Universitat de Girona

TEXTURE RECOGNITION UNDER VARYING IMAGING GEOMETRIES

Xavier LLADÓ BARDERA

ISBN: 84-688-5940-0
Dipòsit legal: GI-215-2004
http://hdl.handle.net/10803/7721

Department of Electronics, Computer Science and Automatic Control

PhD Thesis

Texture Recognition Under Varying Imaging Geometries

Thesis presented by

Xavier Lladó Bardera,

to obtain the degree of:
PhD in Computer Engineering.

Supervisors:
Dr. Joan Martí Bonmatí
Prof. Maria Petrou

Girona, October 2003

To my family, especially to my parents.

A la meva familia, especialment als meus pares.

Agraïments

Després de tant esforç no puc fallar. Agafo la pilota, llenço, i ... bàsquet! Just en el moment en què sona la botzina i s'acaba el partit.

Segurament som molts els que coincidim en afirmar que una de les claus més importants per aconseguir la victòria és "l'entrenador". Jo n'estic convençut! Per això voldria agrair profundament al Dr. Joan Martí i a la Prof. Maria Petrou tot l'esforç que han fet durant aquests anys per dirigir-me. En Joan va ser qui em va donar l'oportunitat d'incorporar-me dins l'equip del Grup de Robòtica i Visió per Computador. Durant tot aquest llarg període sempre m'ha ajudat a superar els problemes que inevitablement han anat sorgint amb el temps. Gràcies per la teva ajuda. A ell també li he d'agrair la possibilitat (encara que fos a correcuita) d'incorporar-me durant uns mesos dins l'equip de la Prof. Maria Petrou. Què dir de la Maria! D'ella destacaria aquest caràcter afable i obert que em va sorprendre des del primer dia que la vaig conèixer. Però per sobre de tot, l'esperit i delit que té pel món de la recerca. És tant gran que es transmet fàcilment als seus "jugadors". A la Maria li he d'agrair especialment els consells, les suggerències i les idees que m'ha donat durant aquests anys. Podríem dir que és d'aquell tipus d'entrenadora que es desviu per la seva feina, que fixa un nivell d'exigència molt alt, però que ajuda als seus deixebles per extreure'n el millor de cadascun. Moltes gràcies.

A més de la importància dels entrenadors, ressaltaria també la figura del "delegat" d'equip. Sovint aquest és el que conviu més el dia a dia amb els jugadors, i el que transmet optimisme, alegria, experiència i moltes ganes de treballar. Sense cap mena de dubte, el Dr. Jordi Freixenet recull totes aquestes grans qualitats. Tot i que també és veritat que ell hi afegeix algun que altre vici com el d'arribar sempre

tard. Jordi, gràcies per tot el que has fet per mi. Per cert, sempre t'has defensat bé en la posició de "pivot", quins partits aquells, eh!

També vull agrair de forma general l'ajuda de tots els membres del Grup de Robòtica i Visió per Computador. Començaria per en Joan Batlle, la figura dura i exigent d'un bon "president", i continuaria pels altres membres de "l'equip tècnic": en Xevi, en Pere, en Rafa, en Quim, en Lluís, en Pep, en Marc, sense deixar-me els "responsables del material": en Toni, en Tomàs i en Lluís. Tots ells també han realitzat una tasca indispensable.

No em puc oblidar dels "jugadors", els que han estat companys d'equip durant tot aquest temps. M'agradaria començar per en Miquel Grabulosa. Ell va assumir el lideratge durant tots els anys en què vam estudiar la carrera. Però després d'això, i quan ja ens havia convençut a tots perquè continuessim estudiant, va fer el "triple" de la seva vida i no va gosar rebutjar una oferta de treball multimilionària. Podríem definir en Miquel com un jugador d'equip excel·lent, d'aquells que fa ser més bons als altres (qui millor que jo per dir-ho), però per sobre de tot, un bon amic. Ja saps, sempre que vulguis podem tornar a repetir aquells famosos u contra u. Això si, ja et vaig deixar guanyar un cop, o sigui que ara et tocarà suar.

També m'agradaria agrair l'ajuda d'en Xavier Armangué. Un jugador amb una gran tècnica, capaç de trobar solucions a les situacions més difícils que se li presentin, i us puc assegurar que no són poques perquè jo mateix he estat el primer d'emprenyar-lo mil i un cops. Sempre recordaré el partit que vam anar a veure plegats al Madison Square Garden. Tot un espectacle!

Un cas particular és el d'en Robert Martí, un jugador de la casa que va decidir provar fortuna a Portsmouth, Anglaterra. Amb en Robert vam començar essent rivals, recordes aquells duels Sta. Coloma de Farners vs Salt? Qui ho havia de dir que al cap d'uns anys seríem companys d'equip. Tampoc oblidaré aquells entrenaments amb l'equip de Portsmouth. Oi que vam deixar el nivell molt alt? Gràcies per la teva ajuda i hospitalitat durant tots aquells mesos a Anglaterra, i també com no, per fer-me de guia a Londres.

Ep! Encara no us he parlat del meu amic i company inseparable, en Xevi (perdò,

des de fa uns mesos li he de dir Dr. Xevi Muñoz). No és que me n'hagi oblidat, ni molt menys. Com bé diu el refrany: "lo bo es fa esperar". En Xevi és d'aquella classe de jugadors que es troben en molt pocs equips. Senzillament brillant, amb talent, gran visió de joc i una gran facilitat per prendre decisions ràpidament. En definitiva, un gran director de joc, un "base" que tothom voldria tenir al seu equip. Jo he tingut la gran sort de gaudir de les seves "assistències" durant tot aquest temps, i sense cap mena de dubte, això m'ha ajudat molt tant en la part professional com en la part personal. De veritat, moltes gràcies.

També voldria destacar altres companys i ex-companys d'equip que m'han acompanyat durant aquests anys. Un dels més carismàtics en Miquel (Monti pels amics). D'ell admiro aquella facilitat innata per convocar sopars i festes, i fer-nos riure sense parar. No em puc oblidar d'en Bernat, l'Enric i en Jeroni, i tampoc de les noves generacions que pugen amb força: en Jordi i l'Ela, una parella explosiva. L'Arnau, tot un mag capaç de fer coses inexplicables, i en David del qui ja vaig veure el seu potencial durant el projecte fi de carrera.

També guardo un gran record dels companys que vaig fer durant les meves estades a Surrey. En especial de l'Svetlana que tant i tant em va ajudar. També d'en Manuel i en Luís que van fer-me de traductors quan més ho necessitava. I com no de l'Andrew i el seu gos Homer, que em van fer sentir com a casa.

Tornant més a prop de casa, voldria agrair a tots els meus amics de sempre el suport que m'han donat. En especial en Francesc, company d'àpat cada dia.

Ja per anar acabant voldria agrair d'una manera molt especial als meus pares: Joan i Dolors, la meva germana Farners i el meu cunyat Carles, tots els ànims i suport que m'han donat. Són els millors "fans" del món! En tot moment els he tingut al meu costat i això m'ha ajudat molt. A vosaltres pares, gràcies per permetre'm estudiar durant tots aquests anys i en definitiva per estimar-me tant. El meu últim agraïment és per una persona molt especial, la Montse. A ella li vull agrair aquelles trucades diàries durant tots els mesos que vaig estar a Anglaterra. No sé pas què hauríem fet sense Europa 15! Una "manager" fantàstica, sens dubte. Gràcies per la teva paciència.

Acknowledgments

First of all, I would like to thank Dr. Joan Martí and Prof. Maria Petrou for their supervision and patience. Joan gave me the opportunity to join the Computer Vision and Robotics Group at the University of Girona and has helped me many times during these years. I would also like to especially thank Maria for offering me the opportunity to work with her group at the University of Surrey, England and for her comments, suggestions and ideas which have been immensely helpful to me.

I want to thank the people from the Computer Vision and Robotics Group for their friendship and support, especially Dr. Jordi Freixenet who always had time to help me. Also the people from the Center for Vision, Speech, and Signal Processing for their patience with my "Catalan-English"! Particularly Svetlana, Manuel and Luís for their constant support.

Also thanks to my PhD-mates Xavier Armangué and Robert Martí. I recall the countless questions I had for Xavier and incredibly he always had an answer. My thanks to Robert for making my stays in England so comfortable and entertaining.

Special thanks to my friend and PhD-colleague Xevi Muñoz. Literally, he has always been at my side. I do not have sufficient words (and it is not a problem of my English) to describe his help and support, not only during our stressful research, but during all the time we spent together.

Finally, I would like to thank my family and my girlfriend for all the support they have given me, for all they have done and continue to do for me.

Texture Recognition Under Varying Imaging Geometries

Resum Executiu

La visió és probablement el nostre sentit més dominant a partir del qual derivem la majoria d'informació del món que ens envolta. A través de la visió podem percebre com són les coses, on són i com es mouen. En les imatges que percebem amb el nostre sistema de visió podem extreure'n característiques com el color, la textura i la forma, i gràcies a aquesta informació som capaços de reconèixer objectes fins i tot quan s'observen sota unes condicions totalment diferents. Per exemple, som capaços de distingir un mateix objecte si l'observem des de diferents punts de vista, distància, condicions d'il·luminació, etc.

La *Visió per Computador* intenta emular el sistema de visió humà mitjançant un sistema de captura d'imatges, un ordinador, i un conjunt de programes. L'objectiu desitjat no és altre que desenvolupar un sistema que pugui entendre una imatge d'una manera similar com ho realitzaria una persona.

Aquesta tesi es centra en l'anàlisi de la textura per tal de realitzar el reconeixement de superfícies. La motivació principal és resoldre el problema de la classificació de superfícies texturades quan han estat capturades sota diferents condicions, com ara distància de la càmera o direcció de la il·luminació. D'aquesta forma s'aconsegueix reduir els errors de classificació provocats per aquests canvis en les condicions de captura.

En aquest treball es presenta detalladament un sistema de reconeixement de

textures que ens permet classificar imatges de diferents superfícies capturades en diferents condicions. El sistema proposat es basa en un model 3D de la superfície (que inclou informació de color i forma) obtingut mitjançant la tècnica coneguda com a *4-Source Colour Photometric Stereo* (CPS). Aquesta informació és utilitzada posteriorment per un mètode de predicció de textures amb l'objectiu de generar noves imatges 2D de les textures sota unes noves condicions. Aquestes imatges virtuals que es generen seran la base del nostre sistema de reconeixement, ja que seran utilitzades com a models de referència per al nostre classificador de textures.

El sistema de reconeixement proposat combina les *Matrius de Co-ocurrència* per a l'extracció de característiques de textura, amb la utilització del *Classificador del veí més proper*. Aquest classificador ens permet al mateix temps aproximar la direcció d'il·luminació present en les imatges que s'utilitzen per testejar el sistema de reconeixement. És a dir, serem capaços de predir l'angle d'il·luminació sota el qual han estat capturades les imatges de test.

Els resultats obtinguts en els diferents experiments que s'han realitzat demostren la viabilitat del sistema de predicció de textures, així com del sistema de reconeixement.

Contents

1 Introduction **1**

1.1 Motivation 1

1.2 Scope of the Research 5

1.3 Original Work 6

1.3.1 Objectives 8

1.4 Thesis Organisation 10

2 Image Formation **15**

2.1 Introduction 15

2.2 Characterisation of the Incident Image 20

2.2.1 Terminology 20

2.2.2 From Physical Surface to Incident Image 23

2.2.3 Dichromatic Reflection Model 25

2.3 Surface Information 30

2.3.1 Gradient Space 30

2.3.2 Surface Roughness Description 32

2.4 Summary 36

3 Recovering Surface Information **39**

3.1 Introduction 39

3.2 Shape from X 41

3.3 Photometric Stereo: Related Work 44

3.3.1 Greyscale Images 45

3.3.2 Colour Images 49

3.3.3 Applications . . . 52
3.4 Use of 4-Source Colour Photometric Stereo . . . 54
3.5 Photometric Stereo: Terms and Notations . . . 56
3.5.1 Greyscale Photometric Stereo . . . 58
3.5.2 4-Source Colour Photometric Stereo . . . 59
3.5.3 The Problem of Highlights and Shadows . . . 62
3.5.4 Texture Rendering . . . 70
3.6 3D Surface Reconstruction . . . 72
3.6.1 Integration Techniques . . . 73
3.7 Summary . . . 76

4 Texture Prediction Seen under Different Imaging Conditions 79
4.1 Introduction . . . 79
4.2 Initial Prediction Approach . . . 80
4.2.1 Normal Vectors under Different Resolutions . . . 81
4.2.2 Colour Information under Different Resolutions . . . 88
4.2.3 Discussion . . . 90
4.3 New Prediction Approach . . . 92
4.3.1 Direct Image Prediction . . . 93
4.3.2 Surface Prediction . . . 99
4.4 Summary . . . 101

5 Texture Recognition 105
5.1 Introduction . . . 105
5.2 Texture Classification System . . . 108
5.2.1 Related Work . . . 108
5.2.2 Discussion . . . 111
5.3 Model-Based Texture Classification . . . 113
5.3.1 Virtual Database Creation . . . 114
5.3.2 Recognition Procedure . . . 116
5.4 Summary . . . 136

6 Experiments **139**

6.1 Introduction . . . 139

6.2 Experimental Data . . . 141

6.3 Texture Prediction Results . . . 149

6.3.1 Accuracy of Photometric Stereo . . . 150

6.3.2 Accuracy of Image Prediction when the Distance Changes . . 152

6.3.3 Accuracy of Surface Shape Prediction . . . 155

6.3.4 Accuracy of Image Prediction from Different Distances 159

6.3.5 Accuracy of Surface Shape Prediction from Different Distances 161

6.4 Texture Classification Results . . . 162

6.4.1 Accuracy of Classification. Images from the Same Distance . . 164

6.4.2 Accuracy of Illuminant Estimation. Images from the Same Distance . . . 168

6.4.3 Accuracy of Classification. Images from Different Distances . . 168

6.4.4 Accuracy of Illuminant Estimation. Images from Different Distances . . . 170

6.5 Summary and Conclusions . . . 172

7 Conclusions **175**

7.1 Contributions . . . 175

7.2 Future Work . . . 177

7.3 Related Publications . . . 180

A Simultaneous Surface Texture Classification and Illumination Tilt Angle Prediction **183**

A.1 Introduction . . . 183

A.2 Co-occurrence Matrix Behaviour . . . 184

A.2.1 Removing the Ambiguity . . . 188

A.3 Illuminant Tilt Angle Prediction . . . 189

A.4 Experimental Trials . . . 190

A.4.1 Accuracy of Tilt Angle Prediction . . . 191

A.4.2 Accuracy of Texture Classification . . . 193

A.5 Conclusions 194

Bibliography **195**

List of Figures

1.1 Three images of the same surface texture captured using different imaging geometries. 4
1.2 Strategy of our texture classification system. 7
1.3 Structure of the thesis. 11

2.1 Rotation-invariant texture descriptors treat an image as if it has an existence of its own. 17
2.2 Specular and diffuse surfaces. 18
2.3 Scenario of image acquisition of an illuminated physical surface. . . . 19
2.4 A surface observed as a collection of planar facets where each one has its own surface orientation **N**. 22
2.5 Reflection process at a surface. 25
2.6 Cast and self shadowing. 29
2.7 Examples of surface gradient distributions. 32
2.8 Surface profile. (a) Surface height map of a surface. (b) Corresponding profile of a vertical cross section. 34

3.1 Indirect methods. (a) Based on stereo vision, (b) Based on optical flow, (c) Based on Shape from X. 40
3.2 Different images of a surface captured from the same viewing position but with different lighting directions. 43
3.3 A review of photometric stereo techniques depending on the use of greyscale or colour images. 51
3.4 Applications of photometric stereo techniques. 53
3.5 Definition of the important vectors and reflectance angles. 57

3.6 Lighting configuration used in the 4-source colour photometric stereo. 58
3.7 Example of how photometric stereo can be computed if more than 3 images are available. 60
3.8 Intuitive interpretation of the colour photometric stereo. 61
3.9 4-source colour photometric scheme. 62
3.10 Example of computed normals affected by a highlight [30]. 64
3.11 Three image samples which present a great variety of shadows. 65
3.12 Input photometric set for the "tomatoes" surface. 68
3.13 Example of reconstruction with highlights and shadows correction. . . 69
3.14 Different ways of relighting surface textures. 71
3.15 Image rendering process. 72
3.16 Classification of integration techniques: local and global integration techniques. 74

4.1 Example of a surface seen in a fine and coarse resolution. 81
4.2 Problem of changing image resolution. 82
4.3 Example: prediction of the surface information for superpixel ij. . . . 83
4.4 The intensity of a pixel is proportionate to the amount of light reflected by the surface patch projected onto the pixel. 86
4.5 General case of image resolution change. 87
4.6 Scheme of our first prediction approach. 91
4.7 Scheme of the direct image prediction approach. 98
4.8 Normal vector of the superpixel obtanined by fitting the plane in the LSE sense. 100
4.9 Scheme of the surface prediction approach. 102

5.1 Example of global and local image properties. 107
5.2 Strategies of light invariant texture classification. 111
5.3 Recognition scheme. It is divided into two main phases: virtual database creation and recognition. 114
5.4 Virtual database creation. 115
5.5 Example of images created in the virtual database. 117

5.6 Recognition procedure. This procedure is divided into two main phases: learning process and classification process. 118

5.7 Example of co-occurrence matrix computation. 122

5.8 Illustrative example of feature spaces. 124

5.9 Example of K-nearest neighbour (K-NN) and mean nearest neighbour (mean-NN) classifier. 134

6.1 Scheme of the platform used to capture the experimental data. 142

6.2 Summary of the notation used in the experimental data. 143

6.3 One image from distance A and one from distance B of each of the twenty five sample textures. 144

6.4 A representative image and the PDF representations of the surface gradients p and q of an isotropic, a directional, and a very rough isotropic surface. 148

6.5 (a) and (c) Images captured with camera (tB). (b) and (d) Corresponding images generated by photometric stereo (BB). 151

6.6 Mean square errors for texture T21 under varying tilt angles. Comparision with the naive case in which the process of surface recovery and rendering has been bypassed. 153

6.7 Accuracy of image prediction for three surface textures under varying tilt angles. 154

6.8 Accuracy of the surface shape predictions for three surface textures. . 156

6.9 Correlation scheme. 157

6.10 Three images of the same surface texture seen from distances A, B, and C, (a), (b) and (c) respectively. 160

6.11 Illumination setup. (a) Illuminant tilt angles used to capture the test images. (b) Illuminant tilt angles used to create the virtual database. 164

6.12 Model-based, best case, and naive case approaches to texture classification. 166

6.13 One image predicted for each of the twenty five textures. 169

7.1 Images of two textures captured while varying the spectral properties of the light source. 179

A.1 One image of each of the fifteen sample textures. 185
A.2 Feature behaviour for isotropic textures 1 (first row), and 9 (second row), and anisotropic texture 15 (third row). 187
A.3 Prediction scheme. 189
A.4 Experimental setup. 191
A.5 Error in the tilt angle prediction for textures 1, 9, and 15, (a), (b), and (c) respectively. 192

List of Tables

2.1 Important types of surface. 36

3.1 Taxonomy of photometric stereo problems. 45
3.2 Greyscale photometric stereo approaches classified according to the required number of images. 49
3.3 Integration techniques: advantages and disadvantages. 76

5.1 Some of the best known texture measures computed from the co-occurrence matrix. 123
5.2 Feature selection methods. 129
5.3 A comparison of evaluation functions provided by Dash and Liu [37]. 131

6.1 Accuracy of Photometric Stereo. Quantitative assessment for each approach over all tilt angles. 152
6.2 Quantitative assessment of shape predictions $imaAB$ and $surAB$. . . 158
6.3 Quantitative assessment for each prediction approach ($imaAB$ and $surAB$) using the $AASR$ parameter. 159
6.4 Overall quantitative assessment for each prediction over all 15 textures and all tilt angles. 161
6.5 Overall quantitative assessment for all 15 textures of shape predictions AC and BC. 162
6.6 Overall quantitative assessment of the $AASR$ parameter for prediction AC and BC for all 15 textures. 163
6.7 Texture and illuminant classification rates obtained for the best, model-based and naive cases. Prediction BB. 167

6.8 Texture and illuminant classification rates obtained for the best, model-based and naive cases. Prediction AB. 171

A.1 Overall quantitative assessment over all 15 textures of illuminant tilt angle predictions. 193

A.2 Texture classification rates and MSE of the tilt angle prediction obtained over the isotropic textures and over all fifteen textures. 194

Chapter 1

Introduction

1.1 Motivation

Vision is our most dominant sense, from which we derive most of our information about the world around us. From images which enter the eyes and passthrough the processing in the brain, we learn where things are, how they move and what they are. In the perceived images our visual system can separate elements such as colour, texture and shape, elements, which allow our brain to recognise objects even when they are observed under totally different conditions.

It is not surprising that our visual system allows us to recognise objects seen from different viewpoints, and distances as well as under different illumination conditions. For instance, it is an easy matter for us to recognise a field of sunflowers when observed from a close distance. Colour, texture, and shape are used for recognition. However, we are also capable of recognising a field of sunflowers even when observed from other viewpoints and distances. Think, for example, of an ariel view of the same field. In this case, although the visual perception is very different, our visual system allows us to recognise it for what it is. Similarly, we are able to recognise this field when seen under different illumination conditions. We recognise it as such at midday when the sun is shining as well as in moonlight.

Computer vision tries to emulate the human vision system by means of an imaging system, a computer and a set of algotithms. The desired goal is to develop

a system which can understand an image in a way similar to a human observer. Among other research topics, texture analysis plays an important role in computer vision and pattern recognition since most real world objects consist of different kinds of surface textures. Texture is a characteristic which surrounds us constantly and refers to the way things feel or would feel if touched. For example, sandpaper looks and feels rough and a cotton ball looks and feels soft.

This thesis is concerned with the application of texture analysis to discriminate between textured surfaces. The main motivation is the problem of classifying textured surfaces imaged under varying geometries, i.e. distance from the sensor and illumination direction, as well as the necessity of finding reliable methods of reducing classification errors caused by changes in the geometry's properties.

In texture analysis one must distinguish between *image* texture and *surface* texture. Image texture is what appears in the 2D image of a physical object, while surface texture refers to the variation of the physical and geometric properties of the imaged surface which give rise to the image texture. Changes in the imaging geometry can significantly alter the appearance of the surface, implying significant variations in the image texture [22]. And one still has to perform the task of recognition from the image texture.

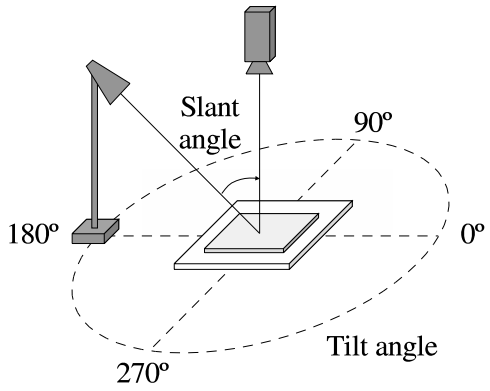
Recognition of 3-dimensional surface textures from 2-dimensional images is not an easy task. The 2-dimensional texture in the image (the image texture) is produced by variation in both surface reflectance and surface relief. While the reflectance properties are intrinsic to a surface and could be dealt with by means of conventional 2D texture classifiers, the surface relief produces a pattern of shadings which depends strongly on the direction of the illumination. Thus, the image texture created by a 3D surface texture changes drastically with the illumination, and a conventional classifier may not cope with such changes even for uniformly coloured surfaces [112].

The effect of varying lighting conditions on supervised texture classification was first investigated by Chantler [21]. It was observed that the direction of the illumination affects the directionality of an image obtained from a given surface and modifies its appearance. Nevertheless, most of the classification approaches proposed in the

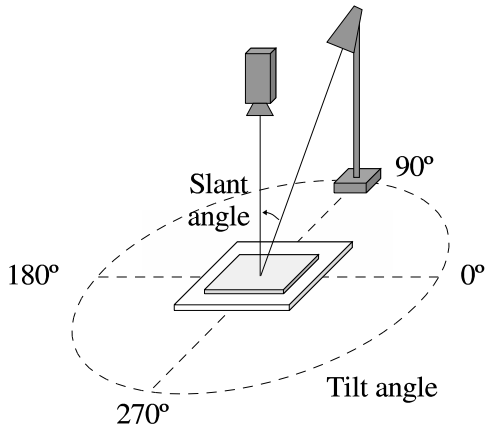
literature do not take into account the effect of light on the imaged scene, thus tacitly assume that the illumination direction is constant. This constraint may induce critical misclassification rates when the illumination is altered between training and classification. Moreover, there is a wide range of applications in which texture classification may have to be performed under varying lighting conditions. For instance, close proximity point lighting, often used for industrial inspection purposes, provides illumination at varying angles throughout the scene. Remote sensing systems which provide their own artificial "light" such as active sonar or radar are non-stationary, hence the "illuminant" direction is dependent on the orientation of the survey platform itself. Other remote sensing devices using natural light are also affected by changes in the illuminant vector according to the time of day, etc.

In order to illustrate the problem, let us consider a simple example. Figure 1.1 shows three images of the same texture illuminated from two different tilt angles with the same slant angle, and obtained from two different distances. The direction of the illuminant with respect to a texture is commonly defined by two polar coordiantes: tilt and slant. If we define our camera axis parallel to the z-axis, the tilt is the angle the illuminant vector makes with the x-axis when it is projected onto the x, y plane. On the other hand, the illuminant slant is the angle the illuminant vector makes with the camera axis. Both angles are illustrated in Figure 1.1. Observe how a shift in the tilt angle is clearly manifested in the perceived images (see Figure 1.1.(a) and Figure 1.1.(b)). Illumination variation attenuates or accentuates the directional information of the image texture. Moreover, changes in camera distance provoke significant variations in the image texture. This is illustrated in Figure 1.1.(c). It should then not be surprising that a classifier would have difficulty in recognising these images as produced from the same textured surface.

(a)



(b)



(c)

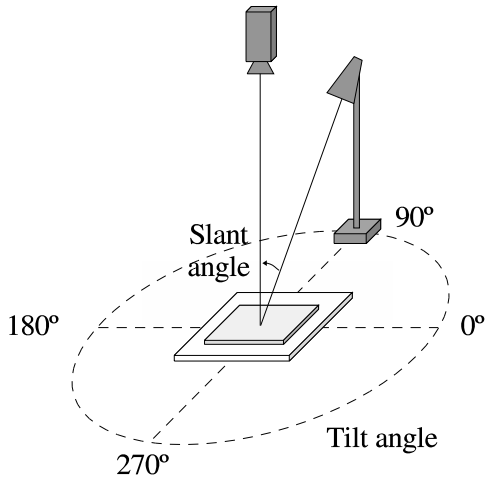


Figure 1.1: Three images of the same surface texture captured using different illuminant tilt angles, (a) and (b), and different camera distance, (b) and (c).

1.2 Scope of the Research

In computer vision, there is an important number of works which propose rotation and scale invariant texture classification [29, 109, 96, 59, 133, 108, 137, 91, 92], as well as similar works on topics like multiscale and scale-space texture analysis [29, 17]. Almost all the work published in the past is on texture features invariant to rotation and scale treated texture as an innate property of a flat surface which remained unchanged when the camera was moved or the illumination changed. I wish to stress that the surface texture which interests us here is distinct from image texture. Image texture is the result of surface relief and surface colour, and if one wishes to avoid its dependence on imaging geometry, we have to go through the fundamentals of image formation and discuss the way surface texture will appear when the imaging geometry changes. So, unless one considers texture as a pattern painted on a smooth surface, it is incorrect to deal with it ignoring the imaging geometry and using the Brodatz album [18] to evaluate the methodology used. The exception to this are the works of Lazebnik et al. [91, 92] for example, who acknowledging that image texture may change significantly with imaging geometry, evaluate their algorithm using recaptured images. These methods assume that under "mild" changes of imaging geometry, the transformation of image texture will be affine. This is true up to an extent. However, depending on how rough a surface is, mild changes in imaging geometry, when referring to the imaged surface as a whole, may be strong changes locally. Under those circumstances, we cannot avoid addressing the issue of texture recognition in conjunction with the process of image formation.

As will be seen in this thesis, texture changes by itself and appears different depending on distance and lighting conditions. Very few works have been published on the topic of texture classification independent from the direction of illumination. However, there are different ways of dealing with this problem:

- The first approach consists of extracting and using explicit separate 3D shape and surface reflectance information. The reflectance property (also known as albedo) and the gradient vector of every visible surface patch describe the surface in a way independent from illumination and the classification can be

done directly on the basis of this explicit information. This was done by McGunnigle and Chantler [113], and Barsky and Petrou [6].

- Another approach is to study the immediate effects produced by illumination direction on the observed 2D texture. Recently, Chantler et al. [23] presented a formal theory which demonstrates that changes in the tilt of the illumination direction make texture features follow super-elliptical trajectories in multi-dimensional feature spaces. Based on their work, Penirschke et al. [125] developed an illuminant rotation invariant classification scheme.

- Finally, one can train a classifier on a wide selection of images of the same surface, obtained from various viewpoints and under various illumination conditions (see the works [33, 36, 32, 56]). Thus, information on changes in surface appearance is explicitly built in the classifier, using both the reflectance and the 3D relief information, which allows it to recognise a surface correctly under novel viewing and lighting conditions [98, 99, 164].

1.3 Original Work

In this thesis, after analysing the three strategies, we integrate the surface texture information derived by colour photometric stereo (CPS) into a complete model-based texture classification system. Photometric stereo [170, 171] is the technique which allows us to obtain surface texture information from a few images of the same surface imaged under various illumination directions. Hence, the photometric technique seems ideally suited to our purposes since the problem itself is caused by variations in the imaging geometry. Basically, the main idea of our strategy consists of creating, by means of the surface texture information, a "virtual" database of image textures against which we compare unknown test images in order to classify them. Note that we do not use the surface texture information directly to perform classification (as in the first strategy), but we use it to create new images which are the references for our training and classification process (see Figure 1.2). Note that this strategy corresponds to the third one of those described in the previous section.

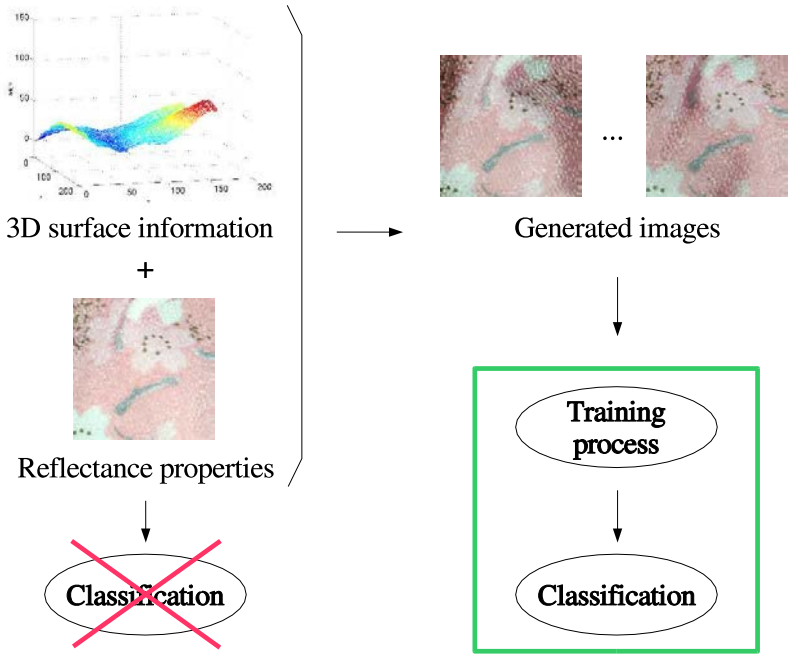


Figure 1.2: Strategy of our texture classification proposal. The surface texture information is not used to perform classification, but to generate new images which are the references for the training and classification process of the system presented in this thesis.

Furthermore, the classification system allows us to guess the approximate direction of the illumination used to capture the test images.

There are various alternative techniques which allow us to obtain 3D information about surfaces, like *stereo vision* [48], *optical flow* [5] and various *Shape from X* methods. These Shape from X methods allow the recovery of the local shape and reflectance properties of the surface which are used in different works to predict the surface appearance as a function of illumination and also to reconstruct 3D shape information [162, 176, 100]. Over all these techniques, the photometric stereo technique was chosen because of its various advantages: it does not suffer from the correspondence problem [177] like conventional stereo does; it does not make strong assumptions about the underlying surface structure like some Shape from X techniques do, and it allows us to recover both local colour and local gradient while flagging the places where some of its assumptions break down and recovery is impossible.

The approach in this thesis uses colour photometric stereo to extract surface information from which a "virtual" database of reference textures is created, compatible with the imaging geometry of the test images, and used to classify the test images. The virtual database creation comes in two "flavours": creation of the virtual database for test images seen from the same distance as the training images, and creation of the virtual database for test images seen from a longer distance than that of the training images. When the test images are known to have been captured from the same distance as the training images, the creation of the virtual database is straightforward: the use of surface texture information extracted by photometric stereo allows us to obtain a rendering of the surface texture directly. Things are less straightforward when the test images have been captured from a longer (but known or hypothesised) distance than the training images. In this case, in order to create the virtual database, we propose a new method which allows us to predict the appearance of a surface texture at a longer distance and for various directions of light. Hence, we deal with the problem of texture recognition under varying geometries such as distance from the camera and light direction.

1.3.1 Objectives

Taking into account the considerations we mentioned above, the main objectives of this thesis are defined as:

To propose a texture prediction framework which allows us to predict the appearance of a surface texture when seen from different imaging geometries, i.e. distance from the camera, sensor sensitivity, directions of illumination, and spectral characteristics of the light source.

To propose a model-based texture classification system which enables the classification of textured surfaces imaged under varying geometries such as distance from the camera and illumination direction. The system also allows one to guess the approximate direction of the illumination used to capture the test images.

Alongside these objectives, there are some points which need to be considered:

- **Prior knowledge (Image database).** There is one important aspect to mention related to the prior knowledge used in this thesis. Before starting the prediction and classification processes, a fundamental knowledge of each surface texture has to be known. This implies that for each texture a database of images should be available with the aim of recovering the surface texture information. This database should contain the required images to which one can apply the colour photometric stereo technique, as well as those images which should be used as training and test images.

  It is well-known that there are many databases of texture images from which it is possible to evaluate a texture recognition system. However, there is no database of images on the topic of texture classificaton invariant to distance from the camera and direction of illumination. For example, the well-known "Columbia-Utrecht database" established by Dana et al. [35] is not suitable for our purposes because the illumination was held constant while the viewpoint and orientation of the samples were varied during data capture. The "Photometric Image Databases" of the Texture lab from Heriot-Watt University [2], do not meet our requirements because they did not provide photometric sets of images captured at different distances. Therefore, we opt in this thesis to build our own image database which provides us with all the information needed to achieve the desired purposes.

- **Robustness.** The proposed classification system should show a robust behaviour and obtain correct classification rates for a large set of textures. Thus, the system will be tested over globally flat surfaces and very rough surfaces, including also directional and isotropic surfaces.

  Along with this point, we will assess the accuracy of the prediction method, evaluating the performance for a wide range of real textures at different distances and under varying lighting conditions. We should investigate a possible relationship between prediction inaccuracy and classification errors.

- **Flexibility.** The proposed classification strategy should be easily extensible. That means the different processes of the classification system may be extended or replaced by other techniques. For instance, it should not be difficult to add/change the feature extraction process, the feature selection process, the classifier, etc. In this sense, the object-oriented approach ensures this flexibility.

1.4 Thesis Organisation

The structure of this thesis is a step by step explanation of the methodology used to carry the work out. Chapter 2 introduces theoretical concepts on image formation and surface description. Next, Chapter 3 reviews the recovering of surface information, detailing the colour photometric stereo technique used in this thesis. The prediction framework which allows us to predict surface information of the same texture seen from different distances is proposed in Chapter 4. In Chapter 5 we propose a strategy which deals with the problem of texture classification seen from different distances and under varying illumination directions. Chapter 6 shows an evaluation of our prediction proposals and an assessment of our model-based classification system. Finally, conclusions and future work are discussed in Chapter 7. Figure 1.3 traces the progression of the investigation from the initial theory and the proposed solutions to the subsequent evaluation and analysis of results.

Chapter 2

Chapter 2 is concerned with the image acquisition process. That is, given a scene description consisting of surface shape and reflectance information, lighting and imaging device, this process describes the formation of an image from the surface. This chapter introduces important concepts such as the Bidirectional reflectance distribution function (BRDF), Gradient space, and the Lambertian model. It also introduces two levels of surface description: single parameter description and histogram representation techniques.

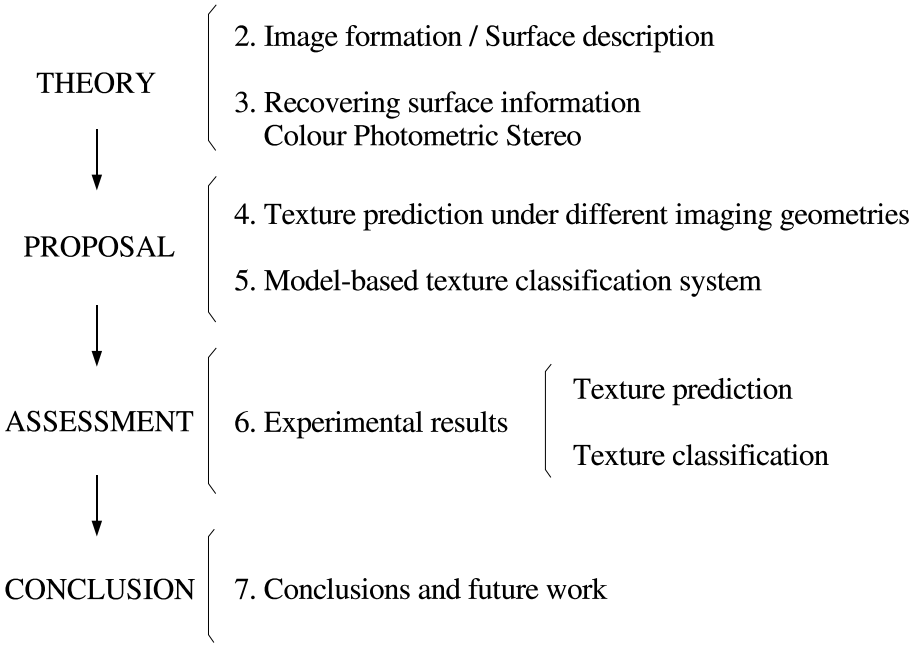


Figure 1.3: Structure of the thesis. Each number indicates the chapter in which each topic can be found.

Chapter 3

This chapter links the learning in Chapter 2 with surface recovery. Main approaches for recovering surface information are reviewed, describing their philosophy. Shape from X techniques are analysed in greater detail, focusing on the photometric stereo techniques. Finally, this chapter suggests colour photometric stereo as a good option to recover surface texture information, describing in depth the colour photometric stereo technique proposed by Barsky and Petrou [8] and the corresponding texture rendering process.

Chapter 4

In Chapter 4 a methodology is proposed to predict how a surface texture will appear if seen from a longer distance. Specifically, two different prediction methods are proposed:

- One which allows us to predict the image intensities under different resolutions directly (direct image prediction).
- Another which allows us to predict first the surface shape information and then the image intensities (image prediction via surface prediction).

The final goal of this prediction process is the generation of a set of textured images for each surface under different imaging geometries. These images are used in our recognition system as models in the training phase of the classification process. That allows us to deal with the problem of distance invariant texture recognition, since we have textured images of the same surface texture at a longer distance and with various light directions.

Summarising, the contributions of this chapter are two prediction techniques which allow us to predict the surface texture information seen under different imaging geometries.

Chapter 5

This Chapter explains the fundamentals of classification, focusing on the principal stages: feature extraction, feature selection and evaluation and also the classification process. After reviewing the theoretical background, we state the problems of texture recognition under varying imaging properties. The described theory and previous work are used to propose a model-based solution which integrates the photometric technique into a complete classification system. This chapter details every stage of the system.

Main contributions of Chapter 5 are:

- The proposal of a model-based strategy to overcome the problem of classifier failure induced by varying imaging properties such as light direction and camera distance.
- The integration of the prediction framework proposed in Chapter 4 into the recognition scheme.

- The capability of the recognition scheme to guess the approximate direction of the illumination used to capture the test images.

Chapter 6

Chapter 6 assesses the performance of our proposals. First, our image database of textured surfaces used as experimental data is presented. Afterwards, an exhaustive evaluation of our prediction methods is performed on these experimental data. Different experiments and error measures are carried out in order to be able to extract and discuss significant conclusions. Next, the proposed model-based classification system is evaluated. Results on estimation of the direction of the illumination are then provided and discussed.

Chapter 6 can be summarised as:

- Presentation of the image database of textured surfaces (experimental data) used in this thesis.
- Evaluation and comparative study of results achieved by the two prediction methods.
- Evaluation of the results obtained on texture classification under varying geometries such as light direction and camera distance.

Chapter 7

Relevant conclusions reached in this thesis work are given in Chapter 7 and future directions are proposed.

Chapter 2

Image Formation

In the last few years some researchers have begun to distinguish between image *texture and* surface *texture. Image texture is what appears in the 2D image of a physical object, while surface texture refers to the variation of the physical and geometric properties of the imaged surface which define the texture in the image. The main objective of this chapter is to describe the transition from physical surface to image texture, taking into account (i) the surface properties of the imaged object, (ii) the illumination arrangement, and (iii) the imaging geometry. Different ways of providing surface roughness description are also presented.*

2.1 Introduction

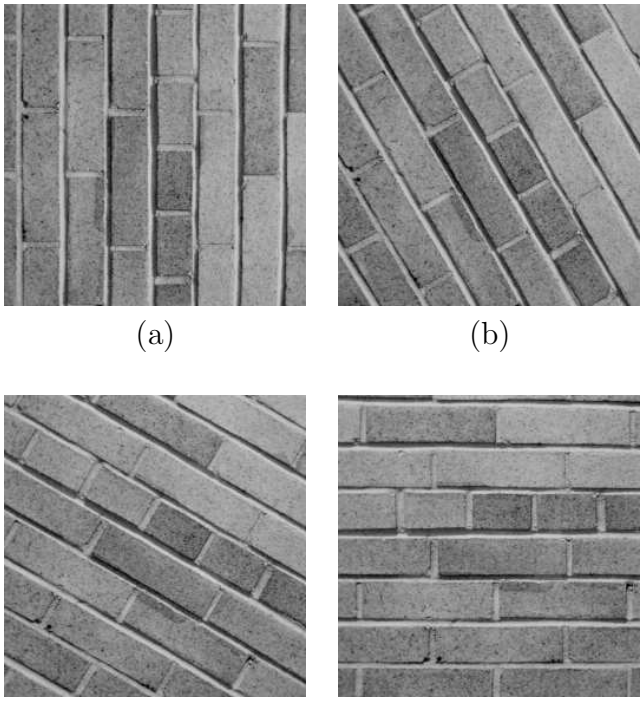
In the past, texture recognition and discrimination has been posed primarily as a 2D problem, assuming that viewpoint and illumination are constant. For instance, some representative techniques are Markov random fields [25, 109] and filter responses [74, 106]. In all of these works, surface normal variations were ignored. Nevertheless, nature shows an abundance of such relief textures and variations produced by the surface relief which cannot be dealt with using a simple brightness normalization or intensity transforms such as histrogram matching. For example, if the surface structure is a ridge, a dark-light transition in one image under one illumination will become a light-dark transition when the light source is moved to the other side of the ridge. Shadows also cause significant problems: two regions will have the

same brightness under one illumination; while the shadowed region will be darker in another.

The complexity in the relationship between the image intensity values to the viewing and lighting settings and the properties of 3D textures led to recent interest in building explict models for 3D textures [33, 34, 85, 97].

In the literature it is often said that texture cannot be defined in a single way [90]. Despite the lack of a universal definition, all researchers agree on two points. First, there is significant variation in intensity levels between nearby pixels; that is, there is non-homogeneity. Second, texture is a homogeneous property at some spatial scale larger than the resolution of the image. Jain et al. [76] proposed the following texture defition: "*Texture is repeating patterns of local variations in image intensity which are too fine to be distinguished as separate objects at the observed resolution*". While Petrou et al. [130] defined texture as: "*a variation in the data at scales smaller than the scales of the objects we wish to identify*". Therefore, it is clear that texture is present in images and, depending on the problem we are trying to solve and the approach we take, it may either be a nuisance or a great help. Texture arises in two ways: due to variation of the reflectance properties of the material of the surface, which could be completely smooth, and due to surface roughness which results in an interplay of shadows and bright patches across the imaged surface.

In realistic situations, both of these factors are usually present. However, in most current texture research studies, texture is treated as an image property and not a property of the imaged surface. In other words, image texture is treated as if it were intrinsic to the imaged scene, and therefore repeatable, and not as something which is the result of the particular combination of the imaged scene and the imaging conditions. As a result, algorithms developed to cope with texture variations usually treat an image as if it has an existence of its own, and in order to develop rotation-invariant texture descriptors, for example, people often rotate the image to check whether the descriptor changes value or not [58, 59, 133, 109, 29]. Figure 2.1 illustrates this fact with a simple example of a texture extracted from the Brodatz album [18].



(a) (b)

(c) (d)

Figure 2.1: Rotation-invariant texture descriptors treat an image as if it has an existence of its own. Therefore, (a)-(d) are considered four different textures although they are the same image rotated in steps of 30°.

In recent years, more researchers have begun to distinguish image texture from surface texture and to realise that the observed image texture is the result of:

- **The surface properties of the imaged object.** These properties can be divided into reflectance properties and surface orientation. The reflectance properties describe how the surface itself reflects incident light. For instance, some surfaces, called *specular* surfaces, have a mirror-like behaviour because the light they reflect depends solely on the angle of the incident light (see figure 2.2.(a)). Alternatively, matte surfaces, called *diffuse* surfaces, reflect light equally in all directions with an intensity which depends on the angle of incident light (see figure 2.2.(b)).

  On the other hand, surface orientation also has an important influence on the image formation. It is obvious that inclining a surface will usually alter the

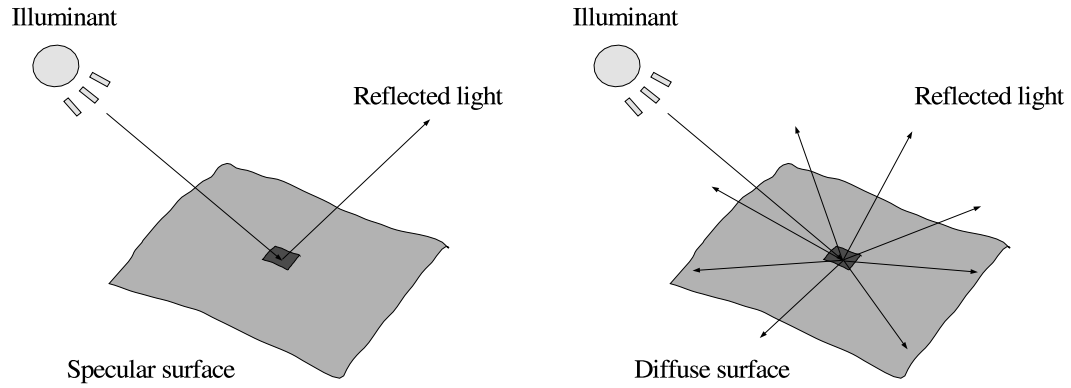


Figure 2.2: Specular and diffuse surfaces. (a) A specular surface reflects all the incident light depending solely on the angle of the incident light. (b) A diffuse surface reflects light equally in all directions.

light reflected towards an observer. Therefore, different intensity values will be captured by the imaging device.

- **The lighting arrangement.** As well as the surface properties of the object, an imaged texture depends on the lighting conditions which affect the analysed surface. For instance, a reduction or an increase in the light intensity may affect the resultant imaged texture. Changes in the direction of the light combined with the surface relief may also modify texture perception. Moreover, other properties such as the spectral distribution of the incident light also contribute to the image formation process.

- **The imaging geometry.** The position of the imaging device with respect to the surface is another factor which plays an important role in the imaged texture formation. For example, an inclination of the camera or a change in distance from the surface imply a different perception of the texture.

  In this thesis, we will assume that the surface's plane always has the same inclination, and specifically will be held perpendicular to the imaging device. See figure 2.3. However, in order to perform texture classification invariant to distance from the camera, we will allow a variation in camera distance.

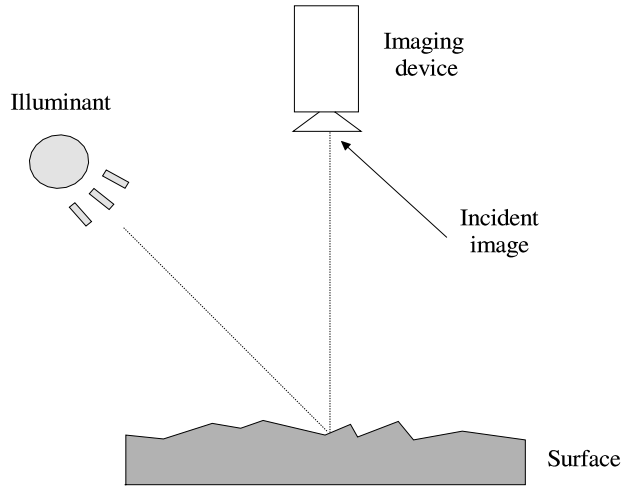


Figure 2.3: Scenario of image acquisition of an illuminated physical surface.

Other mechanisms such as the sensitivity of the acquisition sensor, system optics, and the frame store, have associated problems of a different nature which affect the quality of the final image. Although it is not our intention to investigate the effects of this process in depth, we must take into consideration the image formation process.

It is obvious then that, in order to be able to use texture as a cue to recognition, we must be able to use image texture to extract information about the surface texture irrespective of the lighting and imaging conditions. Two broad approaches may be used for this purpose:

- We may try to identify image statistics invariant to the direction of the light and imaging conditions directly and, therefore, they can be used for surface classification.
- Otherwise, we may try to reconstruct the surface in full detail, and
  - Use this information to compute the statistical properties which can be used directly as cues for classification.
  - Or use the surface information to produce new images under novel conditions and use them as models for the classification system.

