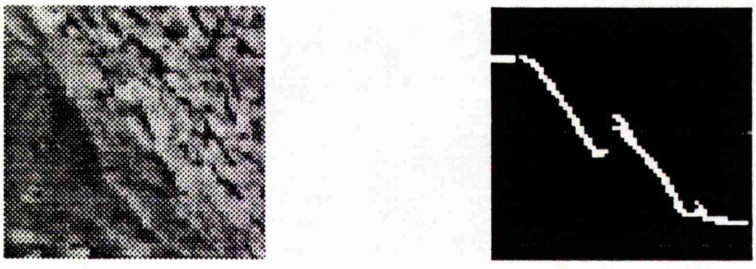
Although there was little room for improvement in the type 2 and 4 and type 3 and 5 images, raster tracking has had the effect of filling in gaps in the edges. The most dramatic result, however, is in the image containing types 1 and 5. Whereas before there were lots of very broken edges, now there is one continuous edge with no gaps. Although there are still a few pixels highlighted which do not belong to the edge, an operator would be able to recognise the edge between the texture types, and erase the remaining pixels.

So far all the segmentation techniques have been tested on the original set of images collected. As mentioned in chapter 3 a new set of images was collected later in the study. There were several differences in the method of capture of these images

(section 3.10) which resulted in difficulties when classifying the two separate data sets. The same problems should not arise when segmenting the images and, because the differences in the texture types are more pronounced, it was hoped that the results would be at least as good as those obtained with the original set of images. To test this, the technique used above was applied to several 64×64 windows taken from the new set of images. A selection of these windows and the resulting segmented binary images are shown in figure 4.16.

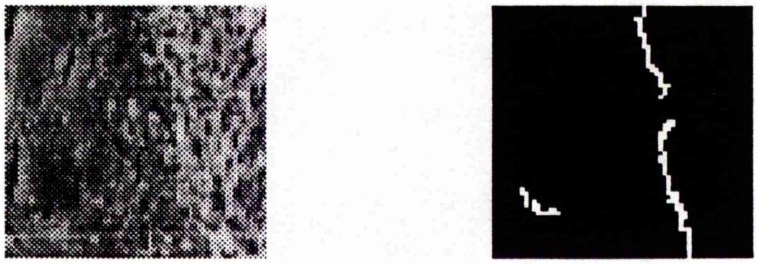


(a)

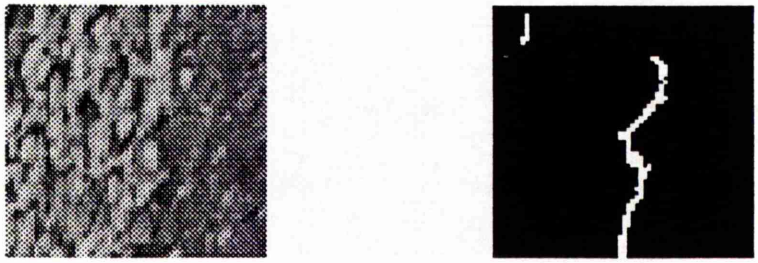


(b)

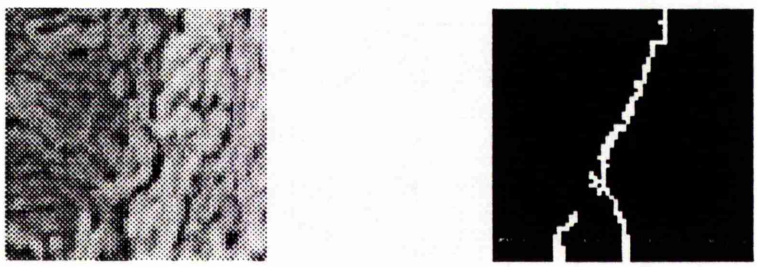
(c)



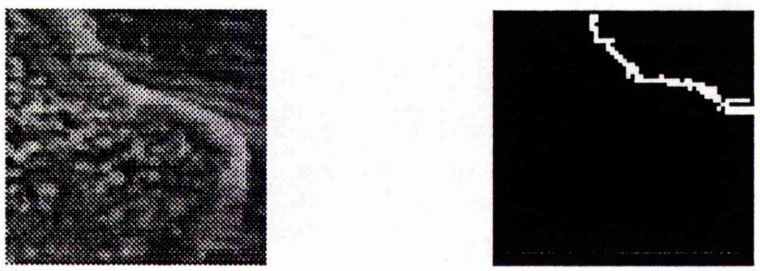
(d)



(e)



(f)



(g)

Fig. 4.16 64×64 windows and their resulting segmented, binary images.

In Fig. 4.16 (a), (c) and (g) the boundaries between the textures are quite clear

and the segmentation algorithm has identified fairly good, well-connected edges. The boundaries between the texture types in Fig. 4.16 (b), (d), (e) and (f), on the other hand, are far less clear, but again the technique has performed well, producing almost continuous edges. Although there are a few broken patches in some of the images, on the whole, the results are good, with fairly continuous edges being detected and very few non-edge pixels being highlighted.

Thus, in summary, for all the images in this study good segmentation results have been obtained when median filter smoothing using an 11×11 window was performed on the images before applying the Rosenfeld & Thurston (1971) edge-detection routine, but restricting the maximum possible edge size to 3, corresponding to 8×8 pixel windows. The edges were then tracked and finally thresholded to produce the segmented binary image.

At this stage, the two thresholds used in the tracking procedure were decided subjectively. Further work was needed to find an objective method of finding the optimum thresholds. This is discussed in the next section.

4.8 Thresholding

Thresholding, which converts a grey-level image into a binary one, is a widely-used tool in image segmentation. The two levels involved may represent objects and background, or two classes in an image. Pixels whose value exceeds a critical value (the threshold) are assigned to one category, and the rest to the other. If the same critical value is used throughout the image the threshold is said to be global, if different values are used the threshold is local. Many algorithms have been proposed for automatically selecting the threshold appropriate for a given image. Comparisons and evaluations of these methods can be found in surveys by Sahoo et al. (1988), Lee et al. (1990) and Glasbey (1993).

Global histogram-based algorithms are the most commonly used despite the possible benefits of using contextual information and allowing the threshold to vary throughout the image. Their main advantages are that they are relatively simple to implement and computationally fast once the histogram has been obtained. Lee et al. (1990) compared three histogram-based algorithms with two contextual ones. The five

methods were: 1) Simple Image Statistics Method (SIS), Kitler & Illingworth (1985), 2) Between Class Variance Method (BSV), Otsu (1979), 3) Entropy Method (ENTROPY), Kapur et al. (1985), 4) Moment Preserving Method (MOMENT), Tsai (1985) and 5) Quadtree Method (Q-TREE), Wu et al. (1982). The authors tested the methods on two 256x256 images - both were objects against a background. The first object was a computer disk and the second a crane. In both images the variances of background and image were quite large and in the crane image the object, i.e. the crane, was relatively small. The conclusion that the authors came to was that all the methods were image-dependent, i.e. they depended on various factors including the object size, the mean difference between background and object, the variances of background and object, contrast and noise. They suggested that SIS and BCV are probably best. They both performed well regardless of the object size, and although BCV is sensitive to noise, SIS works quite well on noisy images.

Glasbey (1993) reviewed eleven histogram-based thresholding algorithms. In this review images were assumed to have bimodal Gaussian distributions. The simplest techniques involved taking the threshold t such that 50% of pixels lie in both categories (MEDIAN), taking t as the MEAN grey-level, choosing t as the minimum grey level in the valley between the two maxima (MINIMUM), and finding the mean of the two maxima (INTERMODES). Other methods were iterative, and Glasbey (1993) found that the method of Kittler and Illingworth (1986), which he called MINERROR (I), was best. This method minimises the number of misclassifications between the classes, and Glasbey (1993) claims that for a bimodal Gaussian distribution this method is one of the simplest to compute, it fails infrequently, and is relatively insensitive to the effects of sampling variability.

In their survey in 1988, Sahoo et al. reviewed both local and global thresholding techniques, as well as histogram-based and contextual algorithms. Point-dependent methods included Otsu's method (1979), based on discriminant analysis, entropic methods of Pun (1980), Kapur et al. (1985) and Johannsen and Bille (1982), a moment-preserving method (Tsai (1985)) and a minimum error method (Kittler and Illingworth (1986)). Region-dependent methods included histogram transformation methods, methods based on co-occurrence matrices and gradient relaxation methods. Sahoo et al. (1988) evaluated nine of the global thresholding methods using 3 images - the

'cameraman', the 'building' and the 'model'. The cameraman and building images were both bimodal, whereas the model was less regular. The authors found the best methods to be the entropic methods of Johannsen and Bille (1982) and Kapur et al. (1985), Otsu's (1979) discriminant analysis method, and the moment-preserving method of Tsai (1985), with Otsu's (1979) method ranked in the top four for all three images.

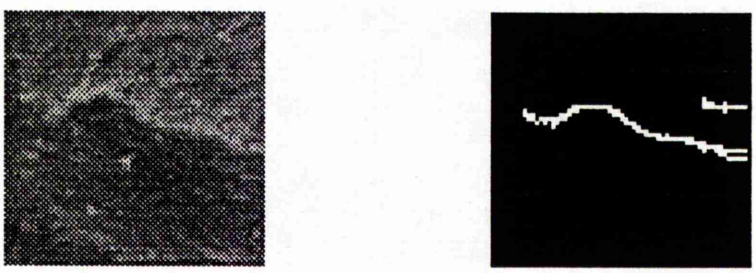
More recently, methods of peak detection have been suggested for threshold selection (Eklundh & Rosenfeld (1979), Olivo (1994)). Segmentation by peak detection methods is based on the clustering of grey levels around peaks of the histogram in order to define homogeneous grey-level areas. In this approach, peaks and valleys are first detected, from which thresholds are set to form grey-level clusters. Olivo's (1994) method is based on the wavelet transform and he used it for multilevel thresholding. He proposed that the detection of zero-crossings and the local extrema of a wavelet transform of the histogram give a complete characterisation of the peaks in the histogram - i.e. the values at which they start, end and are extreme. He tested his method on two images - the 'girl' and the 'garden'. The histogram of the garden was bimodal but that of the girl was multimodal. Using a scale of 2^2 the results were good. The author claimed that the major features of his method are a) it requires no preassigned parameter, so no trial-and-error stage is needed; b) it provides a meaningful sequence of signals which in turn gives a meaningful description of the image histogram; and c) it is computationally effective.

When segmenting images in this study using the tracking technique described above, two thresholds are required to be chosen. These are based on the edge strengths of the image; the first, higher threshold, to detect 'definite' edge pixels, and the second, a lower threshold, to fill in gaps and complete the segmentation. Histograms of edge strengths tend to be unimodal, and skewed to the right, with the majority of edge strengths being low or zero, and relatively few high edge strengths. In some histograms there is a slight jump in edge strength at certain points, whereas in others, there is simply a gradual decline.

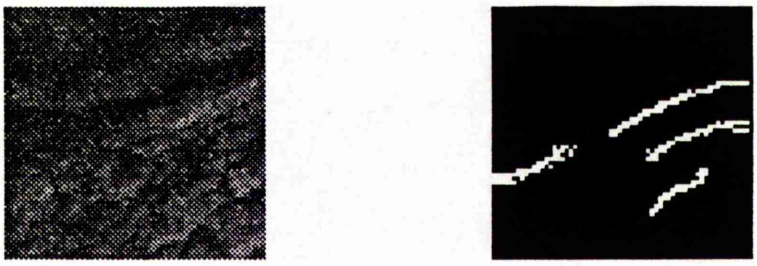
Some of the simplest methods mentioned by Glaseby (1993) such as MEDIAN and MEAN are not appropriate for this problem because it is not possible to specify in advance how much of the image consists of edge points. Glaseby (1993) recommended the minimum error method of Kittler & Illingworth (1986) but he only considered

mixtures of two Gaussian distributions. On the other hand, Sahoo et al. (1988) considered both smooth bimodal and less regular histograms. Otsu's (1979) discriminant method seemed to work well on all the images in their study. Sahoo et al. (1988) used the technique to find one threshold but Otsu (1979) proposed that the method could easily be extended to find multiple thresholds. In addition, Otsu (1979) showed that the method worked well on textured images where grey-level histograms tend to have a peak at one end and long flat valleys. This method was therefore tested on some of the images in this study. Here the method was adapted to find the two thresholds required for tracking. At first the results were very poor with the resulting thresholds being consistently much lower than those chosen subjectively by eye. Due to the edges being 'thinned', however, the vast majority of edge strengths were zero and this had the effect of producing thresholds which were too low. By omitting zero edge strengths, however, the method performed remarkably well on both the older images and the new set of images. The results are shown in Figures 4.17 and 4.18.

(a)

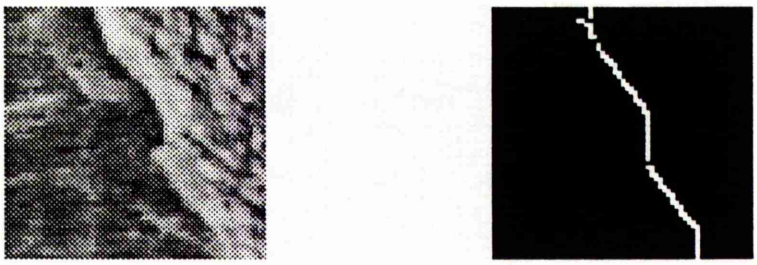


(b)

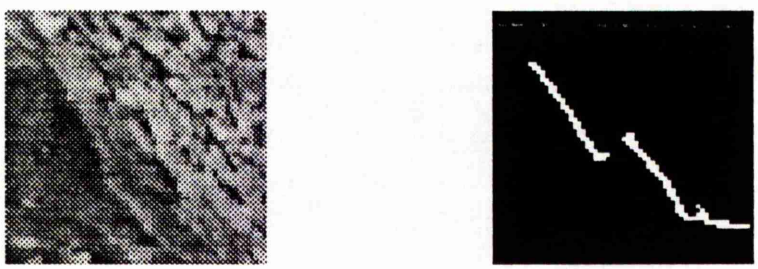


(c)

Fig. 4.17 Original images and their segmented, binary images, after automatic thresholding.

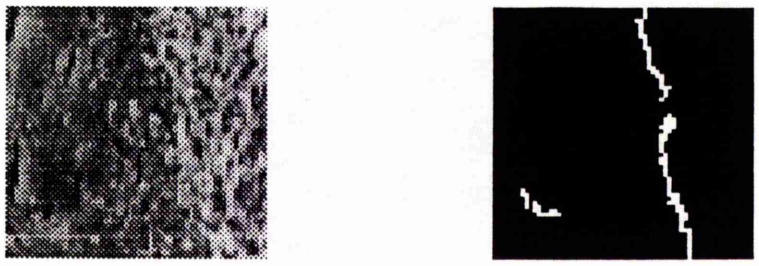


(a)

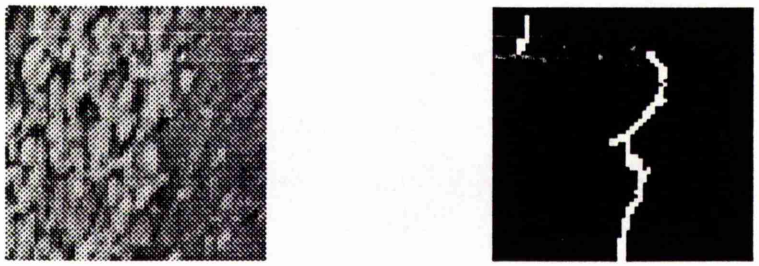


(b)

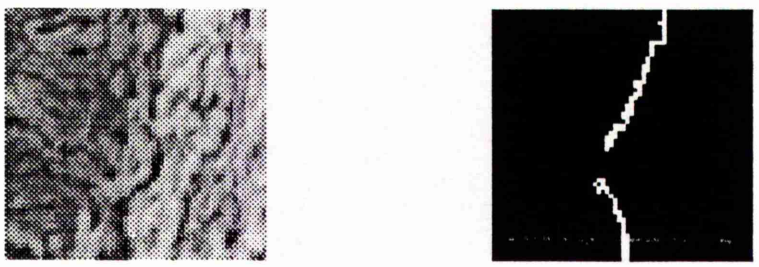
(c)



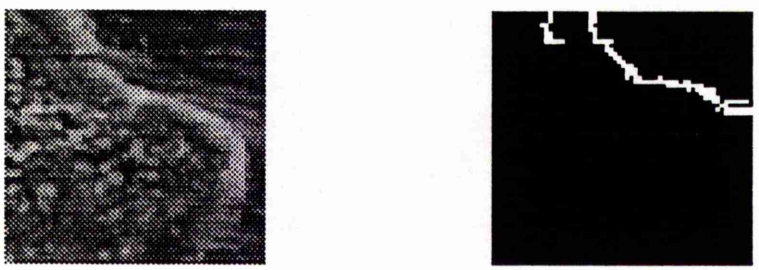
(d)



(e)



(f)



(g)

Fig 4.18 New images and their segmented, binary images, thresholding automatically.

