Colour Texture Analysis in Machine Vision

by

Tele Seng Chu Tan

Thesis submitted to the University of Surrey
for the degree of
Doctor of Philosophy

Department of Electronic and Electrical Engineering
University of Surrey
Guildford, Surrey
United Kingdom

July 1993
Dedicated to my beloved Mother,
and the loving memory of my Father
Acknowledgements

My greatest appreciation and gratitude to Professor Josef Kittler for his constant supervision and guidance during the course of this work. His invaluable suggestions and encouragements have been the cornerstone for the successful completion of this research. I am also indebted to Dr. Maria Petrou for her assistance in this work and especially grateful to her for the careful and critical reading of the draft thesis.

Special thanks to my loving companions, Christina Chew and “Bear” whose encouragement, support and friendship had made life as a research student an enjoyable and memorable experience. Also a fond thank you goes to my many friends, James Song, Petros Papachristou, Dr. Nam Choakjarernwanit, Paolo Remagnino, Dr. Pavel Pudil, Athena Stassopoulou, Alan Cheasley, and Isaac Ng each of whom had helped in their own special ways by providing me with a wealth of technical and moral support.

This research was made possible due to the generous sponsorship of the Ministry of Defence, Singapore. Their financial assistance is greatly appreciated.
Abstract

Texture is an important cue in vision and has been analysed in its own right for the last three decades by researchers in psychophysics as well as computer vision and image processing. The three important early vision roles that texture analysis can play include texture classification, texture description and texture segmentation, all of which are pre-requisites for higher levels of analysis namely image interpretation and understanding. It is well known that colour can aid the human vision system in the analysis of many visual phenomena like shape, motion and texture. This notion coupled with the recent advent of fast computing hardware and the widespread availability of good quality colour cameras, digitizers, and monitors had created a new pathway for improving the performance of traditional grey level texture analysis schemes by incorporating colour information.

In this thesis, the problem of statistical colour texture analysis is addressed. As a pre-requisite to analysing colour textures a review of the main texture analysis techniques available in the open literature is presented.

The local linear transform technique is singled out as the main texture analysis scheme to be used throughout the course of the work. This technique boasts of several advantages; compactness in texture measurement, implementation simplicity, and suitability for stochastic or random texture representation. It is found that the structural property of the local linear transform for texture measurement resembles that of the energy measures based on Gabor functions. This has resulted in the possibility of emulating the latter texture extraction process by a set of quadrature filters like in the case of Gabor filtering. The motivation here is the speed improvement in the computation of the texture representation as the filtering process can be accelerated by Fast Fourier Transform. But unfortunately the number of quadrature filters needed to successfully emulate the
local linear transform measures has been found to be unexpectedly large making the FFT imple- 
mentation very uneconomical to realise.

Two colour texture analysis schemes are developed. The first method advocates the dual trans- 
formation of the colour input image which requires the initial transformation of the tristimulus 
values into several colour co-ordinate systems and then extracting texture attributes from these 
transformed component images. The performance of these features is measured as the percentage 
of correct classification. Feature behaviour under illumination intensity variation will be investigated.

The second approach harnesses the texture and colour information separately in an attempt 
to eliminate redundant or highly correlated features that are usually associated with the first ap- 
proach. The colour histogram is used as an image model from which a colour representation 
scheme of this method can be derived. An efficient and fast way of coding the colour histogram 
by approximate principal component analysis is developed here. This reduces both the mem- 
ory requirement for histogram storage and computation time for colour features by a factor $N_g^2/9$, 
where $N_g$ is the total number of grey level of each channel.

It is shown that features derived from the latter approach perform better in experiments in- 
volving colour granite classification. These colour features are shown to be more robust to illumi- 
nation intensity changes than colour texture features computed from the individual transformed 
channels. Further to this, the overall size of the colour texture feature dimension of the second 
approach is considerable lower than the first approach. The encouraging results gathered here 
indicate the usefulness of a hybrid form of multi-variate feature measurement of colour texture 
using separately, colour and texture attributes.
Contents

1 Introduction ................................................................. 1
  1.1 Definition of Texture .................................................. 2
  1.2 Types of Textures ....................................................... 3
  1.3 Importance of Texture ................................................ 4
  1.4 Roles of Texture Analysis .......................................... 6
  1.5 Texture Cues .............................................................. 7
  1.6 Objectives of This Thesis ........................................... 8
  1.7 Achievements ............................................................ 9
  1.8 Thesis Overview ........................................................ 10

2 A Review of Texture Analysis Techniques .................. 15
  2.1 Preliminaries ............................................................... 15
  2.2 Statistical Texture Analysis ........................................ 16
    2.2.1 Second-Order Statistics ........................................... 16
      2.2.1.1 Grey Level Co-occurrence Matrix ......................... 16
      2.2.1.2 Grey Level Difference Method ............................. 19
      2.2.1.3 Fourier Power Spectrum ..................................... 20
      2.2.1.4 Autocorrelation .................................................. 21
    2.2.2 Higher-Order Statistics ......................................... 21
      2.2.2.1 Grey Level Run Length ......................................... 21
    2.2.3 Multi-Channel Filtering Analysis ......................... 22
      2.2.3.1 Gabor Channel Filtering ...................................... 22
      2.2.3.2 Local Linear Transforms .................................... 23
    2.2.4 Local Geometrical Feature Analysis .................... 24
      2.2.4.1 Edgeness Measure .............................................. 24
4 Colour Texture Feature Extraction

4.1 Introduction ................................................................................................................................ 75
4.2 Colour Vision ................................................................................................................................ 77
4.3 Trichromatic Model .......................................................................................................................... 78
4.4 R_G_N - NTSC Receiver Primary Colour System ................................................................................. 79
4.5 Colour Co-ordinate Systems .......................................................................................................... 80
   4.5.1 XYZ Colour Co-ordinate System ......................................................................................... 80
   4.5.2 YIQ NTSC Transmission Colour Co-ordinate System ......................................................... 80
   4.5.3 YUV Colour Co-ordinate System ......................................................................................... 81
   4.5.4 I_1I_2I_3 Ohta Colour Co-ordinate System ...................................................................... 81
   4.5.5 IHS Colour Co-ordinate System ......................................................................................... 82
   4.5.6 Normalised I_1I_2I_3 Colour Co-ordinate System .............................................................. 82
   4.5.7 Karhunen-Loeve (KL) Colour Co-ordinate System .............................................................. 83
4.6 Spatial Definition of Transformed Images ..................................................................................... 83
4.7 Energy of Component Images ........................................................................................................ 84
4.8 Multi-Band Model for Colour Texture .......................................................................................... 87
4.9 Spatial Texture Representation ...................................................................................................... 87
4.10 Experiment .................................................................................................................................... 89
   4.10.1 Aim and Experiment Setting ............................................................................................ 89
   4.10.2 Results and Discussion ..................................................................................................... 90
4.11 Conclusion .................................................................................................................................... 94

5 Separable Colour and Texture Feature Description

5.1 Introduction ..................................................................................................................................... 99
5.2 The Colour Histogram .................................................................................................................. 101
5.3 Computational Complexity of the Colour Histogram .................................................................... 103
5.4 Dimension Reduction ..................................................................................................................... 103
5.5 Approximating the Colour Histogram .......................................................................................... 104
   5.5.1 Optimal Principal Axes .......................................................................................................... 104
   5.5.2 Sub-Optimal Principal Axes ................................................................................................ 105
   5.5.3 Approximation of Third-Order Probabilities ...................................................................... 106
5.6 First-Order Histograms ................................................................................................................ 108
5.7 Colour Features from First-Order Histograms ............................................................................. 109
5.8  Aptness of Approximation ................................................................. 112
   5.8.1 Estimation-Maximization Approach to Gaussian Mixture Decomposition ................................................................. 112
   5.8.2 The Estimated Mixture Density ......................................................... 114
   5.8.3 The Estimated Density Principal Axes ................................................ 115
   5.8.4 Conclusion .......................................................................................... 120
5.9  Experiments ............................................................................................ 120
   5.9.1 Approximating the Colour Histograms .................................................... 120
      5.9.1.1 Simulation One .................................................................................... 121
      5.9.1.2 Simulation Two ..................................................................................... 122
      5.9.1.3 Simulation Three .................................................................................. 123
   5.9.2 First-Order Colour Feature Invariance Property ....................................... 123
   5.9.3 Colour Texture Classification with First-Order Colour Features .................. 126
   5.9.4 Colour Features computed directly from Colour Histograms ..................... 129
   5.9.5 Use of First-Order Colour Features for Delineating Cover in Landsat Imagery 131
5.10 Discussion ............................................................................................... 132
5.11 Conclusion ............................................................................................... 136

6  Conclusions and Future Work .................................................................... 140
   6.1 Conclusions ........................................................................................... 140
   6.2 Future Work ........................................................................................... 143

Appendix A : Test Images .............................................................................. 146

Appendix B : Test Images .............................................................................. 147
List of Figures

1.1 Textures of left: straw (d15 of Brodatz) and right: grass lawn (d9 of Brodatz). . . . 2
1.2 Illustrating the types of textures: (a) Surface of cork (d4 of Brodatz). (b) A brick wall, (d95 of Brodatz). (c) Knot in a piece of wood (d72 of Brodatz). (d) lace (d41 of Brodatz). ................................................................. 5
3.1 Transform from 2-dimensional space into a vector sequence. ......................... 44
3.2 The different spatial relationships between two pixels in a 3 × 3 neighbourhood. 45
3.3 Block diagram of the local linear transform texture feature extraction scheme. . 46
3.4 Fourier power spectrum of a set of 3 × 3 DCT filters. .......................................... 49
3.5 Set of 12 Brodatz textures used for classification. .............................................. 52
3.6 Textile defects detected using LVDCT features. ................................................... 57
3.7 Definition of a Gabor filter in the spatial frequency domain. ................................. 58
3.8 Gabor filter channel. ........................................................................................... 59
3.9 Test images. ......................................................................................................... 61
3.10 Intensity plot of frequency responses of Gabor channel filters. ......................... 62
3.11 Segmentation results. The images show the segmentation results on the test images using the two feature representation methods. Rows 1 and 2): Bayesian pixel classifier with the energy of Gabor filter outputs as texture features (EOG), Rows 3 and 4): Bayesian pixel classifier with the local variance taken from a 15 × 15 moving window on DCT filter outputs as texture features (LVDCT). ............................................. 64
3.12 1-Dimensional DCT filtering. ................................................................. 65
4.1 A psychophysical model based on the filtering action of three hypothetical filters $S_r(\lambda), S_g(\lambda),$ and $S_b(\lambda).$ The input $I(\lambda)$ is mapped into the 3-dimensional tristimulus space. An arbitrary vector Q in this space can be specified by the three tristimulus co-ordinates. ................................................................. 79
4.2 continue... ............................................................................................................................................. 85
4.2 The colour transformed component images of the “capao bonito” image. ................... 86
4.3 Multi-band colour texture feature extraction model. ............................................... 88

5.1 A new colour texture representation scheme with two independent sources of in­
for­ma­tion. ............................................................................................................................................. 101
5.2 Colour histogram showing shift of cluster due to light intensity changes. ........... 109
5.3 continue... .............................................................................................................................................. 116
5.3 3-dimensional scatter of the observed, $S(r, g, b)$ and estimated, $\hat{S}(r, g, b)$ distribution of the tristimulus values of some colour texture data. The corresponding scatter plots are plotted with axes of the same scale. ........................................................ 117
5.4 Marginal distribution of the “bianco castilla” colour granite image (Right Plot: Ob­
served distribution, Left Plot: Estimated distribution). ................................................... 118
5.5 (a) Mixture density with the same principal component axes can be effectively de­
correlated by the eigenvectors $e_1$ and $e_2$ respectively. (b) This is simply not the case when the mixture has constituent densities with different principal component axes. Data transformation by eigenvectors $e_1$ and $e_2$ does not in this case guarantee un-correlatedness. ........................................................ 119
5.6 Plot of observed and estimated marginal distribution of a single mode Gaussian noise image using the sub-optimal principal axis transformation (Right Plot : Observed, Left Plot : Estimated, Middle Plot : Residual). ................................. 122
5.7 Plot of observed and estimated marginal distributions of a bimodal Gaussian noise image using (a) optimal principal axis transformation and (b) sub-optimal principal axis transformation (Right Plot : Observed, Left Plot : Estimated, Middle Plot : Residual). .................................................. 124
5.8 Plot of observed and estimated marginal distributions of the “bianco castilla” image using (a) optimal principal axis transformation and (b) sub-optimal principal axis transformation (Right Plot : Observed, Left Plot : Estimated, Middle Plot : Residual).125
5.9 Optimal principal axis first-order colour features: Plot of energy, entropy, variance and mean measures with respect to light intensity changes. ................................. 126
5.10 Sub-optimal principal axis first-order colour features: Plot of energy, entropy, variance and mean measures with respect to light intensity changes. ................................. 127
5.11 (a) Constituent densities aligned along a common major axis in 2-dimensional space.
(b) Constituent densities aligned along a common minor axis. 134

5.12 (a) The two modes of a mixture of Gaussian densities that are located along different major or minor axes. \(e_1\) and \(e_2\) are the principal axes. (b) Reproduced densities from the approximated first-order densities that shows additional modes (or 'ghost' densities) being introduced. 135

A.1 12 colour granite images used for classification. 146
B.1 Experiment with Landsat image using colour features. 147
List of Tables

3.1 Classification results expressed in percentage of successful classification for the sub-optimal transform filters of size 3 × 3. M is the number of features. ........................................... 54

3.2 Classification results expressed in percentage of successful classification for the sub-optimal transform filters of size 4 × 4. M is the number of features. ........................................... 54

3.3 Classification results expressed in percentage of successful classification for the sub-optimal transform filters of size 5 × 5. M is the number of features. ........................................... 54

3.4 Classification results expressed in percentage of successful classification for the sub-optimal transform filters of size 3 × 3 using an extended feature set. M is the number of features. ........................................................................................................................ 55

3.5 Estimation of the set of 5 emulated quadrature filters for the LVDCT feature extraction process. DCT filter {−1,0,1} with smoothing region W = 9. ........................................... 68

3.6 Estimation of the set of 5 emulated quadrature filters for the LVDCT feature extraction process. DCT filter {−1,2,−1} with smoothing region W = 9. ........................................... 68

3.7 Top Matrix : matrix for the LVDCT feature extraction using the DCT filter of the form {−1,0,1}. Bottom Matrix : matrix of the same feature extraction process using the emulated quadrature filters. ........................................................................................................................ 69

3.8 Top Matrix : matrix for the LVDCT feature extraction using the DCT filter of the form {−1,2,1}. Bottom Matrix : matrix of the same feature extraction process using the emulated quadrature filters. ........................................................................................................................ 70

4.1 Energy (variance) values computed on each component images of the colour coordinate systems. ........................................................................................................................ 87
4.2 The results of using the features from the *multi-band colour texture* feature extraction schemes on colour granite classification. The classification accuracies using texture features derived individually on the three components are shown in column 2 through to column 4. The last column indicates the accuracy when all three features are linearly combined. .......................................................... 91

4.3 Confusion matrix for the classification of 12 classes of colour granite textures computed on the grey level component of the image. .......................................................... 92

4.4 Confusion matrix for the classification of 12 classes of colour granite textures computed on the I component of the YIQ colour system. .......................................................... 92

4.5 Confusion matrix for the classification of 12 classes of colour granite textures computed on the S component of the IHS colour system. .......................................................... 92

4.6 The classification results on the colour texture images with varying image illumination intensity values. Features used are the 24 combined *multi-band colour texture* features. .......................................................... 93

5.1 Angles between the \(j^{th}\) global optimal principal axis and the \(j^{th}\) principal component axis of the \(i^{th}\) constituent of the mixture density. .......................................................... 120

5.2 Angles between the \(j^{th}\) global sub-optimal principal axis and the \(j^{th}\) principal component axis of the \(i^{th}\) constituent of the mixture density. .......................................................... 121

5.3 Classification results of combining the grey level LVDCT texture feature with the *optimal principal axis first-order colour features* (OPA Colour) and the *sub-optimal principal axis first-order colour features* (SOPA Colour). \(E_T = \{E_1,E_2,E_3\}\), \(H_T = \{H_1,H_2,H_3\}\) and \(\text{Var}_T = \{\text{Var}_1,\text{Var}_2,\text{Var}_3\}\). .......................................................... 129

5.4 Classification results obtained when combining the grey level LVDCT texture features with features computed directly from the colour histogram. .......................................................... 130

5.5 Energy measures of two distinct colour texture images computed directly from the colour histogram and individually from the sub-optimal 1-dimensional histograms. 131
Chapter 1

Introduction

An age of photography is likely to be an age of texture. With photographic skills and processes currently reaching an excellence little dreamed of fifty years ago, it is hardly surprising that our generation has taken a simultaneous interest in the look, touch, and feel of the world around us - an interest reflected in our clothing, our house furnishing, the very textures of our walls, and paintings we hang there.

Phil Brodatz (1966)

Advances in computing technologies over the past three decades has taken the work on visual texture by Brodatz [5] a crucial step forward. Instead of capturing these phenomena on photographic prints for aesthetic appreciation, it is now possible to use image processing techniques to store and process these textures as images in computer databanks. This latter development coupled with the rapid development in computer vision and statistical pattern recognition techniques, has now opened a new world of possibilities with potential use of visual texture processing ranging from art and design to semiconductor manufacturing [7][14][16][23][33][36].

Since any visual phenomenon possesses its own specific texture, human beings can use textures as cues to recognise them. Human observers are also capable of segmentation and discrimination tasks under conditions (such as brief exposure to a test image) [19] which prevent detail scrutiny of the image. This ability of the human vision system is referred to as “effortless” or “pre-attentive” visual discrimination. In this sense, two images (which do not portray particular objects or forms) are said to have distinct ‘texture’ if they are effortlessly discriminable to the human observers, regardless of the image generation procedure. For example, one can identify the two textures [5] in Figure 1.1 as straw and lawn. Even if one is unable to assign a class for the texture (classify them), it is easy to distinguish them by drawing a borderline between the two
Figure 1.1: Textures of left: straw (d15 of Brodatz) and right: grass lawn (d9 of Brodatz).

different textures. Endowing machines with these same visual skills of the human vision system (but not necessarily duplicating them) is not only an interesting problem in computer vision work, but also practically important for a wide range of applications like in the automatic analysis of aerial photographs or medical images and for autonomous robots with artificial vision systems.

1.1 Definition of Texture

In mechanical engineering, surface texture is the repetitive or random deviation from the nominal surface that forms the 3-dimensional topography [39]. In petrology, textures are the geometrical relationships among the component crystals of a rock and any amorphous materials that may be present (size of the component crystals, shape, distribution, and orientation) [23][33]. In art and design, the predominant dimension attached to texture is that of ‘surface quality related to touch’, [16], or ‘the tactile quality of the surface’, [25].

Webster’s dictionary defines texture to be ‘the visual or tactile surface characteristics and appearance of something’. This calls for a distinction between the visual and tactile dimensions of texture. As an illustration of this point, consider the texture of a reptile skin. Apart from being smooth (the tactile dimension) it has also got cellular and specular markings on it that form a well defined pattern (the visual dimension).

Only the visual dimension of texture will be considered in this thesis. Interestingly, most researchers in computer vision do not attempt to define texture in their work. Interpretations of
what constitutes texture vary widely and this has resulted in numerous attempts at its definition.
Here, some of the most common definitions will be presented. According to Haralick, texture is
characterised by tonal primitive properties as well as the spatial relationships between them [13].
Horn [15] considers texture to be ‘detailed structure in an image that is too fine to be resolved, yet
course enough to produce a noticeable fluctuation in the grey levels of neighbouring cells.’ Some
researchers [22] assume the dominant feature of texture to be the repetition of patterns. According
to Pickett [28], ‘the basic requirement for an optical pattern to be seen as a texture is that there is a
large number of elements (spatial variations in intensity or wavelength), each to some degree vis­
ible, and, on the whole, densely and evenly arrayed over the field of view.’ Rao [31] adopts a more
general view of texture, and consider it to be ‘the surface markings or 2-dimensional appearance
of a surface.’ He attempts a constructive definition of texture and considers a textured image as
any image that can be created by either a linear combination or functional composition or both,
of primitive textures followed by an opaque overlap. A primitive texture is either S, W, or D. Here S, W,
and D symbolise strongly ordered textures, weakly ordered textures, and disordered
textures respectively [31].

1.2 Types of Textures

Almost all types of naturally occurring and man-made textures could be grouped into the follow­
ing categories:

- **Disordered Textures** - This type of textures, sometimes called stochastic textures, is most
  commonly found in nature. Neither distinct repetitive pattern nor orientation can be per­
  ceived from disordered textures. Due to its randomness in form, statistical modelling is
  found to be most suitable to characterise them. Of these, attributes like smoothness, rough­
  ness, coarseness, fineness, regularity or irregularity are often used. Figure 1.2(a) shows a
  disordered texture in a form of the surface of a pressed cork (d4 of Brodatz) which may be
described on the basis of its roughness.

- **Strongly Ordered Textures** - Strongly ordered textures, sometimes called deterministic tex­
  tures, exhibit a unique regularity that can be described by a mathematical formula. They are
  composed of regular arrangements of well-defined sub-patterns or primitives according to
  some well-defined placement rules. As such, structural models for texture description are
  most suitable for them. The brick wall (d95 of Brodatz) in Figure 1.2(b) is an example of a
strongly ordered texture. Strongly ordered textures are most commonly found in man-made object.

- **Weakly Ordered Textures** - An important category of textures which can neither be modeled statistically nor structurally is that of weakly ordered, or oriented textures (termed as "observable texture" by Gool et al. [12]). Such textures are characterized by a dominant local orientation at each point of the texture, which can vary arbitrarily. Figure 1.2(c) shows a knot in a piece of wood (d72 of Brodatz) which signifies a weakly ordered texture.

- **Compositional Textures** - This is the most difficult texture type to characterize as full description involves a complex procedure of determining the right compositional rules that allows the creation of more complex textures from the primitive textures belonging to the classes of disordered, strongly ordered, and weakly ordered textures [31]. Figure 1.2(d) shows the texture of a lace (d41 of Brodatz) which is an example of a compositional texture.

Rao [31] had presented a taxonomy for the description of textures using the above descriptions. His taxonomy attempts to provide standardized symbolic descriptors that can be used for qualitative descriptions of a wide variety of textures. The taxonomy is based on appropriate mathematical models for different kinds of textures, and can classify (using visual language), in principle, all the textures from Brodatz [5].

### 1.3 Importance of Texture

Texture is an important element of vision and has been analyzed in its own right for the last three decades by researchers in psychophysics as well as computer vision [2][3][13][19].

There are several disciplines and applications in which texture plays an important role. Texture plays a critical role in inspecting surfaces that are produced at various stages in all types of manufacturing processes [7]. For instance, in the inspection of semiconductor devices, surface texture is an important factor that is used to decide the integrity of a fabricated device [26]. Texture plays an important role in the area of machine parts inspection. In fact, an entire standard [39] is devoted to the specifications related to surface texture. When a metal is deformed, its grains are reoriented with certain orientations being preferred over others. This development of a preferred orientation gives rise to a texture. The analysis of such directional textures is important because many material properties, such as tensile strength, depend on the distribution of grain directionality. There are several areas like geology [34], petrography [23], metallography [36] and
Figure 1.2: Illustrating the types of textures: (a) Surface of cork (d4 of Brodatz). (b) A brick wall, (d95 of Brodatz). (c) Knot in a piece of wood (d72 of Brodatz). (d) lace (d41 of Brodatz).
lumber processing [7] that make extensive use of textural features such as grain shapes, sizes, and distributions for recognising and analysing specimens. Texture is very important in quality control since many inspection decisions are based on the appearance of the texture of some materials [10][30]. Special textures arise when using the method of flow visualisation, which is becoming increasingly popular in diverse areas such as biomedical engineering, tracer methods, oceanography, aerodynamics, and surface flow [38]. In remote sensing, texture analysis is useful for automatic land use classification, crop yield assessment, terrain classification, mining, weather forecasting, etc. [8][27][37]. Finally, in cytology, texture analysis has been used in cell discrimination and classification [20].

Here, only the major examples which highlight the importance of texture analysis are briefly mentioned. There are of course many other areas, that could have been mentioned where texture analysis can be considered important.

1.4 Roles of Texture Analysis

Given the wide spectrum of application for automated texture analysis in the previous section, it is easy to highlight three key roles that texture analysis plays in the scientific and engineering communities. These important roles are in:

- **Texture Classification** - given a textured sample, to which of a finite number of prototype (classes) does the sample belong to?

- **Texture Description** - given a textured region, how can it be adequately described?

- **Texture Segmentation** - given a scene, how can the boundaries between the major textured regions be established?

These three problems are listed in order of increasing difficulty. The first problem, also known as the texture classification problem, is usually approached by using some statistical method for extracting the characteristic parameters for each one of the given classes. The parameters are then the input features for classification using the well known techniques of statistical pattern recognition. The second problem is more difficult because one could easily find two perceptually different textures and it could be still extremely difficult to describe the differences. The third problem, which is also known as the segmentation problem, is difficult mainly because it is usually unknown at which level of complexity a texture is complete up to its physical boundary, and
1.5: Texture Cues

also because it is not well understood what kind of grouping mechanism is needed for outlining regions of uniform texture. Clearly, a partial solution to the segmentation problem could be found in the first two problems.

1.5 Texture Cues

What kind of information or tokens does the human vision system enlist to help in texture discrimination and classification? This question had prompted neurologists and psycho-physicists alike to investigate the perception of texture by humans.

Early experiments performed by Julesz [17] seemed to suggest that differences in the first or second-order statistics (brightness and granularity respectively) allow texture discrimination for a human subject, but that differences in the third and higher-order statistics are irrelevant as long as discrimination is concerned. More recently, Julesz [18] presented a new construction of textures, where all third-order statistics are constant, and yet discrimination is possible. Second-order statistics (i.e. probabilities of the form $P(i,j)$ where $P(i,j)$ represents the likelihood of grey levels $i$ and $j$ occurring at a given displacement $\delta_x$ and $\delta_y$) have been shown to be very important in the human analysis of textures [17][29]. Additional experiments performed by Pratt et al. [29] showed that the second-order statistical measures should be sufficient for texture analysis, but the mean, variance, and autocorrelation function measures, by themselves, although directly or indirectly necessary, are not sufficient. Beck [1] and Beck et al. [2] presented a hypothesis which conflicts with the original Julesz conjecture that textural segmentation is a function of differences in global second-order statistics. They hypothesized that textural segmentation occurs as a result of differences in the first-order statistics of simple properties such as brightness, colour, size, and the slopes of contours and lines of the elemental descriptors (textural elements) of a texture. This hypothesis was further supported by Marr [24] who stated that the statistical process in textural segmentation operates on symbolic assertions that specify the orientation, size, contrast, position, and termination of the intensity changes.

The important experimental findings by Julesz, Beck, and Marr describing the behaviour of the human vision system for texture perception have motivated the development of many texture analysis algorithms suitable for implementation in a computer. Grey level difference histograms and co-occurrence matrices [12] which relate the second-order statistics of the image pixels have been used extensively and very successfully to characterise the statistical properties of the textural elements. Orientation of the textural elements can be characterised by using multichannel spatial
filtering with a bank of Gabor filters or localized spatial filters that are specifically tuned to the desired orientation and frequency of the elements [4]. This form of texture feature extraction is best suited for weakly-ordered textures that are characterised by a dominant local orientation at each point of the texture. Textural edgeness was used by Rosenfeld and Thurston [32] whereby texture was characterised by the edgeness per unit area of the image. Celenk [6] and Dhawan et al. [9] used first order statistics in the form of colour to describe the individual textons or blobs in their work with colour images.

1.6 Objectives of This Thesis

Although extensive research on texture analysis has been conducted in the past three decades, a general form of texture representation has yet to surface. Due to the extreme difficulty in categorising a wide range of textures found in the world around us, this possibility is viewed to be very remote by the research community. In part of this thesis, a review of most existing approaches to texture representation will be carried out. This survey has two aims: firstly, there is a need to examine the relationships between existing approaches and to appreciate any strengths or pitfalls associated with each one of them. Secondly, knowledge of some of the existing techniques is necessary for the presentation of the methods proposed in this work. In order to narrow down the scope of this thesis, only the statistical approaches to texture representation will be considered.

Besides texture, colour information is also a very important cue to the human vision system. From the psychophysical point of view, colour aids the human vision system in two important tasks, namely identification and discrimination [11]. It is also known that colour is identified as one of the features observed by humans pre-attentively, and as such it is computed in parallel with other features to be exploited in object interpretation at a later stage. Beck [1] and Marr [24] further suggest that colour of the primitive textural elements found in texture surfaces help in the visual discrimination of texture by humans. It is then natural to assume that the perception of texture can be aided by colour information. Unfortunately work on colour texture analysis for scene description and identification has not made much progress. Only few attempts on colour texture are reported in the literature. As most of them are very ad hoc in nature, only limited successes have been reported [6][9]. This thesis is devoted to the study of the colour pigment characteristic of colour texture and aims to develop effective and efficient algorithms for colour texture analysis based on this characteristic. The main interest lies primarily in developing colour texture features suitable for use in colour texture classification with possible extension to texture
In summary, the objectives of this thesis are:

- To review and examine existing approaches towards texture classification (Chapter 2).
- To study and implement an existing texture representation method for grey level texture classification (Chapter 3).
- To develop new, effective and efficient algorithms towards colour texture feature extraction based on texture and colour cues (Chapters 4 and 5) and to compare the merits and demerits of the different techniques.
- To identify worthwhile directions for further research (Chapter 6).

1.7 Achievements

In view of the objectives set out in the previous Section, several special contributions from this work can be summarised as follows:

- The spectral properties of the local linear transform for texture characterisation have been investigated. The findings improve the understanding of the advantages and limitations of this family of methods of texture representation and expose why it proved so suitable for disordered (stochastic) grey level textures.

- The development of a new colour texture representation scheme that extracts textural information from each of the transformed colour bands (channels) of the colour image has been accomplished. This scheme allows a fast computation of the features as the texture algorithm used is the local linear transform whose feature extraction operations can be easily realised using DSP hardware. The resulting dual transformation composed of colour transformation followed by texture feature extraction is demonstrated to be a powerful tool for colour texture discrimination.

- The assumption pertaining to the structure of the probability distribution function of the tristimulus values of colour texture phenomena in the application studied, namely that it is a mixture of similarly oriented, normally distributed densities was validated empirically.

- Using the above assumption, a novel approximation of the colour histogram reducing the computation, storage and colour feature extraction time was introduced. In essence the
colour histogram is approximated by a product of three first-order distribution functions estimated along its approximate principal axes. The use of this method will make it possible to compute colour features a few thousand times faster using much less memory as compared to the evaluation based directly on the colour histograms.

- Some robust illumination invariant first-order colour features that could be computed from the first-order histograms have been introduced. These features are capable of retaining most of the histogram information with the added advantage of a reduction in feature dimension. A successful implementation of these features on colour granite image classification was also realised.