In Chapter 3 we will discuss different approaches which allow us to reconstruct surface information from different imaged textures of the same surface.

2.2 Characterisation of the Incident Image

The nature of the process from physical surface to digitised image[1] is illustrated in figure 2.3. This task is defined as follows: given the scene description composed of (1) the surface texture and its topology (the term used to refer to the three-dimensional variation or relief of the physical surface), (2) the surface reflectance function and (3) the light source characteristics and its position, determine the *incident image*. In other words, how the surface appears to the viewer. Finally, given the imaging device, (4) position and other parameters determine the captured image.

Before we start discussing how an incident image is formed, let us introduce some useful terminology.

2.2.1 Terminology

In this section we briefly summarise the basic terms and notations needed to understand the image formation process. Burke presented a survey of all these aspects of image acquisition and image formation in [19].

2.2.1.1 Radiosity Terminology

- *Radiant flux* is the time-rate flow of photons (light energy) from or through a specified location of space, independent of direction.

- *Radiant intensity* is the measure of a source's ability to illuminate an object from a distance. It is defined as the flux through a volume of space, specified as a solid angle, in a specified direction.

[1]The digitised image is the final result of the entire acquisition process, thus this image may be interpreted as a degraded version of the incident image which has been affected by the imaging device. It is important to note that this image forms the basis of all subsequent processes and, therefore, is often referred to as the captured image.

- *Radiance* is the radiant intensity divided by the area of the emitting surface. The radiance of a point source is equal to its radiant intensity.

- *Radiant incidence* or *irradiance* is the total photon flux within a patch divided by the area of the patch. Irradiance describes the flux per unit area which is perpendicularly incident (normal) to a surface.

- *Albedo*, or *surface reflection coefficient*, is the average ratio between the reflected and incident flux, independent of angles.

2.2.1.2 Geometry of the Imaging Setup

In most applications which try to recover surface texture properties, two assumptions are usually made for the imaging setup: first, the surface is lit by a single distant point of light and second, the surface is viewed from a distance. Thus, the light and viewing directions are constant over the whole surface. Moreover, the surface is expected to be small, relative to the distance between the camera and the physical sample, assuming that the image texture contains the three-dimensional variation or relief of the surface. Orthogonal viewing is also often assumed: that is, the principal surface normal coincides with the z axis, which is the viewing direction $\mathbf{R}$ (boldface font will denote vectors). Thus, the coordinates (x, y) on the surface and imaging planes are the same.

It is sometimes convenient to think of a surface as a collection of planar facets, where each one has its own surface orientation $\mathbf{N}$, as shown in Figure 2.4. Let $S(x, y)$ be the functional form of the surface and let us assume that the surface is globally horizontal or flat. In other words, the mean value of $S(x, y)$ (its partial derivative fields) equals approximately 0. The local surface normal $\mathbf{N}$ could be expressed in terms of the gradient (p, q),

$$\mathbf{N} = \frac{1}{\sqrt{p^2 + q^2 + 1}}(p, q, -1)^T \tag{2.1}$$

where

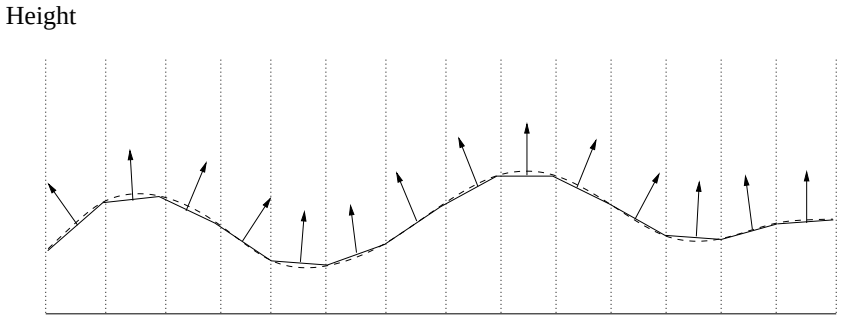


Figure 2.4: A surface observed as a collection of planar facets where each one has its own surface orientation **N**.

$$p(x,y) = \frac{\partial S(x,y)}{\partial x} \qquad and \qquad q(x,y) = \frac{\partial S(x,y)}{\partial y} \tag{2.2}$$

2.2.1.3 Reflectance Maps and BRDF

The intensity of an image at position (x, y) is given by a reflectance map introduced by Horn [66]. The reflectance map combines information about surface material, scene illumination and viewing geometry into a single representation which determines image brightness as a function of surface orientation. Reflectance maps could be derived from formal reflectance models or measured empirically. Thus, for any particular surface and imaging setup, there is a unique reflectance map which is a function of a surface gradient.

The reflectance and geometric information combined in a reflectance map, can be separated into an orientation term and a surface reflectance factor ρ (albedo). The geometric dependence is expressed by introducing the *bidirectional reflectance distribution function* (BRDF) of the surface. The BRDF is a function which characterises the material properties of a surface and depends on the incident and viewing directions. It is defined as the ratio of the radiance of reflected light in the receiving direction to the irradiance of the incident light.

2.2.1.4 Reflectance Models

As we know from everyday experience, different real materials reflect light differently and there is a variety of reflectance models designed to capture the main characteristics of light reflection. Chen et al. [26] presented a comprehensive review of reflectance models.

The simplest form of reflection is the ideal specular reflection (mirror-like reflectance), when all incident light is reflected depending solely on the angle of the incident light. However, one of the most important and widely used reflectance models is the Lambertian model, introduced as early as the 17th century. The Lambertian surface reflects light equally in all directions and the amount of reflected light depends on the angle at which the light falls on the surface. The generalisation of this model is reported by Oren and Nayar in [122].

Reflectance models could be roughly classified into two classes: those based on physical optics and those based on geometrical optics. The models of the first group are based directly on electromagnetic wave theory, whereas the geometrical approach uses the assumption that the characteristic surface roughness is much larger than the wavelength of the incident light and, therefore, light travels in a straight line. This simplification yields a simple and more compact functional form of the reflectance model.

2.2.2 From Physical Surface to Incident Image

The objective of this section is to model the transition from physical surface to incident image. As described in section 2.1, the intensity of a point (or pixel) in a 2D projection of a physical surface depends on surface properties and the lighting and imaging geometries.

Nevertheless, one of the most important terms in the formation of the incident image is the *surface reflectance function*. Roughly speaking, it describes how an elementary surface patch reflects incident light. This is expressed in terms of the bidirectional reflectance distribution function (BRDF) which is constructed from

the ratio of the patch radiance to its irradiance.

The BRDF is spectral-dependent (i.e. it depends on the wavelength λ of the electromagnetic spectrum), but in practice it is often approximated by independent BRDFs, one per colour channel for an RGB image. For grey-level images, there is no need to take into account more than one BRDF since only one channel is used.

The phenomenon of reflection can be produced by one or more different physical processes. These underlying processes, which are beyond the scope of this thesis, lead to various generic behaviours, depending on how a surface reflects the incident light. It is interesting to mention again the two particular cases which illustrate extreme behaviours. First, a perfect mirror or a completely specular material which has a reflectance function equal to 1 in one direction and 0 in all others. Second, a perfect matte or Lambertian surface which reflects equally in all directions with the amount of reflected light depending only on the incident light. Many surfaces with more complex behaviour can have their reflectance described as having both Lambertian and specular reflectance properties, and are commonly known as *hybrid surfaces*.

As it has been previously mentioned, the surface reflectance measures the fraction of emitted light in the observer's direction. However, all surface facets reflect the incident light not only towards the viewer but in other directions as well, including those towards other surface facets. Thus, the light reflected on each facet represents a possible source of secondary illumination for other facets. The effects of illuminating a given surface facet with light reflected from other surface facets is called *interreflection* [121] and corresponds to the situation in which the tight connections between local surface properties interplay: a pixel value contains some contribution driven by properties from other surface facets. This was analysed in depth by He et al. [62], who proposed a classification of the reflection from an arbitrary surface consisting of first-surface reflections and multiple surface and/or subsurface reflections (see figure 2.5). The first-surface reflection process is described by physical optics and is strongly directional, while the multiple surface (light bouncing several times off different facets) and subsurface reflections (light penetrating the surface

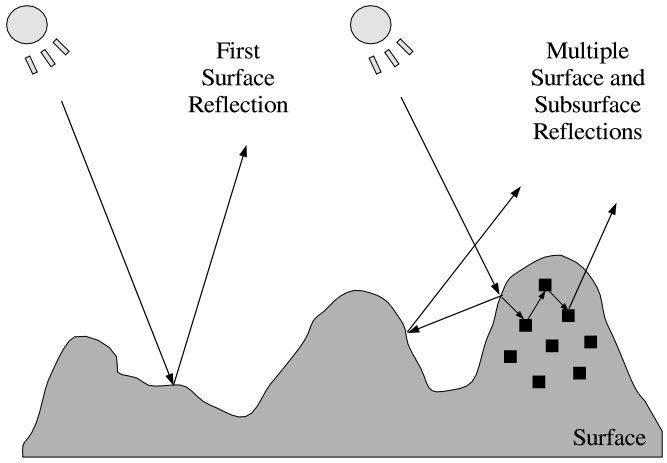


Figure 2.5: Reflection process at a surface [62]. The first-surface reflection process is strongly directional. Multiple surface reflection occurs when the light bounces several times off different facets. Subsurface reflection occurs when light penetrates the surface and is reflected from inner material non-homogeneities.

and reflected from inner material non-homogeneities) are geometrically complex but may be expected to be less strongly directional than the first-surface reflected light. Fortunately, interreflections do not usually affect a pixel value significantly and, in almost all applications, they are assumed not to exist at all. The consequences resulting from this fact depend on the technique used and the desired accuracy of the task, but, at any rate, this assumption always entails some errors.

2.2.3 Dichromatic Reflection Model

In many computer vision and graphics applications, reflectance models are represented by linear combinations of two reflectance components: the diffuse component and the specular component, normally called diffuse reflection and specular reflection respectively.

Consider, for example, a dielectric opaque object which is illuminated by one light source. Some of the incident light is reflected from the surface of the object in a mirror like manner and some enters the object's material body where it is partly absorbed, partly transmitted, and partly reflected from nonhomogeneities inside

the matter. Thus, some of the light which has entered the material body of the object is reflected back and exits from the surface. If the nonhomogeneities inside the object's body have a random nature, there is no preferred direction and the exiting light has a diffuse nature. Therefore, we can talk about the *surface* and *body* reflection components. The theory developed to describe the spectral distribution of reflected light for such surfaces is called the dichromatic reflection model [84] formally introduced by Shafer [144]. We find this reflection model in a number of works and applications, such as the development of the photometric stereo method for non-Lambertian surfaces using colour information [143].

Let us denote the reflectance function of the surface by $\mathcal{R}$. It describes the fraction of light reflected towards the camera depending on spectral and geometric scene parameters: $\mathcal{R}(\lambda, \mathbf{N}, \mathbf{L})$.

It could be separated in the material body (diffuse-like) reflectance component and the material surface (mirror-like) reflectance component:

$$\mathcal{R}(\lambda, \mathbf{N}, \mathbf{L}) = \mathcal{R}_B(\lambda, \mathbf{N}, \mathbf{L}) + \mathcal{R}_S(\lambda, \mathbf{N}, \mathbf{L}) \tag{2.3}$$

The spectral reflectance properties of each of the components can be separated from their geometric reflectance properties. The dichromatic reflection model describes each component as the product of a spectral power distribution $c_\beta(\lambda)$, and a geometric scale factor, $m_\beta(\mathbf{N}, \mathbf{L})$, which determines the intensity of the reflected light, where $(\beta = S, B)$, so that equation 2.3 becomes:

$$\mathcal{R}(\lambda, \mathbf{N}, \mathbf{L}) = m_S(\mathbf{N}, \mathbf{L}) c_S(\lambda) + m_B(\mathbf{N}, \mathbf{L}) c_B(\lambda) \tag{2.4}$$

The matte (body) geometric scaling factor m_S could be successfully modeled by the Lambertian cosine law:

$$m_B(\mathbf{N}, \mathbf{L}) = \mathbf{N} \cdot \mathbf{L} \tag{2.5}$$

The glossy (surface) geometric scaling factor could be modeled by a number of available reflectance models, e.g. the Torrance-Sparrow model [158] and the Phong

model [131], often used in computer graphics, and different generalisations of known models, e.g. Nayar et al. [118], Tagare and deFigueiredo [156], etc.

The intensity of light registered by a sensor (or by an array of sensors) depends on both the intensity and spectral distribution of the reflected light and the spectral sensitivity function of the sensor. If the sensitivity function of the sensor is $\mathcal{S}(\lambda)$, the reflectance function of the surface is $\mathcal{R}(\lambda, \mathbf{N}, \mathbf{L})$, and the spectral distribution of the incident light is $\mathcal{L}(\lambda)$, the intensity value registered by the sensor is:

$$I = \int \mathcal{S}(\lambda)\mathcal{L}(\lambda)\mathcal{R}(\lambda, \mathbf{N}, \mathbf{L})d\lambda \tag{2.6}$$

Using equation 2.4 the above becomes:

$$\begin{aligned} I &= \int \mathcal{S}(\lambda)\mathcal{L}(\lambda)[m_S(\mathbf{N}, \mathbf{L})c_S(\lambda) + m_B(\mathbf{N}, \mathbf{L})c_B(\lambda)]d\lambda = \\ &= m_S(\mathbf{N}, \mathbf{L}) \int \mathcal{S}(\lambda)\mathcal{L}(\lambda)c_S(\lambda)d\lambda + m_B(\mathbf{N}, \mathbf{L}) \int \mathcal{S}(\lambda)\mathcal{L}(\lambda)c_B(\lambda)d\lambda \end{aligned} \tag{2.7}$$

If we denote the surface spectral response by $\mathcal{D}_\beta$, where $(\beta = S, B)$:

$$\mathcal{D}_\beta \equiv \int \mathcal{S}(\lambda)\mathcal{L}(\lambda)c_\beta(\lambda)d\lambda \tag{2.8}$$

Then the pixel value I can be represented by the following sum:

$$I = m_S(\mathbf{N}, \mathbf{L})\mathcal{D}_S + m_B(\mathbf{N}, \mathbf{L})\mathcal{D}_B \tag{2.9}$$

The spectral response values depend on the reflectance properties of the surface and the spectral properties of the illumination and the camera. If the illumination and the sensor sensitivity are constant, then $\mathcal{D}_S$ and $\mathcal{D}_B$ depend only on the reflectivity of the surface and they could be considered as colour. This notion of colour incorporates the spectral properties of the illuminant, the camera and the surface and is the maximum pixel intensity this particular surface patch may produce for this particular camera and this light arrangement. This colour would have been the

intrinsic colour of the surface if the light were white (i.e. $\mathcal{L}(\lambda)$ were a constant) and the sensitivity of the sensor were a delta function ($\mathcal{S}(\lambda) = \delta(\lambda)$).

On the other hand, the scalar functions $m_S(\mathbf{N}, \mathbf{L})$ and $m_B(\mathbf{N}, \mathbf{L})$ show how the pixel intensity depends on the geometrical properties of the scene only and are in fact *shading parameters*.

If one camera cell consists of more than one sensor (i.e. K sensors), we can consider a pixel as a K-dimensional vector. In this case, the spectral surface responses are also K-dimensional vectors:

$$\mathbf{C}_B = (\mathcal{D}_{\beta 1}, \ldots, \mathcal{D}_{\beta K}) \tag{2.10}$$

where ($\beta = S, B$) and the second subscript identifies the spectral band associated with each of the different sensors.

Consider a situation when the same surface patch is imaged under different light directions and/or different orientation with respect to the camera. Then, since we change only the geometry of the scene, and keep its spectral properties the same, all possible pixel vectors form a *dichromatic plane* [84] in the K-dimensional colour space which is spanned by vectors $\mathbf{C}_B$ and $\mathbf{C}_S$.

A simple application of the above ideas was the work on *Shape from shading*, which may be considered as the precursor to photometric stereo: an image irradiance equation is developed to determine image irradiance as a function of surface orientation. For example, for a Lambertian surface $m_S = 0$ and m_B is given by equation 2.5 and, for the case of a single sensor, this equation takes the familiar form

$$I = \frac{pL_x + qL_y + L_z}{\sqrt{p^2 + q^2 + 1}}\rho \tag{2.11}$$

where ρ stands for $\mathcal{D}_B$ and corresponds to the surface albedo, (L_x, L_y, L_z) is the illuminant vector $\mathbf{L}$, while p and q define the gradient components for every point on the surface.

This equation cannot be inverted locally because image brightness provides only

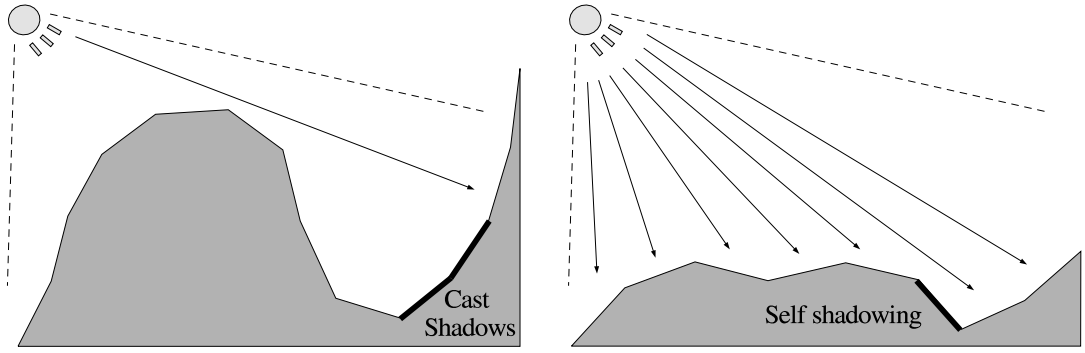


Figure 2.6: Cast and self shadowing. (a) A cast shadow occurs where one part of the surface prevents another from being illuminated by blocking the direct path between light source and shadowed area. (b) A self shadowing occurs when a facet is oriented so that it does not present an area on which light is incident.

one measurement, whereas surface orientation has two degrees of freedom. Brightness values obtained from the same viewpoint, but under different conditions of illumination, allow one to obtain local estimates of surface orientation without requiring additional a priori information such as global smoothness assumptions or prior image segmentation. This is the basic principle of Shape from shading techniques.

The Lambertian model, which is often used for its simplicity and linearity, states that the radiance of a point on a surface falls with the cosine between the surface normal and the viewer direction. Consequently, the camera pixel brightness value is dependent only on the relative relation between the surface normal **N** and the light source **L**. Therefore, assuming that the surface has a Lambertian reflectance function, which is homogeneous over the entire surface, and assuming that the surface is not significantly affected by cast or self shadowing, the pixel intensities of the so-called Lambertian image are obtained by means of equation 2.11.

Figure 2.6 illustrates the concepts of cast and self shadowing. For simplicity's sake, both terms are not often encountered, assuming that they do not affect the analysed surface to any great extent.

2.3 Surface Information

In the context of this thesis, the fact of identifying surface properties and understanding how they affect the image in the presence of light is relevant to the following points.

- First, understanding the mechanism of light reflection is an important issue for many computer vision algorithms. In particular, for photometric based methods which extract differential shape parameters (gradient vectors) and reflectance properties (albedo) from the intensity of reflected light.

- Second, extraction of surface geometry and reflectance properties is crucial for the success of surface modelling. The next chapter will describe how these surface properties can be extracted by means of photometric stereo and used as input data for creating virtual images of the surface texture. These virtual images will be the basis of the texture classification approach proposed in Chapter 5.

- Third, recognising surface characteristics such as roughness, isotropy or directionality may be used to group each surface texture into its corresponding type of surface. This may help us to present the experimental results for each group of surface texture separately, showing, for example, which textures are more difficult to classify.

In what follows we introduce different ways of describing a surface, starting with the concept of gradient space.

2.3.1 Gradient Space

Every surface facet has a single normal vector, which can be described by vector $\mathbf{N} = (n_x, n_y, n_z)$. If only the orientation of the surface is considered, the normal vector can be represented by

$$\left(\frac{n_x}{n_z}, \frac{n_y}{n_z}\right) = (p, q) \tag{2.12}$$

where the pair (p, q) is the two-dimensional gradient representation of the orientation. Interpreting the image plane as $z = 0$, the origin of gradient space corresponds to the vector $(0, 0, 1)$, normal to the image, implying a surface parallel to the image plane.

In essence, the concept of gradient space facilitates the mapping of an array of surface normals to a series of coordinate points (p,q) within the two-dimensional (2D) gradient domain, where p and q describe the local surface slope or gradient in two orthogonal directions at a given location. By mapping such an array of surface normals into gradient space, an indication of the global surface distribution and, hence, a description or signature of the total observed surface shape can be obtained. This is a concept of general use in many analytical problems.

For example, if the topography of the surface can be estimated, it may be possible to improve the performance of an intensity classifier by classifying the surface derivatives instead. This was done by McGunnigle et al. [113], who proposed a surface classification scheme using the gradient information obtained with photometric stereo. In their approach, a local estimate of the bivariate gradient distribution at each point on the surface was calculated. The eigenvalues of this distribution were then computed and used as the basis for classification.

On the other hand, some works [151, 152, 150, 102] used this surface information as a powerful tool in systems devoted to quality control using image processing and computer vision techniques. The mapping of surface normal data to the gradient space domain offers a mechanism for the representation of global surface shape. This information is used to detect surface irregularities such as bumps, scratches, tooling marks and other blemishes. Let us illustrate this by means of a simple example. Consider a planar surface arranged in a specific viewing direction, as shown figure 2.7.(a). In the case of an idealised flat surface, all surface normals will be parallel to the viewing direction. Hence, when plotted in 2D gradient space, and assuming orthogonal projection, all mapped normals will appear as a cluster of

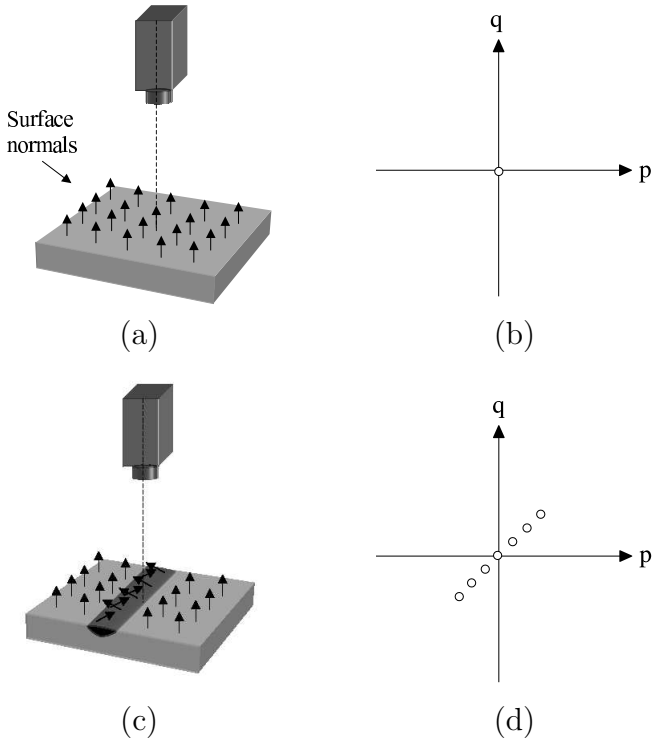


Figure 2.7: Examples of surface gradient distributions.

points located precisely at the origin. Using this representation, we may consider such a grouping as an impulsive distribution, as figure 2.7.(b) illustrates. If we now consider the presence of a local discontinuity, as depicted by the depression in figure 2.7.(c), then the corresponding gradient space will be altered accordingly, and appear as shown in figure 2.7.(d). It is the character of this gradient distribution which provides an indication of the nature of the observed surface and allows us to detect surface defects in industrial processes, for example.

2.3.2 Surface Roughness Description

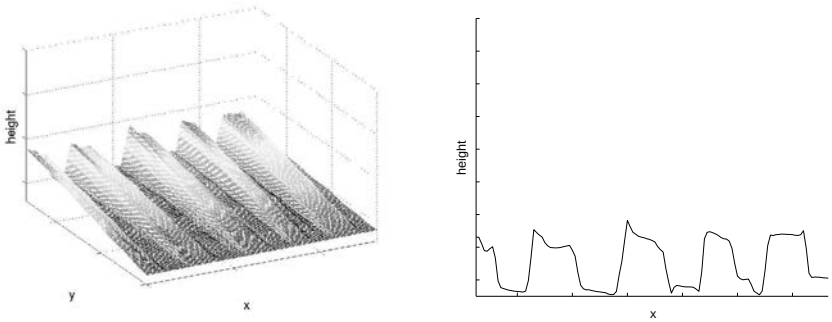
As it has been described, the way in which light is reflected by a surface is dependent on the shape characteristics of the surface, among other factors. A smooth surface, for instance, may reflect incident light in a single direction, while a rough surface will tend to scatter light in various directions, maybe more in some directions than others. In many computer vision techniques, such as Shape from shading, it is often

assumed that the surface has to be globally flat. When this assumption is omitted, and the technique is applied to other types of surface, such as very rough surfaces, the algorithm fails. Therefore, it would be interesting to find a way of characterising and describing a surface in order to distinguish it from different types of surface.

A description of a surface can be made on several levels. For instance, a single parameter may be sufficient to characterise a surface for some purposes, whereas in other cases a much greater degree of description is required. The descriptors introduced in this section will form the basis for modelling all the surface textures used throughout this thesis.

Surfaces can be modelled by their height distribution or their slope distribution [118]. Our work focuses on real surface textures and implicit in this fact is the necessity of knowing or experimentally calculating the surface height map, $S(x, y)$. See equation 2.2 on page 22. This is a complicated task with some associated problems, as the measures of these parameters depend on the instrument and the separation of sampling points [13], or, in contrast, on the reconstruction technique used to recover the height information. As the final goal of this thesis is not in the direction of 3D reconstruction but the classification of surface textures, we basically focus on surface descriptors which use gradient information to characterise a surface.

Three different levels can be used in order to characterise a surface [111]. A first level of description seeks to estimate some property of the surface, i.e. height or gradient, with a single paramenter (first order statistics). On a second level, a statistical model, such as the histogram, is applied to the variation of height or gradient, providing a more visual comprehension of the surface's characteristics. At a third level, there are those techniques which incorporate spatial interaction, such as the Power Spectral Density (PSD) or the Autocorrelation Function (ACF). In this sense, the statistical parameters of the height distribution have been exploited in conjunction with some spatial information. For example, the two point height probability density function, analogous to the co-occurrence matrices used in texture analysis, allows the description of correlation and structure within the surface. This third level of representation will not be considered here.



(a) (b)

Figure 2.8: Surface profile. (a) Surface height map of a surface. (b) Corresponding profile of a vertical cross section.

2.3.2.1 First Order Statistics

Historically, a common method of surface measurement has been to mechanically measure the height of the surface along a line across the surface. The resulting series of values are collectively known as the surface profile (see figure 2.8). Given a surface profile, the most basic descriptor requires the use of only one parameter. Two of the most common measures of roughness are the *root mean square roughness* (RMS) and the *centre line average* or *average roughness* (R_{ave}). Both are defined with respect to surface profile on either the x or the y axis.

$$RMS = \sqrt{\frac{1}{N}\sum_{x=0}^{N-1}\left[s(x) - \overline{s(x)}\right]^2} \tag{2.13}$$

$$R_{ave} = \frac{1}{N}\sum_{x=0}^{N-1}|s(x)| \tag{2.14}$$

where $s(x)$ represents the height of the surface at point x along the profile, $\overline{s(x)}$ is the average height of the surface profile and N is the number of columns of the surface height map.

However, as it has been explained in the context of this thesis, it is more suitable to use parameters which concentrate on the slope of the facets (gradient vectors) rather than their height. Therefore, the parameter known as *absolute average slope*

ratio ($ASSR$), which provides an easy way to evaluate the level of irregularities of a certain surface is of particular interest. Assuming that we have a method which enables us to estimate the derivative fields of the surface, $AASR$ may be calculated as

$$AASR = \frac{1}{2NM} \sum_{x}^{N-1} \sum_{y}^{M-1} |p(x,y)| + |q(x,y)| \tag{2.15}$$

where $N \times M$ is the number of samples contained in each partial derivative. This parameter is used in Chapter 6 to characterise the degree of roughness of a given surface texture providing discrimination between relatively smooth surfaces and very rough ones.

2.3.2.2 Histograms

The histogram description, which can be applied to the variation of height or gradient, can be thought of as an extension of the single parameter descriptor. While the parameters of section 2.3.2.1 concentrate on estimating the standard deviation or mean of the heights of the surface, for example, the histogram represents a statistical model of the surface.

If the gradient components are used to compute the histogram of the surface, since there is a linear relationship between surface gradient and surface height, the characteristics observed in the histogram are also valid for describing the surface height map. Therefore, the use of histogram description not only implies a new degree of visual descrimination between surfaces, but also a certain degree of modelling.

For example, the estimated probability density functions (PDFs) for the surface partial derivatives p and q (i.e. the normalised histograms of these quantities) may be used to characterise the surface shape and roughness of each surface texture. Moreover, this allows us to group the surface textures according to their roughness (relatively smooth and very rough surfaces) and topology (isotropic and directional surfaces).

Table 2.1: Important types of surface.

Diffuse surface; Lambertian surface	A surface which reflects light equally in all directions.	Cotton cloth; many paints and papers; unfinished wood, brick, etc. Surfaces whose apparent brightness does not change with viewing direction.
Specular surface	A surface which reflects all incident light, depending solely on the incident angle (mirror-like behaviour).	Mirrors; polished and bright metals such as chromium and aluminium.
Hybrid surfaces	A surface which has both Lambertian and specular behaviour.	Many diffuse surfaces which have small bright patches (specularities).

2.4 Summary

The aim of this section is to briefly summarise the more important concepts introduced in this chapter. It is important to state that this chapter has presented terms and notations which are basic to understanding the following Chapters of this thesis correctly.

We have seen that in the past few years more researchers have begun to distinguish between image texture and surface texture. However, in many texture studies, texture is treated as an image property and not as something which is the result of the particular combination of the imaged scene and the imaging conditions. In this thesis we use image texture to extract information about the surface texture, irrespective of the lighting and imaging conditions. This surface information will allow us to render new images under novel conditions and use these as models for the texture classification system.

We have explained that an image texture is the result of the **surface properties** of the imaged object, the **lighting arrangement** and the **imaging geometry**. The

surface properties can be divided into reflectance properties which describe how the surface itself reflects incident light and surface orientation. Table 2.1 summarises the different types of surface depending on how they reflect the incident light. The basic terms and notations needed to understand the image formation process have also been defined. For instance, we have introduced the concept of a **reflectance map** and the **bidirectional reflectance distribution function** (BRDF). Since real materials reflect light differently, there is a wide variety of reflectance models designed to capture the main characteristics of light reflection. The **dichromatic reflection model** has been described, detailing how an intensity value is formed. From this model we also describe how the intensities for a Lambertian surface can be obtained.

We have also analysed in this chapter different ways of obtaining a surface roughness description. Basically, two different approaches can be used for this purpose: one based on height distribution, and another based on gradient distribution. As in our work we will focus on real surface textures and, implicit in this fact, there are some associated problems such as correctly calculating the surface height map, we propose to focus on surface descriptors based on gradient information. It is important to remember that the final goal of our work is not 3D reconstruction but the classification of surface textures.

From gradient information, a surface roughness description can be made by using a single parameter such as the **absolute average slope ratio** ($ASSR$) or a statistical model applied to the variation of gradient such as the estimated **probability density functions** (PDFs) for the surface partial derivatives p and q. These two descriptors will allow us to discriminate between relatively smooth surfaces and very rough surfaces as well as to classify them according to their topology such as directionality or isotropy.

Chapter 3

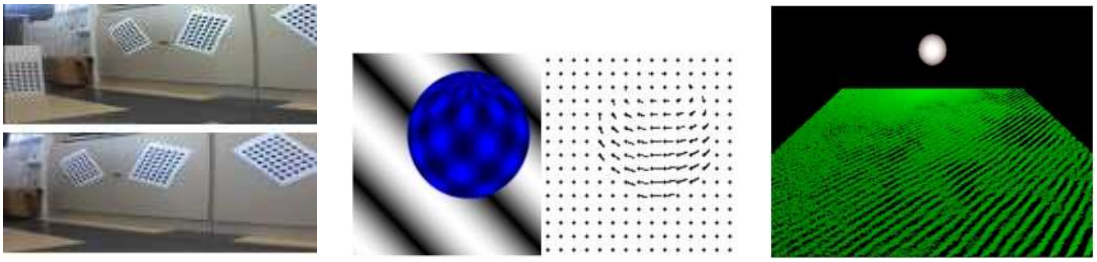
Recovering Surface Information

There are several methods which attempt to determine surface information by measuring parameters calculated from images of the illuminated object. These methods are reviewed in this chapter, focusing especially on photometric stereo approaches. Among these, the 4-source colour photometric stereo technique, which allows us to recover the normal vectors and the reflectance properties for every point of the surface, is described in depth. Although the aim of this technique is to recover surface information, our final goal is focused on reversing the process of surface recovery towards image prediction: to use the recovered surface information to render images of that surface under novel lighting conditions and imaging geometries.

3.1 Introduction

There are many techniques which allow us to obtain 3D information about surfaces. These methods are available in what Woodham [170] refers to as direct methods and indirect methods. Direct methods are those which try to measure distance ranges directly as in pulsed laser based systems, for example, where depth information is the only information available. Indirect methods are those which attempt to determine distance by measuring parameters calculated from images of the illuminated object.

Several indirect methods have been developed; for example, *stereo vision* [48] uses triangulation to compute distance. By using *optical flow* [5], one can compute the relative distance to points on the surface of an object. Therefore, surface vectors



(a) (b) (c)

Figure 3.1: Indirect methods. (a) Based on stereo vision which uses triangulation to compute distance. (b) Based on optical flow which computes the relative distance to points on the surface of an object. (c) Based on Shape from X which obtains surface shape by analysis of the radiometry of image formation.

can be constructed based on how image points flow from one time to the next. Stereo vision and some optical flow techniques require a relationship between surface points in one frame and the same points in another frame, taking us to the well known correspondence problem [177]. Another indirect method for obtaining surface shape is by analysis of the radiometry of image formation. Commonly known as *shape from shading* (although, we should refer to it as *Shape from* X, where X is one of a number of options considering the spread of such technologies in the last few years), this technique can be applied wherever the direction of incident light is known and/or can be controlled. Figure 3.1 illustrates these groups of indirect methods proposed by Woodham [170].

As our main objective is to recover surface texture information (surface shape and surface reflectance properties), it seems more suitable to use a Shape from X technique which allows us to recover both information. Moreover, the other techniques such as stereo vision and optical flow, suffer from the correspondence problem [177] like conventional stereo.

The recovery of shape and material information from images is extremely challenging and is far from being solved in complete generality. A number of approaches which promise to be useful, although not completely general, have been pursued over a long period of time. A well-known objective in computer vision has been to develop lighting models and rendering procedures to produce synthetic images which are vi-

sually and measurably indistinguishable from real-world images. Computer models of architectural scenes have been especially popular subjects and have proven to be entertaining virtual environments as well as valuable visualization tools. While computer vision usually aims to derive a 3D scene description from its 2D images, it is important to clarify that the objective of our approach is not in that direction. Certainly, it is aimed at reversing the process of surface recovery towards image prediction, i.e. using a surface model to render textured images of that surface under novel lighting conditions and imaging geometries; afterwards, these images will become the models of the classifier and, if accurate enough, will provide better classification results.

3.2 Shape from X

Over the last three decades, different methods to extract important information from images have been largely studied for surface recovery purposes. Such techniques, commonly referred to as *Shape from X*, have been developed where X is one of a number of options. This family of techniques includes methods such as shape from stereo, shape from motion, shape from focus/defocus, shape from zoom and shape from intensity. All these methods can be simultaneously itemized into the following groups:

- Techniques based on multiple views, such as shape from stereo and shape from motion. Shape from stereo is based on solving the correspondence problem between two or more views of a given surface taken from different locations, while shape from motion exploits the relative motion between camera and scene.

- Techniques based on modifying the intrinsic camera parameters, i.e. shape from focus/defocus and shape from zooming. In shape from focus/defocus, the images are obtained by changing the camera parameters (usually the focal setting or the image plane axial position), and taken from the same point of view. The difference between shape from focus and shape from defocus is that,

in the first case, it is possible to dynamically change the camera parameters during the surface estimation process, while in the second case, this is not allowed [49]. Besides, shape from zoom considers the use of multiple images taken by a single camera coupled with a motorized zoom.

- Techniques based on extracting shape information from a series of intensity images, assuming each is generated by a single light source. These techniques are commonly known as shape from intensity and can be further divided into three subcategories: shape from shading, shape from photometric stereo and shape from photometric sampling. Basically, the differences between them are in the number of images and arrangement of light sources.

The last of these groups of thecniques is particularly appealing since it does not require additional hardware beyond that used for classification. In the following, we focus on shape from intensity techniques.

Shape from shading uses a single light source, i.e. one image as input, to recover the shape information [65]. It has the advantage of requiring the least amount of input. However, it also introduces evident disadvantages. Since there is less image information available, the method is less accurate: at each pixel, intensity provides only one constraint, whereas the description of surface shape (surface gradient or surface normal) requires at least two parameters. Therefore, many shape from shading techniques introduce constraints, such as smoothness of the surface as well as using optimisation methods to estimate shape [126, 63, 94, 80]. Shape from shading is often not as reliable as other Shape from X techniques since it is very easy to be confused by reflections or for it to fail through poorly modelled reflectance functions.

To overcome some of these problems, shape from photometric stereo was introduced by Woodham [170, 171]. Photometric stereo is based on the fact that image intensities depend on the surface orientation and its reflectance. Hence, if several images are taken from the same viewing position but with different lighting directions, variation of pixel intensities in these images will be due to changes in the relative positions of the light and the surface [172]. This is shown in Figure 3.2,

(a) (b) (c)

Figure 3.2: Different images of a surface captured from the same viewing position but with different lighting directions. Variation of pixel intensities in these images (a)-(c) are due to changes in the relative positions of the light and the surface.

where each image of the scene provides one constraint on the surface shape. Hence, multiple images of the same scene create an overconstrainted system which is solved for the surface shape in order to minimise the total cost. Therefore, these constraints permit us to calculate the normal vectors which represent the surface orientation of any point on the surface and the reflectance factor or albedo which describes the reflection properties of the surface.

Another technique similar to photometric stereo is shape from photometric sampling [142, 120]. It usually uses many light sources instead of a few, and a sequence of images corresponding to each light source. The use of extra light sources eliminates the inaccurate results caused by improper choices of light position in classical photometric stereo. However, it shares so many similarities with photometric stereo that both techniques are commonly considered as the same in the literature.

Over all these Shape from X techniques, the photometric stereo has been chosen in this thesis as a method to recover surface information. It has various advantages over all other methods as it does not require additional hardware beyond that used for classification, it does not make strong assumptions about the underlying surface structure like shape from shading does, for example, and it allows one to recover both local colour and local gradient, which permit us to render new images of the same surface texture under different lighting directions. Moreover, photometric stereo allows us to flag the places where some of its assumptions break down and the

recovery is impossible. As we will see in the next chapter, this is another important factor to consider in our prediction proposals and classification approach.

3.3 Photometric Stereo: Related Work

Shape from photometric stereo was conceived by Woodham in the early eighties and has received an extensive theoretical and experimental treatment [171]. Initially, the method was based on the use of the so called *reflectance maps* (introduced in Section 2.2.1) in the form of look-up tables. These tables were calculated beforehand by means of a calibrating sphere. This was the most common way of empirically measuring the reflectance maps, which determine image brightness as a function of surface orientation. Therefore, if the reflectance factor or albedo at each surface point is known, the surface gradient can be solved by using two input images. In case the albedo is not known, both gradient and reflectance factors are solved for by using an additional image. This method is simple, sensitive to noise and efficient only for Lambertian surfaces.

All approaches published since then can be classified according to the assumptions the authors make about the surface they are dealing with and the type of problem they want to solve [8]. For example, one may assume that a rough surface is Lambertian, or, alternatively, its reflectance is of a more general nature. The reflectance can be constant or varied over the surface. Furthermore, we then may try to recover the surface characteristics using either greyscale or colour images. Although we may think the use of greyscale images is adequate for surfaces with uniform reflectance, it turns out that, unless the surface is Lambertian, the use of colour images has something to offer, even in this case, for dealing with highlights, for example. The photometric stereo technique for greyscale images has been well-researched over the last 20 years [148, 69, 30, 127, 120, 79, 156, 122, 119, 111]. The use of colour, on the other hand, had not been investigated properly until the nineties [42, 43, 87, 27, 44, 45, 7].

In what follows we review the photometric stereo approaches taking into account the above described critera (see table 3.1) and grouping them into techniques which

Table 3.1: Taxonomy of photometric stereo problems proposed by Barsky and Petrou [8].

Method	Type of surface reflectance			
	Lambertian		General	
Greyscale Images	uniform albedo (**I**)	variable albedo (**II**)	uniform albedo (**III**)	variable albedo (**IV**)
Colour Images	uniform albedo (**V**)	variable albedo (**VI**)	uniform albedo (**VII**)	variable albedo (**VIII**)

use greyscale and colour images.

3.3.1 Greyscale Images

Basically, the major part of existing work in the photometric stereo field deals with greyscale images. For Lambertian surfaces (with uniform and variable albedo, cases **I** and **II**) the photometric equations are linear. This allows us to formulate the problem in matrix form. Inverting a system of linear equations makes it possible to recover the unknown albedo as well as the gradient from three image intensities for every surface patch. Therefore, by using three images of a rough Lambertian surface in the absence of shadows, one can successfully separate the surface shape from the pattern on the surface produced by varying albedo. In the case of uniform Lambertian surfaces, the system of photometric equations becomes overconditioned and the surplus information may be used in a variety of ways: for example, to find outliers [172] or reconstruct unknown illumination directions and strengths of the light sources [173].

Traditionally, much work on photometric stereo has assumed that light comes from a single source, generally a point source or a controlled diffused source of light. However, recent aproaches, such as the work of Basri and Jacobs [9], analyse the problem when more general lighting conditions are presented. Their objective was to reconstruct shapes under everyday lighting. For some applications, such as modeling large outdoor structures, it may be impossible to completely control the lighting. Therefore, they proposed a photometric stereo method, applicable to Lambertian

objects, in which the light in each image may be an unknown and an arbitrary combination of diffuse, point and extended sources.

3.3.1.1 General Type of Surface Reflectance

For surfaces with a constant arbitrary reflectance (case **III**), the original method by Woodham [171] works sucessfully, as long as the inspected surface and the calibrating sphere are of the same material. A variation for this method, using extended light sources and specular surfaces, was presented by Ikeuchi [69]. He was the first to obtain the shape of a specular surface employing a photometric stereo technique. In his research, he used a distributed light source obtained by uneven illumination of a diffusely reflecting planar surface and three input images. His solution involved solving a set of non-linear equations and using a look-up table, made from the reflectance map, to perform the numerical inversion of the three reflectance maps. This method assumed a known object position and required accurate measurements of reflected brightness. In a later work, Ikeuchi [70] proposed a dual photometric stereo system in order to determine a depth map from the surface orientation recovered.

Recently, Iwahori et al. [72, 71] proposed a novel approach of photometric stereo for a rotational object with a non-uniform reflectance factor (case **IV**). Their approach was empirical without requiring a separate object calibration to estimate the associated reflectance maps and, therefore, eliminating the need of making any specific assumptions about the surface reflectance. Their technique exploited the ability of neural networks to perform non-parametric functional approximation instead of using the classical look-up tables.

Furthermore, a large amount of research has been devoted to the recovery of reflectance parameters (of a particular reflectance model) alongside the local gradient. Estimation of the reflectance parameters can be performed locally, provided a sufficent number of images are available, and, therefore, such algorithms are suitable for surfaces with variable arbitrary reflectance (case **IV**). Nayar et al. [120] applied photometric stereo to a so called *hybrid reflectance model* which is a linear combination

of Lambertian and specular components. They presented a method for determining the shape of hybrid surfaces without prior knowledge of the relative strengths of the Lambertian and specular components of reflection. It is also a common practice to avoid interreflections, i.e. the mutual illumination between surface facets. The same authors in [121] challenged the interreflection problem applying photometric stereo to Lambertian surfaces. They observed that erroneous shape in the presence of interreflections was a little shallower than the real shape and, therefore, could be iteratively refined.

Tagare and deFigueiredo [156] developed the theory of photometric stereo, introducing a generalisation of reflectance models which they called *m-lobe reflectance maps*. The main question they addressed was the number of images necessary to invert the image formation process for unique reconstruction of surface normals. In a later work they investigated the problem of simultaneous estimation of the shape and reflectance properties for the same class of surface reflectances [157]. An energy function was minimised with respect to the surface normal and the weights of the Lamberitan and specular components. They proved that ten light sources were needed to get a unique solution. This method was based on the assumption that the Lambertian and specular components could be preseparated. Their research was continued by Kay and Caelly [77] who investigated the problem of simultaneous estimation of surface normals and surface reflectance parameters from a practical point of view applying non-linear regression to a large number of input images. Lee and Kuo [93] introduced the concepts of parallel and cascade photometric stereo. In their work, the authors argued that the accuracy of shape from shading algorithms was related to the slope of the reflectance map defined on the gradient space. They suggested two different photometric approaches: first, parallel photometric stereo would take all the images together to produce the best estimation of the surface and second, cascade photometric stereo would take the images one after the other, and the estimated shape from the previous image, computed using a triangular element surface approximation, was used as input for the initial estimate of the next image. They used a two source photometric stereo and concluded that the best results could be obtained with orthogonal light sources.

Another photometric stereo approach for general modelling of reflectance properties was proposed by Clark [28]. His approach, called active photometric stereo, modelled the motion of the light source in infinitesimal steps. He was the first to use perspective instead of orthographic projection, thus removing the necessity of assuming that the light comes from infinity. The computation was local, non-interative, and directly solved for depth by a closed form solution. In order to measure the infinitesimal image gradient with respect to change in lighting, seven images were required to provide an acceptable discrete approximation.

Some of the methods described above solve the problem of recovering the local gradients for surfaces with variable arbitrary reflectance fully, but they are generally fairly complicated. A simple approach to deal with this problem is to ignore the reflectance model. Since many non-Lambertian surfaces exhibit near Lambertian behaviour outside their regions of specularity, it is a very attractive option to apply the linear algorithm, developed for Lambertian surfaces, to surfaces with arbitrary reflectance, and treat highlights as deviations from the Lambertian law. This was done by Coleman and Jain [30] who proposed the 4-source photometric stereo technique. This technique was based on the assumption that only one of the light sources caused specular reflection. Therefore, they used relative deviation to determine the specular source. Basically, the conventional photometric stereo technique is applied for the four triplets obtained combining the four input intensities, giving as a result of each triplet the albedo and gradient of the corresponding pixel. If the surface is Lambertian, the recovered albedos have a small variance. However, if the variance is large, in the absence of shadows, one may assume that highlights are present. The smallest recovered albedo is taken as the true albedo and the corresponding gradient as the true gradient. Further development of this technique was made by Solomon and Ikeuchi [153] who took self-shadowing regions into consideration. They extended Coleman's solution by dividing the surface into different areas, depending on the number of light sources illuminating them.

Observing all the methods described in this section, we notice another important aspect that has to be considered: the number of images used to solve the problem. Table 3.2 presents a classification of the greyscale photometric approaches taking

Table 3.2: Greyscale photometric stereo approaches classified according to the required number of images.

Approaches which use 4 or fewer images	Approaches which use more than 4 images
Woodham [171]	Tagare and deFigueiredo [156, 157]
Ikeuchi [69, 70]	Kay and Caelly [77]
Iwahori et al. [72, 71]	Lee and Kuo [93]
Nayar et al. [120]	Clark [28]
Coleman and Jain [30]	
Solomon and Ikeuchi [153]	

into account the number of images they require. We classified them into those which use a small set of images (4 or fewer images) and those which use a large set of images (more than 4 images). We considered the limit of 4 images since it is the minimum number from which it is possible to deal with shadows and highlights.

3.3.2 Colour Images

Up to the preent, not much work has been done in the field of colour photometric stereo. Information on the colour image of a Lambertian surface illuminated by a single light source is redundant since the photometric equations for individual colour bands are linearly dependent. An efficient way to use this redundancy is to perform a conventional photometric stereo method using a single colour image of a Lambertian surface under complex lighting conditions rather than three greyscale images [42, 45]. The surface should be illuminated by several light sources which are spectrally distinct and their directions do not lie in the same plane. This method is called Shape from Colour. Finlayson et al. [50] applied it to the task of face recognition. Kontsevich et al. [87] considered a similar approach.

Christensen and Shapiro [27] introduced the method of *colour photometric stereo* for surfaces with an arbitrary reflectance (case **VII**). This method uses the notion of the *shading function*, which maps surface normals under a given illumination to the colour space (much in the manner of a reflectance map in the greyscale method where

surface normals are mapped to image intensities). This method was implemented by means of look-up tables representing the inverse shading functions for a given imaging configuration which were built using a calibrating sphere. The disadvantage of this method is that the surface should either be uniformly coloured or its colours should form distinct separable clusters in the colour space which severely restricts the choice of acceptable surfaces. Therefore, this method should be classified as the case of uniform albedo (case **VII**) rather than variable (case **VIII**). Another disadvantage of this method is the need for preliminary calibration.

As explained in Chapter 2, a general reflectance function can be modelled as a sum of a matte (Lambertian) and a specular component (the latter includes both the diffuse specularity and the specular spike). It is a well-known fact that the colour of highlights resembles the colour of the illuminant. This observation leads to the *dichromatic reflection model* [84]. Incorporating spectral information into conventional shape from shading techniques gives a welcome advantage. See, for example, the work of Lee and Bajcsy [95] which used a spectral differencing algorithm to detect specularities from multiple images. In their work, the position of the object was varied rather than the direction of illumination.