Firstly comparing figures 4.15 and 4.17 the binary images are very similar. In

4.17(c) the edge is no longer completely connected but the position of the edge can be clearly seen. In both images 4.17(a) and (b) the thresholding has worked well and the edges are clearly visible. Comparing figures 4.16 and 4.18, again it can be seen that the thresholding algorithm has worked well. For some of the images the results using automatic thresholding are exactly the same as using trial-and-error, and for the rest the results are very similar. In all cases, apart from perhaps 4.18(c) the position of the edge has been detected and any gaps could be filled in by an operator. The fact that the segmentation in some cases is not quite as good using the automatic procedure as when using trial-and-error, is far outweighed by the time saved when thresholding automatically.

4.9 Conclusions

Finding edges between textures is a much more difficult problem than grey-level edge-detection. This chapter has involved investigating several published techniques for texture segmentation and applying them to, and modifying them for use on, the data set in this study. Probably the best known method used for segmenting texture is the method suggested by Rosenfeld & Thurston in 1971. The idea of averaging grey levels in windows of varying sizes and looking for adjacent windows with large differences in these averages is intuitively sensible. This was therefore the first segmentation technique to be tested in this study. For some images the method worked reasonably well and identified the edges between the different textures, but the images were extremely noisy and it was obvious that some smoothing technique was going to be required before the segmentation procedure was applied. A so-called edge-preserving technique (Nagao & Matsuyama (1979)) was tested but this was found to work less well than a simple median-filter smoothing algorithm, using 11×11 windows. Applying the segmentation technique (Rosenfeld & Thurston (1971)) with a few modifications to the smoothed images considerably improved the results and where there was a reasonably clear continuous edge in the original image, this was detected. The only problem then lay with images where the edge was either less clear or broken in places. Relaxation labelling is a method that has been used to strengthen or weaken possible

edge pixels based on the edge strengths of neighbouring pixels. This seemed a sensible approach to take but after fairly extensive testing was found not to improve the segmentation in this case. Gaps in edges were not filled, and all that seemed to happen was that edges became thicker. This was not the desired result and it was decided to investigate other techniques. In fact, the technique that produced the best results was a fairly simple one based on 'tracking' edges. In this technique, initially a high threshold is used to detect 'definite' edge pixels. The neighbours of these pixels are then examined and if they have edge strengths above a slightly lower threshold than the original one they too are considered to be edge pixels.

Thus, in summary, the best results were obtained when median filter smoothing using an 11×11 window was performed on the images before applying the Rosenfeld & Thurston (1971) edge-detection routine, but restricting the maximum possible edge size to 3, corresponding to 8×8 pixel windows. The edges were then tracked and finally thresholded using a discrimination technique to produce the segmented binary image.

Although the methods were tested on a fairly small number of images, the results were good. In images where there was a definite boundary between texture types, fairly fine, continuous edges were detected, and, by applying the tracking algorithm, even where the edges in the original images were not clear continuous edges were detected.

The methods above, however, do not produce 'perfect' segmentation in every case and some further refining would be necessary to make the process fully automatic. At this stage, however, an operator could clearly identify the texture edges and carry out any necessary refining such as filling in gaps in the edges and erasing spurious highlighted pixels. Certainly, considerable steps have been taken towards making the segmentation process automatic. The position of boundaries between texture types can now be found objectively and the automatic thresholding technique makes the whole procedure significantly faster than when thresholding had to be done by trial-and-error.

This chapter has concentrated on segmentation between formative and resorptive types, i.e. between types 1, 2 and 3 and types 4 and 5. Resorptive types occur where bone has been eaten away, they lie at a lower level from formative types, and there is therefore often a boundary between them. On the other hand, formative types 1, 2 and 3 merge into one another and there is therefore no boundary as such. Similarly, texture type 4 merges into type 5. Due to the lack of boundary, segmentation techniques are

not successful in discriminating between types 1, 2 and 3, or types 4 and 5. For this reason classification techniques, as described in chapter 3 will be used to 'segment' them. The results are shown in the next chapter.

Chapter 5 Combining Segmentation and Classification

5.1 Introduction

All the work in chapters 3 and 4 has involved 64×64 windows containing either one texture type, in the images in chapter 3, or two types, in the images of chapter 4 where segmentation was the goal. This chapter now extends the techniques used to 256× 256 images. The aim is to segment into formative and resorptive types using the techniques of chapter 4, partition the images up into smaller regions (64×64), and classify each region into one of the 5 types using the techniques of chapter 3. Also addressed in this chapter will be the problem of how to deal with Haversian channels and bone cells.

5.2 Haversian Channels

Present throughout bone, Haversian or vascular channels contain the blood vessels and nerves of the bone. In cross-section they appear as round dark openings, often with bright edges. Due to the clear boundary, almost any edge-detection routine should be able to detect the edge. Their dark interior, however, suggests that probably the simplest way of detecting vascular channels, and classifying them as such, is by grey-level thresholding. Figure 5.1 below shows 3 examples of 128×128 windows containing vascular channels, and the results of thresholding at 35, after median-filter smoothing. Thus, if

$$IP(i,j) < 35, \text{ set } IP(i,j) = 0,$$
$$\text{otherwise, set } IP(i,j) = 255.$$

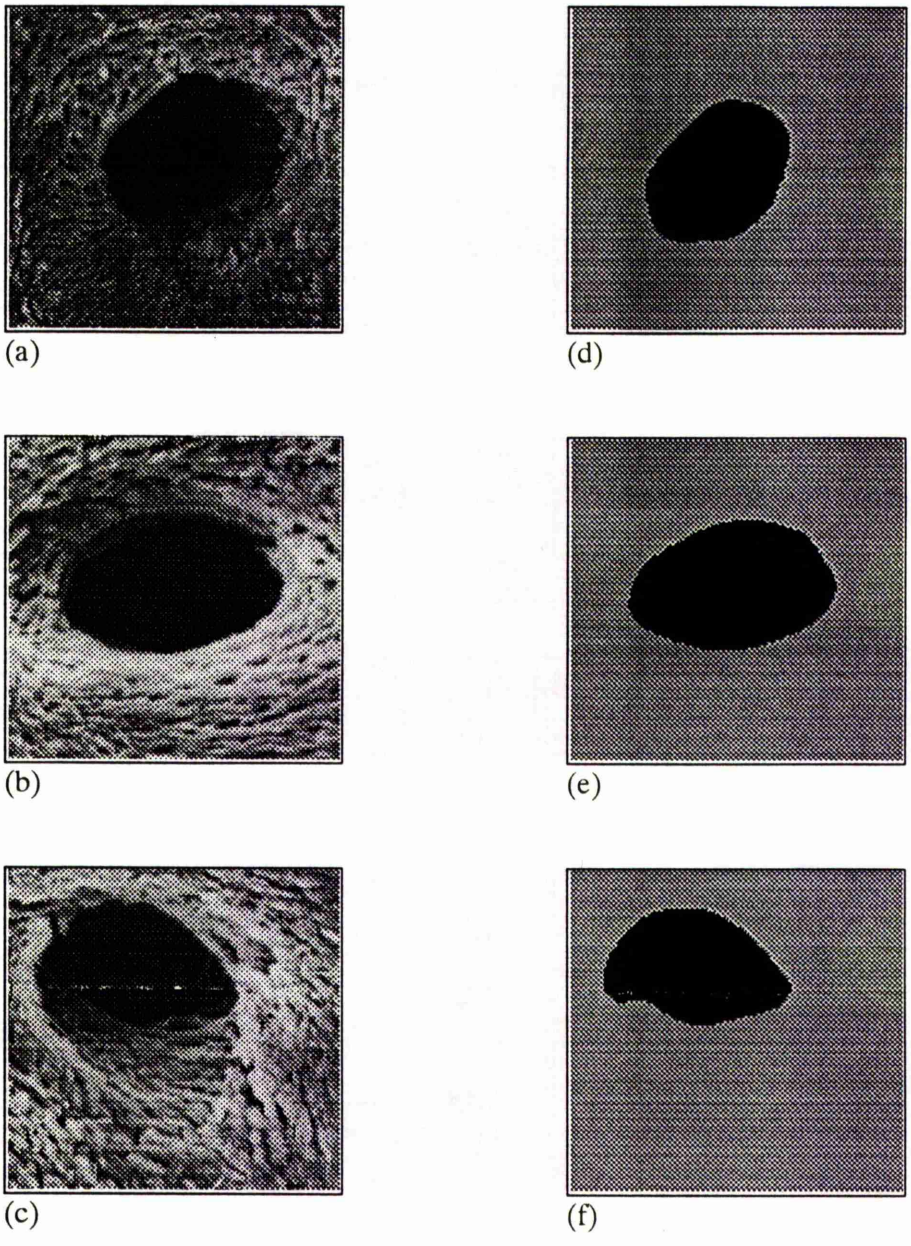


Fig. 5.1 Images containing vascular channels (a,b and c), and their corresponding thresholded images (d,e and f).

In the example above, a threshold of 35 successfully segmented all the images, even though no standardisation had been applied. This, however, will not necessarily be the case for all images in the future, and so the images were standardised using the

histogram equalisation technique described in chapter 3. At this stage, because only thresholding is being used, it is probably sensible to retain as much grey level information as possible. The images have therefore been standardised to contain 256 grey levels. The results, thresholding at 50, are shown below (Fig. 5.2).

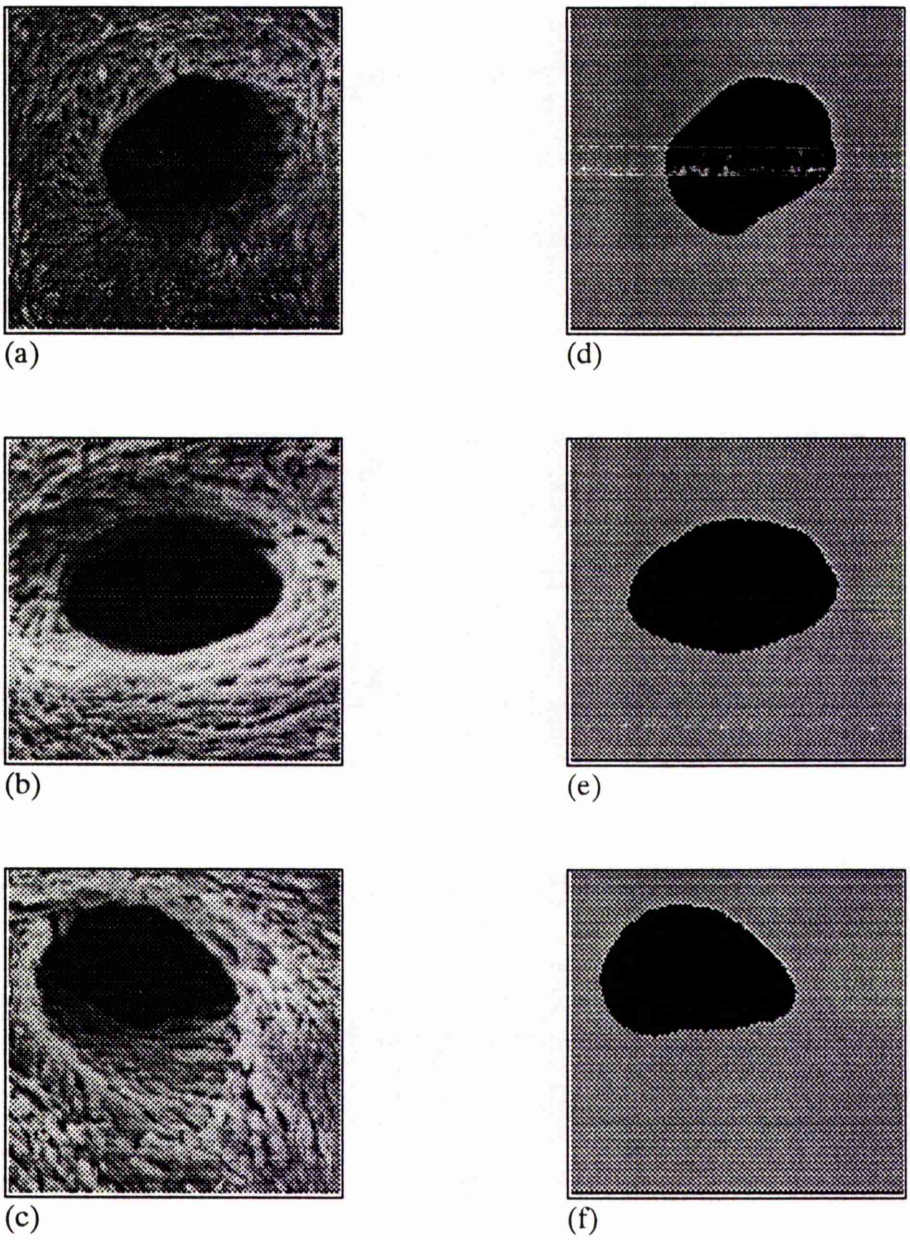
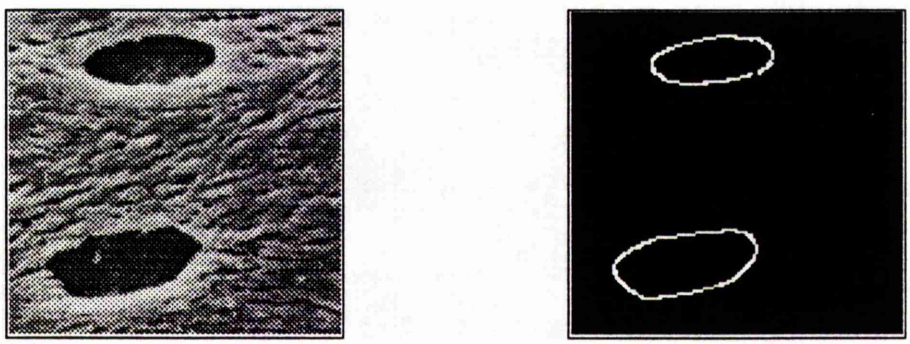


Fig. 5.2 Images containing vascular channels (a,b and c), and their corresponding thresholded images after histogram equalisation (d,e and f).

Thus, because of their dark appearance, vascular channels can be fairly easily detected by thresholding histogram equalised images.

5.3 Bone Cells

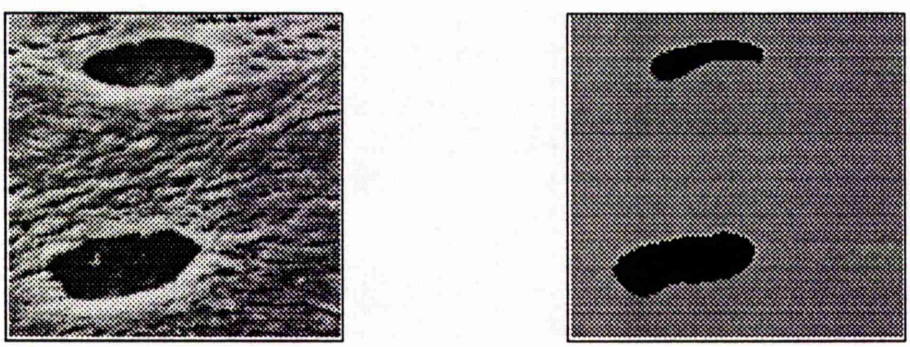
Osteocytes can be identified in a similar way to Haversian channels. Osteocytes also appear as dark, fairly regular ring-shaped regions with bright edges. As with vascular channels, they can be detected by an edge-detector (Fig. 5.3).