- Colour texture classification established that the approximate first-order colour features have a greater discriminating power as compared to features computed directly from the colour histogram. This is understandable as the former feature set spans a feature space of higher dimensionality and thus it encapsulates more descriptive information than the latter feature set.

1.8 Thesis Overview

In computer vision, textures are analysed on two levels: statistical and structural [12][13]. On the statistical level, local features are computed at each pixel in a texture image, and a set of statistics are derived from the distributions of these local features. The local features are defined by the combination of intensities at specified positions relative to each pixel in the image. According to the number of pixels which define a local feature, statistics are classified into first-order, second-order, and higher-order. Statistics give various measures of texture properties. Statistics need not be only for intensities. Statistics of such local geometrical features as edges, peaks and valleys, and spots or blobs, give measures of specific texture properties. On the structural level, texture is considered to be composed of texture elements. The properties of texture elements as well as the placement rules of the texture elements define the texture. The structural analysis is more complex as compared to the statistical analysis, and it derives much more detailed information; it is possible to reconstruct the original texture from the description of the texture. Chapter 2 reviews and describes statistical and structural approaches for texture analysis found in the literature. The principles, merits, and weaknesses of these approaches will be briefly summarised in this chapter.

From the statistical approaches detailed in Chapter 2, the local linear transform for texture anal-
ysis is selected for experimentation. This representation is credited with several merits that are useful for texture analysis:

- The method is almost as powerful for texture discrimination as methods based on co-occurrence measurements [35] which are considered by some as a benchmark and is known to be computationally very involved.
- It enables a more compact description of local texture properties.
- The feature extraction process is computationally less demanding.
- Because of its parallel structure, this method is well suited for an implementation on a specialised parallel architecture.

Chapter 3 introduces the framework on which the local linear transform for texture analysis is based. It also presents some experiments concerning the classification of some Brodatz texture images using features derived from this method.

The processing of colour images is fast gaining momentum in the computer vision community especially with the advent of fast computing hardware and the widespread availability of good quality colour cameras, digitizers, and monitors. In many instances, the addition of colour information is able to enhance the performance of computer vision systems: e.g. in the visual inspection of surface reflectance which is directly related to the quality of surface finish and paint [21]. The traditional texture analysis methods could be adapted by incorporating the element of spectral or colour attributes into their original formulation. First in Chapter 4, the theory and usefulness of the tristimulus model for colour vision is introduced. By transforming the tristimulus co-ordinate system into other colour co-ordinate systems, and computing texture information on each channel component image, one can devise a series of colour texture features based on this dual transformation technique (i.e. colour transformation followed by texture feature extraction). The merits and demerits of this colour texture representation will be discussed in this chapter.

The colour texture representation of Chapter 4 which is no doubt very attractive when speed is an important factor, has some drawbacks; i.e. most features are not illumination invariant hence not suitable for outdoor environments and the dimensionality of the feature space is high. The aim in Chapter 5 is to develop other possible colour features that will eliminate some of the demerits of this earlier method. This is based on the separate representation of both texture and colour information. Texture analysis on the grey level representation of the colour image can be used to estimate texture attributes. Likewise, colour can be computed separately by any suitable models.
The colour model used here for the colour image data will be the colour histogram. The introduction of the approximation of the colour histogram by a product of three first-order distribution functions estimated along its approximated principal axes will be introduced in Chapter 5. This is followed by the definition of an efficient and robust colour feature extraction technique based on this novel approximation. Procedures and experimental results studying the performance of these colour texture features on classification are also presented.

A short summary concludes each Chapter, but a more detailed discussion is presented in Chapter 6. This is followed by the presentation of some ideas for possible future work.

References


References


Chapter 2

A Review of Texture Analysis Techniques

2.1 Preliminaries

Texture classification which involves establishing a class label (from a finite set of labels) for an unknown sample texture, has been the focus of interest in texture analysis for the last thirty years. The important areas involved in accomplishing this task are: texture feature extraction and classifier design. As the methodology for classifier design is well developed [20][23], the critical step here is feature extraction. Feature extraction is the process of choosing texture features that best characterise the texture with the aim of reducing the dimensionality of input data to a computationally reasonable level, and yet preserving the salient information of the underlying texture being considered.

Texture analysis can be performed in two levels: statistical and structural [25][27]. On the statistical level, a set of statistics can be derived from the distributions of local features computed from the texture image. These statistics are classified into first-order, second-order, and higher-order statistics. They can be defined for pixel intensities at specified spatial positions relative to each pixel in the image. Other examples include local geometrical features such as edge, peak and valley, and spot or blob, which all give measures of texture properties. On the structural level, texture is considered to be composed of texture elements. The descriptions of texture elements as well as the placement rules of the texture elements define the texture. It should be pointed out that such a categorisation is rather artificial as the distinction between some of the texture analysis approaches is not always clear. Nonetheless, the grouping is necessary as it provides a means of understanding the fundamentals of the texture analysis tools and their applicability to the type
of texture concerned.

The aim of this Chapter is to review existing statistical and structural texture analysis techniques and to highlight similarities or differences between them. It is beyond the scope of this thesis to present in-depth analysis of these different approaches. Further details could be found in the many references cited therein.

2.2 Statistical Texture Analysis

Most statistical texture analysis methods are primarily motivated by neurological and psycho-physical evidence concerning texture perception by the human vision system. For example, an earlier conjecture by Julesz [31] which hypothesizes that differences in the first or second-order statistics allow for texture discrimination for a human subject propelled the introduction of the co-occurrence matrix by Haralick et al. [26]. Likewise, the belief that the early visual representation of features can be closely modeled by a Fourier-like decomposition into spatial-frequency components [43] lead to the development of filter-based approaches that have good joint spatial and spatial-frequency domain resolution [18][40].

Statistical approaches can be broadly classified into the following main groups: second-order statistics, higher-order statistics, multi-channel filtering analysis, local geometrical feature analysis, descriptor-based analysis, and model-based analysis.

2.2.1 Second-Order Statistics

2.2.1.1 Grey Level Co-occurrence Matrix

Studies of the human vision system relating to texture perception suggested that differences in the second-order statistics allow for the discrimination of textures. This is probably the primary motive for the emergence of the grey level co-occurrence matrix (GLCM) technique for texture analysis.

Let \( \delta = (r, \theta) \) denote a vector in the polar co-ordinates of the image. For any such vector, one can compute the joint probability of the pairs of grey levels that occur at pairs of points separated by \( \delta \). This joint probability takes the form of an array \( P_\delta \), where \( P_\delta(i,j) \) is the probability of the pair of grey levels \( (i,j) \) occurring at separation \( \delta \). This array is called the co-occurrence matrix.

Finding co-occurrence matrices for all \( \delta \) requires a prohibitive amount of computation time. Haralick et al. [26], who first used co-occurrence matrices to classify terrain in aerial photographs, computed just four co-occurrence matrices for \( r = 1 \) and \( \theta = 0^\circ, 45^\circ, 90^\circ, \) and \( 135^\circ \). From each
matrix, they computed 14 features for discriminating between textures, but usually only a set of 5 features are employed. These features are called energy, entropy, local homogeneity, inertia, and correlation. They are defined as follows:

- **Energy:**
  \[
  \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} [P_s(i,j)]^2
  \]  
  (2.1)

- **Entropy:**
  \[
  \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} P_s(i,j) \log P_s(i,j)
  \]  
  (2.2)

- **Local Homogeneity:**
  \[
  \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} \frac{1}{1 + (i-j)^2} P_s(i,j)
  \]  
  (2.3)

- **Inertia:**
  \[
  \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} (i-j)^2 P_s(i,j)
  \]  
  (2.4)

- **Correlation:**
  \[
  \frac{\sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} (i - \mu_x)(j - \mu_y) P_s(i,j)}{\sigma_x \sigma_y}
  \]  
  (2.5)

where

\[
\mu_x = \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} P_s(i,j)
\]  
(2.6)

\[
\mu_y = \sum_{j=0}^{N_y-1} \sum_{i=0}^{N_x-1} P_s(i,j)
\]  
(2.7)

\[
\sigma_x = \sum_{i=0}^{N_x-1} (i - \mu_x)^2 \sum_{j=0}^{N_y-1} P_s(i,j)
\]  
(2.8)

\[
\sigma_y = \sum_{j=0}^{N_y-1} (j - \mu_y)^2 \sum_{i=0}^{N_x-1} P_s(i,j)
\]  
(2.9)

where \(N_x\) is the number of image grey levels. The effectiveness of the GLCM and their features has been extensively illustrated by a number of researchers both theoretically and empirically. The grey level difference method (GLDM) was used by Weszka et al. [71] in a comparative study with the GLCM and the Fourier power spectrum (FPS) method. The grey level difference method and
the Fourier power spectrum method will be discussed in more detail in Section 2.2.1.2 and Section 2.2.1.3 respectively. They concluded that generally, the GLCM is superior to the FPS method. They also found out that the performance of the GLDM and the GLCM methods is quite comparable. This can be explained as follows: if one sums up the elements of GLCM along lines parallel to its main diagonal, one will obtain the total numbers of point pairs having a given grey level difference. Thus features derived from GLCM are likely to be similar to features from the GLDM. As a result, the GLDM is usually preferred to the GLCM as the former approach is computationally less expensive.

To mention a few examples concerning the use of GLCM, Terzopoulos and Zucker [59] used the GLCM features to detect Osteogenesis Imperfecta, which is a highly prevalent genetically determined disease. They reported that an identification accuracy of 90% can be obtained by means of texture analysis whereas the accuracy of visual inspection by medical experts is only 69%. Chen and Pavlidis [8] segmented a textured image into uniform texture regions based on the comparison between GLCM measures computed from adjacent windows. In addition to the extraction of statistical measures, GLCM is also useful for extracting some structural information as shown in [76]. Recently, He et al. [29] proposed a set of 3 new GLCM features, namely diagonal moment, high level moment, and low level moment, to substitute the classical Haralick’s GLCM features. Vickers and Modestino [66] used the co-occurrence matrices directly as features for their maximum likelihood classifier.

Generalised Co-occurrence Matrices (GCM) were studied by Davis et al. [17]. A GCM reflects the shape, size, and spatial arrangement of the texture elements. Three prototypes were defined, pixel-intensity spatial prototype, edge-pixel spatial prototype and extended-edge spatial prototype. The GCM reduces to GLCM when only the pixel-intensity spatial prototype is used.

A point worth noting about the GLCM is that the computation of these matrices is very intensive. This becomes extremely burdensome when the number of grey levels and image size increases. This computation pain can be moderated by reducing the number of grey levels at the price of possible textural information loss. In practice, one usually does not know a priori what displacement vector \( \delta \), should be used, consequently a set of several displacement vectors is often employed.

In another attempt at decreasing the computation burden of the GLCM, Unser [64] approximated the GLCM by defining the principal axes of the joint probability densities of two pixels and arrived at the Sum and Difference histograms (S&D) for texture representation. This method is
able to extract, in a computationally less expensive manner, features that are similar to Haralick's
coop-occurrence method. This follows from the fact that the difference histogram is similar to the
GLDM histogram in its true form.

2.2.1.2 Grey Level Difference Method

The difference statistics are the distribution of probability \( P_\delta(k) \) \( (k = 0, \ldots, N_\delta - 1) \) that the absolute
grey level difference is \( k \) between the points separated by \( \delta \) in the image. Difference statistics are
a subset of the co-occurrence matrix, and can be derived from the matrix by

\[
P_\delta(k) = \sum_{i=k}^{N_\delta-1} P_\delta(i, i - k) + \sum_{i=0}^{N_\delta-1-k} P_\delta(i, i + k) + \sum_{j=k}^{N_\delta-1} P_\delta(j - k, j) + \sum_{j=0}^{N_\delta-1-k} P_\delta(j + k, j)
\]

Weszka et al. [71, 72] computed the following properties from this distribution:

- Angular Second Moment:

\[
\sum_{k=0}^{N_\delta-1} [P_\delta(k)]^2
\]

- Contrast:

\[
\sum_{k=0}^{N_\delta-1} k^2 P_\delta(k)
\]

- Entropy:

\[
- \sum_{k=0}^{N_\delta-1} P_\delta(k) \log P_\delta(k)
\]

- Mean:

\[
\sum_{k=0}^{N_\delta-1} k P_\delta(k)
\]

The appropriateness of the Grey Level Difference Method (GLDM) for texture analysis is easy
to see. For coarse textures, the intensity transitions from pixel to pixel are mostly slow or gradual.
Consequently, the difference statistics will have high probability values around low values of \( k \). As
for fine texture, the distribution will have a large dynamic range and the corresponding histogram
will be more spread out. The effectiveness of the GLDM has been studied both theoretically and
empirically by Weszka et al. [71] and Conners et al. [11]. In practice, a single displacement vector
\( \delta \), is rarely used. A set of displacement vectors are usually employed instead.
2.2.1.3 Fourier Power Spectrum

Textures can be analysed in the frequency domain in which they are represented by the Fourier transform. Let \( f(x, y) \) be the texture image of size \( M \times N \). The 2-dimensional discrete Fourier transform is defined by

\[
F(u, v) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \exp^{-2\pi j \left(\frac{ux}{M} + \frac{vy}{N}\right)}
\]  

(2.15)

The power spectrum which represents the strength of each spatial frequency \((u, v)\) is obtained by

\[
P(u, v) = |F(u, v)|^2
\]  

(2.16)

The Fourier Power Spectrum (FPS) of an image is sensitive to the texture coarseness. A coarse texture will have high values concentrated near the low frequency part of the spectrum while fine texture will have values that are more spread out across the spectrum.

Bajcsy [4] computed the following distributions from the power spectrum \( P(r, \theta) \), (where \( r = \sqrt{u^2 + v^2} \) and \( \theta = \tan^{-1}(u/v) \)) in the polar co-ordinates.

\[
P(r) = 2 \sum_{\theta=0}^{\pi} P(r, \theta)
\]  

(2.17)

\[
P(\theta) = \sum_{r=0}^{\infty} P(r, \theta)
\]  

(2.18)

\( P(r) \) and \( P(\theta) \) are the sums of the powers in ring and wedge regions of the power spectrum space respectively. The peak in \( P(r) \) indicates the size of the dominant texture elements or the texture coarseness. The peak in \( P(\theta) \) indicates the directionality of the texture.

These conventional features were criticised for their poor discriminating power [11][71]. This limitation is not due to the insufficiency of the FPS for texture description but is rather related to the way of defining spectral features and of computing the Fourier spectrum itself as investigated by Dastous et al. [15] and Chen [9]. They suggested a reappraisal of the FPS technique.

Several remarks are now in order regarding the use of the FPS method. In this method, no consideration has been taken into account of the phase aspect of the texture. The neglect seems reasonable as the phase spectrum is not shown to convey much useful information for texture discrimination [14][21][69]. Secondly, the autocorrelation function and the FPS form a Fourier transform pair, and their information content is in principle equivalent. Finally, although FPS can reflect some global features like directionality and regularity, it is incapable of representing local features, which are believed to be critical for preattentive texture perception [33]. This problem may be overcome to a certain extent by applying the Fourier transform to local regions [5].
2.2: Statistical Texture Analysis

2.2.1.4 Autocorrelation

From one point of view, the spatial size of the texture primitive can be used to discriminate textures. Primitives of larger size are indicative of coarser textures, whereas primitives of smaller size indicate fine textures. A function that responds to this change in primitive size is the autocorrelation function. The autocorrelation function $C(i,j)$, of an image $f(x,y)$ (size $M \times N$) can be written as

$$C(i,j) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y)f(x+i,y+j)$$ (2.19)

For coarse texture, the autocorrelation function will drop off slowly when the original image window is moved. If the primitives are small, then the autocorrelation will drop off quickly. Also, directionality of the pattern can be detected from the directional dependence behaviour of this function. Kaizer [34] used the $1/e$ drop off distance as a measure for texture coarseness. This drop off distance in a 1-dimensional case is the distance $d$ such that the autocorrelation function in a particular direction at $d$ assumes the value $1/e$: $C(d) = 1/e$. It can be pointed out that the autocorrelation method and the FPS are closely related as the autocorrelation function and the power spectrum are the transforms of each other. Correlation and the GLDM measures found an industrial application in an automatic visual inspection system reported in [63].

2.2.2 Higher-Order Statistics

2.2.2.1 Grey Level Run Length

The Grey Level Run Length (GLRL) matrix $P_{\theta}(i,j)$ $(i = 1,\ldots,m, j = 1,\ldots,n)$, represents the probability that $j$ points with grey level $i$ will continue in the direction $\theta$. Galloway [24] computed the following 5 properties from the GLRL matrices at $\theta = 0^\circ, 45^\circ, 90^\circ, \text{and } 135^\circ$, for terrain classification of aerial photographs.

- Short Run Emphasis:

$$\frac{\sum_{i=1}^{n} \sum_{j=1}^{l} p_{\theta}(i,j)}{\sum_{i=1}^{n} \sum_{j=1}^{l} P_{\theta}(i,j)}$$ (2.20)

- Long Run Emphasis:

$$\frac{\sum_{i=1}^{n} \sum_{j=1}^{l} j P_{\theta}(i,j)}{\sum_{i=1}^{n} \sum_{j=1}^{l} P_{\theta}(i,j)}$$ (2.21)

- Grey Level Nonuniformity:
Chapter 2: A Review of Texture Analysis Techniques

\[
\frac{\sum_{i=1}^{n} \left( \sum_{j=1}^{l} P_\theta(i,j) \right)^2}{\sum_{i=1}^{n} \sum_{j=1}^{l} P_\theta(i,j)} \quad (2.22)
\]

- Run Length Nonuniformity:

\[
\frac{\sum_{i=1}^{n} \left( \sum_{j=1}^{l} P_\theta(i,j) \right)^2}{\sum_{i=1}^{n} \sum_{j=1}^{l} P_\theta(i,j)} \quad (2.23)
\]

- Run Percentage:

\[
\frac{\sum_{i=1}^{n} \sum_{j=1}^{l} P_\theta(i,j)}{A} \quad (2.24)
\]

where \( A \) is the area of the image.

Intuitively, coarse textures are more likely to have long runs whereas fine ones have short runs. Using these features in conjunction with four directions 0°, 45°, 90°, and 135°, Galloway [24] demonstrated that about 85% classification accuracy can be obtained in classifying swamp, lake, railroad, orchard, scrub, and suburb. However, the GLRL method is extremely noise sensitive which explains why the two comparative studies [11][71] had established the inferiority of this method. To remedy this noise effect, Zhuang and Dunn [74] suggested using the Amplitude Varying Rate Statistics Method (AVRS) which is more robust than the GLRL method and less prone to the effect of noise.

2.2.3 Multi-Channel Filtering Analysis

2.2.3.1 Gabor Channel Filtering

Recent research suggests that the early processing stages embedded in the human vision system involve a set of parallel and quasi-independent mechanisms or channels which are modelled as specially tuned narrow-band filters. The incoming pictorial information is decomposed by this set of channels into a sequence of filtered images, each of which contains intensity variations over a narrow range of frequency and orientation specifying the regularity, coarseness and directionality of the original image. The findings also show that the channel filter responses have optimal joint localization in both the spatial and the spatial-frequency domains and can be mathematically approximated by Gabor functions [18]. A complex 2-dimensional Gabor function can be interpreted as a 2-dimensional Gaussian modulated by a spatial complex sinusoidal grating in spatial domain. The filter frequency response is real and can be specified by four parameters named as the radial center frequency \( f_0 \), the preferred orientation \( \theta_0 \), the radial frequency bandwidth \( B_\omega \), and
the orientation bandwidth $B_\theta$ which are expressed as

$$G(u, v) = e^{-2\pi^2[(u' - f_0)^2 + (v' - f_0)^2]}$$

(2.25)

where $(u', v') = (u \cos \theta_o + v \sin \theta_o, -u \sin \theta_o + v \cos \theta_o)$ are the rotated coordinates. The bandwidths are measured from the half-peak of the Gaussian envelope. The spatial extent of the Gaussian envelope in spatial domain is governed by scaling factors $\sigma_x$ and $\sigma_y$ which are given by

$$\sigma_x = \frac{1}{f_0 \pi \tan \frac{\delta_x}{2}} \sqrt{\frac{\ln 2}{2}} \frac{(2^\delta_u + 1)}{(2^\delta_u - 1)}$$

(2.26)

$$\sigma_y = \frac{1}{f_0 \pi \tan \frac{\delta_y}{2}} \sqrt{\frac{\ln 2}{2}}$$

(2.27)

By varying the four parameters $f_0, \theta_o, B_u$, and $B_\theta$, a bank of Gabor filters can be generated which span the Fourier plane. The joint localization property is important in accurate localization of boundaries among texture regions. The texture measures that were used are the energy of the filtered image [7]. Encouraging results have been reported on texture segmentation using a multichannel spatial filtering approach with a bank of Gabor filters or localized spatial filters [7][73].

### 2.2.3.2 Local Linear Transforms

Texture is a local image property and some kind of local operations are intuitively useful for extracting texture measures. The Local Linear Transforms (LLT) technique to be discussed in this section is of this type. In a general LLT technique, the basic steps are the determination of the convolution mask sizes, the choice of the structure of these masks and the definition of the texture measures.

The transform basis functions can be designed to detect edge, spot, wave and ripple structure as suggested initially by Laws [41]. In his approach, Laws took two steps; microstatistic features were first computed using small convolution masks of sizes $3 \times 3$ and $5 \times 5$, and macrostatistic features were then computed over larger windows. The most useful features were found to be the sums of the squared micro-features which Laws termed as texture energy measures. The sums of the absolute micro-features are usually preferred because they are computationally cheaper. Pietikainen et al. [52] found out that the power of the Laws' masks depends on the general forms of the masks (edge-like, spot-like, etc) rather than on the specific numerical values used in the masks. Cohen and You [10] used a set of selective masks that are tuned to the set of textures to be analysed. This is done at the training stage and the criterion was to maximize the Texture Energy for a specific texture. In [1], Ade introduced an adaptive and automatic way of generating
the convolution masks based on eigen analysis. In each 3 × 3 window, he formed a 9-component feature vector consisting of the 9 pixel element in the window. The covariance matrix can be estimated using as the ensemble the feature vectors found in the remaining image. After this, he computed the eigenvectors of the resulting matrix and substituted the elements of the eigenvectors to form the convolution masks (eigen filters). To extract texture features, he then used these eigen filters in a similar manner as Laws. These filters were later used in a texture defect analysis [2]. The relationship between eigen filters and Laws' Masks can be summed up like this: Both sets span the same feature space; however, whereas eigen filters are mutually orthogonal, Laws' masks are not. The eigen filter approach may be criticised for its high computational demand. This is particularly true for large mask size as one has to solve an eigenvector problem of high dimension.

Most of these approaches can be included in a general computational framework suggested by Unser known as the Local Linear Transformation [65]. In his formulation, a bank of local convolution masks can be viewed as a set of basis operators arranged to form a transformation matrix. The outputs of this bank of filters have the equivalent structure as the transformation coefficients which are obtained by matrix multiplication for all local neighbourhoods in a sliding window fashion. An important contribution of this formulation is that it is capable of unifying a range of linear operator design methods under one framework, for instance, the ad hoc approach proposed by Laws or some well-known digital transformation methodologies such as the discrete cosine transform (DCT) [65] can be considered as similar. Recently Ng et al. [49] conducted a comparative study of the two multi-channel filtering approaches for texture representation. They have shown that the orthogonal masks derived from the Discrete Cosine Transform (Discrete Cosine Transform) have some interesting features that resemble Gabor filters. They then assessed the performances of these approaches by segmentation means. They concluded that features derived from the Discrete Cosine Transform filters gave consistently better segmentation results than the energy features of Gabor filters.

2.2.4 Local Geometrical Feature Analysis

2.2.4.1 Edgeness Measure

Edges are located where the intensity changes abruptly. Let f(x, y) denote the grey level value at the point (x, y) in the image. The strength and the direction of the edge element at this point are
2.2: Statistical Texture Analysis

computed respectively as

\[ |e(x,y)| = \sqrt{f_x(x,y)^2 + f_y(x,y)^2} \]  \hspace{1cm} (2.28)

\[ \angle e(x,y) = \tan^{-1} \frac{f_x(x,y)}{f_y(x,y)} + \frac{\pi}{2} \]  \hspace{1cm} (2.29)

where

\[ f_x(x,y) = f(x + 1, y) - f(x - 1, y) \]  \hspace{1cm} (2.30)

\[ f_y(x,y) = f(x, y + 1) - f(x, y - 1) \]  \hspace{1cm} (2.31)

Tomita and Tsuji [61] derived the following texture properties from the first-order statistics of the distributions of the edge elements:

- **Coarseness**: The density of edge elements is a measure of texture coarseness. The finer the texture, the higher the edge density.

- **Contrast**: The mean of the edge strength is a measure of the contrast.

- **Randomness**: The entropy of the histogram of edge strengths is a measure of randomness.

- **Directionality**: The directionality is detected from the histogram of the edge directions. The entropy of the histogram gives a rough measure of directionality. The dominant orientations in directional textures are given by detecting the clusters in the histogram. If \( 2N \) \((N > 0)\) clusters are detected, they indicate the directionality in \( N \) directions. However, if the distribution is uniform, it indicates no directionality in the texture.

The second-order statistics of the edge directions convey some structural texture properties as well. They are

- **Linearity**: The co-occurrence probability of two edge elements with the same edge direction separated by distance \( k \) in the edge direction indicates the linearity of texture.

- **Periodicity**: The co-occurrence probability of two edge elements with the same direction separated by distance \( k \) along the direction perpendicular to the edge direction indicates the periodicity of the texture.

- **Size**: The co-occurrence probability of two edge elements with the opposite direction separated by distance \( k \) along the direction perpendicular to the edge direction represents the size of the texture elements.
2.2.4.2 Relative Extrema Measure

Psychophysical experimental evidence suggested that for the human visual system, important texture information is contained in the relative frequency of local extrema of various sizes in intensity. Motivated by this conjecture, Rosenfeld and Troy [53] suggested the number of extrema per unit area for a texture measure. Their algorithm looks only within a row for relative extrema. Mitchell et al. [48] used the same principle on a “backlash” smoothed scan line with a variable threshold \( T \). The number of extrema versus \( T \) can be used as characteristic of the texture. This measure can be made invariant to multiplicative gain changes and resolution by taking the log of intensities and using the ratio of the number of extrema at each threshold to the next instead of the extrema themselves. Haralick [27] in his review gave some possible 2-D extensions to the above method.

2.2.4.3 Spot Measure

Rosenfeld et al. [55] proposed a spot (blob) detector of size \( M \times M \) at a point \((x, y)\) in a texture image:

\[
s(x, y) = \frac{1}{(2N+1)^2} \sum_{u=-N}^{x+N} \sum_{v=-N}^{y+N} f(u, v) - \frac{1}{(2M+1)^2} \sum_{u=-M}^{x+M} \sum_{v=-M}^{y+M} f(u, v) \tag{2.32}
\]

where \( N \leq M \). Hayes et al. [28] defined the spot size as \( k \) when the value of a spot detector \((2^k, 2^{k-1})\) \((k = 2, ..., 5)\) is the maximum. Any spot whose value is not the maximal in its receptive field is deleted (non-maximal suppression). The average of the spot size correlates with coarseness of the texture. Davis et al. [16] used the regularity of arrangements of spots in the analysis of aerial photographs of spot-like orchards.

2.2.4.4 Primal Sketch

Marr [45] proposed the symbolic representation of an image which is called the primal sketch. Primal sketch of an image can be produced by filtering the image with edge and bar detector masks to detect texture primitives like edges, lines, and blobs. Each primitive has attributes such as orientation, size, contrast, position, and termination points. The following statistics of the attributes can be computed for texture classification:

- The total amount of contour, and the number of blobs, at different contrasts and intensities.
2.2: Statistical Texture Analysis

- **Orientations**: The total number of elements at each orientation, and the total contour length of each orientation.

- **Size**: Distribution of the size parameters defined in the primal sketch.

- **Contrast**: Distribution of the contrast of items in the primal sketch.

- **Spatial Density**: Spatial density of place-tokens defined in the different possible ways, measured using a small selection of neighbourhood sizes.

Length and orientation of lines are useful properties to discriminate between textures. However, this information is not sufficient in instances where two visually discriminable texture possess the same line and orientation properties. Schatz [56] proposed the use of length and orientation, not only of actual lines but also local virtual line terminators of actual lines to overcome this problem. This captures almost all second-order statistics in the point-and-line textures.

### 2.2.5 Descriptor Based Analysis

This category of texture representation attempts to describe texture properties (using simple visual language terms) that could be extracted computationally.

Amadason and King [3] used the Neighbouring Grey Tone Difference Matrix (NGTDM) and defined computationally five properties of texture, namely, coarseness, contrast, busyness, complexity, and texture strength. These definitions were chosen to approximate human perception of textural properties. The results obtained were quite encouraging but the texture ranking order was not as good as the rank ordering performed by human subjects.

Tamura, Mori and Yamawaki [58] defined a different set of computationally derived texture properties namely coarseness, contrast, directionality, linelikeness, regularity, and roughness. They used these features to find a correspondence with the subject's notion of these properties. They concluded that the properties, coarseness, contrast, and directionality gave correlation values of 0.831, 0.904 and 0.823 respectively. They suggested that a more precise form of description for linelikeness, regularity, and roughness is needed.

Benke et al. [6] proposed a set of Convolution Operators in which elements are adjusted to optimize a specific level of correlation between objective measures and psycho-physical assessments. The properties used by them were bloblikeness, linearity, and regularity.

Descriptor based approaches suffer from the lack of an accurate definition of textural properties that the subjects are expected to measure. Even if such a description exists, the subject's
interpretation of these descriptions and what the subject actually observes in the texture are always prone to discrepancy.

2.2.6 Model-Based Analysis

Originally developed in the texture synthesis field, model-based methods are interestingly useful for texture analysis as well. These methods involve determining the relationship a pixel has with its neighbouring pixels. This can be of linear dependence as with autoregressive models or joint probability as with Markov fields.

In the field of signal processing, time series analysis is a well known method for prediction of a one-dimensional signal from the past signal values. The method is extended to image processing by changing the concept of 1-dimensional time to 2-dimensional space [19] to arrive at the autoregressive model. De Souza [57] presented a method for texture classification based on fitting two 2-dimensional autoregressive models. He remarked that this method does not always allow macrotextures to be adequately specified, since practical considerations only allow the order of the autoregressive model to be at most three. He then proposed a new approach in which he extended the 1-dimensional autoregressive model in order to classify macrotextures. Classification is carried out on the basis of a distance measure between the least square estimates of the coefficients of the sample and the parameters of the prototypes.

Kashyap et al. [36] assumed the texture to be a realisation of an underlying random field model, known as the simultaneous autoregressive model (SAR). To the autoregressive part of the model a white Gaussian sequence with zero mean and unit variance is added. After selecting an appropriate neighbourhood set for the autoregression, the model parameters are estimated and used for classification. Kashyap and Khotanzad in a similar fashion developed a rotationally invariant model in the form of the circular autoregressive model (CAR) [37]. They then used the SAR and CAR models with the aim of segmenting texture images. Two of their features have physical interpretation in terms of the roughness and directionality of the textures. They presented segmentation results on natural textures comprising both microtextures and macrotextures. They also used the parameters of the SAR model to classify nine Brodatz textures [39]. A comparison with six popular grey level co-occurrence features were carried out and it was shown that the SAR model performs significantly better.