Schluns and Witting [143] also used colour information to develop a colour photometric stereo technique for non-Lambertian surfaces without precalibration. They attempted to recover the surface parameters directly from the three input colour images using colour histograms. They worked out the lighting and surface chromaticities directly from the histogram and then broke down the image pixels into linear combinations of matte and specular components. This method was not tried on real surfaces. In an ideal case all image pixels lie on a plane spanned by the chromaticity vectors of the surface and the illumination colours. Such pixels are easy to break down into linear combinations according to the dichromatic reflectance theory. If, however, histograms are not planar, the break down coefficients are not reliable. In real surfaces there are always variations in colour and there are always errors, so histograms are rarely planar. Another difficulty is presented by the saturated pixels. Saturated pixels appear when the irradiance of a surface facet exceeds the capacity of image sensors and the pixel is white. Trying to break down a saturated

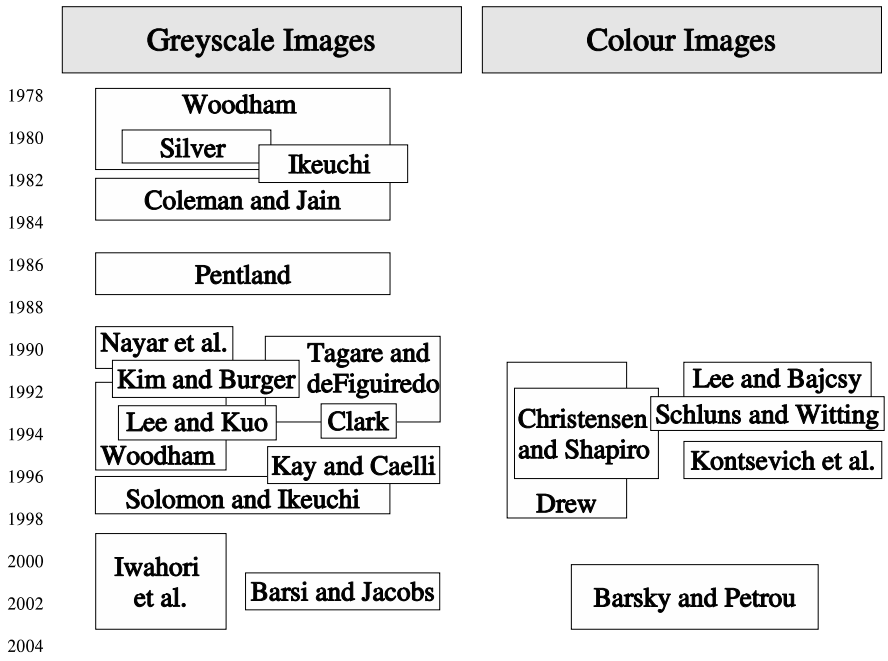


Figure 3.3: A review of photometric stereo techniques depending on the use of greyscale or colour images.

pixel leads to incorrect recovery of the components. There are other considerations which make this method impractical.

Recently, Barsky and Petrou [7, 8] proposed a method which deals with surfaces of arbitrary variable reflectance (case **VIII**). They do not assume any prior knowledge about the imaged surface so they do not use any preliminary calibration. Moreover, they proposed using colour rather than greyscale images because, for non-Lambertian surfaces, the spectral content of the images gives an additional cue for the detection of specularities. Their photometric method was based on the 4-source photometric stereo method proposed by Coleman and Jain [30]. Instead of comparing recovered albedos, they compared chromaticities of input pixels. Usually, some global technique, which often involves building colour histograms of the input images (which was done, for example, by Sanderson et al. [143]), is employed to detect highlights in colour images by spectral difference. Barsky and Petrou, on the other hand, compared colour pixels locally for each surface facet individually. This

allowed them to consider surfaces with variable arbitrary reflectance. However, for certain surfaces, the spectral difference method does not work. For these surfaces, they proposed an alternative method consisting of comparing the recovered normals with specular directions [129]. This method was less reliable than the spectral difference method but it provided reasonable results.

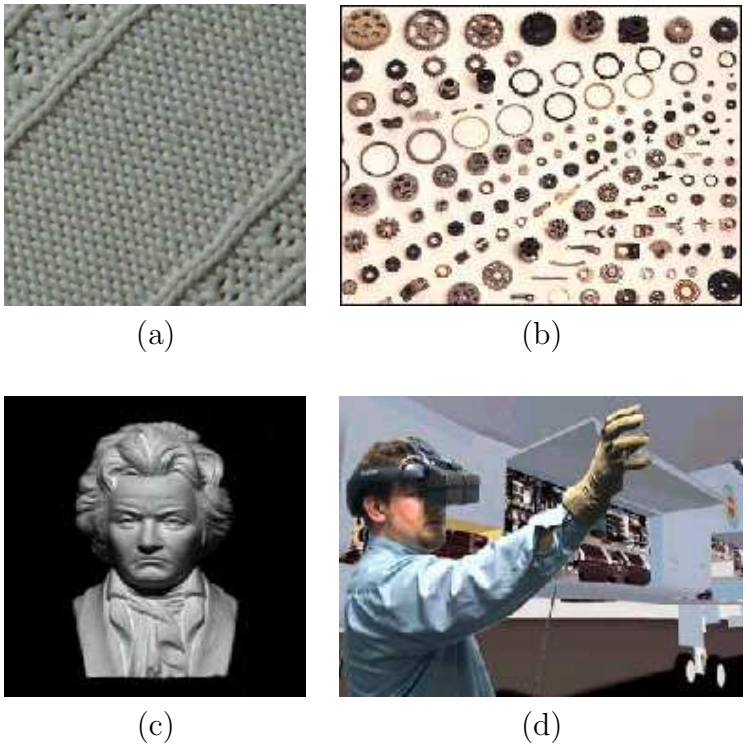
Figure 3.3 summarises the photometric stereo methods described in Sections 3.3.1 and 3.3.2, classifying them according to the use of greyscale or colour images. Note that much work has been done effectively in the field of greyscale photometric stereo, and since the mid-nineties the researchers have started to work with colour photometric stereo.

3.3.3 Applications

In the last few years many of these photometric techniques have been used in a large number of applications which can be grouped in the following categories (see Figure 3.4):

- **To perform surface texture, shape and roughness analysis.** There are many works which use a photometric technique to perform classification of 3D textures. For example, McGunnigle and Chantler [113] used the distribution of surface gradients obtained by means of photometric stereo as a cue for surface texture classification. Other works use photometric stereo to develop classification systems invariant to the rotation of the light direction. For instance, Chantler et al. [24] proposed a classification scheme focused on rotation invariant texture recognition. On the other hand, Kee et al. [78] suggested a photometric-based face recognition system. Their basic idea is to describe each face model using illumination invariant primitives: the surface normal and the albedo recovered from different images of the face captured under different lighting conditions.

- **To develop industrial vision-based inspection systems.** Smith et al. [151] applied photometric stereo to the task of unsupervised industrial inspection



(a) (b)

(c) (d)

Figure 3.4: Applications of photometric stereo techniques: (a) to perform surface texture, shape and roughness analysis, (b) to develop industrial vision inspection systems, (c) to perform 3D reconstruction and (d) to produce synthetic images indistinguishable from real-world images.

of Lambertian surfaces. They used photometric stereo to separate the surface shape from the surface albedo and then analysed the variance in the surface normals to detect defects. Moreover, the recovered surface shape was used to render a synthetic image lit in the best way to show defects. As well as detecting defects, other works proposed by Smith [152, 150] include a global quantification of the surface quality in the inspection system.

- **To reconstruct 3D information.** 3D surface reconstruction is one of the major research areas in computer vision. Several methods in the field of shape reconstruction use the gradient data obtained by photometric stereo techniques to transform it into height or depth maps, or into surface data useful for many applications [53, 67].

- **To render realistic images.** Some works use the surface information extracted by the photometric techniques with the aim of producing synthetic images visually and measurably indistinguishable from real-world images. This implies potential applications in computer graphics and virtual reality such as entertaining virtual environments and valuable visualization tools for medical applications, among others.

3.4 Use of 4-Source Colour Photometric Stereo

It is important to remember that the aim of this chapter is to choose an appropriate technique which allows us to recover surface information. In previous sections we reviewed different approaches to perform this task. Now we will point out the essential features and advantages of the photometric approach we have selected for our purpose.

Note that our goal is to use this surface information towards image prediction, i.e. use the surface information to render textured images of that surface under novel lighting conditions and imaging geometries. These predicted images will become the basis of our model-based classification system.

After analysing the different photometric approaches and taking into account the main objective of this thesis, the classification of colour textures, the 4-source colour photometric technique proposed by Barsky and Petrou [8] was chosen to recover surface information. The main reasons for this decision are the following:

- **Colour instead of greyscale images.** Since we want to perform the classification of colour textures, a virtual database of colour textured images under different geometries will be necessary. So, it is of extreme importance to recover the surface information which allows us to render such colour images. Therefore, we decided to use a colour photometric stereo technique instead of a greyscale approach.

- **Without any preliminary calibration.** As it has been explained, photometric stereo techniques allow us to recover both local reflectance and local

gradient information. There are some techniques which need an initial calibration process for each object. However, the chosen 4-source colour photometric technique permits us to recover both a reflectance map (called *colour map* in the case of colour photometric stereo) and shape information without any preliminary calibration.

- **Lambertian/Specular surfaces.** The 4-source colour photometric stereo technique does not make strong assumptions about the underlying surface structure as other techniques do. Therefore, it can be applied to a general type of surface with a variable albedo.

- **Small set of images.** 4 images of the same surface captured under different directions of illumination are needed to apply the 4-source colour photometric stereo. This is the minimum number of images necessary to deal with surfaces with general reflectance properties. Other techniques such as those proposed by Tagare and deFigueiredo [156, 157], Kay and Caelly [77], and Clark [28] require more images to recover the surface information correctly. Moreover, the selected technique does not present any restriction concerning the direction of the light sources used in the photometric stereo images as other techniques do.

- **Simplicity.** Another important feature of the selected technique is its simplicity. This photometric method is based on the 4-source photometric stereo method proposed by Coleman and Jain [30]. They worked on the fact that many non-Lambertian surfaces exhibit near Lambertian behaviour outside their regions of specularity. Therefore, it is a very attractive option to apply the linear algorithm developed for Lambertian surfaces to surfaces with arbitrary reflectance and treat the specular reflections (highlights) as deviations from the Lambertian law. Note that this is a much simpler solution compared with other algorithms which need to solve a set of non-linear equations.

In the next section we describe the basic terms and notations of the 4-source colour photometric stereo we used, detailing the complete process from which we

obtained the surface description.

3.5 Photometric Stereo: Terms and Notations

Two assumptions are usually made in photometric stereo. The surface is generally flat, parallel to the image plane of the camera, and is lit by a single light source. Moreover, the camera and light are far away from the surface so the viewing direction and light direction are the same for every point on the surface. Usually, most authors choose the coordinate system so that the image plane coincides with the xy plane and the z-axis coincides with the viewing direction. Then the surface can be described by a $2D$ height funcion $z = S(x, y)$ and we can define the gradient components for every point on the surface as

$$p(x, y) = \frac{\partial S(x, y)}{\partial x} \qquad q(x, y) = \frac{\partial S(x, y)}{\partial y} \tag{3.1}$$

and the normal unit vector $\mathbf{N}$ as

$$\mathbf{N} = \frac{1}{\sqrt{p^2 + q^2 + 1}}(p, q, -1)^T \tag{3.2}$$

It is often assumed that all surface facets are visible, i.e. at every point the angle between the local normal $\mathbf{N}$ and the negative direction of the z-axis is less than $\pi/2$.

The surface photometric function for a single source can be defined in terms of three angles called the incident angle (i), the emittance angle (e) and the phase angle (g) as shown in Figure 3.5. These angles quantify the relationship between the light source vector $\mathbf{L}$, the viewer vector $\mathbf{R}$, and the surface normal vector $\mathbf{N}$. All these angles are used to obtain the reflectivity function which measures the reflectance of the surface. In Lambertian surfaces, the reflectivity function is proportional only to the cosine of the incident angle. Under uniform light, these surfaces have the property of looking equally bright from all directions. In contrast, specular surfaces, which have a mirror-like behaviour, appear when the angle between the viewer and the light source is bisected by the surface normal.

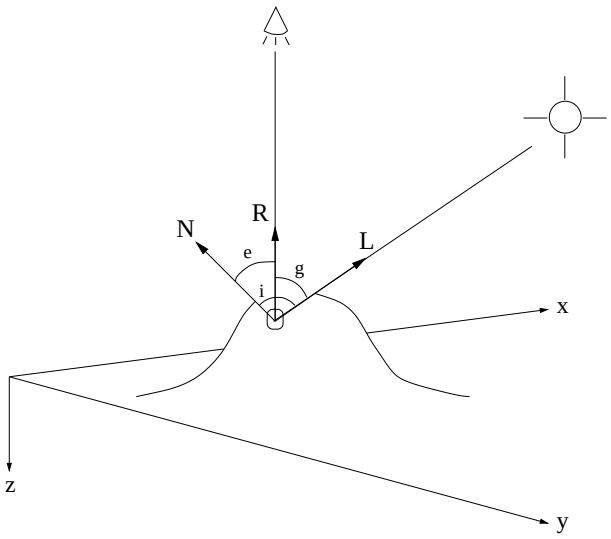


Figure 3.5: Definition of the important vectors and reflectance angles: **R**, viewer vector; **L**, illuminant vector; **N**, normal vector; i, incident angle; e, emittance angle; g, phase angle.

We define a unit illumination vector as $\mathbf{L} = (L_x, L_y, L_z)^T$ which points from the surface towards the light source. If we have several light sources, we denote them by using a superscript $\mathbf{L}^1, \ldots, \mathbf{L}^l$ where l is the number of light sources. A pixel, obtained by a camera with K sensors in each cell, is a vector in a K-dimensional colour space. A colour pixel, obtained by the jth lighting, is denoted by $\mathbf{I}^j = (I_1^j, \ldots, I_K^j)$ where superscripts refer to the lighting and subscripts to the colour component. A greyscale pixel, obtained under the jth lighting, is denoted by I_0^j.

In all the experimental trials presented in this thesis, a colour photometric stereo set is available for each physical texture. Each photometric set is composed of 4 images under the following conditions: (i) the surface is lit by 4 distant point light sources, (ii) the lights are placed in a cross-like manner, all four having the same elevation angle (see Figure 3.6). Thus, the elevation angle (we shall refer to it as θ) is a single parameter which describes a particular lighting configuration. The positions of the lights have been chosen so that the light vectors are:

$$\begin{array}{ll} \mathbf{L}^1 = (0, cos\theta, -sin\theta)^T & \mathbf{L}^2 = (cos\theta, 0, -sin\theta)^T \\ \mathbf{L}^3 = (0, -cos\theta, -sin\theta)^T & \mathbf{L}^4 = (-cos\theta, 0, -sin\theta)^T \end{array} \tag{3.3}$$

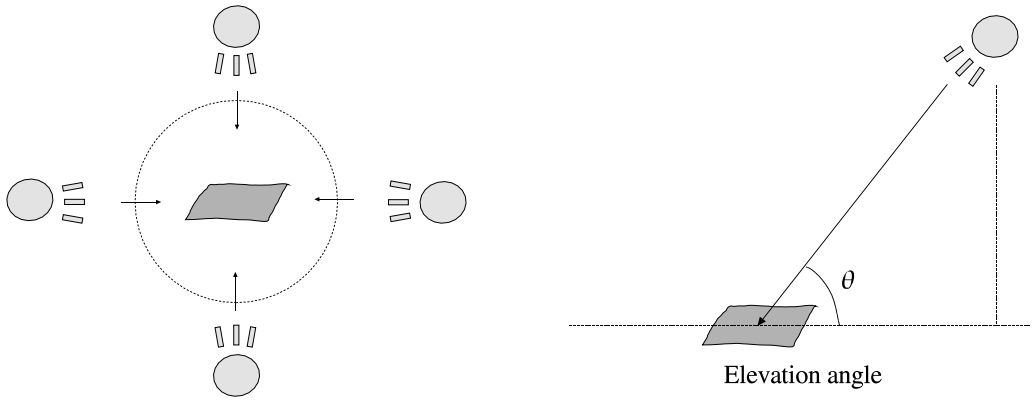


Figure 3.6: Lighting configuration. (a) The lights are placed in a cross-like manner (camera view). (b) All four lights have the same elevation angle θ.

In what follows, we detail the colour photometric stereo proposed by Barsky and Petrou [8]. We first explain the greyscale photometric stereo and then the extension to colour images.

3.5.1 Greyscale Photometric Stereo

Let us consider a Lambertian surface patch with albedo ρ and normal $\mathbf{N}$, lit with only three light sources with directions $\mathbf{L}^1$, $\mathbf{L}^2$, and $\mathbf{L}^3$. In this case, the intensities of the obtained (greyscale) pixels can be expressed as

$$I_0^k = \rho(\mathbf{L}^k \cdot \mathbf{N}) \tag{3.4}$$

where $k = 1, 2, 3$, and $(\cdot)$ represents the scalar product of two vectors.

The pixel intensities can be stacked to obtain the pixel intensity vector $\mathbf{I_0} = (I_0^1, I_0^2, I_0^3)^T$. The light vectors can also be stacked row-wise to form the *illumination matrix* $[L] = (\mathbf{L}^1, \mathbf{L}^2, \mathbf{L}^3)^T$, (square brackets are used to denote matrices). Then, equation 3.4 could be rewritten in matrix form:

$$\mathbf{I_0} = \rho[L]\mathbf{N} \tag{3.5}$$

If the three light directions $\mathbf{L}^k$ do not lie on the same plane, matrix $[L]$ is non-singular and it can be inverted, giving:

$$[L]^{-1}\mathbf{I_0} = \rho\mathbf{N} \tag{3.6}$$

Since $\mathbf{N}$ has unit length, both the normal (as the direction of the obtained vector) and albedo (as its length) can be recovered.

If we have more than three input images, it is possible to add robustness to the scheme as was done by Coleman and Jain [30]. First of all, for every surface patch we can make up a set of all possible intensity triplets, each triplet having an illumination matrix associated with it. Applying the corresponding inverse illumination matrix to each intensity triplet, we obtain a set of recovered vectors. All these vectors are used to estimate the normal vector of the surface patch. More specifically, these vectors are averaged to produce the final one. The albedo is obtained computing the length of the final vector. Figure 3.7 shows an illustrative example in which 4 images (4 pixel intensities) are available. Note that in this case, 4 intensity triplets are used to recover the normal vector and the albedo.

3.5.2 4-Source Colour Photometric Stereo

Let us assume that we have 3 colour images of the same Lambertian surface taken under 3 distinct light sources whose directions are described by the illumination matrix $[L]$. The intensity triplets $\mathbf{I^1}$, $\mathbf{I^2}$, and $\mathbf{I^3}$, produced by a surface patch for each of the 3 lighting arrangements, are described by:

$$\mathbf{I^k} = (I_1^k, I_2^k, I_3^k)^T = (\mathbf{L^k} \cdot \mathbf{N})\mathbf{C} \tag{3.7}$$

where vector $\mathbf{C}$ is the colour of the surface patch $\mathbf{C} = (C_r, C_g, C_b)$.

Let us denote the scalar product $(\mathbf{L^k} \cdot \mathbf{N})$ by s, so we can form a *shading vector* $\mathbf{S} = (s^1, s^2, s^3)^T$. Note that $\mathbf{S} = [L]\mathbf{N}$.

If we stack the pixel vectors row-wise to obtain the *intensity matrix* $[I] = (\mathbf{I}^1, \mathbf{I}^2, \mathbf{I}^3)^T$, we can write:

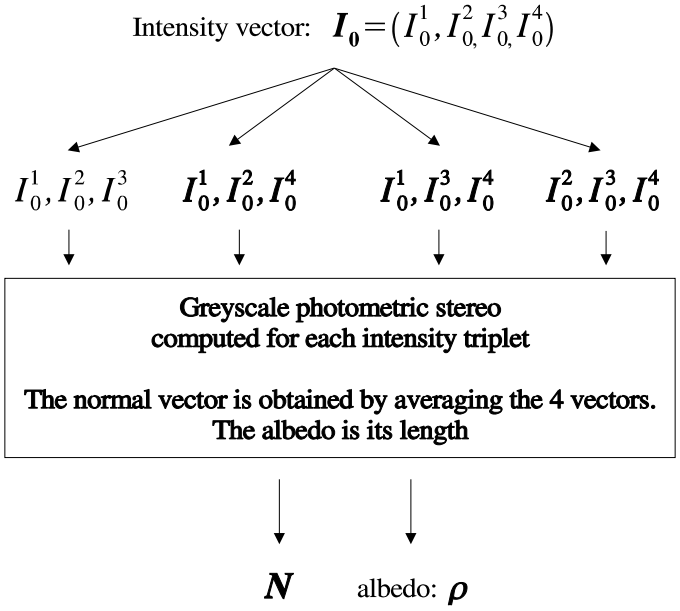


Figure 3.7: Example of how photometric stereo can be computed if more than 3 images are available. By applying photometric stereo to each intensity triplet, a set of recovered vectors is obtained.

$$[I] = \begin{bmatrix} I_1^1 & I_2^1 & I_3^1 \\ I_1^2 & I_2^2 & I_3^2 \\ I_1^3 & I_2^3 & I_3^3 \end{bmatrix} = \begin{bmatrix} s^1C_r & s^1C_g & s^1C_b \\ s^2C_r & s^2C_g & s^2C_b \\ s^3C_r & s^3C_g & s^3C_b \end{bmatrix} = \mathbf{S} \otimes \mathbf{C}^T \tag{3.8}$$

where $\otimes$ represents the Kronecker product.

Note that while the kth row of matrix $[I]$ is the kth input pixel $\mathbf{I^k}$, its lth column is the intensity vector $\mathbf{I_l}$ for the lth colour band. Equation 3.8 describes the intensity matrix in the ideal noiseless case. However, in real data there is always a certain degree of noise, and the "observed" intensity matrix differs from the "ideal" matrix. Barsky and Petrou proposed to find such estimates of $\mathbf{N}$ and $\mathbf{C}$, minimising the error between both matrices. This can be done by applying the Least Square Error technique which results in the desired estimates for colour and shading vectors being the principal eigenvectors of matrices $[I]^T[I]$ and $[I][I]^T$ respectively.

Intuitively, this method can be interpreted as follows. Given a Lambertian surface patch, the three colour pixels corresponding to it are collinear in the RGB

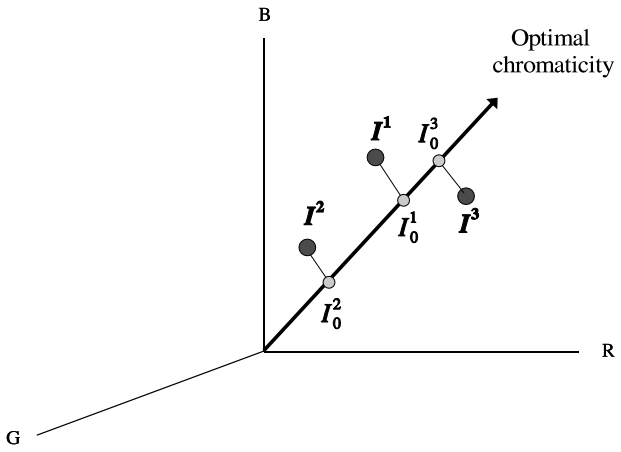


Figure 3.8: Intuitive interpretation of the colour photometric stereo. By computing the chromaticity of each surface patch and projecting all input pixels on the principal colour line, the problem is reduced to the greyscale case.

space and differ only by a scalar factor (the shading of the patch under a particular illumination). Introduced errors may disturb the collinearity, therefore, Principal Component Analysis is used to find their principal direction. The principal direction gives the optimal chromaticity of the surface colour in the Least Square Error sense. By determining the chromaticity of the surface patch and projecting all colour input pixels on the principal colour line, the problem is reduced to the greyscale case where the projections play the role of greyscale intensities. Figure 3.8 illustrates this methodolgy. By applying the above described greyscale photometric algorithm to this intensity vector, we get the optimal estimation of the gradient and the albedo. Note that the colour of the surface is the vector in colour space oriented along the principal colour and has a length equal to the calculated albedo.

This method can be easily extended to more than 3 input images, let us say to M images. It is possible to estimate surface chromaticity using all M colour pixels by finding the principal eigenvector of the corresponding colour correlation matrix. Using this chromaticity, M "intensities" can be produced by projecting all the pixels on the principal colour line. Afterward, these intensities are used as input for the appropriate greyscale method to recover the normal and albedo.

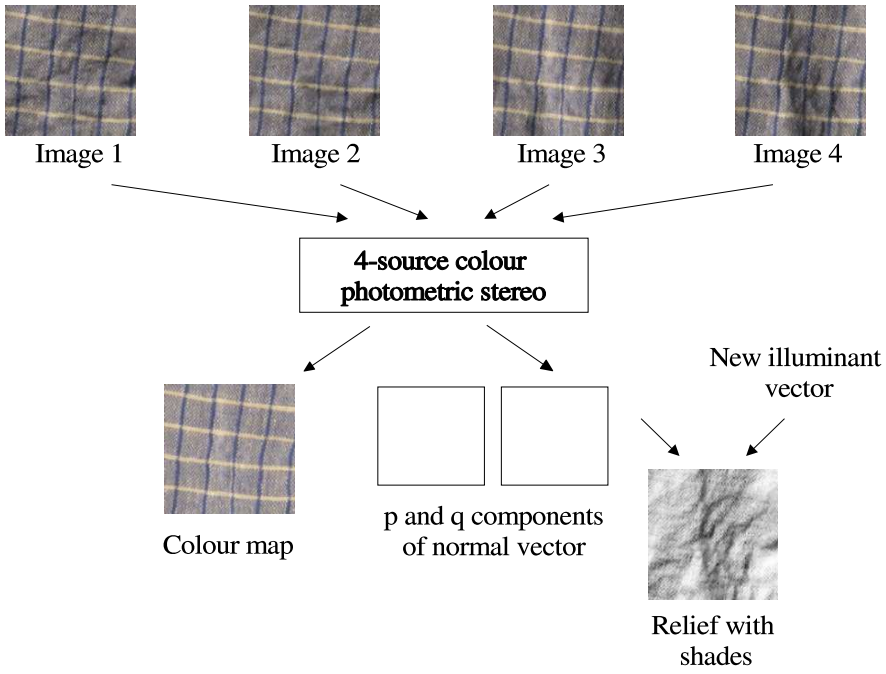


Figure 3.9: 4-source colour photometric scheme. As a result to apply the photometric technique, the colour map and the surface shape information are recovered.

Figure 3.9 summarises the structure of this 4-source colour photometric stereo based virtual image generation approach. Given a photometric set composed of 4 images, this photometric technique allows us to recover the colour information as well as the gradient vector for each point on the surface. Note that for a given particular direction of light and by using the gradient information, the corresponding surface shading can be generated. Note also that the colour information is expressed in terms of the colour map. All of this information is indispensable to the process of rendering new images under different lighting conditions.

3.5.3 The Problem of Highlights and Shadows

Now, since the colour photometric stereo technique has been described, we will analyse what happens with this technique if it is applied to images affected by highlights and shadows.

Basically, if the method described in 3.5.2 is applied to a triplet which has a

highlight or a shadow, the recovery of colour and gradient information is affected. For example, the recovered colour will appear lighter than the true colour if a highlight is present, while the change of colour due to a shadow will depend on the lighting set-up. Moreover, the recovered normal will lean more towards the light source which produced the highlight or away from it accordingly. In what follows, we briefly review some approaches which have studied these problems.

3.5.3.1 Related Work

Coleman and Jain [30] proposed a method using 4 images of the same surface to detect highlights in the absence of shadows. Their basic idea was to compare the albedos recovered from all four possible triplets of pixels under the assumption that the specularity regions do not intersect. It is this redundancy of using four triplets of pixels which allowed them to tag and remove the specular source. To illustrate how this is accomplished, assume that we are given four measured intensity values at a point (x, y) on a surface, one intensity from each image under its corresponding light source. From this quadruple of pixels, we form 4 triplets. We reconstruct the local normal and colour for each of the triplets, thus having 4 normals and 4 albedos. If none of the intensities has a specular component, that is to say the quadruple does not contain a highlighted pixel, the deviation of the recovered values from the four different triplets is insignificant and they will appear as shown in Figure 3.10.(a). However, suppose the intensity value from one image contains a specular component elevating its value. The resulting four normals will be similar to those illustrated in Figure 3.10.(b). It is easy to see that there is a greater deviation in both direction and magnitude of the vectors. Coleman and Jain argued that if the albedos differ significantly, it should be due to a highlight. The three largest albedos must be affected by the highlight, therefore the triplet producing the smallest albedo contains only the Lambertian component and is used for recovery.

Nevertheless, many natural surfaces produce cast and self-shadows when lit by directional light. A variation of the above method, proposed by Solomon and Ikeuchi [153], takes self-shadows into consideration. Solomon and Ikeuchi considered a unit hemisphere of surface normals, simultaneously lit by all four lights at once.

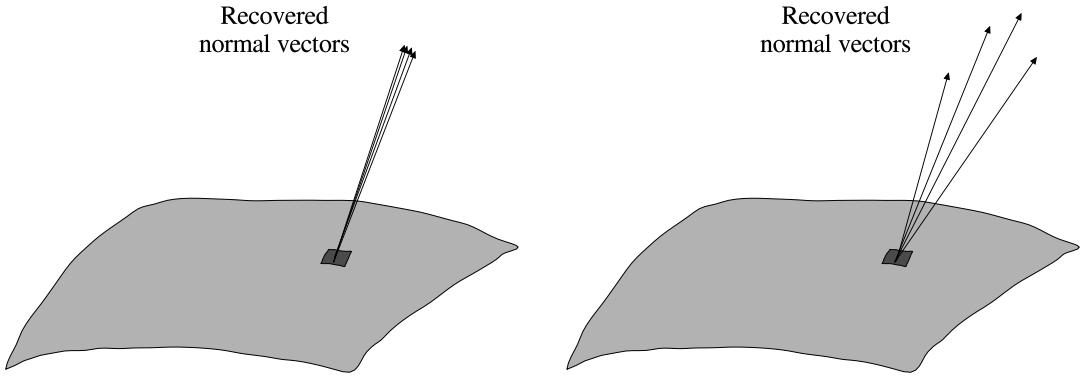


Figure 3.10: Example of computed normals affected by a highlight [30]. (a) Computed normals at a single point on the surface when no specular component is present. The deviation among normals is small. (b) Computed normals at a single point with a source exhibiting specularity. The intensity value from this source is elevated causing a high deviation among these normals.

The hemisphere was divided into regions: those lit by all four lights, by three lights, and by only two lights. Different strategies were suggested for detecting specularities and local surface recovery for each of the regions. This method has several serious shortfalls. First of all, it excludes cast shadows and that can imply an incorrect gradient reconstruction. Moreover, there is no indication of how to detect shadows. In real images, shadows are rarely perfectly black. They may be brightened by secondary reflection, backlight, etc. There is always a range of shadow values even for uniformly coloured surfaces and to detect a shadow we must use some sort of a thresholding algorithm. For arbitrarily textured surfaces the uncertaintly in shadow detection is even greater: shadow values in a brightly coloured area are higher than shadow values in darker areas. See, for example, the variety of shadows presented in the images in Figure 3.11.

As we have seen above, the method proposed by Coleman and Jain assumed that there are no shadows in the input images. If, however, there are shadows, either cast or self-shadows, highlight detection becomes more difficult, as the same spread in the values of the recovered normals and albedos can be caused by shadows as well as highlights.

(a) (b) (c)

Figure 3.11: Three image samples which present a great variety of shadows.

Recently, Petrou and Barsky [129] proposed a new method aimed at reducing the distortion caused by misinterpretation of highlights and shadows. First, they proposed a technique for detecting the "problematic" quadruples (quadruples of pixels containing highlights or shadows) using only pixel values without any information about the surface or the lighting configuration. Their technique works for greyscale and colour images without the need of recovering first the normals and albedos, therefore it is quick. After detecting a problematic quadruple, they decided whether it contained a highlighted pixel or a shadow, using colour information. Finally, having ruled out highlights, they concluded that the remaining problematic quadruples contained shadows. Summarising, for a highlighted quadruple the true normal vector and colour are recovered using the three darkest values and, for a shadowed quadruple the restoration is done using the three brightest values.

3.5.3.2 Methodology for the 4-Source Colour Photometric Stereo

As explained in previous sections, in order to recover the local normal and colour of a surface patch, 4 colour images of the same surface taken from the same viewpoint under light with different directions L^k, $k = 1, 2, 3, 4$, are used. In 3D space, any 4 vectors are linearly dependent, so the 4 light vectors $\mathbf{L}^k$ are also linearly dependent and there is a linear equation expressing the relationship between them:

$$a_1\mathbf{L}^1 + a_2\mathbf{L}^2 + a_3\mathbf{L}^3 + a_4\mathbf{L}^4 = 0 \tag{3.9}$$

Multiplying this equation by the albedo ρ and the local surface normal $\mathbf{N}$, the following equation can be obtained:

$$a_1\rho(\mathbf{L}^1 \cdot \mathbf{N}) + a_2\rho(\mathbf{L}^2 \cdot \mathbf{N}) + a_3\rho(\mathbf{L}^3 \cdot \mathbf{N}) + a_4\rho(\mathbf{L}^4 \cdot \mathbf{N}) = 0 \tag{3.10}$$

Note that this equation is equivalent to:

$$a_1 I^1 + a_2 I^2 + a_3 I^3 + a_4 I^4 = 0 \tag{3.11}$$

In other words, linear dependence of the light vectors leads to the same linear equation for the corresponding pixel intensities if the Lambertian assumption holds. Equation 3.11 can be rewritten in vector form:

$$(\mathbf{A} \cdot \mathbf{I}) = 0 \tag{3.12}$$

where $\mathbf{A} = (a_1, a_2, a_3, a_4)^T$. This means that any non-shadowed Lambertian quadruple of pixel intensities is perpendicular to $\mathbf{A}$, i.e. for a specific lighting configuration, all non-shadowed Lambertian quadruples form a hyperplane in the 4-dimensional intensity space, no matter what albedo and normal the corresponding surface facets have. Therefore, all non-shadowed Lambertian quadruples of intensities in the input images should satisfy Equation 3.12. The parameters of this equation can be estimated even without knowing the lighting configuration by means of the Least Squares technique. Barsky and Petrou proposed that by thresholding the square error with which Equation 3.12 is satisfied, it is possible to separate the problematic pixel quadruples from non-shadowed Lambertian ones. In the case of colour images, they proposed comparing the values of projections of the input colour pixels along the principal colour line rather than using the actual values in each colour band separately. This is because they concluded that, for different surfaces, they could use images with weaker and stronger colour bands, preventing cases in which the method fails if each colour band is analysed separately.

Their method allows us to rule out the majority of purely non-shadowed Lambertian quadruples. It is fast and it does not require any information about the

orientation or the colour of a surface facet. The colour photometric stereo method can be applied to these quadruples without any modification, much in the manner done by Coleman and Jain [30].

Having detected a problematic quadruple, the next step is to decide whether it is a highlight or not. Barsky and Petrou proposed doing that by using colour information. The colour of a highlighted pixel is the combination of "matte" surface colour, and the colour of the light. The more mirror-like the reflectivity of the surface material, the closer the highlight colour is to the light colour. Thus, for white light we can compare the chromaticity of the brightest pixel with the chromaticity of darker pixels (which is the chromaticity of the "matte" surface colour) and, if the difference exceeds a certain threshold, the pixel can be labelled as a highlight.

Nevertheless, this method of detecting highlights fails if the chromaticity of the underlying surface colour is close to the chromaticity of the incident light. That is to say, the variation in pixel colour due to highlighting becomes indistinguishable from the variation due to the imaging process. In this case, they proposed the following technique; the brightest value is disregarded and the normal and surface colour are reconstructed using the three darkest values. If the recovered normal is close to the specular direction of the corresponding imaging configuration and the brightest value is higher than that predicted by the recovered normal and albedo, that value is assumed to be due to a highlight.

In the rest of the problematic quadruples, the darkest value is considered to be due to a shadow and the colour and normal can be recovered using the three brightest values.

3.5.3.3 Example: Highlights and Shadows Correction

In what follows, we present an example to illustrate the performance of the 4-source colour photometric stereo in the presence of highlights and shadows.

We have selected a surface which has highlights and shadows as well as a number of other imperfections, such as interreflection and over-saturation of highlights.

Figure 3.12: Input photometric set for the "tomatoes" surface.

Figure 3.12 shows the four images of the photometric set. The first row in Figure 3.13 shows the output of the 4-source colour photometric stereo without doing highlights and shadows correction. The first image shows the recovered colour map, while the second image shows a shading of the surface lit from a certain direction. Note that both colour and normals are recovered incorrectly in the presence of shadows and highlights producing false highlights and shadows. The third image in this row shows a final reconstruction of the surface under a certain direction of light.

The second row in Figure 3.13 shows the detected "problematic" quadruples. The white areas in the images correspond to both highlights and shadows in all 4 images. These images are obtained by thresholding the error in the linear equation 3.12. The next two images in this row present the highlights and shadows detected by using the methodology described in the previous section. Note that different colours are used to indicate highlights produced by different lights.

When the non-shadowed Lambertian points are detected and labelled, the recovering of the surface normals and albedos is done by using the other 3 values. The

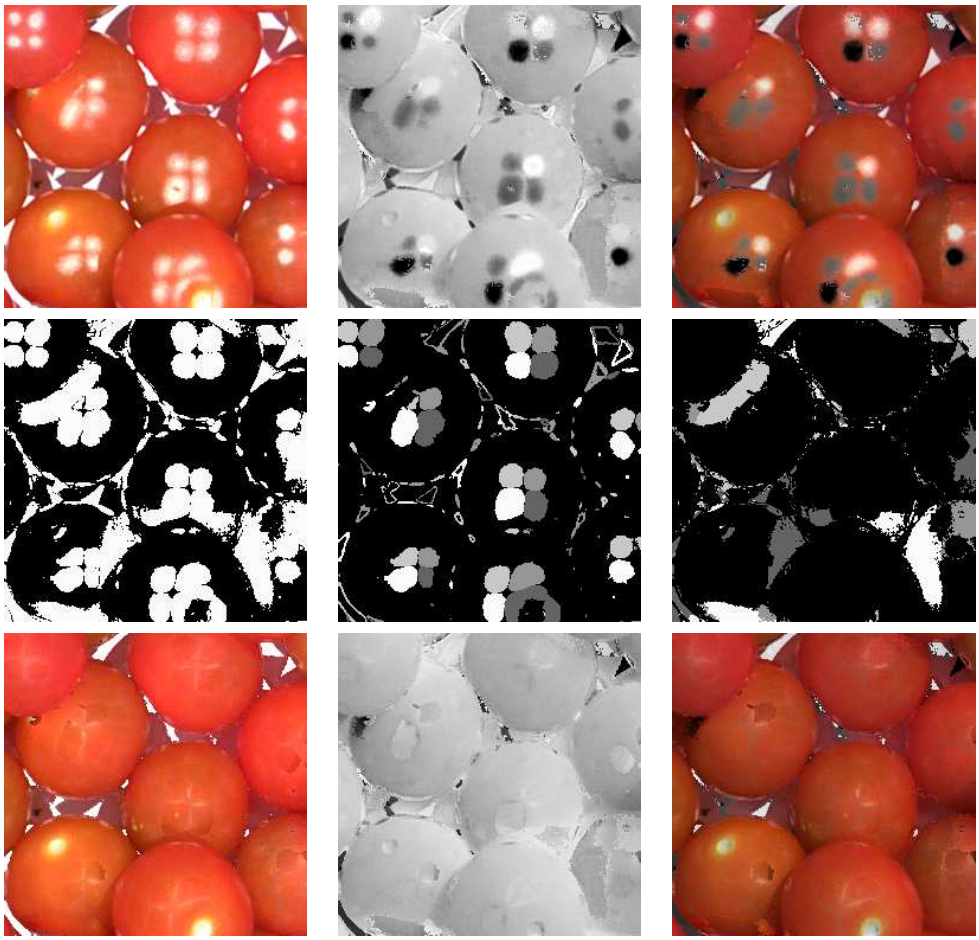


Figure 3.13: Example of reconstruction with highlights and shadows correction (taken from [129]). The first row shows the colour map, a shading of the surface and a surface reconstruction without highlights and shadows correction. The second row shows the "problematic" quadruples, the highlights and shadows detected on the surface. The third row shows the colour map, a shading of the surface and a surface reconstruction with highlights and shadows correction.

third row in Figure 3.13 shows again the colour map, a shading of the surface and also a final reconstruction under a particular light. Note that a number of pronounced highlights and some shadows have been successfully detected and corrected.

However, it is important to observe that from these images another problem arises: when multiple imperfections are present in the same quadruple (for example, shadows from two or more light sources), the method does not have enough information for adequate reconstruction.

3.5.4 Texture Rendering

Rendering is the process of generating an image of a 3D scene using colour and shading to make the object or objects appear solid and three-dimensional. Basically, the rendering processes can be classified into two general approaches: *model-based* and *image-based*. The difference between them is easy. In traditional model-based rendering, a geometric model of a scene together with surface reflectance properties and lighting parameters are used to generate an image of the scene from a desired light point. While in image-based rendering, a set of images taken from known viewpoints are directly processed to create new images.

It must be clarified that we understand the rendering process in a simpler manner. Precisely, we refer to the rendering process as a relighting algorithm which uses as inputs a surface model and a desired light direction and returns the corresponding output image. Note that this rendering method is clearly model-based as a surface model is needed to generate new images.

In a recent work [41], Dong and Chantler discussed three different ways in which the texture may be rendered given the photometric information and lighting conditions. Figure 3.14 illustrates these approaches. The *image-based relighting* method uses a simple linear combination of images which assumes a Lambertian reflectance law in order to relight under a novel direction of light. In contrast, the *gradient-based relighting* uses the surface gradients and the reflectance properties, while the *height-based relighting* uses height-maps instead of using surface gradients. Note that the last two are model-based approaches.

A gradient-based approach will be used in this thesis to render images under novel lighting conditions. Dong and Chantler [41] argued in their work that the three approaches produce very similar results. Therefore, it seems more suitable to our purposes to use a gradient-based approach since both information (gradient vectors and colour) are directly provided by photometric stereo.

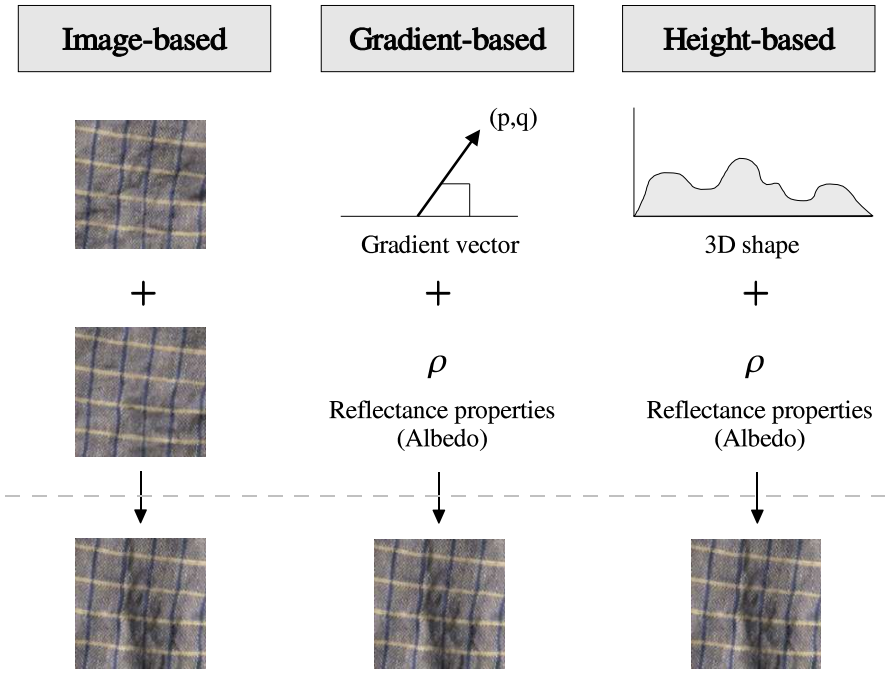


Figure 3.14: Different ways of relighting surface textures.

3.5.4.1 Gradient-based Rendering

We have seen that a surface can be defined as a set of facets which are normal to the gradient vectors of the surface. Using these facets and given a light vector, rendering of the surface shape can be done by showing the *shading* of a facet calculated by taking the dot product of the light vector with the gradient vector.

Moreover, by using the surface shape information and applying the colour map, a rendering of the surface texture can be obtained by multiplying, at every pixel, its colour map with its shading factor. Hence, the image texture is finally generated.

Figure 3.15 summarises this image texture rendering process. First, 4-source colour photometric stereo is applied to recover the surface and colour information. Afterwards, given a light vector and using the recovered gradient vectors, a surface shape rendering is made. Finally, by using the colour map information and the surface shape rendering, the image texture rendereding is obtained. Note that when surface texture information is available, it is possible to render output image textures

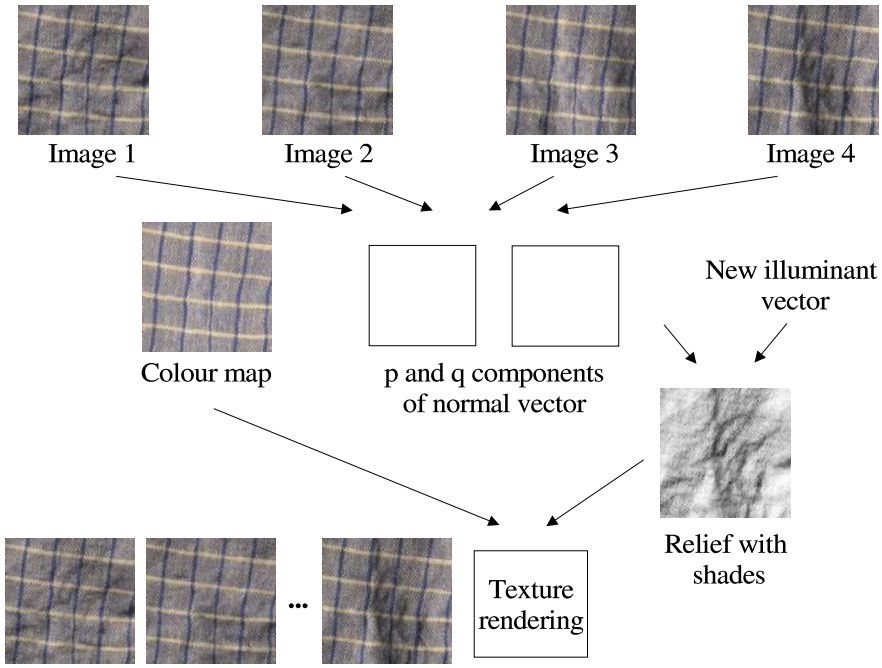


Figure 3.15: Image rendering process.

for every input light vector.

3.6 3D Surface Reconstruction

3D surface reconstruction is one of the major research areas in computer vision. Nowadays, there are many methods which allow us to achieve a 3D reconstruction. For instance, those based on stereo vision [48] which use two images of the same scene captured under the same lighting conditions. Moreover, those based on coded structured light [141, 124] which use a single camera and assume that one image of the scene is lit by multiple and spectrally distinct lights. And among others, those based on photometric stereo using a single camera to obtain several images of the scene captured under different lighting conditions; i.e. lights placed at different positions with respect to the scene and the camera.

The aim of this section is not to describe all of these reconstruction methods. We will be more specific and only focus on those techniques which allow us to obtain a 3D

reconstruction by using the surface gradient information extracted from photometric stereo.

3.6.1 Integration Techniques

Several methods in the field of shape reconstruction lead to gradient data which still have to be transformed into height or depth maps or into surface data for many applications. Thus, reconstruction accuracy also depends upon the performance of these methods which enable the transformation of gradient vectors into height maps.

Essentially, there are two types of known approaches in the literature [83], *local integration along paths* [30, 64, 174] and *global integration techniques* [68, 53, 67]. Figure 3.16 shows the basic idea of both integration approaches.

Path integration techniques are based on local calculations of height increments by curve integrals. The main idea is to use the gradient vectors which give the change in surface height with a small step in either the x or the y direction, to recover a height map of the surface by summing these changes in height along some path. These techniques differ in the way they specify an integration path and in the special procedures (local neighborhood) used to compute the local approximations of height increments. For example, Coleman and Jain [30] start in the middle of the gradient field. Their initial path forms a cross in the array. The integration is then performed in all four quadrants in the column direction. For two points in the sequence the averaged surface normal is calculated defining a surface tangent from the previous point to the next location. Their technique is known as the *two-point* method in the computer vision literature. Healey and Jain [64] have extended this technique to an *eight-point* method. Wu and Li [174] also suggested paths parellel to the x-axis or the y-axis, but averaging gradient values for obtaining increments in height. Moreover, a technique using four different scans through the gradient field starting at the four corners is presented by Rodehorst [140]. This is based on the assumption that the same initial height value is valid in all four corners of the gradient data array. Therefore, by averaging the results of all four scans the final height is obtained.

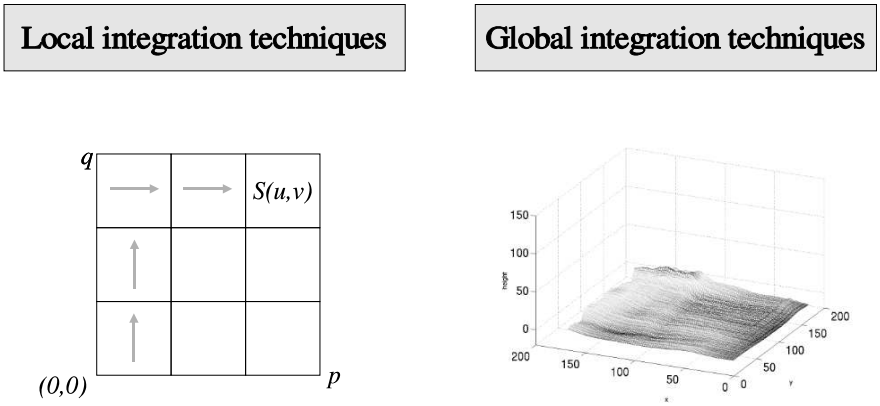


Figure 3.16: Classification of integration techniques: local and global integration techniques.

