Statistical Methods For Facial Shape-from-shading and Recognition

William A. P. Smith

Submitted for the degree of Doctor of Philosophy

Department of Computer Science

THE UNIVERSITY of York

20th February 2007

Abstract

This thesis presents research aimed at improving the quality of facial shape information that can be recovered from single intensity images using shape-from-shading, with the aim of exploiting this information for the purposes of face recognition and view synthesis. The common theme throughout this thesis is the use of statistical methods to offer enhanced accuracy and robustness over existing techniques for facial shape-from-shading. The work presented goes some way to reinstating shape-from-shading as a viable means to recover facial shape from single, real world images.

In Chapter 2 we thoroughly survey the existing literature in the areas of face recognition, shape recovery and skin reflectance modelling. We draw from this review a number of important observations. The first is that existing solutions to the general shape-from-shading problem prove incapable of recovering accurate facial shape from real world images. The second is that statistical models have been shown to be highly effective in modelling facial appearance and shape variation and have been applied successfully to the problem of face recognition. Finally, we highlight the complex nature of light interaction with skin and note that previous attempts to apply shape-from-shading to real world face images have, almost exclusively, discounted these effects.

Chapter 3 presents our first contribution, which is to explore the idea of incorporating a statistical model within an iterative shape-from-shading framework. In order to do so, we first show how a statistical model can be constructed in the domain of fields of surface normals. We overcome problems of modelling directional data using ideas borrowed from directional statistics and cartography. We use the model as a regularisation constraint within a shape-from-shading algorithm which imposes satisfaction of Lambert's law as a hard constraint. We show how the approach provides both a model-based and data-driven solution and how the model-based solution can be used to estimate facial albedo. We use the estimated shape and albedo information for the purposes of novel view synthesis.

In Chapter 4 we extend the ideas presented in Chapter 3 in a number of ways. We begin by reformulating the statistical model for fields of surface normals in terms of a distribution of points on a spherical manifold. We call on techniques from differential geometry and arrive at a model formulation which is more elegant and allows the whole shape-from-shading process to be couched in terms of operations on the tangent plane to the unit sphere. Our second contribution in this chapter is to show how robust statistics can be used to reduce the impact of regions of low albedo and cast shadows. This approach allows us to identify regions in which the image intensity obeys our simple local illumination model. In those regions which do not, we can use the statistical model to complete the surface. The result is improved performance under significantly non-frontal lighting and reduced sensitivity to albedo variation. We explore the use of the recovered shape and albedo information for face recognition.

In Chapter 5 we expand our consideration of facial shape recovery into the domain of surface height. We present two statistical approaches to the problem of recovering surface height from fields of surface normals (the surface integration problem). The first is based on learning the relationship between the parameters of statistical models for the two representations. The second shows how the parameters of a statistical surface height model can be recovered from the field of surface normals directly. We extend this second approach further by showing how a statistical surface height model can be used to provide a constraint on the estimated field of surface normals within a shape-from-shading framework. The resulting algorithm retains the advantages of the techniques described in the preceding two chapters (strict satisfaction of local irradiance constraints) but yields a height map directly without having to integrate the surface normals.

Our final contribution in Chapter 6 is to relax the assumptions made about the reflectance properties of skin. We show how an arbitrary radiance function can be estimated as part of a shape-from-shading algorithm, using the surface height constraint developed

in the previous chapter. Further, we show how spatially varying reflectance properties can be accounted for by estimating a local albedo term as part of the iterative process. By fitting a parametric reflectance model to the recovered data we are able to extrapolate the reflectance properties beyond those present in the input image. We also demonstrate how the method can be applied to colour images and how this provides a route to facial colour constancy.

The work in this thesis suggests that constraints provided by statistical models of face shape render the facial shape-from-shading tractable, even when complex, non-Lambertian reflectance effects are considered. The results suggest that information useful for illumination and pose insensitive face recognition may be recovered from one training image.

Contents

List of Figures

List of Tables

List of Symbols

$.*$	Component-wise product of two vectors of the same length
$\text{Exp}_x(y)$	The Exponential map of point y at base-point x
ϵ	A constant used to determine when convergence is reached
η	Residual of model fit
η_s	Shininess parameter of the Phong [128] model
λ_i	The ith largest eigenvalue
$\text{Log}_x(y)$	The Logarithmic map of point y at base-point x
μ	The intrinsic mean of a distribution of points on a manifold
$\hat{\mathbf{n}}$	The vector $\mathbf{n}$ normalised to unit length
ρ_d	Lambertian diffuse albedo
ρ_s	Specular coefficient of the Phong [128] model
ϱ_σ	A robust smoothing kernel with width parameter σ
ς	Parameter to control influence of prior and data on model fit
$\mathbf{s}$	A vector describing the direction of a distant point light source
S^2	Unit 2-sphere
$T_p\mathcal{M}$	The tangent space of the manifold $\mathcal{M}$ at point p
ψ_σ	Influence function with width parameter σ
Ψ_i	The eigenvector with the ith largest eigenvalue
w_σ	Weight function with width parameter σ

Acknowledgements

I would like to begin by thanking my parents. Not only for their unstinting support throughout my (now rather long) educational career, but also for always being interested in what I'm doing and encouraging me to pursue my PhD in the first place. I'm sure they are, nonetheless, extremely relieved that I finally have a job!

I feel very lucky to have had as inspiring and knowledgable a supervisor as Professor Edwin Hancock. His support has gone far beyond that of a PhD supervisor including being a careers advisor, travel agent, fellow traveller and good friend. His down-to-earth attitude, exemplified by his gentle progress checks of "Okee dokee?", has made completing my PhD a remarkably painless process. I also thank Edwin for supporting my attendance at several overseas conferences which has given me the chance to visit some spectacular places. My thanks to Dr Richard Wilson who, as my assessor, has provided a different perspective on my work and along with my external examiner Professor Tim Cootes had the unenviable task of reading my thesis cover to cover. I would also like to take this opportunity to thank Bill Stoddart and Tim Selkirk for inspiring an early interest in computer science.

I have had the privilege of forming friendships with fellow members of the CVPR group from all over the world. This has provided an excellent environment in which to work and given me the chance to undertake collaborations with a number of fellow students including Mario Castelan, Antonio Robles-Kelly, Yang Li, Jing Wu and Matt Dickens. In particular, Mario's expertise has been invaluable to call on and our discussions helped shape the direction of my research. However, I remain unconvinced by his claims of the brilliance of Mariah Carey. Special mention must go to Gary Atkinson who has provided endless entertainment during my PhD. Working on the desk opposite him has been a great laugh and our travels to New York and China (particularly our trip to the edge of Inner Mongolia) will live long in the memory.

Other people who have provided support, kept me entertained and are generally worthy of thanks are: Niall and Laura, Abigail, Flight Lieutenant Tricklebank, Jonny, Kit and April, all my family, Comp Sci footballers, the York Alpine Club and whoever it is that buys the tea club biscuits.

Finally, I would like to dedicate this thesis to my amazing girlfriend Michelle, without whom it certainly wouldn't have been possible. We've had so many brilliant times over the course of my PhD and the future promises to be just as much fun.

Declaration

I declare that all the work in this thesis is solely my own, except where attributed and cited to another author. Most of the material in this thesis has been previously published by the author. For a complete list of publications, please refer to Appendix A.

Chapter 1

Introduction

> *It is the common wonder of all men, how among so many millions of faces there should be none alike.*
>
> - Sir Thomas Browne [20]

The capacity for humans to distinguish between large numbers of faces under varying conditions of illumination, viewpoint and expression is remarkable, even more so when one considers the subtle variations that differentiate faces. It is not uncommon for a person to instantly recognise a face that has not been seen since childhood, while seemingly minute changes in facial expression can evoke a powerful emotional response, for example the difference between surprise and terror. Clearly, face perception has developed as a powerful element of the human visual processing system in its own right.

Attempts to explain the means by which humans perform these tasks have preoccupied psychology researchers for a number of decades. One of the fundamental questions to arise from this research is whether the neural representation of a face is based on storing an abstract 3D model or many different characteristic views of the face. The model-based paradigm is attractive in its power to provide an elegant explanation for the capabilities of human face processing. However, it presupposes that the human visual system contains

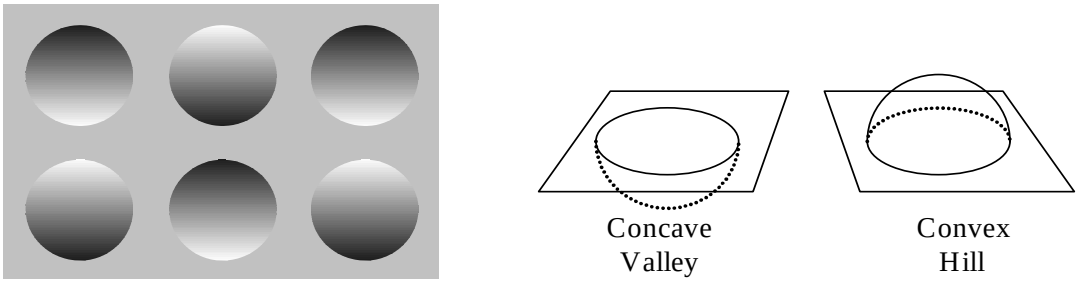


Figure 1.1: Demonstration of the supra-luminance assumption in human SFS. Shaded circles are perceived as convex hills when the image gradient is from top to bottom and as concave valleys when from bottom to top. This is consistent with an interpretation in which the illumination is from above. Rotating the page will invert the effect.

the necessary mechanism for recovering precise, abstract 3D information from one or more views of a face. In contrast, the view- or appearance-based paradigm requires only a mechanism for computing similarity between an observed and stored view. The challenge here is to explain how such a large volume of different views can be efficiently stored and searched.

Although this fundamental question remains unanswered, there are a number of interesting visual phenomena, the implications of which provide a basis for speculation. The human visual system exploits a range of cues in order to infer 3D information from the 2D image projected onto the retina. These include: perspective, stereo, motion parallax, interposition, shadowing and accommodation. However, in the absence of all others some perception of 3D shape is still possible using variations in *shading* as a cue. In the case of observing faces, this cue appears to play a particularly important role. Of interest to psychologists is whether the shading cue is used to infer 3D information or whether it simply forms an important part of an appearance-based representation.

On the left of Figure 1.1 we show a collection of circles shaded with a gradient from white to black. The first thing to note is that from this shading pattern we do infer some 3D shape. There is a strong sense that the circles correspond to either convex hills or concave

Figure 1.2: Bottom-lit face of Frankenstein's monster [58]. Our perception of a face is altered dramatically when it is illuminated from below. The effect is well used in horror films.

valleys. A further observation is that the circles which are brighter at the top are perceived as convex, whereas the circles which are brighter at the bottom are perceived as concave. To further convince the reader, if the page is inverted the effect is reversed. Since this is consistent with an interpretation in which the illumination is from above [95], we can conclude that the human visual system incorporates an assumption of supra-luminance into the process of recovering shape from shading patterns. This appears reasonable when one considers that the human visual system evolved in conditions where the major source of light was from the sun.

When this assumption is breached, our perception of a face is dramatically altered. This effect is well known through its use in cinematography to enhance the frightening appearance of a face. In Figure 1.2 we show an example from the film *Frankenstein*.

In fact, the explanation behind this effect may lie in the disruption of the shape recovery process. It can be shown that faces are treated as a special case by the human shape-from-shading system. There is a sufficiently strong prior constraint on the range of face shapes that are allowable that certain interpretations are considered impossible. This is demonstrated to great effect by the *hollow face illusion* [78] (also known as Gregory's mask [67]). The image on the left of Figure 1.3 shows a mask of a face illuminated from above. The image on the right shows the same mask after a rotation of $180°$, i.e.

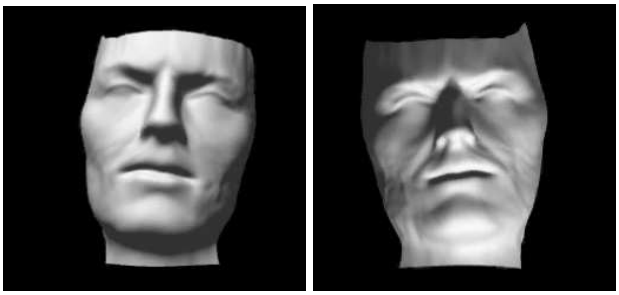


Figure 1.3: The hollow face illusion.

a view into the back of the mask, with the same lighting conditions as on the left. This is perceived as being a convex face with illumination from below, instead of the correct interpretation which is a concave face illuminated from above.

We can draw two conclusions from this result. The first is that the supra-luminance assumption may be overridden if it would result in an interpretation in which a face is globally concave. This implies that in some cases shape-from-shading in humans makes use of prior class knowledge to constrain the shape recovery process. The second conclusion is that lighting from above aids in the derivation of 3D face shape [79]. This is compelling evidence that shape-from-shading is exploited to some degree by the human face processing system.

1.1 The Recovery of Shape from Shading Patterns

This thesis is concerned with the recovery of shape information from the shading patterns in images, specifically face images. Developing a computational method by which to recover shape-from-shading is a classical problem in computer vision. The process was originally known as *photoclinometry* and was considered a branch of *photogrammetry* (the science of measuring objects from images). Its earliest use was to recover information about the surface of the moon in preparation for moon landings [134, 150]. Terrain

Model Acquisition	Virtual reality, computer games, avatars, facial re-animation
Image-based Rendering/Editing	Face re-lighting, expression editing, face re-touching/editing, pose normalisation/adjustment, facial super-resolution
Gaze/Expression Analysis	Virtualised reality/head-up displays, driver safety, crowd counting and dynamics, human-computer interaction
Recognition	See Table 1.2

Table 1.1: Applications of Face Shape Recovery.

analysis using shape-from-shading remains a major area of potential application [17].

An intuitive way in which to conceive the problem is that its aims are the exact inverse of those of computer graphics. In graphics the idea is to produce a realistic image from a description of the scene and imaging conditions. In contrast, shape-from-shading seeks to recover information about the shape of objects in a scene from a single image. This is a considerably more difficult problem since information is lost in the image formation process. This analogy has led to the problem also being referred to as *inverse rendering*.

Although shape-from-shading presents a notoriously difficult challenge, it has nevertheless attracted sustained research over a period of over three decades. The reason for this is that its potential applications are broad ranging and of great utility while it seemingly imposes few restrictions, requiring only a single intensity image as input. In the context of face images, its potential applications are particularly numerous. In Table 1.1 we list some potential applications of facial shape-from-shading. One of the most alluring applications is to use the information extracted from an image using shape-from-shading for the purposes of face recognition. Since shape information and surface markings are independent of viewing conditions, this holds out the possibility of recognising a face, previously seen only once, under dramatically different pose or illumination conditions.

Shape-from-shading for Face Recognition

Machine face recognition is a classical vision problem in its own right. This is primarily because the potential applications are numerous and of great utility. In addition, face recognition represents an achievable subset of the general computer vision problem. Faces are all very much alike, have a consistent orientation and are generally only viewed from a limited subset of the complete viewing sphere. This has allowed the development of systems with acceptable performance in controlled conditions. Commonly these are broken down into three classes of application: authentication ("Am I who I say I am?"), recognition ("Who am I?") and watch list ("Are you looking for me?"). Table 1.2 lists some recognition applications within these categories.

However, face recognition is a very challenging problem. The very fact that all faces are alike makes recognising their differences all the more difficult. In fact, the difference between two images of the same face under grossly different lighting conditions can be greater than between two images of different faces. Yet humans can perform the task well, which shows it is possible.

The appearance of a face in an image is dependent on several factors besides the identity of the subject. These are: pose or viewpoint, lighting conditions, facial expression and skin reflectance properties. Figure 1.4 shows the intrinsic facial features and extrinsic parameters which contribute to the process of forming an image of a face. In particular, the FERET evaluations highlighted variation in pose and lighting as presenting the biggest challenges to face recognition systems [126]. For this reason, there is great inter-

Authentication	Facility or vehicular access, national ID cards, voter registration, driver's licenses, desktop logon
Recognition	Mug-shot database search
Watch list	Advanced video surveillance, most wanted watch list alarm

Table 1.2: Applications of Face Recognition.

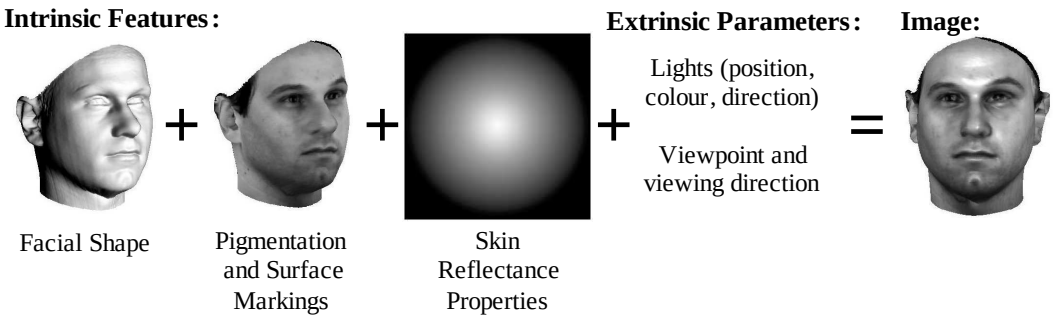


Figure 1.4: The facial image formation process.

est in techniques which can potentially offer pose and lighting invariant recognition, yet only require a single training image (as opposed to multiple images or data from a 3D acquisition device). One way to pose this problem is to attempt to separate these intrinsic facial properties from extrinsic imaging conditions, given only an image of a face (how it 'appears'). This motivates methods that explicitly model the image formation process and account for the underlying physical quantities that give rise to the measured appearance, for example shape and surface reflectance properties. In this thesis we investigate one such technique, and one which is fairly under-represented in the face recognition literature, that of shape-from-shading.

The use of 3D shape information for face perception tasks has been a longstanding goal in computer vision. In addition to biological motivation, 3D shape information would provide invariance to changes in lighting and pose as well as allowing a more principled understanding of the image formation process. Unfortunately, acquiring 3D shape information in a way which maintains real world applicability is an open problem. It is possible to recover accurate high resolution 3D shape information using a laser range finder. However, this is time consuming, requires the subject's participation and requires expensive equipment. An alternative is to make use of one of a class of techniques which attempt to recover shape from an image or images of an object. Such techniques are known as shape-

from-X modules and include shape-from-shading, -texture, -contours, -edges, -stereo and -motion. Of these, shape-from-shading holds out the possibility of recovering 3D shape from a single image and, moreover, allows the recovery of far richer shape information than other shape-from-X modules. This is due to the fact that shape-from-shading uses the information contained in every available pixel as opposed to solely structures such as edges.

Despite being an active area of research in its own right for over three decades, shape-from-shading is still an open problem. Applying shape-from-shading to images of faces has proven challenging due to complex and spatially varying reflectance properties. The use of shape information extracted by shape-from-shading for face recognition has previously been criticised on three counts:

1. Inaccuracy of shape information recovered
2. Prior knowledge of illumination conditions required
3. Assumptions made about reflectance are too general

In this thesis we address these issues and go some way to reinstating shape-from-shading as a viable tool for face recognition.

1.2 Contribution

This thesis presents research aimed at improving the accuracy of the shape information that may be recovered from face images using shape-from-shading. The overarching theme is the use of statistical methods to provide additional constraints for the problem. Noting the success of model-based methods in both face modeling and facial shape recovery evidenced in the literature, we aim to exploit similar techniques for use in facial shape-from-shading. The contributions of the thesis are as follows:

- **Statistical Surface Normal Models**: Our first aim is to construct a statistical model which is suitable for incorporation into a shape-from-shading scheme. The model we construct captures variation in facial shape in terms of a deformable field of surface normals. In order to construct this model, we must tackle the problem of how to model a distribution of unit vectors which are nonlinear in nature. We have developed two approaches to overcoming this problem. The first uses ideas from directional statistics and cartography and is intuitive. This is the model used in Chapter 3. The second uses ideas derived from differential geometry and results in a complex but more elegant formulation. We present this model in Chapter 4.

- **Statistical Shape-from-shading**: We show how to incorporate these models into a shape-from-shading algorithm, by combining the statistical constraint provided by the model with irradiance constraints suggested by Lambert's law. The resulting algorithms offer significant performance improvements over traditional shape-from-shading algorithms that use generic constraints. We then extend this approach to account for the effects of cast shadows and variations in facial albedo using techniques from robust statistics.

- **Statistical Surface Integration**: We take our analysis further by considering how statistical models can aid in the process of recovering surface height from fields of surface normals. We propose two methods for doing this. The first is based on a simplistic statistical analysis which uses linear methods in order to try to learn the relationship between the parameters of a surface normal and a surface height model. The second approach goes further by examining the relationship between surface height parameters and the resulting field of surface normals. We show how surface height parameters can be recovered from a field of surface normals and how this can provide a constraint for accurate surface normal recovery.

- **Estimation of Skin Reflectance Properties**: Central to facial shape-from-shading is a principled understanding of the image formation process and the way in which faces reflect light. We therefore present improved methods for estimating facial reflectance properties and show how these estimates can be incorporated into the shape recovery process. We develop an algorithm that recovers an accurate face surface, field of surface normals and reflectance information from a single image. We also show how the method can be applied to colour images and used for colour constancy.

1.3 Overview of Chapters

The remainder of this Thesis is arranged into the following Chapters:

In Chapter 2 we thoroughly review the relevant literature. This is necessarily broad ranging, since this thesis is concerned with topics from a number of fields: face recognition, statistical face modeling, shape-from-shading and skin reflectance modeling. We also discuss the psychological foundations of human recognition and shape-from-shading.

In Chapter 3 we introduce the geometric shape-from-shading framework and develop a statistical model for fields of surface normals. We show how the model can be used as a constraint within a shape-from-shading algorithm and how the resulting best-fit of the model can be used to estimate facial albedo. We demonstrate that the method offers significant improvements over existing shape-from-shading algorithms and that the estimated shape and albedo recovered from real world images is sufficient for convincing view synthesis.

In Chapter 4 we extend the work in the previous chapter in a number of ways. Firstly, we begin by reformulating the model in terms of the exponential map and use principle geodesic analysis to learn the modes of variation. Secondly, we show how the geometric

shape-from-shading framework can be expressed in terms of operations under the exponential map, allowing the model to be incorporated into the framework in a particularly elegant and efficient manner. Thirdly, we consider the difficulties posed by albedo variations and cast shadows and show how performance can be improved by using robust statistics to downweight areas whose intensity suggests an unlikely normal direction. Finally, we consider how we can use the recovered information to perform face recognition in a number of ways. We build on previous work by performing illumination insensitive recognition given only a single training image per subject.

In Chapter 5 we present our two approaches to recovering 3D facial surfaces from orientation estimates. This leads us to develop a new constraint that can be applied within a shape-from-shading algorithm which enforces both integrability on the field of surface normals and a statistical constraint on the corresponding surface height function.

In Chapter 6, we consider for the first time the issue of non-Lambertian reflectance. Building on the work in the previous three chapters, we develop a facial shape-from-shading algorithm in which the surface radiance function is estimated alongside face shape and a spatially varying albedo term. In relaxing the Lambertian assumption we are able to recover face shape of considerably higher accuracy and by fitting a reflectance model to the estimated data are able to synthesise physically meaningful images in novel poses and under novel illumination.

Finally, in Chapter 7, we review the contributions made in the thesis, highlight weaknesses in the work as it stands and suggest future areas for consideration and ways in which the current work could be extended.

Chapter 2

Literature Review

> *The features of our face are hardly more than gestures which have become permanent.*
>
> - Marcel Proust [132]

In this chapter we provide a thorough review of the literature relevant to this thesis. The review covers psychological foundations, computational shape-from-shading, skin reflectance modelling and machine face recognition. The review focusses on facial shape recovery, particularly where the intended application is face recognition. We also pay special attention to statistical methods, which have proven very successful in a range of face processing tasks. The chapter is organised as follows.

We commence in Section 2.1 with a broad overview of the psychological foundations of face perception, providing biological motivation and highlighting difficulties encountered by the human face processing system. In Section 2.2 we review the face recognition literature, paying greater attention to methods which attempt to use shape or curvature information for recognition. We use the face recognition literature to provide motivation for our work on face shape recovery. Further, the statistical models developed for the purposes of face recognition provide background to the models developed in this thesis.

In Section 2.3 we then review the shape-from-shading literature, specifically where the methods have been applied to the problem of face shape recovery. We discuss previous attempts to use shape-from-shading for face recognition, the challenges encountered and attempts to overcome them. Finally, in Section 2.4 we discuss attempts to model the properties of skin reflectance and highlight work which incorporates these models into a face shape recovery scheme.

2.1 Psychological Foundations

The notion of face recognition as a subset of object recognition and visual processing in general is based on the ability of humans to perform the task. Similarly, performance measures for machine face recognition systems are defined with respect to human face recognition performance. Therefore, an understanding of the human face recognition system is useful in providing inspiration for potential machine face recognition approaches and assessing their successes. Also, an understanding of the limits of face recognition by humans and how a machine system may overcome these limits will help demonstrate the motivation behind and the possibilities made available by the development of machine face recognition.

2.1.1 Human Face Perception

Humans have a remarkably accurate and robust ability to perceive faces, giving them the ability to detect, recognise and make inferences from the expression of another human face [21]. A human face undergoes dramatic transformations over a lifetime, entirely changing itself from birth through to adolescence [87], after which the effects of the environment and aging continue to alter its appearance [98]. Despite this, recognising an adult face from a 40-year-old school photograph is not an unlikely feat for the human

face processing system to perform. This is all the more impressive when it is considered that all faces are very much alike, sharing similar size, proportions, feature locations and skin tone (where the faces are from the same race).

Historically, there was interest in whether facial characteristics reflected traits of personality or character. This idea found particular favour during the Victorian era, as expressed by Schopenhauer [137] who believed that a man's face was "the monogram of all his thoughts and aspirations." The field of phrenology emerged entirely to establish if the characteristics of a face could indicate criminality [152].

From a natural selection perspective, face perception abilities are a clear evolutionary advantage when it is considered that pre-civilised humans cared for their young and existed in co-operative tribes. Correctly identifying known faces would therefore be of great use in making sure the correct young were cared for and differentiating the friendly from the hostile. The ability to read the messages and emotion conveyed in an expression, particularly before the development of spoken communication, would also be of considerable importance.

Evidence for Face Specialisation in Humans

A question of principal interest is whether humans employ a distinct mechanism for recognising faces as opposed to objects in general. If this were established to be the case, we could attempt to distinguish between how the two tasks are accomplished and employ similar approaches in machine face recognition. The weight of evidence suggests that humans do indeed make use of a face-specific process [41, 45, 93]. There is however recent contradictory evidence [61], which suggests face recognition is in fact a process of expert discrimination of similar objects.

Desimone [45] provides a critical review of the neurophysiological evidence for 'face detector cells'. Several independent sets of researchers have found that in monkeys there

are neurons that respond preferentially to facial stimuli which did not respond to other simple stimuli or complex objects. Further evidence of a dedicated face recognition module within the human brain is the existence of *prosopagnosia* [41]. This is an extraordinary condition in which sufferers have a total inability to recognise previously familiar faces, while suffering no other profound agnosia. Prosopagnosia patients retain the ability to visually recognise non-face objects, which suggests the two are separate processes, located in different parts of the brain. Somewhat more observational evidence comes from visual tracking of faces by humans. Infants appear to posses an innate preference for tracking face-like patterns as opposed to those with no pattern or jumbled facial features [113]. This suggests some preconception of a face as a unique object and a pre-wired mechanism for its detection.

Weaknesses of Human Face Recognition

A common observation made of human face processing for a range of perception tasks is the stark difference in performance between 'familiar' faces versus those which are 'unfamiliar'. For example, recognition performance is significantly more robust for faces which are well known to the observer [71]. The implication is that the human face processing system must observe a face many times under a variety of conditions before it becomes sufficiently well-known to be regarded as familiar [30]. What information is extracted in these repeated viewings is unclear but some evidence suggests it does not enable recovery of an abstract 3D representation [147]. The most recent findings go even further, suggesting that unfamiliar faces are not treated as faces at all by the human visual system, but are processed entirely differently [111]. They show that recognition performance on unfamiliar faces, even in ideal conditions is surprisingly poor.

One of the primary limitations that the human system encounters is an apparent capacity limit of familiar faces that can be stored [170], i.e. those that can be reliably recognised

in future. This limit is something machine face recognition systems may aim to exceed. A related weakness is the lack of robustness in human face recognition performance. In particular, large changes in pose or illumination result in significant decreases in performance. Similarly, images of low resolution, high noise, blur or other distortions reduce recognition accuracy. As suggested above, the effect is reduced for familiar faces, but this is nonetheless a weakness on which machine systems could hope to improve.

2.1.2 Human Face Recognition and Shape from Shading

The most convincing evidence to suggest that shading information plays a major role in human face perception is the severe degradation in performance when faces are illuminated from below. Hill and Bruce [79] studied this phenomenon in depth and concluded that the results were consistent with a process in which lighting from above helps the derivation of 3D shape. More evidence of the role played by shape-from-shading in human face perception comes in the form of face specialisation in the shape-from-shading process itself. The supra-luminance assumption may be overridden if the assumption would result in an interpretation where a face is globally concave. This is demonstrated to great effect by the 'hollow face illusion' [78] discussed in the Chapter 1. However, Liu et al [105] present contradictory evidence, suggesting that the role played by 3D shape information is fairly small. They suggest it is the shading information itself which is most important, not the underlying 3D shape which was responsible for the shading patterns.

An alternative answer may lie with an intermediate representation, such as Marr's concept of the $2\frac{1}{2}D$ sketch. The $2\frac{1}{2}$D sketch provides a viewer-centered representation of an observed surface, comprising a field of local surface orientation estimates in a retinocentric coordinate frame. Such a representation could be derived using shape-from-shading. Were such a representation to be used, this may explain the partial viewpoint invariance in human performance [80] without implying the use of a full, abstract 3D

model. There is strong psychophysical evidence that such a representation plays a part in human face processing and that this information is recovered, at least in part, using shape-from-shading [79].

2.2 Automated Face Recognition

Endowing computers with the ability to recognise people would allow the development of a large number of exciting and extremely useful applications. This is the primary reason that face recognition was one of the first areas of computer vision to be significantly researched [16, 92, 94]. A *biometric* is a measurable, physical characteristic which can be used to recognise the identity, or verify the claimed identity, of a subject. Besides the face, there are other biometrics which have proven to be more accurate in machine recognition systems, including fingerprints [106] and retinal or iris scans [43]. The reason that interest in face recognition still continues is that it is the only biometric which may be measured non-intrusively. It does not require the participant to cooperate and it is possible they might not even be aware of the image acquisition process. This improves its social acceptability, whilst increasing the number of potential applications. Another advantage is that face recognition is concerned with data that is readily understandable by humans. Any human would be capable of comparing two face images and making informed deductions, the same would not be true of fingerprints or iris scans.

Face recognition lies within a more general class of *face perception* or *face processing* problems. A number of these other problems are similarly well-known, for example expression analysis [145], aging simulation [98] and facial reanimation [12]. However, of these problems, face recognition is perhaps the most compelling, presents a difficult challenge and, as yet, is unsolved in a general setting. In this part of literature review we restrict ourselves primarily to face recognition. Significant literature surveys have been

carried out in 1995 [27], 2000 [171] and 2003 [170] along with the publication of books surveying the state of the art [69, 104, 169].

2.2.1 Early Work

The earliest attempts at machine face recognition [16, 92, 94] used a seemingly intuitive approach, based on the ideas of physiognomy [21]. The idea was to decide upon a set geometric features which described a face in sufficient detail that they would allow differentiation between faces. These features included the distance between the eyes, length of the nose or width of the face and were chosen so that they could be measured from an image of the face. Recognition was performed by applying classical pattern recognition techniques to the set of features. To overcome the lack of an absolute scale in the input images, Kanade [92] measured all these features relative to the distance between the eyes. However, no attempt was made to compensate for the warping effect of subject pose on the measured features.

The fundamental weakness in the approach is that features which appear intuitively to be important are, in fact, weak in their discriminative power. This is especially true when measured from low resolution images. Consider that humans can recognise faces at very low resolutions, for example down to a width of 15 pixels, where the size of these geometric features varies little between subjects. This suggests that, in humans at least, some other mechanism is being used. Additionally, the use of geometric features alone discounts the important information conveyed by texture, 3D shape and colour.

Kanade's work can now be seen as among the first of what are known as feature-based approaches. More recent work has succeeded in identifying features which allow for robust recognition performance, for example Penev and Atick's Local Feature Analysis [120]. An early influential paper by Brunelli and Poggio [22] compared a feature based system to one in which image templates were used to describe areas of the face. They

concluded that templates outperformed features. This second major class of approaches, based on a more holistic face representation, has proven more fruitful. Subsequently, the majority of research has concentrated on holistic representations.

2.2.2 Image-based Statistical Approaches

Image-based statistical approaches assume that images of faces occupy a subspace, or 'facespace', in the space of all possible images. From a representative training set of face images, a manifold learning technique is chosen to attempt to 'learn' this subspace. By far the best known statistical technique is *Eigenfaces* proposed by Turk and Pentland [149]. Due to ease of implementation and high execution speed this has led to a plethora of work based on extending or improving the technique [9, 40, 75, 123] as well as being the underlying approach to a number of commercial systems [127]. It has also been shown to closely mimic human perceptions of facial similarity and memorability [70].

Eigenfaces

Eigenfaces [149] uses *principal components analysis* (PCA) to derive a basis set of images with which to express a training set of face images in fewer dimensions. PCA or the *Karhunen-Loève transform* is a classical statistical technique for dimensionality reduction [90]. To apply PCA to a set of images, the eigenvectors and eigenvalues of the centred covariance matrix of the data are found. The first to apply PCA to images of faces were Kirby and Sirovich [139]. They used the technique to derive an efficient representation for face images, without using this representation to perform recognition. Turk and Pentland [149] coined the term eigenfaces to describe the eigenvectors which, themselves have a face-like appearance. Intuitively, the first eigenvector will say as much as it is possible to say about a face's appearance in one dimension.

A face is represented by a vector of weights corresponding to each eigenface. An

image from within the training set can be represented or reconstructed accurately or perfectly (depending on whether all the eigenfaces are used). This in-sample truncation error is known theoretically; the R.M.S. error of the reconstruction is simply the sum of the eigenvalues of the unused eigenfaces [5]. The accuracy of representing an out-of-sample image (either a new subject or a subject in the training set but observed under different conditions) depends on whether the training set contained similar images. Recognition using eigenfaces is achieved by calculating the distance between the parameter vector of two faces (e.g. their distance in face space). The identity associated with the closest gallery vector to a given probe vector is reported as the identity.

Recognition using eigenfaces is extremely limited, requiring highly controlled conditions. Fundamentally, the approach suffers from two weaknesses. The first is due to the underlying assumption that face images are linear combinations of a small orthogonal basis set. Were this the case, then the average of any two face images would result in a viable third face image. This is clearly not the case. Secondly, the eigenfaces capture modes of *appearance* variation and so do not distinguish between image variations due to changes of identity and other sources of variation such as lighting, face expression, subject pose etc.

Image normalisation addresses the first problem to some extent, for example by aligning features such as the eye centres [149]. This reduces the problem, but does not ensure that other features will be correctly aligned. The result is that the model is inefficient and capable of representing non-face images [40]. Heseltine et al. [75] assessed whether image pre-processing could be used to normalise for variation in illumination, hence addressing the second problem. Belhumeur et al. [9] noted that the first three eigenfaces appeared to be strongly linked to variation in illumination. By discarding the first three principal components, dependency on illumination conditions is reduced. However, important identity information is also sacrificed.

Rather than providing a realistic route to robust face recognition performance, the real legacy of eigenfaces is that it caused a paradigm shift amongst face recognition researchers. This encouraged focus to move towards developing abstract holistic representations which were suitable features for use in recognition.

Discriminant Analysis

PCA makes no distinction between image variation due to identity and variation due to other sources such as lighting or expression when deriving a face space. Linear Discriminant Analysis (LDA) is another classical statistical technique for dimensionality reduction. However, in contrast to PCA it seeks to maintain the separability of different classes by maximising between-class scatter whilst minimising within-class scatter. By applying the technique to images of faces, the essential idea is to derive a face space which maximises the distance between images of different faces while minimising the distance between images of the same face.

Belhumeur et al. [9] were the first to apply LDA to face recognition and coined the technique Fisherfaces (after R.A. Fisher, who first developed LDA [53]). The method is shown to produce well separated classes in a low-dimensional subspace, even under severe variation of lighting conditions or facial expression. This technique has been empirically shown to far exceed the performance of the eigenfaces algorithm and, given three training images, is robust to gross variations in the conditions of illumination.

The criticism of any discriminant analysis technique is that to perform the analysis requires many training images of each subject, a fact which limits its applicability to real world applications. In the last decade, ever more complex ‘manifold learning’ techniques have been used to attempt to model directly the variability in face images [73, 158, 162]. However, by developing more sophisticated representations which capture appearance in more abstract terms, these complex workarounds can be avoided.

Separating Shape and Intensity

Craw and Cameron [40] were the first to report that images of faces can be more accurately modeled in linear spaces if shape and intensity are considered independently. In this context, by shape we refer to the 2D shape of facial features in the image plane. They proposed morphing a set of training images of faces to an average shape prior to applying PCA to the image intensities. Landmark points were manually identified on each training image, from which an average shape was calculated. The 'shape-free' eigenfaces produced by applying PCA to this morphed training set show more sharply-defined features and capture more subtle variations in intensity across the facial surface.

Intuitively, it is clear that the average of two shape-free faces is likely to result in a viable third face image, since with sufficient landmark points all features will be in alignment. The separation of shape and intensity has led to a body of work under the name Active Appearance Models, which are reviewed in more depth in the following section.

2.2.3 Facial Modeling

The task of machine face recognition is composed of two parts: deriving a representation of the face and then using this representation to perform recognition. If a given face can be accurately represented, it seems likely that the parameters which were used to represent the face could be used for accurate recognition. In particular, if a human can recognise the face after its reconstruction using the recovered parameters, this suggests the parameters contain sufficient information to perform recognition. For this reason facial modeling has, in its own right, been the subject of a large volume of research [15, 33, 36, 37, 39, 40].

A facial model is an attempt to encapsulate facial appearance using a model controlled by a small set of parameters. Ideally, these parameters would explicitly account for effects

such as pose and lighting variation, so as to allow identity to be controlled independently. Such models are generative in nature, in that they allow viable instances of a face to be generated from a set of parameters. Often model parameters are recovered from an input image by iteratively adjusting parameters from their mean position until a chosen error functional is minimised with respect to the input image. This can be a computationally expensive process.

Facial models make possible a number of additional potential applications besides recognition. These include: construction of novel faces from intuitive appearance parameters, synthesis of novel poses, image based rendering, expression analysis, low bandwidth face coding and facial warping such as simulated aging. Previous work on constructing models of facial appearance can be divided into two classes: two-dimensional and three-dimensional models, each of which have associated benefits and drawbacks. The most relevant of these models are surveyed in the following sections.

Two dimensional modeling

The most well known work on two-dimensional facial modeling has been undertaken by Cootes and co-workers over the last ten years [33, 36, 37, 39]. They began by developing a statistical model of shape, using a similar technique to that used to produce shape-free eigenfaces [40]. Instead of measuring the shape of each training face purely to derive an average shape, Cootes et al. built an 'Active Shape Model' (ASM) [36] by applying PCA to the set of vectors describing the shapes in the training set. Just as PCA on intensity images produces a set of eigenvectors describing decreasing amounts of variance, PCA on a set of shape vectors produces a set of eigenvectors which deform the average shape in various ways. When added or subtracted to the average shape, these deformations are known as 'modes of variation'. Often, these modes appear to correspond to identifiable facial features or expressions or describe out-of-plane rotations.

Modeling shape alone may be sufficient for tasks such as expression analysis or pose estimation, however for recognition this is not fundamentally different to Kanade's feature-based approach [92] described above. In order to make use of the intensity information contained within the shape, Cootes et al. developed the Active Appearance Model (AAM) [33]. Here, shape is modeled as for ASMs and intensity is modeled using shape-free eigenfaces. PCA is then applied to the combined parameters, yielding one set of parameters which simultaneously control shape and texture.

To fit an AAM to an image of a face, a gradient descent technique is used to iteratively adjust the appearance parameters until a best fit to the image is found. This can require high computational resources for an accurate fit. Hence, a lot of recent work has focused on improving the efficiency of this process [110]. AAM fitting is also highly dependent on a good initialisation as the search process finds only local minima. There are two fundamental criticisms of AAMs which limit their power as a model of facial appearance:

1. It is not clear how AAMs can model 3D rotations. Landmark points selected for one view may not be visible in another. Cootes et al. attempted to overcome this problem using view-based AAMs [37], by building an AAM for a number of different poses and learning the relationship between poses. More recently, Xiao et al. [160] have shown how to impose constraints of a 3D face model within the 2D AAM fitting process.

2. AAMs still encode global modes of variation. They capture not only identity but also encode changes in illumination and expression. Costen et al. [39] attempted to learn these subspaces using an expectation maximisation algorithm, allowing differentiation between them.

AAMs have no model of illumination or reflectance, hoping to learn these from the data itself. To treat an image of a face as essentially a 'black box' seems simplistic.

The notion of ‘intensity’ can in fact be related to shape, reflectance and illumination, something which three dimensional models have gone some way to exploiting.

Three dimensional modeling

The first notable attempt to build a model of facial appearance using a 3D model was undertaken by Atick, Griffin and Redlich [5, 6]. In analogy with Sirovich and Kirby [139] and Turk and Pentland [149], they showed that human faces whether imaged or as surfaces have few degrees of freedom and thus can be represented with a relatively small number of parameters. They applied PCA to a set of laser range scanned face surfaces represented in cylindrical coordinates to derive a set of eigenvectors describing perturbations from the mean head shape. They coined the modes of variation ‘eigenheads’. Critically, these eigenheads seem to capture modes of variation which are easily identifiable as facial characteristics. They found that an out-of-sample head could be represented with approximately 1% error using 100 modes of variation. Yan and Zhang [161] extended Atick et al’s technique to allow the model to be fitted to non-frontal images.

However, it is clear that rendering a 3D head using Lambertian reflectance without variation in albedo yields very unreal images. Evidently, their model lacks sufficient complexity to realistically capture facial appearance. Nevertheless, they used a minimization technique to fit their model to frontal face images assuming known illumination and found that the recovered shape appeared qualitatively accurate. For synthetic Lambertian images the error was on the order of 2%, though this would obviously be higher for real world images with variation in albedo and which exhibit non-Lambertian reflectance.

Blanz and Vetter [14] enhanced this model by using a device which simultaneously captures shape and albedo (in the three color channels). This allowed them to construct a model whose appearance parameters controlled both 3D shape and surface albedo (or ‘texture’ as they call it). Further, they used a method based on optical flow to find the

dense correspondence between each head. This ensured that every vertex in the model corresponds to the same point on each face in the training sample. When combined with a complex rendering process which simulated camera settings and illumination conditions, near photo-realistic face images can be generated. One of the weaknesses of their approach is the lack of a realistic model of skin reflectance. They used the generic Phong model [128] which combines ambient, diffuse and specular reflectance to capture the reflectance properties of skin. They use a similar technique to Atick et al. [5] to recover model parameters from a given input image of a face, though their optimisation procedure is far more complex. Besides shape parameters, they also adjust albedo, camera parameters, pose and illumination until an optimal match is achieved. This is a very computationally intensive process. Although near photo-realism is achieved for some input images, under extremes of illumination or reflectance the results are poor. It could also be said that these approaches are model dominated. Its ability to represent a face is dependent upon the training set having contained similar faces. The technique also relies heavily on its optimisation procedure, which may return a local rather than a global minimum and is dependent upon a good initialisation.

2.2.4 Two Challenges: Pose and Illumination

In real world trials of face recognition systems, two challenges have consistently proven to present the biggest obstacle to robust recognition performance. The first is the problem of variation in subject pose (or alternatively, variation in viewpoint). The FERET [126] and FRVT [127] evaluations have highlighted the performance degradation introduced by rotations in depth. For example, FRVT2000 [127] found the best recognition rate fell to 68% for rotations of 40° in a database of 200 subjects. The second problem is extreme variation in illumination, particularly the complex illumination conditions encountered in real world images such as multiple and extended light sources. Recently, attention has

focussed specifically on addressing these two issues. In this subsection we summarise this work.

Illumination Variation

Over the last decade, two avenues of research have yielded encouraging results in dealing with the problem of coping with lighting variations in face recognition. The first of these is a class of appearance-based approaches in which a linear subspace is used to capture the variation of a particular face due variations in illumination [8, 64, 100]. The second approaches the problem from a model-based point of view and exploits 3-dimensional facial shape information [15, 19, 164].

3D shape information has the obvious advantage that it is an intrinsic property of the face and hence is invariant to illumination conditions. Although some work has considered the use of non-standard sensing modalities which recover accurate shape [19], to remain widely applicable it is important that the only necessary input is image intensity data [15, 164]. Recovering accurate 3D facial shape information from a single image is a challenging problem that has eluded efforts in both the shape-from-shading and statistical shape modeling literature for a number of decades. The most promising results exploit a morphable model, which captures variation in both 3D shape and texture [15], in an analysis-by-synthesis framework. The difficulty here is that fitting the model to an image requires the costly minimisation of an error functional, the solution of which suffers from model dominance. The result is that the recovered shape only weakly satisfies image irradiance constraints.

In contrast, appearance-based approaches [8, 64, 100] do not aim to recover intrinsic facial features from an image, but rather model the image variability caused by changes in illumination. This work has demonstrated that a low-dimensional subspace can accurately capture the variation in images of a face resulting from arbitrarily complex variations in

illumination. The advantage here is that the basis set can be used in a generative manner to synthesise photorealistic images under arbitrary and possibly extreme lighting conditions. This provides a powerful representation for recognition, in which the identity associated with the subspace which lies closest to a query image is reported as the unknown identity. Impressive illumination-insensitive face recognition performance can be achieved using these approaches. A variety of techniques of varying complexity have been proposed to build these subspaces. The drawback is that these methods either require multiple training images (typically 7-9) or knowledge of the underlying shape and reflectance information (which may be recovered from the multiple training images).

Early work [51, 68] on this subject followed the eigenfaces approach of Turk and Pentland. Rather than attempt to model the variability of all face images using PCA, the idea here was to model only the within class variation in images of one subject caused by changes in illumination. Hallinan [68] and Epstein et al. [51] collected large sets of images of real objects (including faces) with varied lighting conditions. After applying PCA to these sets of images, they found that images of Lambertian objects do lie close to a low-dimensional linear space, the dimensionality of which appears to be approximately 5 ± 2.

The set of images of an object in a fixed pose under all possible illumination conditions forms a convex cone in image space, termed the 'illumination cone' [64]. However, the number of images required to completely define the illumination cone is extremely large (all the possible images obtained with a single light source). Georghiades et al. [64] overcame this problem by assuming that the illumination cone can be well approximated by a low dimensional linear subspace. This assumption is motivated by empirical observations that the typical illumination cone for a human face is flat. The most straightforward way to derive this subspace is by applying PCA to a sample of images which span the illumination cone. Practically this means either acquiring a large number of images of a subject

under a wide variety of illumination conditions, or synthesising such images from a 3D model. Georghiades et al. [64] used photometric stereo to recover an accurate 3D model from a small training sample of images from which they synthesised the large training set necessary. They call the approach a 'few-to-many' training strategy. For face images, it has been found that 11 dimensions of this subspace need to be retained to capture 99% of the variance in the images.

The quality of the approximation to the illumination cone described above depends heavily on the sample images on which the PCA is applied. A biased or small set of samples will produce a subspace which only poorly captures the variance of the cone. Moreover, collecting or synthesising a sufficiently large training sample is cumbersome. In contrast, Basri and Jacobs [8] present an elegant approach based on spherical harmonics in which the low-dimensional subspace is derived analytically from a model. They show that under any lighting conditions, at least 98% of the variability in the reflectance function is captured by the first 9 harmonic images. Their analysis therefore suggests that images of a convex Lambertian surface will lie close to a 9D subspace. This subspace can be derived exactly from the surface normals and albedo of the face without being dependent on the quantity or variability of a sample of training images.

The subspace based on the illumination cone approximation described above requires either a set of training images that samples the entire illumination sphere or the synthesis of such images from a recovered 3D model. Likewise, the subspace based on harmonic images requires the acquisition of a 3D model. Both of these therefore rely on complex and potentially brittle computations. Having made this observation, Lee et al. [100] take the simplest possible approach to constructing a subspace which captures the variation in illumination. Motivated by the spherical harmonics results, they ask whether it is possible to use 9 real images to form the basis images of the low-dimensional space. They determine the best way in which to arrange the physical lighting such that the correspond-

ing images obtain the best possible approximation of the illumination cone. They call these 9 lighting directions the universal configuration, given here in spherical coordinates: $\{(0,0),(68,-90),(74,108),(80,52),(85,-42),(85,-137),(85,146),(85,-4),(51,67)\}$. Their approach requires no training, simply the acquisition of 9 images in which the subject is illuminated by each of the lights from the universal configuration.

Pose Variation

There have been numerous attempts to overcome the problem of variation in pose without explicitly synthesising new views. Pentland et al. [123] extended the Eigenfaces technique of Turk and Pentland [149] in order to achieve pose invariant recognition. In this work, an individual eigenspace is built for each pose. The eigenspace which best describes an input image is selected by finding the lowest residual description error using each viewspace's eigenvectors. The image is then described using the eigenvectors of that viewspace and recognition can be performed. The major criticism of this approach is that many example images of each subject are required to cover all possible views. A feature based approach was proposed by Wiskott et al. [155] which used elastic bunch graph matching. Their system represents a face as a graph with wavelet feature vectors as nodes ("jets"). By learning the way in which these jets transform under face rotation, they were able to perform pose invariant recognition. For rotations of 22° they achieved a recognition rate of 88%, though varying illumination would make feature location more difficult. Cootes et al. [37] extended the Active Appearance Model discussed in Section 2.2.3 to account for variation in pose by building an AAM for a number of different poses, for example 0°, 45° and 90°. For each model, a different set of features is used, which overcomes the problem of different features being visible in different views. They then proposed learning the relationship between models at different views allowing them to represent a face in a particular view, given an image of the face in another view. This essentially 2D

model-based approach was shown to give a reasonable prediction of novel views.