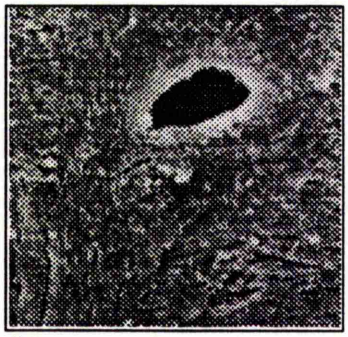
(a) (b)

Fig. 5.3 Image containing 2 osteocytes (a), and corresponding binary image after applying texture edge-detector (b).

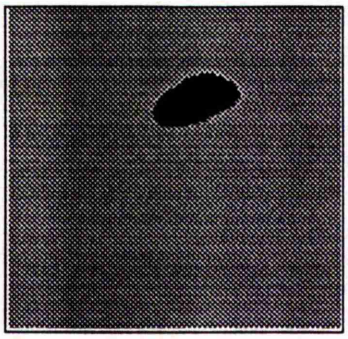
Alternatively, the osteocytes can also be identified by simple thresholding of grey levels after applying median-filter smoothing.



(a) (c)



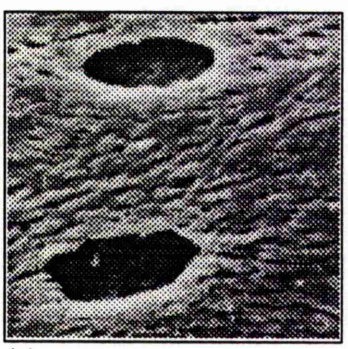
(b)



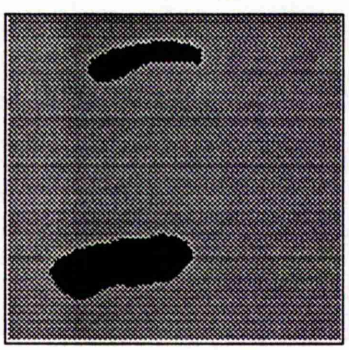
(d)

Fig. 5.4 Images containing osteocytes (a and b) and their corresponding thresholded images (c and d).

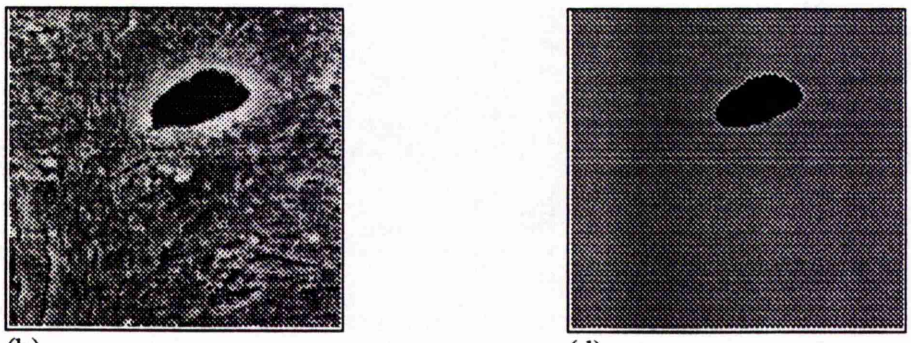
In the case of the osteocytes, the threshold is higher than that for vascular channels. For the two images above the threshold used was 75 but this may not be a suitable threshold for all images. This problem can again be solved by histogram equalisation. The results, thresholding at 30 and using standardised images with 256 grey levels are shown in Figure 5.5.



(a)



(c)



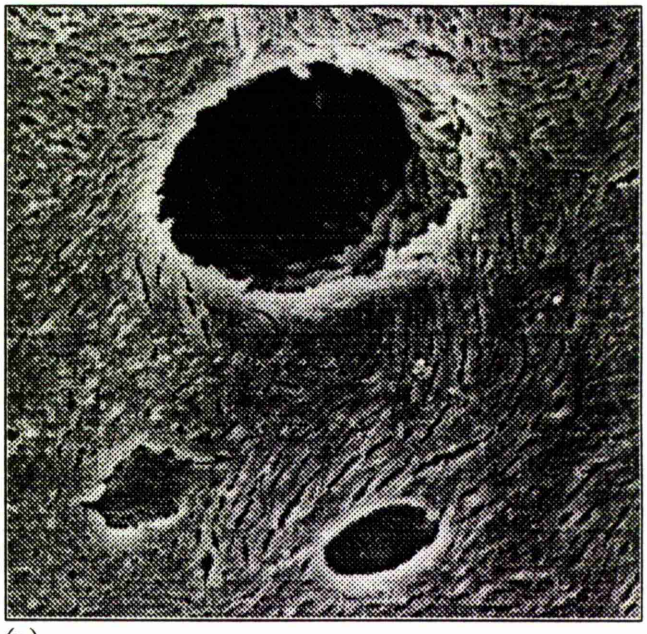
(b) (d)

Fig. 5.5 Images containing osteocytes (a and b) and their corresponding thresholded images after histogram equalisation (c and d).

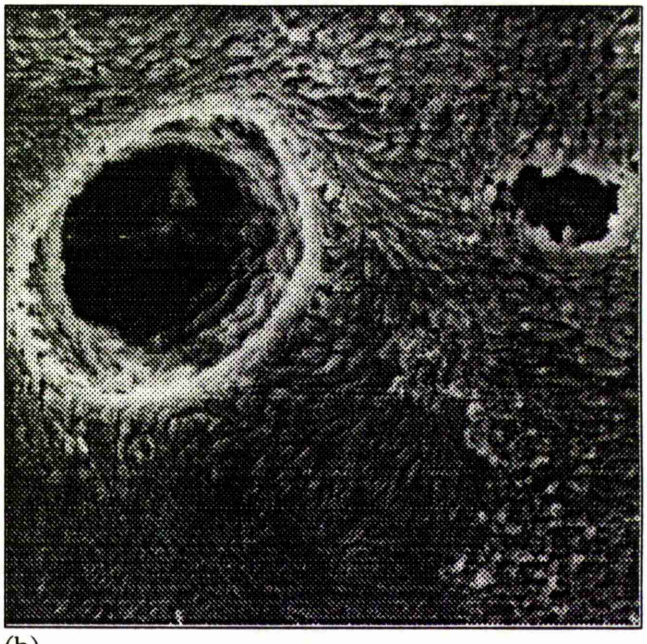
Thus, again, simple thresholding can detect osteocytes. It should be noted though that although vascular channels do appear darker than osteocytes, the threshold used to detect vascular channels from histogram equalised images was actually higher than that used to detect osteocytes. The reason for this is the size of the vascular channels. For example, in figure 5.2(b) the vascular channel takes up about one fifth of the image and will therefore take up about a fifth of the grey levels at the lower end. The osteocytes, on the other hand, are smaller, take up less of the image, and hence less of the grey levels. Thus, the threshold will, in fact, depend on the percentage of the image taken up by the vascular channel.

It is obviously necessary to be able to detect both vascular channels and bone cells in an image. Also osteoblasts are more irregular than the more mature osteocytes, and although osteoblasts do tend to have a bright edge around them they are not as easy to detect as osteocytes. Figure 5.6 below shows two examples of 256×256 images containing both vascular channels and bone cells. Figure 5.6(a) contains one osteocyte and one osteoblast and figure 5.6(b) contains an osteoblast.

Fig. 5.6
Images containing vascular channels and (a) an osteoblast and osteocyte, (b) an osteoblast.



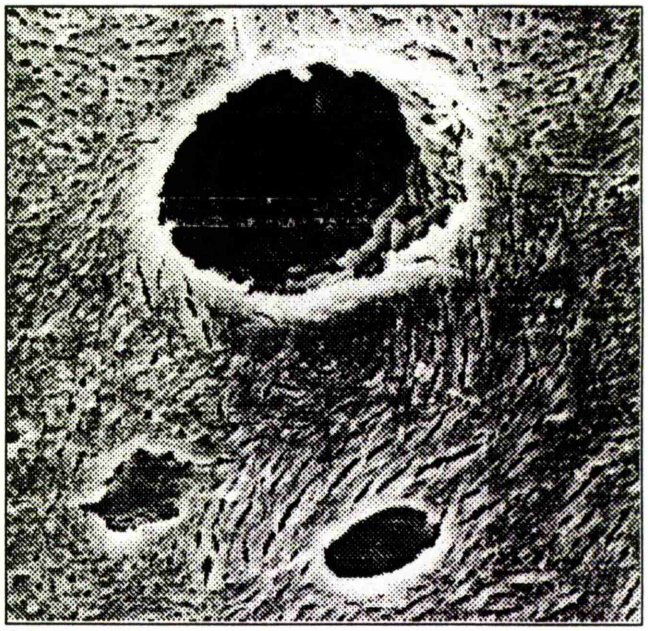
(a)



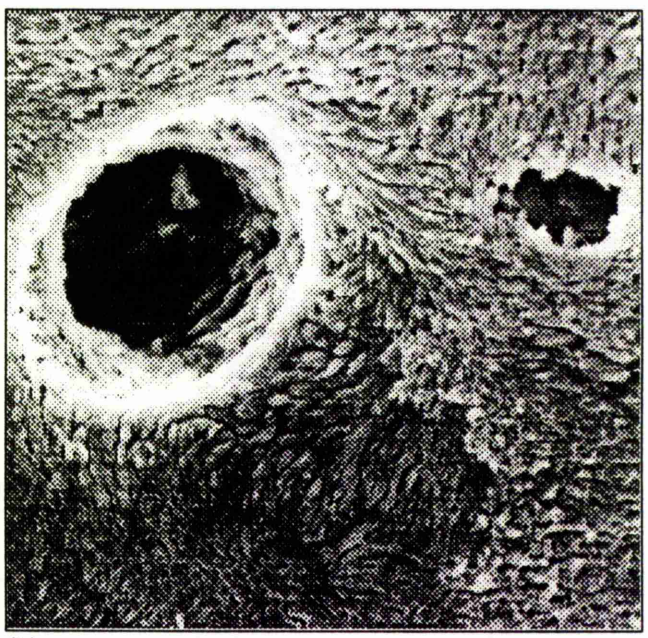
(b)

In both images above, the vascular channels and the bone cells appear darker than the rest of the images. Applying histogram equalisation with 256 grey levels emphasises this further (Figure 5.7).

Fig. 5.7
The images in Fig. 5.6 above after histogram equalisation.



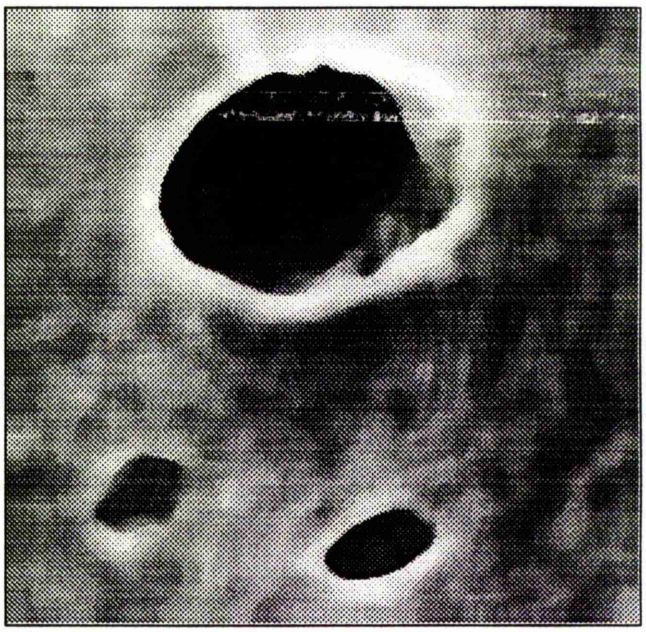
(a)



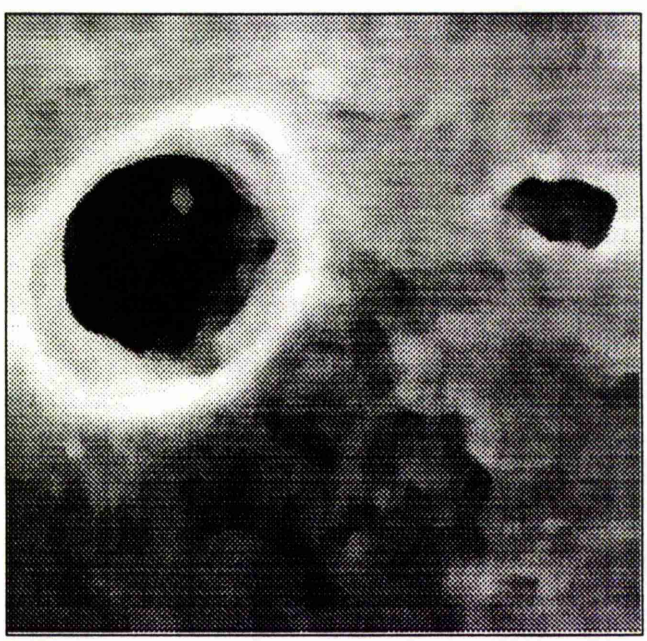
(b)

The interiors of the channels and cells are still dark, and their surrounding edges are brighter than previously. Figure 5.7(b) contains some noise within the vascular channel. This, together with other noise in the images, can be removed by applying median filtering (Figure 5.8).

Fig. 5.8
The result of
applying median
filter smoothing
to the images above.



(a)

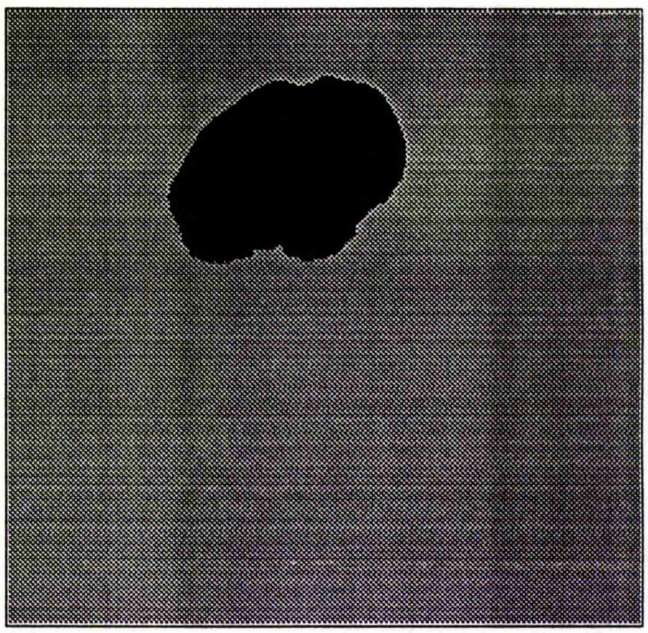


(b)

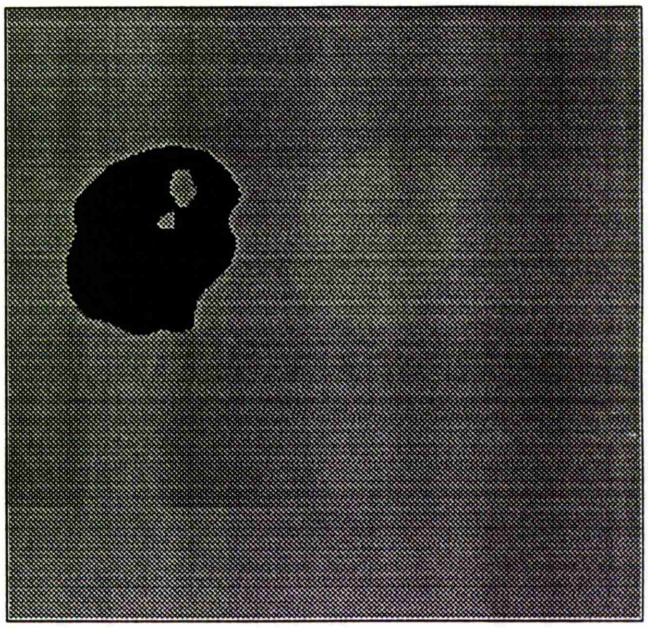
Most of the texture information has been lost, but the bright edges with the dark interiors are still visible. Also, in Figure 5.8(b) the edge between two texture types can be seen.