The dependence of the next generated pixel value on its neighbourhood needs not be restricted to a linear one but can also be based on probability assumptions. Kaneko and Yodogawa [35]
showed that natural textures can be classified on the basis of a Gaussian-Markov field texture model. First the parameters of the Gaussian model are estimated by the maximum likelihood method. Then the mean Bhattacharyya distance between the correct distribution based on the prototype parameters and the distorted distribution based on the parameters of the sample is used as classification measure.

Cross and Jain [13] used the Markov random field as the model and produced blurry, sharp, line-like, and blob-like textures. They also computed the Markov random field model parameters of some natural textures and they further used the parameters to reconstruct the underlying texture. They concluded that Markov random field models are very suitable for microtexture synthesis but not appropriate for synthesising regular and inhomogeneous textures.

Mandelbrot's fractal geometry [44] is a new branch of mathematics providing a very accurate way of modelling the seemingly complex shapes found in nature. Such shapes often possess a remarkable invariance under changes of magnification. This statistical self-similarity may be characterised by a fractal dimension, a number that agrees with our intuitive notion of dimension but spans the non-integer domain. Peleg et al. [50] introduced the “covering blanket approach” for estimating the fractal dimension of a texture. This method uses the concept based on scale varying surface area measurements as suggested by Mandelbrot for fractal objects. They successfully perform classification of Brodatz textures using 48 signatures (features) derived from 48 scales. Pentland [51] used statistics of differences in grey levels between pairs of pixels at varying distances as indicators of the fractal properties of the texture. He successfully applied his models to the problem of 1) texture segmentation and classification, 2) estimation of 3-dimensional shape information and 3) distinguishing between perceptually “smooth” and perceptually “textured” surfaces in the scenes. Keller and Chen [38] suggested an improved method of estimating the fractal dimension of a texture. They also introduced new features based on lacunarity which capture the second order statistics of fractal surfaces.

2.3 Structural Texture Representations

On the structural level, a texture can be viewed as being made up of primitives which appear in near regular repetitive spatial arrangements. To describe this form of texture, one has to describe both the primitive and placement rules [54]. This can be accomplished using the following steps:

- Texture primitive element definition and extraction.
Chapter 2: A Review of Texture Analysis Techniques

- Primitive description and characterisation.
- Primitive placement rule description and characterisation.

Here the primitive or basic textural element (i.e. texel) is an important concept and can be defined as a set of pixels (usually connected) satisfying certain predicates. The grey level value of a single pixel is the simplest primitive. Structural texture analysis approaches found in the literature revolve around the three steps listed above. Their differences lie mainly in their emphasis on one or two of these steps or in the sequence in which these steps are followed. For simplicity, structural texture analysis approaches can be grouped into three classes, namely Strong Texture Analysis, Weak Texture Analysis, and Syntactic Texture Analysis.

2.3.1 Strong Texture Analysis

Strong structural textures are those which have strong or non-random spatial interaction of the primitives. To be adequately represented, all three steps mentioned above are needed in the analysis of a strong texture. However, the order at which this is accomplished is not important. Two common approaches are usually employed. They are the top-down approach and the bottom-up approach. In the top-down approach, the placement rules of the primitives are first computed and the spatial grid structure of texture is derived. Primitive extraction is then subsequently guided. For instance, Conner and Harlow [12] specified the spatial grid structure by two periodic vectors which are chosen as the displacement vectors of the grey level co-occurrence matrices with the smallest moment of inertia. Matsuyama et al. [47] used the energy distribution in the Fourier power spectrum to extract two similar periodicity vectors. Vilnrotter et al. [67] presented another example of top-down approach for strong structural texture. First, they generated an edge map by applying an intensity edge detector to the input texture. From this edge map, they calculated the Edge Repetition Arrays (ERAs) which are similar to the grey level co-occurrence matrices. From some selected ERAs, they extracted the spatial grid structure and placement rules before the corresponding primitives were extracted.

In the bottom-up approach, primitive extraction, primitive description, and primitive interaction description are applied in that order. Matsuyama et al. [46] extracted texture primitives by means of cluster analysis and then computed the regularity vectors which specify the displacement rules of the primitives. Tomita et al. [60] follow a similar methodology. They defined primitives to be image regions having the same grey level intensities. This primitive extraction is equivalent to grey level image segmentation which is achieved via a simple merging technique. After
primitives extraction, they performed shape analysis on these primitives and calculated primitive attributes such as brightness, area, size, elongatedness, and curvature. Based on attribute histograms, they classified primitives into one ground class and several figure classes by a clustering technique. Finally, placement rules were extracted based on some second-order statistics of typical primitives from each class.

In passing, two remarks on strong texture analysis approaches are in order. Usually, primitive extraction is a very difficult process in bottom-up approaches, sensitive to various degradations of images as noted in [47]. On the other hand, primitive extraction could be a very trivial task in top-down approaches, but these approaches may fail completely if the underlying spatial structures are very complex.

2.3.2 Weak Texture Analysis

As the name implies, weak textures consist of primitives that are weakly coupled or randomly placed and therefore the underlying placement rules convey little textural information. As such, some simple statistics such as histograms and co-occurrence matrices of primitives may be sufficient for weak texture discrimination. Therefore, a weak texture analysis approach may only involve primitive extraction and description. Hong et al. [30] defined primitives as dark regions bounded by connected edges on a bright background. They first applied 8 local edge operators to the input texture sample corresponding to the 8 directions in a 3 × 3 window. The resultant edge maps were "cleaned" by grouping appropriate edge points into connected boundaries. After primitive extraction (i.e. extraction of dark regions), they estimated some primitive attributes and form a set of attribute histograms. The means and standard deviations of these histograms were taken as texture measures for texture discrimination. Wang et al. [70] defined primitives as image pixels having intensities above or below some threshold. They described three different ways of extracting primitives. Statistics of primitive attributes were then used as texture features.

In addition to texture classification and discrimination, weak texture analysis has also found many applications in texture segmentation. An example is given in Tsuji and Tomita [62]. They defined primitives to be image regions of constant intensities which were called atomic regions. Since different textures have distinct primitive attributes, they thresholded the attribute histograms to segment the input image into uniform texture regions.
2.3.3 Syntactic Texture Analysis

Zucker [75] considered a real texture as being 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 fashion. Each node corresponds to a cell in a tessellation of the plane. He believes that transformation rules can be defined from the representation of an ideal texture to that of a natural texture. Zucker's model has enjoyed a good implementation in the work by Lu and Fu [42]. They divide a texture sample into non-overlapping windows of size $9 \times 9$, and the pixel intensity relations within each window are represented with some fixed tree structure. Based on these tree structures, they derive a set of production rules which define the three grammars for the underlying textures. These grammars can then be used for texture synthesis as well as texture discrimination. In order to model more realistic textures, they subsequently introduce some distortion on the production rules. More details on syntactic methods can be found in Fu [22].

2.4 Summary

In this Chapter, a review of major techniques for texture analysis and classification was presented. Owing to space limitation of the thesis, this review is by no means exhaustive. Only techniques that are known to be contributing significantly to this area of research were mentioned. Similarities between techniques were highlighted together with the merits and demerits that are associated with them. Interested readers are referred to the texture analysis literature for more details. The reviewed techniques were broadly classified into two major categories, namely Statistical Approaches and Structural Approaches.

From the information content point of view, the grey level co-occurrence matrix, grey level difference, Fourier power spectrum, and autocorrelation methods which are all based on second-order statistics are interrelated [61]. Co-occurrence matrices for all displacements $\delta$, represent all second-order statistics between any two pixels in an image. A co-occurrence matrix for a given $\delta$ is a subset of the second-order statistics. The difference statistics for the same $\delta$ can be derived from the co-occurrence matrix. The Fourier power spectrum and the autocorrelation function are the same subsets of the second-order statistics. Each of them can be mathematically derived from one another. Higher-order statistics do not necessarily give better results in texture discrimination than the lower-order statistics. One reason is that they do not convey as much important infor-
mation as lower-order statistics. This has been confirmed psycho-physically by Julesz [32]. The performance of these second-order statistics methods can be evaluated by testing the correctness of classification of textures. One such study was performed by Weszka and Rosenfeld [71] [72]. The task was to classify terrains in aerial photographs. They concluded that the performance of the grey level co-occurrence matrix method and grey level difference method is quite comparable. In another study, Conner and Harlow [11] concluded that the co-occurrence matrix method is more powerful than the higher-order statistics method in the form of the grey level run length. The co-occurrence matrix technique is the most popular and has been a benchmark which any new technique should be compared with. The major drawback of this technique is its computational complexity and the choice of an appropriate displacement vector, δ. The computational complexity can be reduced when the grey level difference method is used instead. In that respect, the grey level difference method would be a preferred candidate.

Because of its origin from digital signal processing, the multi-channel filtering method provides a sound theoretical framework for the analysis of texture by means of decomposition of the image into filtered components. The attractiveness of the Gabor channel filtering approach stems from the belief that the early processing stages embedded in the human vision system involve a set of parallel and quasi-independent mechanisms or channels which are modelled as spatially tuned narrow-band filters. Unser [65] provides a general formulation framework known as the local linear transform which multi-channel methods developed by Laws (Laws' masks) [41], Ade (eigen filters) [1], and others (i.e. Discrete Cosine Transform filters) [65] could be grouped under. The eigen filter approach suffers from one main problem, i.e. the filters are very dependent on the texture for which the filters are derived. In general texture analysis problems, this approach is unsuitable as many textures may be involved which could lead to the difficult problem of selecting a general set of filters to be used. The general class of sub-optimal transforms introduced by Unser [65] helps to overcome this problem. The comparative study conducted by Ng et al. [49] revealed some interesting findings about the Gabor channel filtering approach and the local linear transform approach. They showed that the orthogonal masks derived from the Discrete Cosine Transform have some interesting features that resemble Gabor filters; the two approaches have similar structural representation and only the parameters of the representation structure differ. They then assessed the performance of these approaches on a segmentation problem. They concluded that features derived from the DCT masks gave consistently better segmentation results than the energy features of Gabor filters. The reason for the poorer performance of the Gabor
filters is that these filters are particularly sensitive to textures which have high concentration of localised spatial frequency. The class of textures tested by them are mainly stochastic in nature, exhibiting no dominant frequency or orientation, hence the poor performance of the Gabor filters.

The lack of a comprehensive understanding of how the later stages of the human vision system process textural tokens (i.e. edges, peaks and valleys, and spots or blobs) and combine them at the interpretation stage has impeded further development of the Local Geometrical Feature analysis method. Only when more is known, can these methods be developed to their potential.

The descriptor base methods are difficult to realise as there is a lack of accurate definition of texture properties as perceived by humans. More work has to be done to determine the precise descriptions of some textural measures that bear a close correlation between the observed and measured descriptions [58]. Probably due to its heuristic nature, this approach has yet to find a widespread use in texture analysis applications.

Model-based approaches involve determining the relationship a pixel has with its neighbouring pixels using a mathematically derived model. Most of the models were originally developed in the texture synthesis field. These methods are in general computationally expensive and their application may be hampered by the difficulty in choosing a suitable neighbourhood system or in determining the order of the model. Fractal analysis methods for texture representation are only of recent development. A lot still has to be done to demonstrate their usefulness as a model for representing textures of varying composition, especially textures that are non fractal themselves.

Structural approaches, on the other hand, are less applicable to real world textures. They require textures to be composed of primitives which are easily identifiable and whose interactions are inferable in some cases. The application of structural approaches has been very limited as textures in reality rarely fit the requirements necessary for the implementation of a structural approach. For stochastic textures, these approaches would be completely useless. In addition, primitive and placement rule extraction and description is usually very complex and hence time consuming. Some structural approaches, i.e. strong texture analysis approaches and syntactic texture analysis approaches, are capable of texture synthesis as well but the synthesized textures are in general monotonous [68].

It should be noted at this stage that the evaluation of texture analysis techniques carried out in the literature should not be taken too literally. It should be appreciated that such results depend on the samples which were used and on the setting of the experiments. They give only an assessment of relative performance. Another important factor for evaluation is whether the method can
be used not only for texture classification but also for texture segmentation. In texture segmentation, texture properties are measured locally in a neighbourhood at each point in the image. The problem lies in the size of the neighbourhood. A method which needs a large neighbourhood to obtain stable values is not recommended because the resolution of the segmentation becomes low. For example, the Fourier Power Spectrum method and Fractal Analysis method may not be good in this sense.

References


Chapter 3

Texture Measurement by Local Linear Transform

3.1 Introduction

Texture characterisation can be seen as a local measurement describing the relationship between a set of neighbouring pixels. This definition has found solid support in the texton theory postulated by Julesz [18]. His theory supports the finding that preattentive texture perception and discrimination is based on some local conspicuous image properties or features that are best computed by means of local operations [19]. This led to the widespread development of many local texture feature measurement approaches [12][15], some of which were briefly mentioned in the preceding chapter.

A fruitful approach, which has been studied by many authors, is to extract local neighbourhood information by means of linear filtering operators [1][2][9][13][20][26][33]. Therefore, the local texture measures of a region can be computed from the energy measures of the output of a bank of filters. The design of the filters is unique to each of the different approaches. Laws recommended the use of local detectors of structural forms like “flats”, “spots”, or “edges” [20] [26]. Directional filters which attempt to model early vision mechanisms of the human vision system were used by Faugeras [9], Granlund [13], and Wermser and Liedtke [33]. Filter design using principal component analysis of the underlying texture fields (eigen filters) were used by Ade [1] [2]. Most of these approaches can be included in a general framework suggested by Unser [28] known as Local Linear Transformation. In his formulation, a bank of local convolution filters can be viewed as a set of basis operators stacked to form a transformation matrix. The outputs from this bank of filters have the equivalent structure as the transformation coefficients which are obtained
Chapter 3: Texture Measurement by Local Linear Transform

by matrix multiplication for all local neighbourhoods in a sliding window fashion. The choice of optimal or sub-optimal linear operators depends on the criteria used for the formulation of the transformation process.

The use of local linear transforms in texture measurements has several advantages:

• These methods are almost as efficient for texture discrimination as methods based on co-occurrence measurements [28] which are considered by many as a benchmark and is known to be computationally very involved.

• They enable a more compact description of local texture properties.

• The feature extraction process is computationally less demanding.

• These features are well suited for texture segmentation based on clustering using multi-resolution techniques [6] [29] [30].

• Because of their parallel structure, these methods could be implemented on parallel machines.

In this Chapter the local linear transform for texture characterisation will be investigated. The essence of this approach is to transform the local neighbourhood characterised by the \( N^\text{th} \)-order probability density function into a series of \( N \) first-order probability density functions. This \( N^\text{th} \)-order density function can be modelled by the mean vector and covariance matrix associated with the pixels of the neighbourhood. Because of stationarity, the diagonal elements of the original covariance matrix are all the same and are of no use for texture analysis. The optimum Karhunen-Loeve transform will fully diagonalise the covariance matrix making the variances as dissimilar as possible. Alternative sub-optimal transforms can be used for a more general implementation. Section 3.2 defines the local neighbourhood and presents a general model for the covariance structure associated with the pixels in the neighbourhood. The need for and the types of neighbourhood transformation will be detailed in Section 3.3. Section 3.4 introduces the texture measures associated with the statistical descriptions of the \( N \) first-order histograms. An experiment involving the classification of a set of Brodatz images will be carried out to test the capability of these texture measures. These measures will then be applied to a defect detection exercise involving textile materials. The experimental results will be presented in Section 3.5. The Gabor filtering approach for texture analysis will be introduced in Section 3.6. It will be shown in Section 3.7 that both the feature representation of the local linear transform and the Gabor filtering approach share
3.2: Local Neighbourhood Vector

A discrete texture image of size $K \times L$ can be considered as the realisation of a bi-dimensional stationary and ergodic stochastic process. Let us denote the image by a 2-dimensional representation of the form $\{x_{k,l}\} (k = 0, ..., K - 1, l = 0, ..., L - 1)$. Since texture is a neighbourhood property, it is practical to consider the pixels in a given neighbourhood centered at the pixel indexed by $(k,l)$ as the components of a local feature vector. For simplicity and ease of implementation, the neighbourhood is defined as a rectangular $N_x \times N_y$ domain centred on $(k,l)$. The $N$ (where $N = N_x \times N_y$) components of the local neighbourhood vector $x_{k,l} = [x_1, x_2, ..., x_N]^T$ are the scan-type ordered pixels belonging to the $N_x \times N_y$ neighbourhood. This formulation transforms the original grey level image $\{x_{k,l}\}$ into a multi-variate 1-dimensional sequence as illustrated clearly in Figure 3.1.

Within this framework, most common texture analysis methods may be considered as estimating the statistics associated with the local neighbourhoods belonging to a homogeneous region of the texture. For example, the grey level histogram provides an estimate of the first-order probability density function of a component of the neighbourhood vector. This information is useless for texture analysis because the stationary nature of the texture generating process makes these first-order histograms identical. Estimates of the second-order probability density functions associated with different pairs of components of the neighbourhood are the well known co-occurrence matrices [14]. Owing to stationarity, the total number of such matrices to fully describe the second-order characteristic of the texture reduces to $2N_xN_y - N_x - N_y$. Gagalowicz [11] has demonstrated that second-order statistics can be used to measure all visually useful textural information. For practical reasons, higher-order probability density functions can only be described with a para-
metric model. The set of second moments can be used to construct the spatial covariance matrix which is defined as

\[ C_x = E \left\{ (x_{ij} - E(x_{ij})) \cdot (x_{ij} - E(x_{ij}))^T \right\} \] (3.1)

and which provides a sufficient statistic for multi-variate Gaussian model. Because of stationarity, the components of the local neighbourhood mean vector are all the same and therefore the covariance matrix exhibits a close-to-Toeplitz structure with only \( 2N_xN_y - N_x - N_y + 1 \) different entries.

For example, let us consider a square neighbourhood of \( 3 \times 3 \) (i.e. \( N_x = N_y = 3 \)) pixels. It admits \( 2 \times 3 \times 3 - 3 - 3 + 1 = 13 \) essentially different spatial relationships between two pixels, counting opposite directions only once and including the null distance. Figure 3.2 gives a pictorial representation of these possibilities. Now if the number of samples used to generate these statistics is large, and under the assumption that the process which generated the texture is stationary, the covariance matrix would take the form of Equation (3.2). The covariance matrix has
Figure 3.2: The different spatial relationships between two pixels in a 3 x 3 neighbourhood.

to be understood symmetrically completed.

\[
C_x = \begin{pmatrix}
c_a & c_b & c_f & c_c & c_d & c_k & c_h & c_m & c_g \\
c_b & c_a & c_e & c_c & c_d & c_l & c_i & c_m & c_g \\
c_f & c_e & c_a & c_c & c_d & c_k & c_l & c_m & c_g \\
c_c & c_c & c_c & c_a & c_d & c_k & c_l & c_m & c_g \\
c_d & c_d & c_d & c_c & c_a & c_k & c_l & c_m & c_g \\
c_k & c_k & c_k & c_k & c_k & c_l & c_l & c_m & c_g \\
c_h & c_h & c_h & c_h & c_h & c_h & c_h & c_m & c_g \\
c_m & c_m & c_m & c_m & c_m & c_m & c_m & c_m & c_g \\
c_g & c_g & c_g & c_g & c_g & c_g & c_g & c_g & c_g \\
\end{pmatrix}
\]

(3.2)

where \( c_p \) represents the covariance of pixels in spatial relationship \( p \). These variances-covariances are estimated over a region in the image the texture of which one wishes to characterise.

3.3 Local Linear Transform

A close look at the covariance matrix of Equation (3.2) reveals that the main diagonal elements (i.e. \( c_a \)) of the matrix are identical. This is uninteresting for texture analysis and classification as first-order statistics of the original local neighbourhood vector \( x_{ij} \) are similar and hence do not convey any useful texture information. This is further exemplified by the fact that the grey level of an image can be non-linearly modified to fit any specified histogram, and yet would not change drastically the local texture property. An interesting alternative is obtained from the linear
transformation of the local neighbourhood vector. A local linear transform is defined by

$$y_{kj} = T_N \cdot x_{kj}$$  \hspace{1cm} (3.3)$$

where $T_N = [t_1, ..., t_N]^T$ is a nonsingular $N \times N$ square matrix with $t_i$ representing the transformation vectors. This Equation may lead to two different forms of interpretation. First of all, the local linear transform can be seen as an $N$ channel correlation (or convolution) by a set of $N$ filters whose spatial coefficients could be found along the $N$ columns of the transformation matrix $T_N$. The corresponding system is shown in the first stage of Figure 3.3; it is equivalent to a bank of $N$ finite impulse response filters. Every channel will extract a particular aspect of local texture property. Secondly, Equation (3.3) can be thought of as a transformation of the original co-ordinate space $x_{kj}$ to another space $y_{kj}$. The efficiency of this analysis method will depend on the choice of the transform matrix $T_N$.

### 3.3.1 Optimal Transformation

The underlying principle is to characterise the $N^{th}$-order probability density function of the pixels in a restricted neighbourhood by a set of $N$ first-order probability density functions estimated along suitably chosen axes. Given a transformation $T_N$, the covariance matrix $C_y$ of the transformed
data is given by

\[ C_y = T_N \cdot C_x \cdot T_N^T \]  \hspace{1cm} (3.4)

The main diagonal elements (i.e. variances) of the covariance matrix \( C_y \), should ideally be as distinct as possible. In other words, the aim is to choose a suitable transform matrix \( T_N \) so that the statistics of the transformed channel histograms like the variances provide the most distinctive texture description. Hence one should choose the transform which produces first-order statistics as “distinct” as possible [28].

The set of all channel variances provides a complete description of the covariance structure of a given texture field when \( T_N \) fully diagonalises \( C_x \). In this situation, \( T_N = [u_1, \ldots, u_N]^T \), where the \( u_i \)'s are the solution vectors to the characteristic equation:

\[ C_x \cdot u_i = \lambda_i \cdot u_i \]  \hspace{1cm} (3.5)

The optimal basis vectors \( (u_i, \ i = 1, \ldots, N) \) are the eigenvectors of the spatial covariance matrix \( C_x \) and the variances are the corresponding eigenvalues \( (\lambda_i, \ i = 1, \ldots, N) \). The optimal transform is known as the Karhunen-Loeve transform (KLT). Ade used this form of principal component analysis method to generate “eigen filters” for texture analysis [1].

Uncorrelatedness is a necessary condition, but not always sufficient for independence. Therefore in general, the \( N^{th} \)-order probability density function of the local neighbourhood vector can only be approximated by a product of \( N \) first-order probability density functions obtained from this optimal transformation. This approximation will be exact in the case of a multi-variate Gaussian distributed texture field.

### 3.3.2 Sub-Optimal Transformations

The optimal transformation of the local neighbourhood vector described in the preceding section is primarily of theoretical value. The transformation matrices are unique to the underlying textures to be analysed or classified and are usually cumbersome to determine: they require the estimation of spatial covariance matrices and the use of standard, computationally expensive, numerical eigenvector extraction method. The optimal transformation had been shown by Ade [1] to work in situations where the structure of a texture one wishes to analyse is known a priori. He used this approach to detect defects on known material surfaces by using a simple classification routine to isolate “out-of-class” regions (i.e. defects) on a test image. In the texture analysis problem, the optimal solution given by the KLT will be generally different from one texture field to
another. This approach is therefore difficult to apply in practical applications which involve the analysis of a large class of textures.

By approximating the close-to-Toeplitz structure of the covariance matrix of Equation (3.2) to a Toeplitz matrix, Unser [27] demonstrated that close-to-optimal performance can be achieved for a wide sense stationary process by using a sub-optimal separable transform, namely Discrete Cosine (DCT), Sine (DST) [16][17], Real Even and Odd Fourier (DREFT and DROFT) [27] transforms. A direct consequence is that these transforms will approximately diagonalise the spatial covariance matrix of a very large class of textures. Further, separable transforms are standard transforms which do not require both the estimation of the covariance matrix and the eigenvectors extraction process; a computationally more attractive alternative. Separable transforms can be computed by successive filtering along the rows and columns. It is shown in [34] that 2-dimensional spatial filtering may be carried out more simply and efficiently by factorising the filters and performing a sequence of 1-dimensional filtering instead. As an example, a 1-dimensional Discrete Cosine basis vector $u_m$ can be expressed as [27]

$$u_m(k) = \begin{cases} \frac{1}{\sqrt{N}} & \text{for } m = 1, k = 1, ..., N_x \\ \sqrt{\frac{2}{N_x}} \cos \left( \frac{2\pi}{N_x} (m-1) k \right) & \text{for } m = 2, ..., N_x, \ k = 1, ..., N_x \end{cases}$$

(3.6)

These 1-dimensional DCT vectors of length $N_x$ can be used to successively filter the image in the two directions or they can be used to generate 2-dimensional transform filters appropriate for images. The latter can be accomplished by simply multiplying the column basis vectors with the row vectors of identical length to produce a set of 2-dimensional filters of $N = N_x^2$ (i.e. $N_x = N_y$) entities. The $i^{th}$ entity in this filter bank representation, $d_i(\cdot, \cdot)$, is given as

$$d_i(\cdot, \cdot) = u_m(\cdot) \cdot u_n(\cdot)$$

(3.7)

where $i = m + (n-1)N_x$ for $1 \leq m, n \leq N_x$. This set of filter bank operation can be seen as the first stage of processing in the block diagram of Figure 3.3. The output from channel $i$ is given as

$$y_i(k, l) = \sum_{a=0}^{N_x-1} \sum_{b=0}^{N_y-1} d_i(a+1, b+1) x(k-a, l-b)$$

(3.8)

where $x(k, l)$ is a real value image function.

The energy preserving property of these sub-optimal transforms is suitable for a wide range of applications including image compression and texture analysis as no information is lost during transformation. This can be explained in the Fourier domain by analysing the Fourier power
3.4 Texture Feature Measurements

Figure 3.4: Fourier power spectrum of a set of 3 x 3 DCT filters.

spectrum of each filter generated by the sub-optimal transforms. The Fourier power spectra of the set of nine 3 x 3 DCT filters are shown in Figure 3.4. It can be seen that these filters occupy non-overlapping positions in the frequency spectrum domain suggesting that there is no duplication of information between adjacent filters. It is also interesting to note that these filters are wide band and that their contributions sum up to a constant at all spatial frequencies. This last point reflects the energy preserving property of the sub-optimal transform.

3.4 Texture Feature Measurements

At this point, it is useful to recall that all filtered channels have identical mean values and that distinct texture regions are presumed to differ mainly in the values of the channel variances. Since most classification and clustering algorithms are designed to distinguish between classes with differences in mean (or centroids), the purpose of the proceeding stage is to translate the differences in dispersion (i.e. variance) characteristics into differences in mean value. This is the purpose of the non-linear transformation that is applied after the initial filtering of Figure 3.3. The channel
variance can be defined as

\[ \sigma^2 = \text{Var}\{y_i\} = E[(y_i - E[y_i])^2] \]  

(3.9)

and provides a sufficient description of the texture. The set of variance measurements characterises the energy distribution at the output of the filter bank. The normalised and centred \( p \)th moments defined by

\[ \mu_p^{\text{nc}} = E[(y_i - E[y_i])^p]/\sigma_p^p \]  

(3.10)

can be used as additional measurements of the texture. Of particular interest are the third (skewness) and fourth (kurtosis) moments which respectively provide useful measures of the amount of skewness and peakedness in the channel distribution. Likewise, if the input image is of zero mean, which can be achieved by pre-processing, the following group of non-linear transform functions \( f(y_i) \) will be suitable: the power function of the form \( |y_i|^\alpha \) where \( \alpha \) is some positive integer \([30]\), or the logarithm function \( \log(y_i) \). For computational reasons the former function with \( \alpha = 1 \) is a preferable choice and has been shown by Laws [20] and Pietikainen et al. [26] to perform reasonably well.

As can be seen in Figure 3.3, the operation that follows the non-linear transformation is the smoothing operation. The purpose of the smoothing operator is to decrease the variances of the feature vectors within the various texture classes, while preserving the inter-class mean differences insofar as possible. Smoothing is usually achieved by linear filtering. This low pass filtering operation can be viewed as an estimation window that is applied to the signal and that provides a local estimate for \( y_i \). The choice of the low pass filter depends a great deal on the implementation. For example, in texture classification, a suitable choice will be a filter with equal weights as every pixel within the window must be treated with equal importance. This correspond to an averaging operation. However in texture segmentation, the choice of a Gaussian filter is particularly attractive because of its effective band rejection and good localisation of texture boundaries. Denoting the variance estimation of Equation (3.9) by a general non-linear operation of \( f(y_i) \), the smoothing operation can be expressed as

\[ \nu_{\text{def}}(k, l) = \frac{1}{W^2} \sum_{p=0}^{W-1} \sum_{q=0}^{W-1} f(y_i(k - p, l - q)) \]  

(3.11)

where a square region (macro-window), centred at \((k, l)\) of size \( W \times W \) is to be smoothed by averaging.
3.5 Experiments

3.5.1 Texture Classification

In this section the capability of the sub-optimal local linear transform of Section 3.3.2 will be tested. Twelve texture images taken from the Brodatz Album [5] were used in the following classification exercise. The images are shown in Figure 3.5. As prototype images are necessary for a supervised classification scheme, each 256 x 256 pixel image was divided into two halves. The prototype image resides in the top half of each image. A collection of local texture measurements estimated within this homogeneous texture region enables the computation of the class statistics $(\mu_i, C_i)$, where $\mu_i$ and $C_i$ are the mean vector and covariance matrix of the multi-dimensional feature of the $i$th texture class. The remaining bottom half of the image was used to generate samples for test (classification) purposes. The macro-windows used here were of variable sizes: 64 x 64, 32 x 32, and 16 x 16 pixels. Samples were gathered from the image by sliding the window with a one-third overlap. This produced 40, 230, and 1104 samples for the respective window sizes of 64 x 64, 32 x 32, and 16 x 16 pixels for each stage.

An optimal Bayesian decision rule was used to classify the texture samples, based on their measured feature values. The class conditional probability density functions of the $D$-dimensional feature vector $z$ are assumed to be multivariate Gaussian distributions with mean vectors and covariance matrices: $(\mu_i, C_i)$ $(i = 1, ..., 12)$. Under such an assumption, the Bayes classifier, which minimises the total probability of error, dictates that a texture sample with feature vector $z$ is assigned to class $j$ satisfying

$$J_j(z) = \min_i J_i(z) \quad \text{where}$$

$$J_i(z) = (z - \mu_i)^T \cdot C_i^{-1} \cdot (z - \mu_i) + \log(|C_i|) \quad (i = 1, ..., 12)$$

In this experiment, the first stage filtering operation was accomplished by a bank of sub-optimal transform filters derived from the Discrete Cosine (DCT) [3], Discrete Sine (DST) [16][17], Discrete Even Sine (DEST), Discrete Real Even Fourier (DREFT), and Discrete Real Odd Fourier Transform (DROFT) [27]. The filters used were of three sizes, namely $3 \times 3$, $4 \times 4$, and $5 \times 5$. The filter with low pass property was excluded from the filter bank as it does not convey any useful texture information. Texture measurements extracted were the variances of the outputs of the different channels. Local smoothing was accomplished by an equal weighted low pass filter. The five texture feature sets resulting from this arrangement are abbreviated as: LVDCT, LVDST,
Figure 3.5: Set of 12 Brodatz textures used for classification.
LVDEST, LVDREFT, LVDROFT. The "LV" preceding each sub-optimal transform signifies that the local variance of the channel outputs is used as a texture measure.