Summarising, these local integration techniques are easy to implement and very efficient in computing speed. However, the locality of calculations causes a high dependency on data accuracy and the propagation of height increments along paths also means the propagation of errors.

In global techniques, surface integration is treated as an optimisation problem. In other words, surface integration can be considered to be a variational problem where a certain function has to be minimised [68, 67]. Usually, a certain representation of the unknown surface: e.g. in terms of Fourier basis functions and the integrability condition are used to constrain the global optimisation process [53]. Two different global techniques are well-known in the literature: the method suggested by Horn and Brooks [68] and the method proposed by Frankot and Chellappa [53]. The first is directed on minimising the error function

$$f(\tilde{p},\tilde{q}) = \int\int |p(x,y) - \tilde{p}(x,y)|^2 + |q(x,y) - \tilde{q}(x,y)|^2 dxdy \qquad (3.13)$$

where p, q denote the given gradient field components and

$$\tilde{p}(x,y) = \frac{\partial \tilde{S}(x,y)}{\partial x}, \qquad \tilde{q}(x,y) = \frac{\partial \tilde{S}(x,y)}{\partial y} \tag{3.14}$$

denote the unknown (ideal) gradient field components which have to be reconstructed. A surface is calculated by the minimisation of f which ensures a maximum consistency of the reconstructed surface with the given data array. The difficulty with this method consists in selecting proper initial values at the boundary of the integration process. To solve this, Horn [67] suggested some boundary conditions in his work.

On the other hand, Frankot and Chellappa [53] assume that the unknown surface function S satisfies the integrability condition, which is to say the surface is continuous at any point. A surface S satisfies the integrability condition when the following equation is valid at all surface points.

$$\frac{\partial^2 S(x,y)}{\partial x \partial y} = \frac{\partial^2 S(x,y)}{\partial y \partial x} \tag{3.15}$$

Furthermore, Frankot and Chellappa assumed a Fourier coefficient representation

$$S(x,y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} S^{(f)}(u,v) e^{-j(ux+vy)} du dv \tag{3.16}$$

of this function, where

$$S^{(f)}(x,y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} S(x,y) e^{j(ux+vy)} dx dy \tag{3.17}$$

denotes the Fourier coefficients of S. Based on these assumptions, they could prove a theorem allowing the reconstruction of function S in the Fourier space. Afterwards, an inverse Fourier transform leads to the desired surface height data.

In general, it is expected that global techniques such as the above described should be more robust to noise in contrast to local path integration because the surface gradient has a global impact on the solution process. Table 3.3 briefly

Table 3.3: Integration techniques: advantages and disadvantages.

Integration technique	Advantages	Disadvantages
Local integration	Easy to implement. Efficient in computing speed.	High dependency on data accuracy. Propagation of errors.
Global integration	Robust against noise.	Computation time.

summarises the main advantages and disadvantages of local and global integration techniques.

As it will be seen in the next chapter, a local integration technique will allow us to recover the surface height information. This information will be used in order to predict the surface shape information when it is viewed from a longer distance.

3.7 Summary

We have seen in this chapter that there are many techniques which allow us to obtain surface texture information. Our objective has focused on reversing the process of surface recovery towards image prediction. In other words, to use a surface texture model to render textured images of that surface under novel lighting conditions and imaging geometries.

Several methods which attempt to determine surface characteristics by measuring parameters calculated from images of the lit object have been presented. Moreover, the photometric stereo approaches have been reviewed, dividing them into those which use greyscale and colour images. Among all these photometric stereo techniques, the **4-source colour photometric stereo** proposed by Barsky and Petrou [8] was chosen as the best way to recover surface texture information. This technique allows us to recover the **gradient vectors** and the **reflectance properties** for every point on the surface.

Main advantages of this photometric approach are the following:

- It uses colour instead of greyscale images.
- It can be applied to a general type of surface (Lambertian or specular).
- It requires a small set of images.
- There is no need to do any preliminary calibration.

We have also seen that this technique can deal with reconstruction problems produced by the presence of highlights and shadows in the original images. This problem has been analysed and reviewed. Furthermore, the solution adopted by the 4-source colour photometric stereo has been described.

The **rendering process** was introduced, classifying such processes into two broad approaches: model-based and image-based. A gradient-based approach, which is clearly a model-based approach since a surface model (obtained by means of 4-source colour photometric stereo) is needed to generate new images, was chosen to perform our image texture rendering.

We have also focused on those techniques which allow us to obtain a **3D reconstruction** by using the surface gradient information extracted from photometric stereo. Essentially, two broad approaches were found in the literature, local integration along paths and global integration techniques. Path integration techniques use the gradient vectors which give the change in surface height in very small steps in either the x or the y direction to recover a height map of the surface by summing these changes in height along a chosen path. In contrast, global techniques treat surface integration as an optimisation problem. Main advantages and disadvantages of both approaches were discussed.

Chapter 4

Texture Prediction Seen under Different Imaging Conditions

A methodology for predicting how a surface texture will appear if seen from different imaging geometries (e.g. distance from the camera, and direction of the light) and even under different spectral characteristics of the incident light and sensitivity of the sensor is presented in this chapter. The 4-source colour photometric stereo is used to extract the surface shape and the reflectance properties of the surface from a certain distance. From this information, a general texture prediction framework composed of two methods is proposed: one which allows us to predict the image intensities directly (direct image prediction) and another which allows us to predict first the surface shape information and then the image intensities (image prediction via surface prediction).

4.1 Introduction

In the last chapter, the main techniques used to recover surface information were presented. Among them, 4-source colour photometric stereo which allows us to recover surface texture information as well as to render images under novel lighting conditions was detailed. It is important to remember that our objective is to use this surface texture information for image prediction.

We have explained in depth how photometric stereo can be used to render new

images of a surface under novel light directions. In this chapter we go further and a prediction framework which allows us to predict how a surface texture will appear if seen from different imaging geometries and different sensor sensitivity and spectral characteristics of the incident light, is proposed.

This chapter is structured in the following way: Section 4.2 describes our first approach to surface texture prediction. In Section 4.3 a more formal approach based on the 4-source colour photometric stereo is proposed. More specifically, two methods for predicting how a surface texture appears when seen from longer distances are presented. The first one analyses the direct relationship between image texture information (image intensities) under two different resolutions. This leads to *direct image prediction*. The second predicts first how the surface itself would be approximated in a lower resolution from the original one and then, from this lower resolution, it predicts the image it would create. This leads to *image prediction via surface prediction*. Finally, some conclusions are presented in Section 4.4.

4.2 Initial Prediction Approach

In 2000 we proposed a first method [101] to predict how a surface texture appears if seen from different distances. In this approach, two important parts were analysed: in the first, the relationship between the surface information under different resolutions and, in the second, the relationship between the colour information under different resolutions. As a result of that study, a method which allows us to generate colour images of a surface under varying geometries such as lighting conditions and distance from the camera[1] was presented.

Basically, the idea of our first approach was to formulate the corresponding set of greyscale photometric equations theoretically and analyse them in order to find the relationship between the surface information in the fine resolution and the surface information in the coarse resolution (see Figure 4.1). Note that for the sake of simplicity we decided to focus on the greyscale case first. This initial approach was based on the intuitive fact that the intensity of a pixel is proportional to the amount

[1] A change in camera distance only implies a change in the z-axis.

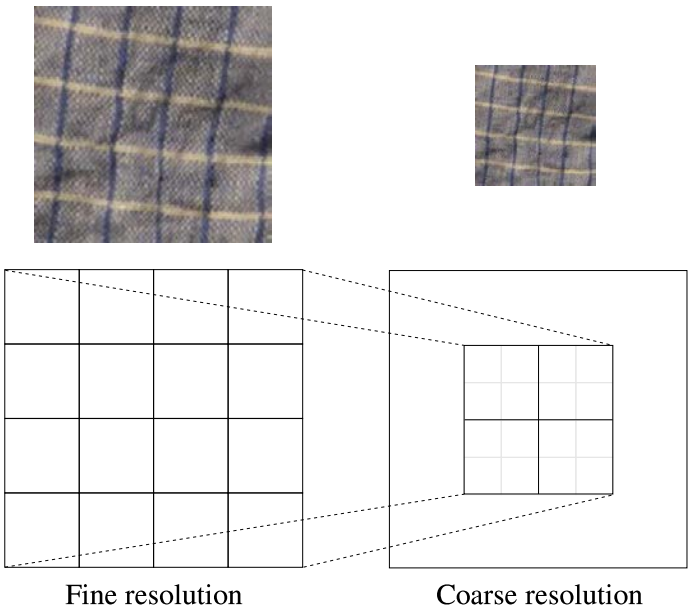


Figure 4.1: Example of a surface seen in a fine and coarse resolution.

of light reflected by the surface patch projected onto the pixel. Moreover, in our case it was assumed that the geometry of the surface and the light source did not vary if the distance of the camera was changed. Hence, the pixel intensity in another resolution was the result of the amount of light received by the new surface patch projected onto the pixel. From this basic principle, the photometric expressions describing the desired relationship were obtained. Furthermore, this approach was extended to colour information applying greyscale photometric stereo for each colour band separately.

4.2.1 Normal Vectors under Different Resolutions

In order to predict how a surface texture looks when seen from different distances, it is necessary to understand what happens with the surface shape information if the distance of the camera is changed. Figure 4.2 illustrates this problem and allows us to formulate the following question: what will the normal vectors be if the camera distance is changed leading to a new image in which every pixel is the union of several old pixels? This is the question we answer in this section, deriving the relationship

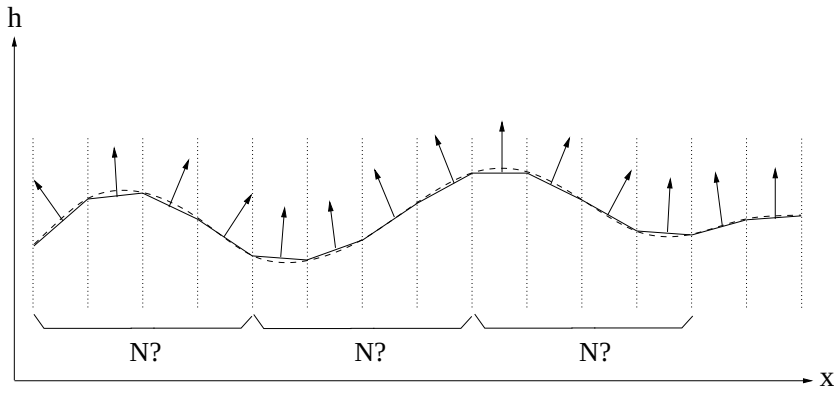


Figure 4.2: A surface is approximated by many flat facets in fine resolution. In coarse resolution these are replaced by large flat facets. What is the gradient vector of a large facet as a function of the gradient vectors of the small facets it replaces?

between the normal vectors when they are calculated in different image resolutions.

To answer this question we must first set up a consistent notation. Let us indicate by indices ij a pixel at location (i, j) in the coarse grid. This pixel is made up of several pixels from the fine resolution grid, some of which contribute only partially. Let us aslo, for the moment, ignore how much each pixel in the fine resolution contributes to pixel ij in the coarse resolution, and let us simply say that "superpixel" ij corresponds to a tile of size $K \times L$ of the fine resolution pixels. We shall refer to the pixels in the coarse resolution as "superpixels" and the term "pixel" will be used only for the fine resolution pixels. Let us now indicate, by indices mn, the position of the pixels in the fine resolution grid. Each superpixel has the intensity value I_{ij}^u in image u of the photometric stereo set. Remember that a greyscale photometric stereo consists of three images, $u = 1, 2$ or 3. Each superpixel also has two values for the components of its gradient vector. Let us call them p_{ij} and q_{ij}. Each superpixel corresponds to a tile of pixels. We want to keep track of the superpixel to which a pixel contributes, so we shall give every pixel three sets of indices: one to tell us to which tile it belongs, one to tell us where in the tile it is, and one set to tell us its location in the fine resolution grid. So, any pixel contributing to superpixel ij will have indices $ij; kl$ and mn, where $k = 1, 2, ..., K$ and $l = 1, 2, ..., L$. Any other quantity associated with pixel $ij; kl$ and mn will be indicated by the same notation

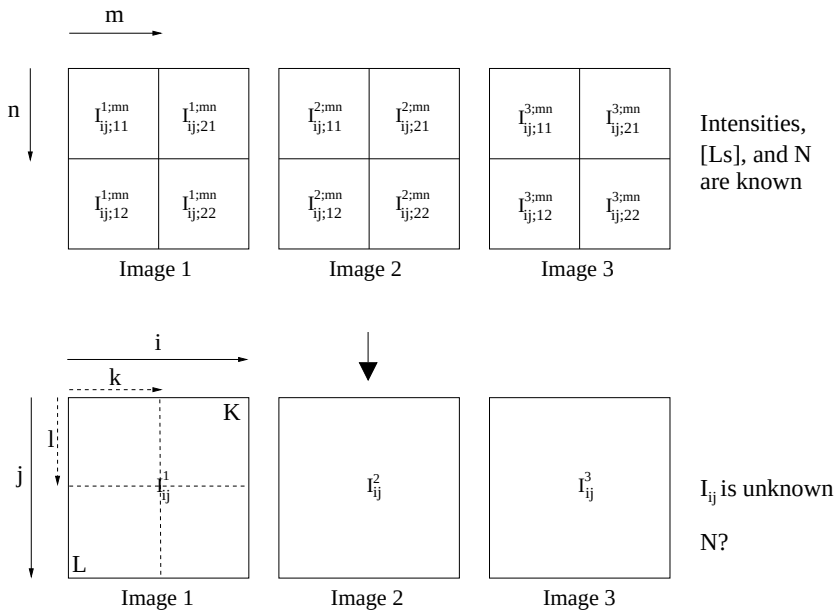


Figure 4.3: Example: prediction of the surface information for superpixel ij, which corresponds to a tile of size 2×2. Indices kl: pixel position inside the superpixel; indices mn: pixel position in the fine resolution grid; indices ij: pixel position in the coarse grid.

as for superpixel ij. That is, pixel $ij;kl$ and mn has the intensity value $I^{u;mn}_{ij;kl}$ in image u and the components of its gradient vector are $p^{mn}_{ij;kl}$ and $q^{mn}_{ij;kl}$. Therefore, we can define our problem as

> **Predict** the intensity value of superpixel I^u_{ij} for a given direction of illumination u and its gradient components p_{ij} and q_{ij}, given the intensity and the gradient components $I^{u;mn}_{ij;kl}$, $p^{mn}_{ij;kl}$ and $q^{mn}_{ij;kl}$ of all the pixels in the fine resolution.

To deal with this problem, and also to illustrate the use of this notation, let us consider a simple example. Suppose that superpixel ij corresponds to a tile of 2×2 pixels as shown in Figure 4.3. Using this notation and applying Equations 3.4 and 3.6, the following set of equations can be written.

$$\begin{bmatrix} I^{1;mn}_{ij;kl} \\ I^{2;mn}_{ij;kl} \\ I^{3;mn}_{ij;kl} \end{bmatrix} = \rho^{mn}_{ij;kl} \begin{bmatrix} L^1_x & L^1_y & L^1_z \\ L^2_x & L^2_y & L^2_z \\ L^3_x & L^3_y & L^3_z \end{bmatrix} \begin{bmatrix} N^{mn}_{x;ij;kl} \\ N^{mn}_{y;ij;kl} \\ N^{mn}_{z;ij;kl} \end{bmatrix}$$
$$\Downarrow$$
$$\rho^{mn}_{ij;kl} \begin{bmatrix} N^{mn}_{x;ij;kl} \\ N^{mn}_{y;ij;kl} \\ N^{mn}_{z;ij;kl} \end{bmatrix} = \begin{bmatrix} L^1_x & L^1_y & L^1_z \\ L^2_x & L^2_y & L^2_z \\ L^3_x & L^3_y & L^3_z \end{bmatrix}^{-1} \begin{bmatrix} I^{1;mn}_{ij;kl} \\ I^{2;mn}_{ij;kl} \\ I^{3;mn}_{ij;kl} \end{bmatrix} \tag{4.1}$$

in which, for this particular example, m and n can be expressed as

$$\begin{aligned} m &= (i-1)K + k \\ n &= (j-1)L + l \end{aligned} \tag{4.2}$$

with $K = L = 2$.

Let us indicate with $\mathbf{I_0}$ the vector of the greyscale intensities corresponding to the same position but in the three different images used in the photometric stereo. So, pixel intensities $I^{1;mn}_{ij;kl}$, $I^{2;mn}_{ij;kl}$ and $I^{3;mn}_{ij;kl}$ can be expressed as $\mathbf{I}^{mn}_{\mathbf{0};ij;kl}$ and the intensities of superpixel ij can be expressed as $\mathbf{I}_{\mathbf{0};ij}$. In what follows, we shall continue using bold face to indicate vectors. Matrices will be indicated by square brackets, while the rest of the quantities will be scalars. Hence, for a tile of size 2×2, the following set of equations can be written.

$$\begin{aligned}
\mathbf{I}^{mn}_{\mathbf{0};ij;11} = \rho^{mn}_{ij;11} [L] \mathbf{N}^{mn}_{ij;11} &\Rightarrow \rho^{mn}_{ij;11} \mathbf{N}^{mn}_{ij;11} = [L]^{-1} \mathbf{I}^{mn}_{\mathbf{0};ij;11} \\
\mathbf{I}^{mn}_{\mathbf{0};ij;21} = \rho^{mn}_{ij;21} [L] \mathbf{N}^{mn}_{ij;21} &\Rightarrow \rho^{mn}_{ij;21} \mathbf{N}^{mn}_{ij;21} = [L]^{-1} \mathbf{I}^{mn}_{\mathbf{0};ij;21} \\
\mathbf{I}^{mn}_{\mathbf{0};ij;12} = \rho^{mn}_{ij;12} [L] \mathbf{N}^{mn}_{ij;12} &\Rightarrow \rho^{mn}_{ij;12} \mathbf{N}^{mn}_{ij;12} = [L]^{-1} \mathbf{I}^{mn}_{\mathbf{0};ij;12} \\
\mathbf{I}^{mn}_{\mathbf{0};ij;22} = \rho^{mn}_{ij;22} [L] \mathbf{N}^{mn}_{ij;22} &\Rightarrow \rho^{mn}_{ij;22} \mathbf{N}^{mn}_{ij;22} = [L]^{-1} \mathbf{I}^{mn}_{\mathbf{0};ij;22}
\end{aligned} \tag{4.3}$$

The albedos $\rho^{mn}_{ij;kl}$, and the normal vectors $\mathbf{N}^{mn}_{ij;kl}$ can be directly obtained by applying photometric stereo. A similar equation can be written for superpixel ij;

$$\mathbf{I}_{\mathbf{0};ij} = \rho_{ij} [L] \mathbf{N}_{ij} \Rightarrow \rho_{ij} \mathbf{N}_{ij} = [L]^{-1} \mathbf{I}_{\mathbf{0};ij} \tag{4.4}$$

However, in this case it is not possible to recover the surface information (ρ_{ij},$\mathbf{N}_{ij}$) because the intensities of superpixel $\mathbf{I}_{\mathbf{0};ij}$ are unknown. It is at this point where we want to find the relationship between the surface information of the four pixels forming the tile and that of superpixel ij.

As seen in Chapter 2, the intensity of a pixel depends on the surface properties (its reflectance function and topology), the light source (its characteristics and position) and the geometry of the imaging device (its position and other parameters). In an intuitive way, and without considering aspects such as the spectral dependence of the surface reflectance, the spectral characteristics of the light source, etc, we can say that the intensity of a pixel is proportionate to the amount of light reflected by the surface patch projected onto the pixel. Thus, if we change the camera distance, the surface patch projecting its light onto one superpixel will be different and the amount of light received by the sensor will also be different (see Figure 4.4).

In our case, the geometry of the surface and the light source do not vary if the camera distance is changed. Hence, the pixel intensity at a longer distance is the result of the amount of light received by the new patch. Therefore, in this simple example, in which the four initial patches form superpixel ij, superpixel ij will have the following intensity vector,

$$\mathbf{I}_{\mathbf{0};ij} = \frac{1}{KL} \sum_{k=1}^{k<K+1} \sum_{l=1}^{l<L+1} \mathbf{I}_{\mathbf{0};ij;kl}^{mn} \tag{4.5}$$

because the amount of light received by superpixel ij is the sum of the light projected by the four patches. Note that the amount of light is proportionate to the area of the patches. Therefore, by substituting $\mathbf{I}_{\mathbf{0};ij}$ in Equation 4.4 and using expression 4.3, the following two expressions can be obtained.

$$\rho_{ij}\mathbf{N}_{ij} = [L]^{-1}\,\mathbf{I}_{\mathbf{0};ij} = \frac{1}{KL}\,[L]^{-1} \sum_{k=1}^{k<K+1} \sum_{l=1}^{l<L+1} \mathbf{I}_{\mathbf{0};ij;kl}^{mn} \tag{4.6}$$

and

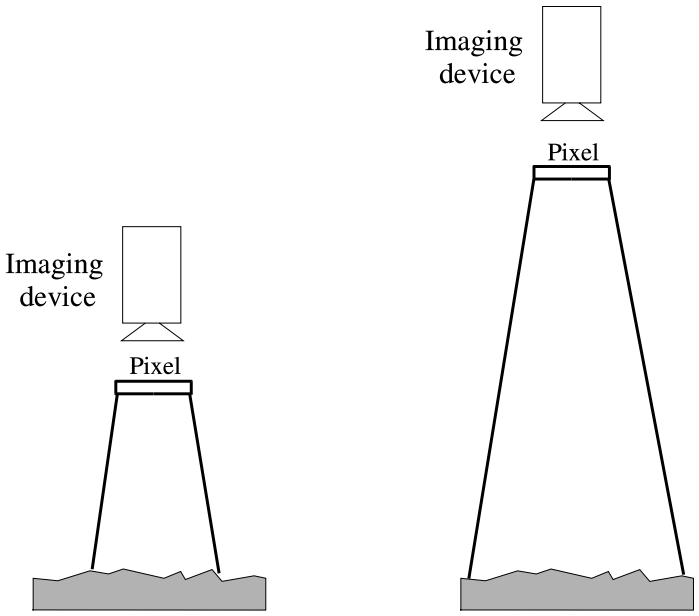


Figure 4.4: The intensity of a pixel is proportionate to the amount of light reflected by the surface patch projected onto the pixel.

$$\frac{1}{KL}\sum_{k=1}^{k<K+1}\sum_{l=1}^{l<L+1}[L]^{-1}\,\mathbf{I}^{mn}_{\mathbf{0};ij;kl}=\frac{1}{KL}\sum_{k=1}^{k<K+1}\sum_{l=1}^{l<L+1}\rho^{mn}_{ij;kl}\mathbf{N}^{mn}_{ij;kl} \tag{4.7}$$

These expressions allow us to obtain the relationship between the surface information of the four pixels forming the tile and superpixel ij, and allow to solve the proposed problem.

$$\frac{1}{KL}\sum_{k=1}^{k<K+1}\sum_{l=1}^{l<L+1}\rho^{mn}_{ij;kl}\mathbf{N}^{mn}_{ij;kl}=\rho_{ij}\mathbf{N}_{ij} \tag{4.8}$$

4.2.1.1 General case

We have seen by means of a simple example how to obtain the relationship between the surface information in a fine resolution and the surface information in a coarse resolution. Nevertheless, we have analysed the theoretical case in which a superpixel is made up of several complete pixels from the fine resolution grid. The next step is to analyse the general case in which a superpixel is made up of several pixels, some

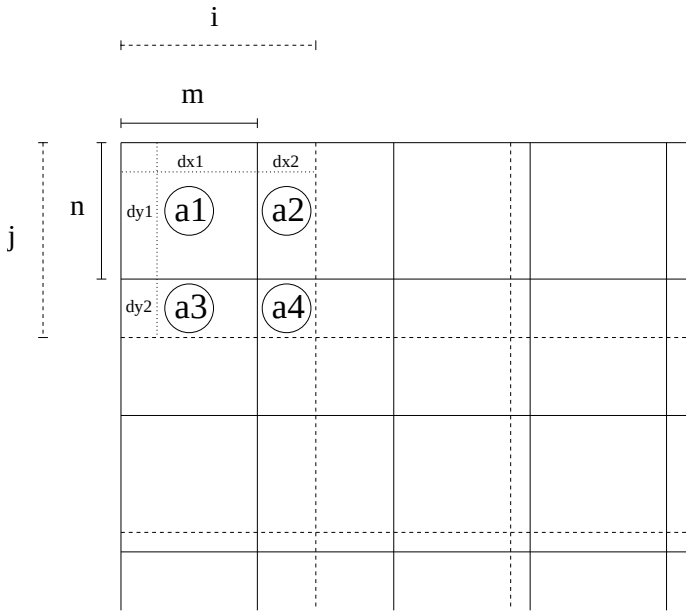


Figure 4.5: General case of image resolution change.

of which contribute only partially (see Figure 4.5). In this situation, it is necessary to know by how much each pixel in the fine resolution contributes to superpixel ij in the coarse resolution. So, the intensity vector of superpixel ij can be expressed as

$$\mathbf{I}_{\mathbf{0};ij} = a_{11}\mathbf{I}^{mn}_{\mathbf{0};ij;11} + a_{21}\mathbf{I}^{mn}_{\mathbf{0};ij;21} + ... + a_{kl}\mathbf{I}^{mn}_{\mathbf{0};ij;kl} \tag{4.9}$$

in which a_{kl} is the area of patch kl which projects an amount of light which forms the intensity vector $\mathbf{I}^{mn}_{\mathbf{0};ij;kl}$.

Note that indices mn can not now be calculated by expression 4.2 because the number of pixels forming the superpixel is not an integer. Therefore, new expressions which allow us to obtain these indices have to be extracted. In order to do that we have assumed the interpixel distances (d_m, d_n) in the fine resolution grid and (d_i, d_j) in the coarse grid are known. Hence, using these distances, areas a_{kl} of the pixels which contribute to superpixel ij can be calculated. When the desired areas are known, the normal vector of superpixel ij and, consequently, the two values p_{ij} and q_{ij} of its gradient vector can be recovered as follows

$$\rho_{ij}\mathbf{N}_{ij} = \sum_{k=1}^{k<K+1} \sum_{l=1}^{l<L+1} a_{kl}\, \rho_{ij;kl}^{mn} \mathbf{N}_{ij;kl}^{mn} \tag{4.10}$$

To sum up, in order to predict the normal vectors in subsequent resolutions, it will be necessary to know the relationship between the pixels in these resolutions. Then, knowing the areas of the pixels forming the superpixel, surface shape information can be predicted.

4.2.2 Colour Information under Different Resolutions

In order to achieve our goal, which is to predict a surface texture when seen from a longer distance, as well as to know what happens with the shape information, it is also necessary to understand what happens with the colour information. This is the problem we want to discuss in this section, explaining how the surface information can be used to predict colour information in different image resolutions. To perform this task, we propose applying the greyscale photometric stereo for each colour band separately.

As defined in Section 4.2.1, each superpixel has the intensity value I_{ij}^{u} in image u of the photometric stereo set and the set of intensities corresponding to the three images u have been called $\mathbf{I}_{\mathbf{0};ij}$. Subscript $\mathbf{0}$ was used to indicate the monochromatic case. We shall continue with the same notation as in the monochromatic case but we shall replace subscript $\mathbf{0}$ with subscript $\mathbf{v}$ to indicate one of the three colour bands, $\mathbf{v} = 1, 2$ or 3 for Red, Green and Blue respectively. So, each superpixel has the intensity value $I_{\mathbf{v};ij}^{u}$ in image u at band $\mathbf{v}$ and each pixel in the fine resolution grid has the intensity $I_{\mathbf{v};ij;kl}^{u;mn}$. Therefore, we can indicate the intensities vector of superpixel ij as $\mathbf{I}_{\mathbf{v};ij}$ at band $\mathbf{v}$ and the intensities vector of pixel $ij;kl$ and mn as $\mathbf{I}_{\mathbf{v};ij;kl}^{mn}$ at band $\mathbf{v}$. The use of photometric stereo for every colour band implies that three albedos and three normal vectors are obtained. So, we shall use index $\mathbf{v}$ with the albedos and normal vectors to tell us from which colour band they were computed. By using this notation along with the general case illustrated in Figure 4.5, the followings equations, one for each colour band ($\mathbf{v} = 1, 2$ or 3), can be written.

$$\mathbf{I}_{\mathbf{v};ij} = \sum_{k=1}^{k<K+1} \sum_{l=1}^{l<L+1} a_{kl}\, \mathbf{I}^{mn}_{\mathbf{v};ij;kl} \tag{4.11}$$

Therefore, using definition 4.1 in this equation, the following expression can be obtained.

$$\mathbf{I}_{\mathbf{v};ij} = \sum_{k=1}^{k<K+1} \sum_{l=1}^{l<L+1} a_{kl}\, \rho^{mn}_{\mathbf{v};ij;kl}\, [L]\, \mathbf{N}^{mn}_{\mathbf{v};ij;kl} \tag{4.12}$$

This equation can also be expressed as

$$[L]^{-1}\, \mathbf{I}_{\mathbf{v};ij} = \sum_{k=1}^{k<K+1} \sum_{l=1}^{l<L+1} a_{kl}\, \rho^{mn}_{\mathbf{v};ij;kl} \mathbf{N}^{mn}_{\mathbf{v};ij;kl} \tag{4.13}$$

As explained in section 3.5.1, the reflectance factor can be calculated by taking the magnitude of the left hand side of Equation 4.13. Therefore, the three albedos $\rho_{\mathbf{1};ij}$, $\rho_{\mathbf{2};ij}$, and $\rho_{\mathbf{3};ij}$ can be obtained using the following equation:

$$\rho_{\mathbf{v};ij} = \left| \sum_{k=1}^{k<K+1} \sum_{l=1}^{l<L+1} a_{kl}\, \rho^{mn}_{\mathbf{v};ij;kl} \mathbf{N}^{mn}_{\mathbf{v};ij;kl} \right| \tag{4.14}$$

When the reflectance factors are known, the normal vectors $\mathbf{N}_{\mathbf{1};ij}$, $\mathbf{N}_{\mathbf{2};ij}$, and $\mathbf{N}_{\mathbf{3};ij}$ can be calculated as

$$\mathbf{N}_{\mathbf{v};ij} = \frac{1}{\rho_{\mathbf{v};ij}} \sum_{k=1}^{k<K+1} \sum_{l=1}^{l<L+1} a_{kl}\, \rho^{mn}_{\mathbf{v};ij;kl} \mathbf{N}^{mn}_{\mathbf{v};ij;kl} \tag{4.15}$$

Finally, normal vector $\mathbf{N}_{ij}$ can be obtained by taking the mean of the normal vectors for each colour band. So, for every superpixel ij the components of its gradient vector p_{ij} and q_{ij} can also be calculated.

$$\mathbf{N}_{ij} = \frac{\mathbf{N}_{\mathbf{1};ij} + \mathbf{N}_{\mathbf{2};ij} + \mathbf{N}_{\mathbf{3};ij}}{3} \tag{4.16}$$

As well as surface shape, colour information for superpixel ij is obtained. We shall indicate this colour information by vector $\mathbf{C}_{ij}$ which allows us to represent the colour map in another resolution.

$$\mathbf{C}_{ij} = (\rho_{\mathbf{1};ij}, \rho_{\mathbf{2};ij}, \rho_{\mathbf{3};ij}) \tag{4.17}$$

Figure 4.6 illustrates the scheme of this strategy to predict surface information when seen from a longer distance. Note that by predicting the normal vectors and the colour map information, a rendering of the surface texture can be achieved in the same way we presented in Section 3.5.4.1. A gradient-based method, based on the notion of the *shading* of a facet and the predicted gradient and colour information, is used to obtain renderings of a surface when seen from different distances and under different directions of light.

4.2.3 Discussion

We have seen that the basic idea of this approach is to deduce the relationship between the surface information from two different resolutions intuitively. For simplicity's sake, the greyscale photometric stereo was used to formulate the problem theoretically. After obtaining the desired relationship, the method was extended to colour images applying greyscale photometric stereo for each colour band separately.

This way to perform colour photometric stereo, commonly known as *separate* colour photometric stereo, has been analysed in various works [27, 143, 7]. In general, the main conclusion provided by the authors is that the method works well for nearly all Lambertian surfaces although there are some cases in which it fails. For example, it can fail if the colour bands have different strengths. Suppose a blue surface in which the blue component is very strong, whereas the red and green components are weak and much more prone to noise corruption. The gradients recovered from the "weak" colour bands may be severely affected by the noise and the averaging over all three bands adds error to the correctly recovered "blue" gradient. Moreover, the algorithm also fails if the surface is not Lambertian or if it is not perfectly opaque

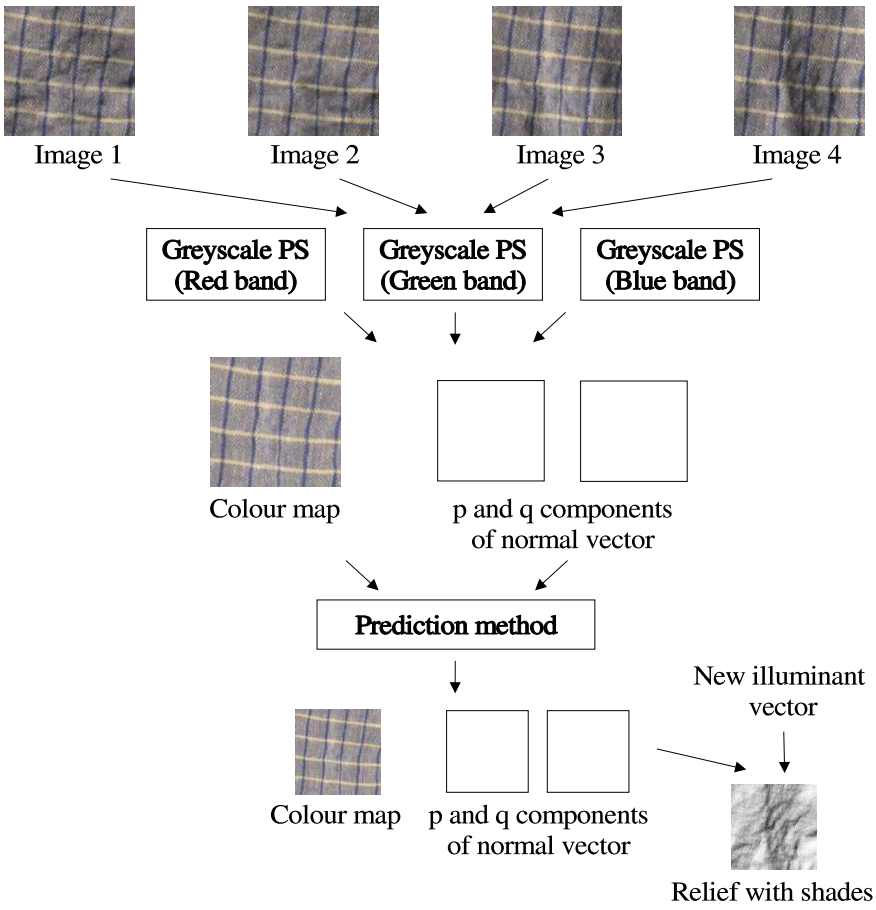


Figure 4.6: Scheme of our first prediction approach. As a result of applying the greyscale photometric technique to each colour band separately, the colour map and the surface shape information are predicted.

and there is a certain degree of translucency.

These problems, which are not directly related to the prediction but to the way of performing the photometric stereo, clearly affect this prediction approach. Note that the errors produced by the photometric stereo are propagated in the surface prediction process. Therefore the predicted images will also contain errors.

This prediction approach, as with all first approaches, also suffers from other important deficiencies: it does not account for the inverse square law of light propagation; it does not permit change in the sensor spectral sensitivity; it does not allow change in the illumination spectrum; it does not allow change in the direction of

illumination. With the idea of correcting these weak points and proposing a more formal theory, a new prediction approach is described in the next section.

4.3 New Prediction Approach

In Chapter 3 a survey of photometric stereo techniques was presented. Moreover, the 4-source colour photometric stereo was analysed in depth. This photometric technique plays an important role in the novel prediction approach, since it is used to recover the optimal gradient and colour estimates instead of using greyscale photometric applied to each colour band separately, as we did in our first prediction approach.

Furthermore, in Chapter 2, some important processes which affect the image formation process were explained. From these, some parameters and definitions were included in our study to deduce a formal theory which allows us to achieve the desired surface prediction under varying imaging conditions. More specifically, we describe two different prediction methods.

- **Direct image prediction.** This prediction method analyses the direct relationship between image texture information (image intensities) under two different resolutions. This leads us to *direct image prediction*. It takes into consideration parameters such as the sensitivity of the sensor, the spectral characteristics of the incident light, the illumination direction, etc. This method provides the possibility of rendering images of a surface under novel imaging geometries. However, it does not predict the surface shape information.

- **Surface prediction.** This second method predicts first how the surface itself would be approximated at a lower resolution from the original one and then predicts the image it would create. Height information is used to deduce the gradient vectors of the surface when seen from different distances. After using this predicted information and that extracted from the original distance, the predicted image would recreate the surface as seen from a longer distance. This

method leads to *image prediction via surface prediction*, but we shall refer to it as *surface prediction* for short.

Both methods were designed with the initial purpose of predicting the surface texture information when seen under different imaging conditions. The first attempts to predict the pixel intensities, while the second is focused on the prediction of gradient vectors. However, as it will be seen in the following sections, both methods allow us to render new images of a surface. A comparision between them will be carried out in the experimental section in order to discuss their performance. We will see how the direct image prediction produces, in general, smaller errors in the images while the surface prediction method produces the best shape predictions.

Summarising, we can define the main contributions of this new texture prediction framework in the following points:

- **Better surface recovery.** The 4-source colour photometric stereo is used to recover the optimal gradient and colour estimates instead of using greyscale photometric stereo applied to each colour band separately. This will allow us to achieve better results from the photometric stereo and, consequently, better predictions.
- **Better modelling.** More aspects and parameters have been included in this prediction proposal. Hence, a general method which allows us to predict the image texture information when seen in a different imaging setup is presented.
- **New shape prediction.** A new method based on the prediction of the planar patches of a surface when seen from a longer distance is proposed to obtain the desired surface shape information.

In what follows, both image and surface prediction methods are described.

4.3.1 Direct Image Prediction

Following the procedure in the previous section, we begin by considering two grids referring to the pixels of two images of the same surface captured from two different

distances. One corresponds to the higher resolution image and is finer than the other. We continue to refer to the pixels of the coarse resolution as "superpixels" and the term "pixel" is used only for fine resolution pixels. Each superpixel is made up of several pixels. Let us, for the moment, ignore by how much each pixel of the fine resolution contributes to the superpixel of the coarse resolution, and let us simply say that it corresponds to a tile of size $K \times L$ of fine resolution pixels. Some of the fine resolution pixels are only partially inside a superpixel, so we do not assume that the two different resolutions are such that a superpixel is replaced by an exact and integer number of fine resolution pixels of that tile. In this case, we define our problem as

Predict the intensity value of a superpixel I^u for a given direction of illumination u, given $\rho(\lambda)$, p and q for all the pixels of the fine resolution.

The values of $\rho(\lambda)$, p and q have been computed by using 4-source colour photometric stereo. Note that we have added the wavelength λ in the reflectance property (albedo). Although the colour photometric stereo scheme we use can deal with non-Lambertian surfaces [8], we assume here that the surface we are dealing with is Lambertian.

Let us begin by considering a facet of the surface of size $R_A \times R_A$ centred at position (α_0, β_0) and a sensor cell of size $X \times Y$ centred at position (x_0, y_0). Let us define the following quantities:

- $\mathcal{L}_A(\lambda)$: The energy per unit time per unit area reaching the imaged surface from a point illuminating source in a certain imaging setup denoted by A.
- $\mathbf{N}_{\alpha\beta}$: Normal vector to the imaged surface at point (α, β).
- $\mathcal{G}(\mathbf{N}_{\alpha\beta}, \mathbf{u}_A)d\alpha d\beta$: Geometric factor determining the amount of incident radiation from the light source at direction $\mathbf{u}_A$, received by an infinitesimal tile of size $d\alpha d\beta$.

- $\rho_{\alpha\beta}(\lambda)$: The fraction of the incident radiation reflected by infinitesimal tile $d\alpha d\beta$ at wavelength λ.
- $\mathcal{G}(\mathbf{N}_{\alpha\beta}, \mathbf{S}_A(x,y))dxdy$: Geometric factor affecting the fraction of reflected radiation reaching the area of the sensor element $dxdy$ centred at (x,y) from surface point (α,β) with the direction of the particular sensor element defined by unit vector $\mathbf{S}_A(x,y)$.
- $\frac{1}{2\pi {d_A}^2}$: Fraction of the reflected radiation reaching a unit area at distance d_A. The quantity $2\pi {d_A}^2$ is the area of the hemisphere over which the opaque point (α,β) is expected to reflect radiation.
- $\mathcal{S}(x,y,\lambda)$: Sensitivity of sensor element (x,y).
- Δt_A: Time interval during which the aperture of the sensor was open when the image at imaging setup A was being captured.
- k_A: A factor encompassing all photomultiplier and quantization processes used to convert the recorded energy into grey values.
- $rec(x-x_0, y-y_0)$: A rectangular of size $X \times Y$ of sensor surface centred at sensor point (x_0,y_0).
- $rec(\alpha-\alpha_0, \beta-\beta_0)$: A rectangular of size $R_A \times R_A$ of the imaged surface centred at point (α_0,β_0).

Then the grey level recorded by sensor $X \times Y$ centred at (x_0,y_0) is:

$$\begin{aligned} I_A(x_0,y_0) &= \frac{\Delta t_A k_A}{2\pi} \int_x \int_y \int_\alpha \int_\beta rec(x-x_0, y-y_0) rec(\alpha-\alpha_0, \beta-\beta_0) \\ &\quad \mathcal{G}(\mathbf{N}_{\alpha\beta}, \mathbf{S}_A(x,y)) \mathcal{G}(\mathbf{N}_{\alpha\beta}, \mathbf{u}_A) \frac{1}{{d_A}^2(\alpha,\beta,x,y)} \\ &\quad \int_\lambda \mathcal{S}(x,y,\lambda) \rho_{\alpha\beta}(\lambda) \mathcal{L}_A(\lambda) d\lambda \; dx \; dy \; d\alpha \; d\beta \end{aligned} \tag{4.18}$$

Let us assume that over rectangle $rec(x-x_0, y-y_0)$ the properties of the sensor are uniform, i.e. $\mathcal{S}(x,y,\lambda) = \mathcal{S}(x_0,y_0,\lambda)$, that over rectangle $rec(\alpha-\alpha_0, \beta-\beta_0)$ the

properties of the surface are uniform, i.e. $\rho_{\alpha\beta}(\lambda) = \rho_{\alpha_0\beta_0}$, $\mathbf{N}_{\alpha\beta} = \mathbf{N}_{\alpha_0\beta_0}$ and that all points of the surface and all points of the sensor are effectively in the same relative orientation defined by unit vector $\mathbf{S}_A(x, y) = \mathbf{S}_A(x_0, y_0)$ and at the same relative distance[2] $d_A(\alpha, \beta, x, y) = d_A(\alpha_0, \beta_0, x_0, y_0)$. We may also assume that $\mathbf{S}_A(x_0, y_0)$ and $d_A(\alpha_0, \beta_0, x_0, y_0)$ are virtually the same for all surface patches and all sensor elements so we can drop the dependence on (x_0, y_0) and (α_0, β_0). Then we can write:

$$I_A(x_0, y_0) = \frac{\Delta t_A k_A}{2\pi {d_A}^2} {R_A}^2 XY \mathcal{G}(\mathbf{N}_{\alpha_0\beta_0}, \mathbf{S}_A)) \mathcal{G}(\mathbf{N}_{\alpha_0\beta_0}, \mathbf{u}_A) \int_\lambda \mathcal{S}(\lambda) \rho_{\alpha_0\beta_0}(\lambda) \mathcal{L}_A(\lambda) d\lambda \tag{4.19}$$

For a Lambertian surface

$$\mathcal{G}(\mathbf{N}_{\alpha_0\beta_0}, \mathbf{S}_A) = 1 \tag{4.20}$$

and

$$\mathcal{G}(\mathbf{N}_{\alpha_0\beta_0}, \mathbf{u}_A) = \frac{\mathbf{N}_{\alpha_0\beta_0} \cdot \mathbf{u}_A}{|\mathbf{N}_{\alpha_0\beta_0}||\mathbf{u}_A|} = \frac{\mathbf{N}_{\alpha_0\beta_0} \cdot \mathbf{u}_A}{|\mathbf{N}_{\alpha_0\beta_0}|} \tag{4.21}$$

Let us now assume a different imaging setup, call it B, for which the light which reaches sensor (x_0, y_0) comes from an extended rectangle of size $R_B \times R_B$. If $rec(\alpha - \alpha_0, \beta - \beta_0)$ is rather extended, with variable shape and reflectance properties, we must use the following equation for the recorded brightness by sensor (x_0, y_0):

$$\begin{aligned} I_B(x_0, y_0) \;=\; & \frac{\Delta t_B k_B XY}{2\pi} \int_\alpha \int_\beta rec(\alpha - \alpha_0, \beta - \beta_0) \mathcal{G}(\mathbf{N}_{\alpha_0\beta_0}, \mathbf{S}_B) \mathcal{G}(\mathbf{N}_{\alpha_0\beta_0}, \mathbf{u}_B) \\ & \frac{1}{{d_B}^2(\alpha, \beta)} \int_\lambda \mathcal{S}(\lambda) \rho_{\alpha\beta}(\lambda) \mathcal{L}_B(\lambda)\; d\lambda\; d\alpha\; d\beta \end{aligned} \tag{4.22}$$

We can still assume that no matter how extended rectangle $rec(\alpha - \alpha_0, \beta - \beta_0)$ is, all of its parts are at the same distance from the sensor, and therefore $d_B(\alpha, \beta) = d_B$.

[2] This is equivalent to assuming that the sensor is at infinity with respect to the imaged surface.

Further, we can still assume Lambertianity, so the two geometric factors again take the form of equations 4.20 and 4.21. Then:

$$I_B(x_0, y_0) = \frac{\Delta t_B k_B XY}{2\pi {d_B}^2} \int_\alpha \int_\beta rec(\alpha - \alpha_0, \beta - \beta_0) \frac{\mathbf{N}_{\alpha_0\beta_0} \cdot \mathbf{u}_B}{|\mathbf{N}_{\alpha_0\beta_0}|} \int_\lambda \mathcal{S}(\lambda)\rho_{\alpha\beta}(\lambda)\mathcal{L}_B(\lambda)\, d\lambda\, d\alpha\, d\beta \tag{4.23}$$

Let us assume that this big rectangle of size $R_B \times R_B$ can be analysed as a superposition of smaller rectangles each of uniform properties for which Equation 4.19 can be written. Then the integral over α and β in Equation 4.23 may be replaced by one sum over all these facets:

$$I_B(x_0, y_0) = \frac{\Delta t_B k_B XY}{2\pi {d_B}^2} \sum_{k=1}^{K} \sum_{l=1}^{L} \frac{\mathbf{N}_{kl} \cdot \mathbf{u}_B}{|\mathbf{N}_{kl}|} A_{kl} \int_\lambda \mathcal{S}(\lambda)\rho_{kl}(\lambda)\mathcal{L}_B(\lambda)\, d\lambda \tag{4.24}$$

where A_{kl} is the area of facet (k, l) inside rectangle $R_B \times R_B$.

Note that this formula is very general: it allows us to predict the intensity value of a superpixel from the information we have on its constituent pixels of the fine resolution, even when seen under a different imaging setup, i.e. a different sensor, under lights with different spectral characteristics and different orientation than those under which the original images from which the surface information was extracted, were captured.

If we assume that the spectral properties of the sensor and the light remain constant then we can redefine the reflectance function of the surface as

$$\tilde{\rho}_{kl} \equiv \int_\lambda \mathcal{S}(\lambda)\rho_{kl}(\lambda)\mathcal{L}_B(\lambda)\, d\lambda \tag{4.25}$$

At the same time, we can assume that the calibration constant k_B is chosen so that all constant factors cancel out. Then, equation 4.24 is simplified to:

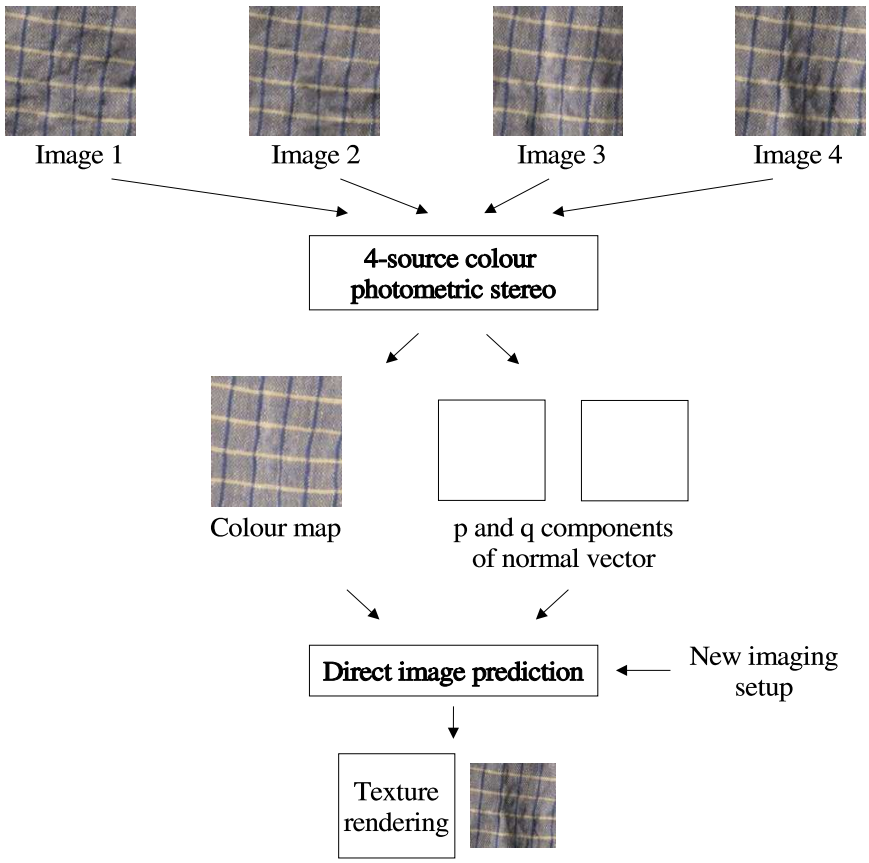


Figure 4.7: Scheme of the direct image prediction approach. As a result of applying the colour photometric technique and the prediction method, novel images under different imaging conditions can be predicted.

$$I_B(x_0, y_0) = \sum_{k=1}^{K} \sum_{l=1}^{L} A_{kl} \frac{N_{kl} \cdot u_B}{|N_{kl}|} \tilde{\rho}_{kl} \tag{4.26}$$

This is the equation we used for our experiments.