Thresholding based on grey levels can still detect the vascular channels but cannot always detect bone cells because often the bone matrix is as dark as the cells (Figure 5.9).

Fig. 5.9
Vascular Channels can be identified by thresholding but bone cells are no longer visible.



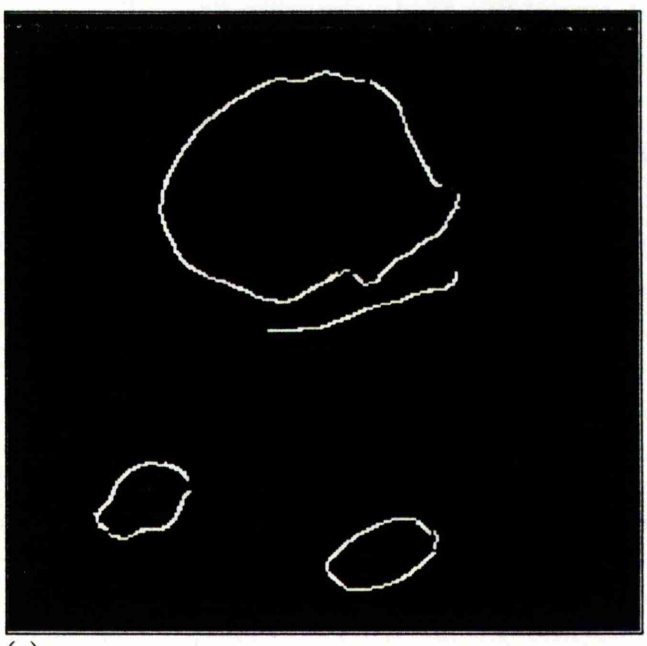
(a)



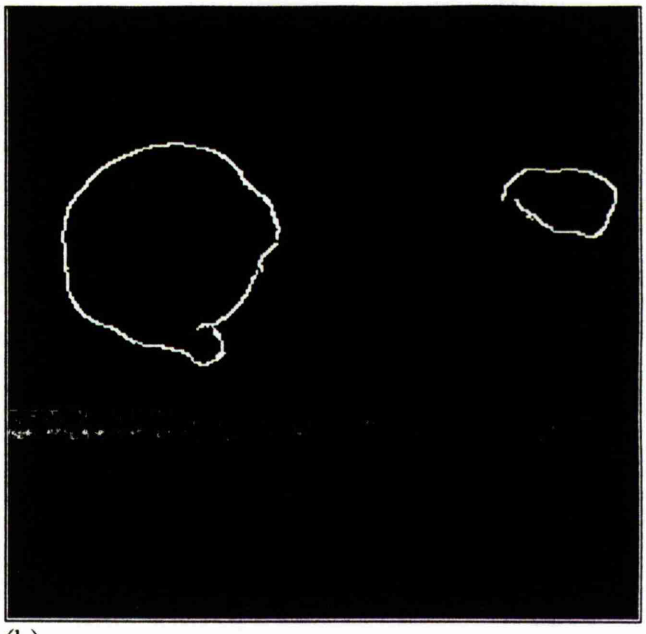
(b)

Edge-detectors can pick up the edge surrounding both the vascular channels and the bone cells, although median filtering often makes the edge around the vascular channels quite thick (Figure 5.10).

Fig. 5.10
Applying a texture edge-detector to the images.



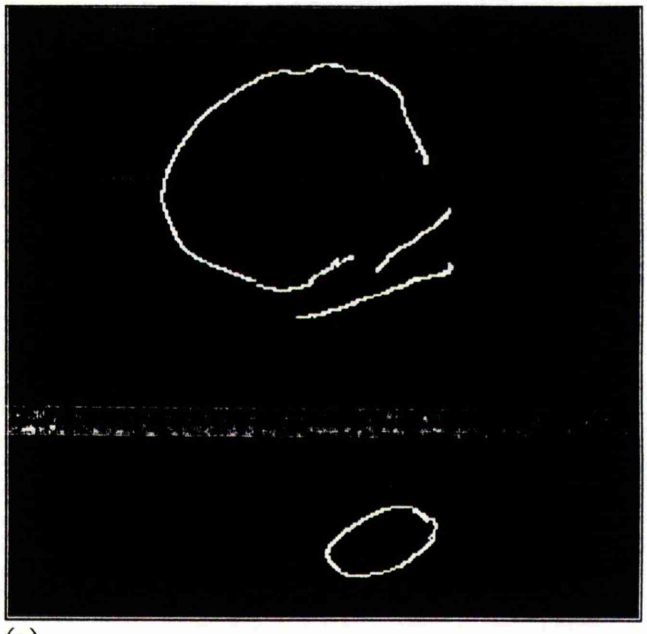
(a)



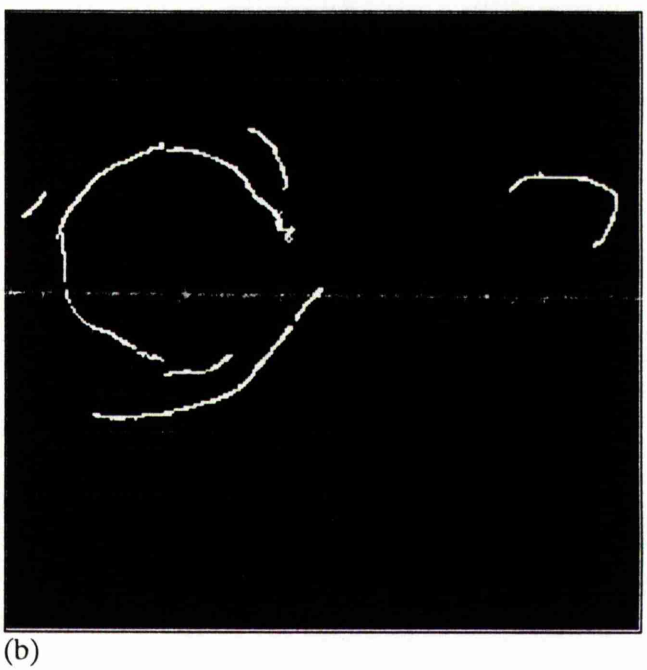
(b)

The thresholds used to identify the edges round the vascular channels and bone cells in the images above were selected subjectively. As with the boundaries between formative and resorptive textures, it would be preferable if the thresholds could be selected automatically. An automatic threshold technique similar to the one that was used in chapter 4 but this time thresholding grey levels instead of edge strengths, and including all 256 grey levels, was applied to the two images. The results are shown in figure 5.11.

Fig. 5.11
Applying a texture edge-detector to the images and thresholding automatically.



(a)



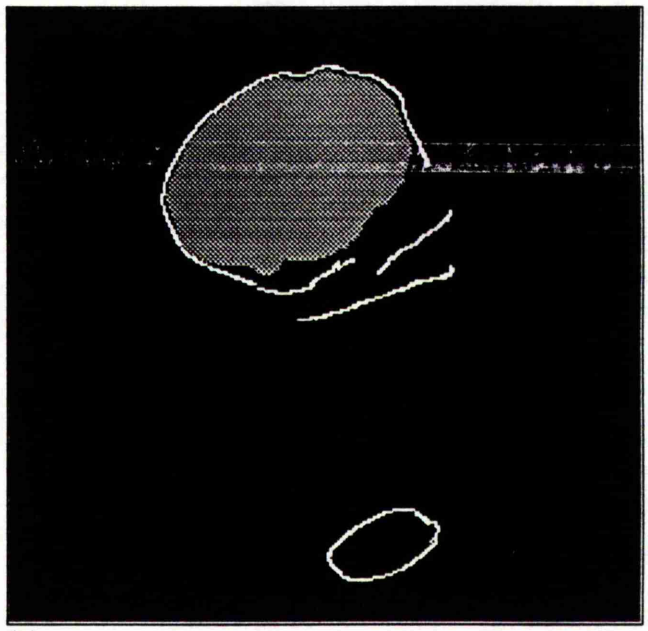
(b)

The results are fairly similar to those obtained by thresholding subjectively. The vascular channels have been identified in both images, although in 5.11(b) part of the outer rim has been highlighted as well as the inner rim. The osteocytes have been at least partly identified in both images but the osteoblast in 5.11(a) has not been highlighted. Thus, although the results are not perfect, it is possible to detect

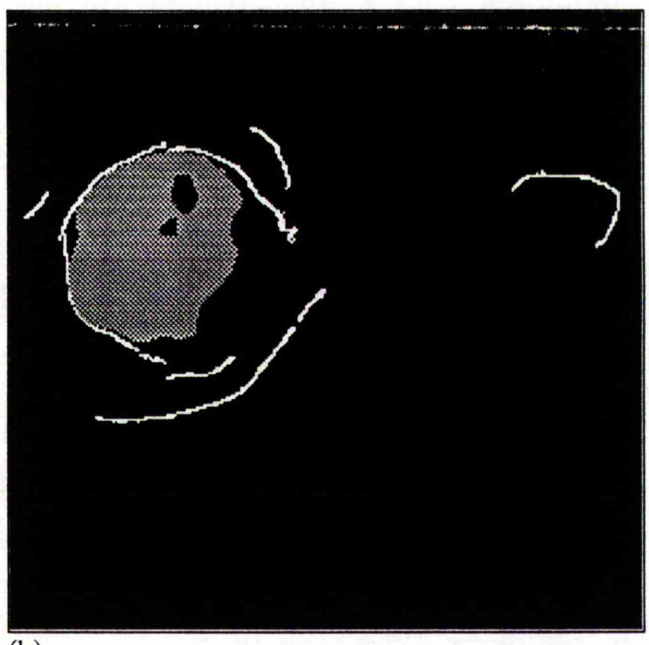
vascular channels and bone cells by automatic thresholding.

The dark vascular channels can then be identified as such by thresholding (Figure 5.12).

Fig. 5.12
The dark interiors of the vascular channels shown in grey with the edges in white.



(a)



(b)

Thus for images which may contain both vascular channels and bone cells the texture edge-detection and automatic thresholding routines should be applied. This will detect the edges around both the vascular channels and bone cells, if present. Then, by thresholding the images, vascular channels can be identified. Other enclosed areas could be classified as bone cells by an operator.

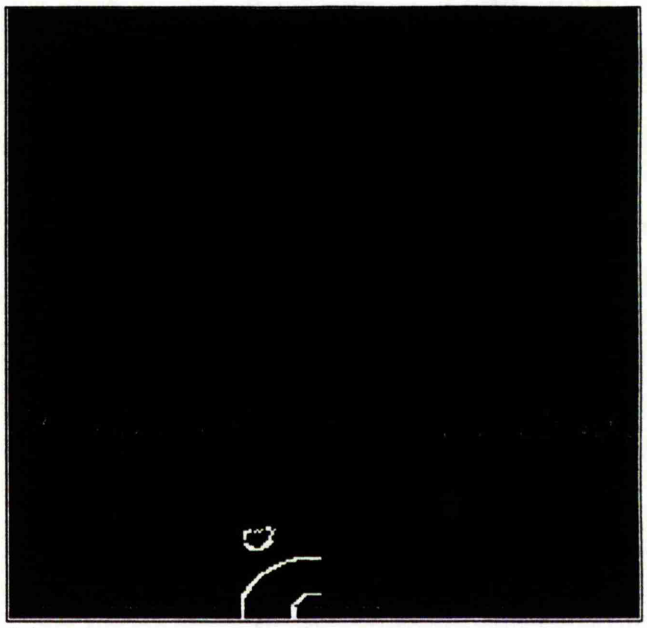
5.4 The Synthesis

The final part of this research is to combine all the techniques described in chapters 3 and 4, plus the earlier part of chapter 5, to produce images fully segmented and classified into one of the 5 texture types, with vascular channels and bone cells identified as such.

The original images saved on disk were 512×512 pixel images, but to save time and storage space these have been divided into 4 256×256 windows. For each window the first step is to segment it into formative and resorptive texture types using the texture segmentation and automatic thresholding techniques described in chapter 4. One image and its resulting binary image are shown in figure 5.13.

Fig. 5.13
An image consisting
of a type 5 texture
on the left and a
type 3 texture on the right
(a) with its corresponding
segmented image (b).

(a)



(b)

As can be seen from figure 5.13(b) the boundary between the formative and resorptive texture types has not been identified and all that has been highlighted are the two small vascular channels at the bottom of the picture. The reason for this is the method of choosing the threshold. The technique is set up to select one threshold and therefore selects the one where there is the greatest difference in edge strengths, and in this case this is around the vascular channels. This will always be the case when an image contains combinations of vascular channels, bone cells and boundaries between formative and resorptive types. A similar problem will also occur in images with no boundaries, i.e. images with either all formative or all resorptive texture types and with no vascular channels or bone cells. Here again the method will select one threshold and will lead to points which do not belong to an edge being highlighted. There is no simple way of getting round this problem other than by use of intervention by an operator.

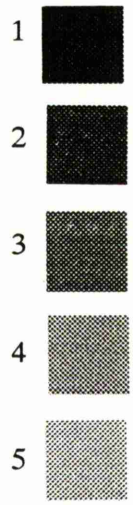
The method used here was to display an image (either 512×512 or 256×256) on the screen and to divide it into 128×128 regions. If a 128×128 region contains neither a boundary between formative and resorptive types nor vascular channels nor bone cells

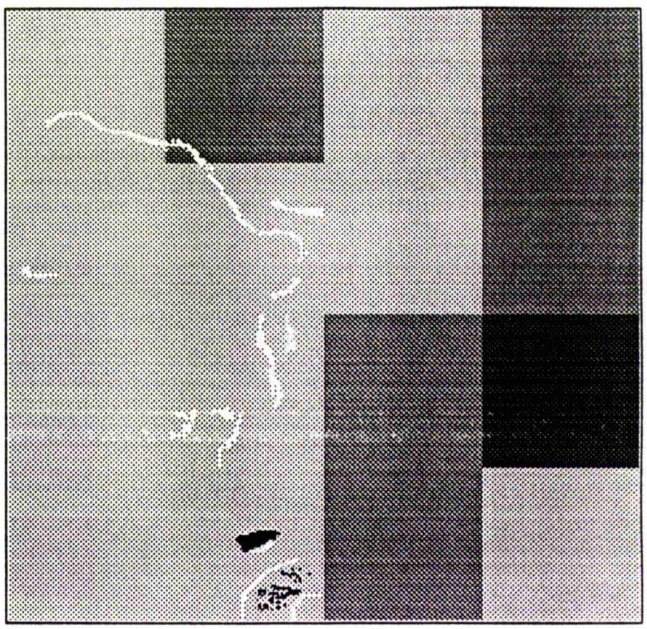
it is not segmented and is set aside for later classification. If a window contains more than one of the above it is further divided into 64×64 regions and the segmentation and automatic thresholding techniques are applied to each of the 64×64 windows. If a 128×128 region contains either a vascular channel or a bone cell or a boundary between formative and resorptive types the segmentation and thresholding techniques are applied. Once the segmentation has been applied to all regions of size 128×128 or 64×64 containing boundaries, all the windows are combined to form either 512×512 or 256×256 regions corresponding to the original image. The result of applying this technique to figure 5.13(a) is shown in figure 5.14 below.

Fig. 5.14
The result of segmenting fig. 5.13(a) using different sizes of windows.

The next stage is classification into one of the 5 texture types. To do this the image is divided into 64×64 windows and classified into one of the 5 texture types using the maximum likelihood method, based on the whole co-occurrence matrix, described in chapters 2 and 3. Vascular channels and bone cells can also be identified at this stage by thresholding the grey levels. The final, segmented and classified image is shown in figure 5.15.

Fig. 5.15
The result of segmenting and classifying Fig. 5.13(a).