### 3.5.2 Results and Discussion

The classification results involving twelve classes of Brodatz textures using the set of texture measures are depicted in Table 3.1 to Table 3.3. DROFT(3 × 3) and DREFT(3 × 3) are not included in Table 3.1 as they are similar to DEST(3 × 3) and DCT(3 × 3) respectively. As can be expected, the classification becomes more accurate as the size of the texture sample increases. Two interesting observations can be made from these results. Firstly it can be observed that for a filter of a given size, the performances of the various texture measures are very similar. Secondly, no significant rise in classification rate was observed as more features were used. If one compares the best performing feature set of the 3 × 3 filter transform (LVDEST-94.27%) with that of the 5 × 5 filter transform (LVDROFT-94.38%), it is noted that these values are almost the same.

The first observation is in close agreement with Pietikainen et al. [26] who found that the performance of these filters does not depend much on the numerical values of the filters' coefficients but rather on the general forms of these filters (edge-like, spot-like, etc.). A closer examination of the transform filters used in this exercise confirmed this to be true. These sets of transform filters share very similar structural forms.

The second observation indicates that the use of smaller filter sizes for micro-texture characterisation is quite adequate. Smaller filters provide a coarser partitioning of the frequency domain (i.e. these filters are wide-band) suggesting perhaps counter intuitively, that multi-channel texture representation does not necessarily require a fine partitioning of the spatial-frequency domain. This further suggests that texture description can be quite compact, resulting in a small feature dimension space.

Next, the classification exercise was performed again but this time with the inclusion of the sets of third and fourth normalised moments of Equation (3.10). Since maintaining the size of the feature dimension to a reasonably small number is invariably the objective of any pattern recognition system design, only transform filters of size 3 × 3 were used. This resulted in 24 features being produced for the classification. The results of classifying the same test images using this extended set of features are shown in Table 3.4. It can be seen that features derived from the higher moment descriptions of the channel histograms provide a slight improvement of the classification results. This is of course achieved at a cost; an increase in the size of the feature
### Table 3.1: Classification results expressed in percentage of successful classification for the sub-optimal transform filters of size $3 \times 3$. $M$ is the number of features.*

<table>
<thead>
<tr>
<th>Transform</th>
<th>Sample Size</th>
<th>M</th>
<th>64 x 64</th>
<th>32 x 32</th>
<th>16 x 16</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>LVDEST (3 x 3)</td>
<td>8</td>
<td>99.58%</td>
<td>96.81%</td>
<td>86.43%</td>
<td>94.27%</td>
<td></td>
</tr>
<tr>
<td>LVDCT (3 x 3)</td>
<td>8</td>
<td>98.75%</td>
<td>96.45%</td>
<td>85.16%</td>
<td>93.45%</td>
<td></td>
</tr>
<tr>
<td>LVDST (3 x 3)</td>
<td>8</td>
<td>98.13%</td>
<td>94.78%</td>
<td>85.97%</td>
<td>92.96%</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3.2: Classification results expressed in percentage of successful classification for the sub-optimal transform filters of size $4 \times 4$. $M$ is the number of features.*

<table>
<thead>
<tr>
<th>Transform</th>
<th>Sample Size</th>
<th>M</th>
<th>64 x 64</th>
<th>32 x 32</th>
<th>16 x 16</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>LVDEST (4 x 4)</td>
<td>15</td>
<td>99.79%</td>
<td>96.38%</td>
<td>87.36%</td>
<td>94.51%</td>
<td></td>
</tr>
<tr>
<td>LVDROFT (4 x 4)</td>
<td>15</td>
<td>96.67%</td>
<td>96.45%</td>
<td>88.55%</td>
<td>93.89%</td>
<td></td>
</tr>
<tr>
<td>LVDREFT (4 x 4)</td>
<td>15</td>
<td>97.50%</td>
<td>96.88%</td>
<td>86.71%</td>
<td>93.70%</td>
<td></td>
</tr>
<tr>
<td>LVDCT (4 x 4)</td>
<td>15</td>
<td>97.29%</td>
<td>96.41%</td>
<td>87.16%</td>
<td>93.62%</td>
<td></td>
</tr>
<tr>
<td>LVDST (4 x 4)</td>
<td>15</td>
<td>98.33%</td>
<td>94.42%</td>
<td>86.81%</td>
<td>93.19%</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3.3: Classification results expressed in percentage of successful classification for the sub-optimal transform filters of size $5 \times 5$. $M$ is the number of features.*

<table>
<thead>
<tr>
<th>Transform</th>
<th>Sample Size</th>
<th>M</th>
<th>64 x 64</th>
<th>32 x 32</th>
<th>16 x 16</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>LVDROFT (5 x 5)</td>
<td>24</td>
<td>97.08%</td>
<td>96.12%</td>
<td>89.95%</td>
<td>94.38%</td>
<td></td>
</tr>
<tr>
<td>LVDEST (5 x 5)</td>
<td>24</td>
<td>98.33%</td>
<td>95.51%</td>
<td>89.00%</td>
<td>94.28%</td>
<td></td>
</tr>
<tr>
<td>LVDCT (5 x 5)</td>
<td>24</td>
<td>96.46%</td>
<td>95.22%</td>
<td>88.61%</td>
<td>93.43%</td>
<td></td>
</tr>
<tr>
<td>LVST (5 x 5)</td>
<td>24</td>
<td>97.67%</td>
<td>94.28%</td>
<td>88.29%</td>
<td>93.08%</td>
<td></td>
</tr>
<tr>
<td>LVDREFT (5 x 5)</td>
<td>24</td>
<td>94.58%</td>
<td>96.01%</td>
<td>87.96%</td>
<td>92.85%</td>
<td></td>
</tr>
</tbody>
</table>

*Samples used here are slightly correlated as they are obtained for partially overlapping windows. The total number of samples used depends on the window size. Accordingly there are 40, 230, and 1104 samples for the respective window sizes of $64 \times 64$, $32 \times 32$, and $16 \times 16$ pixels.
### Table 3.4: Classification results expressed in percentage of successful classification for the sub-optimal transform filters of size $3 \times 3$ using an extended feature set. $M$ is the number of features.

<table>
<thead>
<tr>
<th>Transform</th>
<th>Sample Size</th>
<th></th>
<th></th>
<th></th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>64 x 64</td>
<td>32 x 32</td>
<td>16 x 16</td>
<td></td>
</tr>
<tr>
<td>Extended DCT $(3 \times 3)$</td>
<td>24</td>
<td>100%</td>
<td>97.25%</td>
<td>87.89%</td>
<td>95.05%</td>
</tr>
<tr>
<td>Extended DEST $(3 \times 3)$</td>
<td>24</td>
<td>99.17%</td>
<td>97.08%</td>
<td>87.22%</td>
<td>94.49%</td>
</tr>
<tr>
<td>Extended DST $(3 \times 3)$</td>
<td>24</td>
<td>98.75%</td>
<td>94.49%</td>
<td>86.91%</td>
<td>93.38%</td>
</tr>
</tbody>
</table>

This result indicates that the textures used in these experiments are non-Gaussian and that, in such a case, there is an advantage in using higher-order statistical information to improve the results in classification.

It is worth noting at this point that the manner in which the frequency domain of the DCT filters is partitioned will introduce inherent ambiguity for certain type of textures. A closer examination of the filters' characteristics shown in Figure 3.4 reveals that some of the responses have four spectral peaks in the spatial-frequency domain. This has the effect of making these filters incapable of distinguishing between textures of the following nature:

(a). Medium to high frequency textures, oriented approximately along 45° and 135° to the horizontal.

(b). High frequency textures, oriented approximately along 18° and 162° to the horizontal.

(c). High frequency textures, oriented approximately along 72° and 108° to the horizontal.

However, this is not a serious problem in many texture analysis applications as cases like these do not occur very often. If sensitivity of texture orientations is of the essence, then the use of orientation sensitive filters [10][32] would have to be considered. This would be achieved at the expense of increasing the total number of filters needed to approximately cover the entire frequency spectrum.

The set of eight LVDCT$(3 \times 3)$ texture measurements will be used for texture representation in the remaining part of this Chapter. This is a compromise between achieving a good texture description and keeping the feature dimension as low as possible. The choice of the Discrete Cosine transform is favoured here as there exist numerous methods for its fast computation [21][22][23][24][25][31].
3.5.3 Application: Defect Detection in Textiles

As far as the previous section is concerned, the results obtained on the Brodatz texture classification are only of theoretical interest. It is now pertinent to investigate the performance of the local linear transform features on a real image analysis problem. A set of textile images will serve as a vehicle for testing the applicability of the set of eight 3 x 3 LVDCT features to the problem of defect detection. The images have resolution of 256 x 256 pixels. Generally, one quarter of each non-defective area of each textile image is used for training to estimate the class mean and covariance matrix of the assumed Gaussian model. Texture defects are considered as outliers of the model. They are identified as those feature vectors exceeding a threshold on the Mahalanobis distance $\sqrt{8}$. The results of processing are shown in Figure 3.6. It can be seen that the LVDCT features work well on a wide range of regular textures with different defects and also they are able to detect reliably extended regional defects. These tests have shown that the local linear transform features can be quite effective for texture image analysis applications.

3.6 Texture Representation via Gabor Filtering

Recent research suggests that the early processing stages of the human vision system involve a set of parallel and quasi-independent mechanisms or channels which can be modelled as narrowband filters. The incoming pictorial information is then decomposed, very much like the DCT filtering, by this set of channels into a sequence of filtered images, each of which contains intensity variations over a narrow range of frequency and orientation characterising regularity, coarseness, and directionality of the texture in the image. How a human brain combines the channel outputs to make visual judgements, such as recognition or discrimination, is still unclear. Nevertheless, the multi-channel representation model is widely accepted as a basis for texture description.

The findings also show that the channel filter responses have optimal joint localisation in both the spatial and spatial-frequency domains and can be mathematically approximated by Gabor functions [7]. Hence the Gabor receptive field model can be used for texture representation.

A complex 2-dimensional Gabor function can be interpreted as a 2-dimensional elliptical Gaussian modulated by a spatial complex sinusoidal grating in the spatial domain. The filter frequency response is real as shown in Figure 3.7 and can be specified by four parameters named as the radial center frequency $f_0$, the preferred orientation $\theta_0$, the radial frequency bandwidth $B_0$, and the
Figure 3.6: Textile defects detected using LVDCT features.
Figure 3.7: Definition of a Gabor filter in the spatial frequency domain.

orientation bandwidth $B_\theta$ expressed by

$$G(u, v) = e^{-2\pi^2(u' - f_0)^2 + (v' - \omega_0)^2}$$  \hspace{1cm} (3.14)

where $(u', v') = (u \cos \theta_0 + v \sin \theta_0, -u \sin \theta_0 + v \cos \theta_0)$ are the rotated coordinates. The bandwidths are measured from the half-peak of the Gaussian envelope. The spatial extent of the Gaussian envelope in the spatial domain is governed by scaling factors $\sigma_x$ and $\sigma_y$ which are given by

$$\sigma_x = \frac{1}{f_0 \pi \sqrt{\ln 2}} \sqrt{\frac{2\theta_x + 1}{2\theta_x - 1}}$$  \hspace{1cm} (3.15)

$$\sigma_y = \frac{1}{f_0 \pi \tan \frac{\theta_x}{2}} \sqrt{\frac{\ln 2}{2}}$$  \hspace{1cm} (3.16)

By varying the four parameters $f_0$, $\theta_0$, $B_u$, and $B_\theta$, a bank of Gabor filters can be generated which span the Fourier plane.

The texture measures used here are the energy of the filtered image. Denote by $g_i(\cdot, \cdot) = g_{r_i}(\cdot, \cdot) + j g_{i \cdot}(\cdot, \cdot)$ a complex Gabor filter associated with channel $i$ defined in the spatial domain and let $x(k, l)$ be a real value image function (see Figure 3.8). Applying the complex Gabor filter to $x(k, l)$ yields a complex response $k_i(k, l)$,

$$k_i(k, l) = kr_i(k, l) + j ki_i(k, l)$$

$$= \sum_{k', l'} g_{r_i}(k', l') x(k - k', l - l') + j \sum_{k', l'} g_{i \cdot}(k', l') x(k - k', l - l').$$  \hspace{1cm} (3.17)

with the real and imaginary components $kr_i(k, l)$ and $ki_i(k, l)$. The amplitude of the filtered image, $|k_i(k, l)|$, provides the instantaneous envelope which yields large values over the regions of the
original image containing texture attributes the filter is tuned to. The energy of the filtered image is then given as

\[ v_{gabor}(k, l) = k r_i(k, l)^2 + k i_i(k, l)^2 \]

\[ = \left[ \sum_{k',l'} g_{r}(k', l') x(k - k', l - l') \right]^2 + \left[ \sum_{k',l'} g_{i}(k', l') x(k - k', l - l') \right]^2 \] (3.18)

The complex Gabor filter is approximately analytic and it can therefore be considered as a quadrature filter pair. Hence when taking the amplitude of the filtered image as texture measure, only one half of the Fourier plane has to be covered by the filter set since a real valued image has a Hermitian Fourier transform, i.e. both half-planes convey identical information [4].

### 3.7 Representational Equivalence

In the preceding sections, the texture representation concepts of orthogonal DCT filters based on local variance computation and Gabor filter based texture energy estimation were discussed. Although these two methods seem fundamentally different, we shall show that the representational structure of both approaches is inherently identical.

To start with, consider Equation (3.11). If the non-linear function \( f(y) \) is chosen to be a square
function, this Equation can be re-expressed as

$$v_{\text{det}}(k, l) = \frac{1}{W^2} \sum_{p=0}^{W-1} \sum_{q=0}^{W-1} [y_i(k-p, l-q)]^2$$

(3.19)

Substituting for $y_i(\cdot, \cdot)$ from Equation (3.8) and grouping the summation components, one obtains

$$v_{\text{det}}(k, l) = \frac{1}{W^2} \sum_{p,q=0}^{W-1} \sum_{a,b,c,d=0}^{N-1} \left[ x(k-p-a, l-q-b) x(k-p-c, l-q-d) d_i(a+1, b+1) d_i(c+1, d+1) \right]$$

(3.20)

Rearranging the Gabor texture energy expressed by Equation (3.18), and substituting for $kr_i(k, l)$ and $ki(k, l)$ from Equation (3.17), one has

$$v_{\text{gabor}}(k, l) = \sum_{k', l', k'' l''} x(k - k', l - l') x(k - k'', l - l'') \left[ g_{r_i}(k', l') g_{r_i}(k'', l'') + g_i(k', l') g_i(k'', l'') \right]$$

(3.21)

As can be seen from Equations (3.20) and (3.21), the two approaches have a similar structural representation and only the parameters of the representational structure differ. One can therefore consider the orthogonal mask approach in equation (3.19) to compute the energy of a set of equivalent $W \times W$ filters, very much like in the Gabor filter approach.

### 3.8 Experiment: Performance Comparison

The main interest in this section is to compare the performance of the above two texture feature extraction approaches. All the test images generated for this study contain two texted regions. The images are composed of a set of nine Brodatz textures which are selected to represent the type of textures easily found in nature. The Brodatz texture images are histogram equalised and combined into six pairs as shown in Figure 3.9. Each test image, $256 \times 256$ pixel in size, contains a rotated textured ‘cross’ placed on the foreground with different texture at the background. The images constructed in this way in various texture combinations are assumed to provide a realistic representation of the range of texture segmentation problems encountered in real imagery.

#### 3.8.1 Texture Representation

The main consideration faced in texture representation is to derive a set of channel filters of small cardinality but providing a good coverage of the Fourier plane.

(a). Discrete Cosine Transform. The set of eight $3 \times 3$ DCT filters used in the experiment of Section 3.5.1 is being employed here. The texture measures are defined as the eight local variances of the filter bank outputs computed on a $15 \times 15$ pixel overlapping window.
Figure 3.9: Test images.
(b). *Gabor Filter.* To make the feature set comparable, a set of eight Gabor filters is defined to partition the half plane into four orientations and two frequency bands. The four orientations are equidistant and the two frequency bands are distributed in 1.5 octave steps. The orientation bandwidth $B_\theta$ of the filters therefore equals $45^\circ$ and the preferred orientations $\theta_a$ are chosen to take values from $\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$. The radial frequency bandwidths $B_u$ are 1.5 octave wide and the following values of radial center frequency $f_0$ are used: 0.125 and 0.35. Figure 3.10 shows the frequency responses of this set of Gabor channel filters. Filtering operations are performed in the Fourier domain, in order to save computation time. The energy is evaluated from the inverse Fourier transform of the filter output. The spatial extent of the corresponding complex Gabor filter in the spatial domain can be determined from equations (3.15) and (3.16).

### 3.8.2 Aims and Experiment Procedures

The performance for these two texture extraction approaches is assessed in this section by means of classification accuracy. A supervised Bayesian pixel classifier was used for the two-class supervised segmentation problems. Each test image was subjected to two texture extraction methods:

(a). taking local variance of the DCT channel filter outputs over a $15 \times 15$ sliding window as texture features (LVDCT).
3.8: Experiment: Performance Comparison

(b). taking the energy of the complex Gabor channel filter outputs as texture features (EOG).

Applying the Bayesian pixel classifier to these two different multi-feature representations of the original test image yielded the segmentation results.

The intent is to study how the size of the filter kernel affects the segmentation results. Although the Gabor filtering was performed in the Fourier domain, the spatial size of the filter kernel is fixed implicitly by the four parameters as in equations (3.15) and (3.16). The Gabor filters used have the equivalent kernel size up to 15 x 15 pixels. The question addressed here is whether one should use larger size operators to extract features directly; or use a set of relatively small size operators to describe the local features and take the local statistics, such as local variance, computed from the filtered images over a moving window of larger size.

3.8.3 Results and Discussion

For quantitative comparison, the classification results for all test images and different texture representation schemes are given in the series of images in Figure 3.11. In all cases, the LVDCT features yield better segmentation results. The EOG features are quite good for deterministic textures, for example, Image 3.

In all tests, LVDCT gave high quality results. In contrast the performance of Gabor channel filters is rather disappointing although the Gabor filters used here have a comparable implicit spatial size of about 15 x 15 pixels. However the poor performance of Gabor filters probably relates to the choice of texture images. Most of the Brodatz textures used in experiments are stochastic in nature. They exhibit no dominant frequency or orientation characteristics. In contrast a Gabor filter is particularly sensitive to textures which have high concentration of localised spatial frequencies. It is therefore not surprising to find that the only satisfactory result is obtained on Image 3 composed of highly structural textures. However, LVDCT scheme works well on all test images.

It is only fair to point out that the LVDCT scheme is computationally demanding. No existing method can be used to speed up the local variance estimation in the way the Fast Fourier Transform accelerates Gabor filtering. On the other hand, it has been demonstrated that the local variance representation based on DCT and the energy representation derived from Gabor/quadrature filtering are inherently identical in their representational structure. This raises yet another question: is it possible to derive a set of quadrature filters to be used for feature extraction which are equivalent to the texture characterisation of the DCT filters approach. This will be discussed in
Figure 3.11: Segmentation results. The images show the segmentation results on the test images using the two feature representation methods. Rows 1 and 2): Bayesian pixel classifier with the energy of Gabor filter outputs as texture features (EOG), Rows 3 and 4): Bayesian pixel classifier with the local variance taken from a $15 \times 15$ moving window on DCT filter outputs as texture features (LVDCT).
3.9 Finding Equivalent Filters

As mentioned before in the previous Section, the feature extraction process can be speeded up in the case of the Gabor filtering approach by resorting to Fast Fourier Transform techniques. Unfortunately this could not be applied for the LVDCT features estimation as no fast algorithm exists for the overall computation of the three stages of feature extraction (i.e. filtering, variance estimation, and local smoothing). It has been shown in Section 3.7 that the structural representation of these two approaches is very similar. In principle one could then emulate the LVDCT features by the total sum of energy computed from a set of equivalent quadrature filters just like in the computation of the EOG features. The task now is to determine this set of quadrature filters necessary for the implementation.

3.9.1 Local Neighbourhood

To aid in the explanation, a 1-dimensional LVDCT feature extraction will be used as an example. A sequence of pixels involved in the LVDCT feature extraction process can be denoted by the vector \( \{x_i\}, (i = 1, 2, \ldots, W) \) where \( W \) is the size of the region in which the features are computed (i.e. window size). The first stage filtering by the \( 3 \times 1 \) DCT filters (with coefficients \( f_{-1}, f_0, \) and \( f_1 \)) is shown in Figure 3.12. The second stage operation is to non-linearly (square) transform these filtered components \( \{y_i\}, (i = 1, 2, \ldots, W - 2) \). In the case when \( W = 9 \), the resultant output \( \{u_i\}, \)
can then be expressed as

\[ u_1 = y_1^2 = (x_1 f_{-1} + x_2 f_0 + x_3 f_1)^2 \]  
\[ = x_1^2 f_{-1}^2 + x_2^2 f_0^2 + x_3^2 f_1^2 + 2x_1 x_2 f_{-1} f_0 + 2x_1 x_3 f_{-1} f_1 + 2x_2 x_3 f_{-1} f_1 \]  

\[ u_2 = y_2^2 = (x_2 f_{-1} + x_3 f_0 + x_4 f_1)^2 \]  
\[ = x_2^2 f_{-1}^2 + x_3^2 f_0^2 + x_4^2 f_1^2 + 2x_2 x_3 f_{-1} f_0 + 2x_2 x_4 f_{-1} f_1 + 2x_3 x_4 f_{-1} f_1 \]  

\[ u_3 = y_3^2 = (x_3 f_{-1} + x_4 f_0 + x_5 f_1)^2 \]  
\[ = x_3^2 f_{-1}^2 + x_4^2 f_0^2 + x_5^2 f_1^2 + 2x_3 x_4 f_{-1} f_0 + 2x_3 x_5 f_{-1} f_1 + 2x_4 x_5 f_{-1} f_1 \]  

\[ u_4 = y_4^2 = (x_4 f_{-1} + x_5 f_0 + x_6 f_1)^2 \]  
\[ = x_4^2 f_{-1}^2 + x_5^2 f_0^2 + x_6^2 f_1^2 + 2x_4 x_5 f_{-1} f_0 + 2x_4 x_6 f_{-1} f_1 + 2x_5 x_6 f_{-1} f_1 \]  

\[ u_5 = y_5^2 = (x_5 f_{-1} + x_6 f_0 + x_7 f_1)^2 \]  
\[ = x_5^2 f_{-1}^2 + x_6^2 f_0^2 + x_7^2 f_1^2 + 2x_5 x_6 f_{-1} f_0 + 2x_5 x_7 f_{-1} f_1 + 2x_6 x_7 f_{-1} f_1 \]  

\[ u_6 = y_6^2 = (x_6 f_{-1} + x_7 f_0 + x_8 f_1)^2 \]  
\[ = x_6^2 f_{-1}^2 + x_7^2 f_0^2 + x_8^2 f_1^2 + 2x_6 x_7 f_{-1} f_0 + 2x_6 x_8 f_{-1} f_1 + 2x_7 x_8 f_{-1} f_1 \]  

\[ u_7 = y_7^2 = (x_7 f_{-1} + x_8 f_0 + x_9 f_1)^2 \]  
\[ = x_7^2 f_{-1}^2 + x_8^2 f_0^2 + x_9^2 f_1^2 + 2x_7 x_8 f_{-1} f_0 + 2x_7 x_9 f_{-1} f_1 + 2x_8 x_9 f_{-1} f_1 \]  

The average of these outputs determines the LVDCT feature of this 1-dimensional vector. This LVDCT estimation process can be characterised by the following matrix

\[
\begin{array}{cccccccc}
A & F & G & x_1 & x_2 & x_3 & x_4 & x_5 \\
F & B & I & G & & & & \\
G & I & C & I & G & & & \\
G & I & C & I & G & & & \\
G & I & C & I & G & & & \\
G & I & C & I & G & & & \\
G & I & C & I & G & & & \\
G & I & D & H & & & & \\
G & H & E & & & & & \\
\end{array}
\]

where \( A = f_{-2}^2, B = f_{-1}^2 + f_0^2, C = f_{-1}^2 + f_0^2 + f_1^2, D = f_0^2 + f_1^2, E = f_1^2, F = f_{-0} f_0, G = f_{-1} f_1, H = f_0 f_1, \) and \( I = f_{-1} f_0 + f_0 f_1. \) This matrix, \( S, \) is symmetric with the undefined elements being zero. It indicates the relationships (i.e., \( S_{ij} \)) between any two pixels in the sequence and the filter coefficients necessary to estimate the LVDCT feature.
3.9: Finding Equivalent Filters

3.9.2 Equivalent Quadrature Filter Emulation

An attempt to emulate the process of the LVDCT feature extraction will be carried out in this Section. A set of $M$ quadrature filters of size $W \times 1$ must be determined so that the sum of energies of the filtered channels corresponds to a single LVDCT feature. The feature extraction process is similar to the EOG feature determined earlier. The coefficient matrix, $Q$, can be computed for the feature extraction process of these $M$ quadrature filters. The element of the matrix $Q_{i,j}$ can be expressed as

$$Q_{i,j} = \sum_{k=1}^{M} g_{i}^{k} g_{j}^{k}$$

(3.29)

where $g_{i}^{k}$ is the $i^{th}$ coefficient of the $k^{th}$ filter. $M$ is the total number of quadrature filters used.

Since the matrix for the LVDCT feature extraction process is symmetric, one can consider only the top triangle components. The total number of known components in the $S$ matrix is determined by the size of the smoothing region of the last stage and is equal to $W(W+1)/2$. Hence one can formulate a set of equations which relate the two matrices together. For example if $W = 9$, there is a total of $9(9+1)/2 = 45$ equations to be solved. Each quadrature filter will have $9$ unknown coefficients to be determined. This will require a top limit of $M \times 9 = 45 (M = 5)$ quadrature filters to successfully emulate the LVDCT process.

Since these equations are non-linear, one has to resort to an optimisation technique to solve them. Defining the square error between the two matrices as the cost function,

$$J = \sum_{i=1}^{W} \sum_{j=1}^{W} \left( \sum_{k=1}^{M} g_{i}^{k} g_{j}^{k} - s_{i,j} \right)^{2}$$

(3.30)

the first derivative with respect to $g_{i}^{k}$ can be expressed as

$$\frac{\delta J}{\delta g_{i}^{k}} = \sum_{i=1}^{W} \sum_{j=1}^{W} 2g_{i}^{k} \left( \sum_{k=1}^{M} g_{i}^{k} g_{j}^{k} - s_{i,j} \right)$$

(3.31)

Starting from some initial values of $g_{i}^{k}$, a gradient descent process (minimisation of the cost function $J$) is initiated until the sum of the derivative is close to zero, e.g.

$$\sum_{i=1}^{W} \sum_{j=1}^{W} \frac{\delta J}{\delta g_{i}^{k}} \approx 0$$

(3.32)

This is the position of the local optimum. A standard updating scheme for the filter coefficient at every iteration is

$$g_{i}^{k\text{new}} = g_{i}^{k\text{old}} - \alpha \frac{\delta J}{\delta g_{i}^{k}}$$

(3.33)

where $\alpha$ value is small, i.e. 0.001.
Chapter 3: Texture Measurement by Local Linear Transform

Table 3.5: Estimation of the set of 5 emulated quadrature filters for the LVDCT feature extraction process. DCT filter \([-1, 0, 1]\) with smoothing region \(W = 9\).

<table>
<thead>
<tr>
<th>Filter Coef.</th>
<th>Filter 1</th>
<th>Filter 2</th>
<th>Filter 3</th>
<th>Filter 4</th>
<th>Filter 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(g_1)</td>
<td>0.000000</td>
<td>-0.371748</td>
<td>0.601501</td>
<td>0.000000</td>
<td>0.601501</td>
</tr>
<tr>
<td>(g_2)</td>
<td>0.500000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.601501</td>
</tr>
<tr>
<td>(g_3)</td>
<td>0.000000</td>
<td>0.973249</td>
<td>-0.973249</td>
<td>0.000000</td>
<td>-0.229753</td>
</tr>
<tr>
<td>(g_4)</td>
<td>-1.207107</td>
<td>0.000000</td>
<td>0.000000</td>
<td>-0.707107</td>
<td>0.000000</td>
</tr>
<tr>
<td>(g_5)</td>
<td>0.000000</td>
<td>-1.203002</td>
<td>0.000000</td>
<td>0.000000</td>
<td>-0.743496</td>
</tr>
<tr>
<td>(g_6)</td>
<td>1.207107</td>
<td>0.000000</td>
<td>0.000000</td>
<td>-0.707107</td>
<td>0.000000</td>
</tr>
<tr>
<td>(g_7)</td>
<td>0.000000</td>
<td>0.973249</td>
<td>0.973249</td>
<td>0.000000</td>
<td>-0.229753</td>
</tr>
<tr>
<td>(g_8)</td>
<td>-0.500000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.707107</td>
<td>0.000000</td>
</tr>
<tr>
<td>(g_9)</td>
<td>0.000000</td>
<td>-0.371748</td>
<td>-0.601501</td>
<td>0.000000</td>
<td>0.601501</td>
</tr>
</tbody>
</table>

Table 3.6: Estimation of the set of 5 emulated quadrature filters for the LVDCT feature extraction process. DCT filter \([-1, 2, -1]\) with smoothing region \(W = 9\).

<table>
<thead>
<tr>
<th>Filter Coef.</th>
<th>Filter 1</th>
<th>Filter 2</th>
<th>Filter 3</th>
<th>Filter 4</th>
<th>Filter 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>(g_1)</td>
<td>-0.385676</td>
<td>0.211517</td>
<td>0.491501</td>
<td>0.510221</td>
<td>0.440397</td>
</tr>
<tr>
<td>(g_2)</td>
<td>1.258853</td>
<td>-0.778863</td>
<td>-1.292920</td>
<td>-0.951929</td>
<td>-0.468558</td>
</tr>
<tr>
<td>(g_3)</td>
<td>-1.697719</td>
<td>1.379443</td>
<td>0.905161</td>
<td>-0.111553</td>
<td>-0.525844</td>
</tr>
<tr>
<td>(g_4)</td>
<td>1.161583</td>
<td>-1.759213</td>
<td>0.592064</td>
<td>1.038009</td>
<td>0.214082</td>
</tr>
<tr>
<td>(g_5)</td>
<td>0.000000</td>
<td>1.894232</td>
<td>1.391613</td>
<td>0.000000</td>
<td>0.679846</td>
</tr>
<tr>
<td>(g_6)</td>
<td>-1.161583</td>
<td>-1.759213</td>
<td>0.592064</td>
<td>-1.038009</td>
<td>0.214082</td>
</tr>
<tr>
<td>(g_7)</td>
<td>1.697719</td>
<td>1.379443</td>
<td>0.905161</td>
<td>0.111553</td>
<td>-0.525844</td>
</tr>
<tr>
<td>(g_8)</td>
<td>-1.258853</td>
<td>-0.778863</td>
<td>-1.292920</td>
<td>0.951929</td>
<td>-0.468558</td>
</tr>
<tr>
<td>(g_9)</td>
<td>0.385676</td>
<td>0.211517</td>
<td>0.491501</td>
<td>-0.510221</td>
<td>0.440397</td>
</tr>
</tbody>
</table>

3.9.3 Results and Discussion

The 1-dimensional LVDCT features to be emulated will be from the DCT filters of the form \([-1, 0, 1]\) and \([-1, 2, -1]\). As in previous example, the smoothing region size chosen is \(W = 9\). The coefficients of the five quadrature filters estimated in this manner are shown in Table 3.5 and Table 3.6.