An alternative to developing a pose invariant recognition system is to introduce a preprocessing stage in which a pose corrected image of the input face is generated. This is known as recognition by synthesis. Previous work on facial view synthesis can be divided into those techniques which require more than one example image and those that require only one.

Given more than one image from different viewpoints, a generic stereo technique could be applied [136]. However, the lack of intensity features in a typical face image and the occlusion between views makes finding correspondences difficult. An example of a multi-image technique is the work of Avidan and Shashua [7]. Given two closely spaced model images they are able to synthesise a new view through warping which is far outside the viewing cone of the model images without recovering camera or scene geometry. They do this by deriving the trilinear tensor that describes the transformation from a given tensor of three views to a novel tensor of a new configuration of three views. The synthesised images are realistic, though the method includes no model of lighting or reflectance. Georghiades et al. [64] used their few-to-many approach to recover a 3D model from a sample of training images using a variant of photometric stereo. They were able to synthesise views of face under novel lighting and pose given as few as three images of the face taken under variable lighting.

However, for many applications it is desirable to be able to synthesise novel views of a face given only a single example view. This requires assumptions to be made about occluded areas of the face and therefore the majority of single view techniques use a model which encapsulates prior knowledge of face structure and appearance. Beymer and Poggio [11] proposed such a technique based on image warping. They used optical flow to learn how images of a prototype face change under rotation and then applied the same warp to an image of a novel face. However, any method based on image warping takes

no account of reflectance and lighting, and moreover this technique cannot synthesise occluded areas of the face. More recently, Blanz et al. [13] used their morphable model framework to correct for variations in pose. Having fitted the model to an image, any novel pose can be rendered under any arbitrary lighting conditions. Occluded areas of the input face are implicitly recovered when estimating the face shape parameters that most closely match the visible areas of the face. However, the technique requires manual initialisation and is reliant on the model successfully capturing all possible face shapes.

Shape-from-shading holds out the possibility of recovering 3D shape from a single image, under variable lighting and with a meaningful model of reflectance. This potentially allows the synthesis of new poses from a single image without requiring a prior model of head shape. Previous attempts to use shape-from-shading for novel facial view synthesis have been limited. The most well known work is that of Zhao and Chellappa [166, 167]. They exploited the symmetry in a frontal view of a face to constrain the shape-from-shading process and cancel the effects of variation in albedo, although they did assume Lambertian reflectance. They used this to synthesise novel lighting for frontal images. However, images of rotated faces are not symmetric and hence the constraint cannot be applied. They therefore used image warping based on a single 3D model to generate a frontal view before applying symmetric shape-from-shading. The only results presented were based on pure image warping of very low resolution images, though they did report improved recognition performance.

2.2.5 Depth- or Curvature-based Face Recognition

As described in the previous sections, a representation of a face can be derived from a two-dimensional image of a face in two ways. Either a set of facial features which describe the face can be extracted [22], or the image itself can be projected into a low dimensional space giving an holistic description of the face as a parameter vector [139]. Similarly,

3D shape or curvature information can be used as a source of features [66, 140] or can be described holistically in low dimensional space [5].

However, a representation that uses depth or curvature information will have several advantages over a representation based on a two-dimensional image:

1. Depth or curvature information is implicitly invariant to changes in illumination, reflectance properties and viewpoint

2. Depth or curvature information is better suited to providing a rich description of areas such as the cheeks, forehead and chin

3. There is considerable evidence that humans use shape information for face recognition and very strong evidence that it is used for classification [118]

It is therefore surprising that the attention given to such approaches in the past has been limited. A number of researchers have developed systems for recognising faces from depth information extracted using a laser range scanner [65], stereo camera [76] or other 3D acquisition device [18]. While it is clear that the data recovered by these devices is of high accuracy and resolution, their use severely limits the real world applicability of a face recognition system which uses them. Most importantly, it removes the non-intrusive nature of face recognition by requiring the subject to participate in the data acquisition process.

Gordon [65, 66] presented the first face recognition system based on depth and curvature information. She motivated her work by commenting that traditional approaches have been limited in the feature base available, usually just using the eyes, nose, mouth, and face boundary in their description of a face. It is extremely difficult to describe low contrast areas of the face directly from intensity images with any sort of reliability. However, a depth or curvature-based approach can take advantage of the significant additional

information contained in areas such as the jaw boundary, cheeks and forehead. In general, the larger the feature base, the less strict the accuracy requirements. Therefore an approach that can utilise more features will represent the face more accurately and thus improve its chance of correctly recognising a face.

Gordon used the acquired range data to perform recognition in two ways. The first was based on a global comparison found by calculating the volume between two facial surfaces after an alignment procedure. The second was to extract a set of facial features such as the mean and Gaussian curvature at the tip of the nose in order to describe a face. Both techniques gave good results, with recognition rates above 90% with automated alignment.

Coombes [31, 32] calculated the curvature across the surface of a face from range data acquired using a profile scanner. She used this to produce a shape-based description of a face to measure the effect of facial surgery. Achermann et al. [2] combined data from two coded light cameras to recover facial depth. They applied two classic face recognition approaches to the range images: eigenfaces and hidden Markov models, both of which produced encouraging results. Chang et al. [26] used the surface normal components extracted from range data to build a phase-only vector filter. This allowed recognition between a range face and intensity face in a way that was less sensitive to illumination changes than traditional intensity-based methods.

More recently, Bronstein et al. [18, 19] have developed a technique for representing a facial surface which is invariant to isometric deformations, such as those resulting from different expressions. Using data captured from a structured light camera, they were able to distinguish between identical twins under varying illumination and expression. Heseltine et al. [76] used a fast graph matching algorithm to compare facial surface meshes acquired using a stereo camera.

Besides their limited real world applicability, techniques which use only depth infor-

mation ignore the part played in determining identity by albedo and reflectance properties. For example, the eyes are thought to play a strong role in human face recognition [138], yet without albedo all eyes look very much alike. Likewise, it appears possible that two humans may share very similar facial shape, yet have entirely different coloured skin.

Shape-from-shading holds out the possibility of recovering three dimensional facial shape from a single two dimensional image. This would combine the benefits of using 3D shape information with the real life applicability of working with 2D intensity images. Moreover, as suggested in Section 2.1, there is strong evidence that shape-from-shading is employed by the human face recognition system. Despite this, shape-from-shading has received little attention in the face recognition literature, being described as "not robust enough" [171] and "not shown to be sufficient for the face recognition task" [114]. In the next section we review the shape-from-shading literature and previous attempts to apply shape-from-shading to face images.

2.3 Shape-from-shading

As mentioned in Chapter 1, shape-from-shading has been an active area of research since the 1970s. Even prior to this, photogrammetric work attempted to exploit shading as cue for shape recovery [134, 150]. The computational formulation of the shape-from-shading problem can be stated as follows:

> *From one 2-dimensional grayscale intensity image of an object, compute the shape of the object which gave rise to the image.*

By 'shape', we refer to 3-dimensional shape information of some kind, the means of representation of which may vary. There are few other monocular cues which provide shape information at every point in the image and hence, shape-from-shading as a problem is almost unique. In this section we review the literature concerning attempts to solve the

shape-from-shading problem in a general setting. Two major surveys and evaluations of state of the art shape-from-shading algorithms took place in 1999 [165] and 2004 [50].

2.3.1 Ambiguous Shading Patterns

We have already shown in Chapter 1 that a shading pattern may have more than one interpretation. The example in that case could be perceived as either a convex hill or concave valley. In fact, these ambiguities turn out to be quite fundamental to understanding the problem of shape-from-shading and can be classified as either *bas-relief ambiguities* or *convex/concave ambiguities*.

The example given in Chapter 1 falls into the category of the bas-relief ambiguity. Here, a change in the estimation of the parameters of the lighting results in the perception of a different surface shape. This category of ambiguity was generalised by Belhumeur et al. [10], who proved that when the illumination direction and albedo of a Lambertian surface are unknown, the same image can be produced by a continuous class of surfaces which depend linearly on three parameters. This sort of ambiguity is resolved by the human visual system either by a prior constraint on the illumination direction (lighting from above in the hill/valley example) or by a prior constraint on the recovered shape (global convexity in the case of faces). This ambiguity is often encountered in uncalibrated photometric stereo [64] where additional constraints must be enforced to produce an unambiguous result.

In computational shape-from-shading, it is often assumed that the parameters of the light source are known. In this case, binary convex/concave ambiguities arise in which two interpretations of the shape are valid either side of a critical point. Figure 2.1 provides a contrived 1-dimensional example of this. All of the surfaces in (a) would produce the same 1D image shown in (b) when illuminated from above. Convexity-concavity ambiguities are unavoidable in single image shape-from-shading. Their effect typically

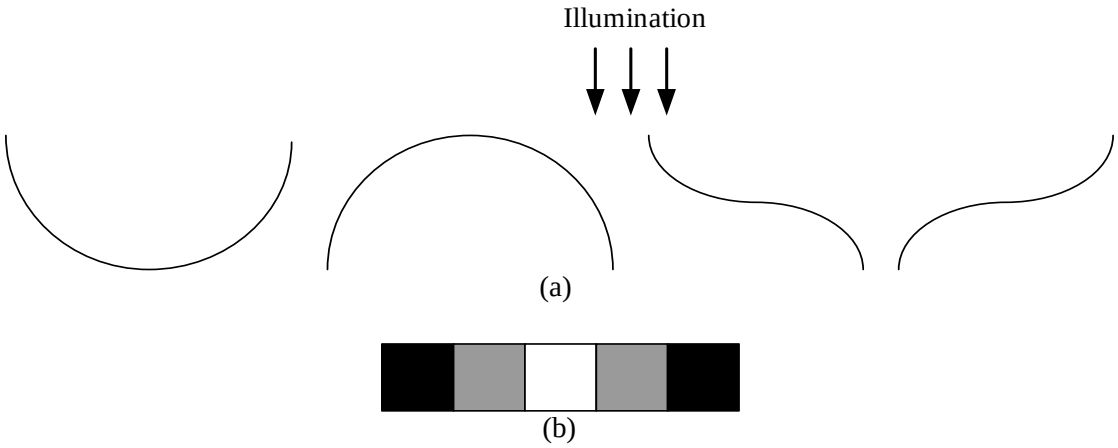


Figure 2.1: All of the surfaces in (a) are valid interpretations of the ambiguous 1D image in (b).

manifests itself in facial shape-from-shading results by imploding the nose and exaggerating the cheeks [25].

2.3.2 Early Work

Van Diggelen [150] was the first to consider the problem of 3D shape recovery from photometric images. Here, the aim was to estimate lunar surface topography and the work focused on the special case in which the angle between the local tangent plane and incident light was very small. With the same application in mind, Rindfleisch [134] proposed a solution applicable only to the special properties of the material in the maria of the moon. The first to give the problem serious attention, provide a general formulation and coin the term "shape-from-shading" was Horn [81,82]. In his seminal work, he posed the problem as that of solving a nonlinear first-order partial differential equation (PDE). If (x, y) is a point in the image plane and the corresponding intensity at this point is $I(x, y)$, the *image irradiance equation* is given by:

$$I(x, y) = R\left(\mathbf{n}(x, y)\right). \tag{2.1}$$

The *reflectance map*, R, relates points in gradient space (i.e. all possible surface orientations) to the measured image intensity $I(x, y)$. Stated in these terms, the image irradiance equation is entirely general, since the reflectance map may describe any surface reflectance properties. However, almost without exception the shape-from-shading literature makes the assumption that the surface reflectance follows Lambert's law (the few examples that do not include [133] and [102], also [59] and [17] in the radar shape-from-shading literature).

Horn [83] showed how to solve the *Eikonal equation* which results from the image irradiance equation using the method of characteristic strips. There are a number of weaknesses in this approach. The first is the problem of error integration, inherent in any line-based integration technique. The second is that determination of the characteristic strips themselves becomes a new problem in itself. The result is that the technique is very sensitive to the chosen boundary conditions.

In retrospect, the earliest work was rather ambitious in its aims, attempting to account for non-Lambertian reflectance effects [82, 134] and trying to directly compute numerical solutions to the shape-from-shading PDE. Subsequently, questions about the uniqueness and existence of solutions have been posed and answered in certain cases [85]. Horn's work [83] is now seen as the first of a class of shape-from-shading techniques based on *propagation*. One of the advantages of these methods is that they do not introduce additional biases by imposing external constraints on the solution. The current state-of-the-art techniques fall into this class and attempt to directly solve the PDE using viscosity solutions.

Other recent shape-from-shading work has typically been classified as either global or local and is based on regularisation techniques, linearisation of the image irradiance equation or intensity gradient constraints. We review a selection of these methods and state-of-the-art viscosity solutions in the following subsections.

2.3.3 Global Methods

Perhaps the most well known solution proposed for the shape-from-shading problem is the variational approach developed by Horn and Brooks [84]. The fundamental idea behind such regularisation approaches is to find a weak solution of the image irradiance constraint, by minimising an error functional which penalises departures from the image irradiance equation and encourages satisfaction of additional constraints. Horn and Brooks used a smoothness term as their additional constraint, which requires adjacent normals to assume similar directions. A similar approach was adopted by Zheng and Chellappa [172] who employ an intensity gradient consistency constraint, while Frankot and Chellappa [59] require that the normals satisfy an integrability constraint.

There are a number of weaknesses that can be leveled at regularisation approaches. Additional parameters must be introduced in order to control the influence of the various constraints. Finding optimum values for these parameters is non-trivial and in order to ensure numerical stability, often conservative values are necessary. This causes the resulting field of surface normals to be oversmoothed and the image irradiance equation to be only weakly satisfied. The estimated shape lacks fine detail and suffers from model dominance. In addition the methods are iterative and hence require convergence criterion to be chosen and are typically slow to execute.

Worthington and Hancock [156] overcame the problem of oversmoothing by enforcing satisfaction of the image irradiance equation as a hard constraint at each iteration of their algorithm. They also investigated the use of more sophisticated smoothness constraints which aimed to preserve discontinuities in the recovered surfaces. Lee and Kuo [101] solve for surface height directly using a stiffness constraint on a triangular mesh and approximating the reflectance map by a locally linear function. This formulation reduces the problem to solving a sparse linear system of equations, but introduces errors through the linearisation of the reflectance function.

2.3.4 Local Methods

Pentland [125] also uses a linear approximation to the reflectance map but in a local setting in which the image intensity and its first and second derivatives are used to recover information about the local surface topography. His approach is non-iterative but is based on the assumption of the local sphericity of the surface. Lee and Rosenfeld [99] avoid using the second derivative of the image intensity, reducing the sensitivity to noise but still require that the surface be locally spherical.

Tsai and Shah [148] use a Jacobi scheme to solve a linear system equations (one for each pixel) to recover a surface, again based on a linear approximation of the reflectance map. Their approach is attractive in that it is straightforward to implement and very efficient. However, because of sensitivity to noise the resulting surface must be smoothed and hence fine detail is lost.

2.3.5 Perspective and Viscosity-based Methods

The most recent approaches to shape-from-shading, the results of which are considered state of the art, use the notion of *viscosity solutions* to first order PDEs. Prados and Faugeras [131] replace the typical modeling assumption of an orthographic camera with a perspective projection. In this case, when the light source is at the optical centre and the surface reflectance follows Lambert's law, they show that the shape-from-shading problem is well-posed. They recover what are perhaps the first usable surface estimates from real world images including: faces, medical images and document surfaces. However, although the recovered shape is, in a global sense, a dramatic improvement over previous work, it is still highly inaccurate and not comparable with the results from multi-image methods.

2.3.6 Shape-from-Shading for Face Recognition

Of the limited attempts to use shape-from-shading for face recognition, the most well known is the work of Zhao and Chellappa [166–168]. Their principal contribution was to exploit facial symmetry in frontal images of faces which allowed them to cancel for the effects of varying albedo. This so called 'symmetric shape-from-shading' allowed them to generate prototype images with novel lighting, which led to improved recognition performance using standard PCA- and LDA-based techniques [166]. They also combined the technique with a generic 3D head model to allow the synthesis of frontal poses from rotated views [167]. Using a single 3D model appears to be a gross simplification and their view synthesis results are of poor quality. In addition, in all their work they assume skin to be a Lambertian reflector composed of piecewise constant albedo patches, which they identify manually.

The eigenhead technique of Atick et al. [5], was posed as a parametric solution to shape-from-shading for face images. They called the minimisation process 'statistical shape-from-shading'. More recently, Dovgard and Basri [47] have combined the technique with the symmetric SFS of Zhao and Chellappa [168], allowing a closed-form solution which satisfies both symmetry and statistical constraints in the best possible way and which accounts for varying facial albedo. However, the technique assumes Lambertian reflectance and requires a frontal view. Nandy and Ben-Arie [115] attempt to learn the relationship between 3D shape and image intensity for a number of face parts. Their shape-from-recognition framework helps to constrain the space of solutions to the image irradiance equation, but relies on statistical methods to learn the effects of illumination variation. However, none of these methods for face shape reconstruction were applied to the recognition problem.

Castelan and Hancock [23,25] combined an integrability constraint with the geometric SFS framework of Worthington and Hancock [156], before applying curvature constraints

to force the convexity of the recovered facial surface. They managed to recover convincing 3D surfaces from single, real world face images, though without a model of skin reflectance or albedo variation.

Worthington and Hancock's geometric SFS [156] has been shown to recover reliable local topographic information. The authors in previous work [140], used histograms of mean and Gaussian curvature extracted by applying the geometric SFS algorithm to images of faces to perform pose invariant recognition. Although the curvature information itself is viewpoint invariant, it is not clear how stable the histogram of curvatures is under facial rotation. In addition, it has not been determined how discriminating these histograms are and whether, even if they were entirely accurate, they would be sufficient for distinguishing between different faces.

2.4 Skin Reflectance Modeling

Variation in images of faces can be conceptually divided into a number of subspaces, which are chosen to reflect useful facial dimensions. Commonly these subspaces are considered to include: identity, expression, pose and lighting. A further and often neglected source of variation is due to the reflectance properties of skin, which can cause dramatic differences in the appearance of a face. In previous attempts at machine face recognition almost exclusively, where skin reflectance has been considered at all, it has been assumed to be Lambertian [5, 28, 63, 140, 166, 167].

Often statistical methods ignore reflectance properties entirely, hoping that they will be learnt implicitly from image intensity [39]. As well as only allowing a simplistic understanding of the image formation process, such an approach also makes no attempt to distinguish between elements of reflectance that may be useful for recognition (such as skin colour and surface markings) and those which vary with time (such as the presence

of perspiration which causes specularities). We suggest that, just as shape free faces have proven more suitable for recognition using principal component analysis [40], reflectance normalised faces would also be more accurately modeled in linear spaces. In addition, the reflectance properties could also be used as a feature vector in their own right which may prove useful for recognition.

Reflectance is of particular importance if image intensity is to be used to recover shape information for recognition, for example using shape-from-shading or photometric stereo. It is for this reason that the topic of skin reflectance is considered in some detail here.

Skin reflectance is extremely complex consisting of Fresnel reflection, sub-surface scattering, asperity scattering and absorption. It is a topic that has been studied extensively in computer graphics and biology and there are some relevant physical models. Potential applications of skin reflectance models include skin detection, photo-realistic skin rendering and detection of skin diseases in vivo. However, only a limited number of these models are applicable to computer vision. A survey of these models is presented in the following sections, preceded by an overview of the biological and optical properties of skin.

2.4.1 The anatomy and optics of skin

Human skin is composed of three distinct layers: the epidermis, the dermis and the hypodermis. Each layer performs a role in the overall function of skin. Skin reflectance is primarily determined by the pigments melanin and hemoglobin which are distributed through these layers. The *epidermis* is the outermost layer of skin. Its thickness varies considerably between different points on the body and between comparable points on different individuals. The epidermis contains melanin which contributes to skin colouring. The *dermis* is mostly composed of collagen fibres. These run in a regular direction over a small area and are responsible for the anisotropic properties of skin reflectance [116]. The

dermis contains the blood vessels which supply the surface of the skin. The hemoglobin contained in this blood plays a significant part in skin reflectance and is the cause of the pink or red colour of light skinned individuals whose lack of melanin allows more light to penetrate to the dermis. Scattering in the dermis is primarily caused by the collagen fibres and its behavior is accounted for by a combination of Rayleigh and Mie scattering [151]. The hypodermis does not play a significant role in skin reflectance.

Approximately 4% to 7% of incident light is reflected by the epidermis at the air-skin boundary [143]. This regular reflectance follows Fresnel's equations relating reflectance to the angle of incidence, plane of polarization and refractive index. The remaining light enters the skin and is either absorbed or, after one or more scattering events, re-emerges from the skin as backscattered light, possibly at a different location to that at which it entered. When the angle of incidence exceeds $60°$, the proportion of light that is regularly reflected increases substantially. This explains why white skin appears pinker when illuminated straight on, whereas with a large incident angle it takes on the color of the illuminant [129].

2.4.2 Other considerations

Besides regular reflectance at the air-skin boundary and back-scattering within the dermis and epidermis, there are a number of other factors which contribute to the reflectance properties of skin and are worthy of consideration. Namely asperity scattering, hair follicles and sweat glands.

An 'asperity' refers to a slight projection from a surface, for example hair tips, dust or fluff. Asperity scattering mostly affects the occluding boundary and shadow terminator and although its effect on outgoing radiance is quantitatively small, it is thought to play a strong perceptual role. Koenderink and Pont [96] suggest it is this that gives female skin its 'peachy' quality whereas subsurface scattering is responsible for more 'milky' skin.

Figure 2.2: Asperity scattering at the face contours.

Sparse asperities result in a single scattering event. The effect is a reflectance function whose behaviour is almost exactly opposite to that of Lambertian reflectance. This is demonstrated in figure 2.2. The area where asperity scattering is prominent is highlighted and enlarged.

The second additional consideration is the presence of facial hair follicles, not only on the scalp, but also in the eye brows, eye lashes and facial hair. There exist numerous reflectance models for hair shading developed by the graphics community, for example the Kajiya-Kay model [91]. However, these are very unlikely to be useable in a vision context. We suggest that it will be sufficient to consider facial hair as dark patches of skin where much of the incident light is absorbed.

Finally, perhaps the most important additional factor to consider is the strong effect of perspiration or oil at the skin's surface [141]. A number of different fluids are produced by glands in the skin and secreted at the skin's surface. Their effect is to cause increased specular reflection, which for highly perspiring faces can dominate skin reflectance. Analyzing such images using shape-from-shading would be highly inaccurate without including specular reflection in the reflectance model.

2.4.3 Empirical skin reflectance measurements

Two notable attempts have been made to learn the BRDF of skin by making empirical estimates [42, 109]. The main problem with this approach is that the BRDF has four degrees of freedom, that correspond to the zenith and azimuth angles for the light source and the viewer. As a result the tabulation of empirical BRDF's can be slow and labour intensive. Moreover, since these methods require a perfectly stationary, flat sample, they are badly suited to taking measurements from skin samples of live subjects.

Dana et al [42] included skin reflectance measurements in their Columbia-Utrecht Reflectance and Texture (CUReT) database. They used an *in vitro* skin sample (extracted from a cadaver) which overcame the problems of stability and shape (the skin could be forced onto a flat surface). However, since skin reflectance is largely dependent on blood flow, it is not clear whether results on dead skin are particularly meaningful, unless the aim is realistic rendering of dead skin.

Marschner et al [109] went some way to solving these problems by using an inverse rendering approach allowing them to take in vivo measurements. Having laser range scanned each subject, they attached a location pattern to their head, allowing them to accurately tabulate surface orientation against measured intensity for a number of different viewpoints and illumination directions. They found that skin reflectance is almost Lambertian at small incidence angles but exhibits strong forward scattering as the incident angle increases. They also found that the application of artificial perspiration had a marked effect but that makeup did not.

Rather than measure the complete BRDF, Angelopoulou [4] instead measured the spectral response of living human skin over the visible spectrum for a fixed incident and viewing angle. She found that the skin of all volunteers showed similar characteristics and all contained a distinct shape at around 575nm to a greater or lesser extent. She was able to derive a simple algorithm for recognizing this shape and showed it could be used

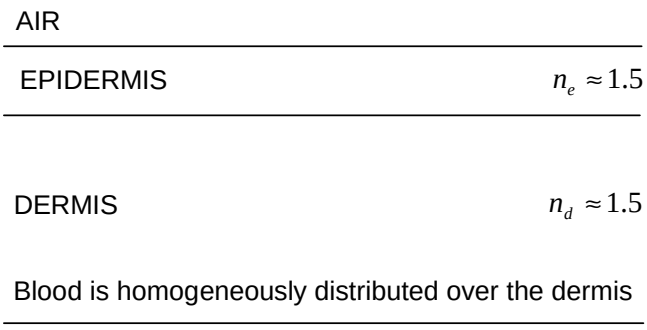


Figure 2.3: The van Gemert [151] simplified model of skin reflectance.

to autonomously distinguish between the spectral response of human skin and that of a mannequin.

However, given the enormous variation in skin reflectance properties both between different subjects and between the same subject at different times, it is unlikely that any single empirically measured set of reflectance properties will suffice to characterise skin reflectance in general. Further, any attempts to use such measurements in vision tasks would require carefully calibrated equipment. For this reason, empirical measurements will not be used to model skin reflectance in this research, though they may serve to validate any reflectance data recovered in other ways.

2.4.4 Physical skin reflectance models

Van Gemert et al. [151] were amongst the first to attempt to build a model of skin reflectance based on measurable physical quantities. They used a simplified model of skin composed of two plane parallel layers as shown in figure 2.3, where n_e and n_d are the refractive indexes of the epidermis and dermis. They assumed that a different density of absorbers and scatterers are distributed throughout the two layers. They used the Henyey-Greenstein phase function to model scattering within the tissue and found a good agreement with experimental data.

However, their model requires a number of parameters to be either known or searched for to minimise an error functional. Moreover, it lacks two of the features of skin reflectance described previously: specular reflection caused by perspiration and asperity scattering. Nevertheless, it may provide a starting point for building a model of reflectance which could be utilised in vision tasks.

A more recent model based on a complete physical model of skin was proposed by Hanrahan and Krueger [72]. They perform a full multiple scattering simulation using a Monte Carlo numerical integration method. They propose a model of subsurface scattering in layered surfaces and use one-dimensional linear transport theory to model the scattering. They use measurable optical properties to describe each layer. The technique is hugely computationally intensive and is not invertible in any sense due to the stochastic sampling used. It is not at all suitable for use in computer vision, but serves to demonstrate the complexity of existing physical models.

2.4.5 Phenomenological skin reflectance models

Photo-realistic skin rendering has been a focus of research for the computer graphics community for many years and is still an unsolved problem. Swerdlow [144] went as far as to say that "Computer Graphics artists will capture the holy grail of animation when they can bring virtual skin to life." The aim in graphics is to develop models which are sufficiently complex to capture many skin types and are controlled by intuitive parameters whilst allowing photo-realistic rendering but which are sufficiently efficient to be useable for real-world applications.

The state of the art in graphical skin rendering is the subsurface transport model of Jensen et al. [88]. Rather than use a BRDF and a model of subsurface scattering, they used the complete Bidirectional Subsurface Reflectance Distribution Function (BSSRDF). This allows them to simulate effects that BRDF models cannot capture, such as colour bleeding

within materials and diffusion of light across shadow boundaries and silhouettes. As with the Hanrahan and Krueger model, it is impossible to envisage how this could be used in a vision context, but as the state of the art is worthy of mention.

It may be that a generic phenomenological reflectance model may suffice to model skin reflectance for computer vision tasks. For example, the rough surface model of Oren and Nayar [117] captures limb brightening effects which could be used to model asperity scattering at the occluding boundary in skin. The simplest analytical model to have been used to approximate skin reflectance is the classical specular reflection model of Phong [128]. Blanz and Vetter [15] estimated the parameters of the Phong reflectance model alongside illumination conditions and camera, shape and texture parameters in order to fit their morphable model to a given image of a face. This is based on the anecdotal observation that skin is shinier than a Lambertian surface, but no physical justification can be made.

2.4.6 Image-based methods

Given the huge variation in skin reflectance properties of a subject over time or between different subjects, and the lack of a model of skin reflectance which is applicable to vision tasks, the attractiveness of an image based method is obvious. There is a clear need to be able to estimate the BRDF of a given subject at a given time if analysis by shape-from-shading is to be performed.

For images of arbitrary surfaces Dror [48] developed a reflectance recognition technique which used single images. This was motivated by the fact that humans seem to be able to estimate reflectance properties accurately without contextual information to specify the illumination [55]. Dror showed that the regular statistical properties of natural illumination lead to predictable relationships between reflectance and certain image statistics. These statistics can be used to classify reflectance properties reliably even though

the problem of reflectance recognition under arbitrary illumination is under constrained. However, Dror's aim was to distinguish between surfaces with gross differences in reflectance properties and the technique is probably ill-suited to the more subtle variations found in the reflectance properties of different subject's skin.

Georghiades [62] presented a technique which was shown to accurately recover shape, albedo and reflectance properties from 12 aligned images of a surface under variable lighting. The technique was in essence an extended uncalibrated photometric stereo technique and used the Torrance and Sparrow reflectance model [146]. He applied the technique to images of faces and confirmed that the recovered BRDF closely matched that of Marschner et al. [109]. The technique also allowed qualitatively accurate view synthesis. A similar technique was proposed by Debevec et al. [44]. However, the need for multiple images in both techniques makes their applicability to real world vision tasks limited.

Claridge et al. [29] derived a model of the spectral response of healthy skin based on three parameters: melanin concentration, haemoglobin concentration and the thickness of the dermis. They then found the relationship between the spectral composition of the light emitted from the skin and these parameters. This allowed the histological parameters of a skin sample to be estimated from a colour image, and abnormal areas to be highlighted. This technique may provide a way to estimate useful model parameters from an input image, helping to initialise a complete model of the BRDF of skin. However, the technique requires calibrated measurements and knowledge of the surface orientation.

Robles-Kelly and Hancock [135] proposed a parameter free technique for estimating a slice of the BRDF of a given surface from a single image. They note that under the retro-reflection condition and assuming an isotropic and homogeneous reflectance function, the reflectance function is dependent only on the angle of incidence at a given point. They use the image gradients to estimate a tabular representation of this function. In this thesis we make use of this technique as a preprocessing step. We discuss the method in more

detail in Chapter 3.

2.5 Conclusions

There is some psychological evidence for the use of shape-from-shading in the perception of faces by humans. However, existing computational methods for shape-from-shading are insufficiently robust to recover usable shape information from real world images of faces. In fact, the most successful methods for automated face processing have been based on statistical models that capture the appearance of a face in an image [33, 149]. The weakness of these approaches is that they make no attempt to decouple the measured image intensity into the physical facial properties (shape and reflectance) and imaging parameters (illumination, viewpoint and camera properties) that produce the resulting appearance. The effect is that the models find it hard to distinguish between changes in identity and changes in imaging conditions.

Recent work using 3D morphable models [14] has gone some way to addressing these issues. Here, the imaging process is modeled explicitly and the statistical model captures only changes in facial shape and reflectance properties. By fitting the model to an image we recover the underlying facial shape and reflectance information, albeit it at the cost of solving a computationally expensive optimisation problem. However, it is the overdependence on this model that proves to be a weakness. By relying entirely on a model-driven parametric process, the information that can be directly extracted from an image is ignored. The problem is model dominance, in which the recovered information relies entirely on the capacity of the model to generalise to unseen examples. The result is that the fitted model only weakly reconstructs the input image, with no notion of local data-closeness, and failure to recover atypical features.

These observations motivate the use of shape-from-shading to recover facial shape

information. Within a shape-from-shading framework, image intensity constraints may be strictly satisfied at every pixel in an image. This ensures accurate recovery of local surface features and that the recovered shape recreates the measured intensities exactly, without having to solve a computationally expensive model fitting optimisation problem. However, the success of statistical models in both facial appearance modeling and facial surface recovery and the face specialisation present in human shape-from-shading provide strong motivation for the use of a statistical constraint within this context. The hope is that such an approach could combine the robustness of statistical methods with the advantages of a classical shape-from-shading algorithm.

One specific weakness of existing shape-from-shading approaches in the context of faces is their lack of a realistic model of skin reflectance. Having reviewed the skin reflectance modeling literature, we conclude that no existing models are suitable for application in a shape-from-shading context. We suggest that image based empirical measurements provide the most flexible way of accounting for varying skin reflectance properties.

Chapter 3

A Statistical Model for Shape-from-shading

I never saw an ugly thing in my life: for let the form of an object be what it may - light, shade, and perspective will always make it beautiful.

- John Constable [103]

Shape-from-shading provides an alluring yet somewhat elusive route to recovering 3D surface shape from single 2D intensity images [165]. In particular, there has been sustained interest in using the method to recover realistic 3D face shape from single face images [5, 24, 47, 130, 168]. Unfortunately, the method has proved ineffective due to problems inherent to shape-from-shading: convexity-concavity ambiguities [156] and the bas-relief ambiguity [10], and challenges specific to face images: variations in pigmentation and facial hair resulting in non-constant reflectance properties. Convexity-concavity ambiguities are unavoidable in single image shape-from-shading and are responsible for a number of illusions, including Gregory's famous inverted mask [67]. Their effect typically manifests itself in shape-from-shading results by imploding the nose and exaggerat-

ing the cheeks [24]. For these reasons, multi-image methods such as photometric stereo have proven more successful [64].

Contribution

Our aim in this chapter is to show how a statistical model of face shape may be embedded within a shape-from-shading framework. The aim here is to recover a field of surface normals (a needle-map) from a single intensity image, by exploiting the direct relationship between surface orientation and measured intensity. Unfortunately, the construction of a statistical model for the distribution of facial needle-maps is not a straightforward task. The statistical representation of directional data has proved to be considerably more difficult than that for Cartesian data [74]. Surface normals can be viewed as points residing on a unit sphere and may be specified in terms of the elevation and azimuth angles. This representation makes the computation of distance difficult. For instance, if we consider a short walk across one of the poles of the unit sphere, then although the distance traversed is small, the change in azimuth angle is large. Hence, constructing a statistical model that can capture the statistical distribution of directional data is not a straightforward task.

In this chapter we provide an intuitive way in which to overcome this problem, drawing on ideas from cartography and directional statistics. Our starting point is the *azimuthal equidistant* or Postel projection [142]. This projection has the important property that it preserves the distances between locations on the sphere. It is used in cartography for path planning tasks. Another useful property of this projection is that straight lines on the projected plane through the centre of projection correspond to great circles on the sphere. The projection is constructed by selecting a reference point on the sphere and constructing the tangent plane to the reference point. Locations on the sphere are projected onto the tangent plane in a manner that preserves arc-length.

We exploit this property to generate a local representation of the field of surface nor-

mals. We commence with a set of needle-maps, i.e. fields of surface normals which in practice are obtained from range images. We begin by computing the mean field of surface normals. The surface normals are represented using elevation and azimuth angles on a unit sphere. At each image location the mean-surface normal defines a reference direction. We use this reference direction to construct an azimuthal equidistant projection for the distribution of surface normals at each image location. The distribution of points on the projection plane preserves the distances of the surfaces normals on the unit sphere with respect to the mean surface normal. We then construct a deformable model over the set of surface normals by applying the Cootes and Taylor [34] point distribution model to the co-ordinates that result from transforming the surface normals from the unit sphere to the tangent plane under azimuthal equidistant projection. On the tangent projection plane, the points associated with the surface normals are allowed to move in a manner which is determined by the principal component directions of the covariance matrix for the point-distribution. Once we have computed the allowed deformation movement on the tangent plane, we recover surface normal directions by using the inverse transformation onto the unit sphere.

Spheres as Manifolds

The general idea of projecting from the Gauss map of a surface to a plane in such a way as to preserve geodesic distance is of course central to the exponential map in differential geometry [46]. In fact, Pennec [122] has recently developed a framework for the analysis of statistical data on manifolds using the exponential map, and has applied the method to the analysis of medical images. In Chapter 4, we revisit the problem of building a statistical model for surface normals and formulate our model using the tools of differential geometry. The resulting representation is both more flexible and elegant. However, in this chapter we adhere to the cartographic analogy since this provides a more intuitive

explanation of the processes involved and is true to the chronological development of the work.

Using the Model To Constrain Shape-from-shading

We use the model to provide a statistical constraint within a geometric shape-from-shading framework. When the surface reflectance follows Lambert's law, the surface normal is constrained to fall on a cone whose axis is in the light source direction and whose opening angle is the inverse cosine of the normalised image brightness. This method commences from an initial configuration in which the surface normals reside on the irradiance cone and point in the direction of the local image gradient. The statistical model is fitted to recover a revised estimate of the surface normal directions. The best-fit surface normals are projected onto the nearest location on the irradiance cones. This process is iterated to convergence, and the height map for the surface recovered by integrating the final field of surface normals. We show how albedo maps can be recovered using the difference between observed and reconstructed image intensity. With the albedo maps to hand we explore how faces can be realistically reilluminated from different lighting and viewing directions.

Chapter Outline

The outline of this chapter is as follows. We begin in Section 3.1 by introducing shape-from-shading and describing a framework which can be used to approach the problem. In Section 3.2 we show how a statistical model may be constructed that captures variations in facial shape in terms of a field of surface normals. In Section 3.3 we describe two methods for incorporating this model into a shape-from-shading framework. Finally, in Section 3.4 we provide experimental results on model construction, fitting and view synthesis on both ground truth and real world data.

3.1 Shape-from-shading

In this section we describe a framework for shape-from-shading within which a wide variety of constraints can be incorporated. We begin by describing the radiometric processes which underpin the shape-from-shading problem and provide the necessary formalisms.

3.1.1 Radiometry and Photometry

Radiometry is the science of the measurement of light in terms of its absolute power. Light may also be measured using the techniques of *photometry*, in which light is measured according to its brightness as perceived by the human eye. The two share analogous quantities, with photometric units weighting the measured power at each wavelength with a factor that represents how sensitive the eye is at that wavelength.

Shape-from-shading is effectively a photometric process (or radiometric, depending on the type of imaging device used) since it seeks to recover information about surface shape from measurements of light made from a scene. We begin this chapter by providing a brief overview of the physical quantities involved and explain how the quantities we measure with a digital imaging device correspond to light in the real world.

Light at Surfaces

When incident light strikes a surface, it may interact with it in a number of ways. The light may be absorbed, scattered or transmitted or a combination of these effects. We make the assumption that reflectance effects are local. That is, light leaving a point is due only to light arriving at this point.

In order to describe the way in which light interacts with a surface we use a function that captures the relationship between incoming and reflected light. This function is known as the *bi-directional reflectance distribution function* (BRDF) and is expressed

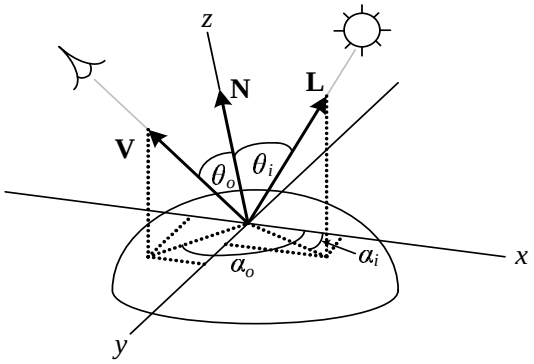


Figure 3.1: The geometry of surface reflectance.

in terms of three unit vectors describing the direction of: the incident illumination $\mathbf{L}$, the viewer $\mathbf{V}$ and the local surface normal $\mathbf{N}$. The vectors are usually written in a local spherical coordinate system whose z axis is given by the surface normal vector. We therefore write the illumination and viewer direction as spherical coordinates: (θ_i, α_i) and (θ_o, α_o) respectively. The geometry of this notation is clarified in Figure 3.1.

The appropriate unit for representing incoming power is *irradiance*, defined as incident power per unit area. We denote the irradiance from direction (θ_i, α_i) by the function $f_i(\theta_i, \alpha_i)$. The appropriate unit for measuring the distribution of light in space is *radiance*, defined as power per unit area per unit solid angle. We denote the radiance in direction (θ_o, α_o) by the function $f_o(\theta_o, \alpha_o)$.

The BRDF describes the ratio of the emitted surface radiance to the incident irradiance. If a surface illuminated by irradiance $f_i(\theta_i, \alpha_i)$ from a differential region of solid angle $d\omega$ was to emit radiance $f_o(\theta_o, \alpha_o)$, its BRDF would be:

$$\rho(\theta_i, \alpha_i, \theta_o, \alpha_o) = \frac{f_o(\theta_o, \alpha_o)}{f_i(\theta_i, \alpha_i) cos(\theta_i) d\omega}. \tag{3.1}$$

Lambertian Reflectance

Surfaces which exhibit perfectly matte reflectance have a BRDF which is independent of direction. In other words, incoming light is scattered equally in all directions. A perfectly diffuse surface such as this is known as *Lambertian* and has a particularly simple BRDF:

$$\rho_{\text{Lambert}}(\theta_i, \alpha_i, \theta_o, \alpha_o) = \frac{\rho_d}{\pi}, \tag{3.2}$$

where ρ_d is the intrinsic reflectivity of the surface, known as the *albedo*. The observed brightness of a Lambertian surface will be independent of viewing direction and shading variations are purely due to foreshortening effects with respect to the incident illumination. More precisely, the outgoing radiance is proportional to the $\cos(\theta_i)$ term. This is known as Lambert's Law [97].

We return to the BRDF in Chapter 6, where we consider more complex models of reflectance that provide better approximations of real world skin reflectance properties. However, in the remainder of this chapter we deal only with Lambertian surfaces.

Measuring Light with a Camera

We assume our camera comprises a thin lens which concentrates the light radiating from a scene onto an image plane. In a digital camera, incident light falling on the image plane is measured using a CCD. In Figure 3.2 we illustrate this process with an example.

We denote the object radiance from a scene patch ∂A centred on P as L and the image irradiance on the image patch $\partial A'$ centred on P' as E. The solid angle subtended by ∂A from the centre of the lens of diameter D is $d\omega$. We can show that the image irradiance at the sensor element is proportional to the outgoing radiance from the surface [57]:

$$E = \left[\frac{\pi}{4}\left(\frac{D}{z'}\right)^2 \cos^4\alpha\right] L. \tag{3.3}$$

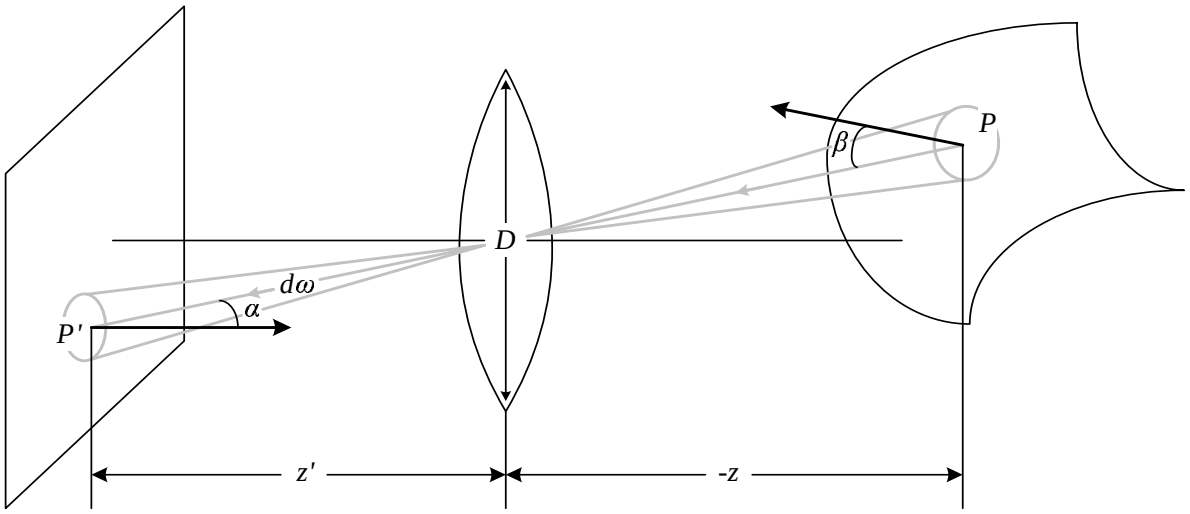


Figure 3.2: Object radiance and image irradiance for a thin lens.

The utility of this relationship is that it shows that what we measure (the image irradiance at the sensor element) is proportional to what we are interested in (the scene radiance). We can further simplify this relationship, since the field of view of a camera is typically small. This means that the effect of the $\cos^4 \beta$ term is small and we can normalise it to unity. Hence, the position on the image plane relative to the optical axis does not effect the image irradiance.

In the case of a Lambertian surface illuminated by a point light source of unit intensity, the result is that the image irradiance is proportional to the albedo and foreshortening term:

$$E \propto \rho_d \cos\theta_i = \rho_d(\mathbf{N} \cdot \mathbf{L}). \tag{3.4}$$

The *camera response function* relates the image irradiance at the sensor element, E, to the measured pixel brightness reported by the camera, I:

$$g_{\text{cam}} : E \mapsto I. \tag{3.5}$$

If the camera response is linear, then the pixel intensity is directly proportional to image irradiance. Otherwise, an estimate of g_{cam}^{-1} is required in order to recover image irradiance values from the image intensity. In our experimental sections we either use images taken by a camera with a linear response function or discuss how it has been estimated.

3.1.2 Geometric Shape-from-shading

Suppose that a facial surface $F \in \mathbb{R}^3$ is projected orthographically onto the image plane and parameterised by the function $z(x, y)$. We denote the local surface normal at a pixel indexed (x, y) on the image plane as $\mathbf{n}(x, y)$ and the corresponding measured intensity as $I(x, y)$. The *image irradiance equation* captures the relationship between the surface normal, light source, $\mathbf{s}$, and measured intensity. We assume that the light source is of unit intensity and therefore that the vector $\mathbf{s}$ is of unit length. For a perfectly diffuse surface this relationship is given by Lambert's law:

$$I(x, y) = \rho_d(x, y)(\mathbf{n}(x, y) \cdot \mathbf{s}), \tag{3.6}$$

where $\rho_d(x, y)$ is the diffuse albedo at the pixel (x, y). Albedo describes the intrinsic reflectivity of the surface, i.e. the proportion of light that is reflected, and lies in the closed interval $[0, 1]$. However, at this stage we disregard the effect of albedo by assuming its value has been normalised to unity.

In general, the surface normal $\mathbf{n}(x, y)$ can not be recovered from a single brightness measurement $I(x, y)$ since it has two degrees of freedom corresponding to the elevation and azimuth angles on the unit sphere. Clearly, Lambert's law provides a partial constraint on the direction of the surface normal, namely that the angle between the light source and normal is:

$$\theta(x, y) = \arccos(\mathbf{n}(x, y) \cdot \mathbf{s}) = \arccos I(x, y). \tag{3.7}$$

Geometrically, this means that the surface normal must lie on a right circular cone whose axis is the light source direction and whose half angle is $\theta(x, y)$. By constraining the surface normal to lie on the cone, we satisfy the image irradiance equation and hence ensure the fullest possible use of the input image. Assuming the direction of the light source vector, $\mathbf{s}$, is known or can be estimated, the task is therefore to choose the correct position of the surface normal on the cone.

Worthington and Hancock [156] provide a particularly simple iterative framework for solving this problem. They employ an iterative process in which the surface normal is free to move to an off-cone position subject to smoothness or curvature consistency constraints. However, the hard irradiance constraint is re-imposed by rotating each surface normal back to its closest on-cone position. This process ensures that the recovered field of surface normals satisfies the image irradiance equation after every iteration.

Suppose that $\tilde{\mathbf{n}}^{(t)}(x, y)$ is an off-cone surface normal at iteration (t) of the algorithm. Let $\Theta\left((u, v, w)^T, \phi\right) \in SO(3)$ denote a rotation matrix which rotates a unit vector about an axis $(u, v, w)^T$ by an angle ϕ, computed using:

$$\Theta\left((u, v, w)^T, \phi\right) = \begin{pmatrix} c + u^2c' & -ws + uvc' & vs + uwc' \\ ws + uvc' & c + v^2c' & -us + vwc' \\ -vs + uwc' & us + vwc' & c + w^2c' \end{pmatrix}, \tag{3.8}$$

where $c = \cos(\phi)$, $c' = 1 - c$ and $s = \sin(\phi)$.

To restore a surface normal to its closest on-cone position it must be rotated by an angle $\phi = \theta(x, y) - \arccos\left[\tilde{\mathbf{n}}^{(t)}(x, y) \cdot \mathbf{s}\right]$ about the axis $(u, v, w)^T = \tilde{\mathbf{n}}^{(t)}(x, y) \times \mathbf{s}$. Therefore the update equation is:

$$\mathbf{n}^{(t+1)}(x, y) = \Theta\left(\tilde{\mathbf{n}}^{(t)}(x, y) \times \mathbf{s}, \theta(x, y) - \arccos\left[\tilde{\mathbf{n}}^{(t)}(x, y) \cdot \mathbf{s}\right]\right) \tilde{\mathbf{n}}^{(t)}(x, y). \tag{3.9}$$

The method is initialised by placing the surface normals on their reflectance cones such that they are aligned in the direction opposite to that of the local image gradient. This initialisation is consistent with the assumption that the object under study is globally convex [156]. The polar angle $\alpha(x, y)$ of the local negative image gradient is given by:

$$\alpha(x, y) = \arctan\left(\frac{-\partial_y I(x, y)}{-\partial_x I(x, y)}\right), \tag{3.10}$$

and hence, each normal is initialised as follows:

$$\mathbf{n}^{(0)}(x, y) = \begin{pmatrix} \sin\theta(x, y)\cos\alpha(x, y) \\ \sin\theta(x, y)\sin\alpha(x, y) \\ \cos\theta(x, y) \end{pmatrix}. \tag{3.11}$$

3.2 A Statistical Model for Surface Normals

Previous work has shown that both images of faces [149] and facial surfaces [5] can be efficiently modeled in a low-dimensional space, derived by applying principal components analysis (PCA) to a training set of images or surfaces. However, fields of surface normals also provide an important source of information from which a statistical model of face-shape can be constructed. A field of surface normals, or needle-map, provides a more detailed description of an object than a corresponding brightness image. Surface normals are invariant to changes in illumination and surface reflectance. Moreover, topographic information such as surface curvature can be computed from a field of surface normals [157]. Using shape-from-shading [156], the field of surface normals is also more easily recovered from an image than the underlying surface height function, since it is orientation and not depth information which is conveyed by variations in image intensity. In the remainder of this section we show how to build a statistical model of surface normal variation by applying PCA to a training sample of fields of surface normals.