The edge between the formative and resorptive types has been identified, although it tails off towards the bottom of the image. Looking at figure 5.13(a) this is not surprising - it is difficult to tell from the image where the boundary is at this point. The boundaries around the vascular channels have also been identified and, from 5.15, it can be seen that thresholding has, to a certain extent, identified them as such. As for the classification, the left-hand side of the image has been correctly classified as type 5 and the top right-hand side correctly as type 3. The two windows in the top right which have been classified as type 5 do look rather like resorptive types - an examination of the images shows what looks like a bone cell in each of the two 64×64 windows. Thus, although the classification techniques have not classified every window correctly, the reason for many of the misclassifications can be seen from the original images and in the case of windows that have been misclassified, the windows do appear very like the texture class to which they have been allocated.

The same techniques were applied to another image which is shown in figure 5.16. The resulting segmented and segmented and classified images are shown in figures 5.17(a) and 5.17(b) respectively.

Fig. 5.16
An image containing
a type 5 texture on
the left and a type
3 texture on the right,
with a vascular channel
in the centre.

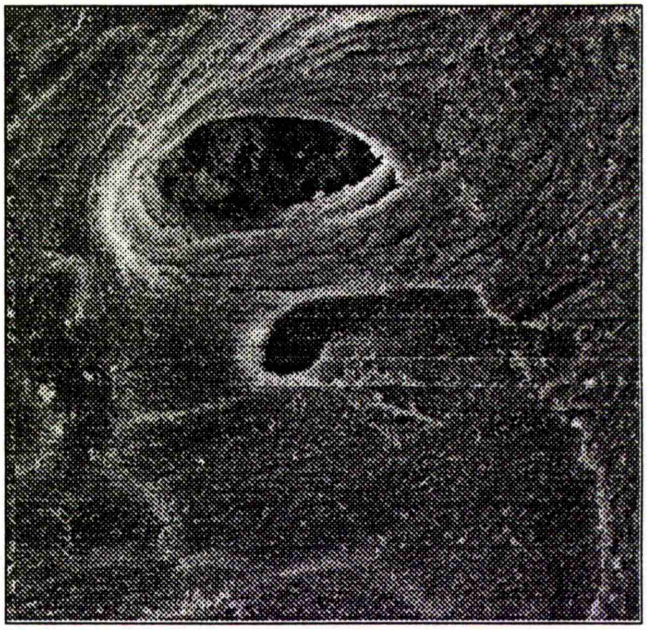
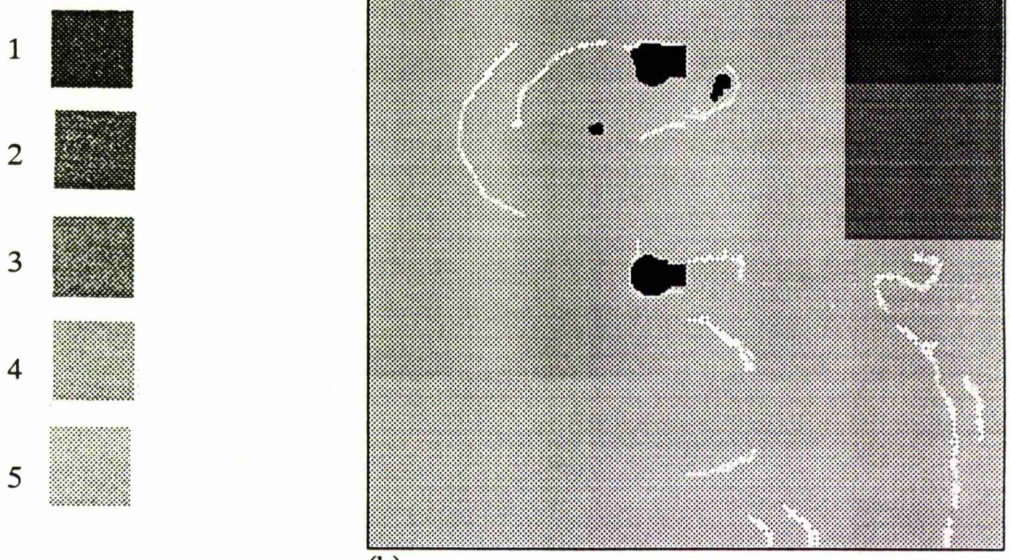


Fig. 5.17
Result of
segmenting (a)
and segmenting
and classifying
(b) figure 5.16.

(a)



(b)

Looking at 5.17(a), the edge round the vascular channel has been identified, although it is slightly broken and the bright outer rim on the left-hand side has been highlighted, rather than the inner rim. Part of the edge around the smaller vascular channel has been identified but once again it is slightly broken. The edge between the formative and resorptive textures in the bottom right of the image has been identified but various other artefacts in the image have also been highlighted. Turning to the classification, most of the image has been classified as type 5, the only exception being two windows at the top right which have been classed as types 2 and 3. On the whole, this is correct. Only smaller windows would allow for a more accurate classification.

The previous two images were examples taken from the original data set. The next step was to see how the methods would perform on images in the new data set. An example of a 256×256 image containing formative and resorptive types is shown in figure 5.18. The results of segmenting and segmenting and classifying it are illustrated in figures 5.19(a) and (b).

Fig. 5.18
An image containing texture type 4 on the left and type 2 on the right.

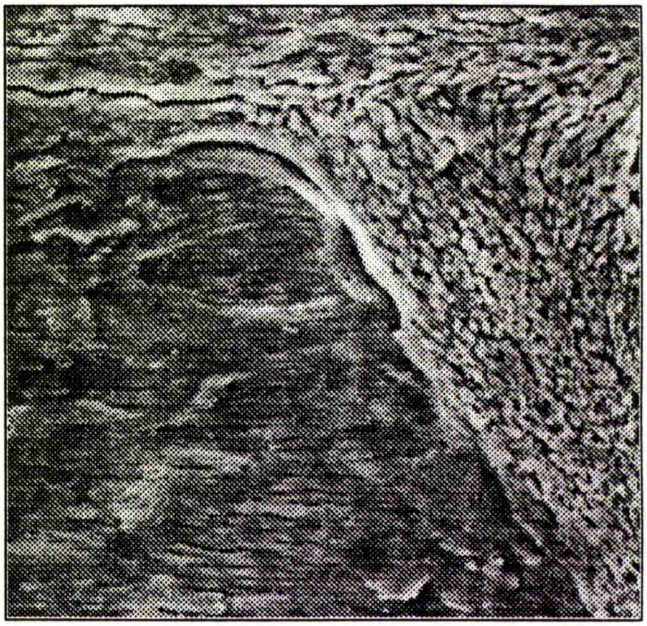
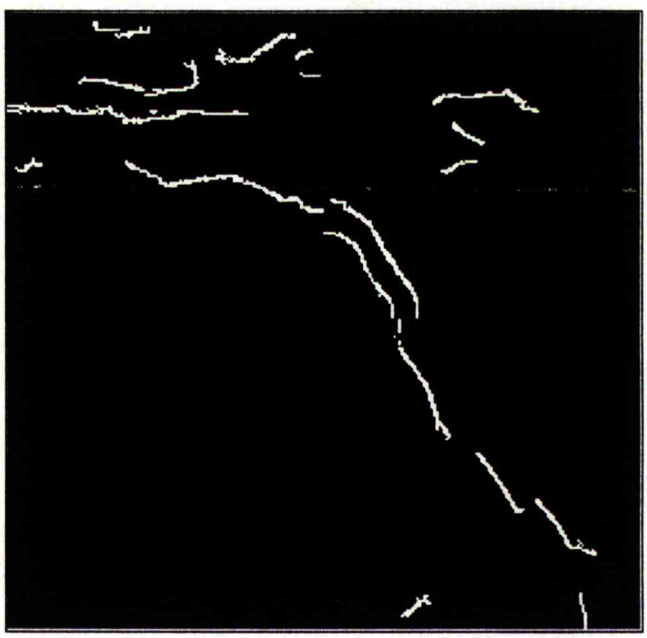
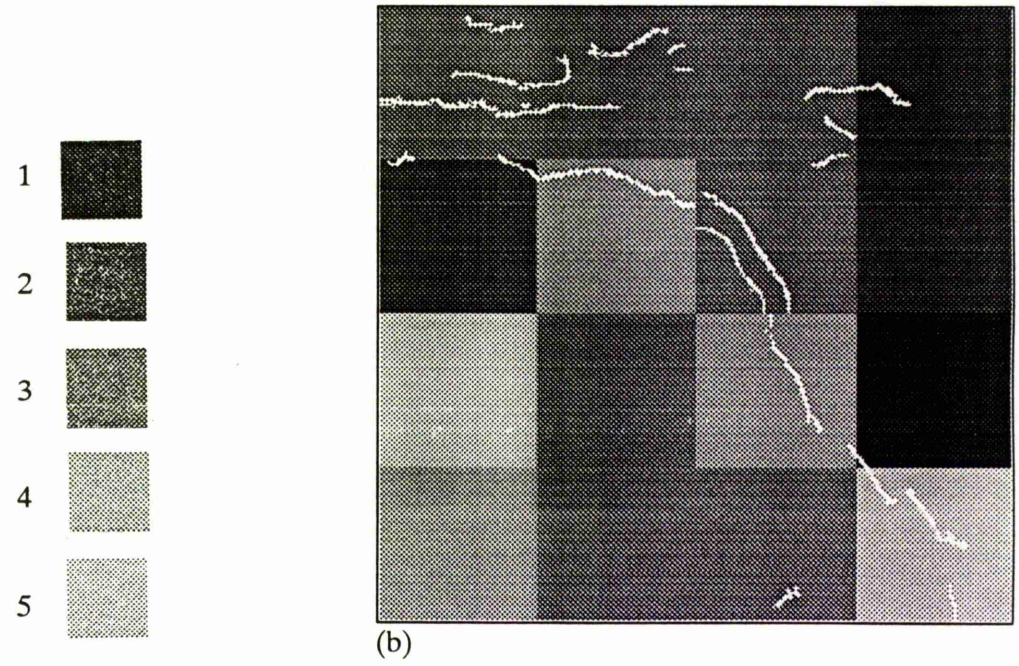


Fig. 5.19
The result of segmenting (a) and segmenting and classifying (b) figure 5.18



(a)

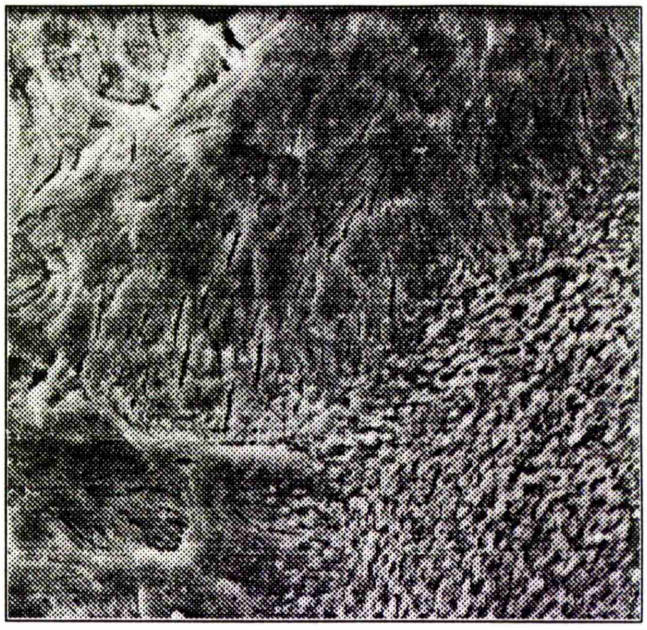


(b)

Again the segmentation and automatic thresholding have worked quite well and a fairly continuous edge has been identified. The classification has identified formative types at the top and to the right of the image and, to a certain extent, resorptive types in the lower left-hand side of the image. As happened in the examples in chapter 3, however, some of the type 5 windows have been misclassified as type 3's. Again, by examining the original image in figure 5.18 the misclassified windows do look very like type 3 textures. Apart from this problem the classification results are good.

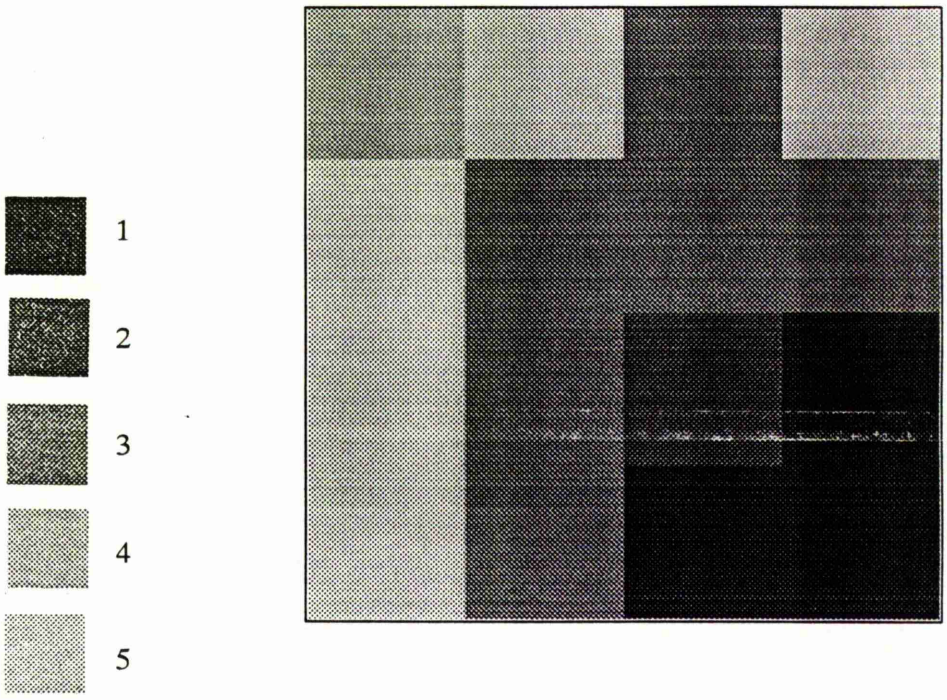
Finally, figure 5.20 shows an image from the new data set which doesn't contain a clear border between formative and resorptive types.

Fig. 5.20
An image containing a type 4/5 texture on the left and type 1/2 on the right.



This is not an easy image to segment or classify - there is no clear border between texture types and the texture on the left could probably be considered as either type 4 or type 5. Similarly, the texture type to the right of the image would probably best be described as a late type 1 or an early type 2. Due to the lack of border, the segmentation techniques were unable to locate an edge between the texture types, and so instead, the image was 'segmented' by dividing into 64×64 windows and classifying each. The results are shown in figure 5.21 below.

Fig. 5.21
The result of classifying fig. 5.20.



The resorptive types at the top and to the left have been classified correctly apart from one type 5 window at the top which has been misclassified as type 3. In addition, the windows in the lower right-hand quadrant have been correctly identified as types 1 and 2. The main problem is in the centre of the image where windows have been classified as type 3. This seems to be due to the same old problem - the similarity between type 3 and type 5 textures. Nevertheless, the boundary round the formative types 1 and 2 can be identified from this segmentation.

5.5 Conclusions

The first impression from this chapter is that the methods do not work as well on larger images as they did when applied to the 64×64 images in chapters 3 and 4. On closer inspection, however, this is not really true - the techniques have performed well considering the images to which they were applied. The images illustrated in this

chapter are not easy to either segment or classify. None of them contains a completely clear continuous boundary between texture types, some of the images contain a considerable amount of noise and the texture types are not always clear cut, making the classification difficult.

As far as the segmentation is concerned, the boundaries between texture types have been identified but often other 'edges' in the images which look no different from the texture edges, but in fact may be cracks in the bone or places where osteoclastic resorption has taken place, are identified as well. The segmentation can identify edges round the vascular channels and thresholding can, to a certain extent, identify them as such.

On the whole the classification is good, the only problem is really distinguishing between type 3 and type 5 textures. The solution to this may be to examine larger windows where osteoclastic resorption in type 5 textures would be more clearly visible. This contradicts the other criticism of the classification in that it would be better if a finer classification could be obtained. It is unlikely that the same methods applied to smaller windows would lead to very high rates of classification. What might work, however, is to use 64×64 moving windows to classify the images.

Overall, considering the images to which the techniques have been applied, the results are good. Boundaries around vascular channels and bone cells and between formative and resorptive textures can be identified, and regions can be classified into one of the 5 texture types with reasonable accuracy. There is still room for improvement, however, and suggestions on how this may be achieved are discussed in chapter 6.