It can be observed that these quadrature filters are uniquely orthogonal with their coefficients summing up to zero, thus retaining the same dc level as in the case of the LVDCT features. To see the success of the emulation, the matrix for both the LVDCT computation and equivalent quadrature filter approach are shown in Table 3.7 and Table 3.8 respectively for the two DCT filters concerned. It can be seen that the corresponding matrices are quite similar indicating that the set of five quadrature filters can successfully emulate the operation of the LVDCT feature extraction.

It is to be noted that the total number of quadrature filters necessary for this implementa-
3.9: Finding Equivalent Filters

Table 3.7: Top Matrix: matrix for the LVDCT feature extraction using the DCT filter of the form \(\{-1,0,1\}\). Bottom Matrix: matrix of the same feature extraction process using the emulated quadrature filters.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>1.00</td>
<td>0.00</td>
<td>-1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>x2</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
<td>-1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>x3</td>
<td>-1.00</td>
<td>0.00</td>
<td>2.00</td>
<td>0.00</td>
<td>-1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>x4</td>
<td>0.00</td>
<td>-1.00</td>
<td>0.00</td>
<td>2.00</td>
<td>0.00</td>
<td>-1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>x5</td>
<td>0.00</td>
<td>0.00</td>
<td>-1.00</td>
<td>0.00</td>
<td>2.00</td>
<td>0.00</td>
<td>-1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>x6</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-1.00</td>
<td>0.00</td>
<td>2.00</td>
<td>0.00</td>
<td>-1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>x7</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-1.00</td>
<td>0.00</td>
<td>2.00</td>
<td>0.00</td>
<td>-1.00</td>
</tr>
<tr>
<td>x8</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-1.00</td>
<td>0.00</td>
<td>1.00</td>
<td>0.00</td>
</tr>
<tr>
<td>x9</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-1.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>0.86</td>
<td>0.00</td>
<td>-1.09</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.09</td>
<td>0.00</td>
<td>0.14</td>
</tr>
<tr>
<td>x2</td>
<td>0.00</td>
<td>0.75</td>
<td>0.00</td>
<td>-1.00</td>
<td>0.00</td>
<td>0.10</td>
<td>0.00</td>
<td>0.25</td>
<td>0.00</td>
</tr>
<tr>
<td>x3</td>
<td>-1.09</td>
<td>-0.00</td>
<td>1.95</td>
<td>0.00</td>
<td>-1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>x4</td>
<td>0.00</td>
<td>1.10</td>
<td>0.00</td>
<td>1.96</td>
<td>0.00</td>
<td>-0.96</td>
<td>0.00</td>
<td>0.10</td>
<td>0.00</td>
</tr>
<tr>
<td>x5</td>
<td>0.00</td>
<td>0.00</td>
<td>-1.00</td>
<td>0.00</td>
<td>2.00</td>
<td>0.00</td>
<td>-1.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>x6</td>
<td>0.00</td>
<td>0.10</td>
<td>0.00</td>
<td>-0.96</td>
<td>0.00</td>
<td>1.96</td>
<td>0.00</td>
<td>-1.10</td>
<td>0.00</td>
</tr>
<tr>
<td>x7</td>
<td>0.09</td>
<td>0.00</td>
<td>0.05</td>
<td>0.00</td>
<td>-1.00</td>
<td>0.00</td>
<td>1.95</td>
<td>0.00</td>
<td>-1.09</td>
</tr>
<tr>
<td>x8</td>
<td>0.00</td>
<td>0.25</td>
<td>0.00</td>
<td>0.10</td>
<td>0.00</td>
<td>-1.10</td>
<td>0.00</td>
<td>0.75</td>
<td>0.00</td>
</tr>
<tr>
<td>x9</td>
<td>0.14</td>
<td>0.00</td>
<td>0.09</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-1.09</td>
<td>0.00</td>
<td>0.86</td>
</tr>
</tbody>
</table>

The number of filters for a window of size \(W\) is \(M = (W + 1)/2\). Therefore, if the size of this region increases, the number of filters needed for the implementation will also increase. This effect will be even more prominent when the argument is extended to the 2-dimensional case. This time the smoothing is performed over a local region of spatial size \(W \times W\). As an example, if a 9 \(\times\) 9 macro-window is used, then the total number of 2-dimensional quadrature filters needed will be \((W \times W + 1)/2\) (i.e. \(M = 41\)). This is computationally very demanding to realise in practice even if the computationally efficient Fast Fourier Transform algorithms is used.

It is possible to use some form of filter reduction process to cut down on the number of filters to be used. However, this is by no means an easy task to realise when the number of filters becomes too large. The benefit of saving in computation using a subset of quadrature filters must be compromised with the extra computational cost for the filter reduction process. At the same time one must be prepared to accept the unavoidable element of error in estimating the LVDCT features. Therefore, it can be concluded here that although the LVDCT feature extraction process can
Table 3.8: Top Matrix: matrix for the LVDCT feature extraction using the DCT filter of the form \{-1,2,1\}. Bottom Matrix: matrix of the same feature extraction process using the emulated quadrature filters.

be implemented by means of filtering using the emulated quadrature filters, the latter approach will be even more computationally demanding to realise.

3.10 Conclusion

It has been shown that access to the $N^{th}$-order probability density function of a collection of pixels can be made possible by means of simple first-order histograms estimated along the principal component axes. The Karhunen-Loeve transformation provides an optimum solution to the computation of transform filters needed to achieve a de-correlated data representation. This is a necessary condition for the approximation of the $N^{th}$-order probability density function to be valid.

The practical importance of the more generally applicable sub-optimal transform processing has been emphasized as it considerably simplifies computation.

Experiments on texture classification have shown that the success rate does not depend on the type of the sub-optimal transform used. Generally speaking, the performance depends on the general form of the filters (edge-like, spot-like, etc.) and not specifically on the numerical
composition of the filters' coefficients. It was also revealed that the size of these transform filters
does not play a major part for texture discrimination. Hence it is recommended that small filters
(i.e 3 x 3) be used as they provide a more compact description of texture resulting in a lower
dimensional feature representation space. If higher classification rate is needed, one can proceed
to include higher-order statistical descriptions of the histograms. The inclusion of third and fourth
moments together with the variance measures has been shown to improve texture classification.
However, the use of higher moments will only be beneficial in the case when the textures are
non-Gaussian. The LVDCT features have been applied very successfully to a problem of defect
detection involving textile materials.

Within the framework of multi-channel filtering, the Gabor filtering approach was shown to be
identical to the orthogonal basis filters approach in terms of their overall representational struc­
ture. However, the performances of these two approaches on texture segmentation are rather
different, with the LVDCT features performing significantly better then the EOG features.

It should be noted that the extraction of the LVDCT features is computationally demanding.
No existing fast method can be used to speed up the feature extraction process in the way the Fast
Fourier Transform accelerates Gabor filtering. Turning on to the advantage of the latter repre­
sentation and noting the similarity in form between the orthogonal filters and the Gabor filtering
schemes, a technique to emulate the LVDCT feature extraction process by using the sum of en­
ergies from a series of equivalent quadrature filters was devised. However, it has been noted
that this technique is not cost effective as the number of quadrature filters needed for a successful
emulation is very large. The filtering of the image with this large amount of filters, even with
the accelerated speed of Fast Fourier Transform entails a huge computational burden. Therefore,
local linear transform measurements for texture analysis should be estimated using the traditional
three-stage approach of filtering, local variance estimation, and region smoothing. If speed is of
the essence, then implementation of the local linear transform on parallel machines would have to
be considered.

References


Chapter 4

Colour Texture Feature Extraction

4.1 Introduction

Research on grey level texture analysis had enjoyed a great deal of attention over the last three decades. This resulted in novel approaches to solving aspects of the multifaceted texture analysis problem [29][38] as cited in Chapter 2. The motivation behind the work on texture analysis is multifold. It includes texture description for image indexing purposes, texture segmentation and texture edge detection as a prerequisite to higher level image analysis and texture classification.

Besides texture, colour information is also a very important cue to the human vision system. Psychophysical evidence had suggested that colour could aid the human vision system in the tasks involving object identification and discrimination [12]. Beck [2] and Marr [30] also suggested that colour of the primitive textural elements found in texture surfaces could help in the visual discrimination of texture by humans. It is then natural to assume that the perception of texture can be aided by colour information. The use of multi-spectral and colour image data is rapidly gaining popularity with the advent of fast computing hardware and the widespread availability of good quality colour cameras, digitizers, and monitors. The traditional texture analysis methods could be adapted to multi-channel data very successfully by incorporating the element of spectral or colour attributes into their original formulation. Rosenfeld et al. [39] used the distribution of absolute differences of pairs of pixels displaced by \( \delta_{xy} \) over all bands of a multi-band image to capture the colour texture content. They gave a specific example involving a 2-band image. Their method can be generally extended to 3-band colour images. However, the main difficulty in using the distribution of absolute differences of pairs of pixels is the choice of \( \delta_{xy} \) which is highly dependent on the underlying structure of the texture to be analysed. Harms et al. [16] considered natural scenes to be made up of small grains or textons with nearly constant intensity.
and colour. Spatial arrangements of these textons relative to the neighbours, produce a visually perceptible texture. They then used colour, shape and size features of textons to characterise stained blood cells. Celenk and Smith [8] used a statistical and structural technique which makes use of the spatial and spectral (colour) information contained in colour images. They defined three paradigms which relate the spatio-spectral grouping property of the human eye for some basic visual patterns. These paradigms were developed in accordance with Julesz's conjecture [20][21][22] regarding continuity, texture and isolation patterns. Colour and texture attributes were used by Dhawan et al. [11] in a multi channel region analysis for detecting early malignant melanoma and predicting its prognosis.

Regardless of the actual application of colour texture, the most important issue in any texture processing is the choice of texture representation. The aim of this chapter is to address these issues in the context of colour texture which involves, in addition to selecting a suitable means of texture characterisation, also the choice of an appropriate colour co-ordinate system. The feasibility of extending the local linear transform approach (i.e. LVDCT) described in Chapter 3 to multi-band colour texture representation will be investigated. This raises the questions of what colour space is most suitable in conjunction with this particular representation, and how much additional information is conveyed by colour. A Colour Texture Feature Extraction scheme will be introduced in this chapter. This representation will require two stages of processing. The first stage involves the transformation of the original tristimulus values into one of various colour co-ordinate system spaces. The choice of a suitable colour co-ordinate system is very much dependent on the nature of the colour image data [23][32][35]. A system reported to work well in one instance may not be suitable for other applications. Hence the need to study the effect of these colour co-ordinate systems on colour texture images which is one of the aims of this chapter. The transformed component images yield various degrees of spatial texture information, hence it is adequate to characterise the colour texture by computing texture information on these components.

Section 4.2 describes briefly the psychophysical aspect of colour vision and how it can play a role in machine vision applications. The CIE trichromatic model and the NTSC receiver primary colour system will be introduced in Sections 4.3 and 4.4 respectively. A set of eight widely used colour co-ordinate systems will be discussed in Section 4.5. Section 4.6 and 4.7 will show that texture information is available in almost all the component images of the transformed co-ordinate systems. The advocated model for colour texture representation will be introduced in section 4.8. Section 4.9 deals with the texture representation aspect of the colour texture model. Experimental
results of using these colour texture features for classifying colour granite images followed by some discussions will be reported in Section 4.10. Section 4.11 concludes this chapter.

4.2 Colour Vision

Colour vision has always been an area which benefited from contributions of different disciplines of science. Developments in the area of physics, physiology, and psychology and their implication on computer vision have led to many widely accepted and hypothesized models of this sensing modality which is considered greatly to enrich the human visual experience. Far beyond its aesthetic effect, colour vision is of great practical value for detecting patterns and objects which would be elusive in a world devoid of colour [14]. Colour has been used in conjunction with machine vision for image analysis and pattern recognition, for example to identify spray paint caps [3], to detect colour codes on resistors [1], to guide a robot to selectively pick up petri dishes [31], and for visual inspection of surface reflectance which is directly related to the quality of surface finish and paint [27]. Colour has also been used to segment images into sets of uniform colour regions [5][9][24][42], and to detect edges in images [32].

Perhaps one of the main contribution to the rapid development of colour for machine vision applications are the discoveries of psychophysical theories which made colour more widely understood. The trichromatic theory of colour vision was first suggested at the beginning of the 19th century by Thomas Young. It considered colour as a combination of three primary colours R (red), G (green), and B (blue). The opponent-colour model was proposed by Ewald Hering who suggested that three mechanisms mediate in colour vision; one that accounts for perception of red and green, the second accounts for yellow and blue, and the third for black and white distinctions [19]. The retinex theory of lightness perception is described by Land and McCann [25] and by Land [26]. It explains how it is possible, in the restricted world of “Mondrian” images, to compute lightness - the psychophysical correlate of surface reflectance - even under highly non-uniform illumination.

Colour does not just add beauty to a scene, it has many functional roles in vision;

- Some tasks humans perform, such as searching, classifying, and predicting are made easier and faster with the use of colour cues and knowledge about objects' colour.
- Colour is identified as one of the features observed pre-attentively by humans once they encounter a scene, and as such, it is computed in parallel with other features so that they
can be integrated during interpretation at a later stage.

- The most evident interaction of colour is with contours - colour seems to fill areas surrounded by contours and can assist in border detection when the stimulus has luminance changes.

- Colour is apparently perceived during motion, which suggests that there is some interaction between the two mechanisms.

The first two roles of colour vision suggest the usefulness of integrating colour with a texture description. The colour model used must be simple as it has to co-exist with the texture analysis algorithm so as to make the overall system of representation efficient to realise in practice. The trichromatic theory of colour suits this requirement very well. Furthermore, this psychophysical theory had been shown to apply very well to the machine vision arena [1][27][31].

4.3 Trichromatic Model

Due to the structure of the human eye, all colours are seen as variable combinations of the three primary colours R (red), G (green), and B (blue). There is a practical need to deal with colour in fields such as television, movies, photography, printing, as well as paint manufacturing. The C.I.E. (Commission Internationale de L'Éclairage) had adopted a standard based on the assumption of colour trichromacy. They designated the following specific wavelength values to the three basic colours; red = 700 nm, green = 546.1 nm, and blue = 435.8 nm. Trichromacy has been taken to imply two basic assumptions:

(a). Any colour can be matched by a vectorially additive combination of three primary spectral responses.

(b). The colour processing is linear, indicating the properties of proportionality and additivity.

The primary colours can be added to produce the secondary colours of light - magenta (red plus blue), cyan (green plus blue), and yellow (red plus green). Mixing the primaries, or a secondary with its opposite primary colour, in the right proportion produces white light.

The trichromatic model is shown in Figure 4.1. Any input into the visual system is mapped into a trivariance space which results in a representation of the 3-dimensional tristimulus space. The equations governing the transformation are [28]:

$$ R = \int \lambda I(\lambda)S_R(\lambda) d\lambda $$

(4.1)
4.4: R<sub>N</sub>G<sub>N</sub>B<sub>N</sub> - NTSC Receiver Primary Colour System

Since the conception of the general purpose C.I.E. spectral primary reference system with three monochromatic primaries, the National Television Systems Committee (NTSC) introduced another primary system that is more suitable to the receiver phosphor employed in cathode ray tubes of colour television and monitors. The C.I.E. spectral primary co-ordinate system RGB is related to the NTSC primary system R<sub>N</sub>G<sub>N</sub>B<sub>N</sub> by a linear co-ordinate conversion [4].

\[
\begin{bmatrix}
R_N \\
G_N \\
B_N
\end{bmatrix} =
\begin{bmatrix}
0.842 & 0.156 & 0.091 \\
-0.129 & 1.320 & -0.203 \\
0.008 & -0.069 & 0.897
\end{bmatrix}
\begin{bmatrix}
R \\
G \\
B
\end{bmatrix}
\]

\[ (4.4) \]
The NTSC primary system is not the same as the NTSC transmission colour co-ordinate system (i.e. YIQ as seen in Section 4.5.2). The latter being a system of transmitting the modulated luminance, hue and saturation components of a colour signal. The NTSC primary system is not affiliated to any mode of transmission system hence it could be used with any form of camera operating in any format (i.e. NTSC, PAL, or SECAM colour television systems).

4.5 Colour Co-ordinate Systems

It was shown that a colour can be matched by its tristimulus values for a given set of primaries (i.e. \(R_n, G_n, B_n\)). Alternatively, the colour may be specified by its chromaticity values and its luminance. A third approach in specifying a colour is to represent the colour by some linear or nonlinear invertible function of its tristimulus or chromaticity values. It can be pointed out that the linear transformation is simply a conversion of the primaries to a new set of primaries. Many different co-ordinate systems have been employed for the specification of colour. Several historical and analytic aspects will be discussed in the following subsections.

4.5.1 XYZ Colour Co-ordinate System

In the NTSC primary system the tristimulus values required to achieve a colour match are sometimes negative. The C.I.E. has developed a standard artificial primary co-ordinate system in which all tristimulus values required to match colours are positive [45]. The XYZ system primaries have been chosen so that the \(Y\) tristimulus value is equivalent to the luminance of the colour to be matched. The linear transform from the NTSC primary system to the XYZ colour co-ordinate system is:

\[
\begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix} =
\begin{bmatrix}
0.607 & 0.174 & 0.200 \\
0.299 & 0.587 & 0.114 \\
0.000 & 0.066 & 1.116
\end{bmatrix}
\begin{bmatrix}
R_n \\
G_n \\
B_n
\end{bmatrix}
\]

(4.5)

4.5.2 YIQ NTSC Transmission Colour Co-ordinate System

In the development of the United States colour television system, the NTSC formulated a colour co-ordinate system for the transmission of colour television signals composed of three tristimulus values YIQ[4]. The \(Y\) tristimulus value is the luminance of a colour. The other two tristimulus values \(I\) and \(Q\) jointly described the hue and saturation attributes of the image signal. The reasons for transmitting the YIQ tristimulus values rather than the \(R_nG_nB_n\) tristimulus values directly
from a colour camera are two-fold: the Y signal alone could be used with existing monochrome receivers to display monochrome images; furthermore, it was found possible to limit the spatial bandwidth of the I and Q signals without noticeable image degradation. As a result of this latter requirement, a clever analog modulation scheme was developed such that the bandwidth of a colour television carrier could be restricted to the same bandwidth as that of a monochrome carrier. The relationship between the NTSC primary system and the YIQ NTSC transmission colour co-ordinate system is:

\[
\begin{bmatrix}
Y \\
I \\
Q
\end{bmatrix} =
\begin{bmatrix}
0.299 & 0.587 & 0.114 \\
0.596 & -0.274 & -0.322 \\
0.211 & -0.523 & 0.312
\end{bmatrix}
\begin{bmatrix}
R_n \\
G_n \\
B_n
\end{bmatrix}
\]  

(4.6)

4.5.3 YUV Colour Co-ordinate System

In the PAL and SECAM colour television systems used in the United Kingdom and many European countries [7], the luminance Y and two colour differences

\[
U = \left( \frac{B_n - Y}{2.03} \right) = -0.148R_n - 0.289G_n + 0.437B_n
\]

(4.7)

\[
V = \left( \frac{R_n - Y}{1.14} \right) = 0.615R_n - 0.515G_n - 0.100B_n
\]

(4.8)

are used as transmission co-ordinates. The YUV co-ordinate system was initially proposed as the NTSC transmission standard, but was later replaced by the YIQ system because it was found [45] that the I and Q signals could be reduced in bandwidth to a greater degree than the U and V signals for an equal level of visual quality. The I and Q signals are related to the U and V signals by a simple rotation of co-ordinates in colour space.

\[
I = -U \sin(33^\circ) + V \cos(33^\circ)
\]

(4.9)

\[
Q = U \cos(33^\circ) + V \sin(33^\circ)
\]

(4.10)

4.5.4 \(I_1I_2I_3\) Ohta Colour Co-ordinate System

Ohta [36] experimented with the eigen vectors computed on several outdoor scenes and concluded that they are quite often dominated by the vectors \(\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right), \left(\frac{1}{4}, 0, \frac{1}{4}\right), \text{ and } \left(\frac{1}{4}, \frac{1}{4}, \frac{3}{4}\right)\). Hence it is possible to define three fixed orthogonal colour features,

\[
\begin{bmatrix}
I_1 \\
I_2 \\
I_3
\end{bmatrix} =
\begin{bmatrix}
0.333 & 0.333 & 0.333 \\
0.500 & 0.000 & -0.500 \\
0.250 & -0.500 & 0.250
\end{bmatrix}
\begin{bmatrix}
R_n \\
G_n \\
B_n
\end{bmatrix}
\]  

(4.11)
that will be helpful in providing a general solution for the Karhunen-Loeve transformation of a wide range of images.

4.5.5 IHS Colour Co-ordinate System

The IHS co-ordinate system [34] has been used within the computer vision community as a quantitative means of specifying the intensity, hue, and saturation of a colour. It is defined by the relations:

\[
\begin{bmatrix}
I \\
V_1 \\
V_2
\end{bmatrix} =
\begin{bmatrix}
0.333 & 0.333 & 0.333 \\
-0.408 & -0.408 & 0.816 \\
0.408 & -0.816 & 0.000
\end{bmatrix}
\begin{bmatrix}
R_N \\
G_N \\
B_N
\end{bmatrix}
\]

\[H = \tan^{-1}\left(\frac{V_1}{V_2}\right)\]

\[S = \left[(V_1)^2 + (V_2)^2\right]^{1/2}\]

By this definition, the colour blue is a zero reference for hue. Hue is the colour aspect of a visual impression, which normally corresponds to the stimulation of the retina by light of narrow range of wavelengths. Saturation signifies the relative white content of a stimulus perceived as having a particular hue. A fully saturated colour is of "pure" hue.

4.5.6 Normalised I_{norm}g Colour Co-ordinate System

In psychophysical terms, the chromaticity of an object, its hue and saturation, do not change with varying light intensity. In order to obtain this effect in a colour co-ordinate system, the following transformation must be applied to the elements of the function value:

\[r = \frac{R_N}{(R_N + G_N + B_N)}\]

\[g = \frac{G_N}{(R_N + G_N + B_N)}\]

\[b = \frac{B_N}{(R_N + G_N + B_N)}\]

These transformed values are called the trichromatic coefficients. Clearly, the trichromatic coefficients always sum to one, allowing all colour information to be contained in any two of the coefficients. The standard approach is to retain the \(r\) and \(g\) coefficients, called the chromaticity co-ordinates. Since the intensity values still contain important information, it can be included to form the third co-ordinate component. The transformation related to the intensity values is the normalised total intensity given by the following:

\[I_{norm} = (R_N + G_N + B_N)/(3 \times 255)\]
The three components $I_{\text{norm}}$, $r$, and $g$, constitute the normalised colour co-ordinate system.

### 4.5.7 Karhunen-Loeve (KL) Colour Co-ordinate System

Typically, the $R_N G_N B_N$ tristimulus values of the NTSC receiver primary system are highly correlated with one another [37]. In the development of efficient quantisation, coding, and processing techniques for colour images, it is often desirable to work with components that are uncorrelated. If the second-order moments of the $R_N G_N B_N$ tristimulus values are known, or at least estimated, it is possible to derive an orthogonal co-ordinate system, in which the components are uncorrelated, by a Karhunen-Loeve transformation of the $R_N G_N B_N$ tristimulus values. The KL colour transform is defined as

$$
\begin{bmatrix}
E_1 \\
E_2 \\
E_3
\end{bmatrix} =
\begin{bmatrix}
u_{11} & u_{12} & u_{13} \\
u_{21} & u_{22} & u_{23} \\
u_{31} & u_{32} & u_{33}
\end{bmatrix}
\begin{bmatrix}
R_N \\
G_N \\
B_N
\end{bmatrix}
$$

(4.18)

where the transformation matrix with general term $u_{ij}$ is composed of the eigenvectors of the $R_N G_N B_N$ covariance matrix with general term $c_{ij}$. The transformation matrix satisfies the relation

$$
\begin{bmatrix}
u_{11} & u_{12} & u_{13} \\
u_{21} & u_{22} & u_{23} \\
u_{31} & u_{32} & u_{33}
\end{bmatrix}
\begin{bmatrix}
C_{11} & C_{12} & C_{13} \\
C_{21} & C_{22} & C_{23} \\
C_{31} & C_{32} & C_{33}
\end{bmatrix}
\begin{bmatrix}
u_{11} & u_{21} & u_{31} \\
u_{12} & u_{22} & u_{32} \\
u_{13} & u_{23} & u_{33}
\end{bmatrix} =
\begin{bmatrix}
\lambda_1 & 0 & 0 \\
0 & \lambda_2 & 0 \\
0 & 0 & \lambda_3
\end{bmatrix}
$$

(4.19)

where $\lambda_1$, $\lambda_2$, $\lambda_3$ are the eigenvalues of the covariance matrix.

### 4.6 Spatial Definition of Transformed Images

The existence in the literature of many co-ordinate systems for colour processing purposes has made it necessary to scrutinise their analytical and representational significance when used in conjunction with colour texture processing. Figure 4.2 contains the transformed images of a typical colour texture image ("capao bonito") using the above colour co-ordinate systems.

The red, green and blue components of the colour image are observed to be highly correlated. The linear transformation systems, YIQ, YUV and I₁I₂I₃, have quite identical component images with one corresponding to the intensity of the image and the other two representing chromatic or colour difference signals. The chromatic planes seem to be lacking in spatial definition, one characteristic that the intensity component exhibits to a great extent. The XYZ components are related to the intensity of the image with specific emphasis on the red, green and blue channels.
respectively. Low saturation values of the S component of the IHS system were observed as natural images of this kind tend to map into small regions circumscribed around the grey points [18]. The hue component image has eliminated most micro-texture details since these structures possess quite identical hue information. The E₂ and E₃ components of the KL colour co-ordinate system have distinctly very low contrast, a consequence of the low variance (energy) values of these components. Most of the texture information is concentrated in the E₁ intensity component.

It can be seen that beside the intensity channel, most of the other two channels possess a certain degree of visual texture content. Though their spatial definition is somewhat limited, their usefulness for colour texture analysis should not be underestimated.

The choice of an appropriate colour co-ordinate system is very much data dependent [23][32][35][36]. It is therefore necessary to determine the candidate colour co-ordinate system that will be suitable to co-exist with the texture representation.

4.7 Energy of Component Images

Except for the normalised colour system Iₐₐₚₐₐₖ, the colour co-ordinate systems used here are all invertible; for example given the transformed values, the original RₙGₙBₙ components could be recovered. This suggests that the co-ordinate transformation systems are energy preserving with different proportion of its total energy being allocated to the various component images. Table 4.1 shows the normalised energy content of all three component images computed for all colour co-ordinate systems using the “capao bonito” colour texture data. It can be seen that the intensity images contain the bulk of the total energy. However the other components still contribute useful information. For example in the YIQ, YUV, and I₁I₂I₃ systems, the second and third components make up about a quarter of the total energy. The XYZ colour system has almost equal energy distributed across the three components. These components emphasize the responses of the three different colour channels. The red Rₙ and green Gₙ components of the tristimulus values contain most of the energy of the data suggesting that the image is reddish-green in colour.

The high energy intensity image does not necessarily provide the only source of textural definition. The ability of the chromatic components to highlight spectral details as seen in Figure 4.2 is also an important aspect which must be treated with equal importance. Again the usefulness of these chromatic components should not be ruled out as so often it appears to be the case.
4.7: Energy of Component Images

Figure 4.2: continue...
Figure 4.2: The colour transformed component images of the “capao bonito” image.
4.8 Multi-Band Model for Colour Texture

The block diagram of the multi-band colour texture representation algorithm is shown in Figure 4.3. The inputs to the system are the tristimulus values of the colour image. This system is quite similar to the common method used for processing multi-spectral images which characterise the textural content by means of texture extraction on the different spectral bands of the image. An extra stage is introduced here prior to the texture extraction process to make up the texture feature extraction model. This stage transforms the tristimulus values either linearly or nonlinearly into other colour co-ordinate spaces. This transformation holds some representation significance as the various colour co-ordinate systems could improve the overall performance of the multi-band colour texture feature extraction model as opposed to just using the tristimulus images. This model is based on the fact that spatial texture information of the colour image is available in all the components of the transformed channels as can be seen in Figure 4.2. Therefore the task of the final stage is to compute the texture features of these transformed component images. The question of which texture representation algorithm to use is also extremely important. The nature of the texture images considered here has warranted the use of the statistical approach.

4.9 Spatial Texture Representation

Any texture operator can be used to characterise the texture structure of the component images. For random textures, which will be the main concern in the ensuing discussion, statistical texture representation methods are generally known to be more appropriate [13][15]. The Local Linear Transform [33][43] was reported to be well suited for such application and for this reason we shall
adopt this representation in our studies. The texture transformation is easy to implement and capable of retaining most of the image texture information as seen in Chapter 3. The linear transform that will be used here is the Discrete Cosine Transform. A set of $3 \times 3$ Discrete Cosine filters can be evaluated (see Chapter 3) and these filters used to filter the three component images. The low pass filter is excluded from the set as it is not suitable for texture analysis. Energies in the form of variances can then be estimated locally over a larger window and used as texture features (LVDCT). The size of the window is chosen empirically so that enough colour texture information is present in the data set.

Therefore the output features of the multi-band colour texture feature extraction scheme will be the result of two stages of processing. The first involves the transformation of the tristimulus values into the various colour co-ordinate spaces. This is then followed by computing the texture information that is available within the component images of the new colour space giving rise to 24 colour texture features. These output features could be abbreviated synonymously by the two stage nature of feature extraction; i.e. $R_nG_nB_n$-LVDCT, $XYZ$-LVDCT, $YIQ$-LVDCT, $YUV$-LVDCT, $I_1I_2I_3$-LVDCT, $IHS$-LVDCT, $I_{norm}rg$-LVDCT and $KL$-LVDCT.
4.10 Experiment

4.10.1 Aim and Experiment Setting

The intent of the experiments here is to

- Demonstrate that colour can help in the discrimination of colour texture.

- Investigate which colour co-ordinate system is best suited for this purpose.