3.2.1 Azimuthal Equidistant Projection

The description of a surface by its field of surface normals $\mathbf{n}(x,y)$ projected onto the view plane is known as a *needle-map*. Suppose we have a training set comprising K such needle-maps and that $\mathbf{n}_k(x,y) = (n_k(x,y)_x, n_k(x,y)_y, n_k(x,y)_z)^T$ is the unit surface normal at the pixel indexed (x,y) in the kth training sample. We assume that the face surfaces in the training set have been aligned, such that the pixel (x,y) corresponds to the same point on each face in the training set.

In directional statistics [107], the measure of the average direction of a set of unit vectors is known as the *mean direction* and is given by

$$\hat{\mathbf{n}}_0(x,y) = \frac{\bar{\mathbf{n}}_0(x,y)}{||\bar{\mathbf{n}}_0(x,y)||}, \tag{3.12}$$

where

$$\bar{\mathbf{n}}_0(x,y) = \frac{1}{K}\sum_{k=1}^{K}\mathbf{n}_k(x,y). \tag{3.13}$$

There are other measures of the average direction on a sphere. If the sphere is treated as a manifold, then the intrinsic mean of a distribution of point lying on a sphere is the *spherical median* [107]. In Chapter 4 we return to these alternative representations. However, we found that aligned facial needle-maps produce surface normal distributions which are highly Fisherian [54] (i.e. have a high concentration parameter) and hence the mean direction is a very good approximation to the spherical median. Because of the computational ease of calculating the mean direction we use this as our definition of the average surface normal in this chapter.

On the unit sphere, the surface normal $\mathbf{n}_k(x,y)$ has elevation angle $\theta_k(x,y) = \frac{\pi}{2} - \arcsin n_k(x,y)_z$ and azimuth angle $\alpha_k(x,y) = \arctan\frac{n_k(x,y)_y}{n_k(x,y)_x}$, while the mean surface normal at the location (x,y) has elevation angles $\theta_0(x,y) = \frac{\pi}{2} - \arcsin \hat{n}_0(x,y)_z$ and azimuth angle $\alpha_0(x,y) = \arctan\frac{\hat{n}_0(x,y)_y}{\hat{n}_0(x,y)_x}$.

We wish to transform the field of surface normals into a representation in which we can apply standard linear PCA. As a rich source of transformations from the sphere to a plane, we turn to the field of cartography and specifically the azimuthal equidistant projection. To construct the azimuthal equidistant projection we proceed as follows. We commence by constructing the tangent plane to the unit-sphere at the location corresponding to the mean-surface normal. We establish a local co-ordinate system on this tangent plane. The origin is at the point of contact between the tangent plane and the unit sphere. The x-axis is aligned parallel to the local circle of latitude on the unit-sphere.

Under the azimuthal equidistant projection at the location (x, y), the surface normal $\mathbf{n}_k(x, y)$ maps to the point with co-ordinate vector $\mathbf{v}_k(x, y) = (x_k(x, y), y_k(x, y))^T$. The transformation equations between the unit-sphere and the tangent-plane co-ordinate systems are

$$\begin{aligned} x_k(x,y) &= k' \cos\theta_k(x,y) \sin[\alpha_k(x,y) - \alpha_0(x,y)] \\ y_k(x,y) &= k' \Big(\cos\theta_0(x,y) \sin\alpha_k(x,y) - \sin\theta_0(x,y) \cos\theta_k(x,y) \cos[\alpha_k(x,y) - \alpha_0(x,y)] \Big) \end{aligned} \tag{3.14}$$

where $\cos c = \sin\theta_0(x,y) \sin\theta_k(x,y) + \cos\theta_0(x,y) \cos\theta_k(x,y) \cos[\alpha_k(x,y) - \alpha_0(x,y)]$ and $k' = \frac{c}{\sin c}$.

Thus, in Figure 3.3, CP' is made equal to the arc CP for all values of θ. The projected position of P, namely P', therefore lies at a distance θ from the centre of projection and the direction of P' from the centre of the projection is retained. The equations for the inverse transformation from the tangent plane to the unit-sphere are

$$\begin{aligned} \theta_k(x,y) &= \arcsin\big(\cos c \sin\theta_0(x,y) - \tfrac{1}{c} y_k(x,y) \sin c \cos\theta_0(x,y)\big) \\ \alpha_k(x,y) &= \alpha_0(x,y) + \arctan\omega(x,y) \end{aligned} \tag{3.15}$$

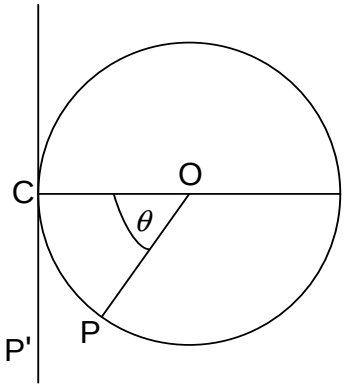


Figure 3.3: The azimuthal equidistant projection.

where

$$\omega(x,y) = \begin{cases} \frac{x_k(x,y)\sin c}{c\cos\theta_0(x,y)\cos c - y_k(x,y)\sin\theta_0(x,y)\sin c} & \text{if } \theta_0(x,y) \neq \pm\frac{\pi}{2} \\ -\frac{x_k(x,y)}{y_k(x,y)} & \text{if } \theta_0(x,y) = \frac{\pi}{2} \\ \frac{x_k(x,y)}{y_k(x,y)} & \text{if } \theta_0(x,y) = -\frac{\pi}{2} \end{cases} \tag{3.16}$$

and $c = \sqrt{x_k(x,y)^2 + y_k(x,y)^2}$.

3.2.2 Point Distribution Model

Suppose that each training example is a range image which consists of an array of depth data. For the pixel indexed (x, y) in the kth training sample the depth is $z_k(x, y)$. Using the range data we estimate the surface normal directions, and the surface normal at the pixel location (x, y) for the kth training image is $\mathbf{n}_k(x, y)$. The components of the vector are transformed into the coordinates $(x_k(x, y), y_k(x, y))$ using the azimuthal equidistant projection. If the range images have $N = X_{\text{res}} \times Y_{\text{res}}$ pixels arranged in X_{res} rows and Y_{res} columns, the surface normal coordinates of each training sample may be represented by the long vector:

$$\mathbf{U}^k = [x_k(1,1), \ldots, x_k(X_{\text{res}}, Y_{\text{res}}), y_k(1,1), \ldots, y_k(X_{\text{res}}, Y_{\text{res}})]^T, \tag{3.17}$$

ordered according to the raster scan (left-to-right and top-to-bottom). Hence, $\mathbf{U}^k$ is a vector of length $2N$ that represents an observed field of surface normals. The first N components of the vector contain the x-coordinates obtained by applying the azimuthal equidistant projection to the surface normals stacked in column order. The second N components of the vector contain the y-coordinates. Since the azimuthal equidistant projection involves centering the local co-ordinate system, the coordinates corresponding to the mean direction are $(0, 0)$ at each image location. Hence, the long-vector corresponding to the mean direction at each image location is zero.

The K training samples can be used to form the $2N \times K$ data-matrix

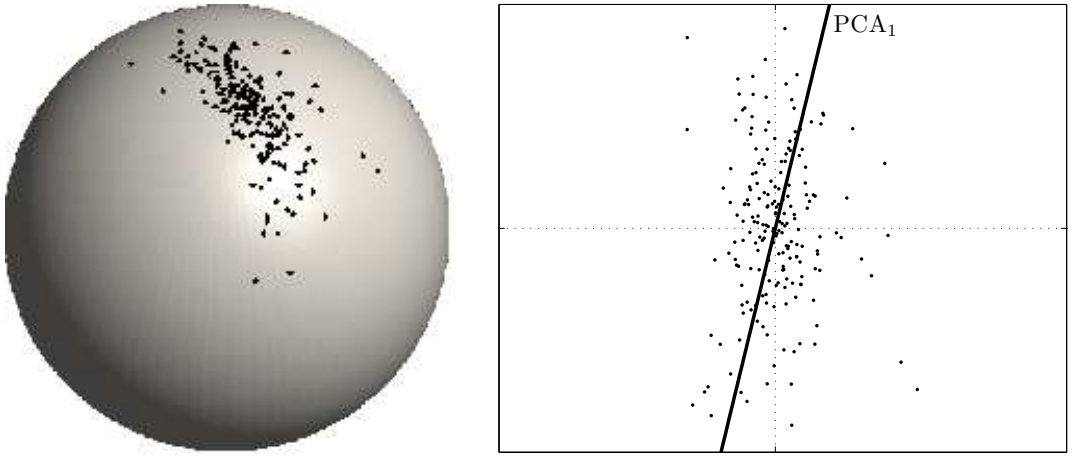
$$\mathbf{D} = [\mathbf{U}^1 | \ldots | \mathbf{U}^K]. \tag{3.18}$$

The $2N \times 2N$ covariance matrix is therefore given by

$$\mathbf{L} = \frac{1}{K}\mathbf{D}\mathbf{D}^T. \tag{3.19}$$

Since in practice, $2N \gg K$, we use the numerically efficient *snap-shot* method of Sirovich [139] to compute the eigenvectors of $\mathbf{L}$. Accordingly, we construct the matrix $\hat{\mathbf{L}} = \frac{1}{K}\mathbf{D}^T\mathbf{D}$ and find the eigenvalues and eigenvectors. The ith eigenvector $\hat{\Psi}_i$ of $\hat{\mathbf{L}}$ can be used to find the ith eigenvector Ψ_i of $\mathbf{L}$ using $\Psi_i = \mathbf{D}\hat{\Psi}_i$. The eigenvalues of $\mathbf{L}$ may also be found using the eigen-decomposition of $\hat{\mathbf{L}}$. Let λ_i be the ith eigenvalue of $\mathbf{L}$ and $\hat{\lambda}_i$ the ith eigenvalue of $\hat{\mathbf{L}}$. For $i \leq K$, $\lambda_i = \hat{\lambda}_i$. For $i > K$, $\lambda_i = 0$.

Figure 3.4 illustrates this process. On the left a distribution of surface normals at one pixel in a model is shown as points on the unit sphere. On the right the azimuthal equidistant projection of the points are shown with the mean point as the centre of projection. The first PCA axis is shown by the line labeled PCA_1. This line corresponds to a great circle on the sphere through the mean direction which minimises the distance on the tangent



(a). (b).

Figure 3.4: Azimuthal equidistant projection of points on the unit sphere (a) to points on the tangent plane at the mean point (b).

plane to each point.

3.3 Fitting the Model to Intensity Images

The ability to fit the model to single intensity images is clearly attractive. The representation of a needle map in the model parameter space provides a compact description of a face which is invariant to changes in illumination and surface reflectance. In addition, using the model to help constrain the fitting process results in an improvement in the shape information recovered from an image and allows an estimate of the albedo map to be made. In this section we present two approaches to fitting the model intensity images. The first of these involves projecting a measured field of surface normals delivered by a geometric shape-from-shading algorithm [156] onto the nearest model configuration in the eigenspace of the training data. The second approach is an iterative one which attempts to ensure that the recovered field of surface normals satisfies Lambert's law. In other words the model provides a statistical constraint which guides the geometric shape-from-shading algorithm.

3.3.1 Projection onto the Model Eigenspace

Worthington and Hancock [156] used a local smoothness constraint regulated by a robust kernel to update the field of surface normals within the geometric shape-from-shading framework. We use this constraint to recover a field of surface normals from an intensity image, and then fit the model to the estimated normals using the matrix multiplication $\mathbf{b} = \mathbf{P}^T\mathbf{U}_o$, where $\mathbf{U}_o$ is the observed needle-map under azimuthal equidistant projection and $\mathbf{P} = [\Psi_1|\ldots|\Psi_K]$ the eigenvectors of the model. The resulting algorithm is described in Algorithm 3.1.

Existing shape-from-shading algorithms are not capable of reliably recovering accurate needle-maps from real world images [165]. Hence the shape information recovered using the geometric shape-from-shading algorithm is likely to be of a low quality. Nevertheless, we investigate this approach in our experimental section to contrast the performance of existing shape-from-shading algorithms with the method we propose below.

3.3.2 Combining the Statistical Model and Geometric SFS

A more attractive alternative is to use the statistical constraint provided by the model itself in the process of fitting the model to an intensity image. Once trained, the statistical model represents the space of valid face shapes. We can exploit this prior knowledge in order to help resolve the ambiguity in the shape-from-shading process. We do this using an iterative approach which can be posed as that of recovering the best-fit field of surface normals from the statistical model, subject to constraints provided by the image irradiance equation. As noted above, when the surface reflectance follows Lambert's law, then the surface normal is constrained to fall on a cone whose axis is in the light source direction and whose opening angle is the inverse cosine of the normalised image brightness.

This method commences from an initial configuration in which the surface normals

Algorithm 3.1: Model fitting by projecting a field of surface normals estimated using shape-from-shading onto the model.

Input: Light source $\mathbf{s}$, image intensities $I(x,y)$, robust smoothing kernel ϱ, statistical model $\mathbf{P}$ and average normals $\hat{\mathbf{n}}_0(x,y)$

Output: Estimated field of on-cone surface normals $\mathbf{n}(x,y)$ and best-fit model normals $\tilde{\mathbf{n}}(x,y)$

Set iteration $t = 0$;

Initialise each $\mathbf{n}^{(0)}(x,y)$ using Equation 3.11;

repeat

  Enforce smoothness constraint using robust regulariser: $\tilde{\mathbf{n}}^{(t)}(x,y) = \varrho\mathbf{n}^{(t)}(x,y)$;

  Rotate back to closest on-cone position: $\mathbf{n}^{(t+1)}(x,y) = \Theta\left(\tilde{\mathbf{n}}^{(t)}(x,y) \times \mathbf{s}, \arccos\left[\mathbf{I}(\mathbf{x},\mathbf{y})\right] - \arccos\left[\tilde{\mathbf{n}}^{(\mathbf{t})}(\mathbf{x},\mathbf{y}) \cdot \mathbf{s}\right]\right)\tilde{\mathbf{n}}^{(t)}(x,y)$;

  Set iteration $t = t + 1$;

until $\sum_{x,y}\left[\arccos\left(\mathbf{n}^{(t)}(x,y) \cdot \mathbf{n}^{(t-1)}(x,y)\right)\right]^2 < \epsilon$;

Each normal in the estimated field $\mathbf{n}(x,y)$ undergoes an azimuthal equidistant projection (Equation 3.14) to give a vector of transformed coordinates $\mathbf{U}_o$;

The vector of best fit model parameters is given by $\mathbf{b} = \mathbf{P}^T\mathbf{U}_o$;

The best fit needle-map $\tilde{\mathbf{n}}(x,y)$ is given by the inverse azimuthal equidistant projection (Equation 3.15) of the best fit vector of transformed coordinates $\mathbf{U} = \mathbf{P}\mathbf{P}^T\mathbf{U}_o$;

reside on the irradiance cone and point in the direction of the local image gradient. The statistical model is fitted to recover a revised estimate of the surface normal directions. The best-fit surface normals are then projected onto the nearest location on the irradiance cones. Our approach to fitting the model to intensity images is hence an iterative process in which we interleave the process of fitting the statistical model to the current field of estimated surface normals, and then re-enforcing the data-closeness constraint provided by Lambert's law by mapping the surface normals back onto their reflectance cones. The height map for the surface is recovered by integrating the final field of surface normals using the method of Frankot and Chellappa [59]. The algorithm is summarised in Algorithm 3.2.

Algorithm 3.2: Model fitting by iteratively combining the cone constraint with the statistical model constraint.

Input: Light source $\mathbf{s}$, image intensities $I(x,y)$, statistical model $\mathbf{P}$ and average normals $\hat{\mathbf{n}}_0(x,y)$

Output: Estimated field of on-cone surface normals $\mathbf{n}(x,y)$ and best-fit model normals $\tilde{\mathbf{n}}(x,y)$

1 Set iteration $t = 0$;

2 Initialise each $\mathbf{n}^{(0)}(x,y)$ using Equation 3.11;

3 **repeat**

4 Each normal in the estimated field $\mathbf{n}^{(t)}(x,y)$ undergoes an azimuthal equidistant projection (Equation 3.14) to give a vector of transformed coordinates $\mathbf{U}_o$;

5 The vector of best fit model parameters is given by $\mathbf{b} = \mathbf{P}^T\mathbf{U}_o$;

6 The best fit needle-map $\tilde{\mathbf{n}}^{(t)}(x,y)$ is given by the inverse azimuthal equidistant projection (Equation 3.15) of the best fit vector of transformed coordinates $\mathbf{U} = \mathbf{P}\mathbf{P}^T\mathbf{U}_o$;

7 Rotate back to closest on-cone position: $\mathbf{n}^{(t+1)}(x,y) = \Theta\left(\tilde{\mathbf{n}}^{(t)}(x,y) \times \mathbf{s}, \arccos\left[\mathbf{I}(\mathbf{x},\mathbf{y})\right] - \arccos\left[\tilde{\mathbf{n}}^{(\mathbf{t})}(\mathbf{x},\mathbf{y}) \cdot \mathbf{s}\right]\right) \tilde{\mathbf{n}}^{(t)}(x,y)$;

8 Set iteration $t = t + 1$;

9 **until** $\sum_{x,y}\left[\arccos\left(\mathbf{n}^{(t)}(x,y) \cdot \mathbf{n}^{(t-1)}(x,y)\right)\right]^2 < \epsilon$;

Initialisation

We have experimented with a number of different initialisations for the fitting process given above and found that the final solution is fairly insensitive to the choice of initialisation, even to the extent that every normal can be initialised to point in the same direction on their reflectance cones. However, in this chapter we retain the negative image gradient of Worthington and Hancock [156]. In Chapter 4 we consider an alternative initialisation which exploits the statistical model and results in a lower starting error and reduces the number of iterations required to achieve convergence.

3.3.3 Practical Considerations

Upon convergence, we have a choice between two solutions. On the one hand, the 'on-cone' surface normal $\mathbf{n}(x,y)$ provides a strict solution of the image irradiance equation

within the framework of geometric shape-from-shading. This represents a data-driven solution in that reillumination of the recovered normals with the input lighting configuration will exactly reproduce the original image. On the other hand, the 'best-fit' surface normal $\tilde{\mathbf{n}}(x, y)$ provides a least squares fit of the statistical model to the shape-from-shading normals. This represents a model-driven solution in which the field of surface normals is globally constrained to correspond to a valid face shape.

In the ideal Lambertian, unit albedo case, the strict on-cone solution is likely to provide a more accurate solution. It is also likely that this solution will recover more of the fine surface detail and discriminating features which are not captured by the model. However, real world face images contain albedo variations caused by skin pigmentation and facial hair. By enforcing data-closeness, pixels of low albedo will be interpreted as having large incident angles. In Figure 3.5 we show the angular change as data-closeness is restored five final best fit needle-maps, i.e. the angular difference between the best-fit surface normal $\tilde{\mathbf{n}}(x, y)$ and the corresponding on-cone surface normal $\mathbf{n}(x, y)$ at every pixel. We show the input images in the first row. From the plots it is clear that the changes are almost solely due to the variation in albedo at the eyes, eye-brows, lips and facial hair. Aside from these regions there is very little change in surface normal direction, indicating that the needle-map has converged to a solution which satisfies the data-closeness constraint except in regions of actual variation in albedo. In this case, the best-fit of the statistical model may in fact provide a more accurate estimate of the underlying facial shape. We provide empirical evidence of this observation in our experimental results.

Albedo and Data-closeness

For a real world image, should we choose the best-fit normals $\tilde{\mathbf{n}}(x, y)$ as our estimate of the underlying facial shape, we can still ensure satisfaction of the data-closeness constraint in some sense. To do so, we allow albedo to vary. In other words we relax the data-

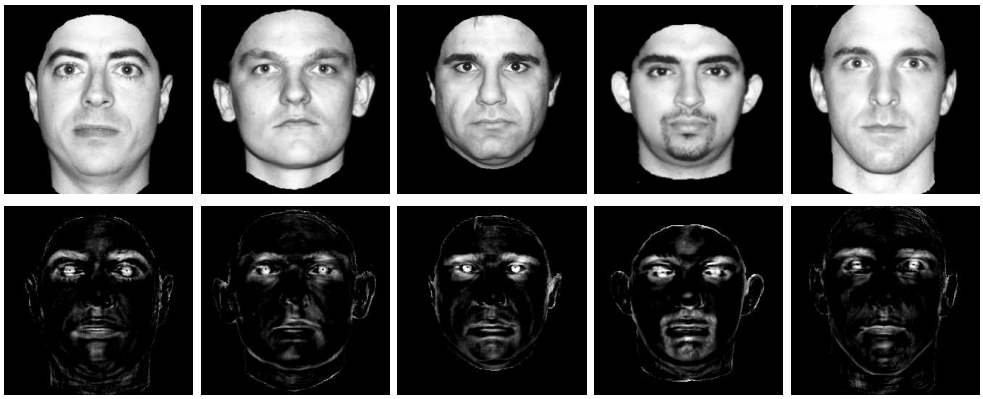


Figure 3.5: Angular difference between $\mathbf{n}(x, y)$ and $\tilde{\mathbf{n}}(x, y)$ at convergence (second row). Input images shown in the first row.

closeness constraint at the final iteration and account for differences between predicated and observed image brightness by variations in albedo. The imposition of data-closeness in previous iterations ensures the model-fit has been encouraged towards a solution which closely satisfies the constraint. If the final best-fit field of surface normals is reilluminated using a Lambertian reflectance model, then the predicted image brightness is given by

$$I(x, y) = \rho_d(x, y)[\tilde{\mathbf{n}}(x, y) \cdot \mathbf{s}], \tag{3.20}$$

where $\rho_d(x, y)$ is the Lambertian albedo at position (x, y). Since I, $\mathbf{s}$, and $\tilde{\mathbf{n}}$ are all known, we can estimate the albedo at each pixel using the formula:

$$\rho_d(x, y) = \frac{I(x, y)}{\tilde{\mathbf{n}}(x, y) \cdot \mathbf{s}}. \tag{3.21}$$

The combination of the best-fit needle-map and corresponding estimated albedo map will still exactly recreate the input image when reilluminated with the original lighting. Additionally, albedo estimated in this manner may vary arbitrarily. This means distinguishing pigmentation or facial hair is accurately recovered.

In Section 3.4 we demonstrate how the strict solution within the geometric shape-

from-shading framework represents a significant improvement over the needle maps estimated using the original curvature consistency constraints proposed by Worthington and Hancock [156]. Moreover, we show how the best-fit needle-map and estimated albedo result in realistic synthesised images under novel illumination and viewpoint, of comparable accuracy to the more computationally intensive analysis-by-synthesis approach of Blanz and Vetter [14].

Alignment and Reflectance Normalisation

In the above analysis we have assumed that during training the sample facial surfaces have been aligned, i.e. that a pixel (x, y) corresponds to the same point on all sample faces. Furthermore, our shape-from-shading algorithm assumes that the input image has been aligned with the model. In this thesis we have chosen not to address the issue of alignment and we have therefore used a combination of existing methods and manual alignment which we describe below. Clearly, there is scope for integrating a fine-scale alignment process into the iterative model fitting. However, this would serve to confuse the main contribution of this chapter, a method for accurate facial shape recovery. Moreover, the fact that good results are obtainable with only an approximate alignment is a strength of the method.

During training we use a database of facial range images in which the surfaces have been aligned using an existing registration algorithm [14]. Prior to applying our shape-from-shading algorithm the input images were manually cropped to remove background, hair and clothing. Finally, the images were aligned and scaled to be brought into correspondence with the model using a number of manually marked feature points.

As is common in the vision literature, we have also assumed that skin reflects light according to Lambert's law. However, to improve the real world applicability of the approach we use a preprocessing step to correct for deviations from Lambertian reflectance.

Rather than incorporate a reflectance model that tries to capture some of the properties of skin reflectance, we exploit a recently reported image based reflectance estimation method [135].

This is a computationally efficient process of complexity $O(N)$, for an N pixel image, which provides a good estimate of an isotropic, monotonic BRDF from a single image. By assuming the surface under study is locally spherical, the method uses the local image gradient to find correspondences between points on the surface and points on a conceptual reflectance sphere. An estimate may then be made of the radiance function which relates the incidence angle to measured intensity. Lambertian correction is then simply a case of inverting the radiance function for each pixel to recover the incident angle, from which the Lambertian intensity can be calculated.

3.4 Experiments

In this section we present experiments with our method. There are three elements to this study. We commence by examining the model when trained on fields of surface normals extracted from range data. Second, we show the results of fitting the model to intensity data, and show the surface height data that can be reconstructed from the fitted fields of surface normals. Third, we illustrate how the fitted models can be used synthesise realistic novel facial views.

3.4.1 Model Training

We begin by describing how our model is constructed from real-world data. We commence by building a "ground truth" model using fields of surface normals extracted from range data. This allows us to show the utility of the model in capturing facial shape in a compact manner when trained on relatively 'clean' data. We used the 3DFS dataset [1]

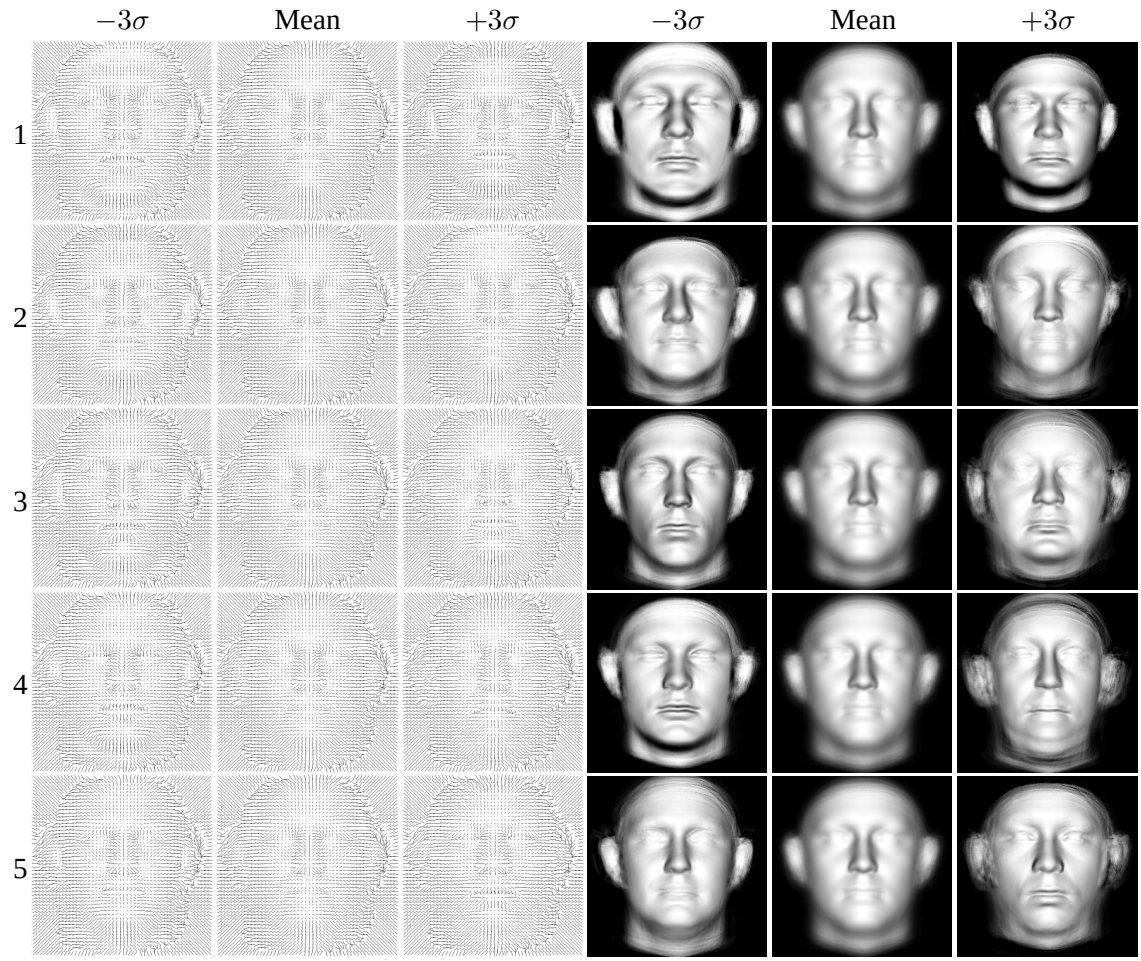


Figure 3.6: The first five modes of variation of a statistical surface normal model trained on a set of facial needle-maps extracted from range data. The deformed needle-maps are visualised as: fields of vectors (first three columns), and rendered with Lambertian reflectance and frontal illumination (second three columns). The mean face is shown in the central column and $\pm$ 3 standard deviations along each of the first 5 principal modes of variation are shown in the first and third columns.

which consists of 100 high resolution scans of subjects in a neutral expression. The scans were collected using a *Cyberware*$^{\text{TM}}$ 3030PS laser scanner. The database is pre-aligned, registration being performed using the optical flow correspondence algorithm of Blanz and Vetter [14]. Fields of surface normals were extracted by orthographically projecting the 3 surface normal components onto a view plane positioned fronto-parallel to the aligned faces.

Figure 3.7: Visualisation of local variation in facial surface orientation. Plot of model concentration shown as $log(\kappa)$, where κ is the von Mises-Fisher concentration parameter.

We begin by examining the principal modes of variation for a model trained on fields of surface normals derived from range images of faces. In Figure 3.6 we show the first 5 modes of variation of this model. In each case we deform the points under azimuthal equidistant projection by ± 3 standard deviations along each of the first 5 principal axes. We then perform the inverse azimuthal equidistant projection and visualise the resulting needle-maps as fields of vectors (first three columns) and rendered with Lambertian reflectance and frontal illumination (second three columns). The modes encode shape only, since the needle-maps are invariant to illumination conditions and the training set contained no variation in expression. The modes clearly capture distinct facial characteristics. For example, mode 1 encodes head size and also seems to be correlated with gender. This is manifested in the broader jaw, brow and nose in the negative direction, all of which are masculine features. The third mode encodes the difference between long, narrow faces and short, wide faces, whereas the second mode encodes the difference between a pointy and a rounded chin. Note that there is a real sense of 3D deformation in the rendered images, caused only by locally rotating the surface normals along great circles.

In Figure 3.7 we provide a visualisation of the variability in the shape of different facial regions. Since it is the natural model for multivariate directional data we fit a von Mises-Fisher (vMF) distribution to the distribution of surface normals at each location. The vMF distribution is analogous to the multivariate Gaussian distribution in $\mathbb{R}^q$ [54].

The distribution is specified by two parameters: the mean direction vector, which in our case is $\hat{\mathbf{n}}_0$, and the concentration parameter κ. For $q = 3$, the distribution over the vector $\mathbf{n}$ is described by the following probability density function:

$$f(\mathbf{n}; \hat{\mathbf{n}}_0, \kappa) = \frac{\kappa^{\frac{1}{2}}}{2\pi^{\frac{3}{2}} I_{\frac{1}{2}}(\kappa)} e^{\kappa \hat{\mathbf{n}}_0^T \mathbf{n}}, \tag{3.22}$$

where $I_{\frac{1}{2}}$ is a modified Bessel function of the first kind and order $\frac{1}{2}$. Using the algorithm in [77] we estimate κ at each location and plot $\log(\kappa)$ in Figure 3.7. The larger the value of κ, the greater the density of the distribution of surface normal directions around the mean direction. Thus, white corresponds to an area of very low variance, and black to an area of very high variance. The plot demonstrates that much of the variation in facial surface orientation is confined to the eyes, nose, lips and chin, while the cheeks and forehead vary less between subjects.

3.4.2 Fitting the Model to Data

In this section we explore the fitting of the model to data using the two methods outlined in Section 3.3. We examine the influence of illumination direction, the number of iterations, the number of model dimensions and non-constant albedo on the accuracy of the recovered surface normals. We use a larger database of range scans for this study in order to provide ground truth data. In this case the model is trained on 180 examples, with 20 retained as ground truth to measure out-of-sample fitting accuracy.

Fitting the model to needle-maps

We begin by fitting the model to fields of surface normals extracted from an image using shape-from-shading, as described in Section 3.3.1. We use the shape-from-shading technique of Worthington and Hancock [156]. In Figure 3.8 we show an example of fitting to

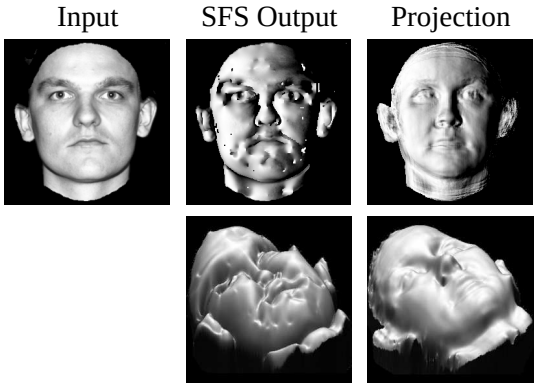


Figure 3.8: Behaviour of the projection fitting process. The input image is shown on the top left. The first row shows the needle map delivered by SFS followed by its projection onto the model eigenspace. The needle maps are reilluminated by a light source with direction $\mathbf{s} = (-1, 0, 1)^T$. The second row shows the surfaces recovered from the corresponding needle maps.

a field of surface normals extracted from a real world image. The input image is shown on the left, followed by the needle-map recovered using shape-from-shading. The projection of the needle-map onto the model eigenspace is shown on the right. Since the recovered needle-map is guaranteed to satisfy data-closeness, it would appear identical to the input image when rendered with a light source from the original direction (**s**=$(0, 0, 1)^T$). For this reason in the top row we show the needle-maps reilluminated with a light source moved along the negative x-axis to subtend an angle of 45° with the viewing direction. From the reilluminations it is clear that there are severe flaws in the initial needle-map recovered using shape-from-shading. Although the gross distribution of intensity appears correct, the resulting image is noisy and not realistic and suffers from convex/concave errors and feature implosions. This is made explicit in the corresponding recovered surface shown in the second row. The projection onto the model eigenspace helps resolve some of these errors, resulting in a more plausible image. From the corresponding recovered surface it is clear that many of the feature implosions are corrected and noise reduced. However, given the gross errors in the initial needle-map, it seems unlikely that this represents the

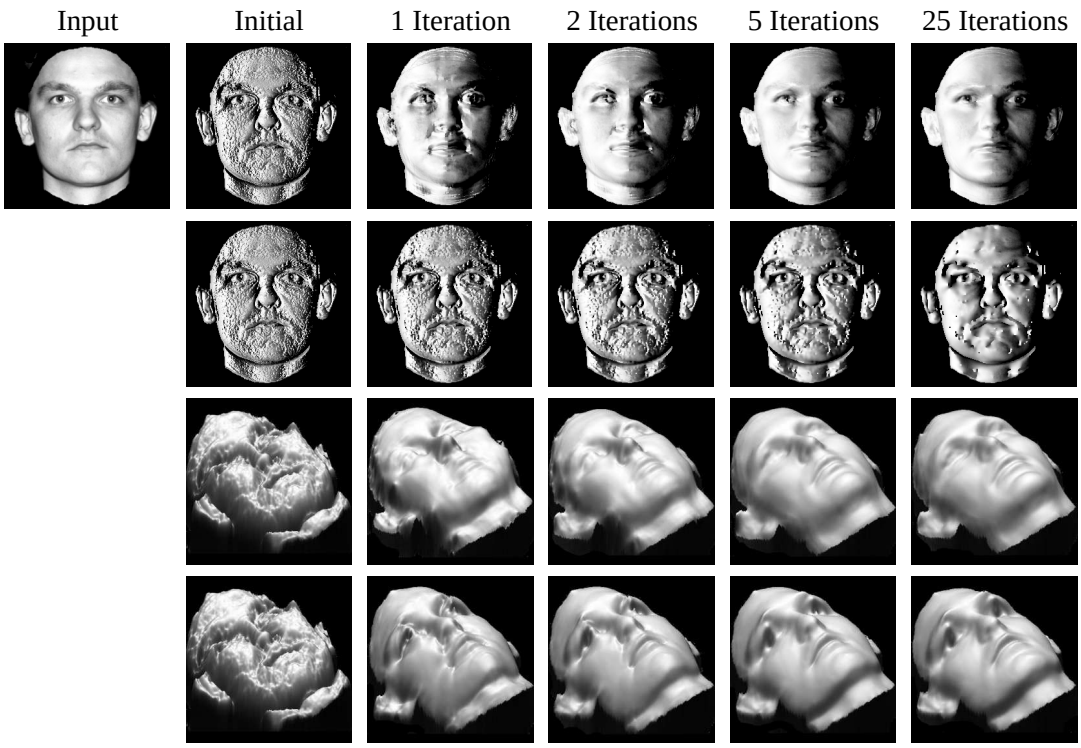


Figure 3.9: Behaviour of the iterative fitting process over 25 iterations. The input image is shown on the top left. The first row shows the recovered needle maps reilluminated by a light source with direction $\mathbf{s} = (-1, 0, 1)^T$. For comparison the second row shows similarly reilluminated needle maps recovered by the Worthington and Hancock algorithm. The third and fourth rows show the surfaces recovered from $\tilde{\mathbf{n}}(x, y)$ (third row) and $\mathbf{n}(x, y)$ (fourth row).

closest model fit of the underlying facial shape of the input face.

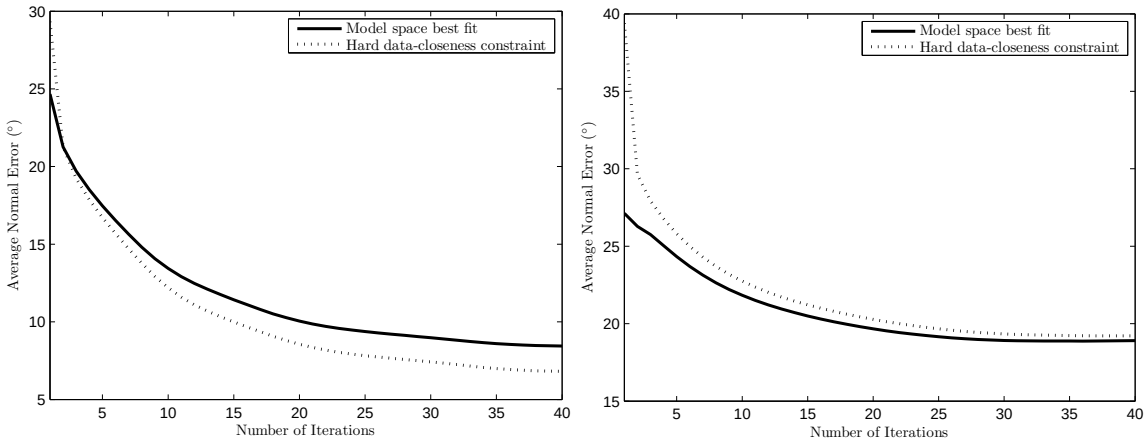
Fitting the model to intensity images

We now show how the statistical model may be fitted to intensity data using the method outlined in Section 3.3.2. We commence by considering the iterative behaviour of the algorithm. The top row of Figure 3.9 shows how a needle map develops over 25 iterations of the algorithm. In this figure the input face is shown on the top left. Since the needle maps satisfy data-closeness at every iteration, they would all appear identical when rendered with a light source from the original direction ($\mathbf{s} = (0, 0, 1)^T$). For this reason in

the top row we show the needle maps reilluminated with a light source moved along the negative x-axis to subtend an angle of 45° with the viewing direction. After one iteration there is a significant global improvement in the recovered needle map. Subsequent iterations make more subtle improvements, helping to resolve convex/concave errors and sharpening defining features. For comparison the second row shows the corresponding needle maps recovered using the original curvature consistency constraint of Worthington and Hancock [156] reilluminated in the same manner. Although there is a steady improvement in the quality of the recovered normals, there are gross global errors as well as feature implosions around features such as the nose.

In Figure 3.9 we also show the surfaces recovered from the current best fit needle maps, $\tilde{\mathbf{n}}(x,y)$, (third row) and the needle maps which satisfy data-closeness, $\mathbf{n}(x,y)$, (bottom row) as the algorithm iterates. Surface recovery is effected using the method of Frankot and Chellappa [59]. As one would expect, the imposition of data-closeness results in errors in the recovered surface where there is variation in albedo, most notably around the eyes and eye-brows. In both sets there is a clear improvement in the recovered surface as the algorithm iterates. The implosion of the nose is corrected, the surface becomes smoother and finer details become evident, for example around the lips.

In Figure 3.10 we provide a quantitative analysis of the iterative behavior of the fitting process using ground truth data. For each of the 20 out-of-sample faces we render the needle-map with a light source situated at the viewpoint to yield an image to which we apply the fitting process. In (a) we render the needle-maps with Lambertian reflectance and unit albedo, whereas in (b) we render the needle-maps with the albedo recorded by the Cyberware scanner. We plot the average normal error across all needle-maps in degrees against the number of iterations. The solid curve shows the error for the best-fit needle map in the model space, $\tilde{\mathbf{n}}(x,y)$, and the broken curve for the needle-map with data-closeness enforced, $\mathbf{n}(x,y)$. From both plots it is clear that the algorithm converges

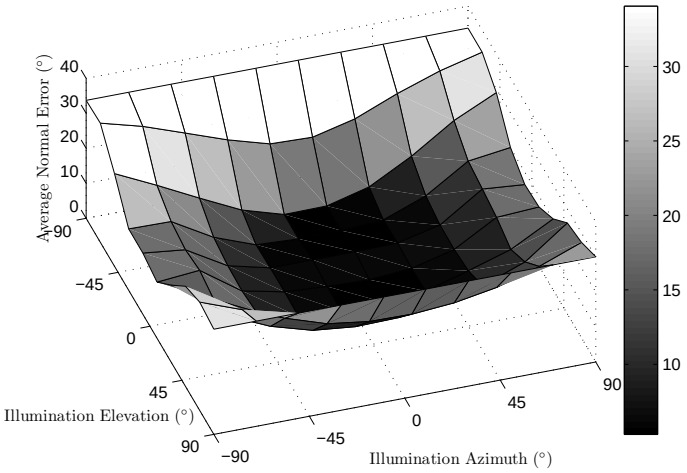


(a). (b).

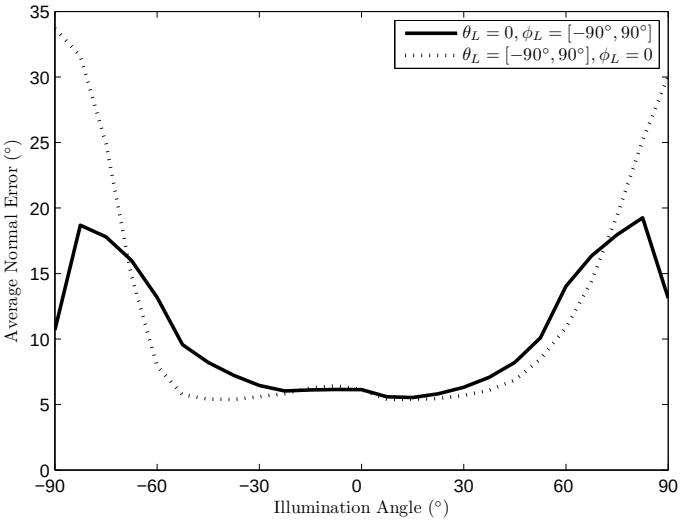
Figure 3.10: Average normal error versus number of iterations of the algorithm when fitting to synthetic data with: (a) constant albedo (b) varying albedo.

rapidly, particularly in the first few iterations. As would be expected, the introduction of varying albedo reduces the accuracy of the recovered normals. An interesting feature to note is that in (a) when fitting to images with constant albedo, the imposition of data-closeness results in higher accuracy. However, in (b) when fitting to images with varying albedo, this is reversed and the best-fit to the needle-map in the model space gives higher accuracy and hence is closer to the underlying shape. This adds quantitative evidence to our observation that on real world images, $\tilde{\mathbf{n}}(x, y)$ is a better estimate of the true facial shape.

In Figure 3.11 we examine the influence of illumination direction on the accuracy of the recovered needle-map. Again, we render the out-of-sample ground truth needle-maps with Lambertian reflectance and unit albedo but we vary the elevation (θ_L) and azimuth (α_L) of the light source through $(-90°, 90°)$. When $\theta_L = \alpha_L = 0$ the light source direction is coincident with the viewing direction, when $\alpha_L < 0$ it is from the left, when $\theta_L < 0$ it is from below and vice versa. The plot in (a) shows the average normal error over all out-of-sample needle-maps in degrees as both θ_L and α_L are varied. For clarity, in (b) we show contours from (a) in which only the azimuth is varied (solid line) and only



(a).



(b).

Figure 3.11: Analysis of the accuracy of the fitting process under varying illumination. (a) Average error in the estimated surface normal direction as both the azimuth and elevation of the light source are varied. (b) Contours from (a) in which the elevation of the light source is set to 0 and the azimuth varied (solid line) and the azimuth is set to 0 and the elevation varied (broken line).

the elevation is varied (broken line). From both plots it is clear that the process is robust to variations in illumination of $\pm 45^\circ$ of both azimuth and elevation (note the plateau of values under 10°). Another feature evident from the plots is that extreme elevation values have a greater detrimental effect than extreme azimuth angles.

Since faces are approximately bilaterally symmetric, the errors introduced by varying α_L (i.e. moving the light source left and right) are approximately symmetric. A more interesting feature is visible when θ_L is varied (i.e. the light source is moved up and down). In both Figure 3.11 (a) and (b) it is clear that greater errors are introduced when θ_L is large and negative than when large and positive. In other words, using our algorithm, it is harder to recover facial shape when the light source is moved below the face than above. This potentially has some interesting psychological implications. Typically, poor human performance when perceiving images of faces illuminated from below has been attributed to an illumination direction estimation process in which the light source is constrained to come from above [89]. However, with perfect knowledge of the illumination direction, our algorithm recovers less accurate facial shape information when the illumination is from below compared to above. This suggests that there is something implicit in human face shape which makes it harder to recover shape accurately when illuminated from below. Hence, at least part of the degradation noted in human performance may be related to this fact.

Finally, in Figure 3.12 we investigate the effect of the number of eigenmodes (or dimensions) retained in the model on the accuracy of the recovered facial shape. We plot the average normal error at convergence versus the number of eigenmodes retained. Increasing the number of eigenmodes retained has the effect of increasing the accuracy. However, the effect is more evident for the best-fit needle map than the on-cone surface normal. Moreover, only the first 40 or so dimensions have a significant impact on the accuracy. In other words, even a limited number of dimensions provides a statistical con-

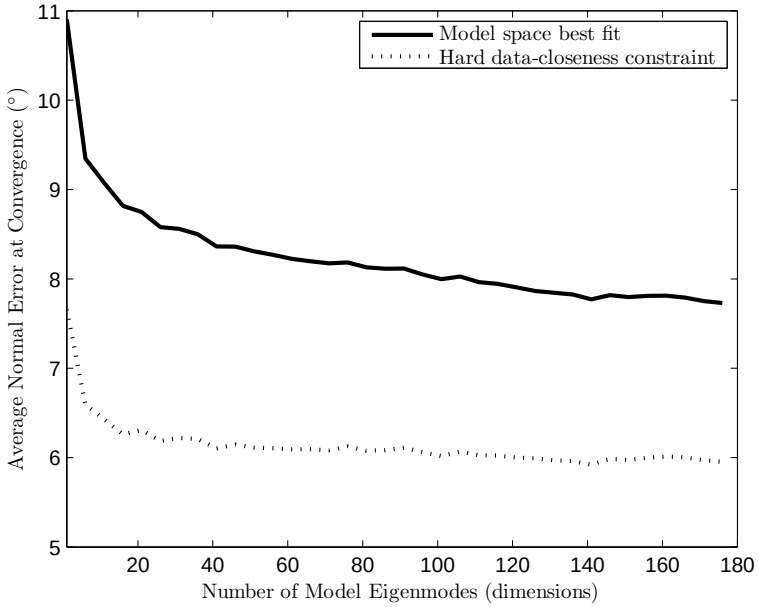


Figure 3.12: Average surface normal error at convergence versus the number of model eigenmodes retained.

straint which is sufficient to guide the shape-from-shading process to an accurate solution when combined with the hard data-closeness constraint.