N.B. The combination of segmentation and classification has so far only been tested on a small number of images. The ones shown in the thesis were chosen to illustrate results obtained from a variety of texture types and for images with and without vascular channels and bone cells. More images should be tested and some way of assessing their performance devised. One way of assessing the classification results would be by computing the percentage area of an image that is correctly classified. As far as the segmentation is concerned, the continuity of the edges could be measured by counting the number of gaps and the length of each one. Similarly, false edges could be quantified by their number and length. The position of the boundaries could be assessed by measuring the area between the fitted boundary and the 'true' one. The main difficulty with this is defining the 'true' boundary and the 'correct' classification. Physiologists will not always agree on these matters but the decision of a panel of expert physiologists would at least provide some means of judging performance.

Chapter 6 Discussion

6.1 Success

This has been a difficult problem. It is one of only a few studies where techniques for texture have been applied to natural images. The majority of researchers who have studied texture analysis have tested their methods on images taken from Brodatz' (1966) texture album, e.g. Vickers & Modestino (1982), Davis et al. (1979), Mitchell et al. (1977), Harwood et al. (1985). The only other types of images that have been analysed to any great extent are aerial photographic images (Haralick et al. (1973), Weszka et al. (1976), Conners et al. (1984), and satellite images (Haralick et al. (1973), Weszka et al. (1976)). Compared with the images used in this study, Brodatz' (1966) images should be relatively easy to segment and classify. Only one image of each type is available, and training and test sets are constructed by dividing up this one image into more manageable regions. On the other hand, no two rats' bones are identical, and therefore the five texture types identified in this study vary, albeit only slightly, from rat to rat, from bone to bone and ultimately from image to image. The appearance of the texture is also highly dependent on the conditions under which the image is captured. This is true of all images, but probably more so where specimens are first viewed under microscope. The magnification used can be stipulated and kept constant, but other factors such as the lighting conditions, and even the operator's eyesight all play a part in obtaining samples - good or not so good - of each texture type. In this study, similar microscopy techniques were used to obtain all samples. The accelerating voltage was kept at 15kW, the brightness was kept as constant as possible, and the same operator (myself) was used to capture all images in the initial data set. The resulting images, however, still vary. Some appear clear, whereas others appear blurred, and there is a considerable difference in brightness, with some images appearing very dark and others very bright. One of the main problems has therefore been to find a way to standardise the images, so that the methods developed can be applied to scanning electron microscopy images of rats' tibiae in the future, regardless of the microscopist.

The objective of this study was to develop techniques to segment, automatically, scanning electron microscope images of rats' bones into different texture types, and then

to classify each type. Neither perfect segmentation nor classification has been achieved, but this is not surprising. Very few authors have managed such a high success rate using Brodatz' (1966) images, and those that did classify all images in their data set correctly, tended to test their methods on the training sets, (Vickers & Modestino (1982), Berry & Goutsias (1991)). It is unlikely that every image in the future would be classified correctly, and anyway there is no way of checking this. Although there is room for improvement, this study has achieved what it set out to achieve. Images can be segmented fairly accurately into formative and resorptive bone types, vascular channels and bone cells can be identified, and areas can be classified with some success into one of five texture types. At this stage, the method is not fully automatic, and an operator is still required to eliminate spurious pixels, but there is now no longer the need to identify boundaries by hand and subjectively classify textures into the appropriate type, as was the case in Dempster's study in 1979.

6.2 Limitations

The ultimate objective of this research is to examine human bone, to be able to segment and classify it and thereby contribute to the understanding of human bone disease. Although vitamin D deficiency diseases such as rickets are less prevalent than they once were, other bone diseases such as osteoporosis are becoming more prevalent (Lindsay (1984)). In osteoporosis there is an imbalance between the rate of resorption and the rate of formation, in favour of the latter and therefore images are likely to contain a larger proportion of texture types 4 and 5 than normal bone. The images used in this study were of rats' tibiae but it is hoped that the techniques used here could be applied, with similar success, to images of human bones captured in the future. There are, however, limitations which must be considered. Before the results can be applied to human bone, the first step is to ensure that the techniques used will achieve reasonable success when applied to rats' tibiae captured in the future. Probably the main consideration has to be the image capture. Not all images used in this study were captured by the same person - images making up the first data set were captured and saved by a different operator from the second data set, although the same microscope

(a JEOL T100 scanning electron microscope) was used and the same basic principles for use of the scanning electron microscope were applied. Even so, differences in lighting and focusing can still be seen in the images within, as well as between, data sets. This can have quite a considerable effect on the success achieved when classifying textures, where for example, as was illustrated in chapter 3, the texture features suggested by Haralick et al. (1973) can vary significantly from image to image depending on image capture. Most authors have dealt with this problem by first histogram equalising the images, i.e. by transforming the grey-level histogram to be uniformly distributed. Thus each image will have equal numbers of each grey level. Visually, after transformation, there appears to be some loss of information, but in fact, experiments have shown that there is little loss of textural information, and methods using second-order statistics seem to work as well on transformed images as on images before transformation. It is, however, no longer possible to use first-order statistics such as mean and standard deviation to classify images.

The images used in this study were saved onto disk by two different methods, and to take account of this, at different magnifications. Images in the initial batch were magnified 350 times, whereas later images were magnified 750 times. The original images were collected using a Joyce Loebl Magiscan computer, and because the Magiscan only uses the central part of each scan, the final digital image magnification is in fact twice the original, i.e. 700. Later images were saved using a digital scanner, and thus the final magnification remained at 750. Neither the change in technique used, nor the difference in magnification, appeared to affect the results in any way, and methods which worked on the original data set also worked on the new data set. Saving images at considerably different magnifications should not affect either the segmentation or classification to any great extent, as long as all images in both the training and test sets are saved at the same magnification. Regardless of magnification, a type 2 texture will always appear coarser, with larger mineral nodules than a type 1 texture, magnified the same number of times, for example. The thresholds may, however, have to be modified slightly. How successful the various algorithms are when tested on images collected by an independent individual, using a different scanning electron microscope, has yet to be investigated, but because the quality of some of the images used in this study is quite poor, and because two operators have been used, it is possible that the

techniques are robust enough to make any differences fairly small.

6.3. Further work

Since the beginning of this research in 1987 there has been a considerable amount of work done in all areas of texture analysis. To test every technique ever published could take a life-time, particularly as the area is still very active and numerous papers are published each year. Although, it is not an exhaustive literature review, several methods for texture classification have been mentioned and commented on in chapter 2. The methods selected for the chapter essentially fall into two, though not mutually exclusive, categories. The first category consists of 'standard' techniques, i.e. those that have been used consistently over the years and have performed well on different data sets. For example, the texture features of Haralick et al. (1973) have been successful in classifying various textures over the past two decades. The 14 features originally suggested by Haralick et al. in 1973 have since been added to, and modifications of their technique have been used. The maximum likelihood method of Vickers and Modestino (1982) uses the whole co-occurrence matrix to classify textures, rather than extracting features, and again, because of its success, this method could be considered a 'standard' technique, worthy of testing before more complicated algorithms are developed. Other statistical techniques such as Fourier transform features and run-length matrices, which have been fairly extensively tested but have received less favourable reviews, are included for completeness. More recently, modelling techniques, particularly those based on Markov and Gibbs random fields, have become an almost standard approach to texture analysis, although a limited amount of work has yet been done on the problem of texture classification.

The second category of techniques in chapter 2 are those which have been less extensively investigated but look promising, in my opinion, for classifying the rats' tibiae in this study. One example of a technique in this category is the relative extrema measures of Mitchell et al. (1977). In the initial stages of this research, line scans of images were taken and analysed subjectively. Although this work was taken no further, results looked promising and could perhaps be extended.

Less work has been done in the area of texture segmentation, although there

have been several publications in the past few years. Until recently, the only well-publicised method was that of Rosenfeld & Thurston (1971). In the past ten years, one active area of research in texture segmentation has been the use of Gibbs and Markov random field models (Derin & Cole (1986), Derin & Elliott (1987), Hansen & Elliott (1982), Geman & Graffigne (1986)). Other techniques involved measuring image properties over windows of varying sizes. These include Chen and Pavlidis' (1979) split-and-merge algorithm, the pyramid approach of Burt et al. (1981) and the quad-tree smoothing approach of Spann and Wilson (1985).

Although it is perhaps unlikely that results in this study could be improved to any great extent using the techniques mentioned above, an investigation into at least some of them, particularly quad-tree smoothing and relative extrema methods, would provide an interesting extension to this research project.

6.3.1 The Quad-Tree Approach

The quad-tree approach is a special case of a split-and-merge algorithm. If an image is inhomogeneous with regard to some texture feature it is divided into quadrants and the process is repeated until the image consists of homogeneous regions. At any stage, homogeneous regions may also be combined. Chen & Pavlidis (1979) used co-occurremce matrices as their features, whereas Spann & Wilson (1985) used a combination of statistical and spatial information to segment synthetic texture images, and in 1988 Wilson & Spann used a similar technique based on finite prolate spheroidal sequences to segment synthetic images and Brodatz' (1966) images. In this study images were divided into quadrants until homogeneous regions were obtained but then the texture segmentation algorithm of Rosenfeld & Thurston (1971) was applied. Further investigation could involve some of the texture features mentioned above.

6.3.2 Relative Extrema

As mentioned above, an initial investigation of the use of scan-lines for classifying the images looked promising, and could be further investigated. For example, a fine type 1 texture would be expected to have numerous peaks, whereas the coarser type 2 would

have less, and the smooth type 5 less still. These properties could possibly also be used to find the boundaries between two textures, e.g. the point where an oscillating line becomes smooth may indicate a boundary between a type 2 and a type 5 texture.

6.3.3 Markov & Gibbs Random Fields

Markov and Gibbs random field models could be tested on the rats' bone data. Little work has been done on texture classification using Markov random field models but Chen and Huang (1993) obtained better results with Markov random field parameters than co-occurrence matrix features when classifying Brodatz' (1966) texture images. On the other hand, Ohanian and Dubes (1992) obtained better results with co-occurrence matrices than Markov random fields, and advised that Markov random fields should only be used on large images. It is perhaps unlikely, therefore, that classification of rats' tibiae could be improved by Markov random fields, but the models are worthy of some investigation.

Segmentation of images using Markov random fields is a much more difficult problem, and one that has only really been addressed using synthetically generated images. An image labelling problem is specified in terms of a set of objects and a set of object labels $L=\{l_1,l_2,...l_G\}$ where the labels denote the pattern classes in the image. For example, if an image contains an object in a background, G=2 labels could be used, one for the object and one for the background. The true pixel labelling is denoted $x^* = \{x_1^*,x^*_2, ...x^*_M\}$ where M is the number of pixels. The object in segmentation problems is to estimate x^*. The set of observable random variables is denoted by $Y=\{Y_1,Y_2,...Y_M\}$ where Y_t is a feature vector associated with the t^{th} pixel, and the true labelling x^* is viewed as a realisation of a Markov random field imposed on $X=\{X_1,X_2,...X_M\}$. Then, given a set of observed feature vectors, Y=y and the contextual information P(X=x), the problem is to find the 'optimal' estimate of the true labelling X^*. The Maximum-A-Posteriori (MAP) method chooses the estimate $\hat{x}$ that maximises the posterior probability of $X=\hat{x}$, given Y=y i.e. $P(X=x|Y=y)$. Using Bayes' Theorem, this can be written

$$P(X=x|Y=y) = \frac{P(Y=y|X=x)P(X=x)}{P(Y=y)}.$$

The computational problems of finding $\hat{x}$ are enormous, and various algorithms, such as simulated annealing and iterated conditional modes (ICM) have been suggested. Knowing the number of pattern classes and some of the parameters simplifies the problem of estimating x^*, but the problem is never easy. Dubes & Jain, in their paper in 1989 state "Good prior information about the true image is a tremendous advantage in segmentaion, as in all Bayesian procedures. Spectacular results can be achieved if enough is known about the image model and the observation process." Thus although authors such as Derin & Cole (1986) and Derin & Elliott (1987) have obtained good segmentation results on synthetically generated images, it is less likely that such success would be achieved on natural images where there is very little prior information. Graffigne (1987) on the other hand, has obtained some promising results using Markov random fields to segment images containing combinations of Brodatz' (1966) texture images. Whether the same techniques would work on bone is a matter for further investigation.

Several other techniques for texture, not already mentioned in this thesis, have been investigated by several authors in recent years. Some of the better-publicised techniques are described below.

6.3.4 Fractals

Fractal geometry was initially explored and developed by Mandelbrot (1982) and has had a major impact in modelling and analysis in the natural and physical sciences. The defining characteristic of a fractal is that it has a 'fractional' dimension. Technically, a fractal is a set whose Hausdorff-Besicovich dimension is strictly larger than the topological dimension. The fractal dimension of a surface corresponds quite

closely to our notion of roughness. Fractals are independent of scale and provide a mathematical framework to study the irregular shapes found in nature, such as coastlines.

Consider a bounded set A in Euclidean n-space. The set A is said to be self-similar when A is the union of N distinct (nonoverlapping) copies of itself, each of which has been scaled down by a ratio r in all coordinates. The fractal or similiarity dimension of A is given by the relation

$$1 = Nr^{D} \quad or \quad D = \frac{\ln N}{\ln \frac{1}{r}} \qquad (1)$$

Natural fractal surfaces do not, in general, possess this deterministic self-similarity. Instead, they exhibit statistical self-similarity; that is, they are composed of N distinct subsets each of which is scaled down by a ratio r from the original and is identical in all statistical respects to the scaled original. The fractal dimension for these surfaces is also given by (1).

There has been increasing interest in the use of fractal models in image analysis, led primarily by Pentland (1984). He yielded a correct classification rate of 84.4% on 8 Brodatz' (1966) textures by calculating the fractal dimensions of 16×16 images. Keller et al. (1989) however, demonstrated that fractal dimension alone does not provide sufficient discriminatory power to classify natural textures. They introduced new features based on Mandelbrot's concept of lacunarity (that characteristic of fractals of the same dimension with different texture appearances) which capture the second-order statistics of fractal surfaces. Using both their new estimate for computing the fractal dimension and four lacunarity features they obtained good results when segmenting images containing combinations of Brodatz' (1966) textures. The results obtained by Pentland (1984) were good considering the images were only of size 16×16. Once again, however, all the analysis mentioned above has been applied to Brodatz' (1966) images. Fractal models and lacunarity features alone may not improve either the segmentation or classification of bone, but it would be interesting to investigate, for example, the use of a combination of lacunarity features and co-occurrence features.

6.3.5 Spatial Filtering

Spatial filtering techniques have achieved good classification results in recent years. The potential power of Fourier analysis for studying human vision was discovered in the 1960s but it wasn't until the 1980s that studies suggested that filtering in spatial frequency channels might be useful for machine pattern recognition, including texture analysis. The method is based on a model of human vision which postulates the existence of channels which decompose an image into several bands of spatial frequency and orientation. The output of the filtering procedure is a sequence of filtered images, each of which contains limited spectral information from the original image. Simple texture features such as a measure of the spread of the grey level frequency histogram of the filtered images, can then be extracted from these filtered images. Coggins & Jain (1985) used Gaussian shaped filters, adapted by Ginsburg (1977) in 4 directions, and extracted a single feature to classify eight texture classes from Brodatz' (1966) texture album. They obtained 98% correct classification on 64×64 images and 91% on 32×32 images. Thus, once again, good results have been obtained on fairly small images, although the images were again Brodatz' (1966) images. Classification of the bone data is unlikely to be as high as 91% but the method is still worth investigating.

6.3.6 Neural Networks

Artificial neural networks are currently enjoying a rapid expansion into all areas of data and image processing and their use for texture classification is an area of research which is fairly active at the time of writing. Artificial neural network models are composed of many nonlinear computational elements operating in parallel and arranged in patterns reminiscent of biological neural networks. Computational elements or nodes are connected via weights that are typically adapted during use to improve performance. Neural networks can be broken down in terms of how they are encoded (how they store knowledge) and how they are decoded (how, once knowledgeable, they process new input data). In its encoding property, a network can be either supervised or unsupervised; in its decoding, it can be either a feedforward or feedback type.