- Investigate the illumination invariance behaviour of the chromatic component image.

The colour textures to be used here are taken from an album of colour granite photographs. These photographs are digitally scanned by the Canon IPU 10/Colour Copier 300 scanning equipment into 24 bit, 256 × 256 pixel colour images. A set of 12 colour images experimented with is shown in Figure A.1 of Appendix A. They represent a variation of both texture and colour attributes that are most likely to be found in nature.

No attempt will be made here to statistically alter (i.e. histogram equalise) the colour data. This is usually done for grey level images to remove artifacts like variations in lighting, lens, A/D converters, quantisation and other factors that might change the mean intensity value of the texture image. Histogram equalisation of natural coloured images usually increases the saturation value of the image thus enhancing the original image [40]. Hence its use in reproduction and display purposes [17]. Its usage should only be restricted to this area of application and not as a means to normalise the first order statistics of the colour texture data for the purpose mentioned earlier. Histogram equalisation of a colour texture image could drastically change the original colour composition of the image. Preservation of the colour information is vital in colour vision work and as such the colour texture images processed here will be examined without prior processing of the original tristimulus values. As such, regular calibration of the imaging system should ensure any drifts in the device components are corrected for.

The supervised Bayesian classification scheme is used here and as such some prototype texture samples are required for training. Hence, the top half of each image is used for such purpose, generating the representative samples for the classifier to train on. A window of 100 × 100 pixels is moved along the image with a 1/20 overlap to generate 192 samples per class for training. Similarly, for the bottom half of each image the same moving window with a 1/5 overlap is used to generate 16 samples per class for testing purposes.
4.10.2 Results and Discussion

The results of using the multi-band colour texture feature extraction scheme for colour texture representation are shown in Table 4.2. Initially, the classification was performed separately on each individual component image of the various colour co-ordinate systems. It can be seen that the performance of the grey level component (i.e. the Y (75.52%) of the YIQ and YUV colour systems) is not as good as some of the chromatic components like the G\textsubscript{N} (85.42%) and B\textsubscript{N} (82.82%) of the R\textsubscript{N}G\textsubscript{R}B\textsubscript{N} system, the I (89.06%) of the YIQ system, the U (84.38%) and V (88.54%) of the YUV systems, the I\textsubscript{2} (78.13%) of the I\textsubscript{1}I\textsubscript{2}I\textsubscript{3} system, the r (83.33%) and g (82.29%) of the normalised co-ordinate system, the S (89.06%) of the IHS system, and finally the E\textsubscript{1} (86.90%) and E\textsubscript{2} (81.25%) of the KL system. This suggests a strong inclination towards texture features computed from the chromatic components of the colour co-ordinate system.

The confusion matrix of the classification results using texture features derived from the grey level image component is shown in Table 4.3. The best performing two components are from the I component of the YIQ system and the S component of the IHS system. Their confusion matrices are shown in Table 4.4 and Table 4.5 respectively. The misclassification associated with the grey level image component is mainly due to the visual closeness of some classes of texture when colour information is absent. As for the chromatic components, the causes of misclassification are mainly due to images with very large primitives. The size of the sample subimages used here (100 x 100 pixels) is insufficient to represent this form of texture with large tokens, hence the discrepancies.

Improvement of the results using single component images can be achieved by combining the 3 sets of features of each colour co-ordinate system to form an extended set of features (see the last column of Table 4.2). It can be seen that the classification rates using such features are in their high 90s with the YIQ-LVDCT and the I\textsubscript{1}I\textsubscript{2}I\textsubscript{3}-LVDCT combined features both achieving 100% correct classification. The performance of the R\textsubscript{N}G\textsubscript{R}B\textsubscript{N}-LVDCT and XYZ-LVDCT features are not as good as the rest because they are all very dependent on the intensity content of the image. The need of some form of chromatic information is necessary here to maintain good classification performance.

In order to determine the relative performance of grey level texture representation and multi-band colour texture representation, the set of features representing the DCT statistics was extended by including the third and fourth moments as described in Chapter 3. This will make the overall feature size to be compatible with the multi-band colour texture feature set (i.e. 24 features). The classification exercise was then performed using these grey level texture features. An accuracy
4.10: Experiment

Colour Features Used    | 1st comp. | 2nd comp. | 3rd comp. | Combined comp. |
------------------------|-----------|-----------|-----------|----------------|
RnGnBn-LVDCT           | 73.44% (Rn) | 85.42% (Gn) | 82.82% (Bn) | 96.36%          |
XYZ-LVDCT              | 69.79% (X)  | 75.52% (Y)  | 80.73% (Z)  | 95.31%          |
YIQ-LVDCT              | 75.52% (Y)  | 89.06% (I)  | 75.00% (Q)  | 100.0%          |
YUV-LVDCT              | 75.52% (Y)  | 84.38% (U)  | 88.54% (V)  | 97.40%          |
I1I2I3-LVDCT           | 73.96% (I1) | 78.13% (I2) | 69.79% (I3) | 100.0%          |
I_normI_r-LVDCT         | 68.75% (I_norm) | 83.33% (r)  | 82.29% (g)  | 98.89%          |
IHS-LVDCT              | 66.15% (I)  | 68.23% (H)  | 89.06% (S)  | 98.44%          |
KL-LVDCT               | 86.90% (E1) | 81.25% (E2) | 70.31% (E3) | 99.48%          |

Table 4.2: The results of using the features from the multi-band colour texture feature extraction schemes on colour granite classification. The classification accuracies using texture features derived individually on the three components are shown in column 2 through to column 4. The last column indicates the accuracy when all three features are linearly combined.

rate of 85.42% was obtained. This result when compared with the set of results obtained using the multi-band colour texture features which range from 95.31% to 100% indicates the following:

- That the grey level DCT texture representation is not able to fully discriminate the granite images.

- Any of the above multi-band colour texture representations improved discriminability. The YIQ-LVDCT and I1I2I3-LVDCT multi-band colour texture features performed best in the experiment.

Note that illumination intensity changes could have a drastic effect on some of the outcome of the multi-band colour texture features. It is important to know the behaviour of the colour coordinate systems under these circumstances. Usually the chromatic components of these colour systems exhibit a certain degree of tolerance over changes in illumination intensity (i.e. I, Q, U, V, I2, I3, r, g, H, S, E2, and E3). On the contrary, the intensity related components (i.e. Rn, Gn, Bn, X, Y, Z, Y, I1, I_norm, I, and E1) are very much affected by this change. The effect of illumination intensity changes is not very critical in a controlled environment where the system has autonomous control over the lighting apparatus and the calibration of the camera. However, in situation where the control of these parameters is not possible, it is necessary to resort to the use of the chromatic components only. The use of the intensity component for texture description is still possible under
Table 4.3: Confusion matrix for the classification of 12 classes of colour granite textures computed on the grey level component of the image.

<table>
<thead>
<tr>
<th>TRUE CLASS</th>
<th>ASSIGNED CLASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. bianco castilla</td>
<td>16 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>2. capao bonito</td>
<td>0 16 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>3. marron guaiba</td>
<td>0 0 16 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>4. new imperial</td>
<td>0 0 0 16 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>5. santiago</td>
<td>0 0 0 0 16 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>6. rosa baveno</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>7. balmoral GF</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>8. rosa monforte</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>9. rosso perla</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>10. cardinal red</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>11. brasil violett</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>12. ghiandonato</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>

Table 4.4: Confusion matrix for the classification of 12 classes of colour granite textures computed on the I component of the YIQ colour system.

<table>
<thead>
<tr>
<th>TRUE CLASS</th>
<th>ASSIGNED CLASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. bianco castilla</td>
<td>16 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>2. capao bonito</td>
<td>0 16 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>3. marron guaiba</td>
<td>0 0 16 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>4. new imperial</td>
<td>0 0 0 16 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>5. santiago</td>
<td>0 0 0 0 16 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>6. rosa baveno</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>7. balmoral GF</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>8. rosa monforte</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>9. rosso perla</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>10. cardinal red</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>11. brasil violett</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>12. ghiandonato</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>

Table 4.5: Confusion matrix for the classification of 12 classes of colour granite textures computed on the S component of the IHS colour system.

<table>
<thead>
<tr>
<th>TRUE CLASS</th>
<th>ASSIGNED CLASS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. bianco castilla</td>
<td>16 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>2. capao bonito</td>
<td>0 16 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>3. marron guaiba</td>
<td>0 0 16 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>4. new imperial</td>
<td>0 0 0 16 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>5. santiago</td>
<td>0 0 0 0 16 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>6. rosa baveno</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>7. balmoral GF</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>8. rosa monforte</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>9. rosso perla</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>10. cardinal red</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>11. brasil violett</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
<tr>
<td>12. ghiandonato</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0</td>
</tr>
</tbody>
</table>
4.10: Experiment

<table>
<thead>
<tr>
<th>Colour Features Used</th>
<th>Class. Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_N G_N B_N$-LVDCT</td>
<td>94.83%</td>
</tr>
<tr>
<td>XYZ-LVDCT</td>
<td>92.71%</td>
</tr>
<tr>
<td>YIQ-LVDCT</td>
<td>98.13%</td>
</tr>
<tr>
<td>YUV-LVDCT</td>
<td>95.98%</td>
</tr>
<tr>
<td>$I_1 I_2 I_3$-LVDCT</td>
<td>97.50%</td>
</tr>
<tr>
<td>$I_{norm} rg$-LVDCT</td>
<td>98.85%</td>
</tr>
<tr>
<td>IHS-LVDCT</td>
<td>97.71%</td>
</tr>
<tr>
<td>KL-LVDCT</td>
<td>96.35%</td>
</tr>
</tbody>
</table>

Table 4.6: The classification results on the colour texture images with varying image illumination intensity values. Features used are the 24 combined *multi-band colour texture* features.

these circumstances only after the component has been properly histogram equalised to remove the first-order statistical artifacts described earlier. Before considering any corrective action, it is pertinent to investigate the sensitivity of the various colour co-ordinate systems to illumination changes. Accordingly, in addition to the test images of Figure A.1, further samples were synthetically generated with different mean intensity values. This was achieved by altering the gamma value of the colour image. Gamma values of 0.8, 0.9, 1.0, 1.1 and 1.2 were used. Gamma value of 1.0 corresponds to the original image. Gamma value of less than 1.0 dims an image and greater than 1.0 brightens it. These effectively set a strict and difficult working environment for the chromatic components to operate in.

The classification results of this new set of colour texture using the same feature sets as above is shown in Table 4.6. In general, classification rate decreases with the introduction of the illumination factor. As expected, the two non invariant feature sets do not perform well at all. For example the $R_N G_N B_N$-LVDCT and the XYZ-LVDCT feature sets managed a classification rate of only 94.83% and 92.71% respectively. As all the components in this feature set are intensity related, it is not surprising at all that its performance is the worst among the features tested. Apart from these feature sets, the performance of the remaining features is still quite good ranging from 95.98% to 98.85%. The normalised $I_{norm} rg$-LVDCT feature set is the most robust of all in terms of exhibiting a degree of illumination invariance. It only shows a drop of 0.04%.

It had been demonstrated that colour conveys complementary information which helps to improve texture discriminability. However, in certain cases when two different colour textures
possess a similar texture form but differ only slightly in their colour composition (or intensity), the intensity related *multi-band colour texture* features will be unable to discriminate them as these features are more sensitive to texture differences rather than colour. However the straightforward extension of the grey level texture representation (LVDCT) can be further modified. In particular it has been shown in [41] that colour information can be retained by means of computing the colour characteristics (mean RGB values) of specific class of textons and using them as features in a second stage of classification.

Kender [23] reported two problems that are associated with the nonlinear transformation defining the saturation, hue, and normalised co-ordinate systems:

- These transformations have non removable singularities, near which a small perturbation of the input tristimulus values can cause a large jump in the transformed values.
- The distribution of the non-linearly transformed values can show spurious modes and gaps.

For these reasons, the use of these nonlinear transformations should be avoided if possible.

The *multi-band colour texture* features can be applied to both the supervised and unsupervised segmentation of texture images. The speed of feature extraction and class decision making could be increased by using a hierarchical (or pyramidal) processing [6][10][44]. To cope with large feature dimensionality, Unser [44] suggested a feature reduction technique which simultaneously diagonalises the scatter matrices evaluated at two spatial resolutions.

### 4.11 Conclusion

In this chapter the role of colour in automatic recognition of colour texture was investigated. This involved the search for an appropriate colour co-ordinate system to provide texture representation. A *multi-band colour texture* analysis scheme was proposed which requires a two stage transformation process. The input tristimulus values are first transformed into the various colour co-ordinate spaces. Next texture attributes are computed on these colour component images and used as colour texture features.

A difficult colour granite classification test was conducted and it was found that colour information does play a very significant role in the discrimination of colour texture. The grey level LVDCT texture representation (85.52%) was outperformed by the *multi-band colour texture* features in the form of YIQ-LVDCT and I₁I₂I₃-LVDCT, both of which achieved a 100% correct classification.
The colour texture classification exercise here was first performed under a controlled environment, the illumination intensity of the light source being fixed during the scanning process. There are many real applications when illumination can be controlled and therein the proposed method is pertinent. If illumination intensity does fluctuate, as often happens in outdoor situations, then the use of features computed from the invariant chromatic component images such as the normalised r and g components could be recommended. The histogram equalised intensity image could also be used to possibly improve the texture aspect of the representation. Some of these chromatic components have been shown to possess a varying degree of resilience to changes in illumination intensity and thus prove useful under these circumstances.

Despite the good performance of the multi-band colour texture features, one disadvantage is the inherent high dimensionality of the feature space associated with this model. Computing texture attributes on the three component images may result in the creation of redundant features or features that are highly correlated. In the next chapter, a novel way of colour texture representation will be introduced. The objective is to develop a new form of colour texture representation that will seek to eliminate some of the problems associated with the multi-band colour texture feature extraction technique. This will include the development and testing of new colour features that have better sensitivity towards illumination changes and have the added advantage of a lower dimensionality in the feature space.

References


Chapter 5

Separable Colour and Texture Feature Description

5.1 Introduction

In Chapter 4, the multi-band colour texture feature extraction technique for colour texture representation was presented. Several observations could be made about this form of colour texture analysis. Firstly, some of the colour texture features computed in this manner are not invariant to illumination intensity changes. Secondly, the performance of the invariant features under various light intensities is quite good but still not good enough for these features to be used under such conditions. Stability of the feature values is extremely vital in many applications where control of ambient lighting intensity is not possible. However if such control is possible for example in an indoor environment, then these colour texture features are most suitable. The main advantage of this method of colour texture representation is that the computation time for feature extraction can be readily reduced using specialised DSP hardware which makes real-time application a possibility. The extra ability of using multi-resolution techniques to estimate the local variances of the filtered channels allows this feature representation to harness the advantages of this particular type of segmentation method [23]. However, one main drawback of using these features is that they contain some redundant information as the colour channels are sometimes highly correlated. This is reflected by the high feature dimension needed for this form of colour texture representation (i.e. 24 features were used).

The objective of this chapter is to develop a new form of colour texture representation that will seek to eliminate some of the problems associated with the former technique (i.e. effect of illumination changes and high dimensionality in feature space). This new form of representation differs
from the earlier one as spatial (i.e. texture) and spectral (i.e. colour) attributes will be treated as independent sources of information. Information about texture can be determined by an existing well proven grey level texture representation technique reviewed in Chapter 2. The development of a robust method of colour representation for texture is the main contribution of this chapter. The colour histogram is used as an image model from which this colour representation can be derived. Its choice is appropriate as a great deal is understood about the behaviour of the colour histogram under changes in the ambient conditions (i.e. changes in illumination intensity, effect of highlights and shadows etc.) [2]. The colour texture representation advocated in this chapter is then obtained by a linear combination of these two separate representations of texture and colour attributes. This results in a more compact form of colour texture description having a smaller feature space dimension then that obtained by the multi-band colour texture feature extraction representation. Figure 5.1 shows such a colour texture analysis scheme.

Features computed from colour histograms in its entirety are computationally intensive to realise even on current machine vision hardware. For example, in order to achieve illumination invariance, one has to resort to correlation of these histograms in a 3-dimensional space; a computationally gruesome task. Even simple invariance properties, such as energy, entropy, and variance computed from such histograms are computationally prohibitive.

A new way of reducing the computational burden when working with 3-dimensional colour histograms will be introduced in this chapter. The essence of this approach is to approximate the third-order distribution function (colour histogram) by a product of three first-order distribution functions estimated along the approximate principal axes of the third-order distribution function. This is followed by statistical descriptions of these first-order histograms in order to arrive at the illumination invariant features. We shall show that this approximation results in a simultaneous reduction of the memory storage and feature extraction time. However, this approximation can only be used subject to the following assumption: that the distribution of the tristimulus values (red, green, and blue) in a colour texture phenomenon is normally distributed and that it consists of a mixture of similarly oriented densities. The validity of this assumption at least for the real colour texture data used in our study will be supported by experimental evidence.

The performance of these invariant features was tested on a difficult colour granite classification exercise. It was found that features computed using the approximated first-order histograms performed better then colour features computed directly from the 3-dimensional histogram. This is found to be attributed to the fact that any form of description using a set of higher dimensional
features (i.e. first-order colour features) has a higher degree of freedom and hence is likely to perform better than when only one feature is used on its own.

Section 5.2 introduces the colour histogram model and presents interesting work in colorimetry that uses such a model. The computational complexity of dealing with colour histogram will be discussed in Section 5.3. Section 5.4 addresses the need of dimension reduction of the colour histogram in order to make its use suitable for computer vision work. A novel technique for approximating the colour histogram will be introduced in Section 5.5. This results in the development of first-order histograms which is the subject of Section 5.6. A set of illumination invariant colour features derivable from the first-order histograms is defined in Section 5.7. Section 5.8 investigates the aptness of using this approximation on real colour texture data. Experiments conducted to test the validity of the approximation and the performance of these invariant features on illumination invariance and colour granite classification as well as cover isolation involving Landsat images will be conducted in Section 5.9. Section 5.10 discusses some of the interesting findings of the experimental work. The conclusions of this chapter can be found in Section 5.11.

5.2 The Colour Histogram

Physical models for light-surface interaction and for sensing are crucial in developing algorithms for an effective colour representation. In colour image sensing using a multi sensor device, the
spectral space is a medium providing a direct transformation of the measured physical phenomenon into a measurement or sensor output space. When a colour camera or a tri-stimuli scanner is used, this spectral space is represented by the quantum catches of the primary sensors (i.e. red, green, and blue). This colour space model will provide sufficient information necessary for the full description of the properties of both the ambient light and the surface reflectance of an object.

Bajcsy et al. [2][8] used the S space model (i.e. scatter plot of the sensor values) to detect specularities from Lambertian reflections using multiple colour images from different viewing directions. Maloney and Wandell [10] used the same model in estimating the surface reflectance function of the objects in a scene with incomplete knowledge of the spectral power distribution of the ambient light. Shafer [14] developed a general model of reflection, based on the S space model which states that two distinct types of reflection - interface and body reflection - occur, and that each type can be decomposed into a relative spectral distribution and a geometric scale factor. Swain and Ballard [15] used colour histograms to index into a large database of models for recognition and searching purposes. Histograms are invariant to translation and rotation about an axis perpendicular to the image plane, and change only slowly under changes of angle of view, change in scale and occlusion, hence their usefulness in situations where such parameter changes are predominant.

Colour texture is usually constituted by a spatial arrangement of a few classes of pigment and such intrinsic characteristic is often adequately portrayed in the colour histogram. A colour texture histogram normally possesses several peaks each corresponding to one of the pigmentation components of the texture. Each peak is associated with a cluster of pixels which are effectively observations on the corresponding pigment component. The groups of clusters corresponding to these peaks are usually unique for different colour texture phenomena. Colour signatures derived from such histograms are therefore suitable for representing colour in texture.

A 3-dimensional histogram, \( H(r, g, b) \), can be generated by observing the number of occurrences of the red, green, and blue sensor values of a colour image. The normalised histogram is simply \( P(r, g, b) = H(r, g, b)/N \), where \( N \) is the total number of pixels used to generate the histogram.
5.3 Computational Complexity of the Colour Histogram

The high memory storage and the time taken to compute the colour histograms make them extremely unattractive to implement in practice. If the colour image is digitised to 24-bit precision per pixel, then the corresponding size of the histogram will be \(2^{24} = 16.8\) million bins. The production of this large amount of data especially during execution time pushes the stack spaces (i.e. RAM) of computers to the limit. The need to store in core memory, all the unique histograms to form a database [15] could prove to be another major problem unless an efficient coding of the histograms is performed.

Pattern matching of histograms is useful when a quantitative measure of the goodness of fit between two histograms is needed. The maximum likelihood classifier can be used in such cases [24]. The complexity and hence speed of this form of classifier increases dramatically when the dimensionality of the histogram increases, like in the case with colour histograms. In situations where the illumination invariant properties are important and when colour histograms are used in their entirety, one needs to correlate these histograms in the 3-dimensional space which is a computationally gruesome task. These factors make the application of colour histogram processing in current vision systems very difficult to realise in practice.

5.4 Dimension Reduction

The primary purpose of reducing the dimensionality of the colour histogram stems from engineering and practical considerations. Features described at a lower dimensional space must contain enough discriminatory information for the classifier to perform reasonably well. The spin-offs of dimensionality reduction include the following:

- Simplification of the classifier design and implementation.
- Minimization of storage capacity.
- Time for and ultimately cost of feature extraction is reduced.

Dimensionality reduction in the case of colour histogram can be achieved in two different ways. Firstly, the colour image can be re-quantised into a smaller number of grey levels across the RGB channels which will result in a histogram of significantly smaller dimension. The understanding of the nature of the blue cones in the human vision system which is found to be less sensitive than the other cones [4][13] would instinctively suggest smaller quantisation level could be
allocated to the blue channel. This decreases further the eventual size of the histogram. Secondly, the 3-dimensional histogram can be approximated very closely by a set of three 1-dimensional histograms which are defined along the principal components of the image data [17]. In this chapter the latter approach will be advocated as it is desirable to maintain full colour resolution of the image.

5.5 Approximating the Colour Histogram

Several assumptions about the distribution of the tristimulus values of colour textures have to be made in order for the following approximation to be possible, namely

- that these variables are normally distributed.
- the distribution consists of a mixture of densities corresponding to the pigmentation of the texture.
- there exist a unique set of principal component axes common to all the constituent components of the mixture density.

The need for these is important and will be justified in the later part of this chapter.

5.5.1 Optimal Principal Axes

Consider the three random variables \( r, g, \) and \( b \) which are associated with the grey levels of the three colour channels. Typically, these tristimulus values are highly correlated with one another [13]. In the development of an efficient processing technique for colour images, it is often desired to work with components that are un-correlated. If the covariance matrix of the tristimulus values are known, then it is possible to determine an orthogonal coordinate system, in which the components are un-correlated, by a Karhunen-Loeve transformation of the tristimulus values. The true covariance matrix of this triplet of random variables can be written as:

\[
\Sigma^o = \begin{bmatrix}
\sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\
\rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\
\rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \\
\end{bmatrix} = \begin{bmatrix}
c_{11} & c_{12} & c_{13} \\
c_{21} & c_{22} & c_{23} \\
c_{31} & c_{32} & c_{33} \\
\end{bmatrix}
\]

(5.1)

The eigen solution of matrix \( \Sigma^o \) satisfies the following relation

\[
\begin{bmatrix}
\lambda_1^o & 0 & 0 \\
0 & \lambda_2^o & 0 \\
0 & 0 & \lambda_3^o \\
\end{bmatrix}
\]

(5.2)
where the transformation matrix with general term $u^*_j$ is composed of the eigenvectors of the covariance matrix, $\Sigma$ and $\lambda_1^*, \lambda_2^*$ and $\lambda_3^*$ (where $\lambda_1^* > \lambda_2^* > \lambda_3^*$) are the eigenvalues of the covariance matrix. The eigenvectors determine the optimal principal axes. The Karhunen-Loeve colour transform space, is defined as

$$e^* = U^* \begin{bmatrix} r \\ g \\ b \end{bmatrix}$$

(5.3)

The transformed values $e_1$, $e_2$ and $e_3$ are un-correlated and have the following properties:

$$\text{Var}\{e_1\} = \lambda_1^*; \quad \text{Var}\{e_2\} = \lambda_2^*; \quad \text{Var}\{e_3\} = \lambda_3^*$$  

(5.4)

$$\text{Covar}\{e_1 \cdot e_2\} = \text{Covar}\{e_1 \cdot e_3\} = \text{Covar}\{e_2 \cdot e_3\} = 0$$  

(5.5)

The orthogonality of the eigenvectors is a nice property, because it ensures that the partial composition (i.e. the principal component images) are un-correlated aspects of the colour image investigated. The principal components do very often improve the separability of classes, hence their popularity, e.g. in land use classification and in segmentation work involving Landsat images.

### 5.5.2 Sub-Optimal Principal Axes

Suppose now that the colour channels have equal variances, $\sigma^2$ which may be ensured by means of pre-processing. Under the additional assumption that the covariances of the respective channels are equal, the covariance matrix of the same triplet of random variables can be written as [22]:

$$\Sigma^{\sigma^2} = \sigma^2 \cdot \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

(5.6)

Such covariance model has been found appropriate for the colour texture images considered in this work (see later part of this chapter). Similarly, the eigen solution of matrix $\Sigma^{\sigma^2}$ can be computed starting from equation (5.2). In this special case, the eigenvalues are given by

$$\lambda_1^{\sigma^2} = \sigma^2 \cdot (1 + 2\rho); \quad \lambda_2^{\sigma^2} = \lambda_3^{\sigma^2} = \sigma^2 \cdot (1 - \rho)$$

(5.7)

with the eigenvectors equal to

$$u_1^{\sigma^2} = [1, 1, 1]^T; \quad u_2^{\sigma^2} = [1, 0, -1]^T; \quad u_3^{\sigma^2} = [1, -2, 1]^T$$

(5.8)
Note that the eigenvalue \( \sigma^2 \cdot (1 - \rho) \) has two corresponding eigenvectors \( \mathbf{w}_1^o \) and \( \mathbf{w}_2^o \). It is now possible to define a set of transformed random variables in the coordinate system defined by this sub-optimal principal axes of the 3-dimensional probability density function (PDF). The linear transform

\[
k = (r + g + b); \quad l = (r - b); \quad m = (r - 2g + b)
\]

produces three other random variables \( k, l \) and \( m \) which are almost un-correlated and have the following properties:

\[
\begin{align*}
\text{Var}(k) &= \lambda_1^o = \sigma^2 \cdot (1 + \rho) \\
\text{Var}(l) &= \lambda_2^o = \sigma^2 \cdot (1 - \rho) \\
\text{Var}(m) &= \lambda_3^o = \sigma^2 \cdot (1 - \rho) \\
\text{Covar}(k \cdot l) &= \text{Covar}(k \cdot m) = \text{Covar}(l \cdot m) = 0
\end{align*}
\]

The sub-optimal principal axes are similar to the Ohta transformation vectors reported in [12]. This is interesting as Ohta et al. arrived at this transformation by systematically observing the eigenvectors of some colour outdoor scenes and affirming that these eigenvectors are mainly dominated by the vectors \( \left[ \frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right]^T, \left[ \frac{1}{2}, 0, -\frac{1}{2} \right]^T, \) and \( \left[ \frac{1}{4}, \frac{1}{2}, \frac{1}{4} \right]^T \). However, the sub-optimal principal axes obtained here were arrived at by approximating the solution to the Karhunen-Loeve transformation of the colour channels.

### 5.5.3 Approximation of Third-Order Probabilities

For simplicity in notation, the sets of optimal (i.e. \( \{e_1, e_2, e_3\} \)) and sub-optimal (i.e. \( \{k, l, m\} \)) random variables are being replaced by a general set of random variables represented by \( \{t_1, t_2, t_3\} \). When these un-correlated eigen transformed random variables \( t_1, t_2, \) and \( t_3 \) are also independent, the joint probability density function can be computed by

\[
P(r, g, b) = c_o \cdot P(t_1, t_2, t_3) \approx c_o \cdot P_1(t_1) \cdot P_2(t_2) \cdot P_3(t_3) = \hat{P}(r, g, b)
\]

This expression is always true for Gaussian random variables. For arbitrarily distributed random variables, un-correlatedness is a necessary but not sufficient condition for independence. The last equality will therefore not always be satisfied. Nevertheless, the product of the first-order PDF's along the principal axes can still be used as a close approximation of the joint PDF of this.
5.5: Approximating the Colour Histogram

form of distribution. The $c_0$ term in equation (5.14) is a normalisation constant chosen in order to
guarantee that

$$\sum_{r=0}^{N_r-1} \sum_{g=0}^{N_g-1} \sum_{b=0}^{N_b-1} \hat{P}(r, g, b) = 1$$

when the support is finite.

The degree of deviation of the true probability density $P(r, g, b)$ from its approximation $\hat{P}(r, g, b)$

08: Approximating the Colour Histogram

H(P, \hat{P}) = \sum_{r=0}^{N_r-1} \sum_{g=0}^{N_g-1} \sum_{b=0}^{N_b-1} P(r, g, b) \cdot \log\{P(r, g, b)l\hat{P}(r, g, b)\}

(5.17)

Substituting equation (5.14) into equation (5.17) we get

$$H(P, \hat{P}) = \sum_{r=0}^{N_r-1} \sum_{g=0}^{N_g-1} \sum_{b=0}^{N_b-1} P(r, g, b) \cdot \log\{P(r, g, b)l[c_0 \cdot P_1(t_1) \cdot P_2(t_2) \cdot P_3(t_3)]\}$$

(5.18)

This entropy function is a measure of independence of variables $t_1$, $t_2$, and $t_3$. The entropy
measure will be zero when these variables are independent. However, when the independence
assumption is not satisfied, then the entropy function will measure the discrepancy between the
original 3-dimensional histogram and its approximation in terms of the marginal histograms in
the transformed space. Rearranging we get

$$H(P, \hat{P}) = -H - \sum_{r=0}^{N_r-1} \sum_{g=0}^{N_g-1} \sum_{b=0}^{N_b-1} P(r, g, b) \cdot \log\{P_1(t_1) \cdot P_2(t_2) \cdot P_3(t_3)\} - \log\{c_0\}$$

(5.19)

Replacing the first term by $-H$ and changing the $r$, $g$, and $b$ variables of the second term to $t_1,$
$t_2,$ and $t_3$ we have

$$H(P, \hat{P}) = -H - \sum_{t_1 \in A_1} \sum_{t_2 \in A_2} \sum_{t_3 \in A_3} P(t_1, t_2, t_3) \cdot \log\{P_1(t_1) \cdot P_2(t_2) \cdot P_3(t_3)\} - \log\{c_0\}$$

(5.21)

$$= -H - \sum_{t_1 \in A_1} \sum_{t_2 \in A_2} \sum_{t_3 \in A_3} P(t_1, t_2, t_3) \cdot \{\log[P_1(t_1)] + \log[P_2(t_2)] + \log[P_3(t_3)]\} - \log\{c_0\}$$

(5.22)
Chapter 5: Separable Colour and Texture Feature Description

\[ H = -\sum_{r \in \mathbb{R}} \sum_{g \in \mathbb{G}} \sum_{b \in \mathbb{B}} \log[P(r, g, b)] \cdot \log[P(r, g, b)] \]

where \( H_1, H_2, H_3 \) and \( H \) are the entropy measures defined by

\[ H_1 = -\sum_{t_1 \in \mathbb{T}_1} \log[P_1(t_1)] \]
\[ H_2 = -\sum_{t_2 \in \mathbb{T}_2} \log[P_2(t_2)] \]
\[ H_3 = -\sum_{t_3 \in \mathbb{T}_3} \log[P_3(t_3)] \]
\[ H = -\sum_{r=0}^{N_r-1} \sum_{g=0}^{N_g-1} \sum_{b=0}^{N_b-1} \log[P(r, g, b)] \cdot \log[P(r, g, b)] \]

The practical interest of the relative entropy measure is that it is possible to test the hypothesis of the approximation by simply comparing the first-order distribution function entropies to the entropy of the associated joint density function.