Comparing Iterative Fitting and Fitting to Shape-from-shading Normals

In Figure 3.13 we quantitatively compare the two approaches to fitting the model using ground truth data. In the top row we fit the model to surface normals recovered using shape-from-shading, in the bottom row we use the iterative fitting process. The estimated needle-map is shown in column 2, reilluminated by a light source with direction $\mathbf{s} = (-1, 0, 1)^T$. For comparison, the ground truth needle map is shown similarly reilluminated in the third column. It is clear that using shape-from-shading normals results in major structural differences between the best fit and ground truth needle map, in particular the exploded cheeks and inaccurate jaw line. In the fourth column the angular error for each surface normal is shown. There are clearly large errors around the nose, jaw and eyes as well as evidence of general noise. The overall average angular error of the fitted

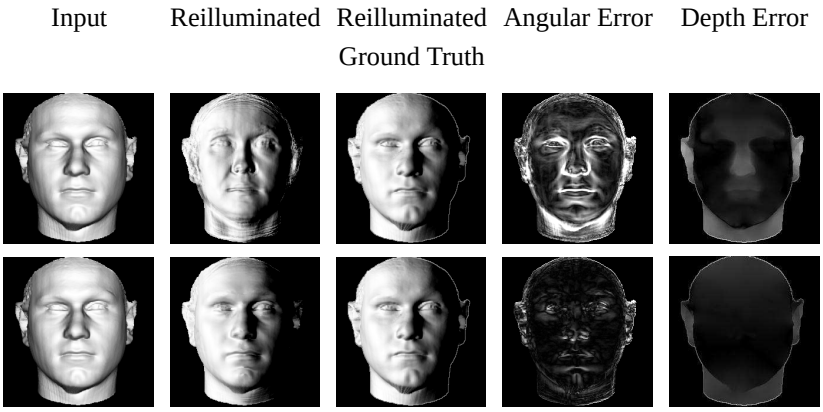


Figure 3.13: Accuracy of fitting process on ground truth data. From left to right: the input image, the recovered needle-map reilluminated by a light source with direction $\mathbf{s} = (-1, 0, 1)^T$, the ground truth needle map similarly reilluminated, the angular error and the height error. The projection fitting method is shown on the top row, the iterative fitting method is shown on the bottom row.

needle-map was 13.25°. In contrast, the needle-map recovered using the iterative fitting process appears very close to the reilluminated ground truth needle-map with no obvious errors. An interesting exception is the dimple in the chin which is not visible in the input image, but which becomes visible when the ground truth needle map is reilluminated from a different direction. The fitting process fails to recover this structure. The angular errors are much reduced, mainly limited to the occluding boundary. The overall average angular error of the fitted needle map was 3.93°. This represents more than 3 times greater accuracy than the projection fitting method. In the final column we also show an error plot of the reconstructed height. Similar results were obtained with a variety of ground truth data.

3.4.3 Synthesising Novel Views

In this section we focus on how the fitted models can be used for the purposes of novel view synthesis. We focus on the results obtained using the iterative method outlined in Section 3.3.2, since from the previous section this appears to outperform the projection

method. The data used in this study comes from the Yale-B database [64]. In the images the faces are in a frontal pose and were illuminated by a point light source situated approximately at the viewpoint, i.e. in direction $\mathbf{s} = (0, 0, 1)^T$.

In Figures 3.14 and 3.15 we present the results of the albedo estimation process and show synthesised images under novel illumination. In the first column we show the input images of the ten subjects from the Yale-B database. In the second column we show the albedo maps estimated using Equation 3.21. The results of the albedo estimation process appear intuitively convincing. For instance, the albedo map identifies the eyes, eyebrows, facial hair, nostrils and lips. Moreover, there are no residual shading effects in the albedo map, for example the nose is assigned approximately constant albedo. The method works well because albedo is approximately constant across much of a face's surface and hence the gross structure of the needle map is not overly disrupted by the imposition of data-closeness at each iteration.

The remaining four columns of Figures 3.14 and 3.15 show the best fit needle maps reilluminated with Lambertian reflectance and the estimated albedo maps. The light source is moved to subtend an angle of 45° with the view direction along the positive and negative x and y-axes. The needle maps show considerable stability under large changes in illumination direction and result in near photo-realistic synthesised images.

In Figure 3.16 we show synthesised images of the input faces in novel pose. In the first and third rows the surfaces are shown rotated 30° about the vertical axis. As with the previous images, the surfaces are rendered with Lambertian reflectance and the estimated albedo maps. The light source remains fronto-parallel with respect to the face (i.e. from the original direction). The resulting synthesised images are near photo-realistic under a large change in viewpoint. Certainly, the results are comparable with those of Georghiades et al. [64] in which 7 input images were used per subject. Rows 2 and 4 of Figure 3.16 show the meshes of the recovered surfaces to allow inspection of the recovered shape

Figure 3.14: Column 1 shows the input images of subjects 1-5 from the Yale B database. Column 2 shows the estimated albedo maps. Columns 3-6 show synthesised views of the subjects under novel illumination. The light source directions are $\mathbf{s} = (-1, 0, 1)^T$, $(1, 0, 1)^T$, $(0, 1, 1)^T$ and $(0, -1, 1)^T$ respectively.

alone. In Figure 3.17 we demonstrate that the recovered surface and albedo map are sufficiently stable to synthesise images in both novel pose and novel illumination. We show the surface of subject 8 rendered as in the previous figure, except that the light source is circled from left profile to right profile.

Finally, in Figure 3.18 we demonstrate the use of our method in a practical application. In (a) we have synthesised novel poses of famous faces from frontal images for use as stimuli in a neuropsychological study. The input images have been texture mapped onto

Figure 3.15: Column 1 shows the input images of subjects 6-10 from the Yale B database. Column 2 shows the estimated albedo maps. Columns 3-6 show synthesised views of the subjects under novel illumination. The light source directions are $\mathbf{s} = (-1, 0, 1)^T$, $(1, 0, 1)^T$, $(0, 1, 1)^T$ and $(0, -1, 1)^T$ respectively.

the recovered shape and so in effect the illumination remains stationary with respect to the face. In this case, little control was exercised over the conditions present in the input images, which include varying illumination direction and facial expression, yet we are still able to synthesise useful images of the subjects in novel poses. In (b) we show the recovered surfaces, rendered with Lambertian reflectance and rotated 30° from frontal. Distinguishing features of the input face shape have been clearly recovered.

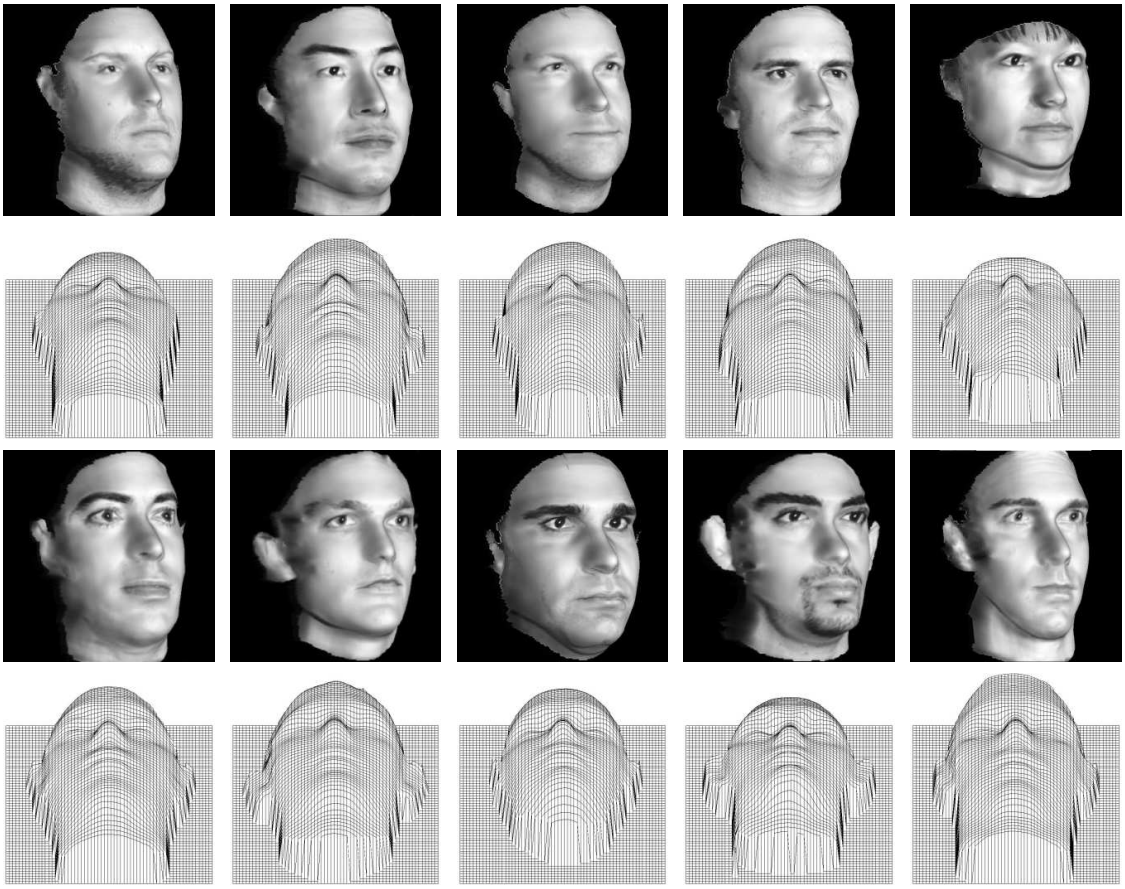
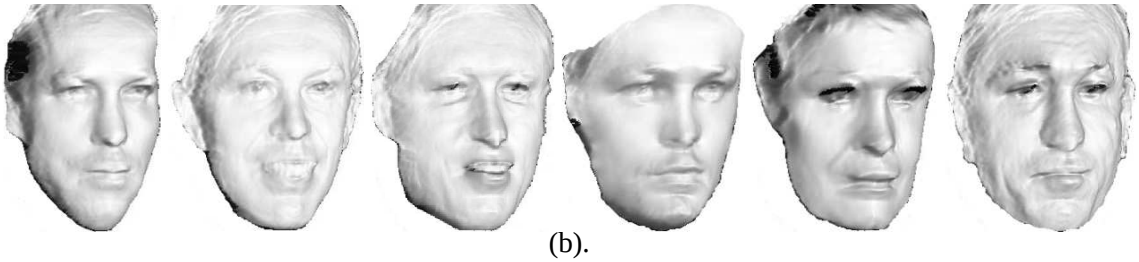


Figure 3.16: Surfaces recovered from the ten subjects in the Yale B database. In the first and third rows, the surfaces are rendered with their estimated albedo maps, Lambertian reflectance and are shown rotated 30° about the vertical axis. The light source remains fronto-parallel with respect to the face. In the second and fourth rows the surface meshes are shown rotated 40° about the horizontal axis.

Figure 3.17: Surface recovered from subject 8 of the Yale B database. The surface is again rendered with the estimated albedo map, Lambertian reflectance and rotated 30° about the vertical axis. The light source is circled from full left profile to full right profile with respect to the face.

(a).



(b).

Figure 3.18: (a) Synthesised views of famous faces. Input images are shown in the central column, synthesised poses of $-24°$, $-12°$, $12°$ and $24°$ from frontal are shown in columns 1, 2, 4 and 5 respectively. (b) Recovered shape shown rendered with Lambertian reflectance and rotated $30°$ from frontal.

3.5 Conclusions

In this chapter we have shown how a statistical model of shape may be constructed from fields of surface normals using the azimuthal equidistant projection. We demonstrated that such a model trained on facial needle maps extracted from range data captures facial shape in a compact manner and is capable of generalising to out-of-sample faces.

We presented two methods for fitting the model to image brightness data. The first of these involves fitting the model to a field of surface normals extracted from the image using shape-from-shading. The second method is an iterative one in which the model is used as a statistical constraint in a geometric shape-from-shading framework. This process can be posed as that of recovering the best-fit field of surface normals from the statistical model, subject to constraints provided by the image irradiance equation. The method proves rapid to converge, is robust to variations in illumination direction of up to 45° from the viewing direction and delivers realistic surfaces when the fields of surface normals are integrated. We then showed how the fitted models could be used to improve the reilluminations of fields of surface normals recovered using shape-from-shading and to estimate an albedo map of the face.

In Chapter 4 we reformulate the statistical model presented in this chapter using the tools of differential geometry. With this more elegant formulation to hand, we show how to improve the accuracy of the recovered shape information by using techniques from robust statistics.

Chapter 4

Incorporating Principal Geodesic Analysis and Robust Statistics

Where there is much light, the shadow is deep.

- Johann Wolfgang von Goethe [153]

The approach developed in the previous chapter allows shape and albedo information to be recovered from single images which is of a significantly higher quality than would be possible using existing shape-from-shading techniques. The key was to construct a statistical model for fields of surface normals and show how to incorporate this into the geometric shape-from-shading framework. The power of the technique can be attributed to the combination of a strict global constraint (projection onto the statistical model) with a hard local constraint (satisfaction of the image irradiance equation).

However, the approach suffers from a serious drawback. During fitting, the effects of albedo are ignored (the image irradiance equation used assumes unit albedo). Similarly, there is no model of cast shadow formation (regions in which the light source is intercepted by another part of the surface). The effects of both of these phenomena are

effectively treated as noise and not explicitly accounted for. The result is that the fit of the statistical model is subject to a systematic error and becomes increasingly inaccurate when regions of low albedo dominate (for instance in the presence of facial hair) or when cast shadows become significant (as the light source direction is more extreme).

Cast Shadows

We have already shown how regions of low albedo cause large differences between the surface normals predicted by the statistical model and those suggested by the image irradiance equation (see Figure 3.5). The presence of cast shadows results in a similar problem.

When a surface is illuminated by a single point light source, a point on the surface is *in shadow* when it is not visible from the light source. In other words, no light reaches the point and the measured intensity at the pixel which corresponds to the point is zero. We restrict our consideration to single point light sources, since extended or multiple light sources produce more complicated effects.

There are two scenarios which result in a point being in shadow and these are modeled in quite different ways. An *attached shadow* (also called a self-shadow) occurs when a point on the surface is oriented away from the light source, thereby occluding itself from the illumination. Shadows of this sort are easily modeled, since they depend only upon the local geometry of the surface (the normal direction). On the other hand, *cast shadows* are caused when an entirely different region of the surface intersects the path from the light source to the point in question. Cast shadows are more complex to model since they are dependent on the global geometry of the surface. In Figure 4.1 we show an artificial example to illustrate the difference between the two types of shadow. We render a surface of Gaussian peaks with a single light source from behind the peaks to the top right. The non-shadow regions have been shaded red with Lambertian reflectance. The regions of

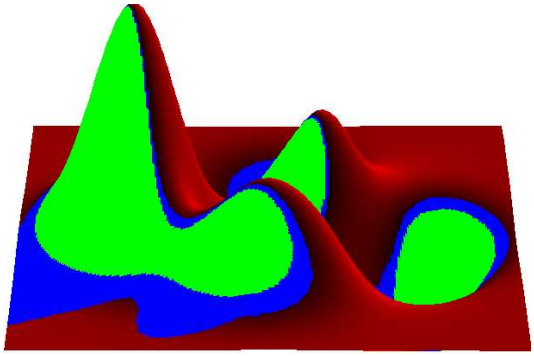


Figure 4.1: Example of cast and attached shadows.

attached shadow are shown in green, while the regions shown in blue are cast shadows caused by the peaks intercepting the light source.

For a given illumination condition, the regions of a surface (whose height function is known) that lie in cast shadow may be calculated using ray-tracing. In the context of shape-from-shading the situation is more complicated. Inter-reflections and the presence of ambient illumination mean that shadow pixels in real world images will not necessarily be measured as having zero intensity. To recover the surface height function (which is necessary to identify cast shadow regions) the surface orientation at every point is required. However, shadow regions convey no information about facial shape or texture and so we cannot make an estimate of the surface orientation at these points. The result is a 'chicken or egg' situation: accurate knowledge of global shape is required to estimate cast shadow regions, but cast shadow regions disrupt the accurate estimation of the global shape from image intensity.

Contribution

In this chapter, we consider ways in which to extend the work of Chapter 3 in order to tackle the problems discussed above. To do so, we provide an alternative formulation for the model based on techniques from differential geometry. We show how to incorporate

robust statistics into the process of fitting this model to an image and how the method can be used to mitigate the effects of varying albedo and cast shadows. Moreover, this robust fitting process provides a means to perform surface completion for those areas in which we cannot recover reliable shape information. Finally, we investigate how the information we can recover from a single image using this technique can be used to perform face recognition. We summarise the contributions of this chapter in the areas of shape representation, facial shape recovery and face recognition below.

Shape Representation: Our representation of the distribution of normal directions is based on the exponential map from a unit sphere onto a Cartesian plane. To capture the statistics of the distribution of the surface normals under the exponential map, we borrow ideas from principal geodesic analysis. This is a technique first proposed by Fletcher, Joshi, Lu and Pizer [56] for the analysis of tensor MRI data, and recently extended to the general problem of statistical analysis on Riemannian manifolds by Pennec [122]. We adopt this approach since the use of the exponential map to represent the distribution of surface normal directions allows us formulate our facial shape-from-shading method in a compact geometric way.

Shape Recovery: With the model to hand, we demonstrate how it may be fitted to image brightness data, so as to satisfy constraints provided by Lambert's law. The resulting iterative update rule for the fitted surface normal is particularly simple and can be couched in terms of rotation operations on the exponential map. To deal with shadowing and variations in albedo, we extend the fitting method to deal with outliers using the apparatus of robust statistics [86]. We use the residuals of model fit to compute a weight for each pixel. The weights are used to exclude shadow regions and regions of low albedo in the shape-parameter estimation process. The parameter update scheme is based on M-estimators [86].

The resulting algorithm has a number of attractive features. In the previous chapter,

we showed how to construct a statistical model for surface normal direction by applying principal components analysis to the distribution of points obtained by transforming the surface normals from a sphere to a tangent-plane using the azimuthal equidistant projection. By using the exponential and logarithmic maps we are able to capture the statistical variations in the distribution of surface normals using principal geodesic analysis. This leads to a way of updating the surface normal directions which is both more efficient and more elegant. In particular it establishes a means of incorporating the differential geometry of geodesics into shape-from-shading. Although this topic is addressed in the work of Oliensis and Dupuis [49] and Fua and Leclerc [60] here we take a step further by demonstrating links to the exponential map. Moreover, by using robust statistics to fit the model to intensity data, we have a method that can deal with both shadowing and facial albedo variations.

We experiment with the resulting shape-recovery method on both synthetic images with known ground truth and real-world images from the Yale-B database. The results indicate that the method works well even when the angle between the light source and the image-normal exceeds 60 degrees. Moreover, the fitting method is able to reliably fill-in shadowed regions of the face.

Face Recognition: The final aim of this chapter is to investigate the use of an intermediate representation, based on *2.5D shape* information. This is a somewhat neglected source of information with which to perform face recognition. By 2.5D shape information, we refer to Marr's [108] concept of the $2\frac{1}{2}$D sketch. The $2\frac{1}{2}$D sketch provides a viewer-centered representation of an observed surface, typically comprising a field of local surface orientation estimates in a retinocentric coordinate frame. There is strong psychophysical evidence that such a representation plays a part in human face processing and that this information is recovered, at least in part, using shape-from-shading [79]. This representation provides a natural extension to view-based approaches which use in-

tensity, whilst incorporating some of the advantages of 3D shape, such as invariance to changes in illumination. The hope is that we will combine the benefits of full 3D shape with the generative power of appearance-based approaches. We show how our shape-from-shading algorithm allows us to recover accurate 2.5D facial shape information and albedo from a single image.

We explore the various ways in which the shape parameters extracted from a single image using the shape-from-shading algorithm can be used for face recognition. Our first recognition approach explicitly compares 2.5D shape information recovered from a probe image with that previously recovered from a gallery image. We do so by measuring the distance between the two sets of data in the space of the principal geodesic parameters. We contrast this approach with using the shape and reflectance information recovered from a probe image to generate prototype illumination-normalised images which can be compared to similarly illuminated gallery images. Here we show that the 2.5D shape information is sufficient to correct for variations in illumination. The most similar work in spirit to this is that of Zhao and Chellappa [166], who used a symmetric shape-from-shading algorithm to generate prototype images. However, their approach required a manual segmentation of the image into regions of piecewise constant albedo.

Finally we show that, from a single input image, we are able to generate the basis images employed by a number of illumination-insensitive recognition algorithms (Illumination Cone [64], Harmonic Images [8] and Nine Points of Light [100]). At the expense of a slight increase in the overall error rate, we remove the need for multiple training images. Furthermore, recent work [164] has shown that the space of harmonic images provides a powerful representation that can capture any face under arbitrary illumination. In this chapter we make a contribution to this work. Specifically, we show how the harmonic images may be expressed directly in terms of the parameters of our surface normal model. We demonstrate that this provides a more efficient and principled means with

which to construct a low-dimensional representation of harmonic images in comparison to applying a linear analysis directly to the harmonic images.

4.1 A Statistical Surface Normal Model using PGA

Most statistical shape models rely on Principal Components Analysis (PCA) to construct an efficient parameterisation of shape variation [5,14,36]. However, as discussed in Chapter 3, a linear combination of unit vectors is not a unit vector. Hence, the analysis of the distribution of surface normals cannot be effected in a linear way. In Chapter 3 we overcame this problem by using the azimuthal equidistant projection to transform surface normals to points on a tanget plane on which a linear PCA could be applied. We now revisit this problem and provide a more formal treatment of the model construction process, turning to the tools of differential geometry. This provides an alternative definition for our notion of the 'average' facial needle-map and a significantly more elegant formulation with which to work. We are also able to express the cone constraint in a similar manner resulting in a particularly simple expression of the shape-from-shading algorithm.

We choose to model a distribution of surface normals as a distribution of points lying on a spherical manifold and exploit the Principal Geodesic Analysis (PGA) approach of Fletcher et al. [56] in order to build our statistical model. In this section we describe PGA, demonstrate how it may be applied to a set of needle-maps and then show how we use the principal geodesics to deform the field of surface normals.

4.1.1 Preliminaries

A unit vector $\mathbf{n} \in \mathbb{R}^3$ may be considered as a point lying on a spherical manifold $n \in S^2$, where S^2 is the unit 2-sphere. The two are related by $\mathbf{n} = \Phi(n)$ where $\Phi : S^2 \mapsto \mathbb{R}^3$ is an embedding. Likewise, a field of surface normals $\mathbf{U} \in \mathbb{R}^{N\times 3}$ describing a surface

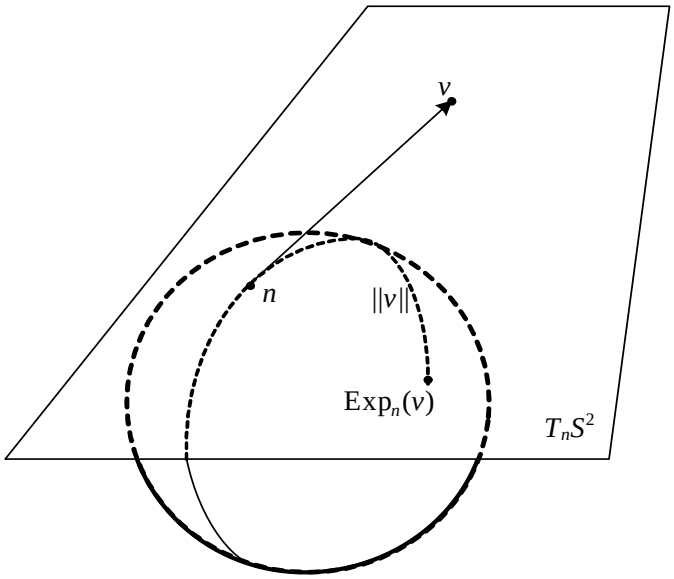


Figure 4.2: The exponential map.

may be considered as a point on a manifold $U \in S^2(N) = \prod_{i=1}^{N} S^2$. Shape variations in Euclidian space are typically analysed using standard linear techniques, i.e. linear averaging and PCA. For data lying on a non-linear manifold, we turn to the concept of the intrinsic mean and PGA [56].

In this section, we describe the application of PGA to needle-map data and begin by introducing some prerequisite theory.

The Log and Exponential Maps

If $v \in T_nS^2$ is a vector on the tangent plane to S^2 at $n \in S^2$ and $v \neq 0$, the *exponential map*, denoted Exp_n, of v is the point on S^2 along the geodesic in the direction of v at distance $\|v\|$ from n. Geometrically, this is equivalent to marking out a length equal to $\|v\|$ along the geodesic that passes through n in the direction of v. The point on S^2 thus obtained is denoted $\text{Exp}_n(v)$. This is illustrated in Figure 4.2. The inverse of the exponential map is the *log map*, denoted Log_n. Therefore, the equality $\text{Log}_n(\text{Exp}_n(v)) = v$ holds. Note that the log map is analogous to the azimuthal equidistant projection de-

scribed in Chapter 3. Likewise, the exponential map is analogous to the inverse azimuthal equidistant projection.

The geodesic distance between two points $n_1 \in S^2$ and $n_2 \in S^2$ can be expressed in terms of the log map, i.e. $d(n_1, n_2) = \|\text{Log}_{n_1}(n_2)\|$. The exponential and log maps for the space of a field of N surface normals, $S^2(N)$, are simply the direct products of N copies of the maps for S^2 given above.

Spherical Medians and Variance

In Chapter 3, we characterised a distribution of spherical directional data $\mathbf{n}_1, \ldots, \mathbf{n}_K \in \mathbb{R}^3$ using the *mean direction* [107] $\hat{\mathbf{n}}_0 = \frac{\bar{\mathbf{n}}}{\|\bar{\mathbf{n}}\|}$ where $\bar{\mathbf{n}} = \frac{1}{K}\sum_{i=1}^{K} \mathbf{n}_i$.

If we consider the distribution of unit vectors as a distribution of points on a spherical manifold $n_1, \ldots, n_K \in S^2$, where $\Phi(n_k) = \mathbf{n}_k$, it is clear that the mean direction is dependent on the embedding Φ and is the *extrinsic mean* of a distribution of spherical data:

$$\mu_\Phi = \underset{n \in S^2}{\arg\min} \sum_{i=1}^{K} \|\Phi(n) - \Phi(n_i)\|^2. \tag{4.1}$$

If we define the projection $\pi : \mathbb{R}^3 \mapsto S^2$ as

$$\pi(\mathbf{n}) = \underset{n \in S^2}{\arg\min} \|\Phi(n) - \mathbf{n}\|^2, \tag{4.2}$$

we may show that the mean direction is the extrinsic mean:

$$\mu_\Phi = \pi(\bar{\mathbf{n}}) = \pi\left(\frac{1}{K}\sum_{i=1}^{K} \Phi(n_i)\right). \tag{4.3}$$

In other words, the extrinsic mean is the Euclidian average (or centre of mass) of the distribution of points in $\mathbb{R}^3$, projected back onto the closest point on the sphere.

A more natural definition of the average of a distribution of points on the unit sphere

uses arc length as the choice of distance measure. Since a 2-sphere is a Riemannian manifold and great circles are geodesics, this is the Riemannian distance $d(.,.)$ between a pair of points, $d(n_1, n_2) = \arccos(\Phi(n_1) \cdot \Phi(n_2))$. Using this definition of distance, we can define the *intrinsic mean*:

$$\mu = \arg\min_{n \in S^2} \sum_{i=1}^{K} d(n, n_i). \tag{4.4}$$

For spherical data, this is known as the *spherical median* [52]. This point cannot be found analytically, but can be solved iteratively using the gradient descent method of Pennec [121]. We initialise our estimate as the Euclidian mean of distribution, i.e. $\mu_{(0)} = \mu_{\Phi}$. The current estimate is then updated iteratively as follows:

$$\mu_{(j+1)} = \mathrm{Exp}_{\mu_{(j)}} \left(\frac{1}{K} \sum_{i=1}^{K} \mathrm{Log}_{\mu_{(j)}}(n_i) \right). \tag{4.5}$$

To find the intrinsic mean $\mu \in S^2(N)$ of a sample of K fields of N surface normals: $U_1, \ldots, U_K \in S^2(N)$, we replace the exponential and log maps in Equation 4.5 with the corresponding maps for the space $S^2(N)$.

We can use the log map and intrinsic mean to define the sample variance of a distribution of points on the sphere:

$$\sigma^2 = \frac{1}{K} \sum_{i=1}^{K} d(\mu, n_i)^2 = \frac{1}{K} \sum_{i=1}^{K} \|\mathrm{Log}_{\mu}(n_i)\|^2. \tag{4.6}$$

Projection onto Principal Geodesics:

In standard PCA, the lower-dimensional subspaces form a linear subspace of the space in which the data lies. In PGA, this notion is replaced by a *geodesic submanifold*. In other words, whereas each principal axis in PCA is a straight line, in PGA each principal axis is a geodesic curve. In the spherical case this corresponds to a great circle.

In PCA a datum may be projected onto a subspace using linear projection. In the case of PGA, a different kind of projection operator is required. Consider a geodesic submanifold H of the Riemannian manifold M. In order to project a point $x \in M$ onto the point on H that is closest to x in terms of Riemannian distance, we use the projection operator $\pi_H : M \to H$ given by $\pi_H(x) = \arg\min_{y \in H} d(x,y)^2$. If $H \subset M$ is a geodesic submanifold at a point μ, we may approximate π_H linearly in the tangent space of M at μ, $T_\mu M$. If $v_1, ..., v_K$ is an orthonormal basis for $T_\mu H$, then the projection operator π_H can be approximated in the tangent plane using:

$$\text{Log}_\mu\left(\pi_H(x)\right) \approx \sum_{i=1}^{K} v_i \left(v_i \cdot \text{Log}_\mu(x)\right) \tag{4.7}$$

Computing Principal Geodesics:

With the preliminaries above to hand, the principal geodesics of $U_1, \ldots, U_K \in S^2(N)$ may be computed. By analogy with PCA, the goal of PGA is to find a sequence of nested geodesic submanifolds which account for decreasing amounts of variance in the data. The principal geodesics are defined by first constructing an orthonormal basis of tangent vectors $v_1, \ldots, v_K \in T_\mu S^2(N)$. These are the principal directions on the tangent plane $T_\mu S^2(N)$.

Using the approximation (4.7), the first principal direction is given by:

$$v_1 \approx \arg\max_{\|v\|=1} \sum_{i=1}^{N} \left(v \cdot \text{Log}_\mu(U_i)\right)^2 \tag{4.8}$$

the remaining principal directions are defined recursively as:

$$v_k \approx \arg\max_{\|v\|=1} \sum_{i=1}^{N} \sum_{j=1}^{k-1} \left(1 - \|v_j \cdot v\|\right)^2 + \left(v \cdot \text{Log}_\mu(U_i)\right)^2 \tag{4.9}$$

This minimisation problem is equivalent to the standard principal components analysis of

the vectors $\text{Log}_{\mu}(U_i)$ in $T_{\mu}S^2(N)$. Hence, the crucial observation made by Fletcher et al. [56] is that the principal geodesics can be approximated by applying standard linear PCA in to the vectors $\text{Log}_{\mu}(U_i)$. In other words, we transform the distribution of unit vectors to points on the tangent plane to $S^2(N)$ at μ and use standard PCA. In the following section we show how to apply principal geodesic analysis to a set of exemplar facial needle-maps for the purposes of learning a statistical model of face-shape.

4.1.2 PGA of Needle-Maps

Suppose that each of the K training examples is a range image which consists of an array of depth data each containing $N = X_{\text{res}} \times Y_{\text{res}}$ pixels. In this chapter we find it convenient to use an alternative notation to index pixel position in the image plane. We refer to a pixel at position (x, y) using a single index $p = (X_{\text{res}} - 1)x + y$, e.g. the surface normal at pixel p is $\mathbf{n}_p$. For the pixel indexed p in the kth training sample the depth is z_p^k. Using the range data we estimate the surface normal directions, and the surface normal at the pixel location p for the kth training image is $\mathbf{n}_p^k = \left((\mathbf{n}_p^k)_x, (\mathbf{n}_p^k)_y, (\mathbf{n}_p^k)_z\right)^T$, where $(\dots)_x$ denotes the x component of the parenthesised vector, and the y and z components are defined similarly.

We calculate the spherical median μ_p of the distribution of surface normals $\mathbf{n}_p^1, \dots, \mathbf{n}_p^K$ at each pixel location p using Equation 4.5. The surface normal $\mathbf{n}_p^k$ is represented by its position on the tangent plane $T_{\mu_p}S^2$ given by the log map: $v_p^k = \text{Log}_{\mu_p}(\mathbf{n}_p^k) \in \mathbb{R}^2$.

The vectors are used as the columns of the data matrix $\mathbf{D} = [\mathbf{v}^1 | \dots | \mathbf{v}^K] \in \mathbb{R}^{2N \times K}$ where $\mathbf{v}^k = [v_1^k, \dots, v_N^k]^T$. From the data-matrix, we compute the covariance matrix

$$\mathbf{L} = \frac{1}{K}\mathbf{D}\mathbf{D}^T. \tag{4.10}$$

We find the eigenvectors and eigenvalues of the covariance matrix $\mathbf{L}$ and use them

to perform PGA. The ith principal geodesic is given by the eigenvector, Ψ_i, with the ith largest eigenvalue, λ_i. To retain c percent of the total model variance we need only retain E eigenmodes, where E is the smallest integer that satisfies: $\sum_{i=1}^{E} \lambda_i \geq \frac{c}{100} \sum_{i=1}^{K} \lambda_i$.

From the E leading eigenvectors $\mathbf{P} = (\Psi_1|\Psi_2|\ldots|\Psi_E) \in \mathbb{R}^{2N \times E}$, we can generate a new facial needle-map, $\mathbf{U} \in \mathbb{R}^{N \times 3}$, from the parameter vector $\mathbf{b} = [b_1, b_2, \ldots, b_E]^T \in \mathbb{R}^E$:

$$\mathbf{U} = \text{Exp}_{\mu}(\mathbf{Pb}). \tag{4.11}$$

So the estimated field of surface normals, $\mathbf{U}$, that lies within the span of the model is found by projecting the linear model from the tangent plane at μ onto the spherical manifold. In fact, the result is quite general and can be used to construct a statistical model over more complex manifold surfaces.

4.2 Incorporating Principal Geodesics into SFS

We now show how to incorporate the statistical constraint provided by the principal geodesics into the geometric shape-from-shading algorithm. To do so, we show how the cone constraint can be expressed in terms of linear operations on the tangent plane. The resulting algorithm is particularly elegant.

Let the long vector $I \in \mathbb{R}^N$ denote the N-pixel input image, formed by stacking the rows of the image into a column vector. We can now state our shape-from-shading formulation in terms of the principal geodesics as recovering the parameter vector whose corresponding field of surface normals minimises the irradiance error:

$$\mathbf{b}^* = \arg\min_{\mathbf{b}} \|I - \text{Exp}_{\mu}(\mathbf{Pb})\mathbf{s}\|^2. \tag{4.12}$$

We assume that the light source direction, $\mathbf{s}$, is known and solve this minimisation prob-

lem using the geometric shape-from-shading framework described above. We begin by recasting the framework in terms of operations on the tangent plane and show how the whole algorithm can be expressed algebraically in terms of matrix operations.

At iteration (t) the parameter vector estimate:

$$\mathbf{b}^{(t)} = \mathbf{P}^T \mathrm{Log}_{\mu}(\mathbf{U}^{(t)}), \tag{4.13}$$

represents the best fit in the model space to the current estimate of the field of surface normals $\mathbf{U}^{(t)}$. Under the log map, the off-cone surface normal at pixel p is given by $\tilde{u}_p^{(t)} = (\mathbf{P}\mathbf{b}^{(t)})_p$.

The unit vector $\mathbf{s}$ describing the direction of the point light source may be considered as a point on the unit sphere $s \in S^2$, such that $\Phi(s) = \mathbf{s}$. In order to satisfy the image irradiance equation (Equation 3.7), the off-cone normal $\tilde{n}_p^{(t)} = \mathrm{Exp}_{\mu_p}(\tilde{u}_p)$ must be transformed such that the geodesic distance on the sphere between the light source s and the normal at iteration $(t+1)$ equals $\arccos(I_p)$, i.e. $\|\mathrm{Log}_s(n_p^{(t+1)})\| = \arccos(I_p)$. The closest normal to $\tilde{n}_p^{(t)}$ which satisfies this condition is given by:

$$n_p^{(t+1)} = \mathrm{Exp}_s \left(\arccos(I_p) \frac{\mathrm{Log}_s(\tilde{n}_p^{(t)})}{\|\mathrm{Log}_s(\tilde{n}_p^{(t)})\|} \right). \tag{4.14}$$

This is equivalent to moving along a geodesic from s in the direction of $\tilde{n}_p^{(t)}$ a distance $\arccos(I_p)$. The result of $\frac{\mathrm{Log}_s(\tilde{n}_p^{(t)})}{\|\mathrm{Log}_s(\tilde{n}_p^{(t)})\|}$ is a unit vector in T_sS^2 in the direction of $\tilde{u}_p^{(t)}$. This is scaled to be of length $\arccos(I_p)$ and transformed back into a point on the sphere by taking the exponential map at s.

The geometry of this process is highlighted in Figure 4.3 which shows the tangent plane T_sS^2. The circle represents the log map of the cone on which the normal must lie to satisfy the image irradiance equation. In other words, points on the circle correspond to surface normals that satisfy $I_p = \mathbf{n}_p \cdot \mathbf{s}$. Hence the intersection of the circle and vector

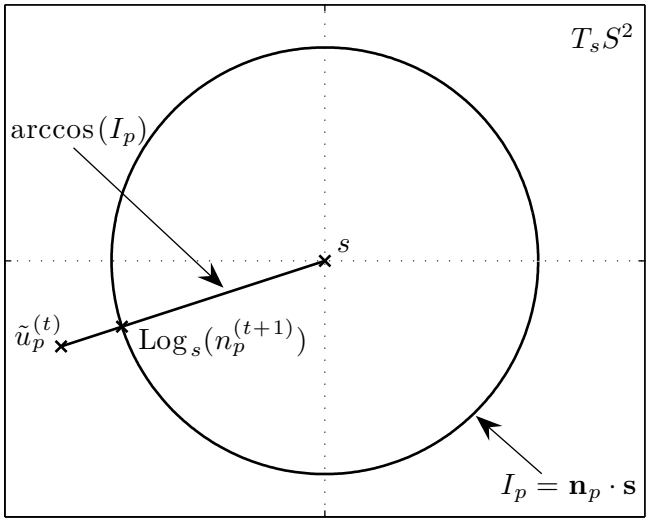


Figure 4.3: Restoring a normal to the closest position on the cone in terms of operations on the tangent plane.

$\tilde{u}_p$ represents the closest normal to $\tilde{u}_p$ which strictly satisfies data-closeness.

4.2.1 Initialisation

We may exploit the statistical model to provide an improvement over the negative gradient initialisation used in Chapter 3. We set the initial off-cone estimate to be the local intrinsic mean surface normal. After enforcing data-closeness, this initialisation is equivalent to selecting the position on the cone closest to the local intrinsic mean direction. This initialisation alone is sufficient to recover coarse, but useful, estimates of facial shape.

4.2.2 Algorithm

The algorithm is summarised in Algorithm 4.1. It should be noted that this is in essence a re-expression of Algorithm 3.2 using the tools of differential geometry. As suggested in the introduction, fitting the model using this model results in a systematic error and becomes increasingly inaccurate when regions of low albedo dominate (for instance in

Algorithm 4.1: Geometric shape-from-shading using principal geodesics as a statistical constraint.

Input: Light source $\mathbf{s}$, image intensities I_p, PGA model $\mathbf{P}$ and intrinsic mean normals μ_p

Output: Estimated field of on-cone surface normals $\mathbf{n}_p$ and best-fit model parameter vector $\mathbf{b}$

1 Set iteration $t = 0$;

2 Initialise parameter vector $\mathbf{b}^{(0)} = (0, 0, \ldots, 0)^T$;

3 **repeat**

4 Enforce image irradiance constraint:

$$n_p^{(t)} = \mathrm{Exp}_s \left(\arccos(I_p) \frac{\mathrm{Log}_s(\mathrm{Exp}_{\mu_p}((\mathbf{P}\mathbf{b}^{(t)})_p))}{\|\mathrm{Log}_s(\mathrm{Exp}_{\mu_p}((\mathbf{P}\mathbf{b}^{(t)})_p))\|} \right);$$

5 Estimate parameter vector: $\mathbf{b}^{(t+1)} = \mathbf{P}^T \mathrm{Log}_\mu(\mathbf{U}^{(t)})$, where $\mathbf{U}^{(t)} = \left[n_1^{(t)} | \ldots | n_N^{(t)} \right]$;

6 Set iteration $t = t + 1$;

7 **until** $\sum_{i=1}^{N} \left\| d\left(n_i^{(t-1)}, n_i^{(t-2)} \right) \right\|^2 < \epsilon$;

the presence of facial hair) or when cast shadows become significant (as the light source direction is more extreme). In this case, a significant portion of the face may be in shadow and fitting the statistical model globally results in erroneous shape parameter estimates. For this reason, we now turn show how to incorporate robust statistics into the process to mitigate these problems.

4.3 Robust Statistics

It is clear that we require a more robust means to fit our statistical model to a potentially noisy observed field of normals, $\mathbf{n}_p$. According to the PGA framework, the quality of this fit can be measured by calculating the distance between the observed and fitted normals on the tangent plane, i.e. the geodesic distance. If $\mathbf{b}$ is the estimated parameter vector, the residual at point p is given by $\eta_p = \|\mathrm{Log}_{\mu_p}(n_p) - (\mathbf{P}\mathbf{b})_p\|$. The standard least squares fit

given above, minimises the quantity:

$$\mathbf{b}^* = \arg\min_{\mathbf{b}} \sum_{i=1}^{N} \eta_i^2. \tag{4.15}$$

This approach is unstable in the presence of outlying data, such as normals erroneously estimated from regions of low albedo or in cast shadow. In particular, the effect of outliers is to severely distort the estimated facial shape.

In this chapter, we turn to the apparatus of robust statistics to help overcome this problem. M-estimators [86] (maximum likelihood type estimators) aim to reduce the effect of outliers by replacing the squared residuals η_i^2 by a kernel function that limits the effects of large residuals:

$$\mathbf{b}^* = \arg\min_{\mathbf{b}} \sum_{i=1}^{N} \varrho_\sigma(\eta_i), \tag{4.16}$$

where ϱ is a robust kernel with width parameter σ.

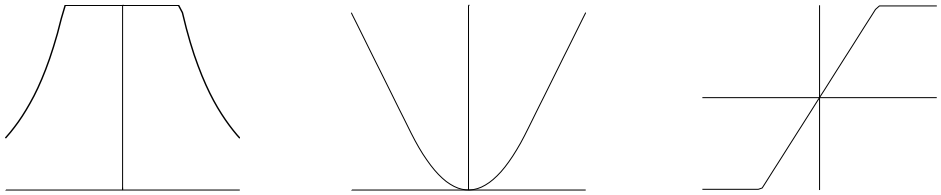
The influence of a residual on the parameter estimate under a given M-estimator can be studied by examining its *influence function*, ψ_σ. This is the derivative of the error kernel:

$$\psi_\sigma(\eta_p) = \frac{\partial \varrho_\sigma(\eta_p)}{\partial \eta_p}. \tag{4.17}$$

Hence, in the least squares case where $\varrho_\sigma(\eta) = \eta^2$, the influence of a datum is $\psi_\sigma(\eta) = 2\eta$ and therefore increases linearly with the size of the error. This is the source of the lack of robustness in least-squares estimation.

4.3.1 Huber's M-estimator

We propose a robust solution to Equation 4.16 using a simple one-step weighted least squares approximation. To do so, we make use of the *weight function*, w_σ, which is



$$w_\sigma(\eta) = \begin{cases} 1 \\ \frac{\sigma}{|\eta|} \end{cases} \qquad \varrho_\sigma(\eta) = \begin{cases} \eta^2 \\ 2\sigma|\eta| - \sigma^2 \end{cases} \qquad \psi_\sigma(\eta) = \begin{cases} 2\eta \\ 2\sigma \text{Sign}(\eta) \end{cases}$$

Figure 4.4: Huber's M-estimator.

related to the influence function by:

$$w_\sigma(\eta_p) = \frac{\psi_\sigma(\eta_p)}{\eta_p}. \tag{4.18}$$

The standard least-squares estimator applies a constant weight to each datum. On the other hand, an error kernel such as Huber's estimator [86] down-weights a datum once its residual exceeds σ:

$$\varrho_\sigma(\eta) = \begin{cases} \eta^2 & \text{if } |\eta| < \sigma \\ 2\sigma|\eta| - \sigma^2 & \text{otherwise} \end{cases} \qquad w_\sigma(\eta) = \begin{cases} 1 & \text{if } |\eta| < \sigma \\ \frac{\sigma}{|\eta|} & \text{otherwise} \end{cases} \tag{4.19}$$

We show the weight function, error kernel and influence function for Huber's M-estimator in Figure 4.4. This is the M-estimator we use in the remainder of this chapter.

4.3.2 Weighted Best-fit

We can incorporate the Huber weights into the least squares expression in Equation 4.13 by constructing a diagonal matrix of weights $\mathbf{W} = \text{diag}\,(w_\sigma(\eta_1), \ldots, w_\sigma(\eta_N))$. Our one-

step weighted least squares approximation of $\mathbf{b}$ is given by:

$$\mathbf{b}^{(t)} = C\mathbf{P}^T\mathbf{W}^{(t)}\mathrm{Log}_{\mu}(\mathbf{U}^{(t)}), \tag{4.20}$$

where C is a constant which compensates for the overall scaling effect of $\mathbf{W}$ on $\mathbf{b}$. If C is set to the sum of the reciprocals of the weights, this amounts to a one-step weighted least squares fit. However, this approach becomes unstable when a large number of normals have been assigned low weights. The result is over-fitting to sparse and potentially noisy data. We overcome this problem by introducing a control parameter, ς, which represents the trade-off between goodness of fit and distance from the mean. Accordingly, we set C as follows:

$$C = \varsigma \sum_{p=1}^{N} \frac{1}{w_{\sigma}(\eta_p)}, \tag{4.21}$$

where ς is allowed to lie in the interval $[0, 1]$. If $\varsigma = 0$, the result at every iteration is the mean field of normals since the parameter vector $\mathbf{b}^{(t)}$ will be zero. If $\varsigma = 1$, a one-step weighted least squares fit is performed. For robust performance on real world data, a value somewhere in between is preferable. The weights are recalculated at each iteration using residuals calculated from the previous iteration.

Computing Equation 4.20 requires an initial estimate of $\mathbf{W}$ and hence the residuals. To do this, we initialise the parameter vector to $\mathbf{b}^{(0)} = 0$ (as in Section 4.2.1) and calculate the residuals when data-closeness is enforced on the resulting field of surface normals. In effect, we are initialising the weights based on measuring departure from the intrinsic mean direction at each pixel. We find this provides a reliable initialisation. For each subsequent iteration t of the algorithm, we can use the weights calculated from the residuals at iteration $(t - 1)$.

4.3.3 Median Absolute Deviation

Implicit in the discussion above is that we have a means to estimate the standard deviation of the residual errors σ, which acts as the width parameter of the error kernel ϱ. A robust estimate of σ is required in order to distinguish outliers from inliers. For this reason, we use the *median absolute deviation* (MAD) estimator:

$$\text{MAD} = \text{median}\left(|\eta_p - \text{median}(\eta_p)|\right), p = 1 \ldots N, \tag{4.22}$$

which is related to the standard deviation by $\sigma = 1.4826 \times \text{MAD}$. We recalculate σ at each iteration.

4.3.4 Algorithm

The robust version of our algorithm, in which a weighted fit of the surface normal model is used to update the estimated normal directions is given in Algorithm 4.2.

It is clear that our algorithm shares much in common with the iteratively reweighted least-squares algorithm [86], in that the weights are calculated using the residuals from the previous iteration. The difference is that in step 4 we adjust the data to which the model is being fitted. The result of this is that during earlier iterations when the majority of residuals are large and the weight function serves only to coarsely mask unreliable areas, convergence is driven primarily by the hard irradiance constraint. During later iterations as reliable pixels contain accurate estimates of the normal direction, the adjustment made in step 4 is minimal and therefore the residuals are small. Hence, the effects of the weighting becomes more subtle and the parameter estimate converges towards a robust solution. The algorithm retains its efficiency, since each operation may still be implemented as a matrix multiplication.

Algorithm 4.2: Geometric shape-from-shading using a robust fit of the principal geodesics as a statistical constraint.

Input: Light source $\mathbf{s}$, image intensities I_p, PGA model $\mathbf{P}$ and intrinsic mean normals μ_p

Output: Estimated field of on-cone surface normals $\mathbf{n}_p$, best-fit model parameter vector $\mathbf{b}$ and robust fit weights $w_\sigma(\eta_p)$

Set iteration $t = 0$;

Initialise parameter vector $\mathbf{b}^{(0)} = (0, 0, \ldots, 0)^T$;

repeat

  Enforce image irradiance constraint:

$$n_p^{(t)} = \text{Exp}_s \left(\arccos(I_p) \frac{\text{Log}_s(\text{Exp}_{\mu_p}((\mathbf{P}\mathbf{b}^{(t)})_p))}{\|\text{Log}_s(\text{Exp}_{\mu_p}((\mathbf{P}\mathbf{b}^{(t)})_p))\|} \right);$$

  Calculate residuals: $\eta_p^{(t)} = \|\text{Log}_{\mu_p}(n_p^{(t)}) - (\mathbf{P}\mathbf{b}^{(t)})_p\|$;

  Calculate standard deviation of residuals:
  $\sigma^{(t)} = 1.4826 \times \text{median}\left(|\eta_p^{(t)} - \text{median}(\eta_p^{(t)})|\right)$;

  Calculate weight matrix: $\mathbf{W}^{(t)} = \text{diag}\left(w_{\sigma^{(t)}}(\eta_1), \ldots, w_{\sigma^{(t)}}(\eta_N)\right)$;

  Calculate robust estimate of parameter vector:
  $\mathbf{b}^{(t+1)} = C\mathbf{P}^T\mathbf{W}^{(t)}\text{Log}_\mu(\mathbf{U}^{(t)})$, where $\mathbf{U}^{(t)} = \left[n_1^{(t)}|\ldots|n_N^{(t)}\right]$;

  Set iteration $t = t + 1$;

until $\sum_{i=1}^{N} \left\| d\left(n_i^{(t-1)}, n_i^{(t-2)}\right) \right\|^2 < \epsilon$;

4.3.5 Combining Evidence and Classifying Shadow Pixels

If upon convergence a pixel p has weight $w_p^{\text{final}} \approx 1$, this indicates a high confidence that the normal $\mathbf{n}_p^{\text{final}}$ is reliable. However, as the weight tends to 0, the normal $\mathbf{n}_p^{\text{final}}$ is likely to be erroneous due to violation of the assumptions of Lambert's law, e.g. non-constant albedo or lying in a cast shadow region. In this case, a more accurate estimate is given by the robust fit of the model to the global field of normals, e.g. $\text{Exp}_{\mu_p}\left((\mathbf{P}\mathbf{b}^{\text{final}})_p\right)$. With reference to Figure 4.3, the normal $\mathbf{n}_p^{\text{final}}$ lies on the circle, while the normal $\text{Exp}_{\mu_p}\left((\mathbf{P}\mathbf{b}^{\text{final}})_p\right)$ lies within the span of the model.