Applications on texture so far tend to have been restricted to Brodatz' (1966)

texture images and satellite images but the methods could easily be applied to bone. Muhamad & Deravi (1992) used a feedforward neural network structure with a single hidden layer to classify 9 texture classes from Brodatz' (1966) texture album. They obtained similar results to those of Vickers & Modestino (1982) using maximum likelihood, although they only used 4 texture features from the co-occurrence matrix whereas the maximum likelihood method uses the whole co-occurrence matrix. Greenspan et al. (1991) suggested a combined neural network and rule-based approach to pattern recognition. They used Gabor filters for the feature-extraction phase and demonstrated a correct classification rate of 95-99% for synthetic textures and 90% for natural texture mosaics. In 1992 they applied their method to 5 texture classes from Brodatz' (1966) texture album and achieved a correct classification of around 90%. In 1993 Greenspan & Goodman applied the techniques to segment Landsat and aerial images and found that their texture classification neural-network achieved 'a correct, yet rough' segmentation based on textural characteristics alone. More recently, Walder, MacLaren & Reid (1994) compared a back-propagation neural network with spectral and textural features with a maximum likelihood classifier to classify satellite images of clouds and obtained 95% correct classification compared with 90% using maximum likelihood. Features from co-occurrence matrices were extracted from the bone data and run through a feed-forward neural net but the results were poor with only 44% of the images in the test set being correctly classified. Different architectures or different learning rules may, however, provide better results.

6.3.7 Continuous Texture Types

In this study five texture types were assumed. Dempster (1979) used six types, the sixth being a mineralisation front over a resorbed area, which is difficult to distinguish from the formative types. It would be possible, however, to extend the research to include this sixth type. It also may be more sensible to treat the texture types as a continuous scale, rather than 6 distinct types. The three formative types merge into one another and physiologists may disagree on whether a bone is, for example, a late type one or an early type two. The same may apply to types two and three, and types four and five. Either treating the types as continuous, or alternatively

increasing the number of texture types, where for example a type 2.5 could be thought of as either a late type two or an early type three, would resolve this problem to some extent.

6.3.8 Back-scattered Electron Images

The images analysed in this study were obtained by viewing specimens using a scanning electron microscope. Scanning electron microscopes are largely used to study surfaces of specimens. They have various advantages over other microscopy methods, including good spatial resolution and a large depth of field. In almost all types of electron microscopy, primary electrons enter the specimen and the same or different electrons leave it again to form the image. Various interactions between the high energy electrons and the atoms of the specimen are possible. The term secondary electrons is used to describe those electrons which escape from the specimen with energies below 50eV. The yield of secondary electrons, i.e. the number emitted per primary electron, can be as high as 1. Secondary electrons are therefore abundant and are the most commonly used imaging signal in scanning electron microscopy. In thick specimens some of the electrons are 'back-scattered' out of the specimen. Back-scattered electrons are not usually as numerous as secondary electrons but most of them have high energies. Scanning electron microscopes normally have facilities for detecting both secondary electrons and back-scattered electrons. Both types of electrons produce images similar to images of solid objects viewed by eye, though back-scattered electron images tend to have more shadows and highlights than secondary electron images. In some circumstances it may be easier to interpret back-scattered images and to distinguish the peaks from the troughs in the specimen, than for secondary electron images. Thus it may, in fact, be easier to segment and classify back-scattered electron images. Further work would be needed to see whether the techniques applied on the secondary electron images used here would improve the results, or whether completely new techniques would have to be developed.

6.4 Conclusions

Several techniques for texture analysis have been reviewed and tested on scanning electron microscope images of rats' tibiae. Reasonably good automatic segmentation and classification have been achieved. Previous studies required an operator to draw boundaries between textures manually, and a physiologist was needed to identify each texture type. Now textures can be identified by computer, with a minimal input required from an operator. Various techniques have been suggested which may improve the method further. Although restricted to rats' tibiae in this study, the techniques could possibly be applied to other natural textures such as human bone, and thereby in some small way aid researchers in the search for cures for bone diseases.

Appendix

Textural Features from Co-occurrence Matrices

1) Angular Second Moment

$$f_1 = \sum_{i=0}^{LVL-1} \sum_{j=0}^{LVL-1} [P(i, j)]^2$$

2) Contrast

$$f_2 = \sum_{k=0}^{LVL-1} k^2 \left[\sum_{|i-j|=k} \sum P(i, j)\right]$$

3) Correlation

$$f_3 = \frac{\sum_{i=0}^{LVL-1} \sum_{j=0}^{LVL-1} [i\ j\ P(i, j)] - \mu_x \mu_y}{\sigma_x \sigma_y}$$

where μ_x,μ_y,σ_x and σ_y are the means and standard deviations of P_x and P_y.

4) Sum of Squares: Variance

$$f_4 = \sum_{i=0}^{LVL-1} \sum_{j=0}^{LVL-1} (i-\mu)^2\ P(i, j)$$

5) Inverse Difference Moment

$$f_5 = \sum_{i=0}^{LVL-1} \sum_{j=0}^{LVL-1} \frac{1}{1+(i-j)^2} P(i, j)$$

6) Sum Average

$$f_6 = \sum_{i=2}^{2LVL} i \, P_{x+y}(i)$$

7) Sum Variance

$$f_7 = \sum_{i=2}^{2LVL} (i-f_6)^2 \, P_{x+y}(i)$$

8) Sum Entropy

$$f_8 = -\sum_{i=2}^{2LVL} P_{x+y}(i) \ln [P_{x+y}(i)]$$

9) Entropy

$$f_9 = -\sum_{i=0}^{LVL-1} \sum_{j=0}^{LVL-1} P(i, j) \ln P(i, j)$$

10) Difference Variance

$$f_{10} = \sum_{i=0}^{LVL-1} [i - \sum_{i=0}^{LVL-1} i\ P_{x-y}\ (i)]^2\ P_{x-y}\ (i)$$

11) Difference Entropy

$$f_{11} = - \sum_{i=0}^{LVL-1} P_{x-y}\ (i)\ \ln P_{x-y}\ (i)$$

12),13) Information Measures of Correlation

$$f_{12} = \frac{HXY - HXY1}{\max\ [HX, HY]}$$

$$f_{13} = (1 - \exp\ [-2.0\ (HXY2 - HXY)])^{\frac{1}{2}}$$

where HX and HY are entropies of P_x and P_y, and

$$HXY = - \sum_{i=0}^{LVL-1} \sum_{j=0}^{LVL-1} P(i, j)\ \ln P(i, j)$$

$$HXY1 = -\sum_{i=0}^{LVL-1}\sum_{j=0}^{LVL-1} P(i, j) \ln [P_x (i) P_y (j)]$$

$$HXY2 = -\sum_{i=0}^{LVL-1}\sum_{j=0}^{LVL-1} P_x (i) P_y (j) \ln [P_x (i) P_y (j)]$$

14) Maximal Correlation Coefficient

f_{14} = (Second largest eigenvalue of Q)$^{1/2}$

where

$$Q(i, j) = \sum_k \frac{P(i, k) P(j, k)}{P_x (i) P_y (k)}$$

Notation:

P(i,j) = (i,j)th entry in the normalised co-occurrence matrix.

$$P_x (i) = \sum_{j=0}^{LVL-1} P(i, j)$$

$$P_{x+y}(k) = \sum_{\substack{i=0 \\ i+j=k}}^{LVL-1} \sum_{j=0}^{LVL-1} P(i, j)$$

k=2,3,...,2LVL.

$$P_{x-y}(k) = \sum_{\substack{i=0 \\ |i-j|=k}}^{LVL-1} \sum_{j=0}^{LVL-1} P(i, j)$$

k=0,1,...,LVL-1.

References

Ahuja, N. & Rosenfeld, A. (1981a). Mosaic Models for Textures. IEEE Trans. Pattern Anal. Mach. Intell., PAMI-3, No.1, 1-11.

Ahuja, N. & Rosenfeld, A. (1981b). Fitting Mosaic Models to Textures. Image Texture Analysis, R.M. Haralick, Ed., New York, Plenum.

Bajcsy, R. (1973). Computer Description of Textured Surfaces. Proc. Int. Joint Conf. Artificial Intelligence, 572-579.

Ballard, D. & Brown, C.M. (1982). Computer Vision. Prentice-Hall Inc., New Jersey.

Berry, J.R. & Goutsias, J. (1991). A Comparative Study of Matrix Measures for Maximum Likelihood Texture Classification. IEEE Trans. Systems, Man & Cybernetics, SMC-21, No.1, 252-261.

Besag, J.E. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems (with discussion). J.R. Statistical Soc., B, 36, 192-236.

Besag, J.E. (1975). Statistical Analysis of Non-Lattice Data. The Statistician, 24, 179-195.

Besag, J.E. (1986). On the Statistical Analysis of Dirty Pictures. J.R. Statistical Soc., B, 48, No.3, 259-302.

Biddlecombe, W.H., Dempster, D.W., Elder, H.Y. (1977). Morphometric analysis by computerised planimetry. J. Physiol., 273, 22-23.

Boyde, A. (1968). Observations on the pattern of mineralisation of various mammalian dentin. J. Dent. Res., 47, 1007, (Abstract no. 123).

Boyde, A., Jones, S.J., & Hobdell, M.H. (1968). The mineralising front of whale dentin. J. Dent. Res., 47, 1007, (Abstract no. 122).

Boyde, A. & Hobdell, M.H. (1969a). Scanning Electron Microscopy of Lamellar Bone. Z. Zellforsch, 93, 213-231.

Boyde, A. & Hobdell, M.H. (1969b). Scanning electron microscopy of primary membrane bone. Z. Zellforsch, 99, 98-108.

Boyde, A. (1972). Scanning electron microscope studies of bone. The Biochemistry and Physiology of Bone, Edn. 2, (Ed. G.H. Bourne) Academic Press, London, N.Y., 259-310.

Boyde, A. & Jones, S.J. (1974). Bone and other hard tissues. In Principles and Techniques of Scanning Electron Microscopy. (Ed. M.Hayat), Vol.2, Van Nostrand Reinhold Co., New York, London, 123-149.

Boyde, A., Maconnachie, E., Reid, S.A., Delling, G. & Mundy, G.R. (1986). Scanning Electron Microscopy in Bone Pathology. Review of methods potential and applications. Scanning Electron Microscopy, 4, Sem. Inc. AMF O'Hare (Chicago), 1537-1554.

Brodatz, P. (1966). Textures - A Photographic Album for Artists and Designers, Dover, New York.

Burt, P.J., Hong, T.H. & Rosenfeld, A. (1981). SEgmentation and Estimation of Image Region Properties Through Cooperative Hierarchial Computation. IEEE Trans. Systems, Man & Cybernetics, SMC-11, No.12, 802-809.

Carlucci, L. (1972). A Formal System for Texture Languages. Pattern Recognition, 4, 53-72.

Carstensen, J.M. (1993). Co-occurrence Feature Performance in Texture Classification. Proc. 8th Scandinavian Conf. on Image Analysis, 831-838.

Chellappa, R. & Chatterjee, S. (1985). Classification of Textures Using Gaussian Markov Random Fields. IEEE Trans. Acoust., Speech, Sig. Proc., ASSP-33, No.4, 959-963.

Chen, C.C. (1988). Markov Random Fields in Image Analysis. PhD. thesis, Michigan State University.

Chen, C.C. & Huang, C.L. (1993). Markov Random Fields for Texture Classification. Pattern Recognition Letters, 14, 907-914.

Chen, P.C. & Pavlidis, T. (1979). Segmentation by Texture Using a Co-occurrence Matrix and a Spli-and-Merge Algorithm. Computer Graphics & Image Processing, 10 172-182.

Chen, P.C. & Pavlidis, T. (1980). Image Segmentation as an Estimation Problem. Computer Graphics & Image Processing, 12, 153-172.

Coggins, J.M. & Jain, A.K. (1985). A Spatial Filtering Approach to Texture Analysis. Pattern Recognition Letters, Vol. 3, 195-203.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational & Psych. Meas., 20, 37-46.

Conners, R.W. & Harlow, C.A. (1980a). Toward a Structural Textural Analyser Based on Statistical Methods. Computer Graphics & Image Processing, 12, 224-256.

Conners, R.W. & Harlow, C.A. (1980b). A Theoretical comparison of texture algorithms. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-2, No.3, 204-222.

Conners, R.W., Trivedi, M.M. & Harlow, C.A. (1984). Segmentation of a High-Resolution Urban Scene Using Texture Operators. Computer Vision, Graphics & Image Processing, 25, 273-310.

Cross, G.R. & Jain, A.K. (1983). Markov Random Field Texture Models. IEEE Trans. Pattern Anal. Mach. Intell., PAMI-5, No. 1, 25-39.

Davis, L.S., Johns, S.A. & Aggarwal, J.K. (1979). Texture Analysis using generalised co-occurrence matrices. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-1, No.3, 259-279.

Davis, L.S., Clearman, M. & Aggarwal, J.K. (1981). An empirical evaluation of generalised co-occurrence matrices. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-3, No.2, 214-221.

Davis, W.L., Mathews, J.L., Martin, J.H. Kennedy, J.W. & Talmage, R.V. (1975). The endosteum as a functional membrane. Calcium Regulating Hormones, Proc. 5th Parathyroid Conference, (Ed. R.V. Talmage, M.Owen, J.Parsons), Excerpta Medica, Amsterdam, American Elsevier Publishing Co.,Inc., New York, 275-283.

Dempster, D.W., Elder, H.Y. & Smith, D.A. (1979). Scanning Electron Microscopy of Rachitic Rat Bone. Scanning Electron Microscopy, Sem. Inc. AMF O'Hare.

Dempster, D.W. (1979). Electron Microscopial and Biochemical Studies on Rachitic Rat Bone. Thesis, University of Glasgow.

Derin, H. & Cole, W.S. (1986). Segmentation of Textured Images Using Gibbs Random Fields. Computer Vision Graphics & IMage Processing, 35, 72-98.

Derin, H. & Elliott, H. (1987). Modeling and Segmentation of Noisy Textured Images Using Gibbs Random Fields. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-9, No.1, 39-55.

Dubes, R.C. & Jain, A.K. (1989). Random Field Models in Image Analysis. J. Applied Statistics, 16, No.2, 131-164.

Duda, R.O. & Hart, P.E. (1973). Pattern Classification and Scene Analysis. John Wiley & Sons, New York.

Dyer, C.R., Hong, T. & Rosenfeld, A. (1980). Texture Classification Using Grey Level Cooccurrence Based on Edge Maxima. IEEE Trans. Systems, Man & Cybernetics, SMC-10, No. 3, 158-163.

Efron, B. (1983). Estimating the Error Rate of a Prediction Rule. J. Amer. Stat. Assoc., 78, 316-331.

Ehrich, R. & Foith, J.P. (1976). Representation of Random Waveforms by Relational Trees. IEEE Trans. Computers, C-25, 725-736.

Eklundh, J.O. & Rosenfeld, A. (1979). Peak Detection Using Difference Operators. (1979). IEEE Trans. Pattern Anal. Mach. Intell., PAMI-1, 317-325.

Galloway, M.M. (1975). Texture Analysis Using Grey Level Run-Lengths. Computer Graphics & Image Processing, 4, 172-179.

Garven, H.S.D. (1957). A Student's Histology. E & S Livingstone Ltd.

Geman, S. & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. IEEE Trans. Pattern Anal. Mach. Intell., PAMI-6, 721-741.

Geman, S. & Graffigne, C. (1986). Markov Random Field Image Models and their Applications to Computer Vision. Proc. International Congress of Mathematicians, Ed. A.M. Gleason. American Mathematical Society, Providence, 1987.

Ginsburg, A.P. (1977). Visual Information Processing Based on Spatial Filters Constrained by Biological Data. PhD. Thesis, University of Cambridge.

Glasbey, C.A. (1993). An Analysis of Histogram-Based Thresholding Algorithms. Computer Vision Graphics & Image Processing, Vol. 55, No.6, 532-537.

Gonzalez, R.C. & Wintz, P. (1987). Digital Image Processing. Addison-Wesley Publishing Co.