One drawback concerning the use of such a measure is the lack of quantitative contribution when the relative entropy measure is non zero. In other words, this measure fails to express the degree of mismatch or match other than to validate that the approximation is an exact one (i.e. \( H(P, \hat{P}) \approx 0 \)). It is understood that the conditions leading to the approximation of the joint PDF by the product of the first-order PDF's along the principal axes cannot be fully satisfied in most cases with real data, hence the ineffectiveness of the relative entropy measure to quantify these discrepancies.

Alternatively, one can resort to a quality measure that gauges the degree of match between the observed joint probability density function \( P(r, g, b) \) and the estimated one \( \hat{P}(r, g, b) \). This quality measure can be accomplished by visually comparing the shape, height and position of the observed and estimated density functions respectively. The disadvantage of this technique is that it does not provide a numerical value from which two matches can be compared and contrasted.

5.6 First-Order Histograms

In colour processing of an image, the ability to decompose the third-order PDFs (i.e. colour histograms) into a product of three first-order PDFs of the transformed random variables is highly...
efficient and desirable. The disadvantages pointed out earlier regarding the processing of the
3-dimensional colour histograms will obviously be eliminated at the expense of introducing po­
tential approximation errors. The two sets of first-order histograms, \( \{H_1^o(e_1), H_2^o(e_2), H_3^o(e_3)\} \) and
\( \{H_1^o(k), H_2^o(l), H_3^o(m)\} \) correspond respectively to the optimal and sub-optimal approximation
of the overall 3-dimensional colour histogram. The corresponding normalised first-order his­
tograms are abbreviated \( \{P_1^o(e_1), P_2^o(e_2), P_3^o(e_3)\} \) and \( \{P_1^o(k), P_2^o(l), P_3^o(m)\} \) respectively. It is
obvious that the first principal component histograms, \( H_1^o(e_1) \) and \( H_1^o(k) \) correspond to the his­
tograms of the grey level image. The rest of the histograms represent the first-order statistics of
the chromatic content of the underlying colour textures.

It can be seen that the dynamic ranges of these histograms are different from the range of
the original variables. For example \( 0 \leq k \leq 3(N_g - 1), -(N_g - 1) \leq l \leq N_g - 1 \) and \(-2(N_g - 1) \leq m \leq
2(N_g - 1). \) The ranges of the components of the optimal principal axes depend on the orientation
of the corresponding eigenvectors.
5.7 Colour Features from First-Order Histograms

The notion of invariance is important in feature measurements as it ensures stability in representation over a wide range of operating conditions. Here, the aim is to seek illumination invariance property of the colour histogram. When the illumination intensity changes, a cluster of points in the colour histogram representing a distinct colour texture either shifts up or down along the axis between the origin and the cluster centre (see Figure 5.2). It is understood that the form of the distribution does not change. Only the positioning of it changes (i.e. mean level). Given this knowledge, it is possible to derive illumination invariant measurements\[17\]. Firstly, the colour histogram in its entirety can be matched with a reference model histogram in a database in a supervised environment. The degree of match between histograms can be measured by a correlation measure, but unfortunately this measure is computationally very burdensome (e.g. a 3-dimensional correlation).

Invariance property at lower dimensional feature space is possible by resorting to simple statistical descriptions of the histogram characteristics \[5\]. These descriptions must preserve, as much as possible, the salient information of the original histogram with the added advantage of a more compact and compressed description of the histogram.

Among the invariant colour features that can be derived from the colour histogram are the Energy, Entropy, Variance and Covariance features. The following shows the computation and description of these colour features.

(a). Energy:

\[
E = \sum_{r=0}^{N_r-1} \sum_{g=0}^{N_g-1} \sum_{b=0}^{N_b-1} [P(r,g,b)]^2
\]  \hspace{1cm} (5.29)

This is a measure of homogeneity of the histogram and it attains large values for histograms with few entries of large magnitude. Smaller measure corresponds to the histogram with a large number of small entries.

(b). Entropy:

\[
H = - \sum_{r=0}^{N_r-1} \sum_{g=0}^{N_g-1} \sum_{b=0}^{N_b-1} P(r,g,b) \log[P(r,g,b)]
\]  \hspace{1cm} (5.30)

This measures the information content of the histogram. A histogram will have a large or small entropy according to the unevenness of the probability density.
5.7: Colour Features from First-Order Histograms

(c). Variance and Covariance (Covariance Matrix):

\[ \Sigma = \sum_{r=0}^{N_r-1} \sum_{g=0}^{N_g-1} \sum_{b=0}^{N_b-1} (x - \mu) \cdot (x - \mu)^T \cdot P(r,g,b) \]  
\[ \mu_r = \sum_{r=0}^{N_r-1} \sum_{g=0}^{N_g-1} \sum_{b=0}^{N_b-1} r \cdot P(r,g,b) \]  
\[ \mu_g = \sum_{r=0}^{N_r-1} \sum_{g=0}^{N_g-1} \sum_{b=0}^{N_b-1} g \cdot P(r,g,b) \]  
\[ \mu_b = \sum_{r=0}^{N_r-1} \sum_{g=0}^{N_g-1} \sum_{b=0}^{N_b-1} b \cdot P(r,g,b) \]

where \( x = [r, g, b]^T \) and \( \mu = [\mu_r, \mu_g, \mu_b]^T \). The Variance and Covariance features can be fully described by the covariance matrix, \( \Sigma \). They are measures of the spread of the distribution in every axis about the corresponding mean.

All the above invariant features can be estimated in terms of the corresponding quantities computed from 1-dimensional probability density functions along the principal axis. These features will be termed as the first-order colour features. Substituting equation (5.29) into equation (5.14) one can estimate the energy feature as

\[ E \approx c_0 \cdot \{ \sum_{t_1 \in A_1} P_1(t_1)^2 \times \sum_{t_2 \in A_2} P_2(t_2)^2 \times \sum_{t_3 \in A_3} P_3(t_3)^2 \} \]  
\[ = c_0 \cdot \{E_1 \times E_2 \times E_3 \} \]  

where \( A_i \) is the finite support of the \( i \)th first-order normalised histogram. Similarly, the entropy feature of equation (5.30) can be re-written as

\[ H \approx c_0 \cdot \{ -\sum_{t_1 \in A_1} P_1(t_1) \log[P_1(t_1)] - \sum_{t_2 \in A_2} P_2(t_2) \log[P_2(t_2)] - \sum_{t_3 \in A_3} P_3(t_3) \log[P_3(t_3)] \} - c_0 \log c_0 \]  
\[ = c_0 \cdot \{H_1 + H_2 + H_3 \} - c_0 \log c_0 \]  

Since the optimal principal axes and sub-optimal principal axes respectively diagonalise fully and approximately the covariance matrix of the original colour space, the covariance term in equation (5.31) needs not be computed. Only the variance features will be non-zero and can be estimated from the 1-dimensional probability density functions:

\[ \text{Var}_1 = \sum_{t_1 \in A_1} (t_1 - \mu_1)^2 \cdot P_1(t_1) \]  
\[ \text{Var}_2 = \sum_{t_2 \in A_2} (t_2 - \mu_2)^2 \cdot P_2(t_2) \]
This approximation of the colour histogram and features resulted in the simultaneous reduction of both the memory storage and feature extraction time. With this new formulation, the memory needed to store the three first-order histograms is approximately only $9N_g$ (where $N_g$ is the total number of grey-levels in each channel) as compared to $N^3_g$ for the 3-dimensional model; a saving of $N^2_g/9$. The feature extraction time is also significantly reduced as a triple summation is being replaced by three single summations. For example consider an image with 256 quantisation levels per channel. The use of the proposed method will make it possible to compute these features 7281 times faster using 7281 times less memory as compared with the evaluation based on the 3-dimensional colour histogram.

### 5.8 Aptness of Approximation

The strength of the previously developed technique to reduce the complexity of the 3-dimensional colour histogram solely relies on one very significant assumption, namely that the distribution of the tristimulus values of a colour texture phenomenon consists of a mixture of similarly oriented normal densities. In this section, an experimental setting will be developed to test the validity of this assumption. This is based on the Estimation-Maximization (EM) approach for estimating and fitting a parametric model to a set of data. The model that will be used is a mixture of multi-normal density functions which is believed to fit the colour texture data.

#### 5.8.1 Estimation-Maximization Approach to Gaussian Mixture Decomposition

Frequently in statistical pattern classification, one may encounter feature vectors from the same class which are bimodally or multi-modally distributed. The estimation of the parameters of this form of probability density model which consists of a linear mixture of several multi-normal distributions [20] will help to simplify the feature selection process and the construction of the classifier. Consequently, it reduces the classification time. One of the most powerful techniques of
5.8: Aptness of Approximation

Parameter estimation is the Estimation-Maximization (EM) algorithm. This algorithm possesses a number of attractive properties:

- The algorithm is guaranteed to converge to a local optimum.
- Low cost per iteration and economy of storage.

If the feature vector is replaced by the colour vector with three elements corresponding to the red, green, and blue values then the EM algorithm can be used to estimate the multi-normal third-order probability density function of the colour image.

Text et al. [19] introduced a working EM algorithm for a linear mixture of two multi-normal functions. To obtain numerical estimates of the parameters of a linear mixture model, they followed Bezdek et al. [3] and Hjort [6] and derived iterative equations of the linear mixture of two multi-normal functions using the EM approach to maximize the log likelihood. Let $N_d(\mu, \Sigma)(x)$ denote the multi-normal density, as a function of a $d$-dimensional $x$, with a mean vector $\mu$ and covariance matrix $\Sigma$. The generalised linear mixture model of a feature vector $x$ can be defined as:

$$ f(x) = \sum_i n_i f_i(x) $$

where

$$ \sum_i n_i = 1 $$

and $f_i(x)$ is the $i^{th}$ multi-normal component of the mixture density $f(x)$, i.e. $f_i(x) = N_d(\mu_i, \Sigma_i)(x)$. The value of the weights $n_i$ must be between zero and unity.

Let the observed feature vectors of a particular class $i$ be $x_1, ..., x_n$ and define

$$ Q(j)(i|x) = n_j^{(j)} f_i(x) f(j)(x) $$

where

$$ \sum_i Q(j)(i|x) = 1 $$

The log likelihood for this linear mixture model is given by

$$ \log L = \sum_{j=1}^n \log \{ n_j f_i(x_j) \} $$

Now defining

$$ n_0^{(j)}(i) = \sum_{j=1}^n Q(j)(i|x_j) $$

and

$$ n = \sum_i n_0^{(j)}(i), $$

(5.45)

(5.46)

(5.47)

(5.48)

(5.49)

(5.50)

(5.51)
the following iterative EM equations can be derived by setting the partial derivatives of the log-likelihood equal to zero and solving with respect to the parameters included [19]:

\[
\begin{align*}
n_i^{(t+1)} &= \frac{n_i^{(t)}(i)/n}{n} \\
\mu_i^{(t+1)} &= \frac{1}{n_i^{(t)}(i)} \sum_{j=1}^{n} Q^{(t)}(i|x_j)x_j \\
\Sigma_i^{(t+1)} &= \frac{1}{n_i^{(t)}(i)} \sum_{j=1}^{n} Q^{(t)}(i|x_j)(x_j - \mu_i^{(t+1)})(x_j - \mu_i^{(t+1)})^T
\end{align*}
\]

These iterative equations could be derived for two or more multi-normal functions. The estimation of the third-order probability density consisting of a mixture of two or more multi-normal functions in the red, green, and blue channel space describing the colour texture can be achieved in a similar fashion.

### 5.8.2 The Estimated Mixture Density

This section is devoted to the testing of the main assumption behind the histogram representation method, namely that the distribution of the tristimulus values in a colour texture is normally distributed and that it consists of a mixture of similarly oriented densities. Before the EM algorithm can operate, some good starting values of the parameters have to be estimated. This can be done by first clustering the colour data into an appropriate number of clusters using the \(k\)-means clustering algorithm [1]. This is followed by computing the mean vector and covariance matrices of each cluster. These statistics could then be used as the starting estimates for the EM algorithm to converge rapidly to at least a local optimum. The number of clusters to be used in this algorithm must be as small as possible to ensure fast convergence. A look at the sample colour texture images used here suggests the value of \(k\) (= 3) to be a good choice.

For every colour texture data tested, the EM algorithm will provide a final statistical description of the mixture in terms of three multi-normal density functions. With this information, the complete estimated density function that fits the model could be realised. Finally the demonstration of the validity of the assumption will be attempted by comparing the estimated and the observed density functions for a closeness of match. The method used for the comparison between densities will be a visual one; namely the form and structure of the 3-dimensional scatter plot generated from the observed colour texture data (i.e. \(S(r, g, b)\)) with the scatter plot from the estimated density function (i.e. \(\hat{S}(r, g, b)\)) will be compared.

The actual and estimated scatter plots of several colour texture data are shown in Figure 5.3.
It can be seen from these plots that $S(r, g, b)$ is often a good estimate of $S(r, g, b)$ for the colour texture images tested. Two corresponding scatter plots can be said to be fully matched if their shapes, positions and densities are similar. It can be seen from the plots of Figure 5.3 that the first two attributes are in agreement. It is not possible to use the scatter plot as a means to match the density of the plots. The set of marginal probability density functions in the form of $P(r, g), P(r, b),$ and $P(g, b)$ is used to gauge the level of match between the densities. Figure 5.4 shows the set of three estimated and observed marginal density functions of the “bianco castilla” colour granite image. Again the densities match reasonably well. This is found to be consistent when tested on the rest of the colour texture images.

5.8.3 The Estimated Density Principal Axes

As far as the approximation is concerned, being normally distributed (the mixture density) is not sufficient to guarantee the simplified representation that leads to equation (5.14). It was mentioned that the global optimal principal axis and sub-optimal principal axis transformation of the normally distributed tristimulus values will de-correlate them. As colour texture data consists of a mixture of such densities, one more condition has to be satisfied in order for equation (5.14) to be valid; that the principal component axes of the constituent densities must be aligned closely with the global optimal principal axes or sub-optimal principal axes. Figure 5.5 gives an example in 2-dimensional space illustrating clearly the need for this requirement.

The purpose of this Section is to determine whether real colour texture data meets this particular requirement. The way to go about determining this can be summarised in these few steps:

(a). compute the estimates of the mean ($\mu_i$) and covariance matrices ($\Sigma_i$) of the three constituent densities of the mixture density using the EM algorithm.

(b). for the $i^{th}$ covariance matrix, compute its eigen solutions to obtain the three principal axes of the estimated density component represented by vectors $u_{ij}$, where the second index represents the $j^{th}$ principal axis of the component density.

(c). compute the global optimal principal axes represented by vectors $e_j$.

(d). for each principal axis ($j$) of every component density ($i$), compute the angles (abbreviated $\theta_{ij}$) between $u_{ij}$ and $e_j$.

Repeat the above using the sub-optimal principal axes in step (c) to check if the above mentioned requirement is met by the sub-optimal axes. The angles computed in step (d) reveal how the
Figure 5.3: continue...
Figure 5.3: 3-dimensional scatter of the observed, $S(r,g,b)$ and estimated, $\hat{S}(r,g,b)$ distribution of the tristimulus values of some colour texture data. The corresponding scatter plots are plotted with axes of the same scale.

* The entropy measure value defined in equation (5.17) is shown in brackets in the subcaption of the scatter plot for each colour texture data.
Figure 5.4: Marginal distribution of the "bianco castilla" colour granite image (Right Plot: Observed distribution, Left Plot: Estimated distribution).
5.8: Aptness of Approximation

Figure 5.5: (a) Mixture density with the same principal component axes can be effectively decorrelated by the eigenvectors $e_1$ and $e_2$ respectively. (b) This is simply not the case when the mixture has constituent densities with different principal component axes. Data transformation by eigenvectors $e_1$ and $e_2$ does not in this case guarantee un-correlatedness.

global optimal principal axes (or sub-optimal principal axes) can be used to substitute the different principal axes of the constituent densities. Small angles signify a good approximation. Table 5.1 shows the values of these angles computed using the optimal principal axes. The majority of them are below 19°, signifying a good approximation. Likewise Table 5.2 shows these angles using the sub-optimal principal axes. Except for the “marron guaiba” and “new imperial” data, these angles are small with the majority of them below 24°. The exceptional case of the two colour texture data can be explained by analysing the covariance matrix of the data. It was noted that the covariance elements of the matrix are very different, which contradicts the pre-requisite for equal covariances in the case of the sub-optimal principal transformation, hence the discrepancy.
Table 5.1: Angles between the $j^{th}$ global optimal principal axis and the $j^{th}$ principal component axis of the $i^{th}$ constituent of the mixture density.

<table>
<thead>
<tr>
<th>Material</th>
<th>$\theta_{11}$</th>
<th>$\theta_{12}$</th>
<th>$\theta_{13}$</th>
<th>$\theta_{21}$</th>
<th>$\theta_{22}$</th>
<th>$\theta_{23}$</th>
<th>$\theta_{31}$</th>
<th>$\theta_{32}$</th>
<th>$\theta_{33}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>bianco castilla</td>
<td>9.3</td>
<td>9.1</td>
<td>2.3</td>
<td>1.6</td>
<td>6.2</td>
<td>6.1</td>
<td>2.3</td>
<td>2.7</td>
<td>2.0</td>
</tr>
<tr>
<td>capao bonito</td>
<td>9.4</td>
<td>12.0</td>
<td>7.4</td>
<td>5.3</td>
<td>10.0</td>
<td>8.6</td>
<td>7.2</td>
<td>8.8</td>
<td>4.9</td>
</tr>
<tr>
<td>marron guaiba</td>
<td>7.3</td>
<td>10.4</td>
<td>12.0</td>
<td>5.4</td>
<td>5.9</td>
<td>2.3</td>
<td>4.1</td>
<td>6.2</td>
<td>4.7</td>
</tr>
<tr>
<td>new imperial</td>
<td>1.2</td>
<td>5.0</td>
<td>4.8</td>
<td>9.1</td>
<td>11.6</td>
<td>7.2</td>
<td>7.5</td>
<td>11.1</td>
<td>8.5</td>
</tr>
<tr>
<td>santiago</td>
<td>7.9</td>
<td>8.9</td>
<td>4.8</td>
<td>3.6</td>
<td>8.0</td>
<td>7.7</td>
<td>1.5</td>
<td>5.5</td>
<td>5.2</td>
</tr>
<tr>
<td>rosa baveno</td>
<td>4.0</td>
<td>4.2</td>
<td>1.2</td>
<td>11.7</td>
<td>11.7</td>
<td>0.7</td>
<td>1.1</td>
<td>4.1</td>
<td>4.0</td>
</tr>
<tr>
<td>balmoral GF</td>
<td>5.9</td>
<td>6.9</td>
<td>3.6</td>
<td>16.5</td>
<td>16.8</td>
<td>3.0</td>
<td>3.6</td>
<td>7.0</td>
<td>6.0</td>
</tr>
<tr>
<td>rosa monforte</td>
<td>5.1</td>
<td>5.3</td>
<td>1.2</td>
<td>9.1</td>
<td>9.5</td>
<td>3.4</td>
<td>2.2</td>
<td>9.7</td>
<td>9.8</td>
</tr>
<tr>
<td>rosso perla</td>
<td>3.8</td>
<td>13.7</td>
<td>13.2</td>
<td>3.8</td>
<td>8.9</td>
<td>9.1</td>
<td>1.1</td>
<td>18.7</td>
<td>18.7</td>
</tr>
<tr>
<td>cardinal red</td>
<td>1.0</td>
<td>3.3</td>
<td>3.2</td>
<td>10.8</td>
<td>10.5</td>
<td>3.3</td>
<td>4.5</td>
<td>11.8</td>
<td>11.0</td>
</tr>
<tr>
<td>brasil violett</td>
<td>4.5</td>
<td>10.3</td>
<td>9.3</td>
<td>3.6</td>
<td>3.6</td>
<td>0.4</td>
<td>7.7</td>
<td>9.4</td>
<td>5.8</td>
</tr>
<tr>
<td>ghiandonato</td>
<td>0.4</td>
<td>11.4</td>
<td>11.4</td>
<td>2.6</td>
<td>2.6</td>
<td>1.1</td>
<td>5.5</td>
<td>7.1</td>
<td>4.5</td>
</tr>
</tbody>
</table>

5.8.4 Conclusion

It was shown here that the distribution of the tristimulus values in a colour texture phenomenon of at least some material surfaces such as granite can be modeled by a mixture of multi-normal density functions. Each constituent of this mixture corresponds to the individual colour pigmentation of the underlying texture. As far as a mixture density is concerned, the optimal principal axis and sub-optimal principal axis global transformation will only work if and only if the constituent densities share the same principal component axes. It was shown here that real colour texture data do in fact conform approximately with this condition, hence the effectiveness of the optimal principal axis and sub-optimal principal axis transformation. Therefore the experimental evidence gathered here provides a good justification for the above assumption leading to the simplifying formulation of equation (5.14).

5.9 Experiments

5.9.1 Approximating the Colour Histograms

Subject to the requirements already mentioned, the validity of the use of three 1-dimensional density functions to approximate the 3-dimensional density functions of the colour data will be tested here. Here, the choice of a suitable matching measure between the observed and estimated density functions is important. As mentioned earlier, the measure that is best suited for this purpose
Table 5.2: Angles between the $j^{th}$ global sub-optimal principal axis and the $j^{th}$ principal component axis of the $i^{th}$ constituent of the mixture density.

<table>
<thead>
<tr>
<th></th>
<th>$\theta_{11}^{(i)}$</th>
<th>$\theta_{12}^{(i)}$</th>
<th>$\theta_{13}^{(i)}$</th>
<th>$\theta_{21}^{(i)}$</th>
<th>$\theta_{22}^{(i)}$</th>
<th>$\theta_{23}^{(i)}$</th>
<th>$\theta_{31}^{(i)}$</th>
<th>$\theta_{32}^{(i)}$</th>
<th>$\theta_{33}^{(i)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>bianco castilla</td>
<td>5.4</td>
<td>16.0</td>
<td>15.1</td>
<td>5.1</td>
<td>10.1</td>
<td>9.2</td>
<td>5.5</td>
<td>13.5</td>
<td>13.8</td>
</tr>
<tr>
<td>capao bonito</td>
<td>10.0</td>
<td>12.9</td>
<td>9.8</td>
<td>14.0</td>
<td>15.7</td>
<td>8.8</td>
<td>12.1</td>
<td>24.1</td>
<td>22.0</td>
</tr>
<tr>
<td>marron guaiba</td>
<td>6.1</td>
<td>81.3</td>
<td>81.0</td>
<td>4.7</td>
<td>9.2</td>
<td>8.4</td>
<td>5.7</td>
<td>5.8</td>
<td>1.6</td>
</tr>
<tr>
<td>new imperial</td>
<td>32.1</td>
<td>42.4</td>
<td>27.1</td>
<td>24.8</td>
<td>46.0</td>
<td>38.7</td>
<td>26.7</td>
<td>35.5</td>
<td>23.1</td>
</tr>
<tr>
<td>santiago</td>
<td>4.2</td>
<td>12.9</td>
<td>13.5</td>
<td>3.1</td>
<td>10.6</td>
<td>10.2</td>
<td>7.8</td>
<td>24.1</td>
<td>22.8</td>
</tr>
<tr>
<td>rosa baveno</td>
<td>1.6</td>
<td>2.9</td>
<td>3.3</td>
<td>7.8</td>
<td>8.9</td>
<td>5.1</td>
<td>5.5</td>
<td>5.4</td>
<td>1.4</td>
</tr>
<tr>
<td>balmoral GF</td>
<td>11.1</td>
<td>15.8</td>
<td>11.1</td>
<td>4.4</td>
<td>17.8</td>
<td>18.3</td>
<td>13.3</td>
<td>16.0</td>
<td>8.9</td>
</tr>
<tr>
<td>rosa monforte</td>
<td>3.3</td>
<td>7.8</td>
<td>8.2</td>
<td>5.4</td>
<td>11.8</td>
<td>12.5</td>
<td>7.9</td>
<td>7.9</td>
<td>0.6</td>
</tr>
<tr>
<td>rosso perla</td>
<td>2.4</td>
<td>2.7</td>
<td>2.3</td>
<td>6.7</td>
<td>23.6</td>
<td>24.1</td>
<td>5.0</td>
<td>6.0</td>
<td>4.0</td>
</tr>
<tr>
<td>cardinal red</td>
<td>9.4</td>
<td>16.8</td>
<td>13.9</td>
<td>4.6</td>
<td>8.8</td>
<td>9.7</td>
<td>4.1</td>
<td>4.1</td>
<td>0.9</td>
</tr>
<tr>
<td>brasil violett</td>
<td>4.9</td>
<td>11.9</td>
<td>11.9</td>
<td>7.3</td>
<td>21.0</td>
<td>21.3</td>
<td>6.4</td>
<td>16.3</td>
<td>15.2</td>
</tr>
<tr>
<td>ghiandonato</td>
<td>1.8</td>
<td>13.2</td>
<td>13.3</td>
<td>4.2</td>
<td>3.4</td>
<td>3.0</td>
<td>5.3</td>
<td>7.9</td>
<td>6.5</td>
</tr>
</tbody>
</table>

is the quality measure that visually gauges the level of match between the observed and estimated density functions.

The test images selected here are: a single mode colour Gaussian noise image, a bimodal colour Gaussian noise image, and a real colour granite image. Both the optimal principal axis and sub-optimal principal axis transformation will be used here to produce the transformed variable set \{ $e_1, e_2, e_3$ \} and \{ $k, l, m$ \} respectively.

5.9.1.1 Simulation One

The example image used here is a simple Gaussian noise that is being extended to all three bands to produce a colour noise image. The mean grey value for all bands is 128, with variance of 20. The aim here is to test the validity of the approximation of the colour histogram using this simple example image.

In order to make the visual comparison possible on a three dimensional graphical plot, the marginal densities were computed for both the observed and the estimated density functions. The marginal density chosen here for illustration is $P(r, g)$. The estimated marginal density $\hat{P}(r, g)$ can be computed from:

$$\hat{P}(r, g) = \sum_{k=0}^{N_s-1} c_k \cdot P_1(e_1) \cdot P_2(e_2) \cdot P_3(e_3)$$

for all values of $r$ and $g$ in the case of the optimal principal axis transformation. A similar expression can also be used for the sub-optimal principal axis transformation by replacing the indices.
Chapter 5: Separable Colour and Texture Feature Description

Figure 5.6: Plot of observed and estimated marginal distribution of a single mode Gaussian noise image using the sub-optimal principal axis transformation (Right Plot: Observed, Left Plot: Estimated, Middle Plot: Residual).

of the 1-dimensional distribution with the $k$, $l$, and $m$ variables.

As the data of a single mode Gaussian image are uncorrelated, any set of orthonormal coordinate axes including the sub-optimal axes will be the principal axes. Therefore the sub-optimal principal axis transformation will be used for this particular example. Figure 5.6 shows the plot of the observed and estimated marginal distribution for the single mode Gaussian noise image using the sub-optimal principal axis transformation. To aid in the comparison, the observed density and the estimated density are shown on the same plot. The residual plot (absolute difference between the observed and estimated densities) is shown in the middle of the plot. It can be seen judging from the small values of the residual plot that the approximation in this case is a very good one, therefore substantiating the validity of the approximation.

5.9.1.2 Simulation Two

Next, the approximation is tested on a bimodal Gaussian noise image which has two modes of unequal population. Figure 5.7(a) and Figure 5.7(b) show the plots using the optimal principal axis and sub-optimal principal axis transformation respectively. It can be seen very clearly that the approximation is very good for the optimal principal axis transformation. The sub-optimal
principal axis transformation is able to reproduce the position and shape of the distribution very well. However, it is unable to register the correct probability level.

The property of fully de-correlating the tristimulus values is an important factor that contributed to the excellent performance of the optimal principal axis transformation. As far as the sub-optimal principal axis transformation is concerned, which only approximates the de-correlation requirement, the estimation is poorer but still quite acceptable.

5.9.1.3 Simulation Three

So far the simulations were performed on synthetic colour images. An arbitrary shaped density function derived from a real colour texture image will now be used to test for the validity of the approximation on the joint density function. The associated plots using the optimal principal axis and sub-optimal principal axis transformation on the “bianco castilla” colour granite image are shown in Figure 5.8(a) and Figure 5.8(b) respectively. It can be seen that the approximation is quite good in both cases, the shape and position of the approximated density function was maintained to a fair degree. As in the previous case, the approximated probability cannot fully match the observed one. There seems to be no favourite transformation that will produce a distinctly good approximation in this case, hence both the optimal principal axis and sub-optimal principal axis transformations could be used to de-correlate the colour channels. Given the almost similar performance of the two transformations on the real colour texture images, it can be noted at this point that the sub-optimal principal axis transformation is preferred as the transformation axes are data independent and hence computationally more attractive.

5.9.2 First-Order Colour Feature Invariance Property

This section investigates the illumination invariance property of the first-order colour features. Images of the “rosso perla” granite were taken with a single chip colour camera under different light intensities from 40 Lux up to 75 Lux. This was achieved by varying the voltage supplies to the light sources and using a photo lux meter to measure the intensity of the light emitted. Care was taken not to have too high an intensity lest intensity saturation occurs.

The optimal principal axis first-order colour features were estimated over a 100 × 100 pixels of the granite image taken at different illumination intensities and their normalised values plotted in Figure 5.9. A similar plot showing the behaviour of the sub-optimal principal axis first-order colour features computed over the same set of images is shown in Figure 5.10. To aid in the comparison,
Chapter 5: Separable Colour and Texture Feature Description

Figure 5.7: Plot of observed and estimated marginal distributions of a bimodal Gaussian noise image using (a) optimal principal axis transformation and (b) sub-optimal principal axis transformation (Right Plot: Observed, Left Plot: Estimated, Middle Plot: Residual).
5.9: Experiments

(a) Using the optimal principal axis transformation

(b) Using the sub-optimal principal axis transformation

Figure 5.8: Plot of observed and estimated marginal distributions of the "bianco castilla" image using (a) optimal principal axis transformation and (b) sub-optimal principal axis transformation (Right Plot: Observed, Left Plot: Estimated, Middle Plot: Residual).
Chapter 5: Separable Colour and Texture Feature Description

Plot of First-Order Colour Features vs Intensity Changes

Figure 5.9: Optimal principal axis first-order colour features: Plot of energy, entropy, variance and mean measures with respect to light intensity changes.