For this reason, our best estimate of the underlying shape of the face is a weighted combination, in which pixels with a low weight are given a higher proportion of the nor-

mal $\text{Exp}_{\mu_p}\left((\mathbf{Pb}^{\text{final}})_p\right)$ and a lower proportion of the normal $\mathbf{n}_p^{\text{final}}$, vice-versa for pixels with a high weight. This gives a weighted average for the estimated surface normal direction which may be expressed in terms of tangent plane operations as:

$$\mathbf{n}_p^{\text{est}} = \text{Exp}_{n_p^{\text{final}}}\left[(1 - w_p^{\text{final}})\text{Log}_{n_p^{\text{final}}}(\text{Exp}_{\mu_p}\left((\mathbf{Pb}^{\text{final}})_p\right)\right]. \tag{4.23}$$

With the estimated facial shape to hand, we may now go further and distinguish between pixels of low albedo and those in cast shadow regions. We may recover the surface height z_p by integrating [59] the field of normals $\mathbf{n}^{\text{est}}$. Using a simple ray-tracing algorithm, we can assign a binary cast-shadow map:

$$\text{shadow}(z, \mathbf{s}, p) = \begin{cases} 0 & \text{if pixel } p \text{ is in cast shadow under illumination } \mathbf{s} \\ 1 & \text{otherwise} \end{cases} \tag{4.24}$$

For non-shadow regions, the albedo $\rho_{d,p}$ can be estimated by rearranging the image irradiance equation:

$$\rho_{d,p} = \frac{I_p}{\mathbf{n}_p^{\text{est}}.\mathbf{s}} \tag{4.25}$$

For pixels lying in cast shadow, we estimate the albedo using facial symmetry.

4.4 Recognition using 2.5D Shape Information

The principal geodesic shape-from-shading approach described above allows us to recover accurate 2.5D facial shape information along with albedo estimates from a single image. Moreover, the parameters of the principal geodesics provide a low-dimensional characterisation of the facial shape. There are a number of ways we can use these intrinsic features of the face to perform recognition. In this section we first present two ways in which the recovered shape information can be used directly to perform recognition. Sec-

ondly, we show how the information recovered from a single image can be used to derive linear subspaces which are suitable for recognition.

4.4.1 Direct Recognition

For a given probe image, assuming the illumination is known, we can recover the parameter vector $\mathbf{b}$ which characterises the field of 2.5D shape information. The simplest recognition strategy is to measure the Euclidian distance between $\mathbf{b}_{\text{probe}}$ and the parameter recovered from each gallery image, $\|\mathbf{b}_{\text{probe}} - \mathbf{b}_{\text{gallery}}\|$. This is analogous to Eigenfaces [149], but in the domain of 2.5D shape as opposed to intensity.

An alternative is to use the shape and albedo information recovered from a probe image to synthesise an illumination normalised prototype image and compare this with a similarly illuminated gallery image. Suppose the 'normal' illumination conditions is $\mathbf{s}_{\text{norm}} = (0, 0, 1)^T$, the prototype image is given by: $I_p = \rho_{d,p}\mathbf{n}_p^{\text{est}}.\mathbf{s}_{\text{norm}}$. Recognition may then be effected by measuring the distance to each gallery image: $\|I_{\text{prototype}} - I_{\text{gallery}}\|$.

4.4.2 Linear Subspaces

Previous work [8, 64, 100] has demonstrated that a low-dimensional subspace can accurately capture the variation in images of a face resulting from arbitrarily complex variations in illumination (the three techniques used are introduced in Section 2.2.4). These linear subspace methods provide a powerful solution to the problem of illumination insensitive face recognition, in which the identity associated with the subspace which lies closest to a query image is reported as the unknown identity. A variety of approaches of varying complexity have been proposed to build these subspaces.

In this section we show how the 2.5D shape information recovered from a single image using the principal geodesic shape-from-shading approach described above complemented with the albedo is sufficient to build each of these subspaces.

Illumination Cone

Using the shape and albedo information recovered from a single image using the approach described in Section 4.3.4, we can synthesise novel images illuminated by a light source $\mathbf{s}_{\text{synth}}$ and which include cast shadows using:

$$I_p = \text{shadow}(z, \mathbf{s}_{\text{synth}}, p)\rho_{d,p}\mathbf{n}_p^{\text{est}}.\mathbf{s}_{\text{synth}} \tag{4.26}$$

Therefore, we can approximate a training set which samples the entire illumination cone by synthesising images with a range of light source directions across the illumination sphere. This is essentially a 'one-to-many' training strategy. In practise we follow [64] and build a training set comprising 45 images each synthesised with illumination from a single point light source from a different direction. We apply PCA to this training set and retain the first 11 dimensions, giving us an $N \times 11$ matrix B in which each column is a basis image. The closest subspace B to an image I is the one which minimises $\|BB^T I - I\|$. The identity associated with this subspace is the recognition result for the image.

Spherical Harmonic Basis Images

Let $\vec{\rho}_d$ denote a vector of length N containing the albedo values across a face's surface, such that $\rho_{d,p}$ is the albedo at point p. Similarly, the x, y and z components of the surface normals are stacked to form a further three vectors of length N: $\vec{n}_x$, $\vec{n}_y$ and $\vec{n}_z$, such that $n_{x,p}$ is the x component of the surface normal at point p. We define: $\vec{n}_{x^2} = \vec{n}_x.*\vec{n}_x$, where we use the operator $.*$ to express the component-wise product of any two vectors of the same length. Similarly for $\vec{n}_{y^2}$, $\vec{n}_{z^2}$, $\vec{n}_{xz}$, $\vec{n}_{yz}$ and $\vec{n}_{xy}$. The first nine harmonic images for

a surface with known normals and albedo are given by:

$$
\begin{aligned}
b_{00} &= \tfrac{1}{\sqrt{4\pi}}\vec{\rho}_d, & b^e_{10} &= \sqrt{\tfrac{3}{4\pi}}\vec{\rho}_d . * \vec{n}_z, \\
b^o_{11} &= \sqrt{\tfrac{3}{4\pi}}\vec{\rho}_d . * \vec{n}_y, & b^e_{11} &= \sqrt{\tfrac{3}{4\pi}}\vec{\rho}_d . * \vec{n}_x, \\
b_{20} &= \tfrac{1}{2}\sqrt{\tfrac{3}{4\pi}}\vec{\rho}_d . * (2\vec{n}_{z^2} - \vec{n}_{x^2} - \vec{n}_{y^2}), & & \\
b^o_{21} &= 3\sqrt{\tfrac{5}{12\pi}}\vec{\rho}_d . * \vec{n}_{yz}, & b^e_{21} &= 3\sqrt{\tfrac{5}{12\pi}}\vec{\rho}_d . * \vec{n}_{xz}, \\
b^o_{22} &= 3\sqrt{\tfrac{5}{12\pi}}\vec{\rho}_d . * \vec{n}_{xy}, & b^e_{22} &= \tfrac{3}{2}\sqrt{\tfrac{5}{12\pi}}\vec{\rho}_d . * (\vec{n}_{x^2} - \vec{n}_{y^2})
\end{aligned}
\tag{4.27}
$$

It is clear that these harmonic images may be derived from precisely the information recovered by our shape-from-shading algorithm. Once again, we form a matrix B containing the basis images as columns (this time of dimension $N \times 9$). However, this basis is not orthonormal. Using a QR decomposition, we find the $N \times 9$ orthonormal basis Q and 9×9 matrix R, such that $QR = B$. We may now compute the distance to the subspace using: $\|QQ^T I - I\|$ and perform recognition as described above.

The Space of Harmonic Images: There has been recent interest in deriving an efficient representation of the space of harmonic images [164]. The aim is to allow the harmonic basis images of any subject to themselves be described by a low-dimensional parameter vector. Zhang et al. [164] represented this space using convex linear combinations of exemplar harmonic bases. However, since the harmonic images contain directional components which are elements of a non-linear space, it is not clear that the application of a linear analysis will result in an efficient representation. Instead, we show that the principal geodesics of our model provide a more efficient representation.

It is clear that the spherical harmonic images may be written in terms of the principal geodesic parameters $\mathbf{b}$ representing a field of surface normals. To do so, we simply need rewrite n_x as: $n_x = \left(\text{Exp}_\mu(\mathbf{Pb})\right)_x$, where $(\dots)_x$ denotes the selection of the x-components of the parenthesised vector. We may similarly rewrite n_y and n_z in terms of $\mathbf{b}$. In order to fully represent the harmonic basis images, our representation must also

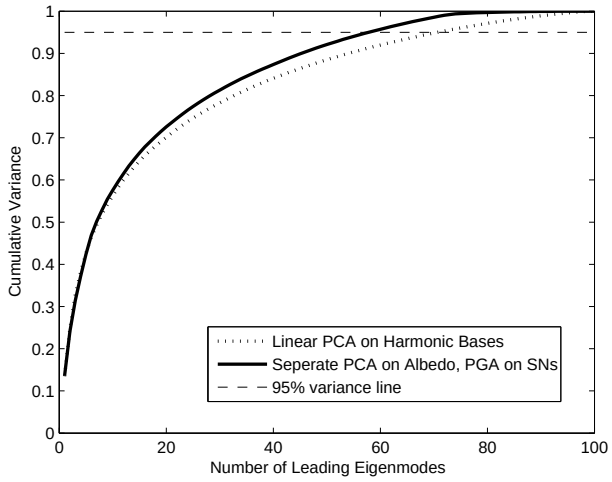


Figure 4.5: Cumulative variance versus number of eigenmodes for two different representations of the harmonic basis images.

capture the albedo. We suggest that it is more justifiable to apply linear analysis to the albedo values and perform PCA on the albedo separately.

In Figure 4.5 we provide empirical evidence of the improved efficiency of our representation using ground truth surface normals and albedo for 100 subjects. We plot the cumulative variance captured as a function of the number of leading eigenvectors retained for two different representations of the harmonic images. The broken line shows the cumulative variance when linear PCA is applied directly to the harmonic bases. The solid line shows the cumulative variance for a combined model in which albedo is modeled linearly using PCA and the surface normals are modeled using PGA (the albedo values are scaled such that they are commensurate with the units of the surface normals). The plot shows that a greater number of eigenmodes are required to capture the variance in the harmonic bases when a linear analysis is applied directly to the basis images as opposed to modeling albedo and surface normal variation separately. Quantitatively, to capture 95% of the variance, the linear model requires 71 dimensions whereas the model which uses PGA requires only 58.

It is also worth noting that the eigenvectors obtained by applying linear PCA directly

to the harmonic images are of size $O(9N)$, whereas the eigenvectors obtained by modeling the albedo and surface normals separately are of size $O(3N)$. Hence, the storage needed for such a model is smaller.

9 Points of Light

Lee et al. [100] use the simplest possible approach to constructing a subspace which captures the variation in illumination, in which 9 real images form the basis images of the low-dimensional space. They determine the best way in which to arrange the physical lighting in these images, which they call the universal configuration. We render images using Equation 4.26 for each of the universal lighting directions, given here in spherical coordinates: $\{(0,0),(68,-90),(74,108),(80,52),(85,-42),\ (85,-137),(85,146),$ $(85,-4),(51,67)\}$. These images are then used to form the basis images of the linear subspace. Recognition is once again effected by choosing the subspace which lies closest to a probe image.

4.5 Experimental Results

In this section we experiment with the methods for facial shape recovery and recognition described above. We commence by evaluating the performance of the method for shadow removal, albedo estimation and facial shape reconstruction. We then explore the use of the shape-parameters for the purposes of face recognition, and compare our method with the alternatives outlined in Section 4.4.2.

We begin by deriving the principal geodesics of a training sample of facial needle-maps. We once again use the 3DFS database [1], comprising laser range scans of 100 subjects collected using a *Cyberware*$^{\text{TM}}$ 3030PS laser scanner, in which the facial surfaces have been registered using Blanz and Vetter's algorithm [15]. From this database we

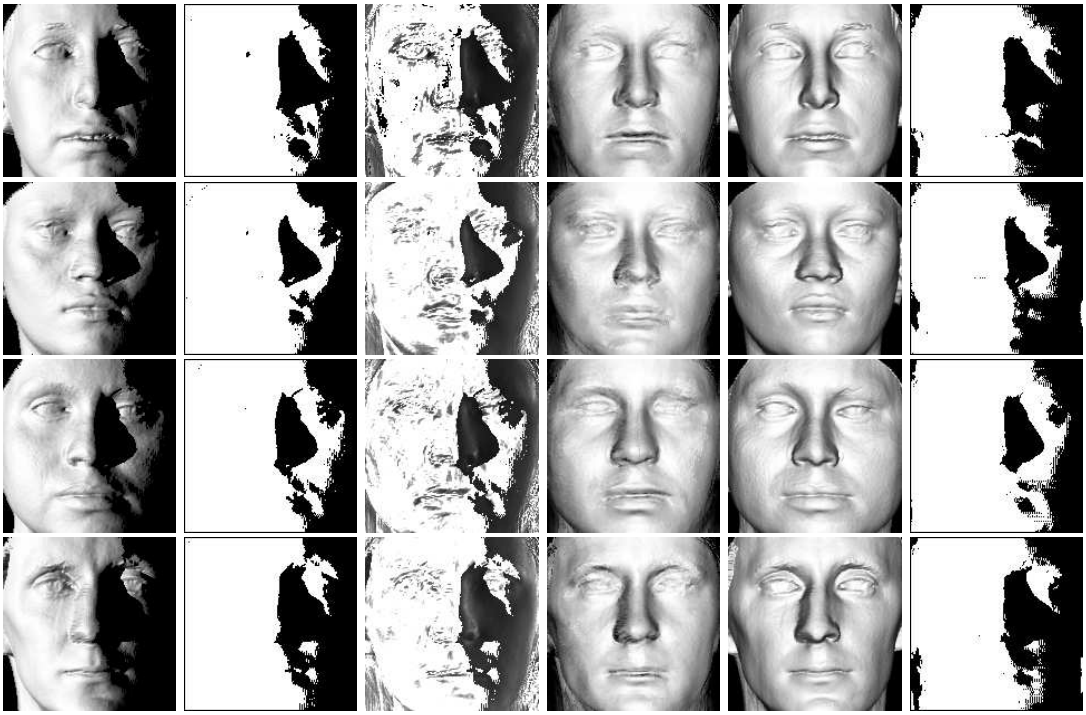
extract fields of surface normals onto a 124×142 pixel image plane. For ground truth, we use a leave-one-out strategy in which we train the model with 99 sets of data, leaving the remaining needle-map as out-of-sample ground truth.

4.5.1 Reconstruction

We begin by applying the method to known ground truth data allowing us to quantitatively assess the performance of the approach. We then apply the method to real world images, demonstrating the robustness of the approach under real world conditions. For the real world images, we show reconstructions and reilluminations of images from the Yale-B database [64]. These contain albedo variation and cast shadows. The algorithm executed in under 5 seconds using Matlab code running on a 2.4GHz Pentium 4 PC.

Ground Truth Data

In Figure 4.6 we demonstrate the performance of our method on ground truth data. We apply our algorithm to a selection of images of rendered ground truth needle-maps including cast shadows. In column (a) we show the input images. The needle-maps of the out-of-sample subjects are rendered with Lambertian reflectance and a point light source with direction $\mathbf{s} = (-1, 0, 1)^T$, i.e. 45° from the viewing direction. We also simulate the effect of cast shadows using the shadow map shown in column (b). We calculate shadow$(z, \mathbf{s}, p)$ from ground truth depth data. In column (c) we show the weight function $w_\sigma(\eta_p)$ for each pixel. It is clear that regions in cast shadow have been successfully down-weighted. In column (d) we show the needle-map $\mathbf{n}^{\text{est}}$ calculated from the input image, rendered with frontal illumination. For comparison, in column (e) we show the ground truth needle-map similarly illuminated. There is a good agreement between the two, even in areas in which no information was present in the input image (i.e. those in cast shadows). This suggests that the robust fit of the model has recovered globally accurate shape



(a) (b) (c) (d) (e) (f)

Figure 4.6: Fitting to images of ground truth needle-maps rendered with Lambertian reflectance and cast shadows.

information, and has filled-in the shadowed areas of the face. The mean surface normal error was typically $< 8^\circ$ across the whole needle-map. Finally, in column (f) we show the shadow map shadow$(z^{\text{est}}, \mathbf{s}, p)$, where z^{est} is the height map integrated from $\mathbf{n}^{\text{est}}$. Again, there is a good agreement between columns (b) and (f), suggesting that this represents a viable means to estimate regions which are in cast shadow.

In Figure 4.7 we examine the influence of illumination direction on the accuracy of the recovered surface normals. Once again we render ground truth, out-of-sample needle-maps with Lambertian reflectance and a point light source for which we simulate cast shadows. The direction of the light source is varied through a horizontal and vertical arc, i.e. from left to right and from top to bottom. Figure 4.7 plots the average surface normal error against the angle between the light source and viewing direction. The arc from left to right ($\mathbf{s} = (-1, 0, 0)^T, \ldots, (1, 0, 0)^T$) is shown as a solid line, while the arc

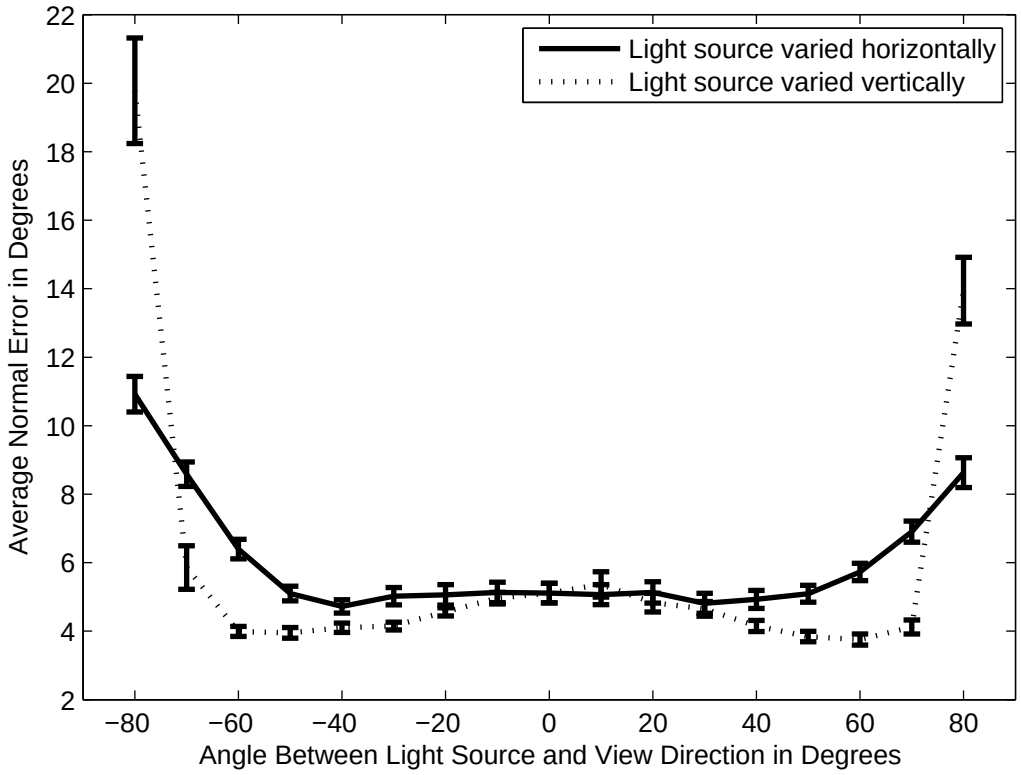


Figure 4.7: Average angular error of the recovered normal versus illumination direction.

from top to bottom ($\mathbf{s} = (0, 1, 0)^T, \ldots, (0, -1, 0)^T$) is shown as a broken line. The plot demonstrates that our method recovers globally accurate shape information, even when the lighting direction is extreme and hence, much of the face is in shadow. For example, illumination from the extreme right or left still results in an average normal error of less than 10°.

Real World Data

In Figure 4.8 we begin by providing a quantitative analysis of the control parameter ς, as defined in Section 4.3.2. We use our method to recover facial shape from an image which contains significant cast shadows and for which ground truth shape information is known. We show the (normalised) total angular error between the estimated and actual surface normals as ς is varied. It is clear there is a minima at approximately $\varsigma = 0.8$ and that any

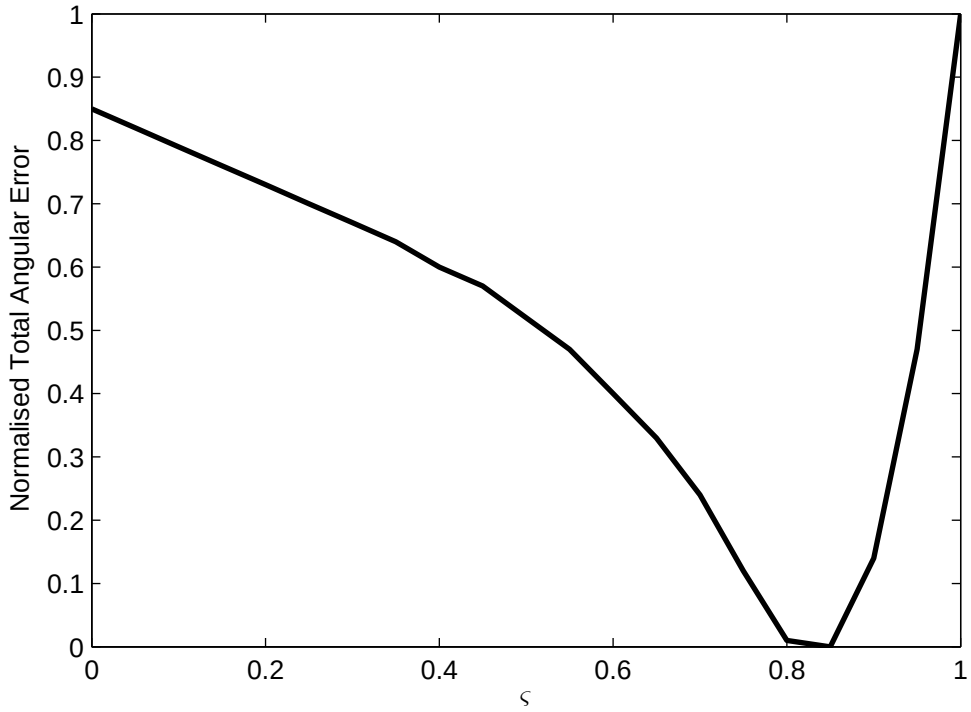


Figure 4.8: Plot of parameter ς versus normalised total angular error.

greater value sharply increases the error due to over-fitting. As ς tends to zero there is a smoother degradation as the recovered shape tends towards the mean face. For real world data, a more conservative setting of ς is required, but its effect is similar. In Figure 4.9 we show the shape recovered from a real world image as ς is varied. The effects of overfitting are clear as ς tends to one.

In Figure 4.10 we show results of applying our method to real world images. In the first column we show input images in which the subjects are illuminated by a light source 50° to the left. In the second column we show the recovered shape, rendered with constant albedo and frontal illumination. It is clear that the surface in the shadowed regions has been convincingly filled in. In the third column we show the shadow map calculated from the recovered shape. These agree well with the cast shadows visible in the first column. In particular, note the bump present in the shadow cast by the nose in the third row.

$\varsigma = 0.4$ $\varsigma = 0.6$ $\varsigma = 0.8$ $\varsigma = 1$

Figure 4.9: Effect of ς parameter fitting to real-world image.

Finally, in the fourth column, we demonstrate the quality of the shape information that can be recovered by our method, even for significantly non-frontal illumination. From the estimated needle-map, $\mathbf{n}^{\text{est}}$, we use the surface integration method of Frankot and Chellappa [59] to recover the surface height function. We show these surfaces rotated to a novel viewpoint and with the input image texture mapped onto the surface. The images show considerable stability under large change in viewpoint.

In Figure 4.11, we apply the method to the task of normalising variations in illumination for real world images. The first row shows the input images of a single subject under varying illumination. The subject is a challenging choice due to the large albedo variations caused by facial hair. The light source is moved in an arc along the horizontal axis to subtend an angle of $-50°$, $-25°$, $-10°$, $0°$, $10°$, $25°$ and $50°$ with the viewing direction. We use our method to estimate the normals, albedo and shadow map. We use facial symmetry to fill-in the missing albedo values for the shadow regions. In the second row we show the recovered needle-maps rendered with the estimated albedo and frontal lighting, effectively correcting for variation in input lighting. These synthesised images are of a good quality, even under large changes in illumination. Moreover, the effects of cast shadows are absent. The method does begin to break down for significantly non-frontal lighting, as can be seen for the results when the subtended angle is $50°$.

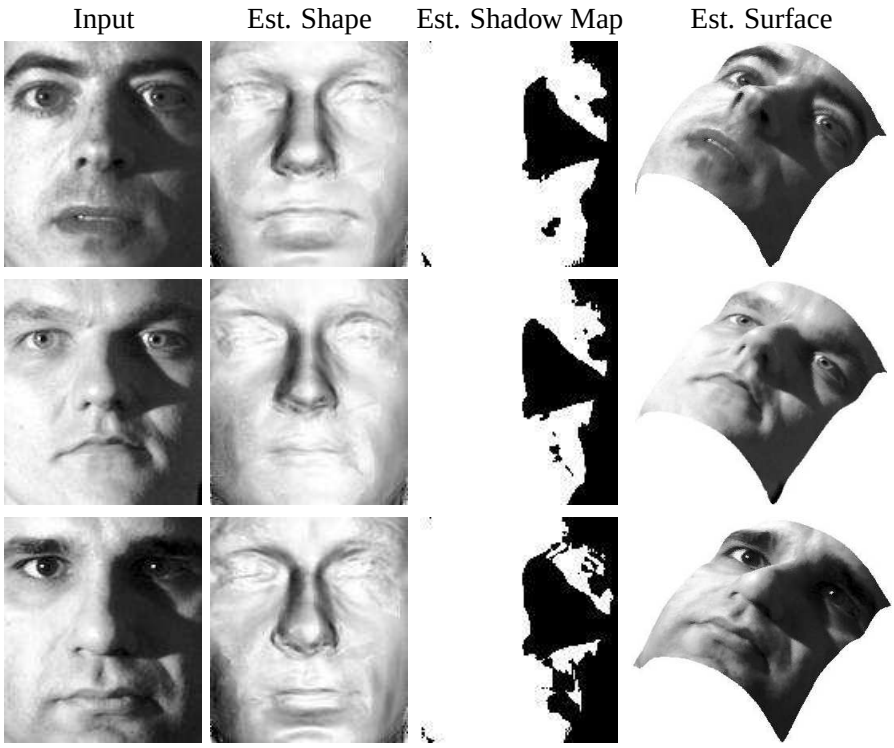


Figure 4.10: Shape and shadow map estimation for real world images.

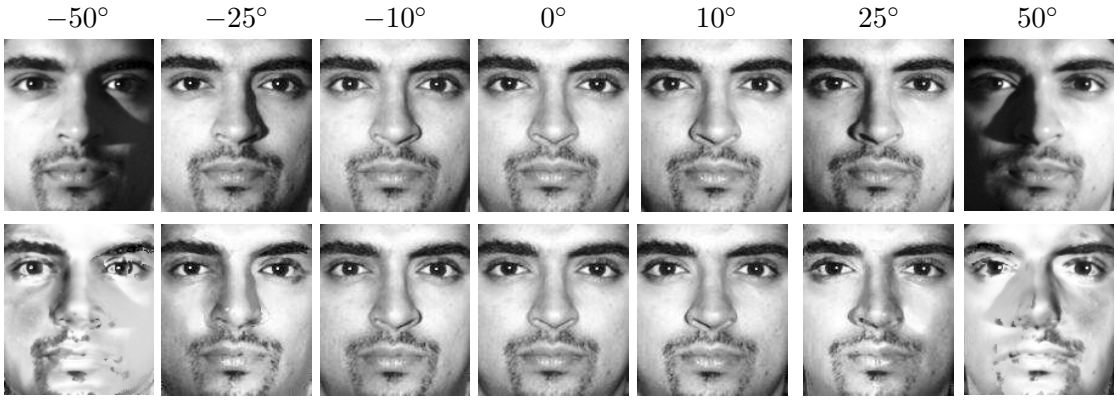


Figure 4.11: Correcting for variations in illumination for real world images.

Figure 4.12: The input images of the 10 Yale B subjects.

4.5.2 Recognition

In this section we present recognition results using our shape recovery algorithm. For testing we have used the Yale Face Database B [64] which contains images of 10 individuals (disjoint from the model training data) under 45 different illumination conditions. We group the lighting variation into 4 subsets of differing extremity (see [64] for details). The images have been manually aligned, cropped and resized to establish correspondence with the model.

In all cases, we use a single gallery image with known lighting. Although we could choose any illumination for the training image, we use frontal illumination for two reasons. Firstly, this configuration ensures that none of the face is in shadow. Non-frontal lighting would result in a degradation in the accuracy of the recovered shape and subsequent recognition performance. Secondly, this corresponds to a useful and realistic scenario in which the training image is captured using a standard camera with a flash lamp. We show the input images used in Figure 4.12. We apply our principal geodesic shape-from-shading algorithm as described in Section 4.3.4 to each training image and use the resulting shape and albedo information according to the recognition strategy employed.

In Figure 4.13 we show an example of each of the subspaces described in Section 4.4.2 constructed from shape and albedo information extracted from a single image. In (a) we

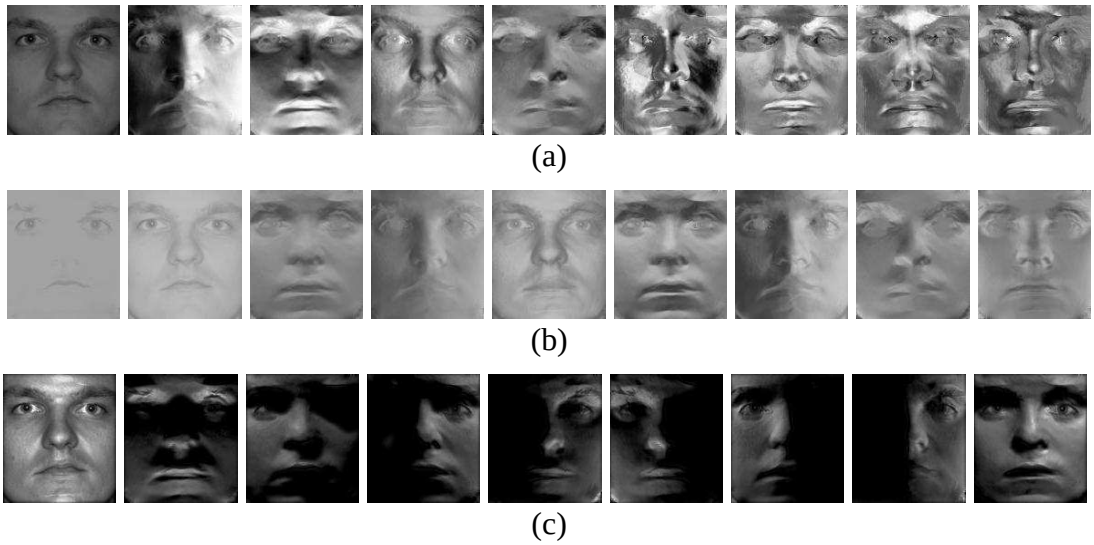


Figure 4.13: Linear subspaces generated from shape and albedo information recovered from a single image: (a) The (re-scaled) first 9 dimensions of an 11D approximation of the illumination cone [64] (b) The first 9 harmonic images [8] (c) Synthesised images of a subject illuminated by lights from the universal configuration [100].

show the first 9 dimensions of an 11D approximation to the illumination cone. These have been rescaled to the interval $(0, 1)$. In (b) we show the first 9 harmonic images scaled such that negative values are shown in black, positive values in white. In (c) we show synthesised images of the subject illuminated by lights from the universal configuration. These images appear plausible, including the estimated cast shadows.

In Table 4.1 we show recognition results using the methods described in sections 4.4 and 4.4.2. The first 6 rows show previously published results. The first three methods are baseline recognition approaches which perform poorly under large variations in illumination. The second three methods show the original results of using the linear subspaces described above for recognition. In each case, either multiple training images were used (> 7) or a 3D model was required. We then show results using the linear subspaces constructed using shape and albedo information extracted from only one training image. These result in an increase of about $3 - 4\%$ in the total error rate over the originally pub-

Comparison of Recognition Methods				
Method	Error Rate(%) vs. Illum			
	Subset 1&2	Subset 3	Subset 4	Total
Correlation [64]	0.0	23.3	73.6	29.1
Eigenfaces [64]	0.0	25.8	75.7	30.4
Eigenfaces w/o 1st 3 [64]	0.0	19.2	66.4	25.8
Harmonic Images [100]	0.0	0.0	3.571	1.1
Cones-cast [64]	0.0	0.0	0.0	0.0
9PL [100]	0.0	0.0	0.0	0.0
Harmonic Images (SFS)	0.0	0.0	12.86	4.0
Cones-cast (SFS)	0.0	0.0	13.57	4.2
9PL (SFS)	0.0	0.0	15.0	4.6
PGA Parameters (SFS)	0.0	15.0	48.6	19.1
Prototype Images (SFS)	0.0	3.57	28.57	16.2

Table 4.1: Recognition results on the Yale B database.

lished results, all the errors being concentrated into the most extreme lighting subset. We believe the reason the Harmonic Images outperform the other two subspaces in this case is because in constructing this subspace we use the recovered shape and albedo information directly, whereas for the Illumination Cone and 9 Points of Light, additional errors are introduced by the extra step of synthesising training images (in particular, estimating the cast shadows from the integrated surface). Finally we show the results using the 2.5D shape information directly. Although both methods outperform the baseline methods, both breakdown under extreme illumination, where extracting accurate shape information is very difficult (for example, the most extreme illumination in Figure 4.11).

4.6 Conclusions

In this chapter we have presented a statistical shape-from-shading framework in which the principal geodesics learnt from an exemplar distribution of surface normals provide a

global constraint which we use to update the surface normal estimates. We fit the statistical model globally, but use robust statistics to ensure that regions of low albedo or which fall into a cast shadow have little or no impact on the parameter estimate. The technique is capable of recovering a useful estimate of facial shape, even when significant portions of the face are entirely in shadow.

We used the technique for a number of face analysis tasks. We began by examining the accuracy of the method on data with known ground truth. We then showed how a cast shadow map can be estimated from both ground truth and real world imagery. This may prove useful to further analysis of the face using other methods. Finally, we demonstrated that the facial shape estimated by our method, combined with an estimate of the albedo and shadow map, is sufficient to synthesise images under frontal lighting from extreme input lighting directions.

We have presented a comprehensive evaluation of the viability of using 2.5D shape information to perform face recognition. In particular we have explored using facial shape recovered using shape-from-shading. We have presented an efficient and robust shape-from-shading algorithm which combines irradiance constraints and constraints on the principal geodesics of the field of surface normals.

We can conclude that shape-from-shading provides a realistic route to building robust, illumination invariant subspaces from single images. Our method may be used to complement existing methods for building these subspaces, in the case where only a single training image exists. However, the poor results obtained when using shape information recovered from probe images suggest that it is significantly simpler to infer the appearance of a face under extreme lighting having first recovered its shape under close-to-frontal lighting rather than the converse.

Chapter 5

Statistical Relations between Surface Normals and Height

She had one of those characteristic British faces that once seen are never remembered.

- Oscar Wilde

In the preceding two chapters, we have focussed solely on estimating surface orientation information from single images of faces. As we have demonstrated, this $2\frac{1}{2}$D shape representation is useful in its own right for synthesising novel illuminations, performing recognition and estimating albedo maps. However, for many purposes, the ultimate aim of facial shape-from-shading is to provide an estimate of the true 3-dimensional surface of the face. This is essential if we wish to synthesise images from a novel viewpoint or for model acquisition for applications such as animation. Up to this point, where we have required surface height information we have relied on existing generic methods to calculate it [59]. In this chapter we extend our consideration of the statistical modeling of surface normals to encompass height information. We show how to use the resulting models to reliably recover facial surfaces from images.

5.1 From Surface Orientation to Surface Height

Suppose that a facial surface $F \in \mathbb{R}^3$ is projected orthographically onto the image plane and parameterised by the function $z(x, y)$. The second partial derivatives of this surface are independent of the order of differentiation:

$$\partial_{xy} z(x, y) = \partial_{yx} z(x, y). \tag{5.1}$$

If the surface gradients of a field of surface normals $\mathbf{n}(x, y)$ are $p(x, y) = \frac{n(x,y)_x}{n(x,y)_z}$ and $q(x, y) = \frac{n(x,y)_y}{n(x,y)_z}$, they are said to be *integrable* if they also satisfy this condition, i.e.

$$\partial_y p(x, y) = \partial_x q(x, y). \tag{5.2}$$

Recovering the surface height function from a field of surface normals or surface gradient estimates is known as the *surface integration* or *height-from-gradient* problem. Given a field of surface normals that satisfy the integrability condition, this problem is straightforward to solve. It simply requires the assignment of an initial height to a chosen starting point, an arbitrary path of integration to be followed through the field of normals and the height to be summed from the surface gradients given by the local surface normal. However, difficulties arise when the surface normals are noisy or contain systematic errors. In this case, the resulting surface will depend on the path of integration chosen.

This problem has received considerable attention in its own right. Typically it is posed in either local or global terms. Local integration methods [159] are conceptually simpler but are highly sensitive to noise and are hence unsuitable for use with normals estimated using shape-from-shading. Global approaches tend to be more robust to noise and recover smoother surfaces. Perhaps the most popular global method is that of Frankot and Chellappa [59]. They project the estimated gradient field into the Fourier domain to impose

integrability constraints and then recover surface height using an inverse Fourier transform. Although this approach is both elegant and numerically efficient it is sensitive to abrupt changes in orientation, for example at the object boundary. This means that careful parameter selection is necessary to recover a stable surface. It also requires that the surface be continuous. Wei and Klette [154] have enhanced this approach by incorporating more complex regularisation constraints in the Fourier domain. More recently, Agrawal et al. [3] have revisited this problem and shown how to pose it algebraically. They exclude gradient measurements which do not have zero curl, i.e. where the surface is not locally integrable, and provide improved results over the original method.

Although these methods are useful, they introduce errors of their own and when combined with imperfect surface normals provided by shape-from-shading, the two sources of error compound each other. Figure 5.1 highlights these problems. On the left of the figure a depth map of a face is shown, which was recovered using a laser range sensor. The second column shows a profile view of the depth map rendered with Lambertian reflectance and a frontal light source. The third column shows the result of using shape-from-shading to recover a field of surface normals from an image of the face in a frontal view and integrating these normals into a surface (we use the Worthington and Hancock [156] shape-from-shading algorithm and Frankot and Chellappa [59] surface integration algorithm). The errors introduced by both processes are compounded resulting in a very inaccurate surface. To illustrate this, the profile view of the surface recovered from ground truth surface normals using the method of Frankot and Chellappa [59] is shown on the right of the figure. Even with noise-free, ground truth normals the recovered surface is inaccurate. Areas which were close to the occluding boundary in the front view have been poorly recovered resulting in a loss of depth (the ear is too close to the nose). Also, the steepness of the surface at the base of the nose has caused the height of the lower half of the face to be exaggerated (in the ground truth surface the chin is at almost the same depth as the

Figure 5.1: Errors compound each other when shape-from-shading is combined with surface integration. From left to right: depth map, ground truth profile view, surface recovered using shape-from-shading, profile view of surface integrated from ground truth surface normals using Frankot and Chellappa's [59] method.

forehead, whereas in the estimated surface the chin is noticeably further forward).

5.1.1 Exploiting a Statistical Model of Surface Height

The improvements in the surface orientation information that may be recovered using the methods described in Chapters 3 and 4, depend upon imposing a global statistical constraint on the field of surface normals. It seems clear that in applying a generic surface integration technique to these normals ignores the important information we have recovered in fitting this statistical model. In other words, the resulting surface height function is free to assume any global shape, without being subject to a similar model-based constraint. Our aim is therefore to show how a shape model may be incorporated into the surface integration process.

A Coupled Model

We begin this chapter by showing how a coupled statistical model may be built that jointly describes variations in surface normal direction and height over the surface of a face. We construct separate eigenspaces for the surface normal and height variations from the covariance matrices of the training data. We fit the surface normal component of this model in the same way as that described in Chapter 3. With the parameters of the surface

normal subspace to hand, we use a quasi-Newton minimisation procedure to find the parameters of the coupled model that give the best match to those in the surface normal subspace. Hence, the model is fitted to needle-map data in the surface-normal subspace, and the recovered parameters are used to project the fitted model into the height subspace to recover the facial surface. This process implicitly recovers the surface height function without having to explicitly integrate the recovered field of surface normals.

This somewhat naive approach relies entirely on the statistical learning method employed (namely PCA) to explain the relationship between the field of surface normals and the surface height function. Nonetheless, we show that the approach provides improved results over using a generic surface integration method.

Imposing Height Constraints on Surface Normals

We then take this analysis one step further by discarding the statistical model of surface normals altogether. Instead we show that a statistical surface height model can be used to provide a similar constraint for surface normals. We do so by examining the relationship between the parameters of a surface height model and the surface normals of the resulting surface. We first show how a field of surface normals may be expressed in terms of the parameters of a height model. We demonstrate that by minimising the error between a field of normals expressed in these terms and an input field of normals, we find the closest field of normals in which both integrability and a statistical constraint on the corresponding surface are enforced.

We then show how the modes of variation of a surface height model can be used to define modes of variation for the corresponding surface normals. By examining the relationship between the height parameters and the surface normals we show that the parameters of the height model may be recovered directly from the corresponding field of normals. We use this observation to develop a robust iterative algorithm for facial

surface recovery in which a field of surface normals is estimated subject to the constraints of the height model. At each iteration the parameters of the surface model are known implicitly and from these parameters we can recover the height map. We term this process 'model-based integration'. The approach aims to retain the advantages of using a surface normal model (speed and good data-closeness) but yields a height map without having to explicitly integrate the surface normals.

5.2 Coupling Height and Surface Normal Variation

In this section we show how a statistical model may be constructed that captures variation in both surface height and surface normal direction. The motivation behind the idea was provided by the Active Appearance Model and Coupled View AAM developed by Cootes et al. [33, 38]. In both cases the idea is to learn the relationship between two representations of the face. In the case of the standard AAM this is between shape and intensity whilst in the coupled view AAM it is between the parameters of different AAMs built for multiple views. Our aim is the same, in that we wish to jointly model two representations of face shape.

The statistical model of surface normal direction is constructed in exactly the same way as in Chapter 3. The matrix containing the K eigenvectors of surface normal variation are captured in $\mathbf{P}_s$. Likewise, a field of surface normals $\mathbf{U}$ may be represented by the parameter vector:

$$\mathbf{b}_s = \mathbf{P}_s^T \mathbf{U}. \tag{5.3}$$

We now describe how the statistical model of variation in surface height is constructed.

5.2.1 A Surface Height Model

The training set comprises K range images of size $N = X_{\text{res}} \times Y_{\text{res}}$ pixels. The kth training sample may be represented by the vector of height values ordered according to the raster scan

$$\mathbf{z}^k = [z^k(1,1), z^k(1,2), \ldots, z^k(X_{\text{res}}, Y_{\text{res}})]^T. \tag{5.4}$$

The mean surface height vector $\hat{\mathbf{z}}$ is given by $\hat{\mathbf{z}} = \frac{1}{K}\sum_{i=1}^{K} \mathbf{z}^i$. We form the $N \times K$ data matrix of height values using:

$$\mathbf{D}_h = [(\mathbf{z}^1 - \hat{\mathbf{z}})|(\mathbf{z}^2 - \hat{\mathbf{z}})|\ldots|(\mathbf{z}^K - \hat{\mathbf{z}})]. \tag{5.5}$$

In precisely the same manner as for the surface normals, we once again use PCA to extract the set of K orthogonal modes of variation $\mathbf{P}_h$ from the covariance matrix $\mathbf{L}_h = \frac{1}{K}\mathbf{D}_h\mathbf{D}_h^T$. Again, a long-vector of height values $\mathbf{z}^k$ can be projected onto the eigenvectors and represented using the vector of model parameters $\mathbf{b}_h = \mathbf{P}_h^T(\mathbf{z}^k - \hat{\mathbf{z}})$.

5.2.2 Coupling the Surface Normal and Height Models

Each training sample can be summarised by the parameter vectors $\mathbf{b}_s$ and $\mathbf{b}_h$, representing the needle-map and height map of the sample respectively. Since surface height and the field of surface normals are closely related (recall that $\mathbf{n}(x,y) = (\partial_x z(x,y), \partial_y z(x,y), 1)$) the two sets of parameters will contain strong correlations. To illustrate this, Figure 5.2 shows the correlation matrix for the first 50 parameters of the surface normal and height models. There are strong correlations between the two, particularly along the diagonal. This suggests that modes of corresponding importance capture similar modes of variation. Intuitively, our argument is that a similar combination of modes of variation will describe a face's shape in both the surface normal and surface height domain. We intend to learn

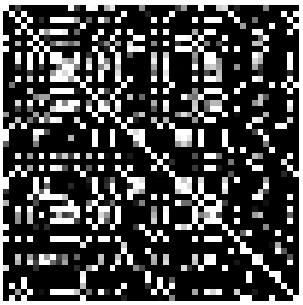


Figure 5.2: Correlation matrix for the first 50 parameters of the surface normal and height models.

this relationship using PCA.

In both models, we may consider small scale variation as noise. Hence, if the ith eigenvalue for the surface normal model is $\lambda_{s,i}$, we need only retain S eigenmodes to retain c percent of the model variance. We choose S as the smallest integer that satisfies $\sum_{i=1}^{S} \lambda_{s,i} \geq \frac{c}{100} \sum_{i=1}^{K} \lambda_{s,i}$. Similarly for the height model we retain H eigenmodes to capture c percent of the variance.

For the kth training sample we can generate the concatenated vector of length $S + H$:

$$\mathbf{b}^k = \begin{pmatrix} \mathbf{W}_s \mathbf{b}_s^k \\ \mathbf{b}_h^k \end{pmatrix} = \begin{pmatrix} \mathbf{W}_s \mathbf{P}_s^T \mathbf{U}^k \\ \mathbf{P}_h^T (\mathbf{z}^k - \hat{\mathbf{z}}) \end{pmatrix}, \tag{5.6}$$

where $\mathbf{W}_s$ is a diagonal matrix of weights for each surface normal model parameter, allowing for the different relative weightings of the surface normal and depth models. The reason for performing this weighting is that the elements of $\mathbf{b}_s$ have units of radians, while $\mathbf{b}_h$ have units of distance, so they cannot be compared directly. We follow Cootes and Taylor [35] and set $\mathbf{W}_s = r\mathbf{I}_S$ where r^2 is the ratio of the total depth variance to the total surface normal variance, i.e.

$$r = \sqrt{\frac{\sum_{i=1}^{H} \lambda_{h,i}}{\sum_{i=1}^{S} \lambda_{s,i}}}, \tag{5.7}$$

and $\mathbf{I}_S$ is the identity matrix of size S. The coupled model data matrix is $\mathbf{D}_c = [\mathbf{b}^1|\ldots|\mathbf{b}^K]$, where $\mathbf{b}^k$ represents the concatenated vector for the kth training sample. Because of the choice of weight matrix, the matrix $\mathbf{D}_c$ is the correlation matrix for the concatenated parameters. In order to find redundancies in the correlations between the two models, we now apply a final PCA to give the coupled model:

$$\mathbf{b} = \mathbf{P}_c\mathbf{c} = \begin{pmatrix} \mathbf{P}_{cs} \\ \mathbf{P}_{ch} \end{pmatrix} \mathbf{c}, \tag{5.8}$$

where $\mathbf{P}_c$ are the eigenvectors and $\mathbf{c}$ is a vector of coupled parameters controlling both the surface normal model and height model simultaneously. The matrix $\mathbf{P}_{cs}$ has S rows, and represents the first S eigenvectors, corresponding to the surface normal subspace of the model. The matrix $\mathbf{P}_{ch}$ has H rows, and represents the final H eigenvectors, corresponding to the height subspace of the model.

Crucially, we may now express the vectors of projected surface normal directions and height values directly in terms of the parameter vector $\mathbf{c}$:

$$\mathbf{U} = \mathbf{P}_s\mathbf{W}_s^{-1}\mathbf{P}_{cs}\mathbf{c}, \tag{5.9}$$

$$\mathbf{z} = \hat{\mathbf{z}} + \mathbf{P}_h\mathbf{P}_{ch}\mathbf{c}. \tag{5.10}$$

For compactness we write: $\mathbf{Q}_s = \mathbf{P}_s\mathbf{W}_s^{-1}\mathbf{P}_{cs}$ and $\mathbf{Q}_h = \mathbf{P}_h\mathbf{P}_{ch}$.

5.2.3 Fitting the Coupled Model to Surface Normal Data

Fitting the model to data involves estimating the parameter vector $\mathbf{c}$ from surface normal data. To do this we seek the coupled model parameters which minimise the error between the measured and reconstructed parameters describing the field of surface normals. In doing so we implicitly recover the surface which is also represented by the coupled model

parameters.