Goodhew, P.J. & Humphreys, F.J. (1988). Electron Microscopy and Analysis. Taylor & Francis Ltd. 2nd Ed.

Graffigne, C. (1987). Experiments in Texture Analysis and Segmentation. PhD. Thesis, Brown Univesity.

Greenspan, H., Goodman, R. & Chellappa, R. (1991). Texture Analysis via Unsupervised and Supervised Learning. Proc. 1991 International Joint Conference on Neural Networks, Vol. 1, 639-644.

Greenspan, H., Goodman, R. & Chellappa, R. (1992). Combined Neural Network and Rule-Based Framework for Probabilistic Pattern Recognition and Discovery. In Advances in Neural Information Processing Systems, 4, 444-452, San Mateo, CA: Morgan Kaufmann Publishers.

Greenspan, H. & Goodman, R. (1993). Remote Sensing Image Analysis via a Texture Classification Neural Network. In Advances in Neural Information Processing Systems, 5, 425-432, Morgan Kaufmann Publishers.

Hall, E.L., Kruger, R.P., Dwyer, I., Hall, D.L., McLaren, R.W. & Lodwick, G.S. (1971). A survey of Preprocessing and Feature Extraction Techniques for Radiographic images. IEEE Trans. Computers, C-20, No.9, 1032-1044.

Ham, A.W. (1965). Histology. Pitman medical publishing Co. Ltd.

Hanaoka, H. Yabe, H. & Bun, H. (1989). The Origin of the Osteoclast. Clinical Orthopaedics and Related Research, 239, 286-298.

Hancox, N.M. (1972). Biology of Bone. Cambridge University Press.

Hand, D.J. (1981). Discrimination and Classification. John Wiley & Sons Inc. Ltd.

Hansen, F.R. & Elliott, H. (1982). Image Segmentation Using Simple Markov Field Models. Computer Graphics Image Processing, 20, 101-132.

Haralick, R.M., Shanmugam, K. & Dinstein, I. (1973). Textural features for image classification. IEEE Trans. Systems, Man & Cybernetics, SMC-3, No.6, 610-621.

Haralick, R.M. (1979). Statistical and Structural Approaches to Texture. Proc. IEEE, 67, No.5, 786-804.

Harwood, D., Subbarao, M., & Davis, L.S. (1985). Texture Classification by Local Rank Correlation. Computer Vision, Graphics & Image Processing, 32, 404-411.

Hassner, M. & Sklansky, J. (1980). The Use of Markov Random Fields as Models of Texture. Computer Graphics & Image Processing, 12, 357-370.

Hong, T.H., Dyer, C.R. & Rosenfeld, A. (1980). Texture Primitive Extraction Using an Edge-Based Approach. IEEE Trans. Systems, Man & Cybernetics, SMC-10, no.10, 659-675.

Huang, T.S., Yang, G.J. & Tang, G.Y. (1978). A Fast two-dimensional median filtering algorithm. IEEE Trans. Acoust., Speech, Sig. Proc., ASSP-27, 13-18.

Ising, E. (1925). Zeitschrift Physik, 31, 253.

Johannsen, G. & Bille, J. (1982). A Threshold Selection Method using Information Measures. Proceedings, 6th International Conference on Pattern Recognition, Munich, Germany, 140-143.

Jones, J.L., Davis, W.L., Jones, R.G., Miller, G.W. & Mathews, J.L. (1977). The Effect of Cytochalasin B on the Endosteal Lining Cells of Mammalian Bone. A Scanning Electron Microscope Study. Calc. Tiss. Res., 24, 1-10.

Jones, S.J. & Boyde, A. (1968). Scanning Electron Microscopy of Cementum. J. Dent. Res., 47, 1007, (Abstract no. 124).

Jones, S.J. & Boyde, A. (1970). Experimental Studies on the interpretation of bone surfaces studied with the SEM. Proc. Third Annual Symposium on Scanning Electron Microscopy, (Ed. O.Johari), IIT Research Institute, Chicago, 193-200.

Jones, S.J. (1974). Secretory Territories and Rate of Matrix Production of Osteoblasts. Calc. Tiss. Res., 14, 309-315.

Jones, S.J., Boyde, A. & Pawley, J.B. (1975). Osteoblasts and Collagen Orientation. Cell. Tiss. Res., 159, 73-80.

Jones, S.J. & Boyde, A. (1976a). Morphological Changes of Osteoblasts in Vitro. Cell. Tiss. Res., 166, 101-107.

Jones, S.J. & Boyde, A. (1976b). Experimental Study of Changes in osteoblastic Shape induced by Calcitonin and Parathyroid Extract in an Organ Culture System. Cell. Tiss. Res., 169, 449-465.

Jones, S.J., Boyde, A. & Ali, N.N. (1986). The Interface of Cells and their Matrices in Mineralised Tissues: a Review. Scanning Electron Microscopy, 4, 1555-1569.

Julesz, B. (1962). Visual Pattern Discrimination. IRE Trans. Inform. Theory, IT-8, 84-92.

Kapur, J.N., Sahoo, P.K. & Wong, A.K.C. (1985). A New Method for Gray-Level Picture Thresholding using the Entropy of the Histogram. Computer Vision Graphics & Image Processing, 29, 273-285.

Keller, J.M., Crownover, R.M. & Chen, R.Y. (1987). Characteristics of Natural Scenes Related to the Fractal Dimension. IEEE Pattern Anal. Machine Intell. PAMI-9, Vol. 5, 621-627.

Keller, J.M., Chen, S. & Crownover, R.M. (1989). Fractal Texture Description. Computer Vision, Graphics & Image Processing, 45, 150-166.

Kirsch, R. (1971). Computer Determination of the consitutuent structure of biological image. Compute. Biomed. Res., Vol. 4, No.3, 315-328.

Kittler, J. & Illingworth, J. (1985). Threshold Selection Based on a Simple Image Statistic. Computer Vision, Graphics & Image Processing, 30, 125-147.

Kittler, J. & Illingworth, J. (1986). Minimum Error Thresholding. Pattern Recognition, 19, 41-47.

Lachenbruch, P.A. & Mickey, M.R. (1968). Estimation of Error Rates in Discriminant Analysis, Technometrics, 10, 1-11.

Lachenbruch, P.A. (1975). Discriminant Analysis. Hafner Press, MacMillan Publishing Co. Inc, New York.

Laws, K.I. (1979). Texture Energy Measures. Proc. Image Understanding Workshop, 47-51.

Laws, K. (1980). Textured Image Segmentation. US-CIPI Report No. 940, Signal & Image Processing Inst. Uni. Southern California, Los Angeles.

Lee, S.U., Chung, S.Y. & Park, R.H. (1990). A Comparitive Performance Study of Several Global Thresholding Techniques for Segmentation. Computer Vision Graphics & Image Processing, 52, 171-190.

Leeson, T.S. & Leeson, C.R. (1970). Histology. W.B. Saunders Co.

Lev, A., Zucker, S. & Rosenfeld, A. (1977). An application of Relaxation Methods to Edge Reinforcement. IEEE Trans. Systems, Man & Cybernetics, SMC-7, No.11, 435-442.

Lindsay, R. (1984). The Osteoporoses. In The Pathological Basis of Metabolic Bone Diseases and their Therapies. Columbia University.

Lu, S.Y. & Fu, K.S. (1978). A Syntactic Approach to Texture Analysis. Computer Graphics & Image Processing, 7, 303-330.

Mandelbrot, B.B. (1982). The Fractal Geometry of Nature. San Francisco, CA: Freeman.

Mitchell, O.R., Myers, C.R., & Boyne, W. (1977). A max-min measure for image texture analysis. IEEE Trans. Computers, C-26, 404-414.

Muhamad, A.K. & Deravi, F. (1992). Neural Networks for Texture Classification. Proc. 4th International Conference on Image Processing, Maastricht, The Netherlands, 201-203.

Nagao, M. & Matsuyama, T. (1979). Edge-preserving smoothing. Computer Graphics & Image Processing, Vol. 9, No.4, 394-407.

Ohanian, P.C. & Dubes, R.C. (1992). Performance Evaluation for Four Classes of Textural Features. Pattern Recognition, Vol. 25, No. 8, 819-833.

Olivo, J.C. (1994). Automatic Threshold Selection Using the Wavelet Transform. Computer vision Graphics & Image Processing, Vol.56, No.3, 205-218.

Ornoy, A., Kim, O.J. (1977). Scanning Electron Microscopy Findings in Osteogenesis Imperfecta Fetalis. Isr. J. Med. Sci., 13, 26-32.

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histogram. IEEE Trans. Systems, Man & Cybernetics, SMC-9, 62-66.

Parkkinen, J. & Selkainaho, K. (1990). Detecting Texture Periodicity from the Cooccurrence Matrix. Pattern Recognition Letters, 11, 43-50.

Pavlidis, T. (1977). Structural Pattern Recognition. Springer-Verlag, Berlin/New York.

Peleg, S., Naor, J. Hartley, R. & Avnir, D. (1984). Multiple Resolution Texture Analysis and Classification. IEEE Trans. Pattern Anal. Machine Intell. PAMI-6, No.4, 518-523.

Pentland, A.P. (1984). Fractal-Based Description of Natural Scenes. IEEE Trans. Pattern Anal. Machine Intell. PAMI-6, No. 6, 661-674.

Pietikainen, M., Rosenfeld, A. & Davis, L.S. (1983). Experiments with texture classification using averages of local pattern matches. IEEE Trans. Systems, Man & Cybernetics, SMC-13, No.3, 421-426.

Pratt, W.P. (1978). Digital Image Processing. John Wiley & Sons, New York.

Prewitt, J.M.S. (1970). Object enhancement and extraction, in Picture Processing and Psychopictorics. (Eds. B.S.Lipkin & A.Rosenfeld), Academic Press, New York.

Pun, T. (1980). A New Method for Gray-Level Picture Thresholding using the Entropy of the Histogram. Signal Processing, 2, 223-237.

Reid, C.A., Moss, V.A., Kay, J.W. & Elder, H.Y. (1990). The Identification of Boundaries and Classification of Mineralised Bone Surface Types in Scanning Electron Microscopy (SEM) Images. Trans. Royal Microscopical Soc., Vol. 1, 543-546.

Reid, S.A. & Boyde, A. (1983). SEM and Osteogenesis Imperfecta, Oxford, 29-31.

Reid, S.A. (1986). A Study of Human Skeletal Maturation using the Scanning Electron Microscope. Thesis, University College London.

Roberts, L.G. (1965). Machine perception of three-dimensional solids, in Optical and Electrooptical Information Processing. (Ed. J.T.Tippett), MIT Press, Cambridge, Mass.

Robinson, G. (1977). Edge detection by compass gradient mask. Computer Graphics & Image Processing, Vol.6, 492-501.

Rosenfeld, A. & Thurston, M. (1971). Edge and curve detection for visual scene analysis. IEEE Trans. Computers, C-20, No.5, 562-569.

Rosenfeld, A, Thurston, M. & Lee, Y.H. (1972). Edge and curve detection: Further Experiments. IEEE Trans. Computers, C-20, No.7, 677-715.

Rosenfeld, A. Hummel, R. & Zucker, S. (1976). Scene Labeling by Relaxation Operations. IEEE Trans. Systems, Man & Cybernetics, SMC-6, No.6, 420-433.

Rosenfeld, A. & Kak, A. (1976). Digital Picture Processing. Academic Press, New York.

Sahoo, P.K., Soltani, S. & Wong, A.K.C. (1988). A Survey of Thresholding Techniques. Computer Vision, Graphics & Image Processing, 41, 233-260.

Schachter, B. J., Lev, A., Zucker, S.W. & Rosenfeld, A. (1977). An Application of Relaxation Methods to Edge Reinforcement. IEEE Trans. Systems, Man & Cybernetics. SMC-7, No.11, 813-816.

Schachter, B.J., Rosenfeld, A. & Davis, L.S. (1978). Random Mosaic Models for Textures. IEEE Trans. Systems, Man & Cybernetics, SMC-8, No.9, 694-702.

Siew, L.W. & Hodgson, R.M. (1988). Texture Measures for Carpet Wear Assessment. IEEE Trans. Pattern Anal. Mach. Intell., 10, No.1, 92-105.

Spann, M. & Wilson, R. (1985). A Quad-Tree Approach to Image Segmentation which Combines Statistical and Spatial Information. Pattern Recognition, 18, No. 3/4, 257-269.

Sun, C. & Wee, W.G. (1983). Neighbouring grey level dependence matrix for texture classification. Computer Vision, Graphics & Image Processing, 23, 341-352.

Tamura, M., Mori, S. & Yamawaki, T. (1978). Textural Features corresponding to Human Visual Perception. IEEE Trans. Systems, Man & Cybernetics, SMC-8, No.6, 460-473.

Teitelbaum, S.L., Kraft, W.J., Lang, R. & Avioli, L.V. (1974). Bone Collagen Aggregation Abnormalities in Osteogenesis Imperfecta. Calc. Tiss. Res., 17, 75-79.

Terzopoulos, D. (1985). Cooccurrence Analysis of Speech Waveforms. IEEE Trans. Acoust., Speech, Sig. Proc., 33, No. 1, 5-30.

Tsai, W. (1985). Moment-Preserving Thresholding: A New Approach. Computer Vision Graphics & Image Processing, 29, 377-393.

Tsuji, S. & Tomita, F. (1973). A Structural Analyser for a Class of Textures. Computer Graphics & Image Processing, 7, 303-330.

Van Gool, L., Dewaele, P. & Oosterlinck, A. (1985). Survey, Texture Analysis Anno 1983. Computer Vision, Graphics & Image Processing, 29, 336-357.

Vickers, A.L. & Modestino, J.W. (1982). A Maximum Likelihood Approach to Texture Classification. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-4, No.1, 61-68.

Walder, P., MacLaren, I. & Reid, C. (1994). Neural Networks and Cloud Classification. Proc. Image and Signal Processing for Remote Sensing, Rome, Italy, 2-11.

Wang, S., Velasco, D., Wu, A.Y. & Rosenfeld, A. (1981). Relative Effectiveness of Selected texture Primitive Statistics for Texture Discrimination. IEEE Trans. Systems, Man & Cybernetics, SMC-11, No.5, 360-370.

Wechsler, H. & Citron, T. (1980). Feature Extraction for Texture Classification. Pattern Recognition, 12, 301-311.

Wechsler, H. & Kidode, M. (1979). A Random Walk Procedure for Texture Discrimination. IEEE Trans. Pattern Anal. Mach. Intell., PAMI-1, No.3, 272-280.

Weszka, J.S., Dyer, C.R. & Rosenfeld, A. (1976). A Comparative Study of Texture Measures for Terrain Classification. IEEE Trans. Systems, Man & Cybernetics, SMC-6, No.4, 269-285.

Wilson. R, & Spann, M. (1988). Finite Prolate Spheroidal Sequences and their Applications II: Image Feature Description and Segmentation. IEEE Trans. Pattern Anal. Mach. Intell., PAMI-10, No.2, 193-203.

Wu, A.Y. & Rosenfeld, A. (1982). Threshold Selection using Quadtree. IEEE Trans. Pattern Anal. Mach. Intell., PAMI-4, 90-94.

Yogesan, K., Albregtsen, F. Reith, A. & Danielsen, H.E. (1993). Cooccurrence and Run Length-Based Texture Analysis of Experimental Liver Carcinogenesis in Mice. Proc. 8th Scandinavian Conf. on Image Analysis, 227-234.

Zucker, S.W., Rosenfeld, A. & Davis, L. (1975). Picture Segmentation by Texture Discrimination. IEEE Trans. Computers, C-24, No.12, 1228-1233.

Zucker, S.W., (1976). Toward a Model of Texture. Computer Graphics & Image Processing, 5, No.2, 190-202.

Zucker, S.W., Hummel, R. & Rosenfeld, A. (1977). An Application of Relaxation Labeling to Line and Curve Enhancement. IEEE Trans. Computers, C-26, No.4, 394-403.

Zucker, S.W. & Terzopoulos, D. (1980). Finding structure in Cooccurrence matrices for Texture Analysis. Computer Graphics & Image Processing, 12, 286-308.