The mean value of the first principal axis corresponding to the intensity of the image is also shown. The mean feature can be seen to increase monotonically as the light intensity increases. The energy, entropy and variance features are quite stable over these ranges of light intensity. These features were found to be consistently stable when tested on other granite images.

5.9.3 Colour Texture Classification with First-Order Colour Features

This section presents the experimental evaluation of the first-order colour feature performance in colour representation of texture. For this purpose, 12 colour granites were selected as sample colour texture images (see Figure A.1 of Appendix A) to be classified. The colour texture samples were digitised with a $256 \times 256$ resolution and 24 bit precision. An equivalent grey level image, $I(x,y)$, can be produced using a linear combination of the colour channels [12] as in

$$I(x,y) = 0.299 \cdot \text{Red}(x,y) + 0.587 \cdot \text{Grn}(x,y) + 0.114 \cdot \text{Blu}(x,y)$$  \hspace{1cm} (5.56)

Histogram equalisation, producing an output image of equiprobable grey levels, was then performed on this grey level image to obtain textures with approximately the same first-order statistics. This has the effect of eliminating the variations in grey levels which might have resulted from
For random textures, which will be the main concern here, statistical texture representation methods are generally known to be more appropriate [11],[16],[21]. It was argued in Chapter 3 that the local linear transform is well suited for such application. The texture transformation is easy to implement and capable of retaining most of the image texture information. The linear transform selected here is the Discrete Cosine Transform. A set of nine $3 \times 3$ DCT masks can be generated from three 1-dimensional DCT basis vectors of the form $\{1, 1, 1\}^T$, $\{1, 0, -1\}^T$ and $\{1, -2, 1\}^T$. The mask with low pass property was excluded in forming a set of channel filters [16]. Texture features can then be defined as the local variance of the filter output (LVDCT) computed over a larger window of the histogram-equalised grey level images. The colour information used here are the first-order colour features derived earlier. The overall colour texture feature set is obtained by linearly combining the texture and colour features. The aim here is to investigate the effectiveness of this new set of colour texture features on colour granite classification. The criterion used to measure the performance success of such features will be the percentage of correct classification.

As before, in addition to the test images of Figure A.1, further samples can be synthetically
generated with different mean intensity values. This is achieved by altering the gamma value of the colour image. Gamma values of 0.8, 0.9, 1.0, 1.1 and 1.2 were used. These effectively set a strict and difficult working environment for the first-order colour features. The LVDCT texture features will not be affected much by these inclusions as histogram equalisation of the grey level representation of the colour images will eliminate this fluctuations in intensity.

A supervised Bayesian classification scheme was used and as such some sample prototype textures will be required for training purposes. The top half of the images with a gamma value of 1.0 (original images) is used as the training data. A moving window of $100 \times 100$ pixels is used to compute both the colour and texture features. In order to get sufficient data samples for the classifier, this window slides across the image with a 1/20 overlap to generate 192 samples of subimages per image for training purposes.

Similarly, for the bottom half of the rest of the images, including the gamma corrected images, the same moving window with a 1/5 overlap is used to generate 16 samples per class producing $5 \times 16 = 80$ samples for classification.

Firstly, the granite images were classified based solely on texture information alone. As such the eight texture features were used. The classification accuracy rate using this form of representation was 73.52%. Classification errors occur between classes of granite images that are perceived to be visually very close. For example, the grey level images of granite samples “bianco castilla” and “rosa baveno” both look very similar and are the cause of most classification errors. Next, the study of the effect on the classification error by incorporating the first-order colour features into the overall representation were conducted. Two sets of first-order colour features corresponding to the optimal principal axes and sub-optimal principal axes were used here. Table 5.3 shows the classification accuracy rate when using the set of first-order features with the LVDCT features.

As expected, an overall increase in classification was observed with the inclusion of these colour attributes. As can be seen, most of the results obtained are in the mid or high 90s. Even a classification rate of 100% was achieved for the features comprising LVDCT+$E_{p} + Var_{p}$.

The performance of the optimal principal axis first-order colour features is generally worse than the sub-optimal principal axis first-order features when used in combination with the LVDCT features. At a first glance this is somewhat surprising since the optimal principal axes are data dependent and therefore should provide a better approximation of the colour histogram. However, by the same token as the transformation is data dependent and therefore different for the various colour texture images, the subsequent feature extraction from the optimal first-order histograms will be
5.9: Experiments

<table>
<thead>
<tr>
<th>Features Used</th>
<th>Classification Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>LVDCT</td>
<td>75.52% 75.52%</td>
</tr>
<tr>
<td>LVDCT+Er</td>
<td>95.42% 97.50%</td>
</tr>
<tr>
<td>LVDCT+Hr</td>
<td>95.63% 97.08%</td>
</tr>
<tr>
<td>LVDCT+VarT</td>
<td>97.50% 97.92%</td>
</tr>
<tr>
<td>LVDCT+Er+Hr</td>
<td>98.13% 99.69%</td>
</tr>
<tr>
<td>LVDCT+Er+VarT</td>
<td>97.50% 100.0%</td>
</tr>
<tr>
<td>LVDCT+Hr+VarT</td>
<td>94.17% 98.75%</td>
</tr>
</tbody>
</table>

Table 5.3: Classification results of combining the grey level LVDCT texture feature with the optimal principal axis first-order colour features (OPA Colour) and the sub-optimal principal axis first-order colour features (SOPA Colour). $E_r = \{E_1, E_2, E_3\}$, $H_r = \{H_1, H_2, H_3\}$ and $Var_T = \{Var_1, Var_2, Var_3\}$.

computed at different feature spaces. Features computed in this multi-domain fashion would not be able to harness the optimality of the Bayes' classifier which minimizes the probability of error in only one feature space. However, the sub-optimal principal axis first-order colour features do not suffer from this problem as the sub-optimal principal axes are the same for all colour texture images.

The results obtained with the sub-optimal principal axis first-order colour features are very encouraging since very strict operating conditions had been imposed on the set of colour texture features by incorporating sample images with a wide intensity range. It can be seen that using both texture and colour attributes, it is possible to discriminate very well, the set of colour granite images of varying illumination intensity values. This form of colour texture representation will generally be applicable to other colour textures (i.e. outdoor scenes).

5.9.4 Colour Features computed directly from Colour Histograms

Though computationally intensive, colour features could be computed directly from the colour histograms as described in equation (5.29), (5.30) and (5.31). The aim of this section is to analyse the performance of these colour features on the above colour granite classification exercise. The experimental settings are unchanged in order to eliminate any bias on the results. The size of the energy and entropy feature sets in this case has been reduced to one as they now describe attributes of the full 3-dimensional colour histogram. The set of $Var$ features describes the three variances of the colour channels.
Table 5.4: Classification results obtained when combining the grey level LVDCT texture features with features computed directly from the colour histogram.

<table>
<thead>
<tr>
<th>Features Used</th>
<th>Classification Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>LVDCT</td>
<td>75.52%</td>
</tr>
<tr>
<td>LVDCT+E</td>
<td>82.40%</td>
</tr>
<tr>
<td>LVDCT+H</td>
<td>84.90%</td>
</tr>
<tr>
<td>LVDCT+Var</td>
<td>99.69%</td>
</tr>
<tr>
<td>LVDCT+E + H</td>
<td>84.90%</td>
</tr>
<tr>
<td>LVDCT+E + Var</td>
<td>99.69%</td>
</tr>
<tr>
<td>LVDCT+H + Var</td>
<td>99.69%</td>
</tr>
</tbody>
</table>

Table 5.4 shows the results of classification using these features. Generally the performance of the combined texture and colour features computed directly from the colour histogram is not as good as the performance of the combined feature set of the texture and first-order colour features obtained earlier. The only exception is with the LVDCT+Var and LVDCT+H + Var feature sets which perform slightly better. The sets of combined features of Table 5.4 were unable to achieve a 100% accuracy rate. This is contrary to the notion that the energy, entropy, and variance features computed directly from the colour histograms should perform better than the approximated first-order colour features. However a post mortem of the results gave a clearer understanding of the nature of these features. It is without a doubt that the features computed directly from the colour histograms gave more accurate measures of the histogram whereas the set of sub-optimal first-order colour features only offer an approximation to these measures. But when these features are to be used for classification, the latter feature set is able to offer more degrees of freedom to the classifier. Consequently, in the higher dimensional space of these features it is possible to obtain a better performance. For example, the two colour granite images (i.e. brasil violett and ghiandonato) have almost the same energy (E) features as can be seen in Table 5.5 but their corresponding 1-dimensional sub-optimal first-order colour features have quite distinct values. This is a fine example in which the direct computation of the features from the colour histogram would not work so well. The use of approximated sub-optimal first-order colour features, despite sacrificing the corresponding true feature values, provides a better alternative.
5.9.5 Use of First-Order Colour Features for Delineating Cover in Landsat Imagery

Besides computing features from the histogram of the tristimulus values of an image, one could estimate the same features from a selected set of three bands out of a possible seven bands that make up Landsat images [18]. It is worth noting that the first three bands correspond very closely to the Blue, Green, and Red visible light spectrum respectively [9] and are useful for some form of analysis work. However, a more powerful discrimination can be achieved by replacing any one of the first three bands with a band that is known to be sensitive to a particular type of cover. This has the effect of visually enhancing this known cover making it more readily detectable by an expert or a vision machine. For example, Bands 6 and 7 (reflected infrared) which are best for delineating water bodies are often being introduced into any one of the visible bands in order to highlight this particular cover in colour [9]. This form of colour coded images can then be processed by any tools developed for colour photograph analysis.

The purpose of this section is to determine how well the first-order colour features perform on colour coded Landsat images. The performance of the features will be assessed by means of classification accuracy. The task is to isolate the area of burnt forest from a Landsat image taken over the area of Korinthos in Greece. A supervised classification scheme is used and as such training samples of the burnt forest area are required. The training sample is subjected to a feature extraction process in order to obtain the class statistics. Next the Mahalanobis distance in the feature space is calculated for every pixel of the image. A threshold can then be chosen to segment out the burnt forest area.

The training sample used is 100 × 100 pixels and it is chosen with the help of an expert to be indicative of the burnt forest area. The size of the image to be classified is 1024 × 1024 pixels. The chosen bands to be used here are Band 1 (Blue), Band 7 (Green), and Band 3 (Red) [18]. The colour

### Table 5.5: Energy measures of two distinct colour texture images computed directly from the colour histogram and individually from the sub-optimal 1-dimensional histograms.

<table>
<thead>
<tr>
<th></th>
<th>$E$ (direct)</th>
<th>$E_1$</th>
<th>$E_2$</th>
<th>$E_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>brasil violett</td>
<td>0.0001998</td>
<td>0.0043620</td>
<td>0.0346327</td>
<td>0.0263021</td>
</tr>
<tr>
<td>ghiandonato</td>
<td>0.0001956</td>
<td>0.0031630</td>
<td>0.0309230</td>
<td>0.0459892</td>
</tr>
</tbody>
</table>
The coded version of this image is shown in Figure B.1.a of Appendix B with the burnt forest being indicated by the rectangular box. A moving square window to localise the feature extraction is chosen empirically to be 32 \times 32 pixels. In this case, the colour coded image will appear greenish for dry land which is characteristic of a burnt area. A look at Figure B.1.a indicates that the isolation of the burnt forest area using the colour coded image can be done quite easily with only colour attributes. It can be seen also that this affected area does not possess a distinct texture variability when compared to the rest of the unaffected areas.

A subset of the first-order colour features in Section 5.7 was chosen for this experiment. The features used here are the Energy, Entropy and Mean of the first-order histograms (i.e. \( E_1, E_2, E_3, H_1, H_2, H_3, \mu_2, \) and \( \mu_3 \)). The classification result of this test image using the above features is shown in Figure B.1.b. The superimposed boundaries of the detected regions are shown in Figure B.1.c. It can be seen that a significant portion of the burnt forest area is being successfully isolated. This implies the usefulness of the first-order colour features in analysing colour coded Landsat images. There exist numerous misclassified regions which can quite easily be removed simply by using size information of the region considering that these regions are often very small.

5.10 Discussion

The distribution of colour (i.e. multi-spectral) values for each pigment component of naturally occurring colour textures can be approximated by a multi-normal density. As texture phenomena often comprise several unique classes of pigments, the overall distribution of the colour texture data will be a mixture of multi-normal density functions. In such situations, a simplification of the 3-dimensional probability density function by the product of three first-order density functions computed along its principal axes is possible. In this work care has been taken not to modify the 3-dimensional PDF of the colour space as colour information will be drastically lost. This view contrasts with the common practice of changing the statistical nature of texture in grey level texture analysis. Histogram equalisation of the grey level is one example of pre-processing that is usually performed before the application of any texture analysis algorithm in order to ensure that no form of first-order statistics affects the analysis. However such modification alters chromatic properties of colour texture.

The optimal principal axis and sub-optimal principal axis transformation can be used to decorrelate the colour channel random variables in a way that make the independent processing of the transformed channels possible. Moreover, the first-order histograms associated with such
channels can then be used for approximating the higher dimensional colour histogram. Features of the colour histogram (i.e. energy, entropy, and variance) can then be estimated much more efficiently using the corresponding individual features derived from these lower dimensional histograms (i.e. first-order colour features).

It was discovered that the set of principal component axes of all possible pigments of most colour texture data are common and that it can be represented by the optimal or sub-optimal principal component axes. The optimal principal axis transformation is the best choice for decorrelating the colour channels of any colour texture data. The sub-optimal principal axis transformation also works quite well in these circumstances, but of course not as well as the optimal principal axes counterpart. The advantage of using the sub-optimal decomposition is that it is not data dependent and hence computationally cheaper to realise in practice.

The validity of the relationship of equation (5.14) was tested on a series of synthetic and real images. The approximation of the third-order density function by the product of three first-order densities was in close agreement when tested initially on a single mode colour Gaussian noise image. In the case of a double mode Gaussian noise image, the approximated density derived from the sub-optimal principal axes matches the observed one quite well but not as well as when the optimal principal axes were used instead for the transformation. The ability of the optimal principal axis transformation to fully de-correlate the channels was shown to be the key factor for the approximation of equation (5.14) to be successful. A similar behaviour was observed when real colour texture data was used. At this point, the question of which transformation (i.e. optimal or sub-optimal principal axes) to use to achieve a good approximation will have to be answered. A look at the estimated density functions of the optimal principal axis and sub-optimal principal axis transformation for real colour texture data suggests that their performance is quite similar. With this in mind, and taking into consideration the computation requirement, the sub-optimal principal axis transformation seems to be the preferred choice.

However, one must consider yet another factor for the approximation of equation (5.14) to be valid. The constituent densities must be aligned either along the major or minor axis as shown in Figure 5.11 for the case in a 2-dimensional space. When the constituent densities do not satisfy this assumption, this approximation will break down. The first-order histograms will then fail to reproduce the joint probability density correctly. This will lead to the inherent introduction of additional densities ('ghost' densities) that will hamper the approximation as can be seen in Figure 5.12. The latter effect of the approximation is more frequently associated with saturated or
enhanced colour images where the local axis of the constituent densities will curve away from the major or minor axis. Natural textures, which might contain areas of green foliage or colour tones, possess very low saturation values [7]. Therefore they tend to map into a small region circumscribed around the white point just like that shown in Figure 5.11. Looking again at Figure 5.3 confirms this fact about the constituent densities, at least for the cases of granite textures investigated here. In any case, provided the classes of textures do not include real textures with chromatic properties corresponding to the "ghost" modes, the proposed approximation will still be effective.

Results from the classification of colour granite images suggested that the sub-optimal principal axis first-order colour features are more suitable for such purposes. As the optimal principal axis first-order colour features are described uniquely in data dependent feature spaces, the Bayes' decision rule which minimizes the probability of error will no longer be optimum in situations where features are computed in disparate feature spaces. This finding further substantiates the use of the sub-optimal principal axis transformation to extract the first-order colour features.

The classification performance of the sub-optimal principal axis first-order colour feature set was compared with the corresponding features computed directly from the colour histogram. It was found that contrary to the notion that the computation of the exact features from the true colour histogram should lead to superior class-feature separability, the first-order colour features performed significantly better. This is attributed to the fact that any form of description using a set of higher dimensional first-order colour features has a higher degree of freedom and hence is likely to perform
5.10: Discussion

(a) Distribution

(b) After transformation

Figure 5.12: (a) The two modes of a mixture of Gaussian densities that are located along different major or minor axes. $e_1$ and $e_2$ are the principal axes. (b) Reproduced densities from the approximated first-order densities that shows additional modes (or 'ghost' densities) being introduced.

better than when only one feature is used on its own.

To summarise, the sub-optimal approximation of the third-order probability density function was shown to perform very well in these four respects:

- ability to approximate quite closely the third-order probability density function of the tristimulus values of colour texture.

- the sub-optimal principal axis first-order colour features are computed on just one transformed space and hence the optimality of the Bayes' classifier could be harnessed.

- the sub-optimal principal axis first-order colour features offer a higher degree of freedom for the classifier.
• computational efficiency.

One problem with this approach is that the colour histogram model is sensitive to the type of imaging device used. The absolute location of the set of clusters in the colour space changes with the mode of imaging. The causes of this are numerous; inconsistency in calibration, system dependent noise (channels, sensors, digitizer etc.), the use of different sensors etc. One way to eliminate this problem is to perform a detailed and accurate calibration on every component of the imaging system so that the colour histograms produced are consistent with that obtained by another imaging system. This can prove to be very difficult and costly to implement in practice. However, a much simpler approach is to ensure that the colour histogram model computed on one particular imaging system should not be used on another system. Eliminating the transportability of the model helps in reducing ambiguity caused by the different systems. Further more the cost of recomputing the first-order colour features and updating the databases for different imaging systems is small compared to the earlier calibration approach.

5.11 Conclusion

In this chapter, a robust colour representation for texture was introduced. This is based on the colour histogram which provides an accurate and stable model for colour image content. Working with the 3-dimensional colour model in image processing tasks is computationally expensive. This drawback motivated the development of a novel method which attempts to approximate this model by the product of the three 1-dimensional histograms computed along the principal axes of the colour histogram. From these decomposed histograms, one can compute statistical invariant colour features in the form of energy, entropy and variance measures. This approximation resulted in the simultaneous reduction of the memory storage and feature extraction time. With this new formulation, the memory needed to store the three first-order histograms is only $9N_g^3$ (where $N_g$ is the total number of grey-levels in each channel) as compared to $N_g^3$ for the 3-dimensional model; a saving of $N_g^2/9$. The feature extraction time is also significantly reduced as a triple summation (in the original 3-dimensional model) is being replaced by three single summations.

All these simplifications were possible as the distribution of the tristimulus values in a colour texture phenomenon roughly satisfies the assumption of being normally distributed and consisting of a mixture of similarly oriented densities. A simulation using the EM algorithm had provided experimental support for the assumptions to hold at least for real colour textures inves-
The stability of the first-order colour features over varying light intensity was also investigated. The classification experiment suggested that the sub-optimal principal axis first-order colour features are more suitable for colour texture representation. As the optimal principal axis first-order colour features are described in distinct data dependent feature spaces, the Bayes' decision rule which minimizes the probability of error will no longer be optimum in situations where features are computed on such disparate feature spaces, hence their poorer performance.

It was also revealed that contrary to the notion that the computation of the exact features directly from the true colour histogram should lead to superior class-feature separability, the sub-optimal principal axis first-order colour features performed significantly better. This is attributed to the fact that any form of description using a set of higher dimensional features (i.e. first-order colour features) has a higher degree of freedom and hence is likely to perform better than when only one feature is used on its own.

The excellent performance of these colour features computed from the approximated colour histogram in a difficult problem of colour granite image classification warrants their use in other image processing tasks. The proposed method of colour texture representation can easily be extended to other spectrally generated phenomena, one of which involving colour coded Landsat images was shown to perform very well in detecting a specified class of cover in the image data. However, the choice of which band to use depends on the type of cover analysed.

References


Chapter 6

Conclusions and Future Work

6.1 Conclusions

Texture is an indispensable characteristic of objects and consequently it plays an important role in vision processing. Texture analysis tasks include texture classification, texture description, and texture segmentation. Together with other early vision cues (i.e. edge and colour etc.) texture is important and sometimes crucial for high level analysis and interpretation of complex scenes.

In this thesis a review of representative approaches to texture representation was carried out. It was noted that statistical texture representation approaches were best suited for stochastic or disordered textures, many of which we commonly found in nature. Structural approaches are more suitable to deterministic or strongly ordered textures.

Among statistical approaches, the second-order statistics techniques namely the grey level co-occurrence matrix, grey level difference, Fourier power spectrum, and autocorrelation methods are known to be interrelated. Also it is well known that second-order statistics are often enough for most texture discrimination by the human vision system. We have found that another technique that is equally effective for texture analysis is the multi-channel filtering method, in particular the local linear transform method. This texture representation has been employed extensively in this research. The advantages of this method are as follows:

- The method is almost as efficient for texture discrimination as the method based on co-occurrence measurements which is considered as a benchmark.
- It enables a more compact description of local texture properties.
- The feature extraction process is computationally less demanding.
- The features are very suitable for multi-resolution texture segmentation based on clustering.
• Because of its parallel nature, its implementation on a parallel architecture is straightforward.

Several interesting observations have been made regarding the local linear transforms for texture analysis. The overall performance of the sub-optimal transforms (i.e. DST, DCT, DSTE, DREFT, and DROFT) does not depend much on the numerical values of the filter coefficients, but rather on the general structural form of the filters (edge-like, spot-like, etc.). Also it was revealed that the size of these transform filters does not play a major part in texture discrimination. Hence small filters (i.e 3 × 3) were recommended. Within the framework of multi-channel filtering, the Gabor filtering approach has been shown to be identical to the orthogonal basis filters approach in terms of their overall representational structure. However, these two approaches performed rather differently on texture segmentation, with the local linear transform features performing significantly better than the Gabor features. It should be noted that the computation of the local linear transform features is demanding. No fast method can be used to speed up the feature extraction process in the way the Fast Fourier Transform can accelerate Gabor filtering. Therefore, the emulation of local linear transform feature extraction process by using the sum of energies from a series of estimated quadrature filters at the first glance provides an interesting fast alternative for feature extraction. However, it has been shown in the thesis that this technique is not very cost effective as the number of quadrature filters needed for a successful emulation is very large. The filtering of the image with this large amount of filters, even with the accelerated speed of Fast Fourier Transform entails a further computational burden. Therefore, local linear transform measurements for texture analysis should be estimated using the traditional three-stage approach of filtering, local variance estimation, followed by region smoothing. If speed is of the essence, then implementation of the local linear transform on a parallel machines would have to be considered.

In information sciences and engineering, it is often wise to resort to additional sources of information to enhance the coding of the vast array of available information. The cost of providing the additional information must be kept low to ensure an economically viable system. In image analysis problems, colour is one such information source, which when properly coded can be used to aid texture representation. Furthermore, from the psychophysical point of view, colour aids the human vision system in two very important tasks, namely identification and discrimination. It is also known that colour is identified as one of the features observed by human pre-attentively, and as such it is computed in parallel with other features to be exploited in object interpretation at a later stage.
Chapter 6: Conclusions and Future Work

The primary objective of this research was to develop effective and efficient algorithms for stochastic colour texture analysis. The main interest lied primarily in developing colour texture features suitable for use in colour texture classification with possible extension to texture segmentation.

The practical need of dealing with colour in fields such as television, movies, photography, printing, as well as paint manufacturing has led the Commission Internationale de L'Eclairage (C.I.E.) to adopt a standard based on the assumption of colour trichromacy. The following specific wavelength values had been designated to the three primary colours; red = 700 nm, green = 546.1 nm, and blue = 435.8 nm. Even with this attempt at standardisation, the introduction of new imaging devices or systems often leads to conflict and incompatibility between standards of different manufacture origin and in different field of applications. In this respect, the National Television Systems Committee (NTSC) introduced another primary system ($R_N$, $G_N$, $B_N$) that is more suitable for the display screen phosphor employed in colour television receivers. An approach in specifying a colour is to represent the colour by some linear or nonlinear invertible function of these tristimulus values.

A brief review of some common colour co-ordinate systems has been carried out. It has been shown that these systems of colour transformation had in some degree preserved the spatial information content of the image thus making the transformed channels suitable for texture analysis. Hence a multi-band colour texture analysis scheme has been proposed which requires a two stage transformation process. The input tristimulus values are first transformed into the various colour co-ordinate spaces. Next, texture attributes are computed on these colour component images and used as colour texture features. These features were shown to work very well for colour texture classification involving granite images taken under fixed lighting conditions. It was observed that some of the features possessed a varying degree of resilience to illumination intensity changes. One disadvantage of the multi-band colour texture features is the inherent high dimensionality of the feature space associated with this model. Computing texture attributes on the three component images often resulted in the creation of redundant features or features that are highly correlated.

These had lead to the development of a novel colour texture representation scheme which is based on the colour histogram. It was noted that analysis based on the 3-dimensional colour model is computationally very expensive. This drawback motivated the development of a novel method which attempts to approximate this model by the product of the three 1-dimensional histograms computed along the principal axes of the colour histogram. This approximation results
6.2: Future Work

in the simultaneous reduction of the memory storage and feature extraction time. With this new formulation, the memory needed to store the three first-order histograms is only $9N_g$ (where $N_g$ is the total number of grey-levels in each channel) as compared to $N_g^3$ for the 3-dimensional model; a saving of $N_g^3/9$. The feature extraction time is also significantly reduced as a triple summation (in the original 3-dimensional model) is being replaced by three single summations. All these simplifications are possible as the distributions of the tristimulus values of colour texture phenomenon has been found, after some experiments with real colour texture data, roughly to satisfy the assumption of being normally distributed and consisting of a mixture of similarly oriented densities. This finding, of course, applies at least to the family of textures investigated. The colour features computed from the approximation of the colour histograms have been found to be quite stable over varying light intensities.

It has been shown that the sub-optimal principal axis first-order colour features are more suitable for colour texture representation than the optimal principal axis first-order colour features. As the latter are described in distinct data dependent feature spaces, the Bayes' decision rule which minimizes the probability of error will no longer be optimum in situations where features are computed on such disparate feature spaces, hence their poorer performance. A comparison of performance in classifying colour granite images of varying intensity values using the multi-band colour texture features (a best result of 98.85% was obtained) and the grey level texture features with sub-optimal principal axis first-order colour features (a best result of 100% was obtained) had shown that the latter set of features is more resilient to illumination intensity variation. Besides its excellent performance, the latter approach also enjoys a lower dimensional feature space.

It was also revealed that contrary to the notion that the computation of the exact features directly from the true colour histogram should lead to superior class-feature separability, the sub-optimal principal axis first-order colour features performed significantly better. This is attributed to the fact that any form of description using a set of higher dimensional features (i.e. first-order colour features) has a higher degree of freedom and hence is likely to perform better than when only one feature is used on its own.

6.2 Future Work

The research here had shown beyond any doubt that colour information can significantly enhance the performance of traditional grey level texture representation. However, this work is far from being complete with many challenging tasks still to be accomplished.
In this research, only the local linear transform measurements were used for the texture analysis part of the algorithms. It would be of interest to investigate the performance of other forms of texture analysis that were reviewed in Chapter 2.

There are a few colour co-ordinate systems that have not been tested here for the obvious reason of their computational involvement. Further work can be carried on other colour space transformations. Some of the colour systems that could be considered are the UVW, U*V*W*, L*a*b*, and L*u*v* colour systems.

Besides the features of the first-order histograms introduced here, the performance of other invariant features that measure the other property of the first-order histograms (like the normalised third and fourth moments etc.) should be investigated. The features used so far are global measures of these histograms. Due to the characteristic of colour textures, some of these histograms are multi-modal. It would therefore be advantageous to harness the local mode information instead of estimating global features of the first-order histograms. It will be interesting also to study the performance of individual features and to see how they fare when used in combination with other features. All these tests should if possible be done for more classes of colour textures.

As colour histograms are invariant to rotation about an axis perpendicular to the image plane, and changes only slowly under changes of scale, the colour features used here can be incorporated into other forms of texture representation that share similar invariance property (i.e. rotation invariance and multi-scale texture analysis schemes).

We used a simple feature grouping scheme for the texture and colour features of the separable colour and texture representation of Chapter 5. In this scheme all features are being treated with equal importance in the final classification stage of the process using the Bayes minimum error classifier. However, some features will tend to have a greater class-feature separability than others. Hence we need to weigh these features according to their capability perhaps in a neural network setup. This will require some a priori knowledge of the performance of the individual colour and texture features.

For colour texture segmentation, each pixel of the image must be distinguished by texture and spectral statistics and they are usually estimated by a sliding window centred on the pixel being examined. The fact that windows centred on adjacent pixels are mostly overlapping can be used to devise an efficient scheme that updates the first-order colour histograms by deleting the histogram entries caused by the removal of the pixels of the previous window position and then updating the entries caused by the inclusion of new pixels. Likewise, the colour statistics could
be updated in the same adaptive manner.

The size of the window for the segmentation exercise involving Landsat images is not selected automatically. The choice of the window size to use is a compromise between reducing the within-region variance and the ability to locate the region boundaries accurately. The window size depends on the content of the image: finer textures (i.e. textures which contain higher spatial frequencies) require smaller window size in order to detect smaller features, whereas coarser textures (i.e. those with lower frequencies) require larger windows. The window size may possibly be related to the frequency content of the Fourier spectrum. This would allow the window size to be selected automatically.

One significant drawback of the colour histogram model is its sensitivity to the type of imaging device used. Other effects like inconsistency in calibration, system dependent noise (channels, sensors, digitizer etc.), the use of different sensors etc. are frequent causes for the colour histogram model to be unstable for different image acquisition systems. One simple and effective solution would be to obtain the colour histogram models for all possible systems. However, a better solution might be to calibrate the colour outputs of the imaging system with a known multi-colour standard (i.e. frequently used standards are the Munsell colour chips). An effective and efficient way of colour calibration which will guarantee a consistent model for a wide range of imaging systems can be developed in the future.

The potential areas of application of the colour texture feature extraction scheme developed in this thesis are numerous. The proposed schemes of colour texture representation are expected to improve the performance of all existing applications involving just texture analysis alone, e.g. automatic classification of minerals and rocks, industrial inspection of materials like timber, granite, fabric etc., calculating colour texture measures in biomedical applications, automatic interpretation of remotely sensed data, and many more.
APPENDIX A: Colour Granite Images

(a) bianco castilla  (b) capao bonito  (c) marron guaiba
(d) new imperial  (e) santiago  (f) rosa baveno
(g) balmoral GF  (h) rosa monforte  (i) rosso perla
(j) cardinal red  (k) brasil violett  (l) ghiandonato

Figure A.1: 12 colour granite images used for classification.
APPENDIX B: Experiment with Landsat Data

(a) Colour coded Landsat image.  (b) Detected regions.

(c) Superimposed boundaries.

Figure B.1: Experiment with Landsat image using colour features.