Suppose that $\mathbf{U}$ is a vector of length $2N$ that represents a field of surface normals under azimuthal equidistant projection obtained by applying the shape-from-shading algorithm of Chapter 3 to the brightness image of a face. We fit the model to data seeking the vector $\mathbf{c}^*$ of length $S + H$ that satisfies the condition

$$\mathbf{c}^* = \arg\min_{\mathbf{c}} \|\mathbf{P}_s^T\mathbf{U} - \mathbf{W}_s^{-1}\mathbf{P}_{cs}\mathbf{c}\|^2. \tag{5.11}$$

The corresponding best-fit vector of depth values is given by

$$\mathbf{z} = \hat{\mathbf{z}} + \mathbf{Q}_h\mathbf{c}^*. \tag{5.12}$$

This provides us with a means by which to recover surface height from a field of surface normals in a model-based way. In practise we use a Matlab implementation of a quasi-Newton minimisation process to solve Equation 5.11. In our experiments, we contrast this with using a generic height-from-gradient technique and show that the method is capable of recovering accurate profiles from frontal images.

5.3 Imposing Surface Height Constraints on Surface Normals

The surface height model described above can be used to efficiently represent a class of surfaces such as human faces in a low-dimensional space. In the previous section we used the model to solve the problem of surface integration. However, such a model also proves useful to the problem of estimating facial shape from images by transforming the ill-posed shape-from-shading problem into one of estimating a low-dimensional parameter vector.

However, since surface reflectance and hence measured intensity is governed by the orientation of the local surface and not directly by its height, the parameter estimate can only be made by optimising the error between predicted and measured appearance. This is the approach adopted by Atick et al. [5] and more recently Blanz and Vetter [15]. There are two problems with this approach. Firstly, the optimisation process is computationally expensive and the error functional may contain local minima. Secondly, such an approach leads to model dominance in which the recovered surface represents a viable facial shape but lacks photorealism and only weakly satisfies data-closeness.

These weaknesses motivate the work presented in Chapters 3 and 4. Here, the normals can be estimated directly from the image intensity and projection into the model space requires only a matrix multiplication. In addition, hard local irradiance constraints can be satisfied by simply rotating the surface normals from their directions given by the model parameters. This approach ensures good data-closeness and hence realistic reillumination under novel lighting. By avoiding posing the problem as a non-linear minimisation the method is also computationally efficient.

One drawback of this approach is that the recovered surface is in the form of surface normals. If the desired output is a depth map (in order to allow the synthesis of novel facial poses or to estimate shadow maps) then the surface normals must be integrated into a surface as described above. This process introduces errors of its own and is potentially computationally expensive.

In this section we take a step further than in Section 5.2, by showing that in fact we can impose the constraints of a surface height model in the surface normal domain.

5.3.1 A Global Statistical Integrability Constraint

We can express the facial surface $F \in \mathbb{R}^3$ projected orthographically onto the image plane and parameterised by the function $z(x, y)$ in terms of a base surface $\hat{z}(x, y)$ plus a linear

combination of K surface functions Ψ_i (or modes of variation):

$$z_{\mathbf{b}}(x,y) = \hat{z}(x,y) + \sum_{i=1}^{K} b_i \Psi_i(x,y), \tag{5.13}$$

where the coefficients b_i are the surface parameters. Two instances of this representation have proven useful to the shape-from-shading problem. The first is that of Frankot and Chellappa [59], where the fourier basis functions are used. These provide a complete orthonormal basis for z. However, the power of representation of the fourier basis means that they form a rather weak constraint on the surface. The second instance in which this formulation has been used is where the basis functions form a statistical model learnt from exemplar facial surfaces. This is the model used in the previous section where the base shape and modes of variation are given by applying PCA to a representative sample of exemplar surfaces. The base surface $\hat{z}$ is the mean average of the training samples and Ψ_i is the eigenvector of the covariance matrix of the training samples corresponding to the ith largest eigenvalue. It is this model that we use in the following sections.

Normals from Height Parameters

We may express the normals of the surface directly in terms of the parameter vector $\mathbf{b}$:

$$\mathbf{n}_{\mathbf{b}}(x,y) = \begin{pmatrix} \partial_x \hat{z}(x,y) + \sum_{i=1}^{K} b_i \partial_x \Psi_i(x,y) \\ \partial_y \hat{z}(x,y) + \sum_{i=1}^{K} b_i \partial_y \Psi_i(x,y) \\ 1 \end{pmatrix}. \tag{5.14}$$

In other words, the normal may be expressed in terms of the sum of the gradient of the mean surface with a linear combination of the gradients of the eigenvectors at the corresponding point. If we write the partial derivatives of the mean surface as: $\hat{p}(x,y) = \partial_x \hat{z}(x,y)$ and $\hat{q}(x,y) = \partial_y \hat{z}(x,y)$ and define $p_i(x,y) = \partial_x \Psi_i(x,y)$ and $q_i(x,y) =$

$\partial_y \Psi_i(x,y)$, i.e. the partial derivatives of the eigenvectors in the x and y directions, we may rewrite this as:

$$\mathbf{n_b}(x,y) = \begin{pmatrix} \hat{p}(x,y) + \sum_{i=1}^{K} b_i p_i(x,y) \\ \hat{q}(x,y) + \sum_{i=1}^{K} b_i q_i(x,y) \\ 1 \end{pmatrix}. \tag{5.15}$$

When we wish to refer to the corresponding vectors of unit length we use: $\hat{\mathbf{n}}_\mathbf{b}(x,y) = \frac{\mathbf{n_b}(x,y)}{\|\mathbf{n_b}(x,y)\|}$.

A field of normals expressed in this manner satisfies two constraints. First, the field of normals will be integrable since they correspond exactly to the surface given by Equation 5.13. Second, the surface corresponding to the field of normals is also constrained to lie within the span of the surface height model. We term this constraint *model-based integrability*.

5.3.2 Relating Modes of Surface Height Variation and Normals

If the mean surface is deformed according to the ith eigenvector Ψ_i by an amount given by the parameter b_i, the resulting height at a point (x,y) is $\hat{z}(x,y) + b_i \Psi_i(x,y)$. The corresponding normal of the deformed surface is therefore given by:

$$\mathbf{n}_{b_i}(x,y) = (\hat{p}(x,y) + b_i p_i(x,y), \hat{q}(x,y) + b_i q_i(x,y), 1)^T. \tag{5.16}$$

Again we write: $\hat{\mathbf{n}}_{b_i}(x,y) = \frac{\mathbf{n}_{b_i}(x,y)}{\|\mathbf{n}_{b_i}(x,y)\|}$ for the unit length normal.

Mode of Surface Variation = Geodesic Submanifold

As the mean surface is deformed according to the ith mode of variation, the surface normal at each pixel is rotated along a great circle. This is of interest because it shows that a

subspace of the surface height model corresponds to a subspace of the field of normals. As discussed in Chapter 4, unit normals may be considered as points lying on a spherical manifold, a subspace of which is a geodesic submanifold (a great circle in the spherical case).

We now prove that a mode of surface height variation corresponds to perturbing the mean surface normal along a geodesic submanifold (a great circle). To do this, we show that the direction of the axis of rotation remains constant as the deformation parameter is varied. In other words, as b_i varies, $\hat{\mathbf{n}}_{b_i}(x,y)$ is rotated about a constant axis and therefore traces out a great circle.

The surface normal of the mean surface is $\mathbf{n}_0(x,y) = (\hat{p}(x,y), \hat{q}(x,y), 1)^T$. The cross product of $\mathbf{n}_0(x,y)$ and $\mathbf{n}_{b_i}(x,y)$ is:

$$\begin{aligned}
\mathbf{n}_0(x,y) \times \mathbf{n}_{b_i}(x,y) &= \begin{pmatrix} \hat{q}(x,y) - (b_i q_i(x,y) + \hat{q}(x,y)) \\ (b_i p_i(x,y) + \hat{p}(x,y)) - \hat{p}(x,y) \\ \hat{p}(x,y)\,(b_i q_i(x,y) + \hat{q}(x,y)) - \hat{q}(x,y)\,(b_i p_i(x,y) + \hat{p}(x,y)) \end{pmatrix} \\
&= b_i \begin{pmatrix} -q_i(x,y) \\ p_i(x,y) \\ \hat{p}(x,y) q_i(x,y) - \hat{q}(x,y) p_i(x,y) \end{pmatrix}.
\end{aligned}$$

Since we can factor out b_i, it therefore has no effect on the direction of the axis of rotation from $\hat{\mathbf{n}}_0(x,y)$ to $\hat{\mathbf{n}}_{b_i}(x,y)$.

The observation we would like to make here is that each subspace of a surface height model corresponds to a subspace of the field of surface normals. This motivates our attempts to enforce the constraints of a surface height model on a field of surface normals in the same way that we have used statistical surface normal models in the previous two chapters. In other words, the modes of variation of a surface height model can also be considered as modes of variation for a surface normal model.

Recovering Surface Parameters

Although we have shown that varying the surface height parameter b_i corresponds to rotating each surface normal locally about a great circle, the relationship between the magnitude of b_i and the magnitude of the surface normal deformation is complex and non-linear. From the field of normals of a surface projected onto the ith mode of surface height variation (i.e. a surface deformed according to the ith eigenvector by an unknown amount), we can recover the magnitude of the surface deformation, b_i, locally at any point using the following equality:

$$\mathbf{n}_0(x,y) \cdot \mathbf{n}_{b_i}(x,y) = |\mathbf{n}_0(x,y)||\mathbf{n}_{b_i}(x,y)|\hat{\mathbf{n}}_0(x,y) \cdot \hat{\mathbf{n}}_{b_i}(x,y). \tag{5.17}$$

Substituting in $\mathbf{n}_{b_i}(x,y) = (\hat{p}(x,y) + b_i p_i(x,y), \hat{q}(x,y) + b_i q_i(x,y), 1)^T$ and $\mathbf{n}_0(x,y) = (\hat{p}(x,y), \hat{q}(x,y), 1)^T$ gives:

$$\hat{p}(x,y)^2 + \hat{p}(x,y) b_i p_i(x,y) + \hat{q}(x,y)^2 + \hat{q}(x,y) b_i q_i(x,y) + 1 = |\mathbf{n}_0(x,y)||\mathbf{n}(x,y)|\hat{\mathbf{n}}_0(x,y) \cdot \hat{\mathbf{n}}_{b_i}(x,y) \tag{5.18}$$

$$b_i = \frac{|\mathbf{n}_0(x,y)||\mathbf{n}_{b_i}(x,y)|\hat{\mathbf{n}}_0(x,y) \cdot \hat{\mathbf{n}}_{b_i}(x,y) - \hat{p}(x,y)^2 - \hat{q}(x,y)^2 - 1}{\hat{p}(x,y) p_i(x,y) + \hat{q}(x,y) q_i(x,y)}. \tag{5.19}$$

If a surface $z(x,y)$ is represented with minimum least squares error by the parameter vector $\mathbf{b}$, we have shown that the parameter b_i may be recovered from the field of normals $\mathbf{n}_{b_i}(x,y)$, i.e. the surface normals of the surface: $\hat{z}(x,y) + b_i \Psi_i(x,y)$. However, it is not clear how the field of surface normals $\mathbf{n}_{b_i}(x,y)$ can be estimated from a field of surface normals $\mathbf{n}(x,y)$ which may not lie within the span of the model. Instead, we pose the problem of imposing statistical integrability on a field of surface normals in terms of operations relating the surface gradients. This results in a linear minimisation problem.

5.3.3 Model-based Integration

To apply the constraint to a (possibly non-integrable) field of surface normals $\mathbf{n}(x, y)$, we seek the parameter vector $\mathbf{b}^*$, whose field of surface normals given by Equation 5.15 minimises the distance to $\mathbf{n}(x, y)$. We pose this as minimising the squared error between the surface gradients of $\mathbf{n}(x, y)$ and those given by Equation 5.15. The surface gradients of $\mathbf{n}(x, y)$ are $p(x, y) = \frac{n_x(x,y)}{n_z(x,y)}$ and $q(x, y) = \frac{n_y(x,y)}{n_z(x,y)}$. The optimal solution is therefore given by:

$$\begin{aligned} \mathbf{b}^* = \quad & \arg\min_{\mathbf{b}} \sum_{x,y} \left[\hat{p}(x, y) + \sum_{i=1}^{K} b_i p_i(x, y) - p(x, y)\right]^2 \\ & + \left[\hat{q}(x, y) + \sum_{i=1}^{K} b_i q_i(x, y) - q(x, y)\right]^2. \end{aligned} \tag{5.20}$$

The solution to this minimisation is linear in $\mathbf{b}$ and is solved using linear least squares as follows. If the input image is of dimension $N = X_{\text{res}} \times Y_{\text{res}}$, we subtract the mean surface gradients from the surface gradients of the field of surface normals $\mathbf{n}(x, y)$ to form a centred vector of length $2N$:

$$\mathbf{G} = \begin{bmatrix} p(1,1) - \hat{p}(1,1) \\ q(1,1) - \hat{q}(1,1) \\ \dots \\ p(X_{\text{res}}, Y_{\text{res}}) - \hat{p}(X_{\text{res}}, Y_{\text{res}}) \\ q(X_{\text{res}}, Y_{\text{res}}) - \hat{q}(X_{\text{res}}, Y_{\text{res}}) \end{bmatrix}. \tag{5.21}$$

We then form the $2N \times K$ matrix of the surface gradients of the eigenvectors, $\mathbf{\Psi}$, whose ith column is $\mathbf{\Psi}_i = [p_i(1,1), q_i(1,1), \dots, p_i(X_{\text{res}}, Y_{\text{res}}), q_i(X_{\text{res}}, Y_{\text{res}})]^T$. We may now state our least squares problem in terms of matrix operations:

$$\mathbf{b}^* = \arg\min_{\mathbf{b}} \|\mathbf{\Psi}\mathbf{b} - \mathbf{G}\|^2. \tag{5.22}$$

The least squares solution is found using the pseudoinverse of $\boldsymbol{\Psi}$:

$$\mathbf{b}^{*} = \left(\boldsymbol{\Psi}^{T}\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}^{T}\mathbf{G}. \tag{5.23}$$

This is a computationally efficient operation since it consists of only matrix multiplications. With the optimal parameter vector to hand, the field of surface normals satisfying the model-based integrability constraint is given by Equation 5.15. Furthermore, we have also implicitly recovered the surface height, which is given by Equation 5.13. This provides a means to recover the surface height from a field of normals without explicitly integrating the field of normals.

We can use this process in two ways:

1. As solely a surface integration technique, applied in the same way as the technique described in Section 5.2. In this case we take as input a field of surface normals recovered using the methods described in Chapter 3 or Chapter 4. We fit the surface height model to this field of surface normals using Equation 5.23 and return the corresponding surface height given by Equation 5.13.

2. Alternatively, we may use the surface height model to provide a constraint to be used during shape-from-shading. The idea is to replace the constraint provided by the statistical surface normal model with one on surface height. This is a stricter constraint, since it requires not only that the normals lie within the span of a statistical model, but also that they correspond to an integrable surface. Yet we can still impose the constraint in the surface normal domain (using the techniques described above) and may hence enforce data-closeness constraints locally as in the previous two chapters.

We describe how this second process is performed in the following subsection.

5.3.4 Shape-from-shading with Model-based Integrability

We incorporate the statistical model-based integrability constraint into geometric shape-from-shading in the same way as in the previous two chapters. To do so we iteratively interleave the two processes of strictly enforcing satisfaction of Lambert's law and enforcing the model-based integrability constraint on the field of surface normals. However, in contrast to the techniques presented in chapters 3 and 4, the method returns both a field of surface normals (which strictly satisfy Lambert's law) and the surface height function (which lies within the span of the statistical surface height model). Moreover, since the model is based on surface height as opposed to surface orientation it is more efficient (fewer dimensions are required to retain the same proportion of the variance) while also offering a stricter constraint.

The shape-from-shading algorithm using model-based integrability as the constraint is described in Algorithm 5.1. Note that this algorithm is similar in structure to Algorithm 4.1 in the previous chapter. The primary difference is that the statistical constraint is now based on a surface height model as opposed to a surface normal model. In addition, Algorithm 5.1 returns an estimate of the surface height function without requiring the use of a generic surface integration technique. We note once again that the imposition of both constraints is implemented using only matrix multiplications and is hence computationally efficient.

5.4 Experiments

In this section we present experimental results using the two methods described in this chapter for surface height recovery.

We begin by comparing the two statistical approaches to the problem of surface integration. We use ground truth fields of surface normals and compare the accuracy of the

Algorithm 5.1: Geometric shape-from-shading using a model-based integrability constraint.

Input: Light source $\mathbf{s}$, image intensities $I(x, y)$, surface gradients of statistical surface height model $\mathbf{\Psi}$ and average surface gradients $\hat{p}(x, y)$ and $\hat{p}(x, y)$

Output: Estimated surface normal map $\mathbf{n}(x, y)$ and surface height function $z_{\mathbf{b}}(x, y)$

1 Set iteration $t = 0$;

2 Initialise parameter vector $\mathbf{b}^{(0)} = (0, 0, \ldots, 0)^T$;

3 **repeat**

4 Enforce local irradiance constraint:

$$\mathbf{n}^{(t)}(x, y) = \text{Exp}_{\mathbf{s}}\left(\arccos(I(x, y))\frac{\text{Log}_{\mathbf{s}}(\hat{\mathbf{n}}_{\mathbf{b}^{(t)}}(x, y))}{\|\text{Log}_{\mathbf{s}}(\hat{\mathbf{n}}_{\mathbf{b}^{(t)}}(x, y))\|}\right);$$

5 Enforce model-based integrability. Estimated surface gradients are given by $p(x, y) = \frac{n^{(t)}(x,y)_x}{n^{(t)}(x,y)_z}$ and $q(x, y) = \frac{n^{(t)}(x,y)_y}{n^{(t)}(x,y)_z}$. Calculate centred matrix of surface gradients $\mathbf{G}^{(t)}$ using Equation 5.21 and find $\mathbf{b}^{(t+1)}$ by solving:

$$\mathbf{b}^{(t+1)} = \left(\mathbf{\Psi}^T\mathbf{\Psi}\right)^{-1}\mathbf{\Psi}^T\mathbf{G}^{(t)};$$

6 Set iteration $t = t + 1$;

7 **until** $\sum_{x,y}\left[\arccos\left(\mathbf{n}^{(t)}(x, y) \cdot \mathbf{n}^{(t-1)}(x, y)\right)\right]^2 < \epsilon$;

surfaces recovered using the coupled model described in Section 5.2 and the model-based integration approach described in Section 5.3.

We then investigate the performance of the shape-from-shading algorithm described in Algorithm 5.1 which uses the model-based integrability constraint. We compare the accuracy of the recovered normals against using the principal geodesic shape-from-shading algorithm described in Algorithm 4.1. In addition, we compare the accuracy of the recovered surface height against that recovered by integrating the fields of surface normals recovered using the principal geodesic shape-from-shading algorithm.

We draw conclusions on which combination of methods appears to provide the best accuracy in terms of both the recovered field of surface normals and surface height function.

Figure 5.3: Comparison of average face surface and average field of surface normals.

5.4.1 Comparing Surface Normal and Height Models

We begin by comparing statistical models of facial shape constructed using surface normals and surface height. The models were trained on the 3DFS database as in chapters 3 and 4 and an additional 200 faces from the Max Planck database [15]. Although trained on the same data and modelling the same underlying structures, the two models are in fact quite different. In Figure 5.3 we show the average face surface on the left and the average field of surface normals on the right (we use the intrinsic mean definition of the average normal as in Section 4.1.1). Both surface representations are rendered with Lambertian reflectance and frontal illumination. It is clear that the two show clear structural differences. This is because they are based on minimising different quantities. The average field of surface normals aims to minimise the difference in terms of surface orientation to all training samples while the average face surface aims to minimise the difference in terms of surface height.

We now examine the efficiency of the two facial surface models and the coupled model in terms of the cumulative variance captured. In Figure 5.4 we plot cumulative variance against the number of eigenmodes retained. It is evident that fewer eigenmodes are required to capture variance in facial height than in facial needle maps. This is because the surface normal at each point has two degrees of freedom whereas the surface height value has only one. We retained 123 dimensions of the surface height model and 250

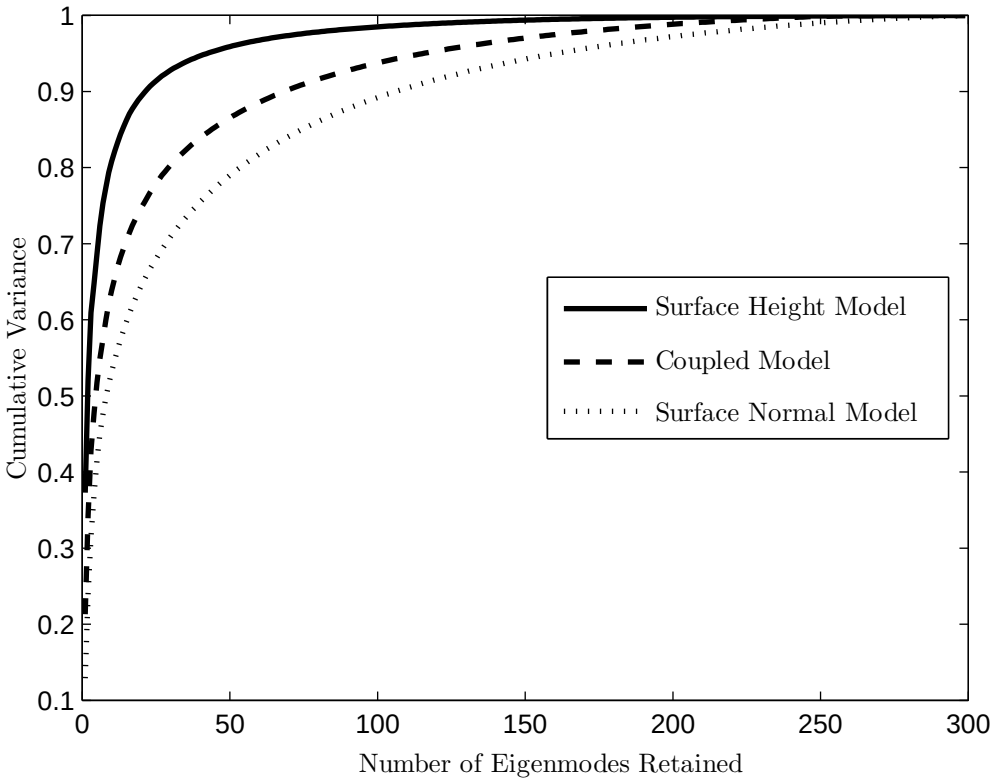


Figure 5.4: Plot of cumulative variance versus number of eigenmodes used for surface height model (solid line), coupled model (dashed line) and surface normal model (dotted line).

dimensions of the surface normal model (each accounting for 99% of the variance). As would be expected, the coupled model lies between the two and requires 207 dimensions to capture 99% of the variance.

5.4.2 Surface Integration

We now evaluate the two surface integration methods proposed in sections 5.2 and 5.3 on ground truth data. We use the generic surface integration technique of Frankot and Chellappa [59] to provide a comparison. We use a leave-one-out validation strategy, in which we train the statistical models with all but one of the facial surfaces. For this out-of-sample subject we compute the field of surface normals which we provide as input to

each of the surface integration techniques. This is done as follows:

- **Coupled Model**: We fit the surface normal model to the input field of surface normals by calculating the best fit parameter vector. We then recover the coupled model parameters that best reconstruct this parameter vector using Equation 5.11. The estimated surface height is then given by Equation 5.12.

- **Model-based Integration**: We find the surface height model parameters that produce a surface whose surface normals best reconstruct the input field of surface normals by solving Equation 5.23. The estimated surface height is then given by Equation 5.13.

For each of 200 subjects we calculate the root mean square (RMS) error between the estimated and ground truth surfaces for the two proposed approaches and the result of applying the Frankot and Chellappa [59] algorithm. In Figure 5.5 we show the RMS errors (in centimetres) for each subject ordered from the best- to worst-case. The results of the Frankot and Chellappa algorithm are shown by the blue dot-dashed line. The results of using the coupled model are shown by the green dotted line and the results using model-based integration are shown by the red dashed line. We also plot the error of projecting the input ground truth surface onto the surface height model. This is the minimum error with which the surface height model can represent the input surface. This represents a baseline error which neither of the proposed methods could outperform.

The first feature to note is that both of the proposed methods substantially outperform the Frankot and Chellappa algorithm. The average RMS error across all subjects for the Frankot and Chellappa method is 0.4749 cm. The coupled model and model-based integration approaches both offer improved performance with model-based integration outperforming the coupled model. The average RMS error across all subjects for the coupled model and model-based integration are 0.2244 cm and 0.1367 cm respectively. For

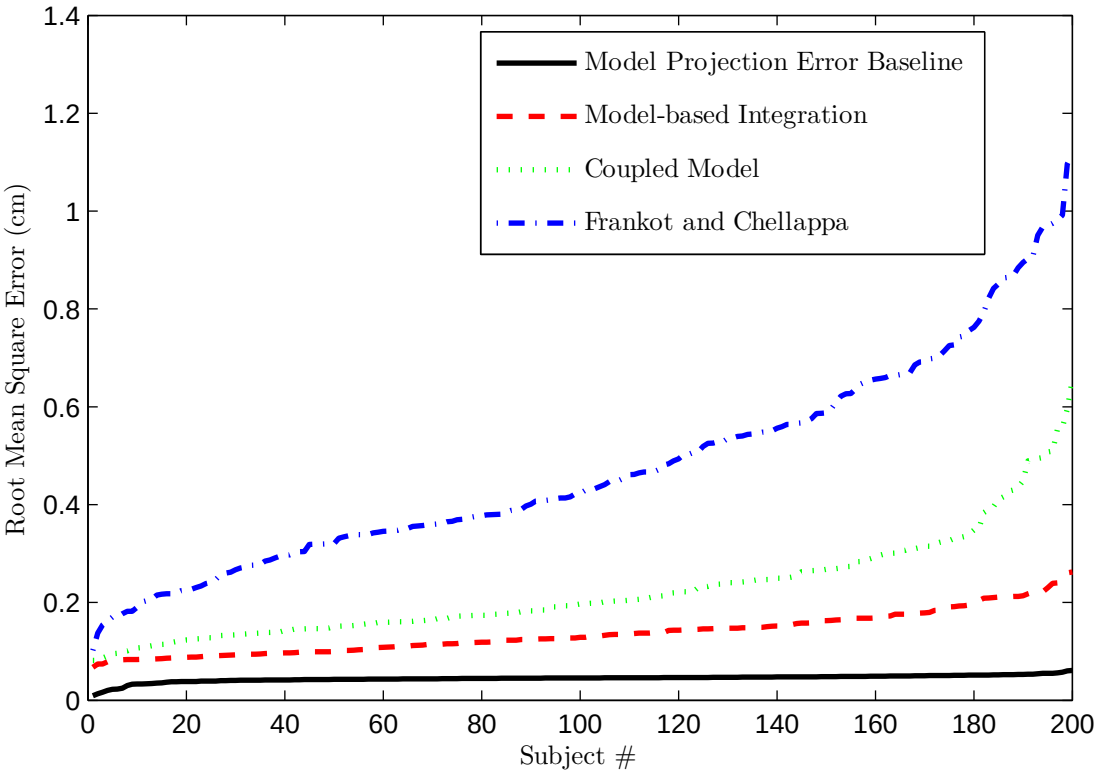


Figure 5.5: RMS surface height errors for 200 subjects using 3 surface integration techniques applied to ground truth normals.

comparison the average RMS error across all subjects for the baseline model projection is 0.0448 cm. In Figure 5.6 we present alternative visualisations of these errors. In the top row we show frequency histograms showing the distribution of RMS errors across the 200 subjects for each of the three methods. In the second row we show error maps in which the colour represents the average RMS error at that pixel (note that different scales are used for each plot, the units are centimetres). These plots highlight the spatial distribution of errors for the three methods.

In Figure 5.7 we show some specific examples from the above experiment. We compare contours of the profile view of the recovered surface (the central vertical scan-line) for 4 subjects. We show the ground truth contour as a solid black line and use the same line labelling convention as in Figure 5.5 for the estimated surfaces. The units of both

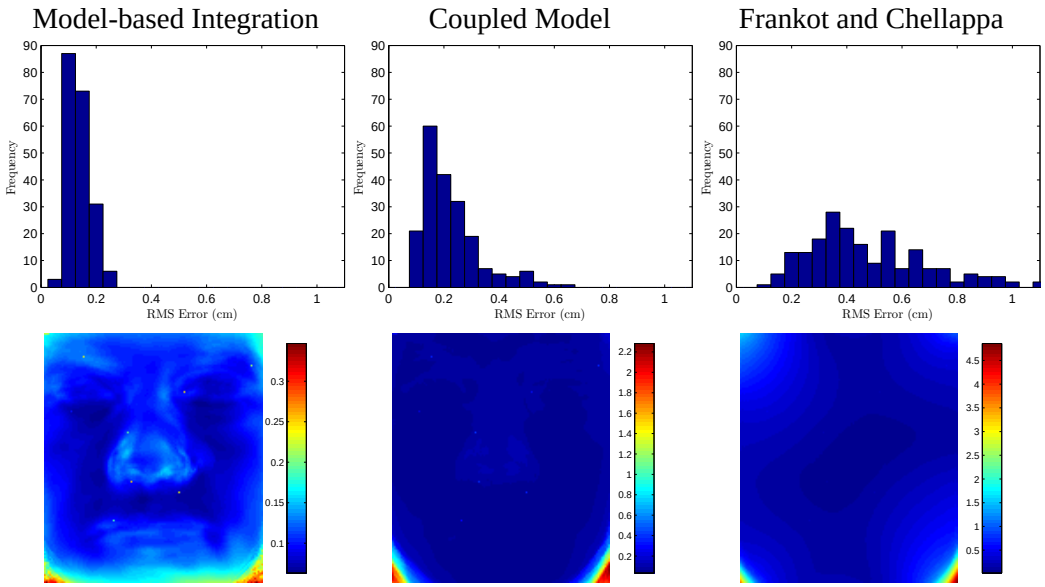


Figure 5.6: Surface integration results: frequency histograms of RMS surface height errors and error maps of average RMS surface height errors at each pixel.

axes are centimetres.

In the first example, all three methods perform well. The second example shows a case where model-based integration performs well, whilst the third example shows a case in which both model-based integration and the coupled model perform well. In these latter two cases the Frankot and Chellappa algorithm performs poorly. Here, the steepness of the surface at the base of the nose has caused the surface to be warped, leaving the lower half of the face estimated too high and the top half too low. This sensitivity to steeply inclined surface patches is a known weakness of the Frankot and Chellappa algorithm.

5.4.3 Shape-from-shading using Model-based Integrability

In this section we evaluate the use of model-based integrability as a constraint for facial shape-from-shading. We apply Algorithm 5.1 to ground truth face images and recover a field of surface normals and facial surface. We compare the recovered field of surface normals with those recovered using the principal geodesic shape-from-shading algorithm

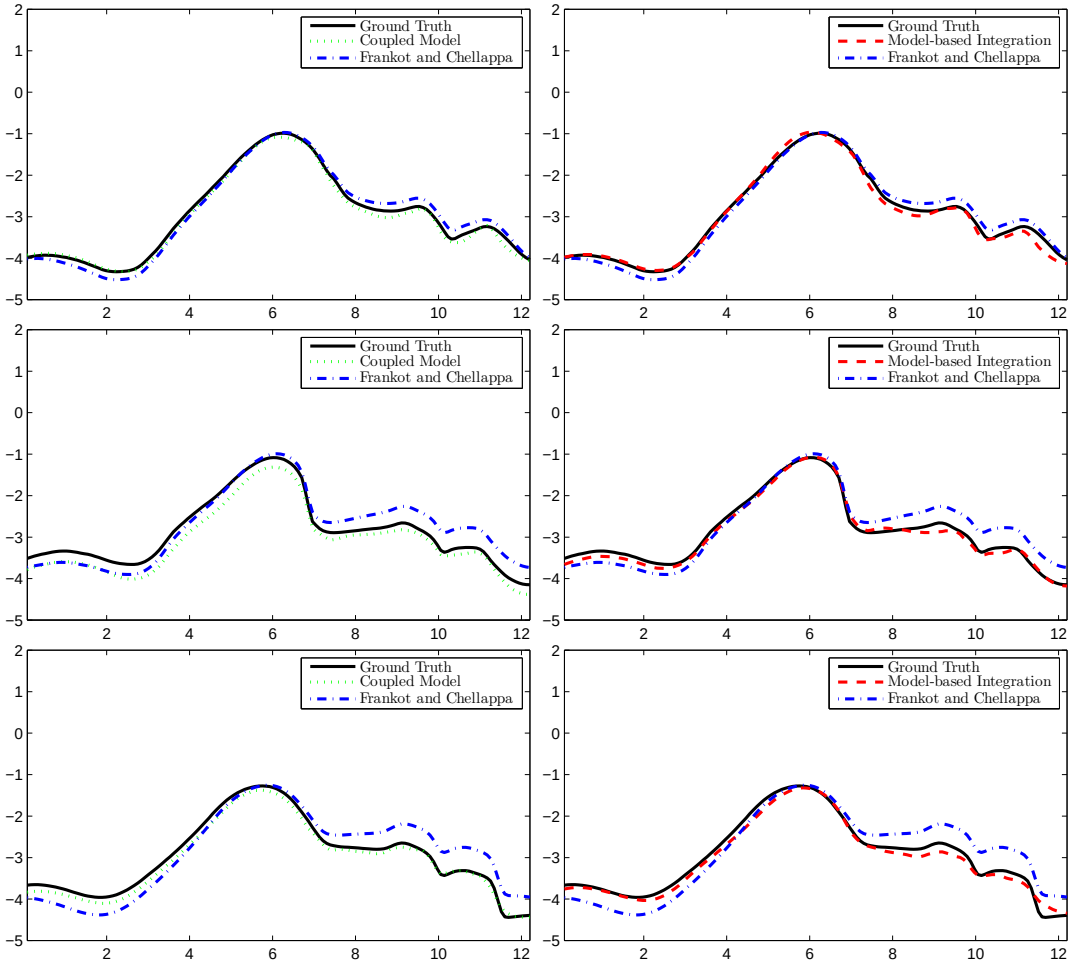


Figure 5.7: Evaluating surface integration results based on contour of profile view of recovered surface. For three faces we compare results using the coupled model (left column) with results using model-based integration (right column). We also show ground truth and the result of the Frankot and Chellappa algorithm for comparison.

described in Algorithm 4.1. We then compare the recovered surfaces with those obtained by applying a surface integration algorithm to the field of surface normals estimated using Algorithm 4.1. As above, we experiment with two surface integration algorithms: the generic method of Frankot and Chellappa and the method based on the coupled model described in Section 5.2.

As for the surface integration experiments in the previous section, we use the same 200 out-of-sample ground truth facial surfaces for our experiments. From these face surfaces we synthesise images using Lambertian reflectance and frontal illumination. To these images we apply the shape-from-shading algorithm which uses model-based integrability. This returns both a field of surface normals and surface height estimate. In Figure 5.8 we once again show the RMS errors (in centimetres) ordered from the best- to worst-case. This is shown by the red dashed line. For comparison, we show the result of applying a surface integration algorithm to the fields of surface normals recovered using the principal geodesic shape-from-shading algorithm. The results using the coupled model as a means of surface integration are shown by the green dotted line. The results using the Frankot and Chellappa algorithm are shown by the blue dot-dashed line. Note that this is effectively the method used in the previous two chapters, i.e. the use of a statistical surface normal model to recover a field of surface normals followed by the application of the Frankot and Chellappa algorithm to recover the surface height function.

From the plots it is evident that the shape-from-shading algorithm which uses model-based integrability recovers the most accurate surfaces (the average RMS error over all subjects was 0.1850 cm). Using the coupled model to recover a surface from the field of surface normals estimated using the principal geodesic shape-from-shading algorithm gives slightly worse results (average RMS error of 0.2438 cm), while the use of the Frankot and Chellappa algorithm gives significantly worse results (average RMS error of 0.4841 cm). In Figure 5.9 we show alternative visualisations of these errors. In the top

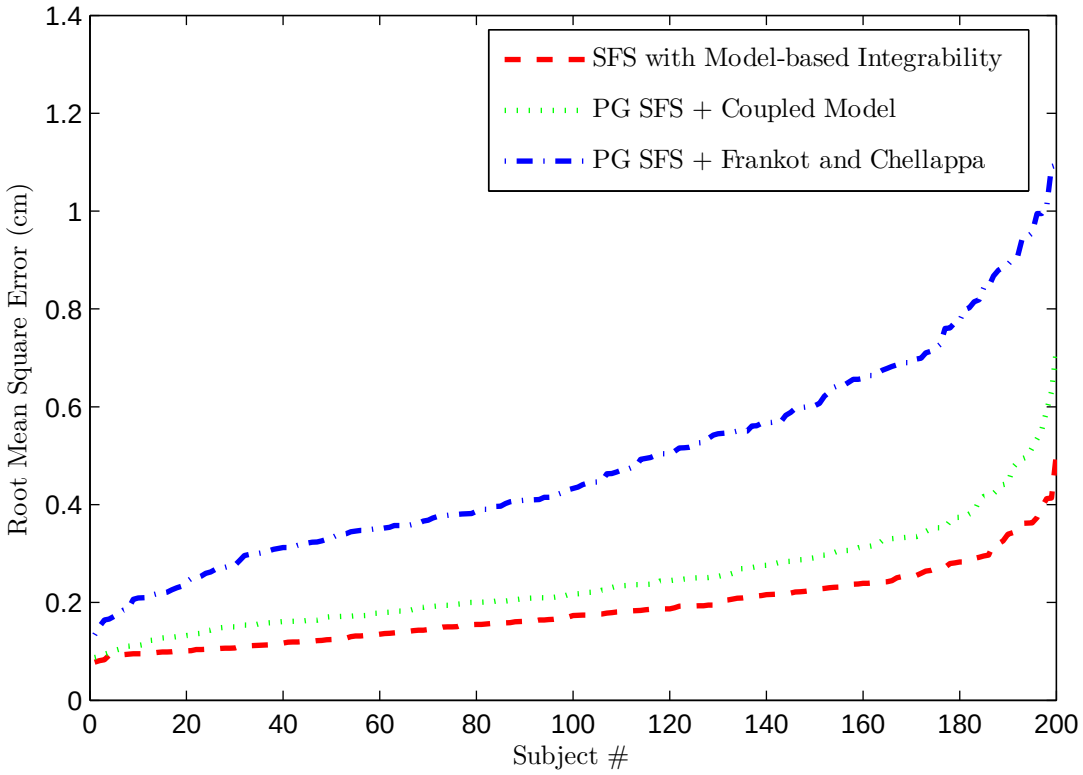


Figure 5.8: RMS surface height errors for 200 subjects using 3 methods for face surface recovery applied to images synthesised from ground truth face surfaces.

row we show frequency histograms of RMS errors, while in the second row we show the spatial distribution of errors.

An interesting question is whether this superior performance is purely related to the surface integration process or whether the surface normals recovered using the model-based integrability constraint are also more accurate. To answer this question we also calculated the average RMS error for the *surface normals* recovered using this method in comparison to the principal geodesic shape-from-shading algorithm. As would perhaps be expected, the surface normals recovered using the principal geodesic shape-from-shading algorithm are more accurate (0.1048 radians as opposed to 0.2136 radians using model-based integrability). This result is intuitive since we would expect the surface normal model is likely to produce more accurate results in the surface normal domain.

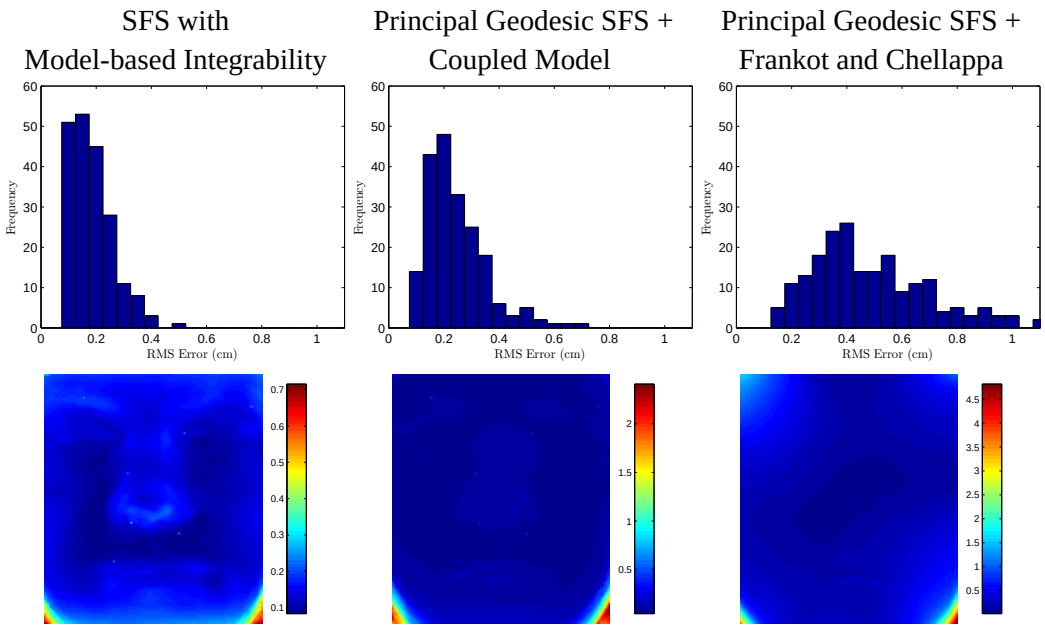


Figure 5.9: Surface recovery results: frequency histograms of RMS surface height errors and error maps of average RMS surface height errors at each pixel.

Finally in Figure 5.10 we show the same profile reconstructions as in Figure 5.7, except that this time they have been recovered from images rather than ground truth surface normals. The most striking feature is the degradation in performance of the Frankot and Chellappa algorithm when applied to imperfect normals. Note, for example, the convex/concave reversal in the mouth of the first subject. The results using the coupled model and shape-from-shading with model-based integrability remain good, with the model-based integrability results more closely following the ground truth profile.

5.4.4 Comparison

In Table 5.1 we summarise all of the results of our ground truth experiments. We show the average RMS errors in centimetres for height errors and radians for surface normal errors (the parenthesised values are the standard errors) for each experiment. The first results are for surface integration on ground truth surface normals. The second results are for

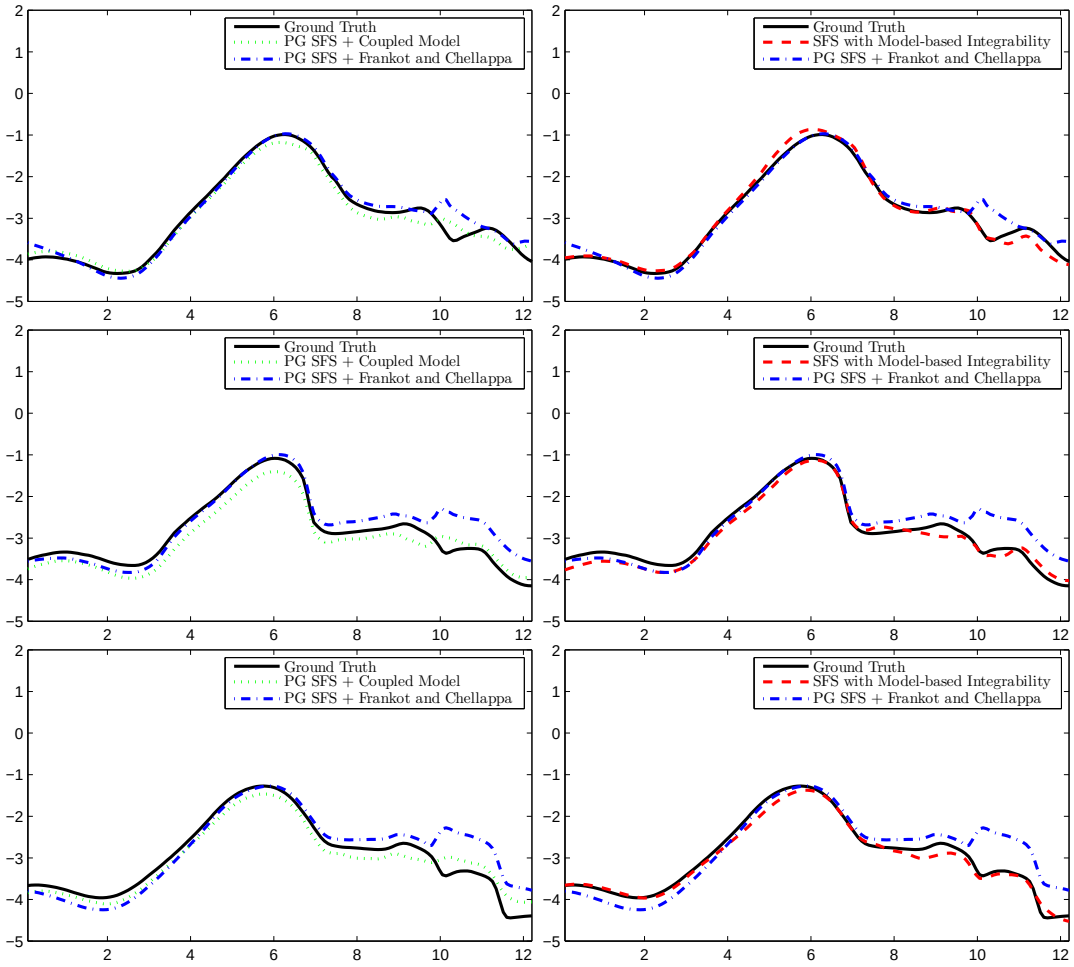


Figure 5.10: Evaluating surface recovery results based on contour of profile view of recovered surface. For three faces we compare results using the coupled model (left column) with results using model-based integration (right column). We also show ground truth and the result of the Frankot and Chellappa algorithm for comparison.

Comparison of Shape-from-shading and Surface Integration Methods		
Surface Integration from Ground Truth Normals		
Frankot and Chellappa	Coupled Model	Model-based Integration
0.4749cm ($S_E = 0.0149$cm)	0.2244cm ($S_E = 0.0075$cm)	$\mathbf{0.1367}$cm ($S_E = 0.0030$cm)
Surface Normal Recovery from One Image		
Principal Geodesic SFS	SFS with Model-based Integrability	
$\mathbf{0.1048}$rad ($S_E = 0.0048$rad)	0.2136rad ($S_E = 0.0038$rad)	
Surface Height Recovery from One Image		
Principal Geodesic SFS + Frankot and Chellappa	Principal Geodesic SFS + Coupled Model	SFS with Model-based Integrability
0.4841cm ($S_E = 0.0147$cm)	0.2438cm ($S_E = 0.0075$cm)	$\mathbf{0.1850}$cm ($S_E = 0.0053$cm)

Table 5.1: Comparing the accuracy of different combinations of shape-from-shading and surface integration methods. Values shown are average RMS errors over 200 samples. The units are centimetres for height errors and radians for surface normals errors. The parenthesised values are the standard errors of the RMS errors.

recovering fields of surface normals from images and the third results are for recovering facial surfaces from images. The table highlights that both the methods developed in this chapter offer an improvement over using the generic surface integration method of Frankot and Chellappa. Moreover, they show that using a statistical surface height model in the context of our model-based integrability constraint offers improved performance over the statistical surface normal models developed in the previous two chapters. This is true both in terms of the accuracy of the recovered field of surface normals and the surface height.

5.5 Conclusions

In this chapter we have extended our consideration of the recovery of facial shape from single images to include the recovery of the surface height function. In the previous two chapters, we used statistical models of surface normal direction to constrain the shape-

from-shading process. We motivated the work in this chapter by arguing that the surface integration process should also be subject to a similar statistical constraint. For this reason we developed two surface integration methods which use a statistical model of surface height.

The first aims to learn the relationship between the parameters of a statistical surface normal model and surface height model based on a linear analysis of the correlation matrix between the two sets of parameters. To perform surface integration, we search for the parameters of this coupled model that best reconstruct the surface normal model parameters of the input field of surface normals.

The second method was based on a more thorough analysis of the relationship between the parameters of the surface height model and the resulting field of surface normals. We derive a method for recovering the surface height parameters directly from a field of surface normals, a process we term 'model-based integration'. In addition, we note that this process can form a constraint on the field of surface normals to be used during shape-from-shading. The resulting shape-from-shading algorithm has a number of advantages over the methods developed in this and the previous two chapters. Firstly, the field of surface normals is subject to a stricter constraint than previously since it must be both integrable and lie within the span of the model. Secondly, the algorithm provides both the surface height and normals as output. Thirdly, model-based integration is accomplished using matrix multiplications only whereas fitting the coupled model requires the solution of a minimisation problem. Finally, the depth model is more efficient in terms of the number of dimensions that must be retained to preserve a given amount of variance. Further, half as many values are required to describe each dimension since only the surface height is being modeled as opposed to the two components of surface orientation.

Our experimental results show that for the purposes of both surface integration and surface recovery using shape-from-shading, the use of the model-based integrability con-

straint outperforms the coupled model and the methods developed in the previous chapters when applied to ground truth data.

Chapter 6

Estimating Skin Reflectance Properties

Often a silent face has voice and words.

- Ovid [119]

In this chapter we address the problem of estimating the reflectance properties of skin from single images. In the work presented in the preceding three chapters, we have made the assumption that skin reflects light according to Lambert's law [97]. In other words, we have assumed that skin is an ideal diffuse reflector and that incoming light is scattered equally in all directions. In fact, the reflectance properties of skin are complex and far from Lambertian in most cases [109].

In Chapter 2, we discussed a variety of reflectance models which may be used to describe skin reflectance. In Table 6.1 we draw from this review a list of features which would be desirable in a skin reflectance model with the application of facial shape-from-shading in mind. We note that no existing models of reflectance satisfy all of these criteria. The reason for this is that as the models incorporate more complex effects and parameters, they inevitably lose their invertability and computational efficiency. Invertability is an important property in order to perform shape-from-shading where we wish to be able to infer some of the reflectance geometry from the measured intensity values.

Skin Reflectance Model Wishlist	
Invertible	It would be beneficial if we can estimate from the measured intensity some information regarding the reflectance geometry, e.g. using the Lambertian model we can recover the incident angle.
Biophysically meaningful parameters	If the parameters have a clear biophysical meaning (for example melanin/haemoglobin concentration, skin layer thicknesses etc), we can use those that we predict will remain constant as part of a biometric. This would also provide intuitive parameters for face image editing.
Captures important features of skin reflectance	These range from simple effects such as "shininess," to more complex phenomena such as off-specular reflectance, asperity scattering and subsurface scattering.
Computationally simple	Particularly important if the model is to be evaluated within an iterative shape-from-shading framework.
Accurate over all possible skin types	Since we wish to apply the model to arbitrary faces, it must not have a bias to model one particular skin type more accurately than another.