Figure 4.7 illustrates the scheme of this prediction approach. First, 4-source colour photometric stereo is applied to the original set of images with the aim of recovering the colour map and the gradient vectors of the surface. Afterwards, this information is used as input for our image prediction method. The result of this prediction process is a new image of the surface seen from a longer distance and under the desired imaging conditions.

4.3.2 Surface Prediction

In contrast to the direct image prediction method described above, we propose, in this section, a new method which tries to predict first how the surface itself would be approximated in a different resolution from the original one, and then predict the image it would create when seen from the new distance. Our intention is to analyse what happens with the surface shape information when the camera distance changes. As in our first approach, we try to answer the following question: What will the normal vectors be if the distance of the camera is changed leading to a new image in which every pixel is the union of several old pixels?

In this proposal we focus on the prediction of the facet by which the surface is approximated locally in the coarse resolution. In other words, assuming that a surface is a collection of planar facets, where each one has its own surface orientation $\mathbf{N}$, we propose to predict the planar facet of the superpixel by computing the average plane in the least square error sense passing through the facets of the fine resolution. Therefore, we first perform a surface integration in the fine resolution using the gradient vectors recovered by photometric stereo. Next, we reconstruct the facet of the superpixel by using the information recovered in the fine resolution. Finally, the normal vector of the recovered facet is the normal vector of the superpixel.

Note that surface integration is necessary in order to compute the height $z = S(x, y)$ at each point on the surface in the fine resolution (*height map*). By using the height information, and knowing the facets which contribute to a superpixel, the planar facet can be computed. Consequently, the recovered normal vector will be the normal vector of the superpixel. Fig. 4.8 illustrates this idea when a superpixel is formed from a tile of 2×2.

If we define a surface as $(x, y, S(x, y))$ where $S(x, y)$ is the height at point (x, y), the normal as a function of (x, y) is

$$\mathbf{N}(x, y) = \frac{1}{\sqrt{p^2 + q^2 + 1}}(p, q, -1)^T \tag{4.27}$$

where the partial derivatives of $S(x, y)$ give us the values of the gradient vector $p = \frac{\partial S(x,y)}{\partial x}$ and $q = \frac{\partial S(x,y)}{\partial y}$.

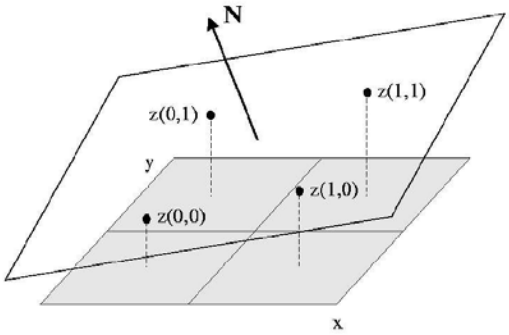


Figure 4.8: Normal vector of the superpixel obtanined by fitting the plane in the LSE sense.

In Section 3.6.1 different integration approaches which allow us to recover the height map of a surface by using the gradient vectors were presented. Basically, there are two types of surface integration approaches [51], *local integration along paths* and *global integration techniques*. Path integration techniques use local calculations of height increments by curve integrals. The main idea is to use the gradient vectors, which give the change in surface height in very small steps in either the x or the y direction, to recover a height map of the surface by summing these changes in height along a chosen path. In global techniques, the surface integration is treated as an optimisation problem. In other words, surface integration is considered as a variational problem where a certain function has to be minimised. Usually, some representation of the unknown surface, e.g. in terms of the Fourier basis functions, and the integrability condition [53, 176], is used to constrain the global optimisation process. For instance, Frankot and Chellappa [53] assume that the unknown surface function S satisfies the integrability condition: that is to say, the surface is continuous at any point. Obviously, this global integration technique works well if the surface is continuous. When the surface presents discontinuities, i.e. places where its derivatives do not exist, it is not integrable.

The photometric stereo surface reconstruction also yields a set of points where the recovery of the gradient vectors is impossible. In general, these are places which are in shadow in more than one of the four images used in the photometric stereo set. Such points are likely to occur at places where the surface has deep "ravines" and

can be associated with places where the surface is not differentiable and continuous. Therefore, as we need to preserve the surface roughness as much as possible and we want to perform the surface integration for these points where the gradient information is available, we apply a local integration technique which uses different paths to obtain the final surface reconstruction. Thus, when we reconstruct the surface, we stop the piecewise integration at these boundaries where the surface is not differentiable and continuous. Note that we only predict the normal vectors of the superpixels for which the surface integration in the fine resolution is done. Therefore, the surface shape we reconstruct and, consequenlty, the image intensity we will predict consist of image patches and not full images. This point will be analysed in depth in the experimental section.

As stated before, when using this surface prediction method, only the surface gradient vectors are directly predicted and not the image intensities. We generate these images by using the predicted gradient vectors and the average reflectance function for each surface tile, assuming that our sensor and illumination source do not change spectrally.

Figure 4.9 illustrates the scheme of this surface prediction approach. First, 4-source colour photometric stereo is applied to the original set of images, recovering the colour map and the gradient vectors of the surface. These gradient vectors are then used as input for the surface prediction process, achieving the corresponding gradient vectors in the coarse resolution. Moreover, by using the colour map extracted by photometric stereo and knowing the facets in the fine resolution which contribute to each superpixel, a new image of the surface seen from a longer distance is generated.

4.4 Summary

A methodology which allows us to predict how a surface texture will appear if seen from different distances and under different imaging conditions has been proposed.

This chapter was structured in two blocks. In the first, our first approach to sur-

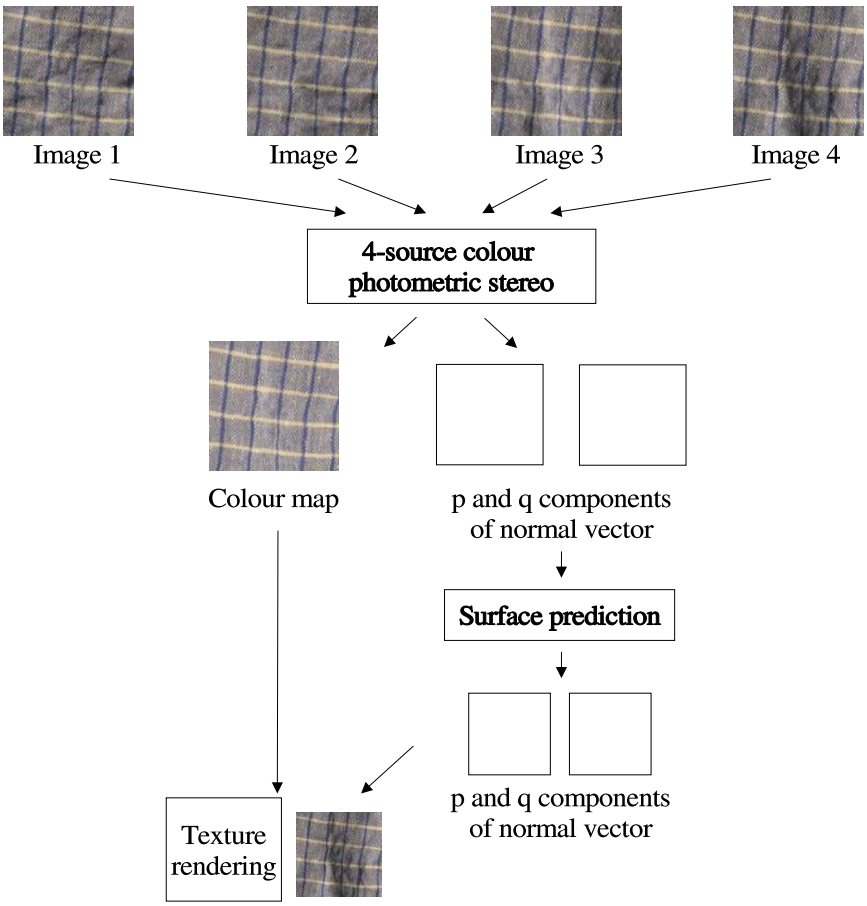


Figure 4.9: Scheme of the surface prediction approach.

face texture prediction was described. This approach applies greyscale photometric stereo to each colour band separately to recover surface information. Using this information, the method predicts how that surface appears as seen from a longer distance. The prediction method has been extended to the general case in which a pixel of the coarse resolution is made up of several pixels of the fine resolution, some of which contribute only partially. We have found that this first approach presents some weak points such as errors related to the way we perform the photometric stereo and the non-inclusion in the prediction process of important aspects in the image formation process. In the second block, a new prediction framework has been detailed. Basically, it presents a more accurate modelling of the image formation process which allows the prediction of image intensities as well as surface shape

information. This modelling was based on the **4-source colour photometric stereo**. Two methods called **direct image prediction** and **surface prediction** were proposed for predicting how a surface texture appears from longer distances. The first method analyses the direct relationship between image texture information (image intensities) under two different resolutions. The other predicts how the surface shape itself would be approximated in a resolution lower than the original one. This information can then be used to predict the image the surface would create.

These prediction approaches have to be tested and evaluated over a large number of surface textures in order to demonstrate the ability of predicting the image texture a particular surface texture will create when seen from a distance longer than the original one. This will be discussed in the experimental section. We will also analyse the role of surface properties such as roughness, directionality and specularity in our image and shape predictions.

Chapter 5

Texture Recognition

A model-based texture recognition system which classifies image textures seen from different distances and under different light directions is presented in this chapter. This system works on the basis of a surface model obtained by means of 4-source Colour Photometric Stereo (CPS) used to generate 2D image textures as they would have appeared if imaged under different imaging geometries. The proposed recognition system combines co-ocurrence matrices for feature extraction with a Nearest Neighbour classifier. Moreover, the recognition process allows one to guess the approximate direction of the light used to capture the test image.

5.1 Introduction

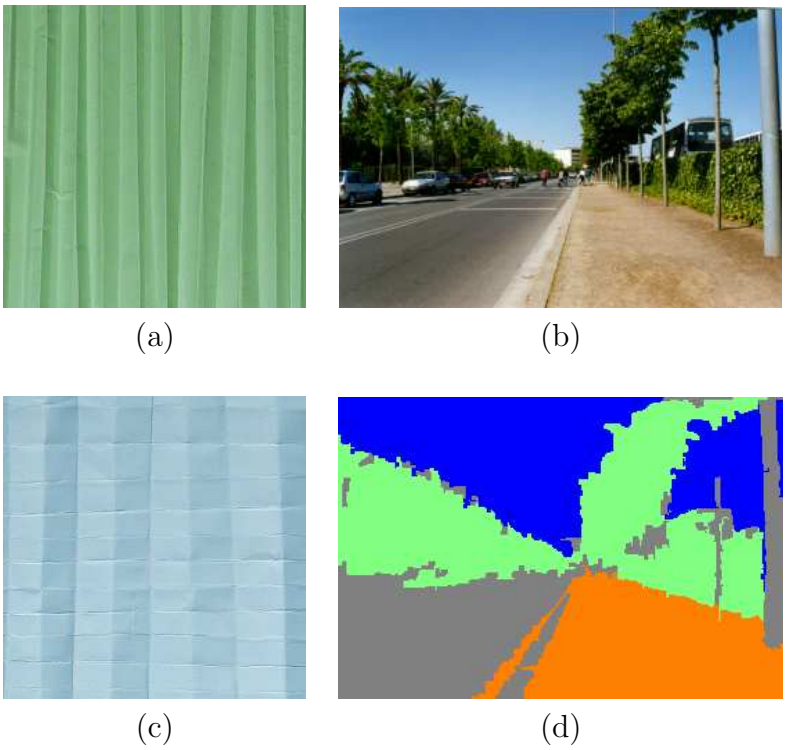
In its simplest form, image classification is the process of assigning similar images or regions of an image to the class to which they belong. It is well-known that there are two types of classification, namely *supervised* and *unsupervised*. In supervised classification, classes are specified a priori by an analyst, whereas in unsupervised classification, classes are automatically clustered into sets of prototype classes where in some cases a user specifies the number of desired categories. The recognition problem here is being posed as a classification or categorisation task where the classes are either defined by the analyst or are learned based on the similarity of patterns.

Many recognition/classification systems require a previous learning or modeling

task to obtain the *feature vectors* (represented by a set of d features) which characterise the objects of interest. Before classification can take place, some homogeneity or similarity criteria must be defined. These criteria are normally specified in terms of a set of feature measures, with each one providing a quantitative measure of a certain characteristic. The process which allows us to compute these features, called *feature extraction*, is concerned with the detection and localisation of particular image patterns which represent significant features in the image. These features are dependent on the application and are generally of two different origins: a global image property or a region of the image with a relevant measurable property.

While a global property tries to describe an image by extracting a single feature over the whole image (see Figure 5.1.(a) and Figure 5.1.(c)), a region property tries to describe the different regions of the image individually. Figure 5.1.(b) shows an example in which the main regions of the original image, such as road, sky, trees, and ground, are charactarised by a set of relevant measurable properties in order to perform their recognition (see Figure 5.1.(d)). This was done by Martí et al. [10, 110], proposing an object recognition system for outdoor scenes. Their work is based on a top-down strategy which segmented the main objects in the image, taking colour and texture features into account.

Texture is a fundamental characteristic of natural images which, in addition to colour, plays an important role in image understanding and scene interpretation. There are many researchers in image processing and computer vision areas who have considered the concept of feature vectors to cope with texture classification. In the area of database retrieval for instance, texture features are used to search an image database to find images similar to the image submitted by the user [128, 20]. In texture segmentation, many algorithms partition the image into a set of regions which are visually distinct and uniform with respect to textural properties [74, 107, 116]. In remote sensing and radar applications, texture features have been used to identify forest regions and their boundaries and to identify and analyse various crops [54, 166]. In industrial vision inspection, texture features have been used to perform the classification of different surface materials [105]. Obviously, there are many other applications in which texture is used to carry out a recognition or



(a) (b)

(c) (d)

Figure 5.1: Example of global and local image properties. In (a) and (c) a global image property is extracted in order to distinguish between the two images. In (b) each region of interest illustrated in (d) is characterised by a set of relevant measurable properties in order to perform object recognition.

classification task.

The aim of this chapter is to present a novel proposal of performing texture classification under varying imaging geometries. It will focus on the use of textural feature vectors computed over the global image to characterise the textures of interest. Note that texture features seem ideally suited to our purposes since the problem analysed is itself caused by variations in texture perception. Furthermore, this will be a supervised classification approach since prior knowledge about the analysed texture classes would be needed to perform their recognition.

Summarising, the main contributions of this chapter are:

- To propose a texture classification system to overcome the problem of classifier failure induced by varying imaging properties such as light direction and

camera distance.

- To integrate the prediction framework proposed in Chapter 4 into a texture classification system.
- To provide the capability of guessing an approximate direction of the light used to capture the test images.

The remainder of this chapter has been structured in the following way. In the next section we review the most important work on the topic of texture classification under varying imaging geometries. In Section 5.3, our model-based approach for texture classification is presented, describing first the general scheme then detailing the steps individually. Finally, some conclusions are presented in Section 5.4.

5.2 Texture Classification System

As illustrated in the introduction of this thesis (see Section 1.1), changes in the imaging geometry can significantly alter the appearance of the surface, implying significant variations in the image texture. These variations in the image may introduce critical misclassification rates in a typical texture classification system. Therefore, our main challenge is the problem of classification of textured surfaces imaged under varying geometries as well as the necessity of finding reliable methods of reducing classification errors caused by these changes in the geometry.

5.2.1 Related Work

Studying the dependence of texture on viewing and light directions is fairly new in texture research, therefore, there is only a limited amount of related work.

In computer graphics, traditional methods of 2D texture synthesis and texture mapping make no provision for the change in texture appearance with viewing and light directions. For example, in 2D texture mapping, when a single digital image of a rough surface is mapped onto a 3D object, the appearance of roughness is usually lost or distorted. Bump mapping [11, 14, 15] preserves some of the appearance

of roughness, but knowledge of the surface shape is required. Many of the problems associated with traditional texture mapping and bump mapping techniques are described by Koenderink and Van Doorn in [85].

In computer vision, very little work has been published on the topic of texture classification independent from the direction of the light. On the other hand, there are a great number of works which propose rotation and scale invariant texture classification [29, 52, 133, 108, 137, 91], as well as similar works on topics like multiscale and scale-space texture analysis [29, 17]. All these methods tend to take a texture obtained from a single view and change it for various scales and different angles, performing multiscale texture classification. In general, they assume the imaged texture is the same for all scales and rotations and, therefore, they do not consider that seeing a texture from a different physical distance and under different lighting conditions may change it entirely.

While analysing the published work on the topic of texture classification independent from the direction of light, we observed three different ways of dealing with this problem:

- The first strategy consists of extracting and using explicit separate 3D shape and surface albedo information. The colour and gradient vector of every visible surface patch describe the surface in a way independent from the light, and the classification can be done directly on the basis of this explicit information. For example, Barsky and Petrou [6] proposed an illumination-invariant classification scheme based on 5 descriptors for each surface patch obtained by means of colour photometric stereo: two gradient components and three colour components.

  McGunnigle and Chantler [113] proposed a rough surface classification scheme which extracts rotation invariant statistics from photometric estimates of the surface derivatives. Their method assumes that the surface is uniformly coloured. On the other hand, Chantler and Wu [24] proposed a novel classification scheme which is surface rotation invariant. They based it on the magnitude spectra of the surface derivatives extracted from photometric stereo.

- Another strategy to solve this problem is to study the immediate effects introduced by light direction to the observed 2D texture. Chantler [22] has shown that this effect can be described as a directional filter and, in principle, could be inverted. Recently, Chantler et al. [23] presented a formal theory demonstrating that changes in the tilt of the light direction make texture features follow super-elliptical trajectories in multi-dimensional feature spaces. Based on this work, Penirschke et al. [125] developed an illuminant rotation invariant classification scheme which uses photometric stereo for the detection of surface relief and Gabor features for feature extraction.

  In Appendix A, a study analysing the immediate effect of light direction over features extracted from the co-occurrence matrix is presented. This is used to perform a simultaneous surface texture classification and illumination tilt angle prediction.

- Finally, we can train a classifier on a wide selection of images of the same surface, obtained from various viewpoints and under various lighting conditions [33, 36, 32, 56]. Thus, the information on changes in surface appearance is explicitly built in the classifier using both the reflectance and the 3D relief information which allows it to recognise a surface correctly under novel viewing and lighting conditions. Leung and Malik [98, 99], following this strategy, developed a texture classification scheme which identifies 3D "textons" in the Columbia-Utrecht database for the purpose of light and viewpoint invariant classification. Basically, a 3D texton is an item in a vocabulary of prototypes of tiny surface patches with associated local geometric and photometric properties. More recently, Varma and Zisserman [164] proposed a new classification system which uses a distribution over textons obtained from training images as a texture model. On the other hand, Gonzalez [56] proposed a supervised statistical classification scheme which combines a bank of Gabor filters for feature extraction with a linear Bayes classifier to deal with changes in the light direction.

Figure 5.2 summarises these three different ways of dealing with the problem of

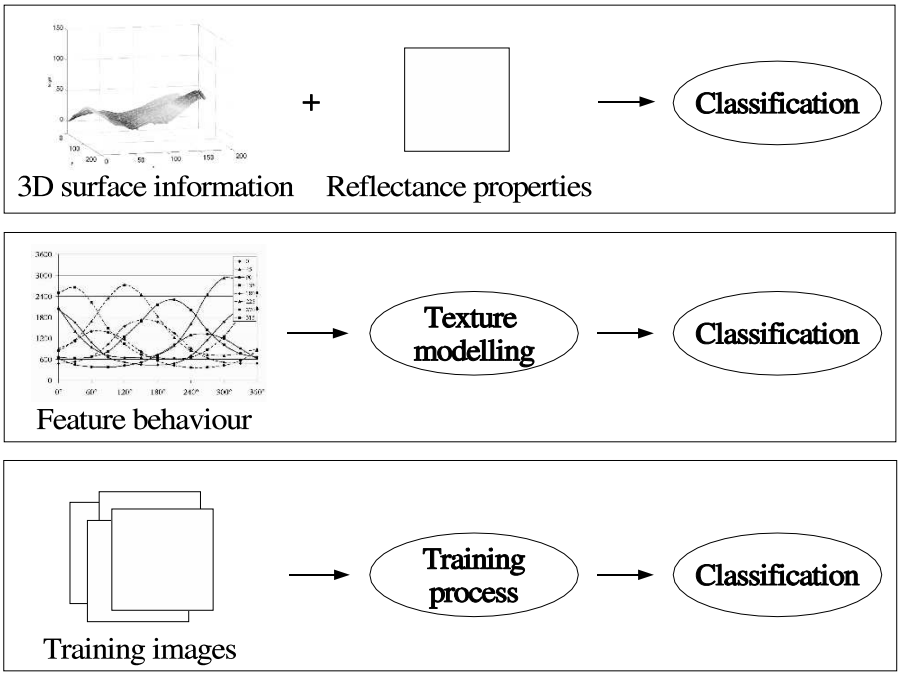


Figure 5.2: Strategies of light invariant texture classification.

classifying textures imaged with different light directions.

5.2.2 Discussion

We reviewed three different ways of dealing with the problem of texture classification independent from the direction of light. The first strategy uses the 3D shape and surface albedo information to characterise a surface texture. Although we saw that this strategy can be used to perform surface classification, it always requires the recovering of surface information. Therefore, we can not directly classify an unknown imaged texture since surface information is needed. That can be solved by using the second strategy which tries to study the immediate effects introduced by light direction over texture features. This approach requires one to work out theoretically the relationship between changes in the light direction and the values of the texture features used.

In this thesis, we have opted to follow the third strategy based on a set of training images. Roughly, we train a classifier on a selection of images. The main reason

of choosing this strategy in lieu of the others is to exploit the capability of colour photometric stereo to render images of a surface under novel lighting conditions as well as the capability of the prediction method proposed in Section 4.3.1 to create images of a surface texture under novel imaging geometries (e.g. camera distance).

The ability of performing the classification using a set of texture images is very important. Clearly, the quality of the classifier depends on the quality and size of the training set which is always finite. It is obvious that we can not think of using all the images corresponding to every single light direction in order to design a classifier.

In general, the process of designing a classifier must be inductive in the sense that the information obtained from the elements of the training set must be generalised to cover the whole feature space, implying that the classifier should be near optimal for all feasible textures, even for those it had never seen before. When tackling classification under varying lighting conditions, the problem is considerably increased. Not only should the classifier be able to discern different textures, it should also be able to perform robustly when faced with changes in the appearance of identical textures. In that case, it is not viable to train on every single light direction and store up an infinite number of training statistics to secure a more successful classification. One potential solution is a *model-based* approach. This takes advantage of the fact that it is not based on the actual surface but on various images obtained from its model. Therefore, it is a much more reasonable approach because it merely requires a finite and rather small data set to model the surface. Essentially, this strategy aims at modelling the texture variability in the feature space by means of a primary training stage.

Therefore, following this idea, we design a supervised model-based texture classification system in which a set of images under a variety of imaging configurations is used as a prior knowledge. In the next section this proposal is exhaustively described.

5.3 Model-Based Texture Classification

My approach is to integrate the surface texture information derived by colour photometric stereo, as described in Section 3.5.2, into a complete model-based texture classification system. Basically, the main idea consists of creating, by means of this surface texture information, a "virtual" database of image textures against which we compare the unknown test images in order to classify them.

In Chapter 3, we saw that photometric stereo is the technique which allows us to obtain surface texture information from only a few images of the same surface imaged under various light directions. Moreover, we have analysed various alternative techniques which also allow us to obtain 3D information about surfaces [170]: *stereo vision* [48], *optical flow* [5] and various *Shape from X* methods [65]. Over all these techniques, the photometric stereo technique was chosen because it has various advantages over all other methods. We have seen that it does not suffer from the correspondence problem like conventional stereo does, it does not make strong assumptions about the underlying surface structure like some Shape from X techniques do, and it allows us to recover both local colour and local gradient while flagging the places where some of its assumptions break down and recovery is impossible.

In supervised classification systems, a previous learning or modeling task is usually required to obtain the feature vectors which characterise the objects of interest. In our case, we use a virtual database of images as the input of this learning process in which the goal is simple: to extract those texture feature vectors which should be used as models in the classification process. It is important to note that surface information plays an important role in our proposal since it is used in a preliminary step to create image textures compatible with the imaging geometry of the test images.

Figure 5.3 illustrates the whole procedure of the proposed model-based texture classification system. Note that it is divided into two main phases:

- **Virtual database creation.** The process in which different images of each

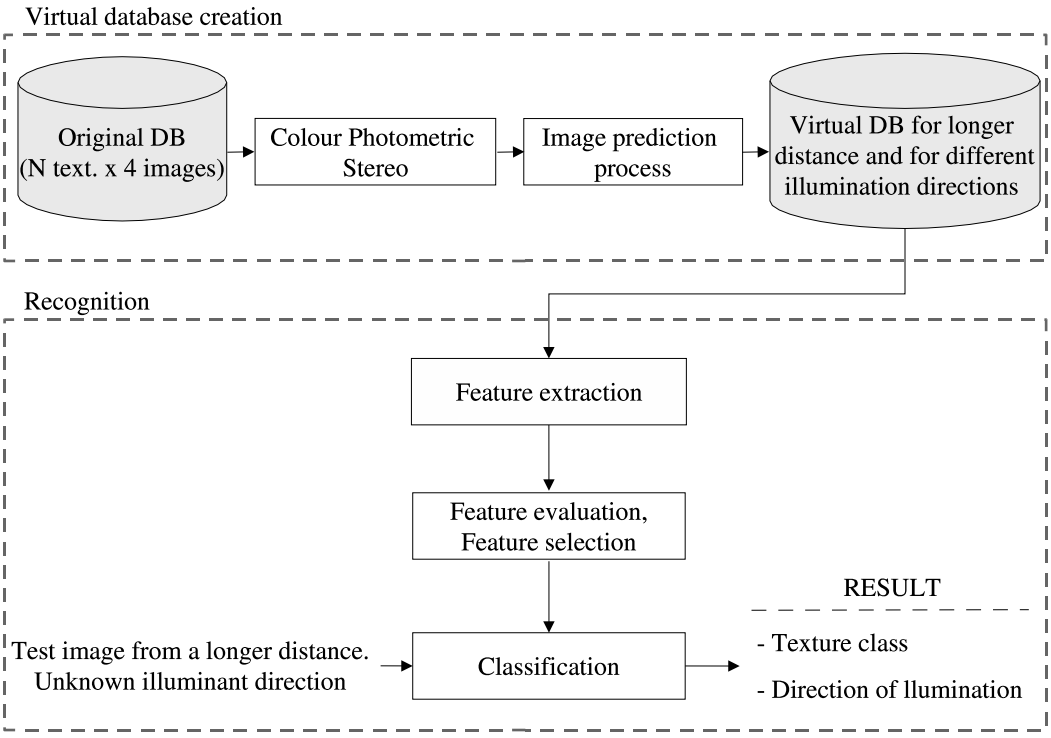


Figure 5.3: Recognition scheme. It is divided into two main phases: virtual database creation and recognition.

texture are created to be used as references in the classification process.

- **Recognition procedure.** Initially, each texture class in a learning process is modelled by a feature vector. Afterwards, in the classification process, an unknown test image is classified into the texture class to which it belongs.

We will also show how the proposed recognition procedure allows us to guess the approximate direction of the light used to capture the test images.

5.3.1 Virtual Database Creation

The virtual database creation comes in two "flavours": creation of the virtual database for test images seen from the same distance as the original images and creation of the virtual database for test images seen from a longer distance than that of the original images. When the test images are known to have been captured at

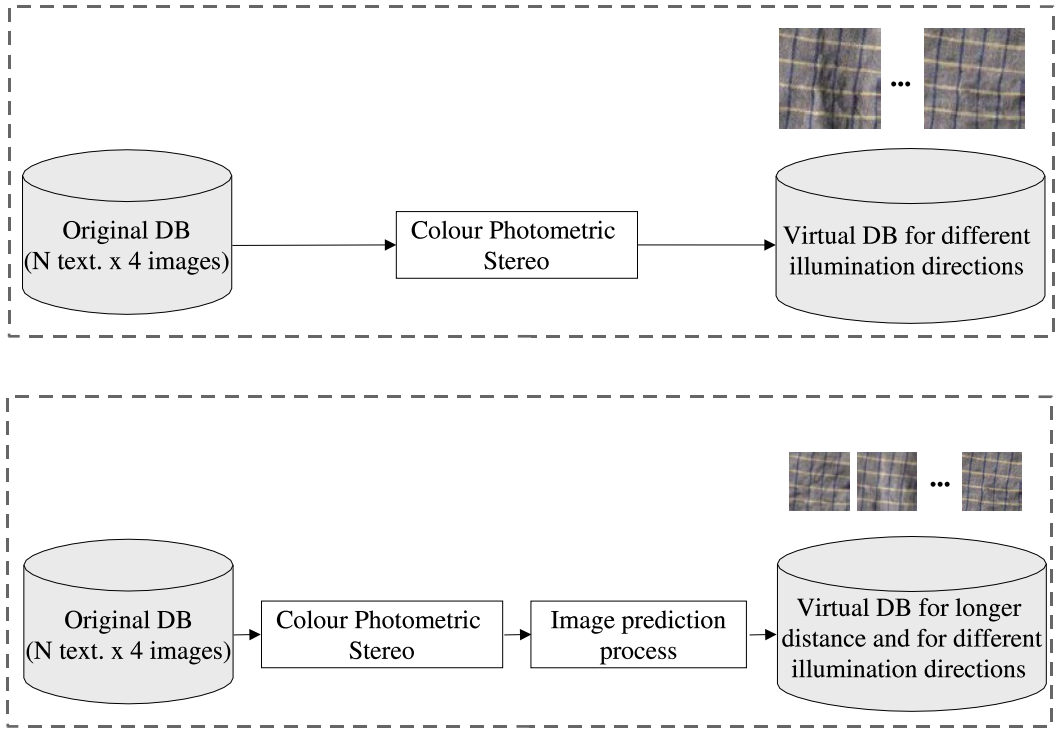


Figure 5.4: Virtual database creation. This can be performed in two different ways: creation of the virtual database for test images seen from the same distance as the training images and creation of the virtual database for test images seen from a longer distance than that of the original images.

the same distance as the originals, the creation of the virtual database is straightforward: the use of surface texture information extracted by photometric stereo allows us to directly obtain a rendering of the surface texture (see Section 3.5.4.1). However, things are less straightforward when the test images have been captured from a longer (but known or hypothesised) distance than the originals. In this case, in order to create the virtual database, we use the method presented in Section 4.3.1, which allows us to predict the appearance of a surface texture as seen from a longer distance and for various directions of light. Hence, we deal with the problem of texture recognition under varying geometries such as camera distance and light direction.

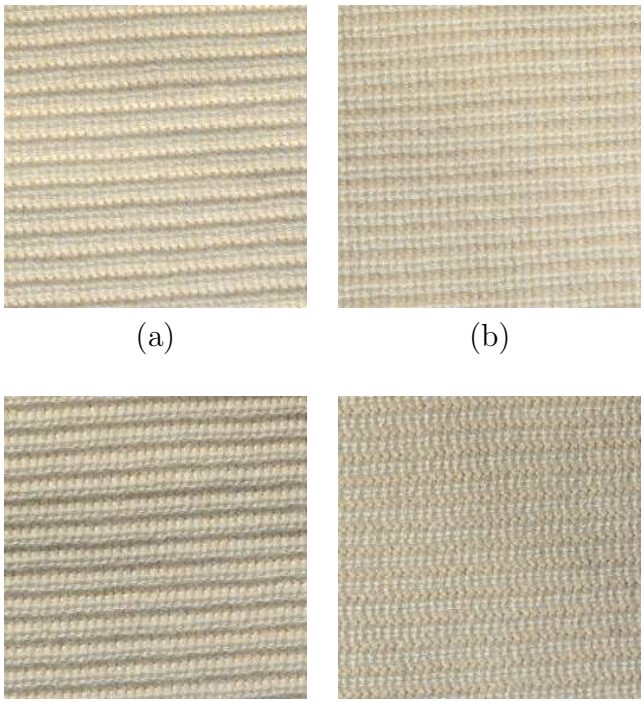
Figure 5.4 shows these two ways of performing the virtual database creation. Note that in both cases, 4-source colour photometric stereo is used to compute the detailed shape and colour of a surface which is used to create the desired images.

All the photometric stereo sets are constructed from 4 images illuminated at a fixed elevation angle and with 4 different tilt angles, 0°, 90°, 180° and 270°. This is the original database illustrated in Figure 5.4. Note that in the first case, the virtual database is created directly from the photometric information, while in the second case, the prediction method is needed. In that case, the virtual database contains a set of texture images imaged from a longer distance and under different directions of light. This will be the basis of dealing with the recognition of textures seen from a longer and known or hypothesised distance.

As it will be explained in more detail in the experimental section, the virtual database will contain 4 representative images for each texture class. Each image will correspond to a different direction of light from that used for training and testing. Note that reference and training images could be the same, however we want to add robustness to the features we will extract, on the basis that the test images are bound to be of different illumination than the reference images. It is important to remember that another objective of my work, apart from performing texture classification, is to provide an estimation of the light direction used to capture the test images. Specifically, we focus on the tilt angle estimation, assuming that all the images have been captured with the same elevation angle. With the idea of performing an accurate evaluation of the illuminant tilt angle estimation, we will create the images of the virtual database under 4 illuminant tilt angles which will be different from those used for the training and test images. This way, we will ensure that there is no correspondence between the light directions in the images in the virtual databases and the training and test images. Figure 5.5 shows the 4 images created in the virtual database for one texture.

5.3.2 Recognition Procedure

This section details each step in the recognition procedure. Basically, as it is shown in Figure 5.6, our recognition procedure is divided into two main phases: the learning process and the classification process. In the learning process, each texture class presented in the virtual database is modelled by a texture feature vector. In order



(a) (b)

(c) (d)

Figure 5.5: Example of images created in the virtual database. 4 images corresponding to 4 different illuminant tilt angles: 10°, 100°, 190° and 280°, (a)-(d) respectively.

to carry out this learning task, it is necessary to perform first a feature extraction process in which different texture features are computed for each image texture. Afterwards, a feature selection and evaluation process will allow us to choose from all the computed features those which could discriminate between the different classes best. When the learning process is finished, the classification process starts. This extracts the feature vectors for the unknown image textures (test images) and assigns them, by means of a classifier, to one of the classes in the virtual database, at the same time estimating the light direction in the test images.

In the remainder of this section, we describe how feature extraction, feature selection, feature evaluation and classification can be performed. Moreover, we detail the solution adopted for each of these processes.

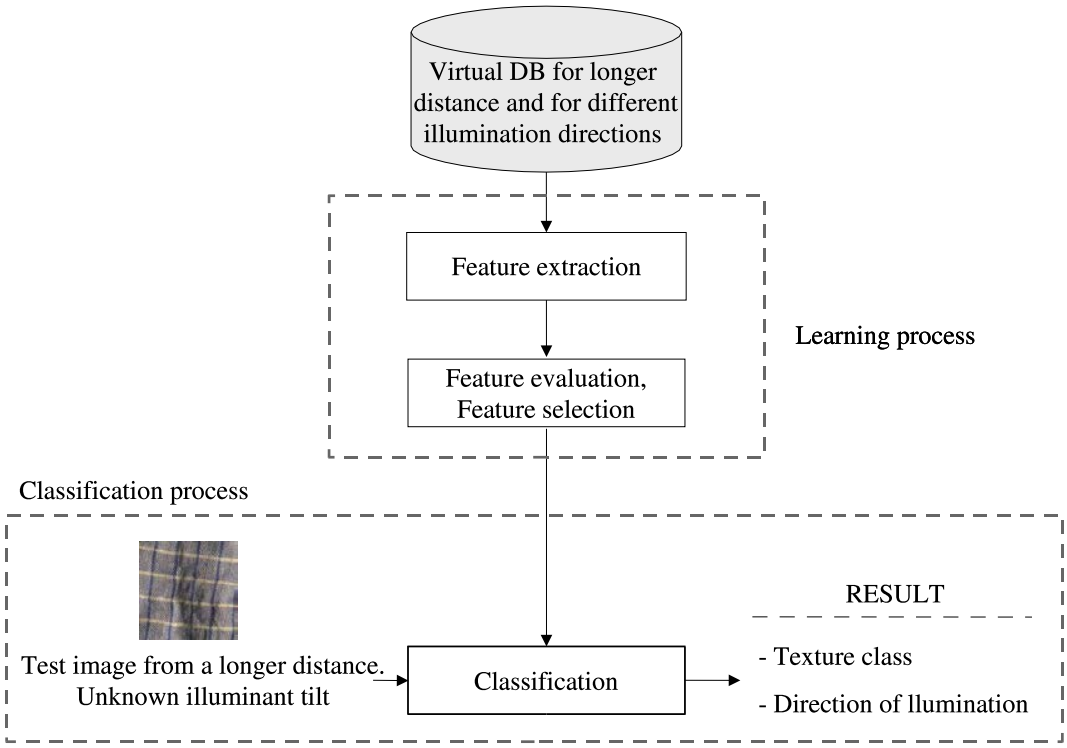


Figure 5.6: Recognition procedure. This procedure is divided into two main phases: learning process and classification process.

5.3.2.1 Texture Feature Extraction

Haralick provided a classic survey of texture measures [60]. Moreover, many types of representations and texture features have been proposed in the past few decades. Essentially, two major categories of texture measure methods have been identified: structural and statistical. This same basic classification was later adopted by other authors like Wechsler [168], Van Gool et al. [163], Grau [57], and Al-Janobi [4].

In structural methods, texture is considered as the repetition of some basic primitive patterns with certain rules of placement which are formally defined by grammars of various types. Nevertheless, since natural textures are not very regular, structural techniques are not very popular as argued by Wang and Liu [165]. On the other hand, statistical methods are based on the characterisation of the stochastic properties of the spatial distribution of grey levels in the image. Statistical approaches include a wide number of methods to extract texture measures. For example:

- **Laws's texture energy filters.** With the main desire of producing a computationally efficient method, Laws developed a coherent set of texture energy masks [90]. These masks allow us to achieve texture characterisation in two steps: first, the image is convolved with a set of small sized masks, second, statistics are computed from the outputs of these convolutions.

- **Random field models.** It is well-known that a number of random field models have been used for modelling and synthesising texture. If a model proves to be capable of representing and synthesising a range of textures, then its parameters may provide a suitable feature set for the classification of textures. Popular random field models include fractals, autoregressive models and Markov random fields. An extensive review of these approaches can be found in [139].

- **Frequency domain methods.** Some authors argue that many naturally occurring textures exhibit a combination of regularity such as approximate periodicity and variation which is hard to model using straightforward repetition or traditional statistical techniques. Hence, features related to the local spectrum have been proposed in the literature and used for the purpose of texture classification and/or segmentation. In most of these studies the relationship with the local spectrum is established through features obtained by filtering with a set of the two-dimensional Gabor filters to highlight frequency bands of two-dimensional spectra. This filter is linear and local and is characterised by a preferred orientation and spatial frequency, which cover the spatial frequency domain adequately. Roughly speaking, it acts as a local band-pass filter with certain optimal joint localisation properties in both the spatial domain and the spatial frequency domain [38].

- **Co-occurrence matrix.** The grey-level co-occurrence matrices (GLCM) are essentially two-dimensional histograms of the occurrence of pairs of grey-levels for a given displacement vector. Formally, the co-occurrence of grey levels can be specified as a matrix of relative frequencies P_{ij}, in which two pixels separated by a distance d along direction θ are found in the image, one with gray

level i and the other with gray level j. These GLCMs depend on the angular relationship between neighbouring pixels as well as on the distance between them. GLCMs are not generally used as features, rather a large number of textural features derived from the matrix have been proposed starting with the original fourteen features described by Haralick et al. [61].

Many researchers have attempted to carry out comparative studies to evaluate the performance of textural features. Weszka et al. [169] compared features derived from GLCMs on terrain images and found that co-occurrence features obtained the best result. Moreover, a theoretical comparison of four types of texture measures which Weszka et al. investigated was reported by Conners and Harlow [31]. They measured the amount of texture context information contained in the intermediate matrices of each algorithm and their conclusions were similar to those obtained by Weszka et al. Focusing on frequency domain methods, Pichler et al. [132] compared wavelet transforms with adaptive Gabor filtering feature extraction and reported superior results using the Gabor technique, although its higher computational cost was a drawback. Recently, Singh and Sharma [149] compared five different texture analysis methods in terms of recognition ability. They performed the experiments on the image benchmarks Meastex [1] and Vistex [3], concluding that features extracted from co-occurrence matrices are the best.

Summarising, the results of comparing the relative merits of the different types of features have been inconclusive and not a single method has emerged as being acceptable in all cases [138]. Comparative works have resulted in different, and sometimes contradictory, conclusions. The reason can be found in the use of different test images and evaluation or classification methods as well as some aspects related to code implementation.

It is important to note that in a previous work related to illuminant invariant texture classification, Penirschke et al. [125] developed a classification scheme based on the use of photometric stereo for the detection of surface relief and the use of Gabor features as texture measures. The production of features by filtering, however, requires the use of all points of the surface/image. The surface gradient of some of

these points may have been wrongly calculated by the photometric stereo technique and their inclusion in the feature extraction process may affect the performance of the classifier. That is why in our approach, we do not use filtering but co-occurrence matrices which allow us to work only with the pixels for which reliable information is available. In addition to this advantage, the co-occurrence matrix is an intuitive measure of texture and it is straightforward to compute. For all of these reasons it has been chosen in our recognition system as a feature extraction method.

Co-occurrence Matrix

As noted above, the co-occurrence matrix can be described as a matrix conveniently representing the occurrences of pairs of grey levels of pixels separated by a certain distance d and lying in a certain direction (angle θ) in the image.

Suppose that two grey values are separated by distance d along angle θ in an image which has m gray values. Typically, the angles are quantised in intervals of 45°: horizontal, first diagonal, vertical, and second diagonal (0°, 45°, 90°, and 135° respectively). Also, suppose that these grey values occur in this configuration of d and θ, N times in the image. The co-occurrence matrix is therefore the $m \times m$ matrix containing the number of occurrences N for each combination of grey values in the image. Note that a new co-occurrence matrix can be created for each different choice of d and θ.

Let us illustrate this with a simple example. Using the 4×4 image of Figure 5.7, we compute the co-occurrence matrices $P_{0°,1}$ and $P_{135°,1}$. Note that the subscripts indicate angle θ and distance d, respectively. As shown in the figure, the image has only 4 distinct grey levels, so the co-occurrence matrix is 4×4. For every horizontal grey level pair (horizontal in the first case in which we are computing $\theta = 0$) the number of instances of this pair in the entire image is counted up and then recorded in the co-occurrence matrix at the index corresponding to the two grey levels. Therefore, the number 6 at $P_{0°,1}(2,2)$ means that occurence $(2,2)$ is found 6 times in the original image. The elements of the co-occurrence matrix are usually normalised by dividing each entry by the total number of pixel pairs. This

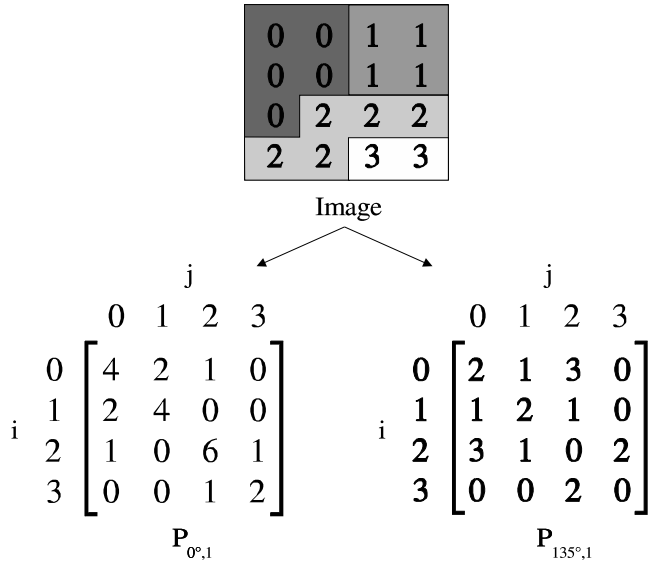


Figure 5.7: Example of co-occurrence matrix computation. The co-occurrence matrices at distance 1 and angle 0° and 135° are computed over the original image.

way, all values are between 0 and 1 and may be thought of as probabilities.

In general, a smooth texture gives a co-occurrence matrix with high values along the diagonals if d is small compared with the texture spatial variation (distance and angle). This is because the pairs of points at distance d should have similar grey levels. Conversely, if the texture is macrotexture and distance d is comparable to the texture scale, then the grey levels of points separated by distance d should often be quite different, so the values in the co-occurrence matrix should be spread out relatively uniformly.

Many researchers used statistics based on the co-occurrence matrix proposed by Haralick et al. [61] in their work [31, 46, 169]. In our approach, we will also use three of these tipical texture features shown in Table 5.1.

The main drawback of the GLCM technique is the dependence on the parameters used. The number of matrices needed to obtain good texture features is related to the angle and distance used. This number can be potentially enormous. As it will be seen in the experimental section, the co-ocurrence matrices will be used to extract features as contrast, homogeneity and energy for 20 different values of distance d

Table 5.1: Some of the most known texture measures computed from the co-occurrence matrix. $P_{\theta,d}(i,j)$ is the probability that grey level j follows grey level i at distance d and angle θ.

Contrast= $\sum_{i,j}(i-j)^2 P_{\theta,d}(i,j)$	This is the measure of the amount of local variation in the image. A low value results from uniform images whereas images with greater variation produce a high value.
Homogeneity= $\sum_{i,j} \frac{P_{\theta,d}(i,j)}{1+\|i-j\|}$	This is high when the GLCM concentrates along the diagonal. This occurs when the image is locally homogeneous.
Energy= $\sum_{i,j}[P_{\theta,d}(i,j)]^2$	This is also known as Uniformity or Angular second moment. It is high when the GLCM has few entries of large magnitude and low when all entries are almost equal.

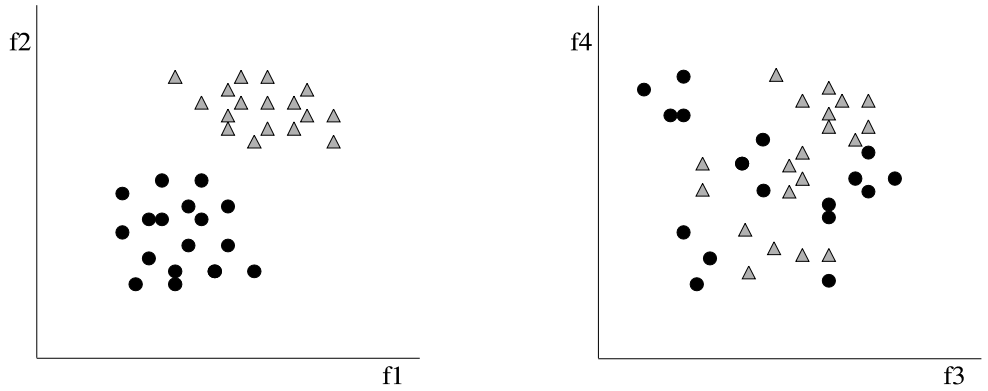
and for 4 different angles: 0°, 45°, 90°, 135°. This means that the co-ocurrence matrices are implemented in an anisotropic way. Note also that we compute in all 240 texture features (3 features × 4 directions × 20 distances).

Obviously, this is a large number of features to be used for a classification process. Moreover, some of these features can introduce redundant information, implying a decrease in classification accuracy. Therefore, we propose from all of these features, to select those which could best discriminate between the different classes.

5.3.2.2 Feature Selection

The aim of this section is to introduce the feature selection process as well as summarise the most relevant methods.

The problem of choosing the best features to discriminate between the different classes is known as feature selection. In the last few decades, this has been analysed from various points of view: statistics [47, 115], pattern recognition [39, 82, 75] and machine learning [16, 86]. These areas define the problem of feature selection as follows: given a set of d features, select a subset of size m (where $m < d$)



(a) (b)

Figure 5.8: Illustrative example of feature spaces. (a) Feature space in which the samples of each class are clearly separate. (b) Feature space in which the samples of each class overlap.

which obtains the highest value of a criterion function, assuming that a higher value indicates a better feature subset. In other words, choosing those features which allow feature vectors belonging to different classes to occupy compact and disjoint regions in a d-dimensional feature space. Figure 5.8 shows two examples of feature spaces. Note that the feature space in Figure 5.8.(a), composed of features f1 and f2, is preferred over that in Figure 5.8.(b), composed of features f3 and f4. The reason is simple: the first feature space has more discriminative power than the second in which the samples of the two classes overlap.

The simplest way of finding the best feature set is to perform an exhaustive search. However, this may be too costly and practically prohibitive even for a medium sized-feature set size. In order to solve this problem, different methods have been proposed in the literature in an attempt to reduce computational complexity by compromising performance. Obviously, these methods affect how well the patterns fit reality, and these become a central issue in the learning process. The importance of feature selection in a broader sense is due to its potential for speeding up the processes of learning and classifying objects, reducing the cost of recognition and, in some cases, improving the quality of classification.