Table 6.1: A wish list of desirable properties of a skin reflectance model for use in computer vision tasks.

In this chapter, we do not make use of an existing model of skin reflectance. Instead, we aim to demonstrate that we can empirically measure skin reflectance properties from a single image within a shape-from-shading framework.

We provide a method for estimating facial shape and reflectance properties from a single image, which allows the surface reflectance function and albedo to vary arbitrarily. The number of unknowns here is large and the problem is highly under-constrained. The reason for this is that from each single intensity value we wish to estimate three local values. These are the two components of the surface normal and the local albedo. Additionally we wish to make a global estimate of the surface reflectance function. We show that it is sufficient to enforce the model-based integrability constraint developed in the previous chapter on the field of surface normals. In addition we employ standard photometric constraints.

The idea underpinning the approach is to iteratively apply two processes: estimation

of surface reflectance properties and surface shape estimation:

1. We use the estimated surface shape to calculate the reflectance geometry at each pixel in the input image. This provides a scattered and potentially noisy sampling of the global surface reflectance function, to which we fit a smooth and monotonic curve. This is an estimate of a slice of the surface radiance function.

2. With this function to hand, we improve our estimate of the surface shape. This consists of two steps. First, we locally enforce consistency with the global estimate of the reflectance function. This ensures the estimated shape and reflectance function will exactly recreate the input image. Second, we enforce a global statistical constraint on the field of surface normals. Model-based integrability ensures that the surface normals are both integrable and correspond to a surface which lies within the span of a statistical surface height model.

By iterating and interleaving these two processes we arrive at stable estimates of both the surface shape and reflectance. We extend the method to real world images and, by calling on additional photometric constraints, we may also simultaneously make an estimate of the albedo at each pixel location. The resulting algorithm provides a means to perform reliable non-Lambertian shape-from-shading on face images using a statistical constraint on the surface height function.

Finally, we turn our attention to colour images of faces. We apply the same process as for grayscale images to each colour channel independently. However, there is an additional constraint here which we may exploit. The underlying surface shape is fixed and hence may not differ between colour channels. For this reason, we take as our estimate of the surface shape at each iteration the average of the three fields of surface normals recovered from each of the three colour channels. As an additional application, we demonstrate how our method can be used to solve the problem of facial colour constancy. Our technique provides a means for estimating the colour and strength of the illuminant. We show

how the estimated surface shape and reflectance properties can be used to render an image in which the surface is illuminated by a white light, effectively removing the effect of the original coloured light source.

6.1 Skin Reflectance

Light interaction with skin is a complex process involving reflection at the air-skin boundary following Fresnel theory and subsurface scattering within the layers of the skin [151]. The process has been studied from several perspectives by the biology [163], graphics [72], vision [44], psychophysics [96] and medical imaging communities [29]. Skin reflectance is controlled by biological parameters such as melanin and haemoglobin concentration and the thickness of the component skin layers [151]. The most comprehensive models have been developed by the graphics community, which explicitly account for subsurface light transport [88].

However, for a wide range of tasks the complexities of subsurface scattering may be discounted. Hence, the assumption is made that light enters and leaves the surface at the same point. Under this assumption skin reflectance properties can be explained by a bidirectional reflectance distribution function (BRDF). Attempts have been made to empirically measure the BRDF of skin [42, 109]. However, this exhaustive process requires either a stationary, flat sample (difficult to achieve with an *in vivo* sample) or accurate knowledge of the 3D shape. For this reason, the computer vision community has tended to employ simple, analytical reflectance models that can capture a range of surface BRDFs. In particular, the Lambertian model has been widely used [8, 47, 64, 164, 168]. Examples of the use of more complex models of skin reflectance within a vision context are rare. The examples being Blanz and Vetter [14] who used the Phong [128] model in their analysis-by-synthesis framework, while Georghiades [62] incorporated the Torrance

and Sparrow [146] model into an extended uncalibrated photometric stereo algorithm. In this chapter, we do not call on an existing reflectance model and seek to estimate its parameters. Instead, we aim to show that it is possible to make empirical reflectance measurements using shape-from-shading.

As discussed in Chapter 3, the BRDF describes the ratio of the emitted surface radiance to the incident irradiance. If a surface illuminated by irradiance $f_i(\theta_i, \alpha_i)$ from a differential region of solid angle $d\omega$ was to emit radiance $f_o(\theta_o, \alpha_o)$, its BRDF would be:

$$\rho(\theta_i, \alpha_i, \theta_o, \alpha_o) = \frac{f_o(\theta_o, \alpha_o)}{f_i(\theta_i, \alpha_i)\cos(\theta_i)d\omega}. \tag{6.1}$$

In the context of facial shape-from-shading, the focus is on single images and we restrict ourselves to situations in which only a single source of illumination is present. In this scenario, only a particular subset (or 'slice') of the BRDF is relevant. In such a setup, the light source and viewing direction are fixed relative to each other with respect to each point on the surface. Hence, the slice of the BRDF that we are interested in has only two degrees of freedom.

For this purpose, it is more convenient to use a viewer centred coordinate system, in which the spherical coordinates of the view vector are $(0, 0)$. The light source vector may be written in spherical coordinates as (θ_L, α_L). The reduced slice of the BRDF, $\hat{\rho}$, which is observed in a single image varies over the range of surface normal directions (θ, α) (also in viewer centred spherical coordinates) present in the image:

$$\hat{\rho}(\theta, \alpha) = \rho\left(\theta - \theta_L, \alpha - \alpha_L, \theta, \alpha\right). \tag{6.2}$$

We further assume that illumination is provided by a point light source situated at infinity and that its direction is fixed across our slice of the BRDF measurement. We may therefore set the irradiance and solid angle terms to a constant, $k = f_i(\theta_i, \alpha_i)d\omega$, and write our

reduced reflectance function as:

$$\hat{\rho}(\theta, \alpha) = \frac{1}{k} \frac{f_o(\theta, \alpha)}{\cos(\theta)}. \tag{6.3}$$

In Section 3.1.1, we showed that the image irradiance at the sensor element, E, is proportional to the outgoing radiance from the surface. If we subsume the parameters related to the optical properties of the camera into a constant: $\tau = \frac{\pi}{4}\left(\frac{D}{z'}\right)^2$, we may relate the irradiance at the sensor to the BRDF directly:

$$E = \tau f_o(\theta, \alpha) = \tau k \hat{\rho}(\theta, \alpha). \tag{6.4}$$

As discussed in Section 3.1.1 the response function of the camera, g_{cam}, determines the relationship between the reported pixel brightness and the irradiance incident on the CCD. Assuming this function is linear, the image intensity is directly proportional to the radiance function:

$$I \propto E = \tau f_o(\theta, \alpha). \tag{6.5}$$

The relationship between image intensity and the surface reflectance function is now explicit:

$$I \propto \hat{\rho}(\theta, \alpha) \cos(\theta). \tag{6.6}$$

Therefore, estimating the radiance function from the measured image intensities provides a means to estimate a slice of the BRDF (upto a scaling by constants related to the optical properties of the camera). Repeating the process on multiple images with multiple light source directions would provide a means to estimate a full BRDF.

6.1.1 Radiance Function Estimation

If the surface normal directions at each point in the image are known then the surface radiance function $f_o(\theta, \alpha)$ can be estimated directly. If I is the image brightness at the surface location where the surface normal has elevation and azimuth angles θ and α, then $f_o(\theta, \alpha) = \frac{I}{\tau}$.

In practice, the surface normals at each pixel are unknown and hence the radiance function cannot be estimated directly. Instead, at each iteration of the algorithm we have an estimate of the surface shape provided by shape-from-shading. Our aim is therefore to make an estimate of the surface radiance function from the noisy surface normal estimates and the input image intensities.

It is useful to couch the radiance function in terms of the Gauss map of the facial surface onto a unit sphere. If we consider the face as an orientable surface $F \in \mathbb{R}^3$, the Gauss map $G : F \mapsto S^2$ maps points on the surface F onto locations on the unit sphere S^2 which have identical surface normal directions. Points on the surface with identical normal directions have identical brightness values. By using the surface normal directions to map brightness values from the image onto a unit sphere, we arrive at a representation of the radiance function. The normalised polar distribution of brightness values on the unit sphere S^2 is the radiance function $f_o(\theta, \alpha)$ for the surface. Intuitively, the image of the Gauss sphere will be an image of a sphere exhibiting the same reflectance properties as the surface under study.

6.1.2 The simplest case: Retro-reflection

If the viewer and light source directions are coincident, i.e. $(\theta_L, \alpha_L) = (0, 0)$ (a case known as *retro-reflection*), the situation is simplified further. Assuming the radiance func-

tion is isotropic:

$$\partial_\alpha f_o(\theta, \alpha) = 0, \tag{6.7}$$

then circles of latitude on the unit sphere will have constant brightness. Further, assuming the radiance function decreases monotonically with θ, brightness on the unit sphere will fall away from a maximum value at the visible pole to zero at the equator. In this case, the latitude itself will be uniquely determined by a function of the image brightness. Hence, the radiance function can be expressed in terms of a reduced function in one variable:

$$f_o(\theta, \alpha) = g(\theta). \tag{6.8}$$

Since we require g to be smooth and monotonic, we can calculate the inverse and for a given brightness value I, the value of θ can be recovered using the fact that $\theta = g^{-1}(I)$.

We now show how to make an estimate of the function g from a noisy estimate of the field of surface normals.

Fitting to a noisy radiance function sampling

Suppose that we have to hand an estimate of the surface height function of the face, $z(x, y)$, and the corresponding field of surface normals $\mathbf{n}(x, y)$. If the scene is illuminated by a distant point light source from direction $\mathbf{s}$, the estimated incident angle for the current surface normal estimate is given by: $\theta(x, y) = \arccos(\mathbf{n}(x, y) \cdot \mathbf{s})$. From the input image intensity at the corresponding pixel, we know the corresponding value of the surface radiance function: $g(\theta(x, y)) = I(x, y)$. Tabulating these two values against each other provides a dense but noisy sampling of the function g. By fitting a function to this scattered data we derive an estimate of the surface radiance function. We assume this function is smooth and monotonic. We use the following process in order to fit a function to the scattered data:

1. The incident angle θ lies in the closed interval $[0, \frac{\pi}{2}]$. We bin the values of $g(\theta(x, y))$ into n bins of width $w = \frac{\pi}{2n}$. Let $Q_i = \{(x, y) \mid (i-1)w \leq I(x, y) < iw\}$ be the set of pixels where $\theta(x, y)$ falls into the ith bin. For each bin we find the median value of g:

$$h(i) = \underset{(x,y)\in Q_i}{\text{median}}\ I(x, y). \tag{6.9}$$

   The distribution of median values of $g(\theta(x, y))$ is stored as a vector

$$\vec{h} = (h(1), \ldots, h(n))^T. \tag{6.10}$$

2. Bins in which no data points lie within the associated range of values of θ will be empty. We perform piecewise linear interpolation of the adjacent non-zero elements in order to fill in these missing values of $\vec{h}$.

3. Due to noisy surface normal estimates, the estimate of the radiance function g is likely to be noisy. We therefore apply a 1D Gaussian smoothing operator to $\vec{h}$.

4. Finally, we enforce monotonicity on g. We discard values of $\vec{h}_i$ where $\vec{h}_i \not\geq \vec{h}_{i-1}$. The remaining values are both smooth and monotonic. Our final estimate of the radiance function g is given by fitting a piecewise cubic hermite interpolating polynomial (PCHIP) to these remaining values. We use a PCHIP function in order to preserve the smoothness and monotonicity of the sampled points.

In Figure 6.1(a)-(d) we demonstrate the effect of each of these steps on an artificial sample of radiance data. The scattered radiance measurements are shown in black, while the estimate of the radiance function is plotted as a solid blue line. Values of $g(\theta)$ are plotted on the y-axis as a function of θ on the x-axis. In Figure 6.1(a) the median bin values are shown. In (b), linear interpolation has been applied to fill-in missing values, (c) shows the result of the 1D smoothing, while in (d) we show our final estimate with monotonicity enforced.

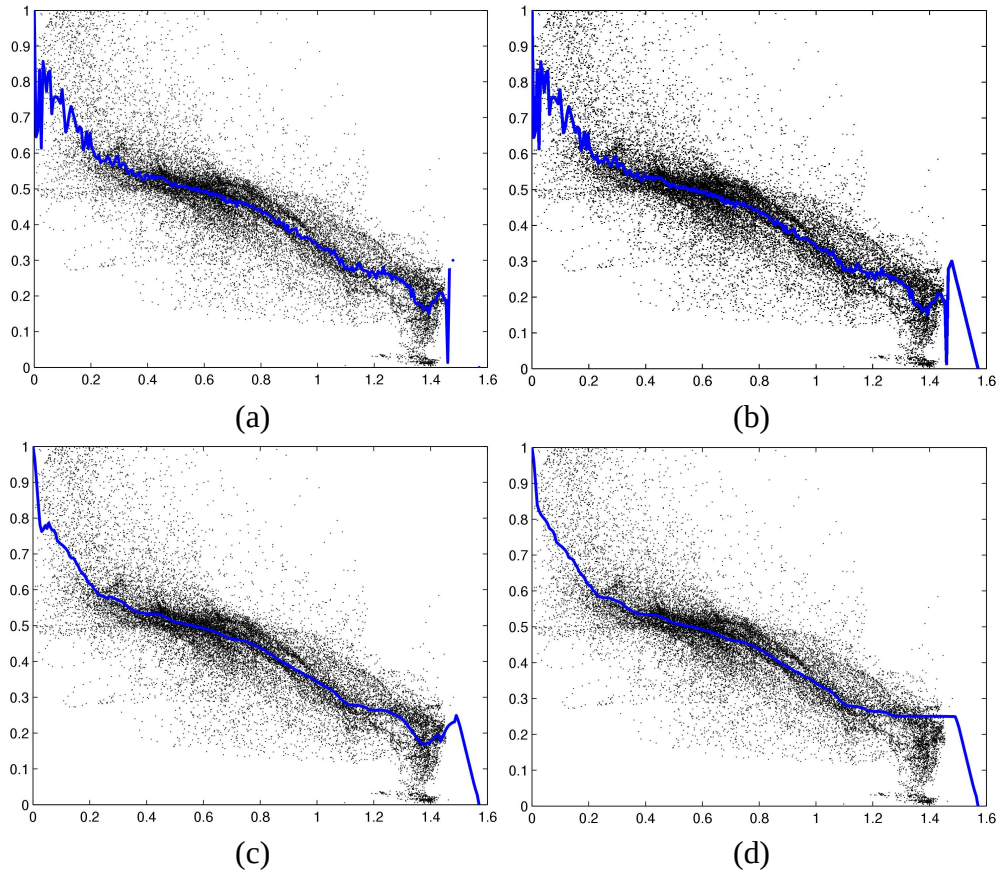


Figure 6.1: Process of fitting BRDF function to scattered reflectance measurements: (a) Calculate median bin values (b) Linear interpolation (c) Smoothing (d) Enforce monotonicity and fit PCHIP for final function.

6.1.3 Enforcing Consistency with the Surface Radiance Function

Our aim in this chapter is to estimate both the surface radiance function and surface shape from a single intensity image. In this case the surface normal directions at each point on the surface are unknown. Estimating surface normal directions from intensity measurements is the shape-from-shading problem, as presented in the previous three chapters. In this chapter we exploit the model-based integrability constraint developed in the previous chapter. Satisfaction of this constraint ensures not only integrability of the field of surface normals, but also that the corresponding surface height function lies within the span of a statistical model learnt from example facial surfaces. We combine this with an irradiance

constraint which locally enforces consistency between the predicted and measured image intensity.

With our estimate of the surface radiance function to hand, we can enforce a local constraint on the current estimate of the field of surface normals, by enforcing consistency with the input image intensities. The incident angle implied by the current estimate of the surface radiance function and the image intensity at a pixel (x, y) is given by $\theta(x, y) = g^{-1}(I(x, y))$. Suppose that the surface height parameter vector estimated using the method described in the previous chapter is $\mathbf{b}$, then we update each normal $\hat{\mathbf{n}}_{\mathbf{b}}(x, y)$ to ensure consistency with this constraint.

We again turn to the apparatus of the Log and Exponential maps to simplify the notation involved. The normal $\hat{\mathbf{n}}_{\mathbf{b}}(x, y)$ can be written in terms of a vector on the tangent plane to the light source direction as:

$$v(x, y) = \text{Log}_{\mathbf{s}}(\hat{\mathbf{n}}_{\mathbf{b}}(x, y)). \tag{6.11}$$

Straight lines through the centre of this tangent plane correspond to geodesics on the sphere. Therefore, surface normals that lie in the direction given by the unit vector:

$$\hat{v}(x, y) = \frac{\text{Log}_{\mathbf{s}}(\hat{\mathbf{n}}_{\mathbf{b}}(x, y))}{\|\text{Log}_{\mathbf{s}}(\hat{\mathbf{n}}_{\mathbf{b}}(x, y))\|}, \tag{6.12}$$

will provide the closest normal to $\hat{\mathbf{n}}_{\mathbf{b}}(x, y)$ for a desired value of $\theta(x, y)$.

In order to ensure that the constraint $\theta(x, y) = g^{-1}(I(x, y))$ is satisfied, we update $\hat{\mathbf{n}}_{\mathbf{b}}(x, y)$ as follows:

$$\mathbf{n}(x, y) = \text{Exp}_{\mathbf{s}}\left(d(x, y)\hat{v}(x, y)\right), \tag{6.13}$$

where d is the distance we move in the direction of $v(x, y)$. The distance $d(x, y)$ is simply given by $d(x, y) = g^{-1}(I(x, y))$.

The updated normals $\mathbf{n}(x,y)$ satisfy the constraint that $g(\arccos(\mathbf{n}(x,y)\cdot\mathbf{s})) = \mathbf{I}(\mathbf{x},\mathbf{y})$. In other words for a given normal, the intensity predicted by the current estimate of the radiance function is equal to the measured input intensity.

6.2 Combining Radiance Function and Surface Shape Estimation

We now show how the process of radiance function estimation described in Section 6.1.1 can be combined with a shape-from-shading algorithm which employs the two constraints described in Section 6.1.3. We interleave the two processes in an iterative framework. Intuitively, our aim is to find the field of surface normals and surface radiance function which best recreate the input image, subject to the surface normals obeying the model-based integrability constraint. We commence from an initialisation in which we assume average facial shape, i.e. we set $\mathbf{b} = 0$ and hence, $z_{\mathbf{b}}(x,y) = \hat{z}(x,y)$. We have found that, although coarse, this initialisation allows us to make a reliable initial estimate of the surface radiance function. The algorithm takes as input the light source direction vector, $\mathbf{s}$, and the image intensities, $I(x,y)$, and is summarised in Algorithm 6.1.

Upon convergence we output our estimate of the surface radiance function g and our surface shape estimates. This is in the form of the surface $z_{\mathbf{b}^{(\text{final})}}(x,y)$ which lies in the span of the surface height model and the normal map $\mathbf{n}^{(\text{final})}(x,y)$ which satisfies the local irradiance constraint.

It is important to note that the field of surface normals $\mathbf{n}^{(\text{final})}(x,y)$ will not correspond exactly to the surface normals of the surface $z_{\mathbf{b}^{(\text{final})}}(x,y)$. However, since this field of surface normals satisfies the irradiance constraint, they are significantly more stable for the purposes of rendering novel views and our experiments suggest they are more accurate. Further, since they are not constrained by the statistical model they are free to accurately

Algorithm 6.1: Iterative estimation of facial shape and surface radiance function.

Input: Light source $\mathbf{s}$, image intensities $I(x, y)$, surface gradients of statistical surface height model $\mathbf{\Psi}$ and average surface gradients $\hat{p}(x, y)$ and $\hat{p}(x, y)$

Output: Estimated surface normal map $\mathbf{n}(x, y)$, surface height $z_{\mathbf{b}}(x, y)$ and surface radiance function g

Set iteration $t = 0$;

Initialise parameter vector $\mathbf{b}^{(0)} = (0, 0, \ldots, 0)^T$;

repeat

  Calculate the incident angle for each pixel: $\theta^{(t)}(x, y) = \arccos(\hat{\mathbf{n}}_{\mathbf{b}^{(t)}}(x, y) \cdot \mathbf{s})$;

  From the estimated incident angles $\theta^{(t)}(x, y)$ and predicted radiance values $g(\theta^{(t)}(x, y)) = I(x, y)$, estimate the surface radiance function g using the method described in Section 6.1.1;

  Enforce local irradiance constraint:

$$\mathbf{n}^{(t)}(x, y) = \text{Exp}_{\mathbf{s}} \left(g^{-1}(I(x, y)) \frac{\text{Log}_{\mathbf{s}}(\hat{\mathbf{n}}_{\mathbf{b}^{(t)}}(x, y))}{\|\text{Log}_{\mathbf{s}}(\hat{\mathbf{n}}_{\mathbf{b}^{(t)}}(x, y))\|} \right);$$

  Enforce model-based integrability. Estimated surface gradients are given by $p(x, y) = \frac{n^{(t)}(x,y)_x}{n^{(t)}(x,y)_z}$ and $q(x, y) = \frac{n^{(t)}(x,y)_y}{n^{(t)}(x,y)_z}$. Calculate centred matrix of surface gradients $\mathbf{G}^{(t)}$ using Equation 5.21 and find $\mathbf{b}^{(t+1)}$ by solving:

$$\mathbf{b}^{(t+1)} = \left(\mathbf{\Psi}^T\mathbf{\Psi}\right)^{-1} \mathbf{\Psi}^T\mathbf{G}^{(t)};$$

  Set iteration $t = t + 1$;

until $\sum_{x,y} \left[\arccos\left(\mathbf{n}^{(t)}(x, y) \cdot \mathbf{n}^{(t-1)}(x, y)\right)\right]^2 < \epsilon$;

capture local or atypical shape features.

6.2.1 Spatially Varying Reflectance Properties

The analysis presented above assumes that surface reflectance properties are homogenous across the surface. Clearly, for real world images this is not the case due to spatial variance in pigmentation, blood flow and facial hair distribution. We now show how to relax this assumption by introducing an additional albedo parameter which we allow to vary

arbitrarily over the surface.

Albedo is the intrinsic reflectivity of the surface and represents the ratio of reflected to absorbed light. We use albedo to account for the absorption of light by melanin and haemoglobin within the skin and facial hair on the skin's surface. Although this formulation is simplistic, it allows us to recover useful surface shape and reflectance estimates from real world images.

We introduce a 'combined' albedo parameter, which acts in a multiplicative manner on the image irradiance equation as follows:

$$I(x, y) = \rho_c(x, y) g(\theta(x, y)), \tag{6.14}$$

where $\rho_c(x, y)$ is the combined albedo at pixel (x, y). We use the term 'combined' since, were the radiance function considered to be comprised of a number of components (e.g. diffuse and specular), then the albedo acts on this combined function. The algorithm given above requires slight modification in order to account for this albedo term. At the radiance function estimation step, the incident angles are calculated in the same way. However, the value of g against which we tabulate the incident angles now includes the albedo: $g(\theta_i(x, y)) = \frac{I(x,y)}{\rho_c(x,y)}$. It is to this data that we fit the function g. Similarly, in order to enforce the local irradiance constraint we update the distance term in Equation 6.13 to $d(x, y) = g^{-1}\left(\frac{I(x,y)}{\rho_c(x,y)}\right)$, such that the incident angle is consistent with both the radiance function and local albedo.

Finally, we add an additional albedo estimation step after enforcing model-based integrability on the field of surface normals. This is done by simply rearranging Equation 6.14. However, there are two additional constraints we may impose here. First, the albedo may not be greater than 1 (a surface cannot reflect more light than was incident upon it) and it may not be less than the intensity at the same point (this would mean the predicted

intensity was less than the measured intensity regardless of the normal direction). The albedo is therefore estimated using:

$$\rho_c(x,y) = \max\left(\min\left(\frac{I(x,y)}{g(\arccos(\hat{\mathbf{n}}_\mathbf{b}(x,y)\cdot\mathbf{s}))},1\right), I(x,y)\right). \tag{6.15}$$

We initialise the albedo at every pixel to be the maximum intensity value across the image: $\rho_c^{(0)}(x,y) = \max_{x,y} I(x,y)$. The updated algorithm in which we account for varying albedo is described in Algorithm 6.2. The differences between this algorithm and Algorithm 6.1 which assumes constant albedo are as follows. In step 5, the estimated values of the radiance function take account of the current estimate of the local albedo. Similarly in step 6, the magnitude of the update term takes into account the reduction in intensity caused by the current estimate of the albedo. Finally, step 8 is added which calculates the albedo values based on the current shape and radiance function estimates and the input image.

6.2.2 Novel Pose Synthesis

The slice of the BRDF that we have measured is that in which the viewer and light source directions are coincident. Using this information we can synthesise physically meaningful images by rotating the estimated surface, but keeping the viewer and light source direction constant. The incident angle at each pixel is given by the angle between the light source and the local surface normal after rotation to the new position. We can use this angle, the combined albedo and radiance function to calculate the intensity at each point. In our experimental section we show that this approach results in synthesised images that are stable under a large range of novel viewpoints. However, should we wish to synthesise images in which the illumination and viewing direction are no longer coincident, we must rely on a model to extrapolate the reflectance properties. We describe this approach in the

Algorithm 6.2: Iterative estimation of facial shape, the surface radiance function and spatially varying albedo term.

Input: Light source $\mathbf{s}$, image intensities $I(x, y)$, surface gradients of statistical surface height model $\mathbf{\Psi}$ and average surface gradients $\hat{p}(x, y)$ and $\hat{p}(x, y)$

Output: Estimated surface normal map $\mathbf{n}(x, y)$, surface height $z_{\mathbf{b}}(x, y)$, surface radiance function g and combined albedo $\rho_c(x, y)$

Set iteration $t = 0$;

Initialise parameter vector $\mathbf{b}^{(0)} = (0, 0, \ldots, 0)^T$ and albedo $\rho_c^{(0)}(x, y) = \max_{x,y} I(x, y)$;

repeat

Calculate the incident angle for each pixel: $\theta^{(t)}(x, y) = \arccos(\hat{\mathbf{n}}_{\mathbf{b}^{(t)}}(x, y) \cdot \mathbf{s})$;

From the estimated incident angles $\theta^{(t)}(x, y)$ and predicted radiance values $g(\theta^{(t)}(x, y)) = \frac{I(x,y)}{\rho_c^{(t)}(x,y)}$, estimate the surface radiance function g using the method described in Section 6.1.1;

Enforce local irradiance constraint:

$$\mathbf{n}^{(t)}(x, y) = \text{Exp}_{\mathbf{s}} \left(g^{-1} \left(\frac{I(x, y)}{\rho_c^{(t)}(x, y)} \right) \frac{\text{Log}_{\mathbf{s}}(\hat{\mathbf{n}}_{\mathbf{b}^{(t)}}(x, y))}{\|\text{Log}_{\mathbf{s}}(\hat{\mathbf{n}}_{\mathbf{b}^{(t)}}(x, y))\|} \right);$$

Enforce model-based integrability. Estimated surface gradients are given by $p(x, y) = \frac{n^{(t)}(x,y)_x}{n^{(t)}(x,y)_z}$ and $q(x, y) = \frac{n^{(t)}(x,y)_y}{n^{(t)}(x,y)_z}$. Calculate centred matrix of surface gradients $\mathbf{G}^{(t)}$ using Equation 5.21 and find $\mathbf{b}^{(t+1)}$ by solving:

$$\mathbf{b}^{(t+1)} = \left(\mathbf{\Psi}^T\mathbf{\Psi}\right)^{-1} \mathbf{\Psi}^T\mathbf{G}^{(t)};$$

Estimate combined albedo:

$$\rho_c^{(t+1)}(x, y) = \max \left(\min \left(\frac{I(x, y)}{g(\arccos(\hat{\mathbf{n}}_{\mathbf{b}^{(t+1)}}(x, y) \cdot \mathbf{s}))}, 1 \right), I(x, y) \right);$$

Set iteration $t = t + 1$;

until $\sum_{x,y} \left[\arccos\left(\mathbf{n}^{(t)}(x, y) \cdot \mathbf{n}^{(t-1)}(x, y)\right)\right]^2 < \epsilon$;

next section.

6.3 Model Fitting

The radiance function estimation process described above can be used to estimate an arbitrary radiance function that is monotonic and smooth. In some instances, it is useful to fit a reflectance model to these estimated curves. There are a number of reasons for doing this:

1. We can use the fitted model to extrapolate different slices of the BRDF other than that which was estimated from the input image. For example, to allow the synthesis of images under illumination from a different direction.

2. It allows us to separate the image into the components provided by the model, for example a diffuse and specular component.

3. In the case of colour images (discussed below), we can estimate the colour and strength of the light source as part of this model fitting process. Having done so we can reilluminate the face with arbitrary strengths and colours of illuminant, allowing us to normalise for the effects of varying illumination properties.

In this work we use the Phong [128] model of reflectance, which comprises a Lambertian diffuse term and a specular term controlled by a shininess parameter. In principal any parametric reflectance model could be used and the same method as we describe here could be applied. Although the Phong model is rather simplistic, we choose to use it here for a number of reasons. The first is that it is straightforward to compute and can be easily manipulated into its component parts. The second reason is that it fits our data well, probably because the deviations from Lambert's law present in skin reflectance can be explained well with the addition of a simple specular reflection model. The obvious

weakness of the Phong model is that is has no physical basis and hence the parameters do not have physically meaningful units. For this reason, we suggest future work could include investigating incorporating alternative models of reflectance into the proposed framework.

6.3.1 The Phong Model

The radiance function we use for the Phong model is as follows:

$$g_{\text{phong}}(\theta_i, \theta_h, \rho_d, \rho_s, \eta_s, \|\mathbf{s}\|) = \underbrace{\rho_d \|\mathbf{s}\| \cos(\theta_i)}_{\text{Diffuse term}} + \underbrace{\rho_s \|\mathbf{s}\| \cos^{\eta_s}(\theta_h)}_{\text{Specular term}}, \tag{6.16}$$

where ρ_s is the specular coefficient, η_s is the shininess and $\|\mathbf{s}\|$ is the intensity of the point light source. θ_h is the angle between the surface normal $\mathbf{N}$ and the vector $\mathbf{H} = \frac{\mathbf{L}+\mathbf{V}}{\|\mathbf{L}+\mathbf{V}\|}$ which bisects the light $\mathbf{L}$ and view $\mathbf{V}$ direction vectors, e.g.

$$\cos(\theta_h) = \mathbf{H} \cdot \mathbf{N}. \tag{6.17}$$

6.3.2 Fitting the Phong Model

We wish to estimate the parameters of the Phong reflectance model that best fit our estimated data. We fix our surface normal estimates as those that locally satisfied the irradiance constraint, i.e. $\mathbf{n}(x, y)$. We make the assumption that the specular parameters (ρ_s and η_s) are fixed across the image. This reduces the number of degrees of freedom of the problem while still providing a good fit to the data. We allow the diffuse albedo ρ_d to vary arbitrarily across the surface and use the combined albedo values ρ_c estimated using Algorithm 6.2 as an initialisation.

The minimisation that we wish to solve in order to fit the Phong reflectance model is

Algorithm 6.3: Iterative fitting of Phong reflectance model.

Input: Light source $\mathbf{s}$, image intensities $I(x, y)$, estimated field of surface normals $\mathbf{n}(x, y)$ and estimated combined albedo $\rho_c(x, y)$

Output: Fitted global Phong parameters $\|\mathbf{s}\|$, ρ_s and η_s, and spatially varying diffuse albedo $\rho_d(x, y)$

1 Set iteration $t = 0$;

2 Initialise the diffuse albedo from the combined albedo values estimated using Algorithm 6.2: $\rho_d^{(0)}(x, y) = \rho_c(x, y)$;

3 Initialise global Phong parameters: $\rho_s^{(0)} = 0.5$, $\eta_s^{(0)} = 10$ and $\|\mathbf{s}\|^{(0)} = 0.8$;

4 **repeat**

5 Solve minimisation of global parameters using a Matlab implementation of Newton's method:

$$(\rho_s^{(t+1)}, \|\mathbf{s}\|^{(t+1)}, \eta_s^{(t+1)}) = \underset{\rho_s, \|\mathbf{s}\|, \eta_s}{\arg\min} \sum_{x,y} \left\| I(x, y) - g_{\text{phong}} \left(\theta_i(x, y), \theta_h(x, y), \rho_d^{(t)}(x, y), \rho_s, \eta_s, \|\mathbf{s}\| \right) \right\|^2,$$

values at iteration $(t - 1)$ are used as the initialisations;

6 Calculate diffuse albedo for each pixel by rearranging Equation 6.16:

$$\rho_d^{(t+1)}(x, y) = \frac{I(x, y) - \rho_s^{(t+1)} \|\mathbf{s}\|^{(t+1)} \cos^{\eta_s^{(t+1)}}(\theta_h(x, y))}{\|\mathbf{s}\|^{(t+1)} \cos(\theta_i(x, y))};$$

7 Set iteration $t = t + 1$;

8 **until** $t = 20$;

as follows:

$$\underset{\rho_d(x,y), \rho_s, \|\mathbf{s}\|, \eta_s}{\arg\min} \sum_{x,y} \|I(x, y) - g_{\text{phong}} (\theta_i(x, y), \theta_h(x, y), \rho_d(x, y), \rho_s, \eta_s, \|\mathbf{s}\|)\|^2 \quad (6.18)$$

We solve this minimisation by iteratively applying two processes: we use Newton's method to solve for the three global parameters ($\|\mathbf{s}\|$, ρ_s and η_s), we then invert the radiance function to calculate the diffuse albedo. The process is summarised in Algorithm 6.3. We find this process converges within approximately 20 iterations and the resulting fit recreates the original image to within a negligible tolerance.

6.3.3 Image Synthesis

With the estimated parameters to hand we can synthesise images under any pose and illumination conditions. To do so, we simply calculate the intensity for each normal $\mathbf{n}(x, y)$ with Equation 6.16 under the desired illumination and viewing directions using the fitted values for the global Phong parameters and diffuse albedo. The resulting intensities are texture mapped onto the surface $z_{\mathbf{b}}(x, y)$ which is then rotated to the specified pose. In our experimental section we demonstrate that this approach yields stable images under a wide variety of conditions in which the position of specularities and other complex reflectance effects are realistically synthesised.

6.4 Colour Images

Prior to presenting our experimental results, our final contribution is to show how the methods developed in this chapter may be applied to colour images. If an RGB colour image comprises three colour channels $I_{\text{red}}(x, y)$ containing intensity values in the red channel, $I_{\text{green}}(x, y)$ in the green channel and $I_{\text{blue}}(x, y)$ in the blue channel, we may apply the radiance function estimation process to each channel independently, yielding three radiance functions g_{red}, g_{green} and g_{blue}. Enforcing the local irradiance constraint described above to each channel provides three estimates of the field of surface normals $\mathbf{n}_{\text{red}}(x, y)$, $\mathbf{n}_{\text{green}}(x, y)$ and $\mathbf{n}_{\text{blue}}(x, y)$. We choose as our estimate of the true field of surface normals the average of the three, i.e.:

$$\mathbf{n}(x, y) = \frac{\mathbf{n}_{\text{red}}(x, y) + \mathbf{n}_{\text{green}}(x, y) + \mathbf{n}_{\text{blue}}(x, y)}{\|\mathbf{n}_{\text{red}}(x, y) + \mathbf{n}_{\text{green}}(x, y) + \mathbf{n}_{\text{blue}}(x, y)\|} \tag{6.19}$$

It is to this estimate that we apply the model-based integrability constraint. From the resulting field of surface normals, $\hat{\mathbf{n}}_{\mathbf{b}}(x, y)$, we calculate the combined albedo in each

channel:

$$\rho_{c,\text{red}}(x,y) = \max\left(\min\left(\frac{I_{\text{red}}(x,y)}{g(\arccos(\hat{\mathbf{n}}_{\mathbf{b}}(x,y)\cdot\mathbf{s}))},1\right), I_{\text{red}}(x,y)\right) \tag{6.20}$$

Similarly for the green and blue channels.

6.4.1 Fitting the Phong Model to Colour Images

We can fit the Phong model in exactly the same way as described in Section 6.3.2. For each colour channel, we estimate the diffuse albedo, e.g. $\rho_{d,\text{red}}(x,y)$, the strength of the light source at this wavelength, e.g. $\|\mathbf{s}\|_{\text{red}}$, and the two specular paramters, e.g. $\rho_{s,\text{red}}$ and $\eta_{s,\text{red}}$. In each case we use the field of surface normals estimated from all three colour channels, $\mathbf{n}(x,y)$, and initialise the diffuse albedo using the combined albedo estimated for that channel, e.g. $\rho^{(0)}_{d,\text{red}}(x,y) = \rho_{c,\text{red}}(x,y)$.

Colour Constancy

In fitting the Phong model to a colour image, we estimate the intensity of the light source in each of the colour channels: $\mathbf{s}_{\text{rgb}} = [\|\mathbf{s}\|_{\text{red}}, \|\mathbf{s}\|_{\text{green}}, \|\mathbf{s}\|_{\text{blue}}]^T$. In other words, we have an estimate of the colour of the light source in terms of an RGB vector. If we reilluminate the input face using Equation 6.16 with a white light source of unit intensity, i.e. $\mathbf{s}_{\text{rgb}} = [1,1,1]^T$, we effectively normalise for the effects of the colour and intensity of the light source. This may be a useful preprocessing step for algorithms which use colour face images. In our experimental section we demonstrate that this process is stable across different subjects and can be used to realistically remove the effects of coloured lighting from real world images.

6.5 Experiments

In this section we present experimental results of our reflectance estimation process. This study is broken down into four parts. In Section 6.5.1 we begin by examining the iterative behaviour of the algorithm. In Section 6.5.2 we then examine the quantitative performance of the algorithm on ground truth data. In Section 6.5.3 we present results on real world images. In Section 6.5.4 we show the results of applying our algorithm to colour images. Finally, in Section 6.5.5 we compare the accuracy of the shape information recovered from images of faces with non-Lambertian reflectance using a variety of the methods developed in this thesis.

The surface height model we use was trained in exactly the same way as in Chapter 5.

6.5.1 Algorithm Behaviour

We begin by observing the iterative behavior of the algorithm. In Figure 6.2 we show the scattered radiance data overlayed by the estimated radiance function (shown in blue). From left to right we show results after mean face shape initialisation, 1 iteration and 10 iterations. As the algorithm iterates, the radiance data becomes more tightly clustered around the radiance function estimate. In the final column we compare the estimated function with ground truth (shown by the broken line). There is a good agreement between the two. In addition, we show the result of the Robles-Kelly and Hancock radiance function estimation procedure (dotted line). It is clear that our method provides improved accuracy. We obtained similar results for a range of synthetic data.

6.5.2 Synthetic Data

We begin with a qualitative analysis of the radiance function estimation process. In Figure 6.3 we show synthetic input images (first column) rendered with Phong parameters: $(\rho_d =$

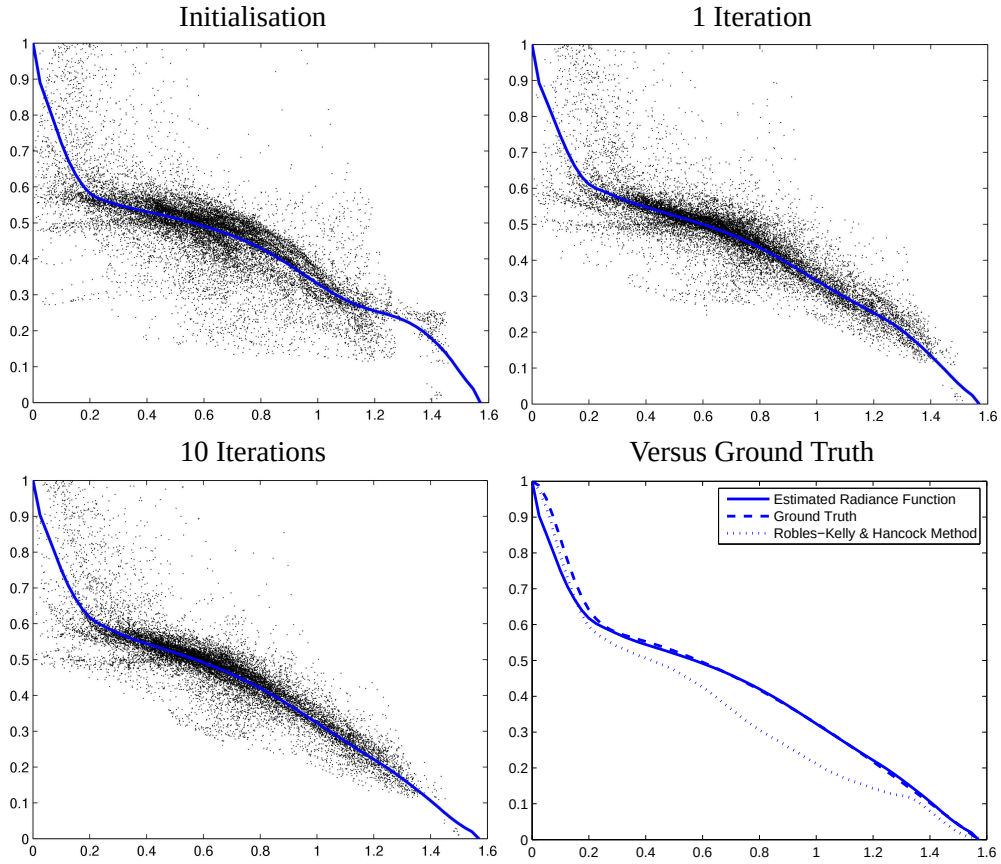


Figure 6.2: Iterative behavior of radiance estimation algorithm. Estimated radiance functions plotted over current estimate of the scattered radiance measurements for differing iteration numbers. In the top left panel we show the initial estimate using the average face surface. In the top right and bottom left panels we show results after 1 and 10 iterations respectively. Finally, in the bottom right panel we show the estimated radiance function plotted against ground truth and the estimate provided by an existing technique.

$1, \rho_s = 0, \eta_s = 0)$, $(\rho_d = 0.7, \rho_s = 0.3, \eta_s = 100)$ and $(\rho_d = 0.6, \rho_s = 0.4, \eta_s = 10)$ respectively. In the second column we show unit spheres rendered with the radiance function estimated from the corresponding input images using Algorithm 6.1. In the third column we show ground truth radiance spheres. Visually, there is a good qualitative agreement between the estimated and ground truth spheres. In Figure 6.4 we plot these estimated radiance functions against ground truth for the same three input images (in other words, the functions that were used to render the spheres in Figure 6.3). On each plot we

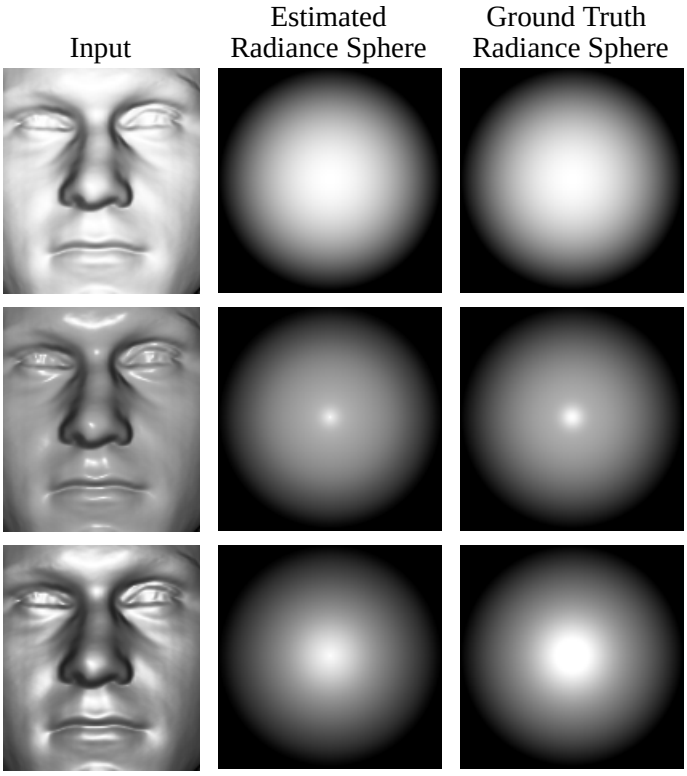
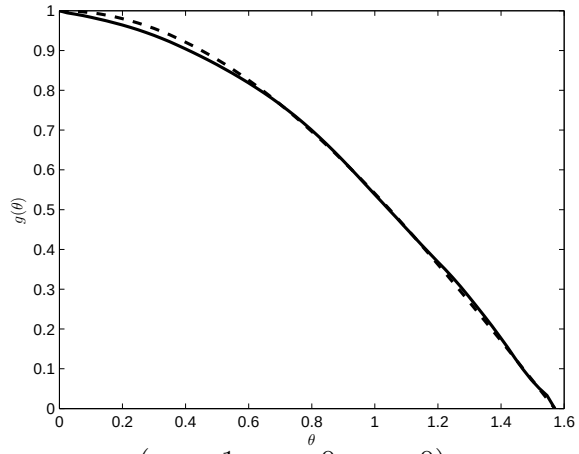


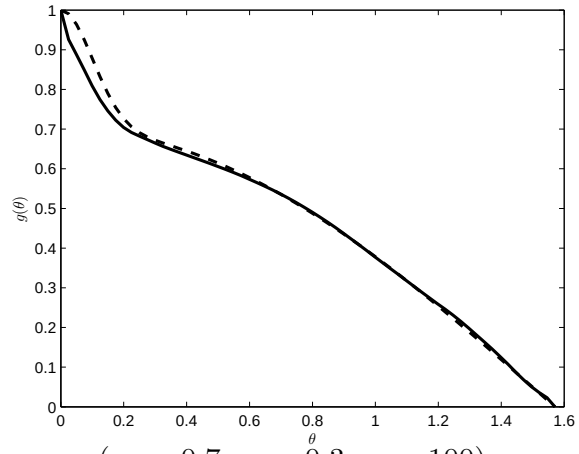
Figure 6.3: Example results of the radiance estimation process on synthetic images. First column: synthetic Phong images with parameters: $(\rho_d = 1, \rho_s = 0, \eta_s = 0)$, $(\rho_d = 0.7, \rho_s = 0.3, \eta_s = 100)$ and $(\rho_d = 0.6, \rho_s = 0.4, \eta_s = 10)$ respectively. Second column: image of a unit sphere rendered with the radiance function estimated using our method. Third column: image of a unit sphere rendered with the ground truth radiance function.

show the estimated radiance function with a solid line and ground truth with a broken line. The estimate appears accurate in all three cases. Errors in the estimated radiance functions appear to be concentrated towards lower values of incident angle, i.e. when $\theta \approx 0$.

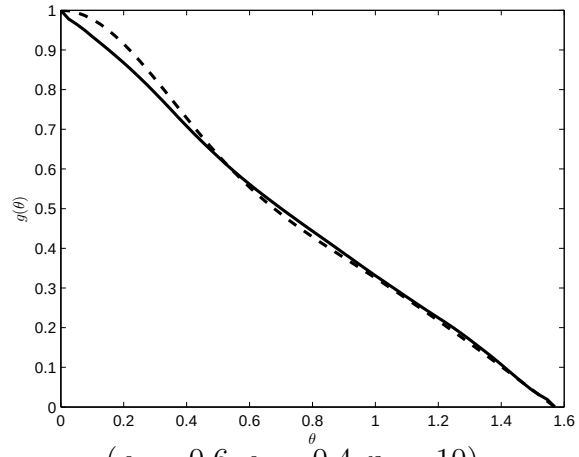
We now provide quantitative results. As in Chapter 5 we used a leave-one-out testing strategy in which we rebuilt the surface height model, excluding the surface under study to avoid biasing the process. In other words, all testing data was out-of-sample of the model. We rendered each facial surface with a point light source whose direction was coincident with the viewing direction. We used the Phong reflectance model and varied



$(\rho_d = 1, \rho_s = 0, \eta_s = 0)$



$(\rho_d = 0.7, \rho_s = 0.3, \eta_s = 100)$



$(\rho_d = 0.6, \rho_s = 0.4, \eta_s = 10)$

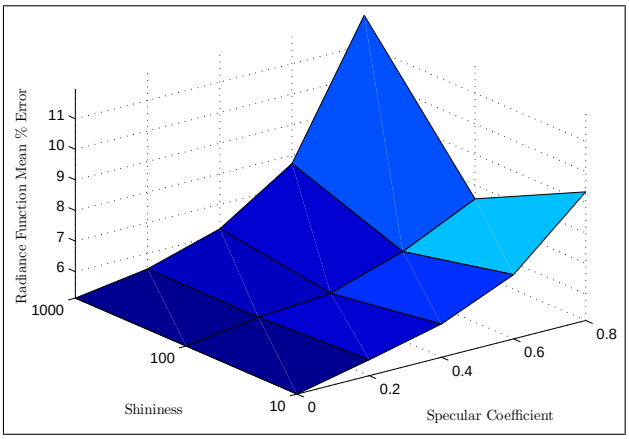
Figure 6.4: Examples of estimated radiance functions. Estimated (solid line) and ground truth (broken line) radiance functions for synthetic Phong images. From top to bottom, the images were rendered with Phong parameters $(\rho_d = 1, \rho_s = 0, \eta_s = 0)$, $(\rho_d = 0.7, \rho_s = 0.3, \eta_s = 100)$ and $(\rho_d = 0.6, \rho_s = 0.4, \eta_s = 10)$ respectively.

the parameters to study their effect on the accuracy of the estimated radiance function. Perhaps more interestingly, we can also examine the relationship between the reflectance function parameters and the accuracy of the recovered shape. The radiance function used was as described in Equation 6.16. For each surface we generated synthetic images with three values of shininess $\eta = 10$, 100 and 1000 and five values of the specular coefficient $k_s = 0$, 0.2, 0.4, 0.6 and 0.8. Note that when $k_s = 0$ the reflectance model is Lambertian. We set the diffuse coefficient as $\rho_d = 1 - \rho_s$.