Many authors have proposed different classifications of feature selection methods.

For instance, Siedlecki and Sklandsky [146] discussed the evolution of feature selection methods and grouped them into past, present, and future categories. Doak [40] identified three categories of search algorithms: exponential, sequential and randomised. Jain and Zongker [75] presented a taxonomy of available feature selection algorithms, dividing them into those based on statistical pattern recognition techniques and those using artificial neural networks. In a more recent survey, Kudo and Sklansky [88] focused on a comparative study of algorithms for large scale feature selection.

In what follows, we briefly review the best known feature selection methods, classifying them into three broad categories according to the generation procedure: those which perform a complete search, heuristic search, or a stochastic search.

- **Complete search.** The first algorithm which can be used to carry out a complete search and find an optimal subset is the exhaustive search. This consists of exploring all possible feature subsets (feature combinations) and selecting the best one. However, it has exponential complexity in the number of features and is frequently prohibitively expensive to use.

  Some heuristic algorithms, which introduce backtracking in the search, are used to reduce this exhaustive search without jeopardising the chances of finding the optimal subset. For instance, the Branch and Bound (B&B) algorithm, proposed by Narendra and Fukunaga [117], can be used to find the optimal subset of features much more quickly than through an exhaustive search. However, the main drawback of this algorithm is the requirement of a criterion function which has to be monotone. This means that the performance of a feature subset should improve whenever a feature is added to it. Furthermore, this algorithm is still impractical for problems with very large feature sets because the worst case complexity of this algorithm is exponential.

- **Heuristic search.** In each iteration of this generation procedure, all remaining features yet to be selected or rejected are considered for selection or rejection. There are many variations of this simple process and all of them

find a suboptimal solution. The search space order is, in general, polynomial, so these procedures are very simple to implement and very fast.

In this generation procedure we find popular methods such as Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS) [81]. These methods begin with a single solution and iteratively add or remove features until some termination criterion is achieved. These are the most commonly used methods for performing feature selection. SFS, also called the "bottom-up" method, starts from the empty set and generates a new subset in each iteration by adding the feature which best improve the quality of the selected subset (that is measured by some evaluation function). SBS, also called the "top-down" method, starts from the complete feature set and generates a new subset in each iteration by discarding a feature selected by some evaluation function. Both methods suffer from the so-called "nesting effect". This means that in the case of the "top-down" search, the discarded features can not be reselected, while in the case of the "bottom-up" search, the features once selected can not be later discarded. Note that since they do not examine all possible subsets, these algorithms are not guaranteed to produce the optimal result. With the aim of preventing this problem, Stearns [155] developed the "Plus l-take away r" (PTA(l,r)) search method. This method goes forward l stages by adding l features by SFS and goes backward r stages by deleting r features by SBS and repeating this process. The main drawback of this method is that there is no theoretical way of predicting the values of l and r to achieve the best feature subset. Research in this direction was concluded by introducing the generalisation of SFS and SBS (GSFS/GSBS) and the GPTA(l-r) algorithms proposed by Kittler [81]. Pudil et al. [134, 135] updated this study by introducing the two "floating" selection methods, SFSS and SFBS. Besides avoiding the nesting of features, one of their distinctive characteristics is that, during the backtracking process, the values of the criterion function are always compared only with those related to the same cardinality of the feature subset. Somol et al. [154] presented a more sophisticated version of "classical" floating search algorithms (adaptative floating methods) which attempt to remove

some of their potential deficiencies and facilitate the finding of a solution even closer to the optimal one. The adaptative floating search is called "adaptive" because of its ability to adjust the limit under which the actual generalisation level can be automatically set.

- **Stochastic search.** This search procedure is based on a random generation. Although the search space is exponential, these methods typically search fewer subsets by setting a maximum number of possible iterations.

  The best known method is based on the use of a genetic algorithm (GA). In 1989 Siedlecki and Sklansky introduced the use of GA for feature selection [147]. Since then, many works have used this strategy to perform feature selection [160, 55, 175, 161, 145]. In GA, a feature subset is represented by a binary string with length n, called a *chromosome*, with a zero or one in position i denoting the absence or presence of feature i. Each chromosome is evaluated for fitness through an optimisation function in order to survive to the next generation. Unlike classical hill-climbers, it does not evaluate and improve a single solution but, instead, analyses and modifies a "population" of solutions at the same time. The optimisation process is carried out in cycles called generations. During each generation, the population of chromosomes is maintained and evolved by two operators: (i) crossover: where parts of two different parent chromosomes are mixed to create an offspring, and (ii) mutation: where the bits of a single parent are randomly perturbed to create a child, increasing the variability of the population. In each generation only a few of the best chromosomes survive to the next cycle of reproduction. When the maximum number of cycles is achieved, a set of feature subsets is provided. The main drawback of this feature selection method is that it needs a proper assignment of values of different parameters: initial population size, crossover rate and mutation rate.

We have seen that quite a few methods have been proposed for feature selection. However, choosing the proper method for a particular problem is a difficult task, as Pudil and Novovicová argued in [136]. The best choice depends on a large number of

conditions related to the user's knowledge of the problem. The following questions describe perhaps the most important aspects we must keep in mind when choosing a feature selection method: What exactly is your aim? What is the dimensionality of your problem? What is your a priori knowledge? Do you have a criterion for subset evaluation appropriate to your knowledge of the problem? In Table 5.2, we provide recommendations for each feature selection method with the aim of facilitating the election of one of them for a particular problem.

As it will be seen in the experimental section, we will use a heuristic search algorithm in order to perform our feature selection process. Basically, the reason for this choice can be found in the reduction of the search cost to that of polynomial complexity. Moreover, the heuristic search implies a fast generation procedure and, as it has been mentioned, a simple implementation.

5.3.2.3 Feature Evaluation

The process of choosing an appropriate criterion function is known as feature evaluation. The main goal of this process is to measure the "goodness" of a subset produced by some generation procedure. In this context, goodness means the capability of a feature subset to distinguish the different class labels and the ability of providing compact and maximally distinct descriptions for every class. Geometrically, this constraint can be interpreted to mean that this feature takes on (i) nearly identical values for all samples in the same class and (ii) different values for all samples of the other classes. This is illustrated in Figure 5.8.

Feature evaluation has been studied for many years and different measures have been proposed. Ben-Bassat [12] grouped the existing criterion functions up to 1982 into three categories: information measures, distance measures, and dependence measures. Another viewpoint was introduced by Langley [123], who grouped different feature selection methods into two broad groups, *wrapper* and *filter*, based on their dependence on the feature evaluation used.

Table 5.2: Feature selection methods.

Method	Description	Comments
Exhaustive Search	Evaluates all possible subsets.	Guaranteed to find the optimal subset; not feasible for even moderately large values of m and d.
Branch and Bound Search (B&B)	Only a fraction of all possible feature subsets need to be enumerated to find the optimal subset.	Guaranteed to find the optimal subset provided the criterion function satisfies the monotonicity property; the worst case complexity of this algorithm decision is exponential.
Sequential Forward Selection (SFS)	Selects the best single feature and then add one feature at a time which in combination with the selected features maximizes the criterion function.	Once a feature is retained, it can not be discarded; computationally attractive; it examines only $(d-1)$ possible subsets.
Sequential Backward Selection (SBS)	Starts with all the d features and successively deletes one feature at a time.	Once a feature is selected, it cannot be brought back into the optimal subset; requires more computations than sequential forward selection.
"Plus l-take away r" PTA(l,r) Selection	First enlarge the feature subset by l features using forward selection and then delete r features using backward selection.	Avoids the problem of feature subset "nesting" encountered in SFS and SBS methods; need to select values of l and $r(l > r)$.
Generalized Sequential Forward Selection (GSFS) and Generalized Sequential Backward Selection (GSBS)	SFS and SBS are generalized in such a way that a number (n) of features are evaluated at the same time and the best n features subset is chosen for addition or deletion.	Their performance is a little better than SFS and SBS; they are effective when the number of features is very small; they are very time consuming.
Sequential Forward Floating Search (SFFS) and Sequential Backward Floating Search (SBFS)	A generalization of "plus-l take away-r" methods; the values of l and r are determined automatically and updated dynamically.	Provides close to optimal solution at an affordable computational cost for small-scale and medium-scale problems; in order to avoid excessive computation, the maximum level of backtracking could be constrained.
Adaptive Sequential Forward Floating Search (ASFFS) and Adaptive Sequential Backward Floating Search (ASBFS)	They are called adaptive because of their ability to adjust the limit under which the actual generalization level can be automatically set.	Adaptive Sequential Floating Search yields better results than classical Sequential Floating Search; eliminates the lack of an explicitly specified termination condition.
Genetic Algorithms (GA)	A feature subset is represented by a binary string with length n, called a chromosome, with a zero or one in position i denoting the absence or presence of feature i; a population of chromosomes is maintained and evolved by two operators of crossover and mutation.	It is very useful to find a compromise between maximum criterion value and minimum size subset in large scale; Needs proper assignment of values to different parameters: iterations, initial population size, crossover rate, and mutation rate.

- The idea behind the wrapper approach is simple: a classifier is used as a criterion function in order to obtain a metric (classification accuracy) for guiding the feature subset selection.
- On the other hand, the filter approach attempts to assess the merits of features from the data alone and the selection is performed independently of the classifier.

The filter approach normally uses a criterion function which is simple and fast to compute, so it is generally computationally more efficient. Its major drawback is that an optimal selection of features may not be independent of the classifier used. Moreover, the wrapper approach provides a better estimate of accuracy for a feature subset, but involves a computational overhead by executing the classifier.

A recent research made by Dash and Liu [37] considering the latest developments, divided the evaluation functions into five categories: distance, information (or uncertainty), dependence, consistency and classifier error rate.

- The first category called distance, is also known as separability, divergence, or discrimination measure, where, for a two class problem, a feature A is preferred to another B if A induces a greater difference between both classes. If the difference is zero, it is impossible to distinguish the two classes. An example is the Euclidean distance measure.

- The information (or uncertainty) measures determine the information gained from a feature. Therefore, one feature is preferred to another if the information gain is greater. An example of this category is the entropy measure.

- The dependence measures or correlation measures qualify the ability of predicting the value of one variable from the value of another. The coefficient is a classical dependence measure and can be used to find the correlation between a feature and a class. If the correlation of feature A with a class is higher than the correlation of feature B, then feature A is preferred to B. A slight variation of this is to determine the dependence of a feature on other features. This value indicates the degree of redundancy of the feature.

- The consistency measures find the smallest sized subset which satisfies the acceptable inconsistency rate set by the user. This rate is usually calculated as the sum of inconsistency counts, divided by the total number of samples. Two samples are considered inconsistent if their attribute values are the same (they match).

Table 5.3: A comparison of evaluation functions provided by Dash and Liu [37].

Evaluation function	Generality	Time complexity	Accuracy
Distance measure	Yes	Low	–
Information measure	Yes	Low	–
Dependence measure	Yes	Low	–
Consistency measure	Yes	Moderate	–
Classifier error rate	No	High	Very High

- The last category is the classifier error rate. This corresponds with the idea of wrapper methods in which the accuracy of the classifier is used as an evaluation measure.

Obviously, an optimal subset in feature selection is always relative to a certain criterion function. Hence, choosing an optimal subset using one criterion function may not be the same as using another one. Therefore, a good evaluation function is a key factor in choosing the best feature subset for a recognition process. In their work Dash and Liu compared the evaluation functions using different properties; (*i*) generality: how suitable is the selected subset for different classifiers, (*ii*) time complexity: time taken for selecting the subset of features, and (*iii*) accuracy: how accurate is the prediction using the selected subset. Table 5.3 shows a comparison of evaluation functions realised by Dash and Liu. The "–" in the last column means that nothing can be concluded about the accuracy of the corresponding evaluation function. In all of these cases, the accuracy of the evaluation functions depends on the data set and the classifier used. They discovered an unsurprising trend, the more time spent, the higher the accuracy. The table also shows us which measure should be used under different circumstances such as with time constraints, when given several classifiers to choose from, etc.

Classification error rate and distance measures are usually chosen as a criterion function in many works. It is difficult to estimate the correct recognition rate of a classifier on the basis of a limited number of training samples. This is one reason why most comparative studies on feature selection used distance as a criterion function.

Nevertheless, a recent study made by Kudo and Sklansky [88] compares feature selection algorithms using a recognition rate measure. On the other hand, if it is known which classifier will be used in the problem under consideration, the best criterion is, in general, the classification error rate because it makes feature selection procedures specific for the classifier and the sizes of the training and test sets used. As it will be seen in the experimental section, the classification error rate will be used as a feature evaluation in our feature selection process.

5.3.2.4 Classifiers

Up to now, we have described the learning process of our recognition system. We have seen that each texture class presented in the virtual database is modelled by a texture feature vector extracted from the co-occurrence matrices. The necessity of performing feature selection/evaluation in order to choose the best feature subset which distinguishes the different classes best has also been stated. At this point, when the learning process is finished, the classification process proper begins. The objective of this process is to assign, by means of a classifier, the feature vectors of the unknown image textures (test images) into one of the classes of the virtual database.

It is well-known that a classifier can be designed using a large variety of possible approaches. In practice, the choice of a classifier is a difficult problem and is often based on which classifier happens to be available or best known to the user. As Jain et al. presented in [73], there are three different approaches to design a classifier.

- The simplest and most intuitive approach to classifier design is based on the concept of similarity: patterns which are similar should be assigned to the same class. Therefore, once a good metric has been established to define similarity, patterns can be classified by template matching or the minimum distance classifier using a few prototypes per class. The choice of metric and prototypes is crucial to the success of this approach.

- The second main concept used for designing pattern classifiers is based on the probabilistic approach. These classifiers are commonly known as Bayesian

classifiers. They are based on probabilistic information on the populations from which a sample of training data is to be randomly drawn. Randomness in sampling is assumed and it is necessary for a better representation of the sample of the underlying population probability function. There are different approaches to Bayesian classifiers. See [114] for a survey of these classifiers. Among them, the optimal Bayes decision rule, used in many works, assigns the unknown patterns to the class with the maximum posterior probability, taking the class conditional probabilities and the a priori probabilities into account.

- The third category of classifiers is to construct decision boundaries. A classical example of this type of classifier is the linear discriminant analysis (LDA) [114] which uses linear boundaries between data distributions to discriminate between samples. Another example is the single-layer perceptron where the separating hyperplane is iteratively updated as a function of the distance of the misclassified patterns from the hyperplane. It is important to note that neural networks themselves can lead to many different classifiers depending on how they are trained.

  A decision tree is a special type of classifier which is trained by an iterative selection of individual features which are most salient at each node of the tree. The criteria for feature selection and tree generation include the information content, the node purity or Fisher's criterion (a linear discrimination which minimises the MSE between the classifier output and the desired labels). During classification, only the features needed for the test pattern under consideration are used, so feature selection is implicitly built-in. The main advantage of the tree classifier, apart from its speed, is the possibility of interpreting the decision rule in terms of individual features.

Nearest Neighbour Classifier

Considering the first approach described above, we use the Nearest Neighbour classifier (NN) to perform our texture classification.

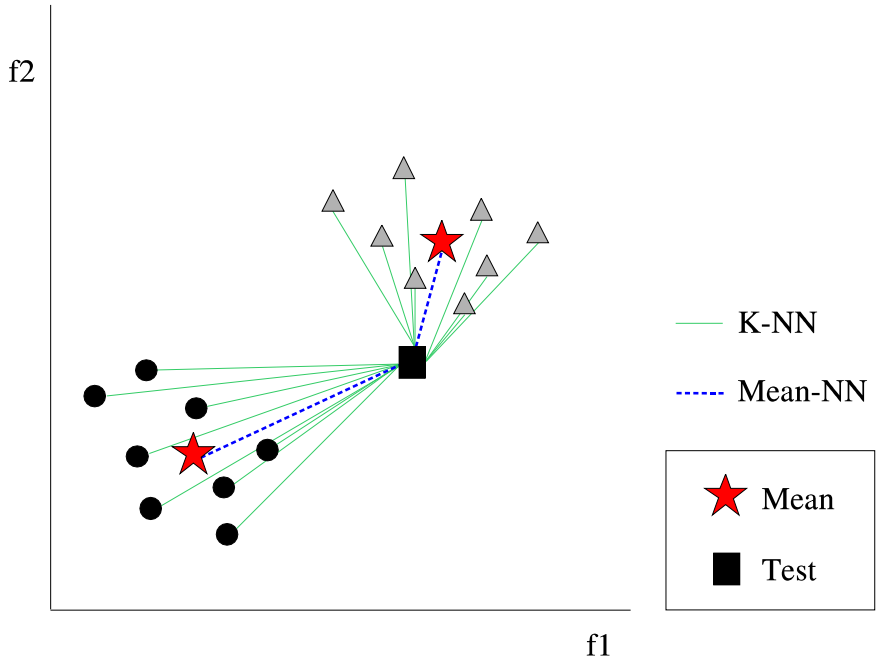


Figure 5.9: Example of K-nearest neighbour (K-NN) and mean nearest neighbour (mean-NN) classifier. Note that the test sample is classified in both cases to the same class.

Nearest neighbour methods have been used as an important pattern recognition tool [47, 89, 114]. The one-nearest neighbor (1-NN) classifier is the most natural classification method we can think of. It consists of comparing an unknown observation x with all the N cases in the training set. N distances between a pattern vector x and all the training patterns are calculated and the label information contained in the training set with which the minimum distance results is assigned to the incoming pattern x. The K-nearest neighbour (K-NN) is the same as the 1-NN rule except that the algorithm finds K nearest points within the points in the training set from the unknown observation x and assigns the class of the unknown observation to the majority class in the K points. Note that another criterion has to be defined in case of a tie. Another common approach is the nearest mean classifier (mean-NN) in which each class is represented by a single prototype (feature vector) which is the mean vector of all the training samples in that class. Figure 5.9 intuitively illustrates how the nearest neighbour classifiers (K-NN, Mean-NN) classify a test sample.

The nearest neighbour classifiers are simple and robust. The most straightforward one-nearest neighbour decision rule (1-NN) is conveniently used in many works as a benchmark for all the other classifiers since it appears to provide a reasonable classification performance in most applications. The 1-NN classifier only has to define the distance metric used to find the nearest neighbour. In general, Manhatan or Euclidian distance are commonly used. The main drawback of this classifier is the high computational cost.

In terms of typical texture classification the different approaches of nearest neighbour classifiers perform in a similar way. Nevertheless, when dealing with texture classification independent from light direction, things are different. Changes in the imaging geometry and lighting conditions can significantly alter the appearance of the surface, implying significant variations in the image texture and, therefore, in the feature vectors. This fact causes the data to suffer high intraclass variation. In other words, the feature values for the same texture class are not stable. Hence, the data distribution for each class does not make a compact cluster. It is this intraclass variation which introduces critical misclassification rates in the texture classification process. In that case, it is reasonable to think that the mean-NN is not an adequate approach since the variation of a texture can provide mean values which do not reflect the nature of the data (i.e. one texture with more than one cluster). Analysing these facts, in our experimental results we use a 1-NN classifier in which an unknown test image is compared with all the cases contained in the virtual database and is classified with the one which has minimum distance. At the same time it performs texture classification, this classifier allows us to provide an estimation of the illumination direction of the test images directly. It is important to remember that for each texture class, different images corresponding to different light directions are available in the virtual database. Therefore, using the 1-NN classifier, as well as assigning the test image to the texture class to which it belongs, the light direction of the unknown test image is approximated by one of those contained in the virtual database. This will be specified in more detail in the experimental section related to texture classification and light estimation.

5.4 Summary

A novel methodology of texture classification under varying imaging geometries (i.e. light direction and camera distance) was presented in this chapter. Basically, the main contributions of this method are: (i) to overcome the problem of classifier failure induced by varying imaging properties, (ii) to integrate the texture prediction framework proposed in Chapter 4 into a texture classification system, and (iii) to provide the capability of guessing an approximate direction of the light used to capture the test images.

This chapter has been structured in two blocks. In the first, works on the topic of texture classification under varying imaging geometries were reviewed. Moreover, we discussed the different ways of realising this texture classification system, concluding that a model-based approach is a potential solution to the problem. In the second block, our model-based approach for texture classification was presented, describing the general scheme and detailing each step individually.

Roughly speaking, the main purpose of our method has been to train a classifier on a selection of images of the same surface obtained under different imaging geometries. Hence, variability is modelled for each surface texture. The main reason for this choice is our interest in exploiting the capability of colour photometric stereo to render images of a surface under novel lighting conditions as well as the capability of the image prediction method proposed in Section 4.3.1 of creating images of a surface texture under novel imaging geometries such as camera distance.

The whole procedure of the proposed **model-based texture classification system** was divided into two main phases:

- **Virtual database creation.** The aim of this phase is to create, by means of the surface texture information and the image prediction method, a "virtual" database of image textures against which we compare the unknown test images in order to classify them.
- **Recognition procedure.** The recognition procedure was divided into two steps: the learning process and the classification process.

- The learning process has the goal of modelling each texture class by means of a representative feature vector.
- The classification process has the goal of classifying an unknown test image into the texture class it belongs.

In this chapter, we have seen how the learning process requires the application of other processes. First, the feature extraction process is needed to extract different texture features for each image in the virtual database. Different techniques to extract textural properties were reviewed and, from them, the **co-occurrence matrix** was chosen to compute the texture features used in our classification system. We also explained that it is very important to choose from all the computed features those which can distinguish the different classes best. For this reason, we analysed and discussed various methods which allow us to perform the feature selection and feature evaluation process.

When the learning process is finished the classification process starts, extracting first the feature vectors for the unknown image textures (test images) and assigning them by means of a classifier into one of the classes of the virtual database. We described how the **Nearest Neighbour classifier** is used to perform the texture classification and, at the same time, to provide the approximate direction of the light used to capture the test images.

In the next chapter, we will test and evaluate this texture classification approach over a large set of surface textures with different properties (i.e. smooth surfaces, rough surfaces and directional surfaces). Different experiments will be performed and, after discussing them, we will extract the corresponding conclusions.

Chapter 6

Experiments

The proposed prediction methods, as well as the model-based texture classification system, are tested and evaluated in this chapter. A set of real surface textures containing a wide variety of relatively smooth and very rough surfaces are used in this thesis as our image database (experimental data). Different experiments and error measures are used in order to carry out an exhaustive evaluation. The validity of the texture classification system is demonstrated by classifying texture images captured under imaging geometries different from the reference images in the database. The process of recognition allows us also to guess the approximate direction of the light used to capture the test images.

6.1 Introduction

The aim of this chapter is to assess the performance of the methodology presented in this thesis. As it has been seen in previous chapters, there are certain properties of the theory which must be accurately analysed. For instance, the ability of the colour photometric stereo technique to use images captured with a certain light direction in order to predict novel images referring to the same camera distance, but with different light directions. It is also important to analyse the abilities of the prediction methods presented in Section 4.3 in order to predict the surface/image information when seen from a longer distance. An exhaustive evaluation of these prediction methods is presented in this chapter. As it will be seen, different experiments and

error measures are used to extract conclusions and discuss them.

On the other hand, we must also evaluate the model-based texture classification system proposed in Section 5.3. Texture classification results obtained on a broad set of real surface textures are presented and analysed. Moreover, results related to the estimation of the direction of the light are also provided and discussed.

In order to obtain all these experimental results, it is necessary to define which experimental data will be used. It is well-known that there are many texture image databases available to perform and evaluate a typical texture classification system: the Brodatz album [18], the Meastex database [1] and the Vistex database [3]. Although there are various databases of images to test texture classificaton algorithms invariant to the direction of illumination, there is no database of images to test algorithms for changes of camera distance. For example, the "Photometric Image Databases" from the Texture lab at Heriot-Watt University [2] are not suitable to our purposes because they do not provide photometric sets of images captured from different distances. Moreover, the "Columbia-Utrecht database" established by Dana et al. [35] is also unsuitable because the light was held constant while the viewpoint and orientation of the samples were varied during data capture. Obviously, in order to perform the set of experiments which allows us to evaluate and extract conclusions for our proposals, an appropriate experimental data set is required. Therefore, acknowledging the lack of a texture database which accomplishes our requirements, we opted to build our own image database which provided all the information needed for our purposes.

The main contributions of this chapter can be summarised with the following points:

- The presentation of the image database of textured surfaces used in this thesis. This database contains the required images to apply the photometric stereo technique correctly, as well as those images which should be used as training and test images.

- The evaluation of results achieved by our prediction methods (direct image prediction and surface prediction), providing a comparison between them.

- The evaluation of results obtained on texture classification under varying geometries, such as light direction and camera distance.

The remainder of this chapter is structured in the following blocks. In the following section we describe the experimental data. In Section 6.3 and 6.4 all the experimental trials are analysed and discussed individually. Finally, conclusions are drawn in Section 6.5.

6.2 Experimental Data

Twenty five physical texture samples were used throughout the experimental trials presented in this thesis. Basically, we used real textures commonly found in our every day life. For example, simple textures based on different cloth, paper, etc, and other textures formed by repeating primitives of foods such as beans, lentils, spaghetti, chips, etc.

For each texture sample two photometric sets composed of 4 images each were available. The two sets were captured at two different distances; distance A and distance B (longer than A). All photometric stereo sets consisted of 4 images lit at an elevation angle of 55° with 4 different tilt angles 0°, 90°, 180° and 270°. Figure 6.1 illustrates the scheme of the platform used to capture the experimental data. In addition to the 4 images used in the photometric sets, different images for each surface taken at the two distances were captured for testing. We call these sets of images tA and tB. Thirteen surfaces were captured using 12 illuminant tilt angles between 0° and 360° incremented by steps of 30°. The remaining surfaces were captured using 24 illuminant tilt angles between 0° and 360° incremented by steps of 15°. All surfaces were lit at an elevation angle of 55°. It is important to state that the lights or the object position were not changed when the distance of the sensor was changed. For fifteen texture samples, in addition to these images, another set of images including the photometric images and the test images were captured at distance C (longer than B).

Virtual images constructed from photometric sets of the same distance are re-

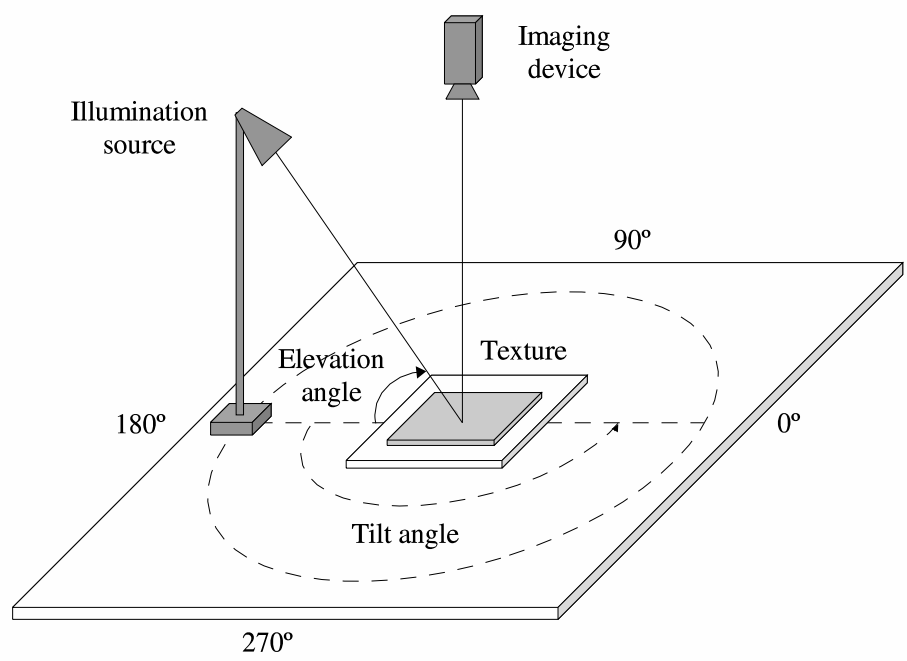


Figure 6.1: Scheme of the platform used to capture the experimental data. Different images for each texture were captured under different illuminant tilt angles and a fixed elevation angle of 55°.

ferred to as images AA, BB and CC. Virtual images constructed from photometric set A for distance B are referred to as predictions AB. In a similiar way, virtual images constructed from photometric set A for distance C are referred to as predictions AC. In order to distinguish between the direct image prediction method of Section 4.3.1 and the image prediction method via the surface prediction of Section 4.3.2, we shall refer to them as prediction $imaAB$ and $surAB$ respectively. Figure 6.2 summarises the notation described above.

Figure 6.3 shows one image for each surface texture captured from distance A and one from distance B. These images include two major groups of surface textures:

- One group of surfaces consists of a wide variety of relatively smooth surfaces which may be further divided into:
  - Isotropic surfaces and
  - Directional surfaces.

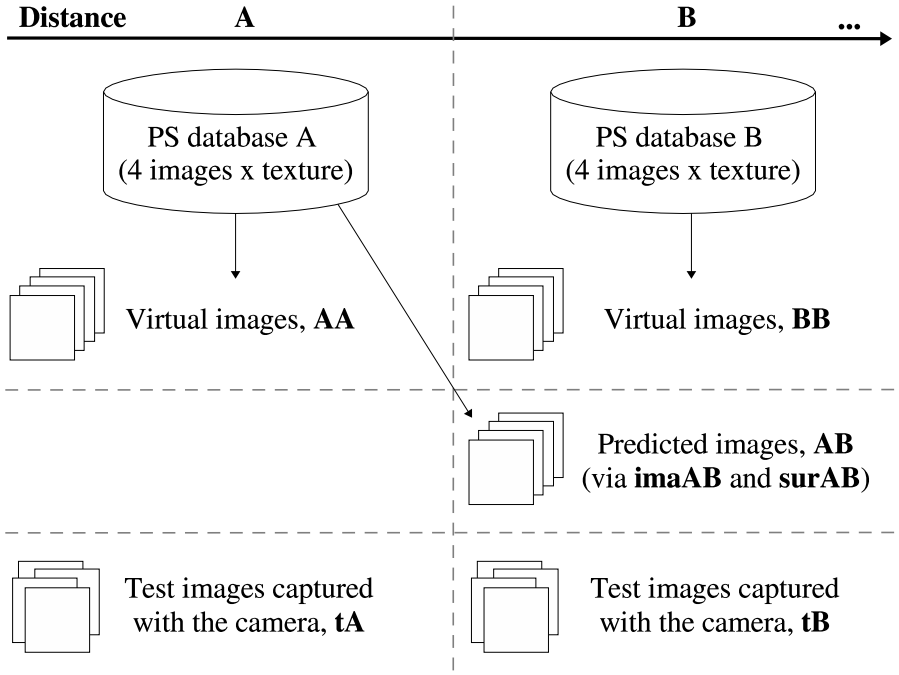


Figure 6.2: Virtual images constructed at the same distance are referred to as images AA and BB. Virtual images contructed from photometric set A with distance B are referred to as predictions AB. Images captured for testing at distances A and B are called tA and tB respectively.

- The other group of surfaces consist of a variety of very rough surfaces, for which the assumption on which photometric stereo is based is violated, (i.e. relatively smooth surface with low roughness). We do not expect CPS to work well for such surfaces, but we included them in order to test the proposed method to the extreme.

As it was seen in Section 2.3.2, the description of a surface can be stated in different ways. For example, a single parameter may be sufficient to characterise a surface for some purposes. This is the case of the *absolute average slope ratio* ($AASR$) which provides an easy way to characterise the degree of roughness of a given surface texture. Remember that $AASR$ is calculated as

$$AASR = \frac{1}{2NM} \sum_{x}^{N-1} \sum_{y}^{M-1} |p(x,y)| + |q(x,y)|$$

where $N \times M$ is the number of points for which the (p, q) values are known. For

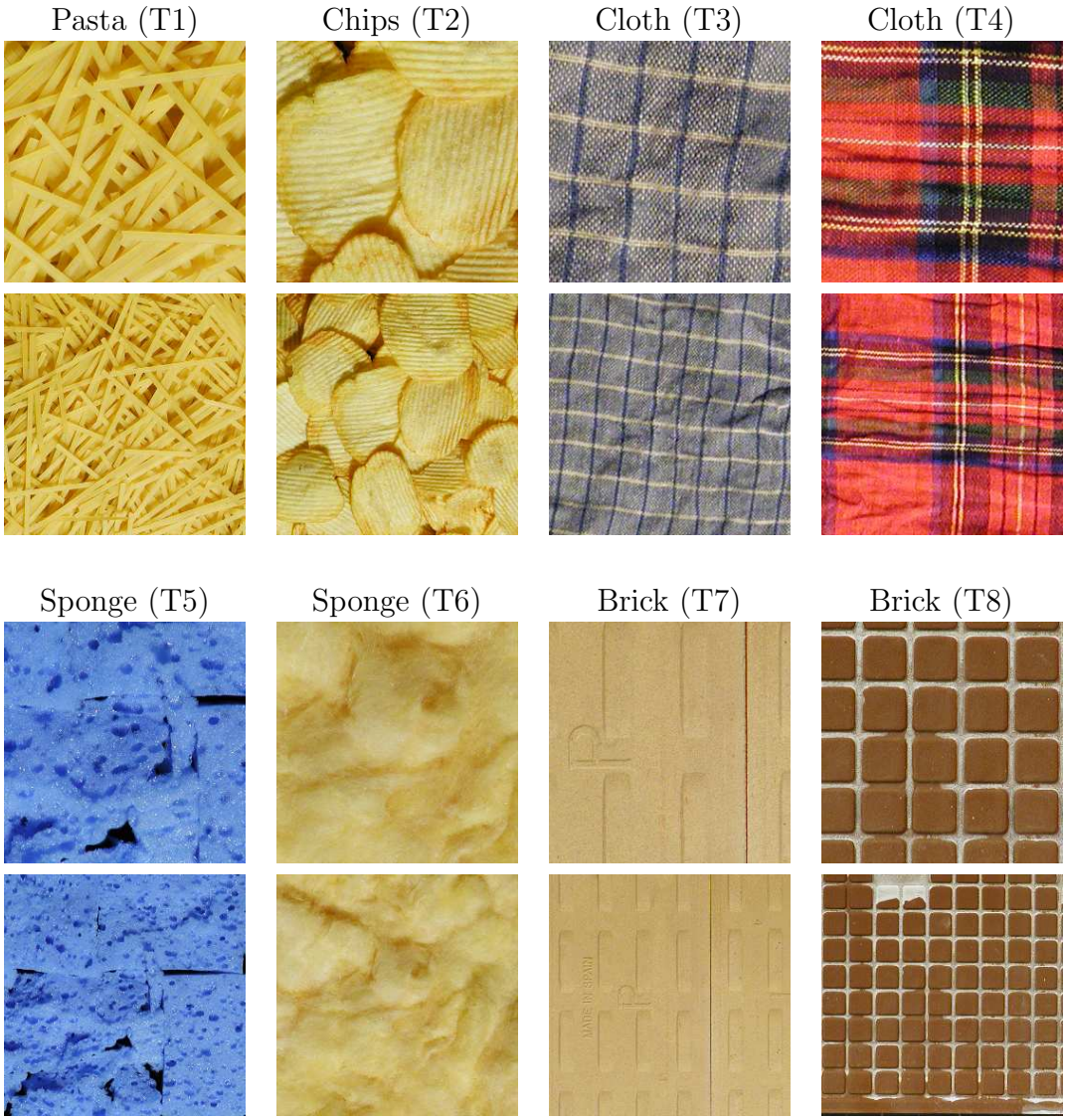


Figure 6.3: One image from distance A and one from distance B of each of the twenty five sample textures.

Figure 6.3: (continued).

Figure 6.3: (continued).

other purposes in which a more accurate description is required, statistical models such as the histograms of the values of the components of the gradient vectors may provide better descriptors. In this chapter we use both descriptors to characterise the surface shape and roughness of each texture, namely the $AASR$ parameter and the estimated probability density functions (PDFs) for the surface partial derivatives p and q (i.e. the normalised histograms of these quantities). The p-map and q-maps of the surface textures are estimated using CPS. The histograms are 256 point discrete approximations of the PDFs in the range $[-1, 1]$. These histograms represent statistical models of the surfaces. Moreover, as there is a linear relationship between surface gradient and surface height, the characteristics observed here are also valid for describing the surface height map.

After analysing all the PDFs, we identified three different types of gradient distribution corroborating the diversity of the dataset. For instance, consider three textures (T5, T14, and T15) which provide representative examples of each texture class defined above (see first column of Figure 6.4).

We observed that some surfaces, are essentially random textures as in the first example (T5), which was formed by a fracture process. Note that the PDFs for this texture could be modelled by Gaussian distributions (see first row of Figure 6.4). Similiar behaviour has been observed, for instance, for textures T1, T6, and T10. The second example, which corresponds to a directional texture, has one gradient component close to a typical distribution of a surface with a sinusoidal height profile (see the PDF of p in the second row of Figure 6.4). The q component concentrates all its values close to 0 with a maximum probability density of 0.1148. This distribution is shown out of scale in order to maintain the same scale for the y-axis so as to allow comparison between histograms. We observed that all the directional surfaces, such as textures T11, T12, and T13, show the same behaviour. Other textures (very rough isotropic surfaces) do not generally fit any particular distribution. They look like normal distributions which have been flattened out, presumably because the estimation is affected by severe shadowing, as shown in the third example of Figure 6.4. Other textures which follow this behaviour are textures T17, T18, T20, and T21.

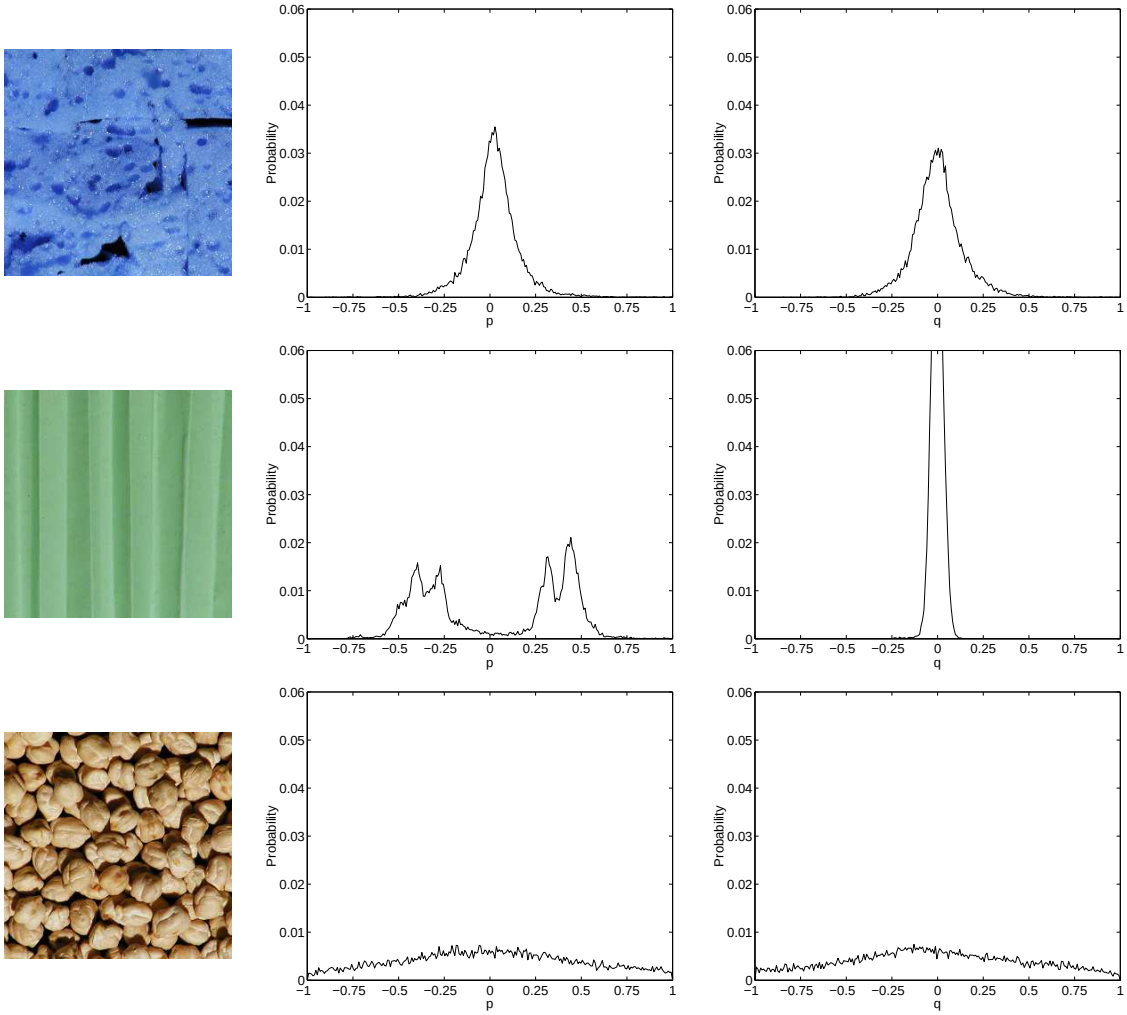


Figure 6.4: One representative image of an isotropic (T5), a directional (T14) and a very rough isotropic surface (T15). The second and third columns show the PDF representations of the surface gradients p and q.

Moreover, by analysing the $AASR$ parameter, we can also establish a simple classification scheme of surfaces based on their degree of roughness. Relatively smooth surfaces have small $AASR$ values. For example, textures T1, T5, and T12 have $AASR$ values of 0.0897, 0.1003, and 0.1129 respectively. On the other hand, very rough surfaces have large $AASR$ values. Textures T15, T20, and T22, have $AASR$ values of 0.3960, 0.3530, and 0.2470 respectively. Hence, this parameter allows us to distinguish between a smooth and a rough surface in a simple way.

Considering both surface descriptors, the $AASR$ and the estimated PDFs, we have classified our textures into one of the three groups described earlier: isotropic surfaces (from texture T1 to texture T10), directional surfaces (from texture T11 to texture T14), and very rough surfaces (from texture T15 to texture T25).

Another aspect we analysed is how chromatically distinct the surface textures are. To quantify this we used the Euclidean metric in conjunction with the Luv colour space. First, we calculate the average Luv colour value from each surface. Next, using these values, we find the distance between all pairs of surface textures, choosing the minimum value. This value indicates a measure of how distinct the surfaces are. We find that the minimum distance is 14.3471, corresponding to textures T1 and T11. We also observe that the texture most chromatically distinct from the others is texture T5, with a minimum distance value of 1273.2258.

6.3 Texture Prediction Results

After describing the experimental data, this section will focus on different experimental trials related to the texture prediction described in Chapter 4. All experiments performed have been to check various aspects of our methodology, namely:

1. To check the accuracy of the photometric stereo technique to use images captured with a certain light direction in order to predict images referring to the same camera distance but with different light directions. The results of this experiment allow us to demonstrate the validity of the photometric stereo technique in order to create virtual images of the same surface seen under different light directions.

2. To check the accuracy of *image prediction* using a photometric set captured at distance A to predict images captured at distance B ($>$ distance A). The predictions will be compared with real images captured at distance B and with images produced from a photometric set captured at the same distance B. The results of this experiment allow us to evaluate how well an *image* can be predicted from one resolution to another.

3. To check the accuracy of *surface shape prediction* using a photometric set captured at distance A to predict the surface as it would appear at the resolution of distance B and compare it with the surface reconstructed from a photometric set captured at distance B with the same light orientations as for the set at distance A. The results of this experiment are to be used as a benchmark of how well a *surface* can be predicted from one resolution to another under ideal conditions where the lighting does not change in orientation.

4. To check the accuracy of image prediction using photometric sets captured from distances A and B to predict images captured from distance C ($A < B < C$). The results of this experiment allow us to analyse the effect of using two different resolutions in the original sets to perform the same prediction when seen from distance C.

5. To check the accuracy of surface shape prediction using photometric sets captured from distances A and B to predict the surface as it will appear from the longer distance C. The results of this experiment are to be used as a benchmark of how well a surface shape can be predicted from different distances and from original sets with different resolution.

6.3.1 Experiment 1: Accuracy of Photometric Stereo

This experiment analyses the accuracy of the photometric stereo technique for creating virtual images referring to the same camera distance but under different light directions. In order to examine the performance of the photometric technique, we have decided to compare the set of test images captured from distance B (tB) with the corresponding images for the same lighting conditions created from photometric stereo information (BB). Figure 6.5 shows one image of the set tB and the corresponding image generated from photometric stereo BB for textures T3 and T22. Note that to evaluate the quality of the generated images visually is not an easy task since the differences between the original and the generated images are difficult to perceive.

In order to perform this evaluation, we propose to quantify the difference between

(a) (b) (c) (d)

Figure 6.5: (a) and (c) Images captured with camera (tB). (b) and (d) Corresponding images generated by photometric stereo (BB).

a captured colour image and a generated one using the mean square error of colour differences computed over all the pixels. To compute the colour difference between the predicted and true colour values for a pixel, we use the Euclidean metric in conjunction with the Luv colour space [159]. This way the estimated error in colour reflects the perceived difference in colour since the Luv space is assumed to be perceptually uniform.

In the first row of Table 6.1 we give the average mean square error and its standard deviation for each type of surface used and for all twenty-five textures together for all the images captured for testing with different tilt angles (i.e. over 13 textures $\times$ 12 tilt angles + 12 textures $\times$ 24 tilt angles = 444 test images). From all these tests we observed that, in general, the MSE is larger when the degree of surface roughness increases. For example, rough surfaces such as textures T16, T20, and T21 had an average MSE for all tilt angles of 13.9003, 13.5633 and 12.7423 respectively, while relatively smooth surfaces such as T3, T7 and T10 had an average MSE of 5.7843, 4.1196 and 4.2268. We conclude that the photometric stereo technique introduces errors and the generated images are not perfect. However, for many textures we may consider these results as acceptable. We must remember that our goal is to use generated images as models in the classification process, therefore, the key question is whether the generated images are accurate enough to enhance the classification performance with respect to a naive classification system in which the same image texture captured under a particular light direction is always used as the reference image. Figure 6.6 shows an illustrative example of the MSE obtained when

Table 6.1: Quantitative assessment for each approach over all tilt angles. Average MSE and its standard deviation for the colour difference between the predicted and true values, using the Euclidean metric in conjunction with the Luv colour space.

	Isotropic		Directional		Rough		Overall	
Approach	Avg	Std	Avg	Std	Avg	Std	Avg	Std
tB vs BB	7.2346	2.9259	6.7650	1.8441	11.9886	2.9665	9.2512	3.6690
tB vs $imaAB$	8.5669	3.5692	7.5840	3.0038	12.4888	2.5783	10.1352	3.6501
tB vs $surAB$	8.6240	3.6385	7.6447	3.0654	12.6897	2.5231	10.2526	3.7084

comparing the test images tB with the images generated by photometric stereo BB, along with the MSE obtained in the naive case, in which only the captured image under the tilt angle 0° is used as a model to perform the comparison with all tB images. This testing was carried out on texture T21 which is one of the cases with the least accurate generated images with an average MSE of 12.7424. This means that is an example in which photometric stereo introduces a large error in image generation. We observe that the average MSE for the naive case is much higher, i.e. 20.1357, with a maximum MSE of 29.0195.

Note that the result of this experiment allows us to show how the use of a single image as the reference image produces a larger difference between images. This will imply errors in the computation of features and, therefore, errors in classification. Hence, we demonstrated that a model-based system using virtual images generated by photometric stereo, even if they are not very accurate, would be preferable over the use of a fixed reference image for the classification of textures under varying lighting conditions.

6.3.2 Experiment 2: Accuracy of Image Prediction when the Distance Changes

The purpose of this experiment is to evaluate the accuracy of image prediction using a photometric set captured from distance A and to predict how an image will appear from distance B. Using the image prediction method in Section 4.3.1, the image

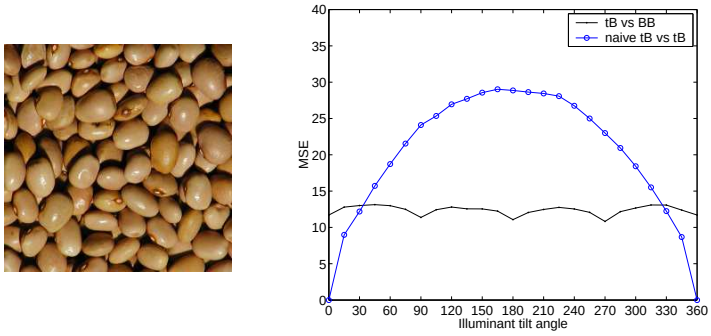


Figure 6.6: Mean square errors for texture T21 under varying tilt angles. MSE for the evaluation tB vs BB and the naive case tB vs BB, in which the process of surface recovery and rendering has been bypassed and the image at tilt=0° is compared directly with images tB.

intensities are directly predicted from the photometric information extracted from distance A. Therefore, we call these predicted images $imaAB$. However, using the surface prediction method in Section 4.3.2, only the surface gradient vectors are directly predicted but not the image intensities. As it was explained in Section 4.3.2, we generate predicted images by using the predicted gradient vectors at distance B and the average reflectance function for each surface tile assuming that our sensor and light remain spectrally unchanged. We call these predicted images $surAB$. Therefore, in this experiment we perform a comparison between the test images tB and the predicted images $imaAB$ and $surAB$.

Several results obtained over three textures are shown in Figure 6.7, where we plot the MSE over all the tilt angles. Note that for each texture, three different comparisons are shown: two curves show the results of comparing the predictions $imaAB$ and $surAB$ with the test images tB captured with the camera at distance B. The third curve shows the results of comparing the image created by using the information extracted by photometric stereo for distance B and rendering, with the test images tB. This curve presents the error produced exclusively by the photometric stereo technique, as no change in distance is involved.

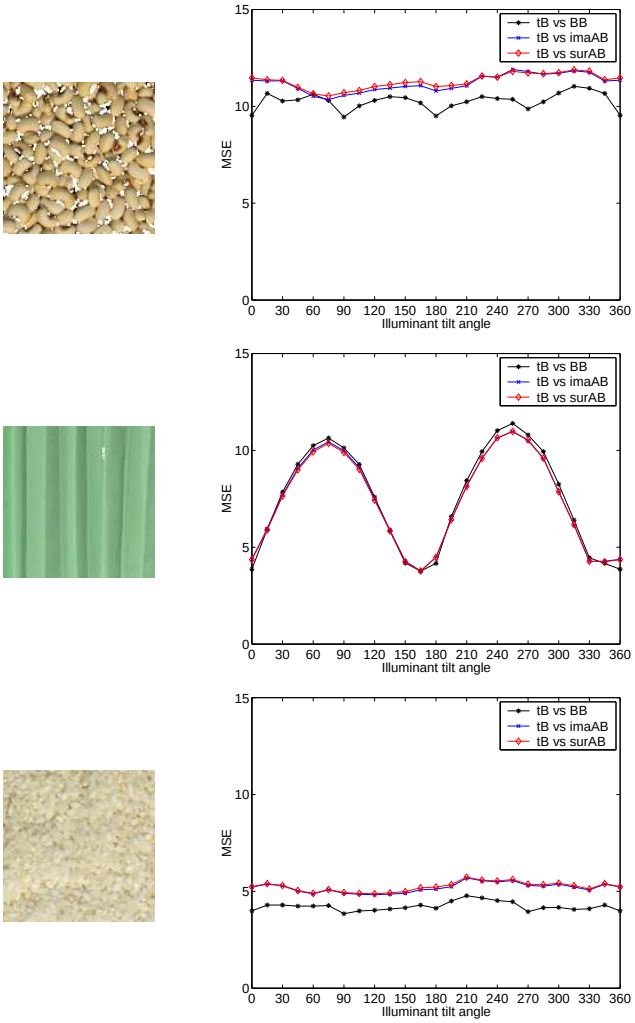


Figure 6.7: Accuracy of image prediction for three surface textures under varying tilt angles (textures T22, T14, and T10). The curve tB *vs* BB compares real images captured from distance B with those created using the information extracted by photometric stereo at distance B as well. The curves tB *vs* $imaAB$ and tB *vs* $surAB$ compare predicted images with data from distance A with real images captured from distance B.

Observing the results obtained over the twenty-five textures we conclude that in almost all cases, the performance of both prediction AB methods is very similar, producing very small differences. In Table 6.1 we give an overall quantitative assessment for each method by computing the average MSE and its standard deviation over all textures and tilt angles. The average MSE is similar for each prediction ap-

proach, although the image prediction method, which predicts the pixel intensities directly, gives, in general, smaller errors in the image. Note that most of the error can be accounted for as being produced by the photometric stereo technique and not by the step dealing with the distance change.

We also observe that surface roughness has an influence on the accuracy of the image predictions. For rougher surfaces, the error of the prediction is increased (see Table 6.1). Other surface properties, such as directionality or specularity, may contribute to the errors as well. For example, in the directional texture T14 in Figure 6.7, some orientations of the light source provoke more difficulties than others and, therefore, the error of image prediction may vary significantly depending on the light direction.

6.3.3 Experiment 3: Accuracy of Surface Shape Prediction

The goal of this experiment is to perform an evaluation of surface shape prediction, comparing the predicted gradient vectors AB with those obtained using the original photometric set for distance B.

Using the surface prediction method described in Section 4.3.2, the gradient vectors are directly predicted from the photometric information extracted for distance A. However, using the image prediction method (see Section 4.3.1), only the intensity values can be predicted but not the gradient vectors. Therefore, to make the surface shape evaluation possible, photometric stereo was applied to these predicted images (four images corresponding to four directions of the light) in order to compute the gradient vectors.

The first column of Figure 6.8 shows three examples of image predictions $imaAB$ with a particular light direction. Note that the image intensity we predict consists of image patches and not full images since there are points which photometric stereo can not recover correctly. We flag these points with white values.

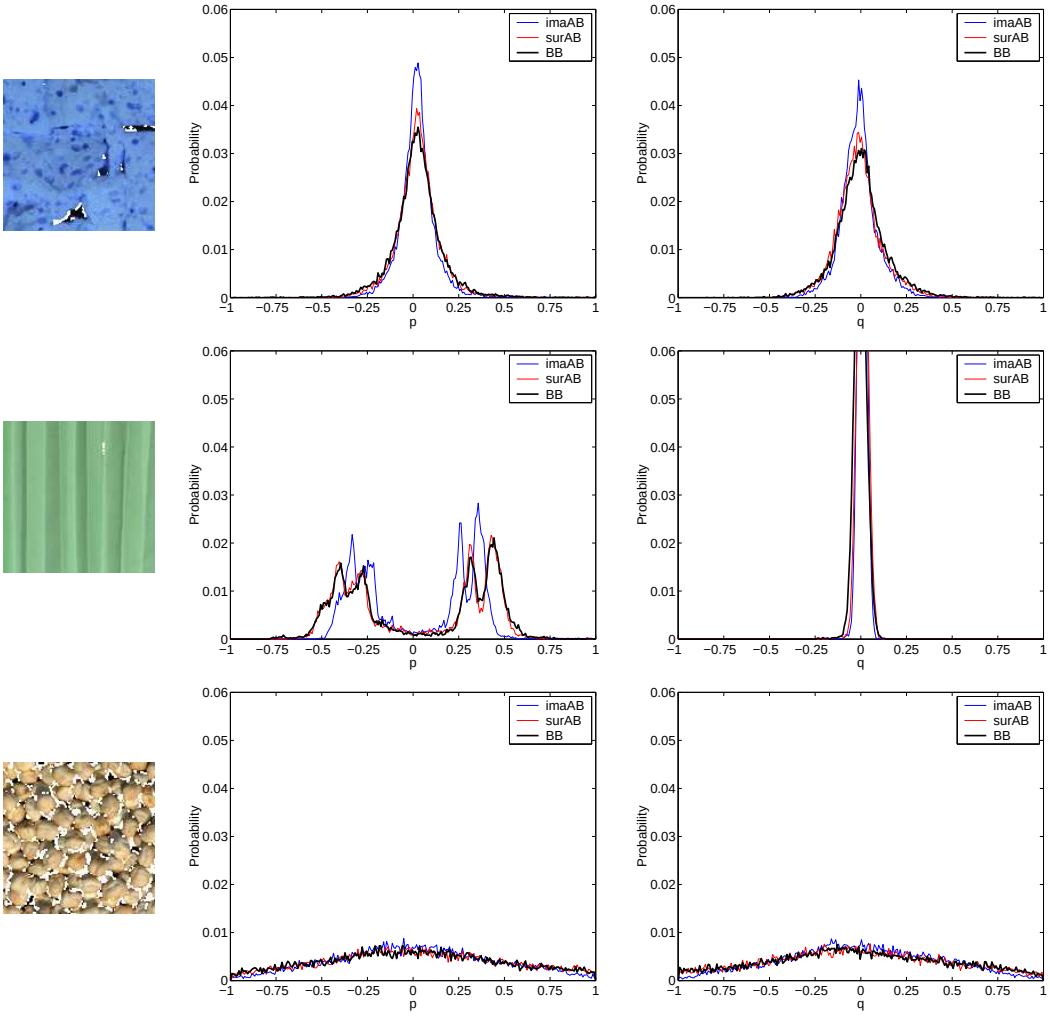


Figure 6.8: Accuracy of the surface shape predictions for three surface textures (T5, T14, and T15). Second and third column show the distributions of the surface gradients p and q obtained by applying photometric stereo directly to the original images for distance B, by applying photometric stereo to the images predicted by the method in Section 4.3.1 ($imaAB$) and by applying the surface prediction method in Section 4.3.2 ($surAB$).

Before we perform the evaluation, it is necessary to solve the problem of localising which region of the original set of distance B corresponds exactly to the region of the prediction AB. We do this by computing the correlation of surface shape (gradient components p and q) between results obtained by applying photometric stereo to the original set directly and results obtained with our prediction AB. When computing

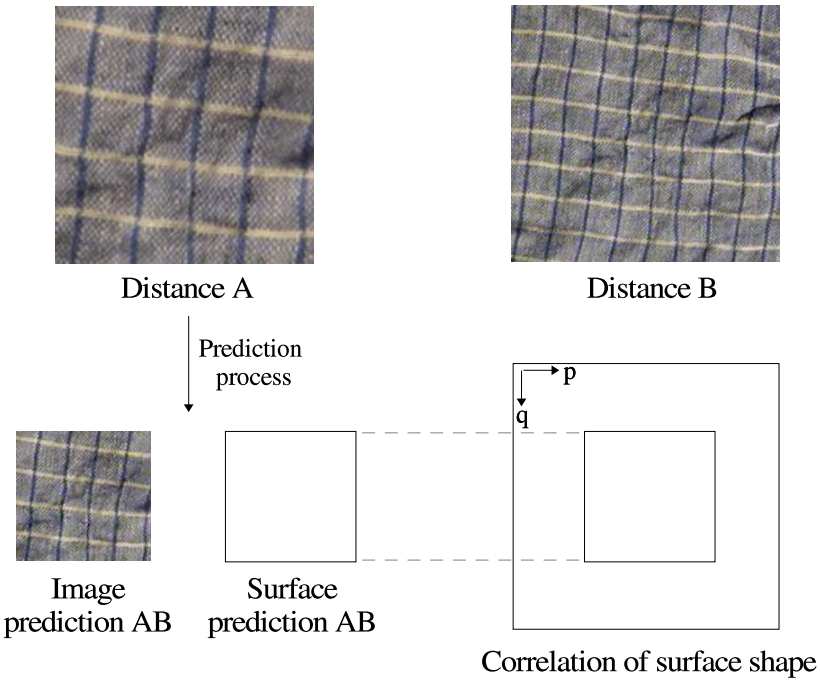


Figure 6.9: Correlation scheme. The correlation of surface shape (gradient components p and q) between results obtained by applying photometric stereo directly to the original set of distance B and results obtained with our prediction AB.

the correlation function, we exclude all points which are flagged as not reconstructed. The correlation method is applied separately for the gradient components p and q, obtaining a set of possible relative shifts between the corresponding images from p and another set of possible relative shifts between the corresponding images from q. Then, the common shifting which maximises both correlations of p and q is chosen, localising the region of the original set exactly. Figure 6.9 summarises the scheme of this correlation process. After that, the PDFs of p and q are used in order to compare the surface shape information.

Each plot in Figure 6.8 shows the surface distributions obtained by the prediction methods and those obtained by applying photometric stereo to the original images captured from distance B.

Analysing the PDFs of this figure, we observe that the results obtained by the surface prediction method fit the original PDF distributions better. On the other

Table 6.2: Quantitative assessment of shape predictions $imaAB$ and $surAB$. Two quantitative measures are used: (1) the average MSE of the PDFs and their standard deviation and (2) the average MSE per pixel of gradient components p and q and their standard deviation.

			Isotropic		Directional		Rough		Overall	
	Prediction		Avg	Std	Avg	Std	Avg	Std	Avg	Std
PDF error	$imaAB$	p	0.0008	0.0002	0.0018	0.0008	0.0017	0.0016	0.0013	0.0011
	$surAB$	p	0.0005	0.0001	0.0008	0.0003	0.0012	0.0017	0.0008	0.0011
	$imaAB$	q	0.0009	0.0002	0.0023	0.0007	0.0017	0.0018	0.0014	0.0012
	$surAB$	q	0.0005	0.0001	0.0011	0.0003	0.0012	0.0018	0.0009	0.0011
Error per pixel	$imaAB$	p	0.0708	0.0592	0.0623	0.0699	0.0849	0.0371	0.0757	0.0507
	$surAB$	p	0.0620	0.0389	0.0627	0.0507	0.0914	0.0274	0.0751	0.0377
	$imaAB$	q	0.0679	0.0476	0.0507	0.0464	0.0998	0.0330	0.0792	0.0442
	$surAB$	q	0.0605	0.0384	0.0572	0.0491	0.0896	0.0268	0.0728	0.0373

hand, surface information extracted from the predicted images ($imaAB$) introduces more error in the predicted gradient vectors. In general, the gradient values are smaller than those obtained by the surface prediction method. That can be clearly observed, for instance, in the q distribution of the second texture. In Table 6.2 we confirm this conclusion providing an overall quantitative assesment over all these histogram comparisons. We computed the average MSE of each histogram and its standard deviation over all textures. We also included the average MSE per pixel of the gradient components p and q. In both cases, better results are obtained with the surface prediction approach. The reason for these results can be found in the error introduced by the image prediction ($imaAB$), which is propagated when photometric stereo is applied to the generated images in order to recover shape information.

As mentioned earlier, the *absolute average slope ratio* ($AASR$) provides another way of characterising the degree of roughness of a given surface texture with a single parameter. We used this ratio as an alternative measure to evaluate our surface shape predictions. We compared the $AASR$ of our predictions AB with the estimated values computed using the shape information extracted by the photometric set captured from distance B. Table 6.3 gives an overall quantitative assessment over all textures computing the average MSE of the $AASR$ parameter obtained using

Table 6.3: Quantitative assessment using the $AASR$ parameter. Average MSE and its standard deviation for each prediction approach ($imaAB$ and $surAB$).

	Isotropic		Directional		Rough		Overall	
Prediction	Avg	Std	Avg	Std	Avg	Std	Avg	Std
$imaAB$	0.0273	0.0095	0.0382	0.0257	0.0369	0.0170	0.0334	0.0161
$surAB$	0.0061	0.0035	0.0129	0.0125	0.0124	0.0097	0.0095	0.0085

both prediction approaches. Note that the values obtained by the surface prediction method ($surAB$) are again better than those obtained by the photometric stereo approach applied to the predicted images ($imaAB$).

From this experiment we conclude that the surface prediction method provides the best shape estimation. Moreover, we observe that, in general, surface roughness has a strong influence on the accuracy of the surface shape predictions. For rougher surfaces the error of the predictions is increased (see Tables 6.2 and 6.3). For instance, the MSE per pixel of the gradient components p and q for texture T15, which is a surface with a high degree of roughness ($AASR$ of 0.3960), are 0.1019 and 0.1573 respectively, while for texture T5, a relatively smooth surface ($AASR$ of 0.1003), are 0.0470 and 0.0558, respectively.

6.3.4 Experiment 4: Accuracy of Image Prediction from Different Distances

The purpose of this experiment is to evaluate the accuracy of the direct image prediction method proposed in Section 4.3.1, using a photometric set captured at distance A and another at distance B to predict images captured at the longer distance C ($A < B < C$). The main goal of this experiment is to analyse the effect of using two different resolutions in the original sets to perform the same prediction when seen from distance C. Two different prediction distances (AC and BC) are used in the image prediction method.

Note that now three different sets of images taken from distances A, B, and C are used. In our image database we have fifteen textures which were captured for

(a) (b) (c)

Figure 6.10: Three images of the same surface texture seen from distances A, B, and C, (a), (b) and (c) respectively.

these three distances: textures T1, T2, T3, T4, T10, T11, T12, T13, T14, T15, T16, T17, T18, T19 and T20. Figure 6.10 shows 3 images of texture T15 seen from distances A, B and C. Over these fifteen textures, we have performed a comparasion between the test images tC and the predicted images obtained from distances A and B. To quantify the difference between a captured colour image (test image) and a predicted colour image, we use again the mean square error of colour differences computed over all pixels.

Observing the results obtained over the fifteen textures, we conclude that in almost all cases, the performance of both predictions is similar, although the predictions AC (those obtained from the information extracted at the closest distance A) give, in general, smaller errors in the image. In Table 6.4 we give an overall quantitative assessment for each prediction (AC and BC) by computing the average MSE and its standard deviation over all textures and all tilt angles which confirm this conclusion. The reason for these results can be found in the superior resolution of the images from distance A and therefore in the superior amount of information available to perform the predictions. In fact, with prediction AC, more surface patches at distance A are used to compute the intensity value corresponding to a surface path of distance C. On the other hand, there is less information available from distance B to predict the corresponding value for distance C. Note that the major error is again produced by the photometric stereo technique and not by the inclusion of the distance prediction.

Table 6.4: Overall quantitative assessment for each prediction over all 15 textures and all tilt angles. Average MSE and its standard deviation for the colour difference between the predicted and true values using the Euclidean metric in conjunction with the Luv colour space.

Overall assessment	Avg.	Std
tC vs CC	10.4412	3.5115
tC vs AC	11.4239	3.3770
tC vs BC	13.2937	3.9568

6.3.5 Experiment 5: Accuracy of Surface Shape Prediction from Different Distances

The goal of this experiment is to perform an evaluation of the surface shape prediction method described in Section 4.3.2, comparing the predicted gradient vectors AC and BC with those obtained using the original photometric set for distance C. In this experiment, we have considered the same experimental setup as for previous experiment. Therefore, we use the set of fifteen textures taken from distances A, B, and C.

Before we perform the surface evaluation, we have seen that it is necessary to solve the problem of localising which region of the original set of distance C corresponds exactly to the region of the predictions AC and BC. As explained in experiment 3 in Section 6.3.3, we do this by computing the correlation of surface shape (gradient components p and q) between results obtained by applying photometric stereo to the original set directly and results obtained with our predictions.

By analysing the PDFs for all the predictions, we observed that results obtained by prediction AC fit the original PDF distributions better. On the other hand, surface shape information extracted from distance B introduces more error in the predicted gradient vectors. In general, the gradient values are smaller than those obtained from distance A. In Table 6.5 we confirm this conclusion by providing an overall quantitative assesment over all these histogram comparisons. As in the third experiment, we computed over all the textures the average MSE of each histogram and its standard deviation. Moreover, we also included the average MSE per pixel

Table 6.5: Overall quantitative assessment for all 15 textures of shape predictions AC and BC. Two quantitative measures: (1) the average MSE of the PDFs and their standard deviation and (2) the average MSE per pixel of gradient components p and q and their standard deviation.

		p		q	
	Pre.	Avg.	std	Avg.	std
PDF error	AC	0.0008	0.0005	0.0006	0.0002
	BC	0.0010	0.0004	0.0011	0.0006
Pixel error	AC	0.0794	0.0525	0.0873	0.0449
	BC	0.1130	0.0606	0.1343	0.0638

of the gradient components p and q. In both cases, better results are achieved when the prediction is carried out from the original distance A. This is due to the fact that for prediction AC more surface patches are used to compute the desired normal vector corresponding to the surface patch of distance C.

We also used the $AASR$ ratio to evaluate the surface shape predictions. We compared the $AASR$ of our predictions AC and BC. Table 6.6 gives an overall quantitative assessment for all the textures, computing the average MSE of the $AASR$ parameter obtained using both predictions. Note that the values obtained by the surface predictions AC are again better than those obtained by the predictions BC.

As a result of this experiment we conclude that prediction AC, which has a larger distance prediction but also the most resolution among the original images, provides better shape estimation than prediction BC. We demonstrated that the accuracy of the original information used to perform the predictions has a strong influence on the results of our predictions.

6.4 Texture Classification Results

After describing the experimental trials performed on texture prediction, we will focus on the experimental trials related to the texture classification system proposed in Chapter 5. All experiments performed have been for the purpose of checking

Table 6.6: Overall quantitative assessment for all 15 textures of the $AASR$ parameter. Average MSE and its standard deviation for prediction AC and BC.

Prediction	Avg.	std
AC	0.0091	0.0078
BC	0.0263	0.0114

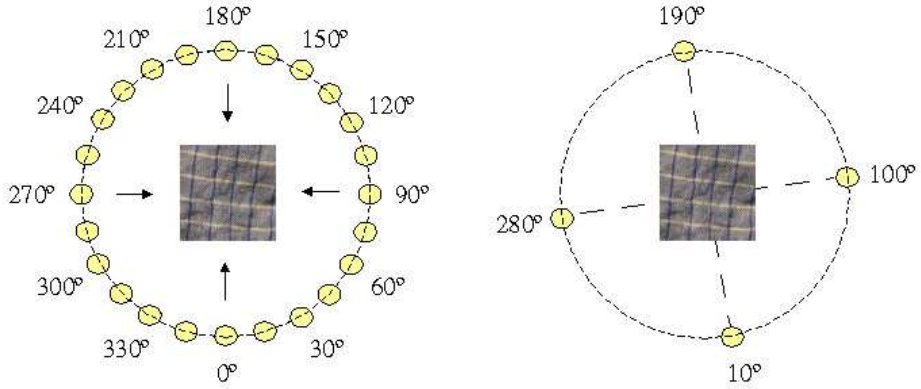
various aspects of the theory, namely:

1. To check the accuracy of the classifier when photometric sets of distance B are used to produce virtual model images for the same distance, in terms of which images captured from distance B are to be classified. This experiment provides a quantitative evaluation of our model-based texture classification when we use model images and test images from the same distance B and when the test images have been captured under different light directions from those used by the photometric stereo.

2. To check the accuracy of the light direction estimation provided by our model-based classification, giving a measure of how well the light direction in the test images can be estimated when reference and test images have been captured from the same distance.

3. To check the accuracy of the classifier when photometric sets from distance A are used to produce virtual model images for distance B, in terms of which images captured from distance B are to be classified. The results of this experiment are to be used to show how well our system can classify test images captured from the longer distance B and under different light directions.

4. To check the accuracy of the light direction estimation provided by our model-based classification, giving a measure of how well the light direction in the test images can be estimated when reference and test images have been captured from different distances.



(a) (b)

Figure 6.11: Illumination setup. (a) Illuminant tilt angles used to capture the test images. (b) Illuminant tilt angles used to create the virtual database. The unknown illuminant tilt angle of the test image has to be classified to the nearest one among these four angles used for the creation of the virtual database.

6.4.1 Experiment 1: Accuracy of Classification when Photometric and Testing Images are Captured from the Same Distance

This experiment analyses the accuracy of the texture classifier when photometric sets from distance B are used to produce model images for the same distance B. For this experiment, and all successive ones, we always use 25 texture classes in the classification process. However, in order to show which textures are more difficult to classify, we present the classification errors separately for each group of surface textures (isotropic, directional, and rough surfaces).

For each surface texture, 4 images were rendered using an elevation angle of 55° and 4 illuminant tilt angles: 10°, 100°, 190° and 280° (see Figure 6.11.b). This is the virtual database of images used as references for classification. It is composed of 100 texture images (25 surfaces × 4 illuminant tilt angles). The four illumination tilt angles used are different from those used for the test images. Therefore, when classification is performed we do not have exact correspondence between the tilt angles of the images in the virtual database and the test images.

When the virtual database is created, the recognition procedure starts. This procedure is divided into two steps: the learning process and the classification process. The goal of the learning process is to model each texture class of the virtual database by means of a representative feature vector. The goal of the classification process is to classify an unknown test image into the texture class to which it belongs.

The learning process starts by feature extraction, i.e. computing a representative feature vector for each texture image in the virtual database. Co-ocurrence matrices [61] are used to extract as features the contrast, homogeneity and energy for 20 different values of a distance d (distances between $[1 \dots 55]$ incremented in steps of 3). The pixels labeled as un-reconstructed points (points the shape and colour of which are unreliably calculated by photometric stereo) are not used in the computation of the co-ocurrence matrices. The co-ocurrence matrices are implemented in an anisotropic way. We analyse 4 different directions: 0°, 45°, 90° and 135° so that we have in all 240 texture features (3 features $\times$ 4 directions $\times$ 20 distances d). From all the computed features, those which can distinguish best between the different classes are chosen. We use the Sequential Forward Selection (SFS) algorithm [81] and a set of training images in order to select the best feature set for discrimination. The feature evaluation is performed by applying the Nearest Neighbour classifier over the set of virtual training images. This training set is composed of 3 virtual images for each surface texture and each illuminant tilt angle. We have in all 600 rendered images = 25 textures $\times$ 8 illuminations $\times$ 3 images.

When the learning is finished, the classification process starts by first extracting the feature vectors for the unknown test (real) images and classifying them by means of the Nearest Neighbour classifier into one of the classes in the virtual database. At the same time, this classification process allows us to approximate the illuminant tilt angle of the test image by identifying it with one of its nearest neighbours in the virtual database.

Due to the large size of the captured images, we produced from each one different subimages of size 133 $\times$ 133 to be used for testing. The test set consists of 9 real images for each surface texture and each illuminant tilt angle. We have

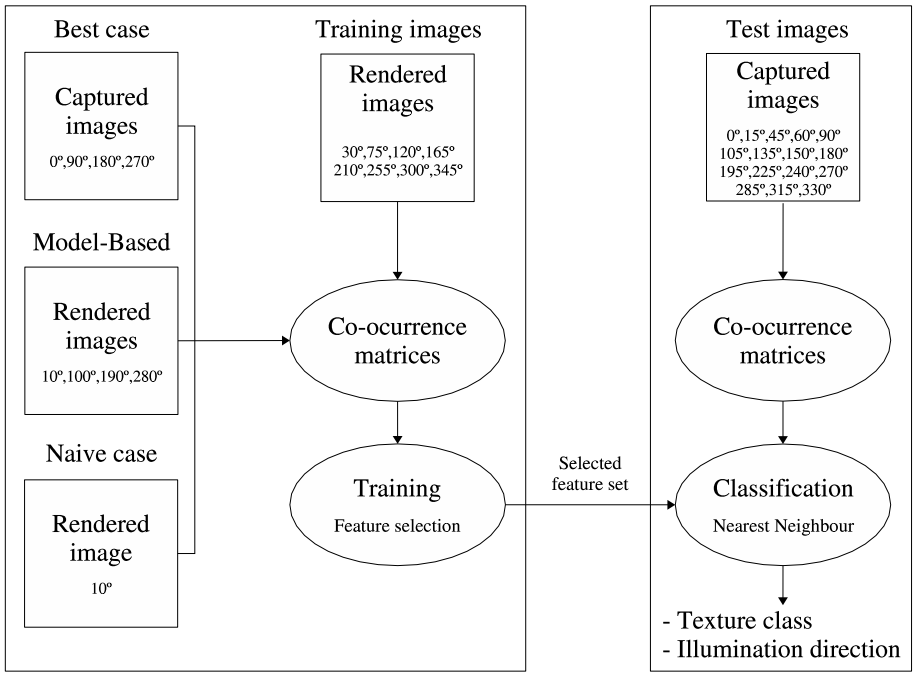


Figure 6.12: Model-based, best case, and naive case approaches to texture classification.

in all 2664 images (13 textures $\times$ 8 illuminations $\times$ 9 images + 12 textures $\times$ 16 illuminations $\times$ 9 images). Note that different tilt angles are used for training and testing; as in a real situation we do not know the true tilt angle of the test image.

In this experiment, we compare the results of our model-based approach with the results obtained by the "best" case in which the 4 reference images of each texture are real images captured with a camera and not images rendered by photometric stereo. We also compare the results with those obtained by the "naive" case in which just a single captured image of one light direction (tilt angle 0°) is used to characterise each texture class. Figure 6.12 illustrates the classification system configuration for these three different cases.

Using our model-based approach, the system obtained a 97.04% accuracy of texture classification. However, in the "best" case in which original images were used as references for the purpose of classification, the texture classification accuracy was 100%. For the "naive" case in which only one light direction was used for training, the texture classification accuracy was 87.73% (see Table 6.7). Analysing the mis-

Table 6.7: Texture and illuminant classification rates obtained for the best, model-based and naive cases when photometric sets from distance B are used to create images for distance B.

	Experiments 1 and 2				
	Best case		Model-based		Naive case
Texture	Texture	Illuminant	Texture	Illuminant	Texture
T1	100%	100%	100%	75.01%	46.67%
T2	100%	100%	100%	31.29%	100%
T3	100%	100%	100%	88.20%	100%
T4	100%	100%	100%	75.01%	73.33%
T5	100%	100%	100%	81.26%	64.44%
T6	100%	100%	100%	100%	76.29%
T7	100%	100%	100%	81.26%	100%
T8	100%	97.04%	100%	75.71%	46.67%
T9	100%	100%	88.89%	14.63%	73.33%
T10	100%	100%	100%	66.68%	97.04%
T11	100%	100%	100%	91.67%	46.67%
T12	100%	100%	100%	57.66%	100%
T13	100%	100%	100%	100%	100%
T14	100%	100%	100%	100%	100%
T15	100%	100%	100%	63.91%	100%
T16	100%	87.41%	91.85%	41.00%	100%
T17	100%	95.56%	99.56%	68.07%	100%
T18	100%	94.07%	59.70%	59.74%	100%
T19	100%	94.07%	100%	61.83%	100%
T20	100%	82.96%	85.93%	76.40%	100%
T21	100%	98.52%	100%	80.57%	93.33%
T22	100%	95.56%	100%	72.24%	100%
T23	100%	97.04%	100%	84.73%	100%
T24	100%	93.33%	100%	63.91%	75.56%
T25	100%	91.11%	100%	43.09%	100%
Overall	100%	97.06%	97.04%	70.15%	87.73%

classification error of the model-based approach (2.96%), we concluded that 84.98% of this error was produced by the group of very rough surfaces, while the remaining 15.02% was produced by the isotropic surfaces. This indicates that the missclassification errors of the model-based approach are mainly due to the image generation by colour photometric stereo information. Moreover, the results demonstrate that the model-based approach significantly reduces the texture classification errors caused by changes in light direction compared with the "naive" case.

6.4.2 Experiment 2: Accuracy of Illuminant Estimation when Photometric and Testing Images are Captured from the Same Distance

The purpose of this experiment is to evaluate the accuracy of the light direction estimation provided by our model-based classification when reference and test images have been captured from the same distance B.

As well as classifying the test images captured from distance B into the corresponding texture class, we can also classify the illuminant tilt angle under which they were captured, into one of the 4 illuminant tilt angles in the images in the virtual database. Note that by using only 4 virtual images for each surface, we can estimate the illuminant tilt angle with an accuracy of $\pm 45^\circ$ only.

With this in mind, in this experiment, the illuminant tilt angle could be estimated corectly in 70.15% of the cases by the model-based approach. However, in the "best" case, in which captured images are used as references, an accuracy of 97.06% was obtained (see Table 6.7). So, the accuracy of the illuminant tilt angle classification is considerably lower compared with the "best" case. This is presumably due to the errors introduced by the process of virtual image generation. These errors have a major influence on the illumination classification but less on texture classification because the differences between features extracted from images of the same texture under different tilt angles are smaller than the differences between features extracted from different texture classes.

6.4.3 Experiment 3: Accuracy of Classification when Photometric and Testing Images are Captured from Different Distances

This experiment analyses the accuracy of the texture classifier when photometric sets from distance A are used to produce model images for distance B.

Now, the virtual database of images corresponding to the longer distance B is generated using the direct image prediction method described of Section 4.3.1. As discussed in the experiment in Section 6.3.2, this image prediction method produces

Figure 6.13: One image predicted for each of the twenty five textures. The points flagged with white are those which cannot be correctly recovered by the photometric stereo technique.

better results than those obtained by the image prediction via the surface prediction method proposed in Section 4.3.2.

As for experiment 6.4.1, the virtual database is composed of 100 image textures (25 surfaces $\times$ 4 illuminant tilt angles). Figure 6.13 shows a predicted image for each of the 25 surface textures. In the experiment, 600 images are used for training, while 2664 images are used for testing.

After applying the feature selection algorithm and choosing the appropriate feature set, we apply the classifier to the unknown test images. Using the model-based approach, the system classified 89.57% of them into the correct texture class (see Table 6.8). We had 100% correct texture classification in the "best" case when reference and test images were captured from distance B (note that this case cor-

responds exactly to the "best" case used in experiment 6.4.1). The "naive" case here is the case when we use as the reference image one captured from a different distance from that of the test image. In this case, the classifier achieved only an accuracy of 34.35%.

Analysing the missclassification error of the model-based approach (10.43%), we concluded that 92.61% of this error was produced by the group of very rough surfaces while the isotropic surfaces contributed only 7.39% of this error.

Table 6.8 summarises the obtained texture classification rates using the model-based approach, the "best" case and the "naive" case. Comparing these results with those obtained in experiment 1 in Section 6.4.1, we conclude that the performance of the classifier is decreased due to the error introduced by the image prediction method in Section 4.3.1. However, we also demonstrate that the model-based approach increases the accuracy of the texture classification significantly compared with the "naive" case.

Note that results obtained with the "naive" case justify the creation of the virtual database of images seen at the longer distance as well as the specific learning process for these images. In [103] we presented similar experimental results related to the naive texture classification. We tried to classify test images captured from distance B using the features extracted from the virtual database of images from distance A. In this "naive" case, we confirmed the conclusion stated above since we classified correctly only 21.83% of the test images.

6.4.4 Experiment 4: Accuracy of Illuminant Estimation when Photometric and Testing Images are Captured from Different Distances

The goal of this experiment is to evaluate the accuracy of the light direction estimation provided by the model-based classification when reference and test images are from different distances.

Table 6.8: Texture and illuminant classification rates obtained for the best, model-based and naive cases when the direct image prediction method is used to generate images for distance B from the photometric sets from distance A.

	Experiments 3 and 4				
	Best case		Model-based		Naive case
Texture	Texture	Illuminant	Texture	Illuminant	Texture
T1	100%	100%	100%	77.87%	0%
T2	100%	100%	100%	77.87%	35.55%
T3	100%	100%	100%	73.56%	0%
T4	100%	100%	93.33%	28.07%	0%
T5	100%	100%	100%	83.40%	53.33%
T6	100%	100%	100%	100%	0%
T7	100%	100%	87.41%	100%	85.76%
T8	100%	97.04%	100%	61.27%	100%
T9	100%	100%	100%	17.00%	46.67%
T10	100%	100%	100%	55.73%	100%
T11	100%	100%	100%	39.13%	0%
T12	100%	100%	100%	77.87%	23.61%
T13	100%	100%	100%	92.62%	40.40%
T14	100%	100%	100%	77.87%	5.92%
T15	100%	100%	100%	100%	26.67%
T16	100%	87.41%	95.56%	60.65%	65.93%
T17	100%	95.56%	87.41%	81.56%	44.44%
T18	100%	94.07%	48.15%	82.17%	73.11%
T19	100%	94.07%	100%	58.81%	85.67%
T20	100%	82.96%	9.63%	23.76%	0%
T21	100%	98.52%	23.70%	35.44%	0%
T22	100%	95.56%	100%	77.87%	71.84%
T23	100%	97.04%	94.07%	78.48%	0%
T24	100%	93.33%	100%	75.41%	0%
T25	100%	91.11%	100%	33.60%	0%
Overall	100%	97.06%	89.57%	66.80%	34.35%

We used the same experimental setup as that used in experiment 3 in Section 6.4.3. Using the model-based approach, the system estimated the illuminant tilt angle with an accuracy of 66.80% while the light classification accuracy of the "best" case[1] was 97.06% (see Table 6.8). Comparing the results of this experiment with those obtained when the reference and testing images were captured from the same distance, we conclude that the accuracy of the tilt angle classification is de-

[1]Note that for the "best" case, the texture and light classification rates are the same as those obtained in experiment 1 of Section 6.4.1. In both experiments, we used the same reference and test images from distance B.

creased due to the error introduced by the step of the algorithm when dealing with the change in distance.

6.5 Summary and Conclusions

The prediction methods presented in Section 4.3 and the model-based texture classification system proposed in Section 5.3 were tested and evaluated in this chapter.

After analysing different texture databases we observed the lack of a database to accomplish our principal requirements: textures seen under various directions of light and various distances from the camera. Due to this fact, we opted to build our **own image database** to perform the experimental trials presented in this thesis. The image database contains real surface textures which can be classified into relatively smooth surfaces (which may be further divided into isotropic surfaces and directional surfaces) and very rough surfaces. For each surface texture, as well as the 4 required images to apply the photometric stereo technique, the images which should be used as training and test images are available in our image database.

The experimental trials presented in this chapter are divided into two blocks: one related to texture prediction, and the other to texture classification. In the first block we performed different experimental trials in order to evaluate the two prediction methods in Section 4.3:

- One which allows us to predict image intensities directly (direct image prediction).
- Another which allows us to predict the surface shape information first and then the image intensities (image prediction via surface prediction).

Both methods were tested and evaluated over a set of twenty-five surface textures, demonstrating the ability to predict the image texture a particular surface texture will create when seen from a distance longer than the original. The direct image prediction method produces, in general, smaller errors. We also concluded that most errors can be accounted for as being produced by the photometric stereo technique

and not by the step dealing with distance change. On the other hand, the surface prediction method produces the best shape predictions. Several error measures were used in order to evaluate the surface shape predictions: the absolute average slope ratio ($AASR$) which measures the degree of roughness of a surface, the MSE of the estimated probability density functions for the surface partial derivatives p and q, and the MSE per pixel of the gradient components p and q.

As a result of these experiments we concluded that surface roughness plays an important role in the accuracy of image and shape prediction. The rougher the surface, the larger the error of the prediction. Other surface properties, such as directionality or specularity, may also contribute to errors.

Another set of experimental trials related to the texture classification system were presented in the second block. We observed that even if the predicted images are not perfect, they may still make a significant difference in the accuracy of a texture classifier. We compared the results of our approach with those obtained by the "best" case where the reference and test images of each texture were real images captured under exactly the same imaging geometries and by the "naive" case where only a single real image was used as reference. We saw how the results of our texture classification approach were significantly better than those obtained by the "naive" case, while the results were not much inferior to those obtained by the "best" case. An improvement was also obtained in cases for which the assumptions made by the photometric stereo technique were violated and the photometric stereo results were not as accurate as we might have wished. In all cases, an estimate of the unknown light orientation under which the test image was captured was also obtained, although not with very high accuracy.

Although we presented results of our texture classification system when seen from different distances and under different illuminant tilt angles, it is interesting to mention that our approach could be used in other situations. For instance, changes in the illuminant slant angle and in the camera direction. If we have information on the surface relief and the surface albedo and we use the general equations derived in Section 4.3, then we could render the surface to create model textures as required.

Chapter 7

Conclusions

Conclusions extracted from this thesis work are presented. Moreover, possible further work is analysed considering the different directions in which this research line could go. Finally, publications related to this thesis are listed.

7.1 Contributions

In this thesis we have presented a model-based texture classification system able to classify textures when seen from different imaging geometries such as the distance from the camera and the direction of light. The main contributions of this thesis work are summarised in the following points:

- **Texture prediction framework.** We presented a general framework for describing textures when seen from different imaging geometries. The 4-source colour photometric stereo was used in order to obtain the reflectance and shape information of the surface from a close distance. From this surface information, the proposed prediction framework allows us to predict how the surface texture will look even when seen under different imaging conditions: different sensor, distance from the sensor (the camera positioned at distances at least as long as the distance used for capturing the training images), different spectral characteristics of the light source, and different directions of light. This is based on the assumption of Lambertian surfaces, but it can easily be

generalised to other types of surface.

Two different prediction methods have been proposed, one which allows us to predict image intensities directly (direct image prediction), and another which allows us to predict the surface shape information first and then the image intensities (image prediction via surface prediction). We observed that the direct image prediction method produces, in general, smaller errors in the images while the surface prediction method produces the best shape predictions. We extracted unsurprising conclusions: (i) a large amount of the error in the predictions can be accounted for as being produced by the photometric stereo technique, and (ii) surface roughness plays an important role in the accuracy of image and shape prediction. The rougher the surface, the larger the error of the prediction.

- **Model-based texture classification system.** A model-based classification system able to classify textures when seen from different distances and under different directions of lights was presented. The main motivation of this methodology has been to overcome the problem of classifier failure induced by varying these imaging properties. Moreover, we have exploited the capability of the classification system to guess an approximate direction of the light for the unkown test images. The whole procedure of the proposed texture classification system was divided into two main phases. In the first phase, *"virtual" database creation*, a virtual database of textured images under novel imaging geometries was created by using the surface texture information and the proposed texture prediction framework. The second phase, the *recognition procedure*, was divided into two steps: the learning process in which each texture class of the virtual database is modelled by means of a representative texture feature vector and the classification process in which an unknown test image is classified into the texture class to which it belongs. In the learning process, the images of the virtual database were used to extract texture features derived from the co-occurrence matrices. Moreover, a feature selection process was used to chose the best feature models for each texture class. On the other hand, classification was achieved by determining the nearest neighbour

in the feature space populated with feature vectors from the virtual database computed during training.

We saw, in the experimental trials, how the texture classification results of our approach are significantly better than those obtained with a typical texture classification system in which the variability produced by changes in light direction and camera distance are not considered.

- **A novel texture database.** Different databases were considered, for example, the Brodatz album [18], the "Photometric Image Databases" from the Texture lab at Heriot-Watt University [2], and the "Columbia-Utrecht database" established by Dana et al. [35]. None of them could fulfill all the requirements needed for our purposes. Therefore, we opted to build our own image database, providing for each surface texture different sets of images captured from different distances of the camera and under different directions of light. This image database is of public domain and we think it can be useful for the research community to check and test techniques related to scale invariant and light invariant texture analysis and classification. It can be downloaded from *http://eia.udg.es/∼llado/database.html*.

7.2 Future Work

The design of a texture classification system invariant to the imaging geometries involves the consideration of a wide number of questions. In addition to the various solutions which have been adopted and described in this thesis, different ideas and different approaches have been dicussed throughout this thesis work. Moreover, a large number of ideas have remained as undeveloped tasks which need to be further analysed and worked on in depth and we therefore suggest them as future work.

Basically, we have divided this future work into two blocks. The first is composed of ideas which can directly improve our texture classification approach. The second suggests some ideas for further work to be considered as future research lines.

Further Work on the Proposed Classification System

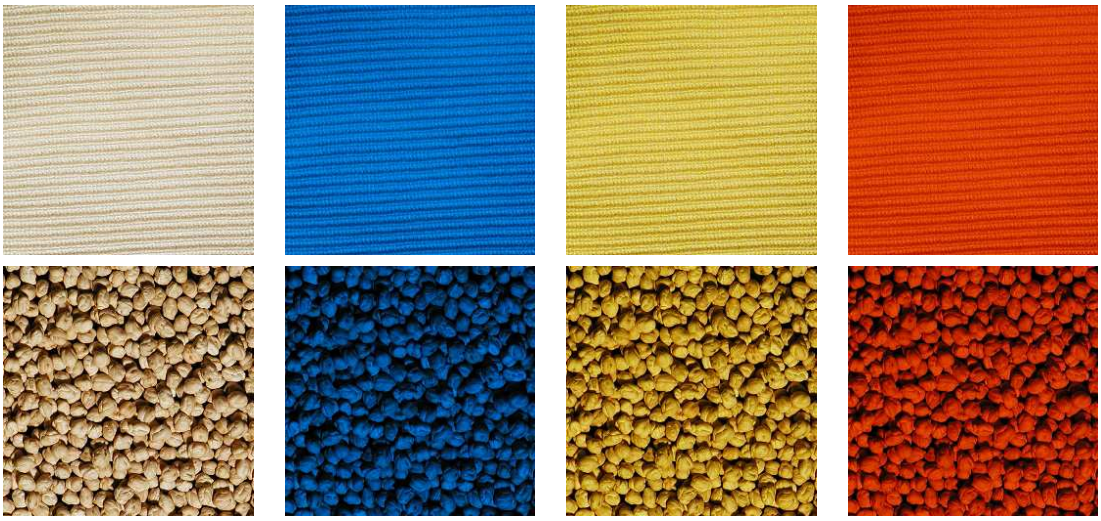
- **Different sensors and light sources.** We propose as further work to analyse the effect of parameters such as the sensitivity of the sensor and the spectral properties of the light source which have been assumed to be the same for all the experimental trials presented in this thesis. Figure 7.1 illustrates the effect of varying the spectral properties of the light source on two textures. Three different colour filters (blue, yellow and red) have been used in order to obtain these images.

- **Changes in slant angles.** Obviously, large slant angles will intensify the level of shadowing, providing more information for the photometric stereo technique but also deterioring the precision of the surface estimation. On the other hand, low slant angles (close to a perpendicular to the surface plane) will provide less information since the perceived image textures would be more similar. It could be very interesting to analyse what happens in all of these cases.

- **Tolerance in the estimate of the guessed distance of the camera.** What if we do not known exactly the longer distance from which the test image was captured? How accurately do we have to guess it for the performance of the classifier not to deteriorate significantly? These are important aspects which could be studied.

- **Different texture features.** As it has been noted, there are different texture analysis techniques which allow the extraction of texture information from an image. We propose a future project to consider different types of features and compare them with those extracted from co-occurrence matrices.

- **Extend our texture database.** We also suggest, as further work, to extend our texture database. We propose to include more surface textures captured under different imaging geometries i.e. camera distance, light directions (including different slant angles), different spectral properties of the light source and using cameras with different sensitivity.



(a) (b) (c) (d)

Figure 7.1: Images of two textures captured under varying the spectral properties of the light source. (a) Images captured with the original light source, (b)-(d) images captured using three different colour filters: blue, yellow, and red, respectively.

Further Work along this Research Line

- **Analyse the texture features behaviour.** In Appendix A we presented a first step in order to deal with the illumination invariant classification analysing the behaviour of the features extracted from the co-occurrence matrices. This has been used to perform a simultaneous surface texture classification and illumination tilt angle prediction. We think that this is an interesting idea to continue this research line. This can provide important knowledge and open new possibilities.

- **Viewpoint independent texture classification.** Recently, Lazebnik et al. [91, 92] introduced a texture representation suitable for recognising images of textured surfaces under a wide range of transformations, including viewpoint changes and non-rigid deformations. Obviously, camera direction is an important imaging parameter which has not been considered in this thesis. We think this subject should be studied in depth and exploited, starting from the work presented here.

- **Applications.** Some of the ideas proposed in this thesis can be used in different Computer Vision applications. We would like to mention some possible applications closely related to the research developed within the Computer Vision and Robotics group at the University of Girona. Considering two of the basic research lines of this group, 3D perception and Underwater robotics, it could be interesting to study the use of photometric stereo in order to achieve 3D reconstruction. Moreover, and more ambitiously, we could attempt to work it into an underwater reconstruction basically focusing on the idea of recovering distances and depth maps of the underwater floor.

7.3 Related Publications

Ideas and projects of this work have evolved from initial stages to obtain the final version presented in this thesis. The following publications are a direct consequence of the research carried out during the elaboration of this thesis and give an idea of the progress achieved.

Journals

- X. Lladó, M. Petrou, and J. Martí. Surface texture recognition by surface rendering. *IEEE Transactions on Image Processing* **IEEE T-IP**, in second revision 2003.

International Conferences

- X. Lladó, A. Oliver, M. Petrou, J. Freixenet, and J. Martí. Simultaneous surface texture classification and illumination tilt angle prediction. In *British Machine Vision Conference* **BMVC 2003**, volume 2, pages 789-798, Norwich, England, September 2003.

- X. Lladó, J. Martí, and M. Petrou. Classification of textures seen from different distances and under varying illumination direction. In *IEEE International Conference on Image Processing* **ICIP 2003**, Barcelona, Spain, September 2003.

- X. Lladó, and M. Petrou. Classifying textures when seen from different distances. In *IAPR International Conference on Pattern Recognition* **ICPR 2002**. Quebec, Canada, August 2002.

- X. Lladó, and M. Petrou. Predicting surface texture when seen from different distances. In *International Workshop on Texture Analysis and Synthesis* **ECCV-Texture 2002**, pages 83-86, Copenhagen, Denmark, June 2002.

- X. Lladó, J. Martí, and M. Petrou. Image texture prediction using colour photometric stereo. In *Lecture Notes in Artificial Intelligence* **LNAI 2002**, pages 355-363. Springer-Verlag, 2002.

- X. Lladó, J. Pagès, J. Martí, J. Batlle, and C. Dragoste. Quality control by using surface shape analysis. In *IEEE-TTTC International Conference on Quality Control, Automation and Robotics* **QCAR 2002**, volume 1, pages 444-449, Cluj-Napoca, Romania, May 2002.

- J. Freixenet, X. Lladó, J. Martí and X. Cufí. Use of decision trees in colour feature selection. Application to object recognition in outdoor scenes". In *IEEE International Conference on Image Processing* **ICIP 2000**, volume 3, pages 496-499, Vancouver, Canada, September 2000.

- X. Lladó, J. Martí, J. Freixenet, and Ll. Pacheco. A novel criterion function in feature evaluation. Application to the classification of corks. In *International Conference on Quality Control, Automation and Robotics* **QCAR 2000**, volume 2, pages 275-280. Cluj-Napoca, Romania, May 2000.

National Conferences

- X. Lladó, J. Martí, J. Freixenet, and J. Batlle. Automated industrial inspection using photometric stereo. In *Catalonian Conference on Artificial Intelligence* **CCIA 2001**, pages 269-275, Barcelona, Spain, October 2001.

- J. Martí, J. Freixenet, X. Lladó, X. Muñoz, and X. Cufí. A new approach to natural object labelling in colour aerial images. In *Spanish Symposium on Pattern Recognition and Image Analysis* **SNRFAI 2001**. volume 1, pages 67-72, Benicàssim, Spain, May 2001.

- X. Lladó, J. Martí. J. Freixenet, and V. Ila. Object characterization in outdoor scene analysis. In *Spanish Symposium on Pattern Recognition and Image Analysis* **SNRFAI 2001**, volume 2, pages 113-118, Benicàssim, Spain, May 2001.

- X. Lladó, J. Martí, J. Freixenet, and X. Muñoz. Feature selection using genetic algorithms. In *Catalonian Conference on Artificial Intelligence* **CCIA 2000**, pages 152-156, Vilanova i la Geltrú, Spain, October 2000.

- X. Lladó, J. Freixenet, J. Martí, J. Forest, and J. Salvi. Object recognition in outdoor scene using a bottom-up strategy. In *Catalonian Conference on Artificial Intelligence* **CCIA 1999**, pages 299-307, Girona, Spain, October 1999.

Technical Reports

- X. Lladó, J. Martí, J. Freixenet, and J. Batlle. A review of feature selection and feature evaluation. Technical Report 02-19-RR, Institute of Informatics and Applications. University of Girona, 2002.

- X. Lladó, and J. Martí. Surface texture prediction using colour photometric stereo. Technical Report 01-14-RR, Institute of Informatics and Applications. University of Girona, 2001.

Appendix A

Simultaneous Surface Texture Classification and Illumination Tilt Angle Prediction

In this Appendix we investigate the effect of the illuminant tilt rotation over surface textures by analysing a set of image texture features extracted from the co-occurrence matrix. From the behaviour of each texture feature, a simple method, able to predict the illuminant tilt angle of test images, is developed. Moreover, the method is also used to perform a texture classification invariant to the illuminant tilt angle rotation. This study, and experimental results over different real textures, show that the illumination tilt angle can be accurately predicted as part of the texture classification process.

A.1 Introduction

Very little work has been published on the topic of illumination invariant texture classification. One strategy to solve this problem is to study the immediate effects introduced by light direction on the observed 2D texture. This was done recently by Chantler et al. [23] who presented a formal theory which demonstrates that changes in the tilt of the light direction make texture features follow super-elliptical trajectories in multi-dimensional feature spaces. Based on this work, Penirschke et al. [125] developed an illuminant rotation invariant classification scheme which

uses photometric stereo for the detection of surface relief and Gabor features for feature extraction. In other applications of Computer Vision, a correct estimation of light can play an important role. For instance, light estimation is apparent in all applications which use photometric measurements which obviously depend on light. Weber and Cipolla [167] focus their attention on reconstruction problems and the estimation of light-source.

In this Appendix, we investigate the effect of the illuminant tilt rotation over surface textures by analysing a set of texture features extracted from the co-occurrence matrix. From the behaviour of each texture feature, we develop a simple method able to predict the illuminant tilt angle of unknown test images. Moreover, we use this prediction method to perform the classification of textures under varying illuminant tilt angles.

The remainder of this Appendix is organised as follows. In section A.2 we analyse the behaviour of texture features in the feature space. The method to predict the illuminant tilt angle is explained in section A.3. In section A.4 the performace of the method is evaluated in two experimental trials: illuminant tilt angle prediction and texture classification. Finally, the Appendix ends with conclusions and further work.

A.2 Co-occurrence Matrix Behaviour

The co-occurrence matrix [61] is a well known method used to extract texture features. In this work, the co-ocurrence matrices are implemented in an anisotropic way. That is, we analyse 4 different directions: 0°, 45°, 90°, and 135° computing for each the contrast feature for a fixed distance 1.

A set of 15 different textures have been used throughout the experimental trials presented in this work. Figure A.1 shows one image of each texture captured under a specific direction of light. Note that almost all are isotropic textures (see the first 10 textures). Therefore, the only anisotropy in the image is induced by the anisotropic light. It is this anisotropy we wish to detect in order to identify the light direction.

Figure A.1: One image of each of the fifteen sample textures.

We have included in the database some anisotropic textures for comparison. In order to detect anisotropy in an image we use the rotationally sensitive co-occurrence matrix from which we construct features.

Figures A.2.(a), A.2.(c) and A.2.(e) illustrate the behaviour of these 4 features for textures 1, 9, and 15 respectively. Each plot shows how one output of one contrast feature varies when it is applied to the same texture sample, but under varying illuminant tilt angles (we use steps of 30° degrees from 0° to 360°).

Note that the graph of the anisotropic texture 15, which consists of long rough structures (see figure A.1), shows some strange behaviour at first glance: three of the features behave in a very similar way (i.e. note the three similar curves in figure A.2.(e)) while one feature is totally constant. The catastrophic behaviour of this feature is due to the fact that the direction used to compute the corresponding co-occurence matrix coincides with the direction of the long structures which make up the imaged surface. In this particular case, the images captured under all the tilt angles have no significant intensity changes with respect to the axis of the local structures. Therefore, we obtain constant feature values for all tilt angles. On the other hand, when the long rough structures are lit from lateral directions, we expect some invariant behaviour as one side of these structures remains lit through a broad range of incident angles. This is the reason three of the features behave in a very similar way.

Moreover, observing the feature behaviour for isotropic textures, we can see that the contrast feature at distance 1 has a symmetric behaviour approximately between ranges $[0°, 180°]$, and $[180°, 360°]$. Therefore, it is difficult to distinguish whether the illuminant tilt angle is in one or the other range. That is, by measuring one of these features, which is tilt angle sensitive, we may be able to identify the illuminant tilt angle only up to an ambiguity of $\pm 180°$.

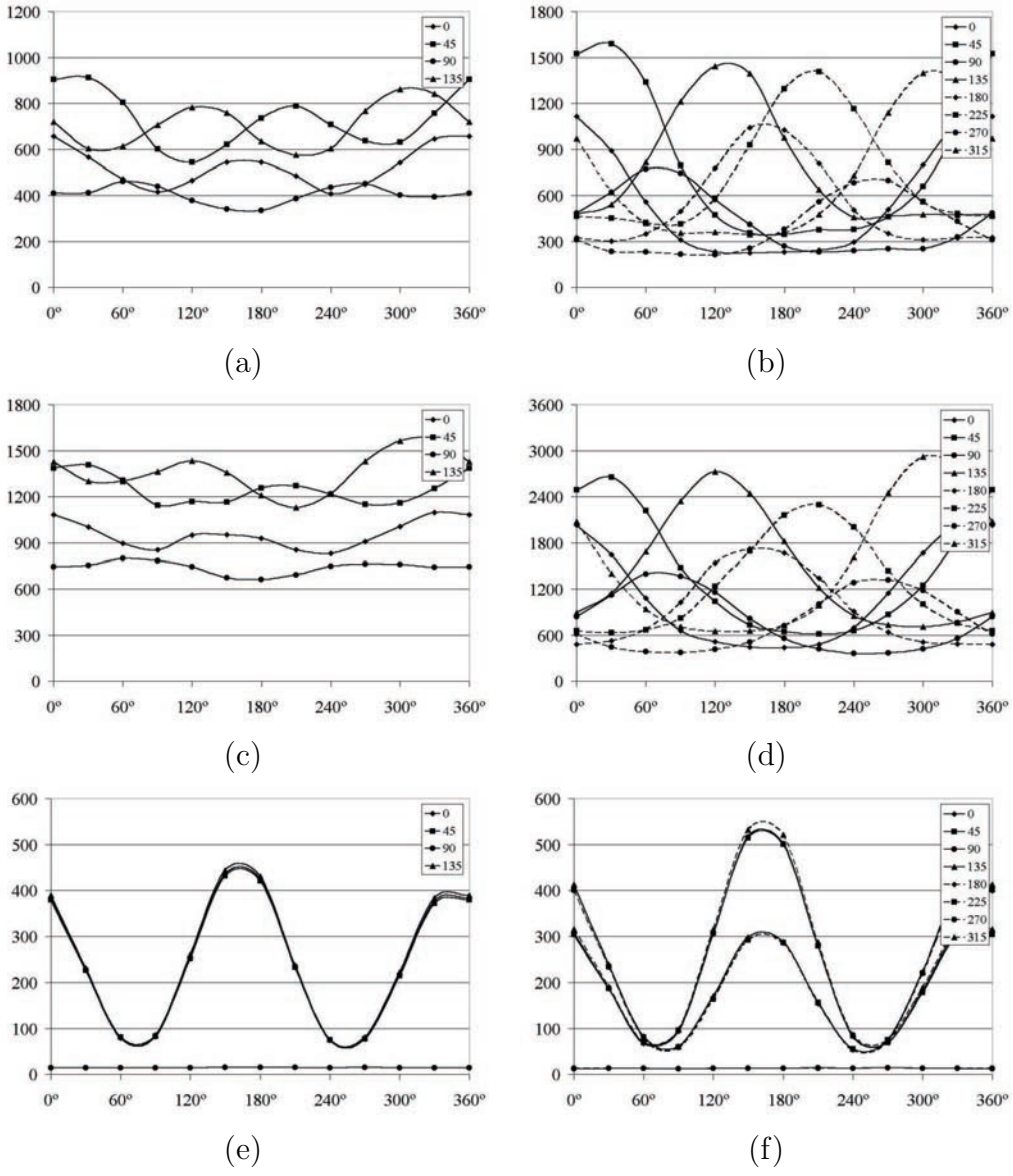


Figure A.2: Feature behaviour for isotropic textures 1 (first row), and 9 (second row), and anisotropic texture 15 (third row). Each plot shows how one output of one feature varies when it is applied to the same physical texture sample, but under varying illuminant tilt angles. (a), (c), and (e) Co-ocurrence matrices computed for 4 different directions. (b), (d), and (f) Co-ocurrence matrices computed for 8 different directions.

A.2.1 Removing the Ambiguity from the Estimation of the Tilt Angle

With the aim of improving the illuminant tilt angle prediction, allowing the estimation over the whole tilt rotation of 360°, we propose to introduce the use of 4 new directions in the computation of the co-occurrence matrix. Therefore, we compute the co-occurrence matrix using 8 different directions: 0°, 45°, 90°, 135°, 180°, 225°, 270° and 315° . Note that in this approach the co-occurrence matrix obtained for 180°, 225°, 270°, and 315° are the transposed matrices of 0°, 45°, 90°, and 135° respectively. Hence, computing the classical contrast feature for these 8 matrices we only obtain 4 different values since the contrast feature gives us the same value for a matrix and its transposed.

As our objective is to distinguish the sense of the directions used in the co-occurence matrix, we propose to compute the contrast feature from the upper triangular matrix only. We do that with the idea of counting the pairs of pixels in which the intensity value increases (transitions of darker pixels to brighter pixels).

This can be seen in figures A.2.(b) and A.2.(d) where we plot the variation of these 8 new features for the isotropic textures 1 and 9. Note that for each tilt angle, the maximum value of the contrast feature is attained by the feature which has been computed from the co-occurrence matrix with the same orientation angle as the tilt angle. This is because we always find more transitions of darker pixels to brighter pixels when the orientation of the light source "coincides" with the orientation of the co-occurence matrix. It is important to clarify that we use the term "coincide" when we refer to the same angle, but under two reference systems: one which defines the direction used in the co-occurence matrix, and one which defines the direction of the incident light.

In contrast, the anisotropic textures do not follow this behavior (see figure A.2.(f)), although they still exhibit an approximately symmetric behaviour over the two ranges of values: namely $[0°, 180°]$, and $[180°, 360°]$.

The feature behaviour described in this section will be used as a model to predict the tilt angle of an unknown texture image. Each texture model is created using a

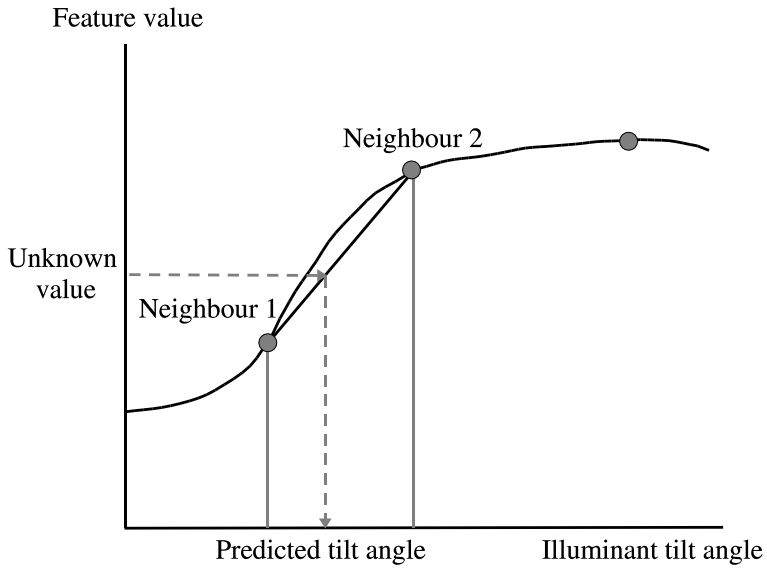


Figure A.3: Prediction scheme. In a first step, we find the two feature vectors (neighbour 1 and 2) closest to the analysed feature vector. Then, the tilt angle is obtained in the current interval applying a linear interpolation individually for each single feature.

small number of different tilt angles. Specifically, we have used 12 tilt angles 30° apart. Hence, a texture model is composed of 12 vectors of 8 features each (12 tilt angles, 8 directional co-occurrence matrices for each).

A.3 Illuminant Tilt Angle Prediction

The process of predicting the illuminant tilt angle, given an unkown test image, starts by extracting a representative feature vector for this image which implies the computation of 8 texture features. Note that the model of each texture is composed of 12 different feature vectors, one for each reference tilt angle.

When the feature vector for the unknown image is obtained, the prediction consists of looking for the most similar feature vector among the model textures. Then, a simple method of three steps is proposed:

- First, we obtain a first approximation of the predicted angle with one of the known angles used to describe the model texture. The nearest neighbour

classifier is used to find the closest feature vector.

- After that, we localise the angle interval which contains the test feature vector. We use again the nearest neighbour classifier to find the second most similar feature vector of the same model texture (see figure A.3). In this second step we ensure that the angle interval is of 30°. This means that the second known angle is ±30° with respect to the first approximation.

- Next, the exact tilt angle is found in the current interval applying linear interpolation for each single feature. The results are averaged to produce the final predicted tilt angle. The tilt angles computed from individual features for which the linear interpolation provides slopes close to zero are not considered in the averaging process.

A.4 Experimental Trials

The proposed prediction method was tested with 15 different textures. The first 10 were isotropic textures, the remaining five were anisotropic. For each one, 4 complete sets of 24 images corresponding to the illuminant tilt angles between 0° and 360° incremented in steps of 15° are available. All these images are lit at an elevation angle of 55°. From these images we create 2 different sets: one used for modeling the light behaviour, and another for testing the classification process.

The first set is composed of 15 textures×12 light directions = 180 images. The 12 illuminant tilt angles used for modeling are 0°, 30°, 60°, 90°, 120°, 150°, 180°, 210°, 240°, 270°, 300°, and 330°. On the other hand, the testing set is composed of 15 textures×12 light directions×3 images = 540 test images. The 12 illuminant tilt angles used for testing are different from the reference tilt angles mentioned above. Figure A.4 illustrates the configuration used in our experimental trials.

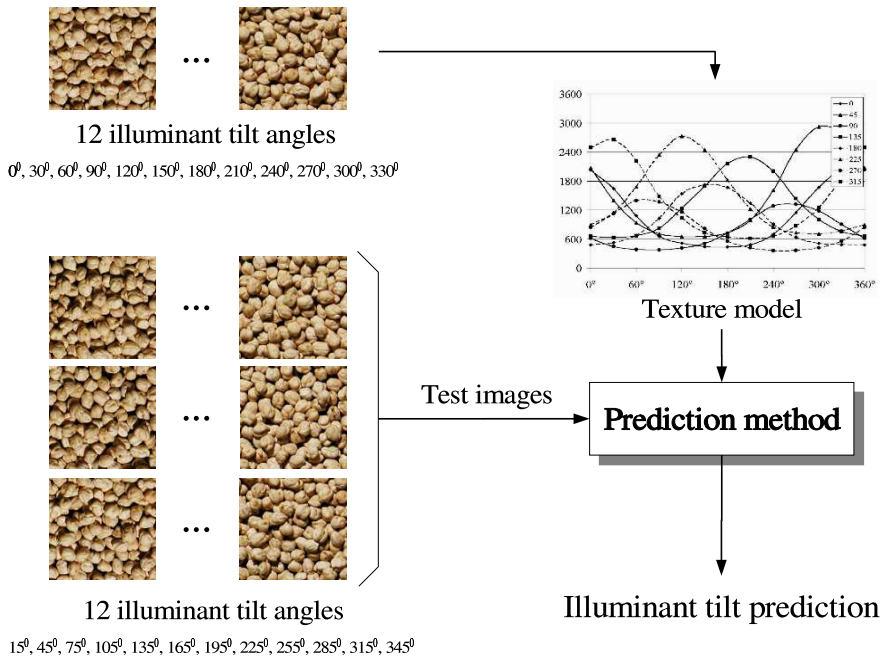


Figure A.4: Experimental setup. Four different surface patches are available for each texture. One is used for modelling the feature behaviour, the remaining ones for testing.

A.4.1 Accuracy of Tilt Angle Prediction

The purpose of this experiment is to evaluate the accuracy of the illuminant tilt angle prediction. After computing the model behaviour, we individually apply for each texture, the prediction method for all the corresponding test images.

Figure A.5 shows the error distributions for textures 1, 9, and 15 of the absolute tilt angle difference between our predictions and the correct values. Each plot has been computed for the 36 test images available for each texture and shows the percentage of times a particular error value was observed. Note that for isotropic textures 1 and 9, we predict the tilt angle for almost all the test images with an error of a few degrees. However, for texture 15, the errors are significantly larger.

From the error distributions of all 15 textures we conclude that for isotropic textures the illuminant tilt angle may be predicted quite accurately. However, poor results are obtained for anisotropic textures. We confirm this conclusion providing

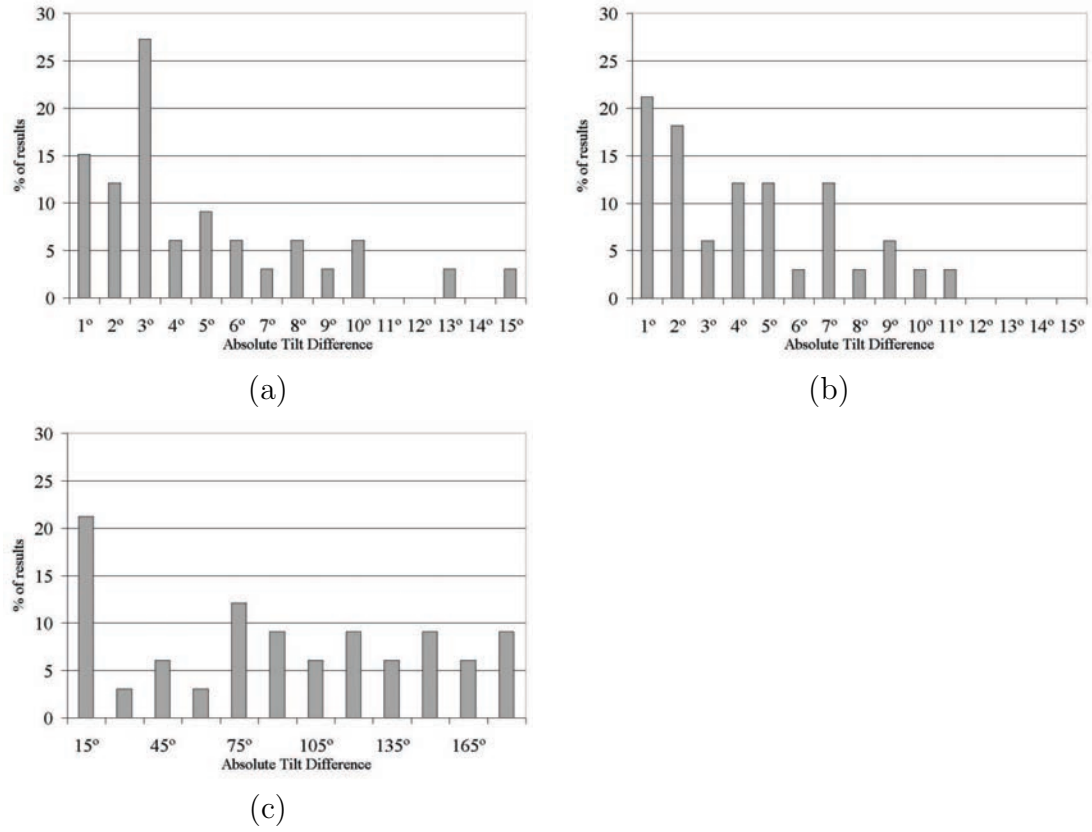


Figure A.5: Error in the tilt angle prediction for textures 1, 9, and 15, (a), (b), and (c) respectively.

an overall quantitative assessment over all these predictions (see table A.1). We have computed over all the textures the average MSE of the tilt angle prediction and its standard deviation. In the same table we present a quantitative assessment for the isotropic and anisotropic textures separately. It is important to note that we predict the tilt angle for isotropic textures with an average error of 6°. Nevertheless, the prediction error increases when anisotropic textures are considered. This is because the anisotropy of the image cannot be solely attributed to the light direction and, therefore, the detected anisotropy does not give us the clear and unambiguous information which is needed by our prediction method.

Table A.1: Overall quantitative assessment over all 15 textures of illuminant tilt angle predictions. Average MSE and its standard deviation for the tilt angle difference between the predicted and true values.

Isotropic		Anisotropic		Overall	
Avg	Std	Avg	Std	Avg	Std
5.96	1.85	59.61	22.57	23.84	28.86

A.4.2 Accuracy of Texture Classification

This experiment analyses the accuracy of the texture classification when our feature behaviour models are used as references for classification.

The method described in section A.3 may be used not only to predict the illuminant tilt angle of a test image, but also to classify the unkown test image into one of the texture classes present in the database. Specifically, the first step in which the nearest neighbour classifier is used to find the closest feature vector to the model, allows us to perform texture classification as well.

Table A.2 summarises the obtained texture classification results when all fifteen models are used in the classification process. The texture classification accuracy is 82.63%, while the illuminant tilt angle is predicted with an average MSE of 24.04° and standard deviation of 43.07°. We have repeated the same experiment, but using only the isotropic textures, achieving a texture classification accuracy of 79.09%. In this case, the illuminant tilt angle is predicted with an average MSE of 5.88° and standard deviation of 4.80°. Note that when using all fifteen textures we obtain better classification results than those using only isotropic textures. That is because isotropic textures have similar feature behaviour. For instance, test images of textures 2 and 4 have been misclassified. In contrast, as is shown in figure A.2.(f), anisotropic textures have different feature behaviour. This fact causes an improvement in the classification rate, when anisotropic textures are included in the reported results. However, it is important to notice that accurate tilt angle predictions are only obtained for isotropic textures.

Table A.2: Texture classification rates and MSE of the tilt angle prediction obtained over the isotropic textures and over all fifteen textures.

	Texture classification	Tilt angle MSE	
		Avg	Std
Isotropic textures	79.09%	5.85	4.80
All textures	82.63%	24.06	43.07

A.5 Conclusions

We presented a simple method able to predict the illuminant tilt angle of unknown test images. This method is based on behaviour models of texture features extracted from the co-occurrence matrix. It works under the assumption that the only anisotropy in the image is induced by the anisotropic light. The experimental results over different real textures, including some anisotropic textures for comparison, show that the illumination tilt angle can be accurately predicted. As well as predicting the illuminant tilt angle, this method is used to perform texture classification. The results show that anisotropic textures may be classified more accurately but their illuminat tilt angle may not be predicted so well, while isotropic textures cause more confusion to the classifier but allow us to predict the direction from which the imaged surfaces were lit very accurately, as long as reference images for 12 different tilt angles of each surface are available. Such reference images may be created either by direct image capture when creating the reference database, or from surface and colour information recovered by 4 source colour photometric stereo [8] and subsequent image rendering [104].

Bibliography

[1] Maestex database. *http://www.cssip.uq.edu.au/staff/meastex/meastex.html*.

[2] Photometric image databases. Photometric Image Databases maintained by the Texture lab. of the Heriot-Watt University, Edinburgh. *http://www.cee.hw.ac.uk/texturelab/database/*.

[3] Vistex database. Vision Texture database maintained by the Vision and Modelling group at the Massachusetts Institute of Technology, MIT Media Lab. *http://www-white.media.mit.edu/vismod/imagery/VisionTexture/vistex.html*.

[4] A. Al-Janobi. Performance evaluation of cross-diagonal texture matrix method of texture analysis. *Pattern Recognition*, 34:171–180, 2001.

[5] J.L. Barron, D.J. Fleet, S.S. Beauchemin, and T.N. Burkitt. Performance of optical flow techniques. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 236–242, Champaigne, Illinois, USA, 1992.

[6] S. Barsky and M. Petrou. Classification of 3d rough surfaces using colour and gradient information recovered by colour photometric stereo. *SPIE Proceedings. Visualization and Optimization Techniques*, 4553:10–19, October 2001.

[7] S. Barsky and M. Petrou. Colour photometric stereo: Simultaneous reconstruction of local gradient and colour of rough textured surfaces. In *International Conference on Computer Vision*, volume II, pages 600–605, 2001.

[8] S. Barsky and M. Petrou. The 4-source photometric stereo technique for 3-dimensional surfaces in the presence of highlights and shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, November 2003.

[9] R. Basri and D.W. Jacobs. Photometric stereo with general, unknown lighting. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages II:374–381, 2001.

[10] J. Batlle, A. Casals, J. Freixenet, and J. Martí. A review on strategies for recognizing natural objects in colour images of outdoor scenes. *Image and Vision Computing*, 18(6–7), May 2000.

[11] B. G. Becker and N.L. Max. Smooth transitions between bump rendering algorithms. In *Computer Graphics SIGGRAPH*, pages 183–190, Anaheim, CA, 1993.

[12] M. Ben-Bassat. *Pattern recognition and reduction of dimensionality. In: Handbook of Statistics*. P.R. Krishnaiah and L.N. Kanal, North Holland, 1982.

[13] J.M. Bennet and L. Mattson. *Introduction to surface roughness and scattering*. Optical Society of America, Washington D.C., 1989.

[14] J.F. Blinn. Models of light reflection for computer synthesized pictures. In *Computer Graphics SIGGRAPH*, volume 11, pages 192–198, 1977.

[15] J.F. Blinn. Simulation of wrinkled surfaces. In *Computer Graphics SIGGRAPH*, pages 286–292, 1978.

[16] A. L. Blum and P. Langley. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1–2):245–271, 1997.

[17] A. Bradley, P. Jackway, and B. Lovell. Classification in scale-space: Applications to texture analysis. In *Information Processing in Medical Imaging*, page 375.

[18] P. Brodatz. *Textures, A photographic album for artists and designers*. Dover Publications, New York, 1966.

[19] M.W. Burke. *Image Acquisition: Handbook of Machine Vision Engineering.* Volume 1. New York. Chapman and Hall, 1996.

[20] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: Color and texture-based image segmentation using em and its application to content-based image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1026–1038, August 2002.

[21] M.J. Chantler. *The effect of variation in illuminant direction on texture classification. Ph.D. thesis, Dept. Computing and Electrical Engineering, Heriot-Watt University.* PhD thesis, Department of Computing and Electrical Engineering, Heriot-Watt University, August 1994.

[22] M.J. Chantler. Why illuminant direction is fundamental to texture analysis. *IEE Proceedings-Vision Image and Signal Processing*, 142(4):199–206, August 1995.

[23] M.J. Chantler, M. Schmidt, M. Petrou, and McGunnigle G. The effect of illuminant rotation on texture filters: Lissajous's ellipses. In *European Conference on Computer Vision*, volume 3, pages 289–304, Copenhagen, Denmark, May 2002.

[24] M.J. Chantler and J. Wu. Rotation invariant classification of 3d surface textures using photometric stereo and surface magnitude spectra. In *British Machine Vision Conference*, pages 486–495, 2000.

[25] R. Chellappa and S. Chatterjee. Classification of textures using gaussian markov random fields. *IEEE Transactions on Acoustics Speech and Signal Processing*, 33:959–963, 1985.

[26] C. Chen, L. Pau, and P. Wank. *Handbook of Pattern Recognition and Computer Vision.* World Scientific, 1992.

[27] P.H. Christensen and L.G. Shapiro. Three-dimensional shape from colour photometric stereo. *International Journal of Computer Vision*, 13(2):213–227, 1994.

[28] J.J. Clark. Active photometric stereo. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 29–34, 1992.

[29] F.S. Cohen, Z. Fan, and M.A. Patel. Classification of rotated and scaled textured images using gaussian markov random field models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:192–202, 1991.

[30] E.N. Coleman and R. Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Computer Graphics and Image Processing*, 18(4):309–328, April 1982.

[31] R.W. Conners and C.A. Harlow. A theoretical comparison of texture algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(3):204–222, May 1980.

[32] O.G. Cula and K.J. Dana. Recognition methods for 3d textured surfaces. In *SPIE Conference on Human Vision and Electronic Imaging VI*, volume 4299, pages 209–220, January 2001.

[33] K.J. Dana and S.K. Nayar. Histogram model for 3d textures. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 618–624, June 1998.

[34] K.J. Dana and S.K. Nayar. Correlation model for 3d texture. In *International Conference on Computer Vision*, pages 1061–1067, 1999.

[35] K.J. Dana, B. Van Ginneken, S.K. Nayar, and J.J. Koenderink. Reflectance and texture of real-world surfaces. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 151–157, 1997.

[36] K.J. Dana, B. Van Ginneken, S.K. Nayar, and J.J. Koenderink. Reflectance and texture of real-world surfaces. *ACM Transactions on Graphics*, 18(1):1–34, January 1999.

[37] M. Dash and H. Liu. Feature selection for classification. *Intelligent Data Analysis*, 1(3), 1997.

[38] J. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America*, 2(7):1160–69, July 1985.

[39] P. A. Devijver and J. V. Kittler. *Pattern Recognition. A Statistical Approach*. Prentice-Hall, Englewood Cliffs, NJ, 1982.

[40] J. Doak. An evaluation of feature selection methods and their application to computer security. Technical report, Davis, CA: University of California, Department of Computer Science, 1992.

[41] J. Dong and M.J. Chantler. Capture and synthesis of 3d surface texture. *International Workshop on Texture Analysis and Synthesis*, pages 41–45, June 2002.

[42] M.S. Drew. Shape from color. Technical Report CSS/LCCR TR 92-07, Simon Fraser University School of Computing Science, 1992.

[43] M.S. Drew. Optimization approach to dichromatic images. *Journal of Mathematical Imaging and Vision*, 3:189–205, 1993.

[44] M.S. Drew. Direct solution of orientation-from-color problem using a modification of pentland's light source direction estimator. *Computer Vision and Image Understanding*, 64:286–299, 1996.

[45] M.S. Drew. Photometric stereo without multiple images. *Human Vision and Electronic Imaging*, 3016:369–308, February 1997.

[46] J.M.H. Du Buf, M. Kardan, and M. Spann. Texture feature performance for image segmentation. *Pattern Recognition*, 23(3/4):291–309, 1990.

[47] O. Duda and P. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, Inc, 1973.

[48] O.D. Faugeras. *Three-Dimensional Computer Vision*. The MIT Press, Cambridge, Massachusetts, 1993.

[49] P. Favaro and S. Soatto. Learning shape from defocus. In *Proceedings of European Conference on Computer Vision*, volume 2 of *Lecture Notes in Computer Science*, pages 735–745, Copenhagen, Denmark, 2002. Springer.

[50] G.D. Finlayson, J. Dueck, B.V. Funt, and M.S. Drew. Colour eigenfaces. In *International Workshop on Image and Signal Processing*, Manchester, England, November 4-7 1996.

[51] D. Forsyth and J. Ponce. *Computer Vision - A modern approach*. Prentice Hall, Upper Saddle River, N.J., 2002.

[52] S. Fountain, T. Tan, and K. Baker. A comparative study of rotation invariant classification and retrieval of texture images. In *British Machine Vision Conference*, 1998.

[53] R.T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10:439–451, 1988.

[54] S. Fukuda and H. Hirosawa. A wavelet-based texture feature set applied to classification of multifrequency polarimetric sar images. *IEEE Transactions on Geoscience and Remote Sensing*, 37:2282–2286, May 1999.

[55] E. S. Gelsema. Special issue on genetic algorithms. *Pattern Recognition Letters*, 16(8), 1995.

[56] A. Gonzalez. Model-based texture classification under varying illumination. Technical report, Research Memorandum RM/02/8, Dept. of Computing and Electrical Engineering. Heriot-Watt University, Edinburgh, September 2002.

[57] A. Grau. *Mètode d'Extracció Multiparamètrica de Característiques de Textura Orientat a la Segmentació d'Imatges*. PhD thesis, Departament d'ESAII-UPC, Universitat Politècnica de Catalunya, May 1997.

[58] G.M. Haley and B.S. Manjunath. Rotation-invariant texture classification using modified gabor filters. In *IEEE International Conference on Image Processing*, pages 262–265, 1995.

[59] G.M. Haley and B.S. Manjunath. Rotation-invariant texture classification using a complete space-frequency model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(2):255–269, February 1999.

[60] R.M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67:786–804, 1979.

[61] R.M. Haralick, K.S. Shanmugan, and I. Dunstein. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6):610–621, 1973.

[62] X.D. He, K.E. Torrance, F.X. Sillion, and D.P. Greenberg. A comprehensive physical model for light reflection. *Computer Graphics*, 25(4):175–186, 1991.

[63] G. Healey and T.O. Binford. Local shape from specularity. *Computer Vision, Graphics and Image Processing*, 42:62–86, 1988.

[64] G. Healey and R. Jain. Depth recovery from surface normals. In *IAPR International Conference on Pattern Recognition*, pages 894–896, July 1986. Montreal, Canada.

[65] K.P. Horn. *Obtaining shape from shading information,* in The Psychology of Computer Vision. P.H. Winston, Ed. McGraw-Hill, New York, 1975.

[66] K.P. Horn. Understanding image intensities. *Artificial Intelligence*, 8:201–238, 1977.

[67] K.P. Horn. Height and gradient from shading. *International Journal of Computer Vision*, 5:37–75, 1990.

[68] K.P. Horn and M.K. Brooks. The variational approach to shape from shading. *Computer Vision, Graphics and Image Processing*, 33:174–208, 1986.

[69] K. Ikeuchi. Determining surface orientation of specular surfaces by using the photometric stereo method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3(6):661–669, November 1981.

[70] K. Ikeuchi. Determining a depth map using a dual photometric stereo. *International Journal of Robotics Research*, 6(1):15–31, 1987.

[71] Y. Iwahori, Y. Watanabe, R.J. Woodham, and A. Iwata. Self-calibration and neural network implementation of photometric stereo. In *IAPR International Conference on Pattern Recognition*, pages IV: 359–362, 2002.

[72] Y. Iwahori, R. Woodham, M. Bhuiyan, and N. Ishii. Neural network based photometric stereo for object with non-uniform reflectance factor. In *International Conference on Neural Information Processing*, volume III, pages 1213–1218, 1999.

[73] A.K. Jain, R. P.W. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37, January 2000.

[74] A.K. Jain and F. Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, 24:1167–1186, December 1991.

[75] A.K. Jain and D. Zongker. Feature selection: Evaluation, application, and small sample performace. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):153–158, February 1997.

[76] R. Jain, R. Kasturi, and B.G. Schunck. *Machine Vision*. McGraw-Hill, 1995.

[77] G Kay and T. Caelli. Estimating the parameters of an illumination model using photometric stereo. *Graphical Models and Image Processing*, (5):365–388, September 1995.

[78] S.C. Kee, K.M. Lee, and S.U. Lee. Illumination invariant face recognition using photometric stereo. *Machine Vision and Applications*, (7):1466–1474, July 1999.

[79] B. Kim and P. Burger. Depth and shape from shading using the photometric stereo method. *Computer Vision, Graphics and Image Processing*, 54(3):416–427, November 1991.

[80] B. Kim and R. Park. Shape from shading and photometric stereo using surface approximation by legendre polynomials. *Computer Vision and Image Understanding*, 66(3):255–270, June 1997.

[81] J. Kittler. Feature set search algorithms. In *Pattern Recognition and Signal Processing*, pages 41–60, Sijthoff and Noordhoff, Alphen aan den Rijn, The Netherlands, 1978.

[82] J. Kittler. Feature selection and extraction. In T. Y. Young and K. S. Fu, editors, *Handbook of Pattern Recognition and Image Processing*, pages 59–83, Orlando, FL, 1986. Academic Press.

[83] R. Klette and K. Schlüns. Height data from gradient fields. *SPIE Proceedings Machine Vision Applications, Architectures, and Systems Integration*, 2908:204–215, November 1996.

[84] G.J. Klinker. *A Physical Approach to Color Image Understanding*. A.K. Peters, 1993.

[85] J.J. Koenderink and A.J. Van Doorn. Illuminance texture due to surface mesostructure. *Journal of the Optical Society of America - Optics Image Science and Vision*, 13(3):452–463, 1996.

[86] R. Kohavi and G. H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1–2):273–323, 1997.

[87] L.L. Kontsevich, A.P. Petrov, and I.S. Vergelskaya. Reconstruction of shape from shading in colour images. *Journal of the Optical Society of America*, 11(3):1047–1052, March 1994.

[88] M. Kudo and J. Sklansky. Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition*, 33(1):25–41, 2000.

[89] L. I. Kuncheva and L. C. Jain. Nearest neighbor classifier: Simultaneous editing and feature selection. *Pattern Recognition Letters*, 20:1149–1156, 1999.

[90] K.I. Laws. Textured image segmentation. Technical Report 940, Image Processing Institute, University of Southern California, Los Angeles, California, 1980.

[91] S. Lazebnik, C. Schmid, and J. Ponce. Affine-invariant local descriptors and neighborhood statistics for texture recognition. In *International Conference on Computer Vision*, Nice, France, October 2003.

[92] S. Lazebnik, C. Schmid, and J. Ponce. A sparse texture representation using affine-invariant regions. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume II, pages 319–324, Madison, WI., June 2003.

[93] K.M. Lee and C.C.J. Kuo. Shape reconstruction from photometric stereo. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 479–484, 1992.

[94] K.M. Lee and C.C.J. Kuo. Shape from shading with a generalized reflectance map model. *Computer Vision and Image Understanding*, 67(2):143–160, August 1997.

[95] S.W. Lee and R. Bajcsy. Detection of specularity using color and multiple views. *Image and Vision Computing*, 10:643–653, 1992.

[96] M. Leung and A.M. Peterson. Scale and rotation invariant texture classification. In *International Conference on Acoustics, Speech and Signal Processing*, pages 461–465, 1992.

[97] T. Leung and J. Malik. On perpendicular texture: why do we see more flowers in the distance? In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 807–813, June 1997.

[98] T. Leung and J. Malik. Recognising surfaces using three-dimensional textons. In *International Conference on Computer Vision*, pages 1010–1017, September 1999.

[99] T. Leung and J. Malik. Representing and recognising the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1):29–44, June 2001.

[100] X. Liu, Y. Yu, and H. Shum. Synthesizing bidirectional texture functions for real-world surfaces. In *SIGGRAPH'2001*, pages 97–106, Los Angeles, CA, August 2001.

[101] X. Lladó, J. Martí, and M. Petrou. Surface texture prediction using colour photometric stereo. Technical Report 01-14-RR, Institute of Informatics and Applications. University of Girona, 2001.

[102] X. Lladó, J. Pagès, J. Martí, J. Batlle, and C. Dragoste. Quality control by using surface shape analysis. In *International Conference on Quality Control, Automation and Robotics*, Cluj-Napoca, Romania, May 2002.

[103] X. Lladó and M. Petrou. Classifying textures when seen from different distances. In *IAPR International Conference on Pattern Recognition*, pages 83–86, August 2002.

[104] X. Lladó, M. Petrou, and J. Martí. Surface texture recognition by surface rendering. *IEEE Transactions on Image Processing*, in second revision, 2003.

[105] F. Lumbreras, R. Baldrich, M. Vanrell, J. Serrat, and J.J. Villanueva. Multiresolution colour texture representation for tile classification. In *Simposium Nacional en Reconocimiento de Formas y Análisis de Imagen*, volume I, pages 145–152, Bilbao, Spain, May 1999.

[106] J. Malik and P. Perona. Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America*, 7(5):923–932, 1990.

[107] B.S. Manjunath and R. Chellappa. Unsupervised texture segmentation using markov random field models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(5):478–482, 1991.

[108] R. Manthalkar, P.K. Biswas, and B.N. Chatterji. Rotation and scale invariant texture classification using gabor wavelets. In *International Workshop on Texture Analysis and Synthesis*, pages 87–90, Copenhagen, Denmark, June 2002.

[109] J. Mao and A.K. Jain. Texture classification and segmentation using multiresolution simultaneous autoregressive models. *Pattern Recognition*, 25(2):173–182, 1992.

[110] J. Martí, J. Freixenet, J. Batlle, and A. Casals. A new approach to outdoor scene description based on learning and top-down segmentation. *Image and Vision Computing*, 19(14):1041–1055, December 2001.

[111] G. McGunnigle. *The Classification of Textured Surfaces Under Varying Illuminant Direction. Ph.D. thesis, Dept. Computing and Electrical Engineering, Heriot-Watt University.* PhD thesis, Department of Computing and Electrical Engineering, Heriot-Watt University, June 1998.

[112] G. McGunnigle and M.J. Chantler. Rotation invariant classification of rough surfaces. *IEE Proceedings-Vision Image and Signal Processing*, 146(6), December 1999.

[113] G. McGunnigle and M.J. Chantler. Rough surface classification using point statistics from photometric stereo. *Pattern Recognition Letters*, 21:593–604, 2000.

[114] E. Micheli-Tzanakou. *Supervised and Unsupervised Pattern Recognition. Feature extraction and computational intelligence.* CRC Press. New York, Washington, D.C., 2000.

[115] A. J. Miller. *Subset Selection in Regression.* Chapman and Hall, 1990. Mos H517.23 M647S.

[116] X. Muñoz. *Image segmentation integrating colour, texture and boundary information.* PhD thesis, Department of Electronics, Informatics and Automation. University of Girona, February 2003.

[117] P.M. Narendra and K. Fukunaga. A branch and bound algorithm for feature subset selection. *IEEE Transactions on Computers*, 26(9):917–922, September 1977.

[118] S. K. Nayar, K. Ikeuchi, and T. Kanade. Surface reflection: Physical and geometrical perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):611–634, 1991.

[119] S.K. Nayar and R. Bolle. Reflectance based object recognition. *International Journal of Computer Vision*, 1996.

[120] S.K. Nayar, K. Ikeuchi, and T. Kanade. Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE Transactions on Robotics and Automation*, 6(4):418–431, August 1990.

[121] S.K. Nayar, K. Ikeuchi, and T. Kanade. Shape from interreflections. In *International Conference on Computer Vision*, pages 2–11, December 1990.

[122] M. Oren and S.K. Nayar. Generalization of Lambert's reflectance model. *Computer Graphics*, 28(Annual Conference Series):239–246, 1994.

[123] Langley P. Selection of relevant features in machine learning. In *AAAI Fall Symposium on Relevance*, pages 140–144, New Orleans, 1994.

[124] J. Pagès, J. Salvi, and C. Matabosch. Implementation of a robust coded structured light technique for dynamic 3d measurements. In *IEEE Int. Conference on Image Processing*, Barcelona, Spain, September 2003.

[125] A. Penirschke, M.J. Chantler, and M. Petrou. Illuminant rotation invariant classification of 3d surface textures using lissajous's ellipses. In *International Workshop on Texture Analysis and Synthesis*, pages 103–107, Copenhagen, Denmark, June 2002.

[126] A.P. Pentland. Local shading analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(2):170–187, March 1984.

[127] A.P. Pentland. Shading into texture. *Artificial Intelligence*, 29:147–170, 1986.

[128] A.P. Pentland, R.W Picard, and S Sclaroff. Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3):233–254, 1996.

[129] M. Petrou and S. Barsky. Shadows and highlights detection in 4-source colour photometric stereo. In *IEEE International Conference on Image Processing*, pages 967–970, October 2001.

[130] M. Petrou, S. Barsky, and M. Faraklioti. Texture analysis of 3d surface roughness. *Pattern Recognition and Image Analysis*, 2(3):616–632, 2001.

[131] B.T Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975.

[132] O. Pichler, A. Teuner, and D.J. Hosticka. A comparison of texture feature extraction using gabor filter, pyramidal and tree structured wavelet transforms. *Pattern Recognition*, 29(5):733–742, 1996.

[133] M. Pietikainen, T. Ojala, and Z. Xu. Rotation-invariant texture classification using feature distributions. *Pattern Recognition*, 33(1):43–52, 2000.

[134] P. Pudil, F. Ferri, J. Novovicova, and J. Kittler. Floating search methods for feature selection with nonmonotonic criterion functions. *Pattern Recognition*, 2:279–283, October 1994.

[135] P. Pudil, J. Novovicova, and J. Kittler. Floating search methods in feature selection. *Pattern Recognition Letters*, 15(11):1119–1125, November 1994.

[136] P. Pudil and J. Novovičová. Novel methods for feature subset selection with respect to problem knowledge. *IEEE Intelligent Systems & Their Applications*, 13(2), March/April 1998.

[137] C. Pun and M. Lee. Log-polar wavelet energy signatures for rotation and scale invariant texture classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):590–603, May 2003.

[138] T. Randen and J.H. Husoy. Filtering for texture classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21:291–310, 1999.

[139] T.R. Reed and J.M.H. Du Buf. A review of recent texture segmentation and feature extraction techniques. *Computer Vision, Graphics and Image Processing - Image Understanding*, 57(3):359–372, 1993.

[140] V. Rodehorst. Vertiefende analyse eines gestalts-constraints von aloimonos und shulman. Technical report, CV-Bericht 8, Institut für Technische Informatik, TU Berlin, 1993.

[141] J. Salvi, J. Batlle, and E. Mouaddib. A robust-coded pattern projection for dynamic 3d scene measurement. *International Journal of Pattern Recognition Letters*, (19):1055–1065, September 1998.

[142] A.S. Sanderson, L. Weiss, and S.K. Nayar. Structured highlight inspection of specular surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(1):44–55, January 1988.

[143] K. Schluns and O. Witting. Photometric stereo for non-lambertian surfaces using colour information. In *International Conference on Computer Analysis of Images and Patterns*, pages 444–451, Budapest, Hungary, September 1993.

[144] S.A. Shafer. Using color to separate reflection components. *Color Resolution Applications.*

[145] J. Sherrah. *Automatic Feature Extraction for Pattern Recognition.* PhD thesis, University of Adelaine. South Australia, 1998.

[146] W. Siedlecki and J. Sklansky. On automatic feature selection. *International Journal of Pattern Recognition and Artificial Intelligence*, 2:197–220, 1988.

[147] W. Siedlecki and J. Sklansky. A note on genetic algorithms for large-scale feature selection. *Pattern Recognition Letters*, 10:335–347, November 1989.

[148] W. Silver. *Determining Shape and Reflectance Using Multiple Images.* PhD thesis, MIT, Cambridge, MA., 1980.

[149] S. Singh and M Sharma. Texture analysis experiments with meastex and vistex benchmarks. *Lecture Notes in Computer Science*, (2013):417–424, 2001.

[150] M.L. Smith. The analysis of surface texture using photometric stereo acquisition and gradient space domain mapping. *Image and Vision Computing*, 17:1009–1019, 1999.

[151] M.L. Smith, T. Hill, and G. Smith. Surface texture analysis based upon the visually acquired perturbation of surface normals. *Image and Vision Computing*, 15:949–955, 1997.

[152] M.L. Smith, G. Smith, and T. Hill. Gradient space analysis of surface defects using a photometric stereo derived bump map. *Image and Vision Computing*, 17:321–332, 1999.

[153] F. Solomon and K. Ikeuchi. Extracting the shape and roughness of specular lobe objects using four light photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:449–454, November 1996.

[154] P. Somol, P. Pudil, J. Novovicova, and P. Paclík. Adaptative floating search methods in fetaure selection. *Pattern Recognition Letters*, 20:1156–1163, 1999.

[155] S.D. Stearns. On selecting features for pattern classifiers. In *IAPR International Conference on Pattern Recognition*, pages 71–75, 1976. Coronado, CA.

[156] H. Tagare and R. deFiguiredo. A theory of photometric stereo for a class of diffuse non-lambertian surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(2):133–152, February 1991.

[157] H. Tagare and R. deFiguiredo. Simultaneous estimation of shape and reflectance map from photometric stereo. *CVGIP: Image Understanding*, 55(3):275–286, May 1992.

[158] K.E. Torrance and E.M. Sparrow. Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America*, 57(9):1105–1114, 1967.

[159] D.C. Tseng and C.H. Chang. Color segmentation using perceptual attributes. In *IAPR International Conference on Pattern Recognition*, volume A, pages 228–231, 1992.

[160] H. Vafaie and K. De Jong. Robust feature selection algorithms. In *International Conference on Tools with Artificial Intelligence*, pages 356–363, Los Alamitos, CA, USA, November 1993. IEEE Computer Society Press.

[161] H. Vafaie and K. De Jong. Feature space transformation using genetic algorithms. *IEEE Intelligent System*, 13:57–65, 1998.

[162] Van Ginneken, B. and Stavridi, M. and Koenderink, J.J. Diffuse and specular reflectance from rough surfaces. *Applied Optics*, 37(1):130–139, 1998.

[163] L. Van Gool, P. Dewaele, and A. Oosterlinck. Texture analysis anno 1983. *Computer Vision, Graphics and Image Processing*, 29:336–357, 1985.

[164] M. Varma and A. Zisserman. Classifying images of materials: achieving viewpoint and illumination independence. In *European Conference on Computer Vision*, volume 3, pages 255–271, May 2002.

[165] L. Wang and J. Liu. Texture classification using multiresolution markov random field models. *Pattern Recognition Letters*, 20:171–182, 1999.

[166] D. We and J. Linders. A new texture approach to discrimination of forest cleancut, canopy and burned area using airborne c-band sar. *IEEE Transactions on Geoscience and Remote Sensing*, 37(1):555–563, 1999.

[167] M. Weber and R. Cipolla. A practical method for estimation of point lightsources. In *British Machine Vision Conference*, pages 471–480, Manchester, CA, September 2001.

[168] H. Wechsler. Texture analysis - a survey. *Signal Processing*, 2:271–282, 1980.

[169] J. Weszka, C. Dyer, and A. Rosenfeld. A comparative study of texture measures for terrain classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 6(4):269–285, 1976.

[170] R. Woodham. Reflectance map techniques for analyzing surface defects in metal castings. Technical Report AI-TR-457, MIT A.I. Laboratory, June 1978.

[171] R. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.

[172] R. Woodham. Gradient and curvature from the photometric-stereo method, including local confidence estimation. *Journal of the Optical Society of America*, 11(11):3050–3068, November 1994.

[173] R. Woodham, Y. Iwahori, and R. Barman. Photometric stereo: Lambertian reflectance and light sources with unknown direction and strength. Technical Report 91-18, University of British Columbia, Laboratory for Computational Intelligence., 1991.

[174] Z. Wu and L. Li. A line-integration based method for depth recovery from surface normals. *Computer Vision, Graphics and Image Processing*, 43:53–66, 1988.

[175] J. Yang and V. Honavar. Feature subset selection using a genetic algorithm. *IEEE Intelligent Systems & Their Applications*, 13(2), March/April 1998.

[176] A. Yuille, D. Snow, R. Epstein, and P. Peter, Belhumeur. Determining generative models for objects under varying illumination: shape and albedo from multiple images using svd and integrability. *International Journal of Computer Vision*, 35(3):203–222, 1999.

[177] Z. Zhang. The matching problem: The state of the art. Technical Report 2146, Institut National de Recherche en Informatique et en Automatique, December 1993.