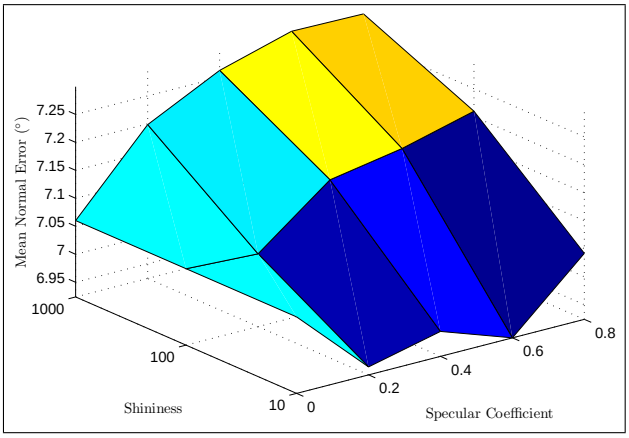
In Figure 6.5(a) we show the percentage error in the estimated radiance function. In (b) we show the mean angular surface normal error, i.e. the average angular difference between the estimated and ground truth surface normals. In (c) we show the mean absolute error between the estimated and ground truth surface height function. From the surface shape error plots we note that the error tends to increase with shininess, i.e. as the specular spike becomes narrower. This adds empirical support to the anecdotal evidence that it is harder to recover shape-from-shading for shiny objects [133]. There is also a trend for the error to increase as the specular coefficient increases, i.e. as the specular spike becomes taller. The reason for this is that the effect of a larger specular coefficient is to compress a wider range of values of incident angle into a smaller range of intensity values. Hence, quantisation noise becomes more significant. In other words, fewer gray levels are used to represent a certain range of incident angles as the specular coefficient increases.

6.5.3 Real World Data

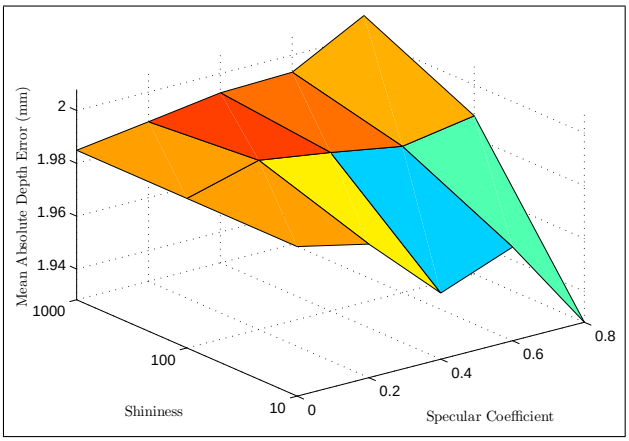
We now present the results of applying Algorithm 6.2 to real world face images. In the first column we show a selection of input images from the Yale Face Database B [62]. These images clearly show variation in surface reflectance properties (some are shinier than others and some are darker or lighter) as well as variations in albedo. In the sec-



Error in estimated radiance function.



Error in estimated field of surface normals.



Error in estimated surface height.

Figure 6.5: Error plots for synthetic data rendered using the Phong model.

ond column we show unit spheres rendered with the estimated radiance functions. These appear to capture the corresponding reflectance properties well. In the third column we show Lambertian corrected images, in which the estimated shape has been rendered with Lambertian reflectance. The estimated shape appears qualitatively good (distinguishing features are sharply defined) and the effect of varying reflectance properties has been removed. Finally, in the last two columns we show the results of fitting the Phong reflectance model to the estimated data using Algorithm 6.3. We show the estimated diffuse and specular components of the Phong model. The estimated specular reflectance agrees well with empirically measured results [44].

In Figure 6.7 we show how changes to skin reflectance in the real world are recovered using our method. We applied increasing amounts of artificial perspiration, in the form of a water-based spray, to a subject's face. The input images as the amount of simulated perspiration is increased are shown in the first column. In the second column we show unit spheres rendered with the estimated radiance functions. Note how the specular spike becomes more sharply defined as the perspiration increases. In the third column we show the estimated radiance functions. Again, note that the gradient of the radiance curve increases with increasing perspiration when the incident angle is close to zero. It is this effect that is responsible for the sharper specular spike. Finally, we fitted the Phong reflectance model to this data using Algorithm 6.3. From top to bottom, the values of the shininess parameter η_s were 35.17, 173.7 and 184.9 respectively. There is therefore a strong correlation between the amount of simulated perspiration and the shininess parameter estimated using our technique.

Direct Synthesis

We now show the results of synthesising images in a novel pose directly using the estimated radiance function and combined albedo values as described in Section 6.2.2. We

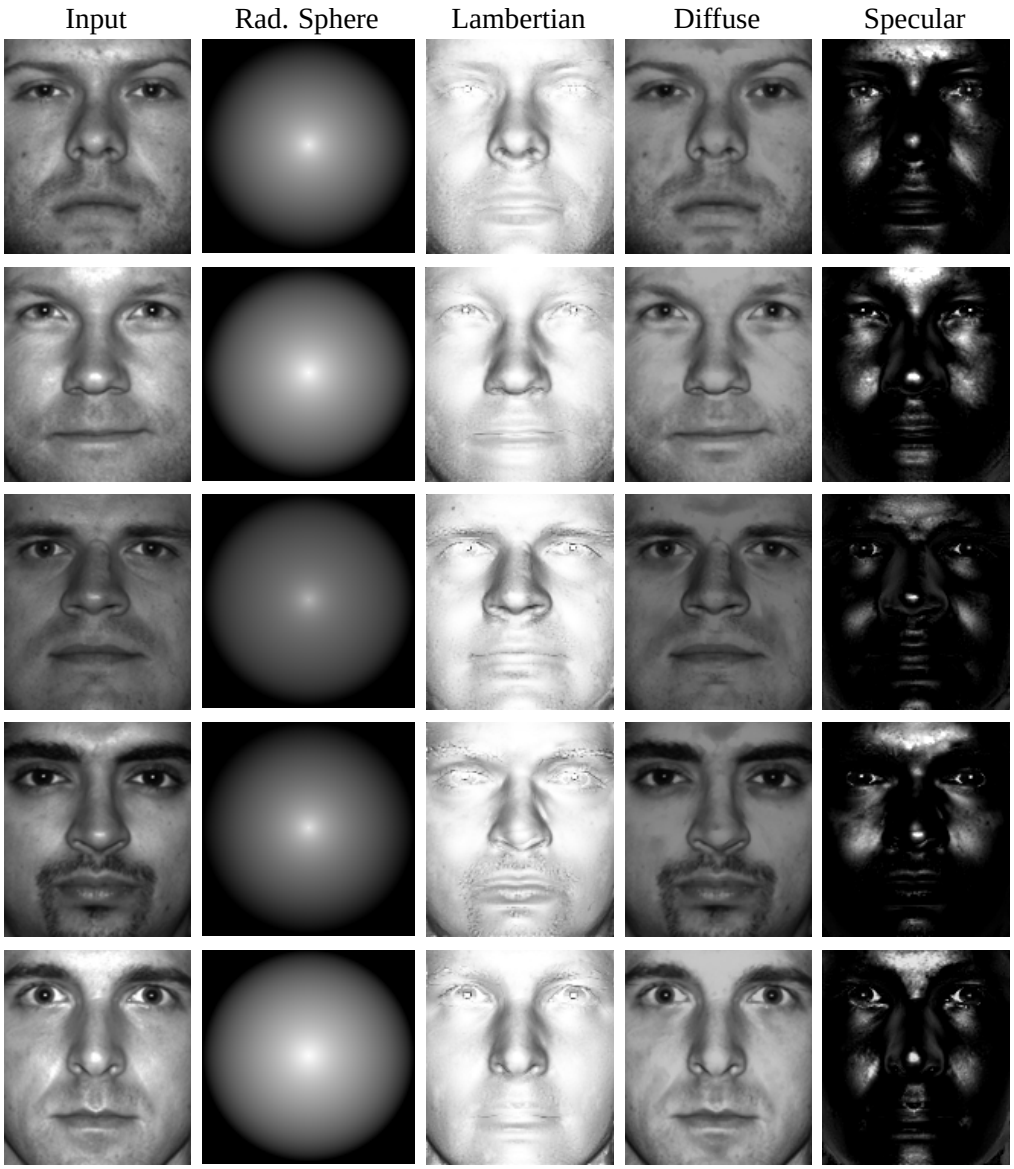


Figure 6.6: In the first column we show the input images and in the second we show a unit sphere rendered with the estimated radiance function. In the third column we show the estimated shape rendered with Lambertian reflectance. Finally we show the estimated diffuse (column 4) and specular (column 5) components of the surface reflectance having fitted the Phong model to the estimated data.

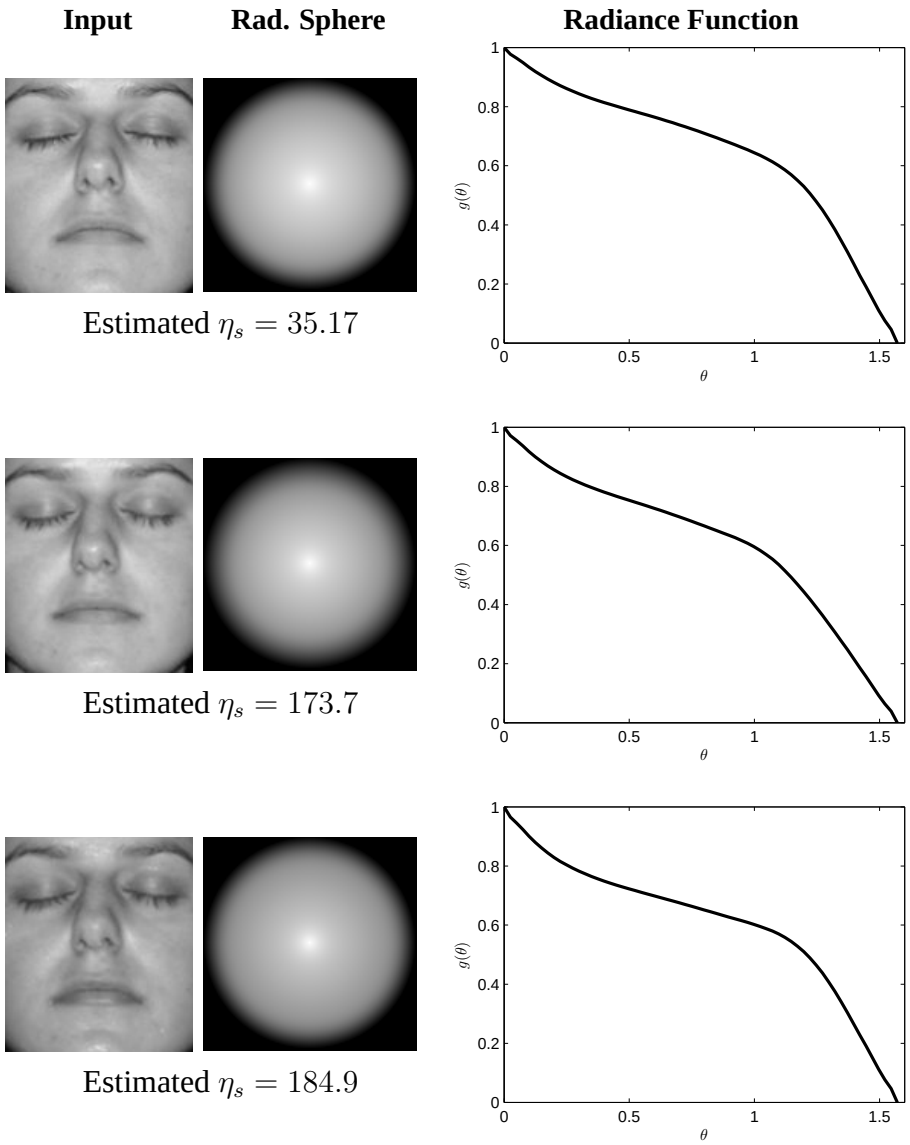


Figure 6.7: Reflectance measurements of simulated perspiration. From top to bottom the first column shows input images of a subject with increasing amounts of artificial perspiration added. To simulate the effect of perspiring, a water-based solution was sprayed onto the subject's face. In the second column we show unit spheres rendered with the estimated radiance functions. In the third column we show estimated radiance functions. The shininess parameter of the fitted Phong model correlates well with the increased perspiration.

show synthesised views in which the face is rotated about the vertical axis by an angle of $-45°$, $-22.5°$, $0°$, $22.5°$ and $45°$. The viewing and light source directions remain coincident and hence the position of specularities varies as the face rotates. Qualitatively, the effect is convincing and shows that we can synthesise realistic images using the estimated radiance function directly.

In Figure 6.9 we synthesise images under varying pose in the same manner. However, this time we provide real views in approximately the same pose in order to provide a comparison between the predicted and actual view. In many cases the position of the specularities and the global shape estimate appear accurate.

Synthesis using fitted Phong Reflectance Model

By fitting the Phong reflectance model to the estimated data we are able to synthesise images using a different slice of the BRDF to the one that was present in the original input image. In Figure 6.10 we show each of the 10 subjects in the Yale B database under 4 different synthesised illumination conditions. In each case the light source subtends an angle of $45°$ with the viewing direction and, from left to right, is moved left, right, above and below the subject. The input images are shown in the first column. Note how the synthesised images capture the complex distribution of specular reflectance. For comparison, the reader is referred back to Figures 3.14 and 3.15 in Chapter 3. Here, the images are synthesised using the Lambertian reflectance model. Note how the results in Figure 6.10 are substantially more life-like and more realistically capture the reflectance properties of the different subjects.

6.5.4 Colour Images

We now present results related to the estimation of the colour of the illuminant and normalising for its effect. We begin by examining the stability of this process on the 68 sub-

Figure 6.8: Synthesising novel viewpoints directly using the estimated radiance functions and albedo maps. The viewer and light source direction remain coincident. Note that the specularities change position as the head rotates.

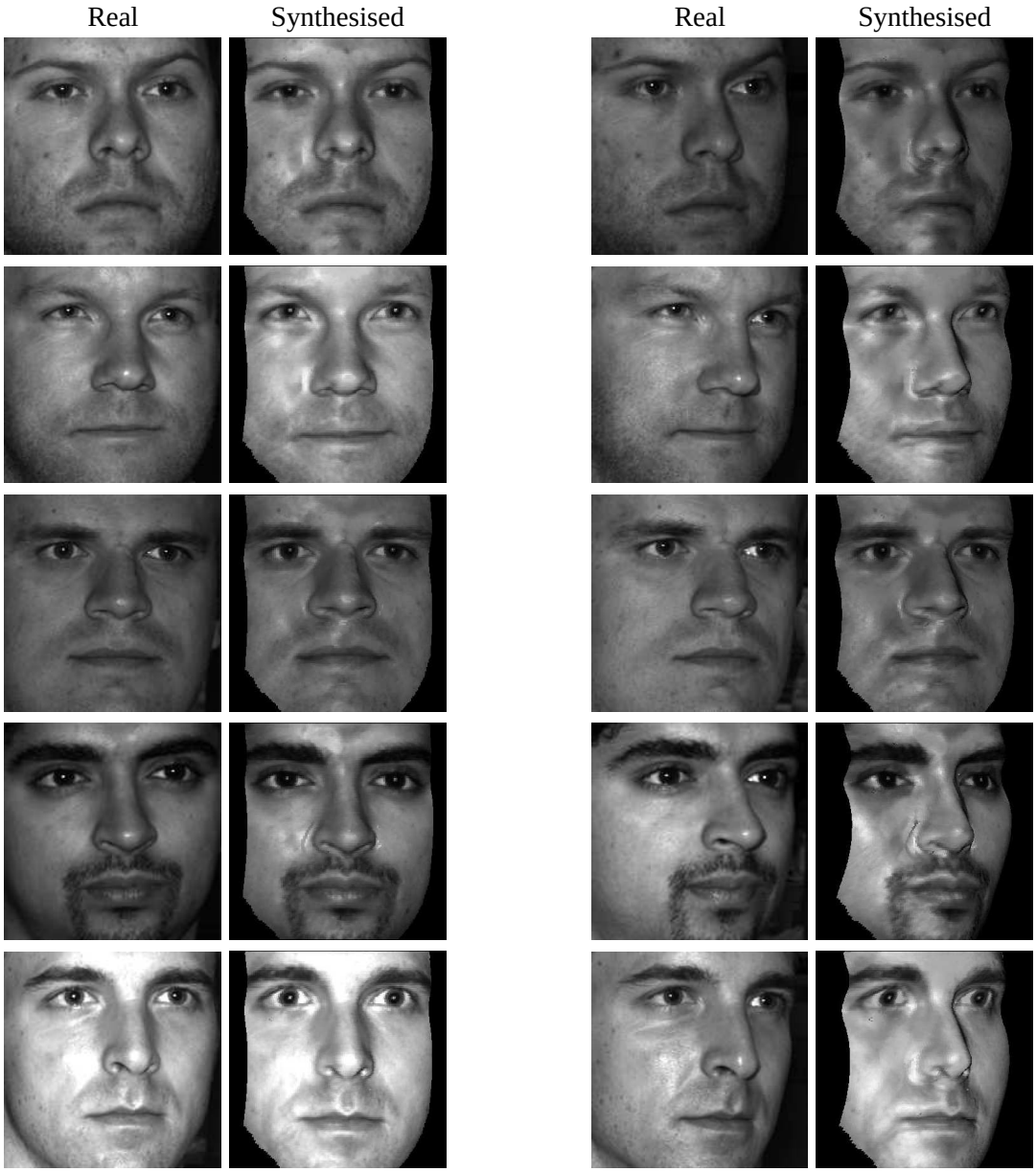


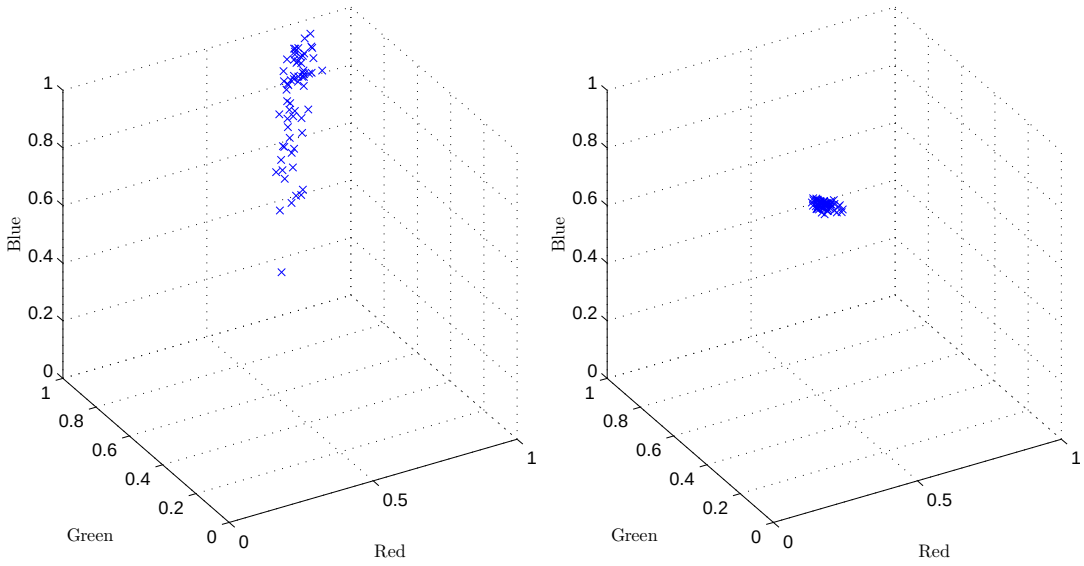
Figure 6.9: Comparing real views against synthesised for changes in pose.

Figure 6.10: Synthesising images under novel illumination using the fitted Phong Reflectance Model. The first column shows the input images. From left to right, the remaining four columns show images synthesised with a light source 45° left, right, above and below the subject.

jects in the CMU PIE database. For each subject, we apply the colour shape-from-shading algorithm described in Section 6.4 to input images in which the subject is illuminated by a frontal light source. The illumination in this database is provided by Minolta 220X flash bulbs, which remain constant between subjects. All gain and gamma correction on the cameras was switched off. The colour of the flashes is biased towards shorter wavelengths, and hence is strongest in the blue channel (this is evident in the input images in which the subjects appear unnaturally blue). For this reason, these images are ideal for analysing the performance of colour constancy algorithms.

We fit the Phong model to each colour channel of each image, as described in Section 6.4.1. We therefore estimate a diffuse albedo map for each colour channel, as well as the shininess parameter, specular coefficient and the strength of the light source in each channel. In Figure 6.11(a) we show the estimated light source colour vectors as a scatter plot in RGB space. Because of the unknown scaling factor present in Equation 6.5, these vectors cannot be compared directly. Instead, we consider the hue of the light source (i.e. its intrinsic colour) by normalising the RGB vectors to unit length. We show a scatter plot of these vectors in Figure 6.11(b). It is clear that the hue of the light source estimated from the 68 input images is very stable (the points are tightly clustered). The intrinsic mean of this distribution of unit vectors is: $\mathbf{s}_{\text{rgb}} = [0.4801, 0.5276, 0.7008]^T$. This agrees well with the information available about the illumination used, i.e. it is strongest in the blue channel and the intensity of the light source increases as the wavelength decreases. The distribution of estimated hues also cluster well, the standard deviation across the 68 samples is only 0.0459 radians or 2.63° (these values are based on the directional errors of unit vectors in RGB space).

In Figure 6.12 we show the results of our facial colour constancy algorithm. Having recovered the facial shape and fitted the Phong reflectance model to the input colour images as described above, we synthesise an image in which the strength and colour of the



(a) (b)

Figure 6.11: Distribution of estimated light source colour vectors. In (a) we plot the raw estimated RGB vectors of the colour of the illuminant for all 68 subjects in the CMU PIE database. In (b) we plot the hue of these colour vectors (i.e. length normalised to unity). The estimated hue of the light source is stable across all input images.

illuminant is normalised to unity. To do so, we simply render the estimated field of surface normals using the fitted Phong parameters and a white light source of unit intensity. The resulting images effectively normalise for the colour and strength of the illumination. Qualitatively, the results demonstrate that the method works. The illumination-normalised images appear more natural and traces of the blue illumination have been removed. Also note that because of the increase in the strength of the illuminant, the specularities increase in size.

6.5.5 Comparisons

We conclude our experimental section by demonstrating the effect of non-Lambertian reflectance on the various shape-from-shading techniques presented in this thesis. We use a synthetic input image rendered using Phong reflectance with parameters ($\rho_d = 0.7, \rho_s =$

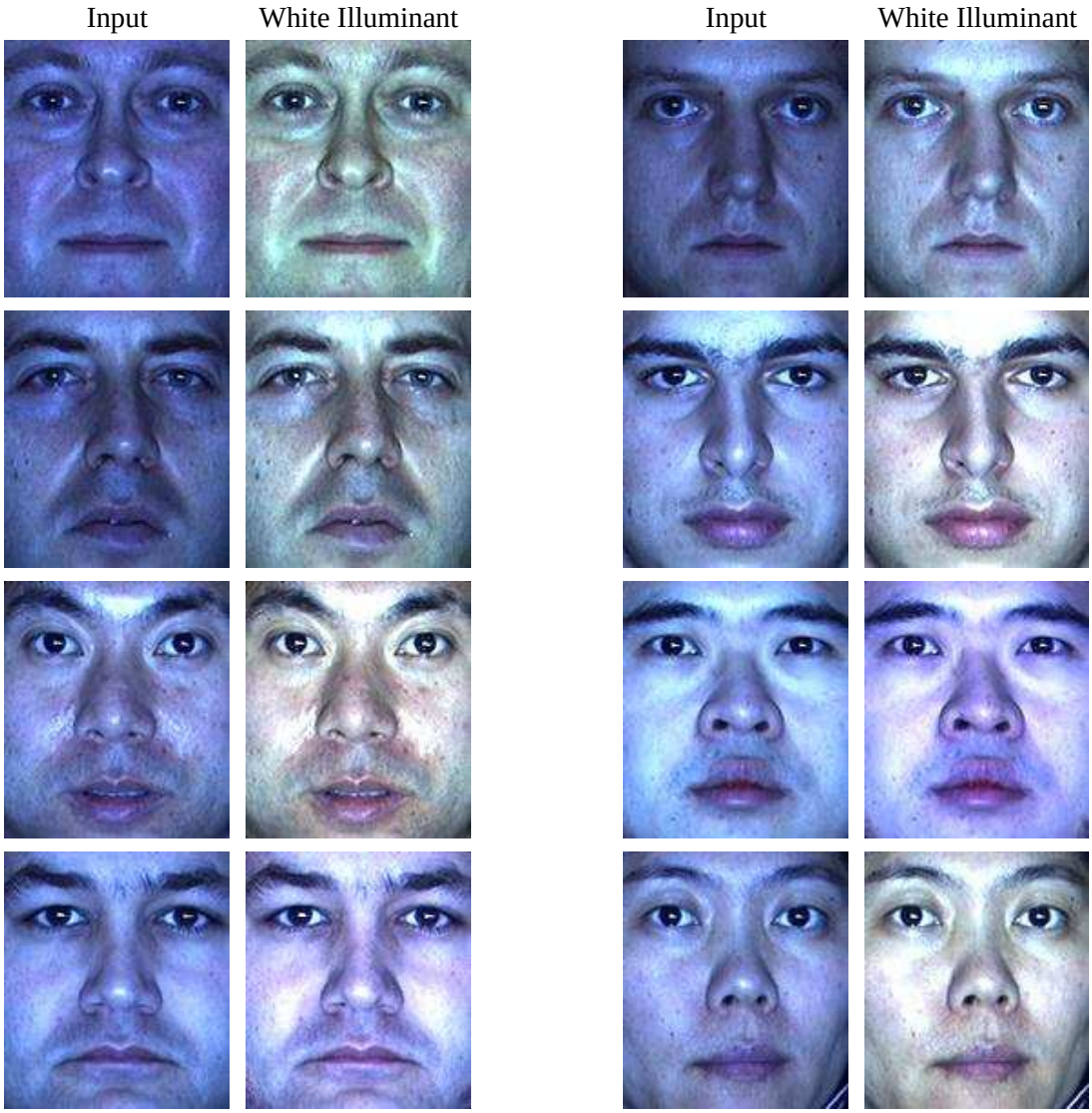


Figure 6.12: Colour input images are shown in the first and third columns. It is evident that the illuminant is stronger in the blue channel than the red or green. In the second and fourth columns we show the results of our colour constancy algorithm. Here the estimated field of surface normals have been rendered with the Phong model using the fitted parameters and a white light source of unit intensity.

$0.3, \eta_s = 100)$ (the input image is the same as in the second row of Figure 6.3). This image contains strong specular reflection. In Figure 6.13 we provide a comparison between the results of applying four shape-from-shading techniques to this image. The first is the principal geodesic shape-from-shading algorithm given in Algorithm 4.1. The second approach uses the same shape-from-shading algorithm, but the input image is preprocessed with the Lambertian correction technique of Robles-Kelly and Hancock [135]. This is the same approach as is adopted in the experimental sections of Chapters 3 and 4. The third approach is the Lambertian shape-from-shading algorithm which uses the model-based integrability constraint, as described in Algorithm 5.1. The fourth and final approach is the method described in this chapter in which the radiance function is estimated as part of the shape-from-shading process (Algorithm 6.1).

In the second column we show error maps for the estimated fields of surface normals. The scale is consistent across the four plots and the units of the colour bars are degrees. The error plotted is the angular difference between the estimated and ground truth surface normal. In the third column we show errors maps for the estimated surface heights. For the first two techniques, the surface integration algorithm of Frankot and Chellappa is used. The errors shown are the magnitude of the difference in height to the ground truth surface and the units are centimetres. Finally, in the fourth column we show a profile view of the recovered surface. A profile view of the ground truth surface is shown in the fifth row for comparison.

The two techniques which assume Lambertian reflectance (rows one and three) show significant errors in both the estimated surface and surface normals. In both cases, from the surface normal error maps, it is clear that the largest errors fall in the regions containing significant specular reflections, for example the tip of the nose, forehead and lips. The effect of these errors is to warp the recovered surface, exaggerating features such as the nose and forehead. In contrast, both methods that take account of non-Lambertian

reflectance (rows two and four) have significantly reduced errors. In the case of the Lambertian corrected input (row two), the surface normal errors are distributed fairly evenly across the face. This is due to the inaccuracy of the one-shot reflectance estimation process. On the other hand, the technique which incorporates radiance function estimation (row four) has lower surface normal errors across much of the face. Where errors are more significant (primarily around the lips and eye sockets) it is believed this is due to the poor generalisation of the surface height model in these areas. From the profile views of the recovered surfaces, it is clear that the approach developed in this chapter recovers the most accurate surface. We obtained similar results on a wide range of ground truth data.

6.6 Conclusions

In this chapter we have significantly relaxed the assumptions made in the previous three chapters regarding the shading model used. We have shown how an estimate of the radiance function can be made from single, frontally-illuminated face images as part of a shape-from-shading algorithm which incorporates the model-based integrability constraint developed in Chapter 5. We demonstrated on ground truth data that these estimated radiance functions are quantitatively accurate. In addition, we allow the local albedo to vary arbitrarily (subject to physical constraints). Where in Chapter 3 we ignored albedo variations during fitting and in Chapter 4 used robust statistics to try and down-weight its effects, in this chapter we have explicitly accounted for the albedo and calculate it as part of the shape recovery process. We are able to estimate accurate fields of surface normals using this approach and, using the estimated radiance function, can directly synthesise images in which the light source and viewing direction remain coincident.

We then showed how a parametric reflectance model (in this case the Phong model, though in principle it could be any such model) may be fitted to the estimated data. In do-

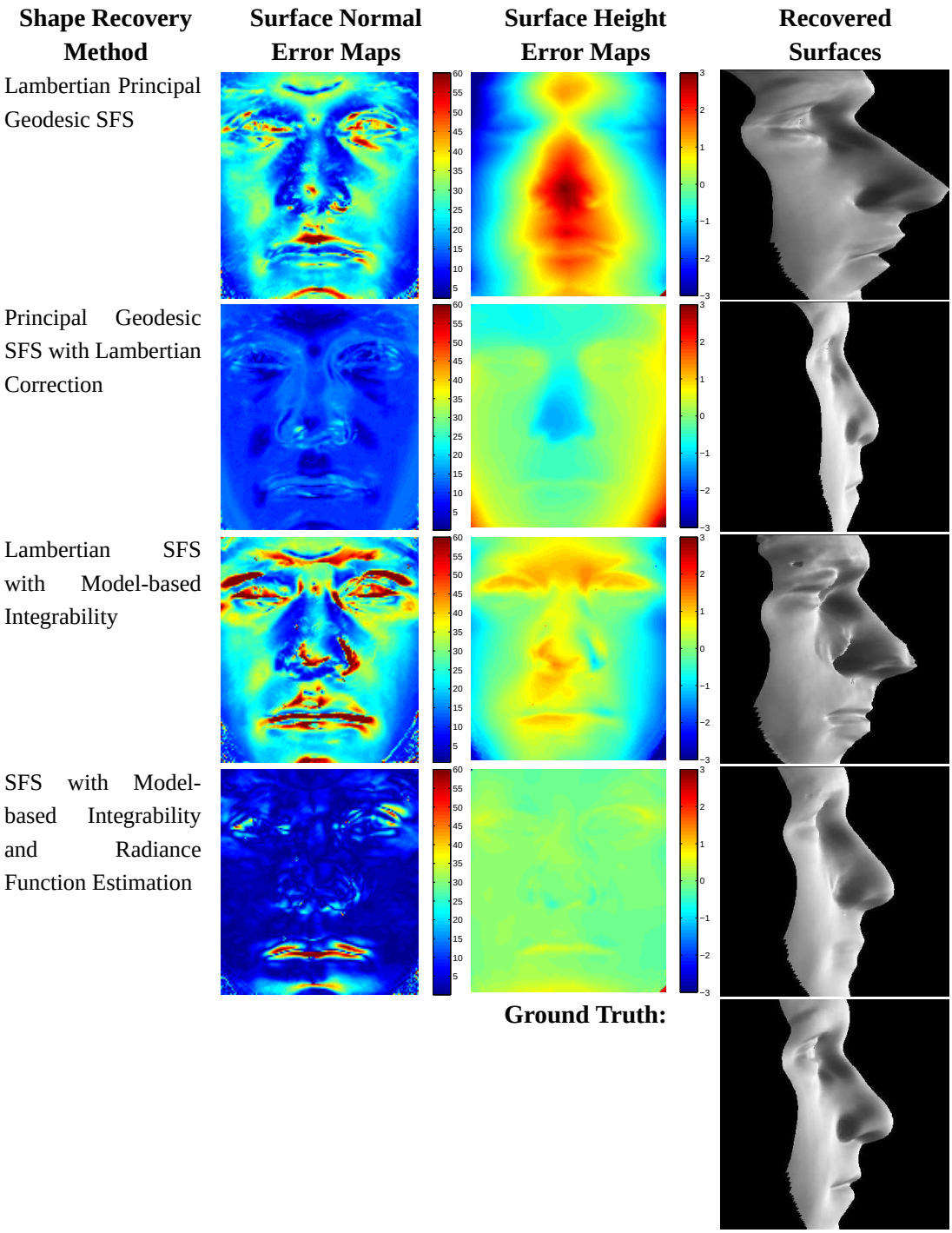


Figure 6.13: Comparison between shape-from-shading techniques on non-Lambertian images. We compare the performance of four shape-from-shading algorithms on a synthetic image rendered with Phong reflectance and parameters $(\rho_d = 0.7, \rho_s = 0.3, \eta_s = 100)$. Column two shows error maps in the estimated field of surface normals. The units are degrees and the error is the angular error. Column three shows errors maps in the estimated surface height. The units are centimetres and the error plotted is the magnitude of the difference in surface height. Finally, in the fourth column we show a profile view of the estimated surface.

ing so, we can synthesise images in which the reflectance properties must be extrapolated beyond those which were present in the original image. This allows us to synthesise images under arbitrary illumination and pose. Comparisons show that the predicted shading patterns agree well with real world images.

Finally, we also showed how the same approach can be applied to colour images. The number of unknowns relative to the amount of data is reduced here, since the shape estimate is fixed across the three colour channels. We demonstrated that we can synthesise realistic colour images as well as correcting for the colour of the illuminant. This shows that the method provides a route to facial colour constancy, which may be of use as a preprocessing step prior to applying existing colour face image analysis techniques.

Chapter 7

Conclusions

In vain we fondly strive to trace
The soul's reflection in the face;
In vain we dwell on lines and crosses,
Crooked mouths and short probosces;
Boobies have looked as wise and bright
As Plato and the Stagyrite
And many a sage and learned skull
Has peeped through windows dark and dull.

- Thomas Moore [112]

In this chapter we summarise the main contributions of the Thesis in the areas of facial shape-from-shading, estimation of reflectance properties and face recognition. These are: the development of two formulations of a statistical model for fields of surface normals, the incorporation of these models into a geometric shape-from-shading framework, the use of robust statistics to reduce the influence of areas of low albedo and cast shadows, constructing linear subspaces for face recognition using the recovered shape and albedo information, the development of statistical methods for surface integration and finally a non-Lambertian facial shape-from-shading scheme which estimates reflectance properties

alongside facial shape. In addition, weaknesses are highlighted in the work as it stands and we provide suggestions for areas that warrant future attention.

7.1 Summary of Contributions

In the literature review in Chapter 2, we highlighted the fact that solutions to the classical shape-from-shading problem perform poorly on real world images, including faces. The work presented in this thesis has shown that statistical models of face shape can significantly improve the quality of the shape information that can be recovered from a single image.

7.1.1 Statistical Models for Field of Surface Normals

Our starting point was to consider the problem of constructing a statistical model that captures variations in facial shape in the domain of fields of surface normals. The motivation behind this aim was that shape-from-shading is concerned with recovering surface orientation information as opposed to 3D surface shape. This is because it is the orientation of a surface that determines the corresponding image brightness as opposed to the surface height directly. To construct the model required overcoming the problem that a distribution of unit vectors is nonlinear in nature. We proposed two ways in which to do this.

The first was described in Chapter 3. This was based on the intuitive observation that our model should linearly deform each surface normal from its average position along great circles on the unit sphere. We defined the average field of normals using the notion of the mean direction as defined in the directional statistics literature. We turned to the field of cartography as a rich source of projections from the sphere to a plane and showed how to transform a distribution of unit vectors from the sphere to the plane using the

azimuthal equidistant projection. The resulting points retain their distance to the mean and can be analysed using linear methods.

The second model formulation described in Chapter 4 was based on a more rigourous analysis using techniques from differential geometry. We define the average field of surface normals by calculating the intrinsic mean of a distribution of points on a spherical manifold. We find the modes of variation of the data using principal geodesic analysis. This second formulation results in a more elegant algorithm in which the whole shape-from-shading process can be couched in terms of operations on tangent planes to the unit sphere.

7.1.2 Statistical Shape-from-shading

We use these models to impose a statistical constraint on the recovered field of surface normals within the framework of a geometric shape-from-shading algorithm. The method proposed in Chapter 3 iterates between imposing Lambert's law (that the cosine of the angle between the estimated normal and light source vector must equal the corresponding image intensity) as a hard constraint and the constraint provided by the model (by finding the best fit of the model to the estimated field of surface normals). The method offers significant improvements over the generic constraints employed within the original geometric shape-from-shading work, enabling plausible facial shape information to be recovered from single images. Where we required the 3D surface height function, we applied an existing surface integration algorithm to the estimated field of surface normals.

Moreover, the algorithm provides as output two estimates of the field of surface normals which satisfy the two different constraints. The field of surface normals that satisfies Lambert's law is more likely to recover fine surface details and atypical local shape features. However, the model-based estimate is likely to be less affected by areas of low albedo. By taking this as our estimate of the underlying facial shape we can make an

estimate of the albedo. We have shown that qualitatively, these albedo maps successfully recover surface marking such as the eye brows and lips. By combining the estimated shape and albedo we are able to synthesise images under widely varying conditions of illumination and pose with considerable stability.

In Chapter 4 we extend this algorithm in a number of ways. First, by calling on techniques from differential geometry we show how the constraint provided by both the statistical model and Lambert's law can be imposed using operations on the tangent plane. Second, we use robust statistics to down-weight areas of low albedo or regions which lie within a cast shadow. The result is that we effectively fit the statistical model only to surface normals that satisfy our simple, local Lambertian shading model. A further advantage of this technique is that the weights provide a principled means to combine the two sets of estimated normals (satisfying the irradiance and model-based constraints respectively), allowing the surface to be infilled across shadow regions.

7.1.3 Face Recognition

Much of the motivation for developing techniques for recovering intrinsic facial features from single images comes from the potential of applying this information to the problem of face recognition. We have explored this possibility to a limited degree in this thesis. In Chapter 4 we considered a number of ways in which the recovered surface normals and albedo information could be used for recognition. The first two techniques were based on comparing the shape information recovered from each gallery image to that recovered from a probe image. This strategy fails, since the accuracy of the recovered shape degrades significantly for highly non-frontal lighting. The conclusion here is that it is easier to predict how a face will appear under extreme illumination given an image under frontal lighting than vice versa. The second group of techniques are based on constructing linear subspaces which capture the image variation due to changing illumination for one subject.

Previously, these subspaces were constructed either from multiple training images (¿7) or using information from a laser range scanner. Using the shape and albedo estimated from a single training image we are able to synthesise these subspaces. The cost for reducing the training requirements to a single image, is an overall increase in the error rate of around 4% across a wide range of illumination conditions.

7.1.4 Statistical Surface Integration

In Chapter 5 we broadened our consideration of facial shape recovery to include estimation of surface height in addition to surface orientation, once again in a statistical manner. We present two surface integration techniques which make use of a statistical surface height model. The first is based on a statistical coupling between the parameters of a surface height and surface normal model. The second is based on fitting a surface height model to a field of surface normals by minimising the error between the fitted and actual surface gradients. We show that both techniques outperform the generic surface integration algorithm used in Chapters 3 and 4.

We then take this analysis further by showing that each subspace of a surface height model corresponds to a subspace of the surface normals. In effect we show that a surface height model may be used to impose a statistical constraint on a field of surface normals. Using this constraint within a shape-from-shading algorithm retains the benefits of using a statistical surface normal model (the ability to locally impose strict irradiance constraints and computational efficiency) but brings a number of additional advantages. A surface height model provides a stricter constraint, since the surface normals must not only satisfy a statistical constraint, but must also correspond to a valid surface. The surface height model is also more space efficient and allows the shape-from-shading algorithm to provide an estimate of the surface height function without calling on an additional surface integration step.

7.1.5 Estimation of Skin Reflectance Properties

Our final contribution was to relax the typical modelling assumptions made in the shape-from-shading literature and in Chapters 3-5 of this thesis, namely of homogeneous, Lambertian reflectance properties. In Chapter 6, we show that the model-based integrability constraint developed in Chapter 5 is sufficiently strict that we can make an estimate of the surface radiance function and a spatially varying albedo term as part of the shape-from-shading process. The idea is to fit a curve to the scattered and noisy sampling of the radiance function which arises from the current facial shape estimate and the measured intensities. We do this at each iteration before enforcing consistency with this function locally, in a similar way to enforcing Lambert's law strictly in the previous three chapters.

We then showed how a parametric reflectance model can be fitted to the recovered shape and reflectance data allowing the reflectance properties to be extrapolated beyond those present in the input image. This allows us to synthesise realistic images under a wide range of conditions of illumination and viewpoint. Finally, we show how the same techniques can be applied to colour images and how, as part of the model fitting process, we can make an estimate of the colour of the illuminant. This provides a route to facial colour constancy.

7.2 Critical Analysis

There are a number of angles from which criticisms may be levelled at the work presented in this thesis. Some of these weaknesses could be addressed by further work or by incorporating the techniques we have developed into other frameworks. We now discuss some of these weaknesses, after which we provide a balanced comparison between the merits of our work and the state of the art in the field of face shape reconstruction.

7.2.1 Weaknesses

Registration and Alignment: Throughout this thesis we have chosen to disregard the problems of registration and alignment. When we construct the statistical surface normal models in Chapters 3 and 4 and surface height models in Chapters 5 and 6, we assume that there is a dense correspondence between every point on each facial surface in the training set. In fact, a particular point in our model does not correspond to the same point on every face (this is not possible in a Cartesian parameterisation). We mitigate the problems caused by this to a certain extent by aligning a sparse set of features (eye centres) and (from Chapter 4 onwards) by cropping the faces such that areas of large spatial variance (such as the occluding contour) are not included. We also do not consider the problem of aligning an input image to the model (or vice versa) when we attempt to recover the facial shape. Again, a coarse manual alignment is all that is used. Although clearly, a fine scale alignment update could be included in the shape recovery process, this was not the theme of the thesis and the fact that a coarse alignment is sufficient for good results is a strength of the approach.

Iterative Fitting: The shape-from-shading, surface integration and reflectance model fitting algorithms described throughout this thesis have been posed in iterative terms. This is largely due to the use of the geometric shape-from-shading framework being used as the foundation for much of the work. Also, given that many of the problems tackled are essentially complex, nonlinear minimisation processes, this is inevitable to a certain degree. The weakness of this approach is that iterative procedures are ill-suited to applications which must operate in real time or on live streams of data. This problem is mitigated to some extent by the fact that all the algorithms presented converge quickly (less than 20 iterations are typically sufficient) and by the numerical efficiency of the steps involved.

Unconstrained Albedo: In Chapters 3, 4 and 6 we present methods for estimating surface albedo. Although we apply photometric constraints to this process (for example,

not allowing albedo values of greater than 1), we do not enforce any global constraints on the form that the estimated albedo maps take. It is believed incorporating such a model would help regularise the albedo estimation process and may significantly improve the results presented in Chapter 6. However, there are a number of persuasive reasons for allowing the estimated albedo to vary arbitrarily. It is not clear that a statistical model is able to accurately describe all variations in albedo. Surface marking such as moles, freckles and facial hair can appear at any position on the surface of a face and could not be efficiently represented by a statistical model. The techniques described in Chapters 3, 4 and 6 are able to recover such surface markings. Yet it is these variations that are very often the distinguishing feature of a face. Further, relying entirely on a statistical model to describe albedo variations would result in a corruption of the estimated shape where the albedo has not been correctly recovered.

Viewpoint Dependency: The statistical models used in this thesis are viewpoint specific, since the surface shape information is projected onto a fronto-parallel viewplane. This means that the shape recovery techniques described are only applicable to images in which the pose of the subject is the same as that of the model. The benefit of this approach is computational efficiency and simplicity since we do not need to estimate pose parameters as part of the shape recovery process or establish correspondence between the model and the image. Nevertheless, the restriction on the range of acceptable inputs is clearly a weakness. In future, there is no reason that our work could not be incorporated into a viewpoint independent model such as the morphable model of Blanz and Vetter [14]. The only significant difference would be a change from a Cartesian to a cylindrical coordinate system and modification to some elements of the shape-from-shading algorithms.

Restrictions on Illumination Conditions: We have assumed throughout this thesis that illumination is provided by a single, distant light source whose direction is known. In Chapter 6 we make this restriction even stronger by requiring input images in which

the viewer and light source directions are coincident. However, techniques do exist for estimating the direction of a point light source from a single image [124], some of which have been developed specifically for faces [5]. Hence, in a real world implementation of our methods, a light source estimation module could be applied to the input image and its results passed to the shape-from-shading module. It is also worth noting that our techniques are fairly robust to poor estimates of the light source direction. For example, the results in Figure 3.18 were achieved with a coarse manual estimate of the light source direction. In addition, the techniques presented for face recognition using linear subspaces only require knowledge of the light source direction at the training stage. At the recognition stage, the probe image may contain illumination of arbitrary complexity which does not need to be estimated explicitly.

7.2.2 Shape-from-shading versus Analysis-by-synthesis

The work described in this thesis in principle tackles the classical shape-from-shading problem, in that we focus on recovering fields of surface normals directly from image intensity and recover surface height through surface integration techniques. The common theme through the work is the use of statistical models to provide constraints which can be incorporated into a shape-from-shading framework. The fields of surface normals recovered using these techniques represent a significant improvement over previous attempts to apply shape-from-shading to images of faces [165, 168] using generic constraints.

However, the use of a statistical model limits the application of the methods we develop to a particular class of objects (faces). Therefore our work must be judged with respect to state of the art face reconstruction approaches as well as generic shape-from-shading algorithms. For this reason, we now provide a comparison between our work and the morphable model used by Blanz and Vetter [14] to recover facial shape.

In contrast to Blanz and Vetter's model, our techniques require a more restrictive setup

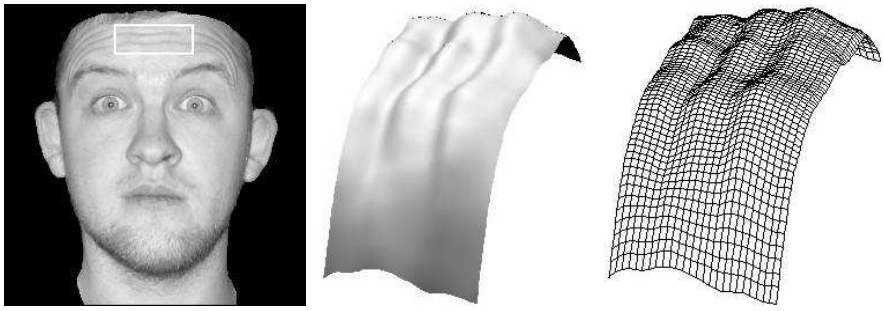


Figure 7.1: Demonstration of recovery of local shape features. The (exaggerated) surface is shown on the right (both as a mesh and a rendered image), recovered from the highlighted region of the image on the left.

(frontal face view and known illumination). However, our work offers a number of potential advantages. Since a field of surface normals may be estimated directly from an image using shape-from-shading, the model parameters are recoverable using a matrix multiplication. This avoids posing the fitting process as a minimisation problem and we therefore avoid the pitfalls of local minima and slow convergence. In addition, data-closeness can be restored locally by rotating each surface normal in the best fit needle-map onto the irradiance cone (or to the position implied by the radiance function estimate in the case of Chapter 6). The advantage brought to bear by the imposition of this local constraint is clearly demonstrated in Figure 7.1. On the right we show the surface recovered by our method from the highlighted region of the image on the left. The surface is displayed as both a wire frame and rendered surface. Discriminating local surface features (in this case wrinkles) not captured by the model have been accurately recovered. The shape of the recovered surface is in good agreement with the highlighted intensity. This would not be possible using a morphable model unless the training set contained examples of similarly positioned wrinkles.

In Table 7.1 we compare the strengths and weaknesses of the two approaches. The potential advantages of our method include: an efficient implementation (all of our shape-from-shading algorithms consist of steps implemented as matrix multiplications), faster

Comparison of Face Reconstruction Methods		
	Vetter and Blanz	Non-Lambertian SFS (Chapter 6)
Illumination:	**Unknown, single point**	Known, single point (could be estimated [124])
Pose:	**Arbitrary**	Frontal
Minimum number of images:	**1**	**1**
Albedo:	Linear Statistical Constraint	**Arbitrary**
Reflectance:	Phong	**Any isotropic, monotonic BRDF.**
Satisfaction of Data-closeness:	Weak	**Strict**
Alignment:	**Manually initialised with sparse feature points, fitting improves**	Manual
Approximate shape recovery time:	4.5 minutes [12]	**10 seconds**

Table 7.1: Comparison between shape-from-shading and analysis-by-synthesis.

convergence (we obtain a stable solution from a variety of initialisations in 20-30 iterations) and the potential for higher accuracy (enforcing data-closeness as a hard constraint avoids model dominance and allows accurate recovery of atypical surface features). Finally, our model is consistent with psychological observations regarding human face perception.

7.3 Future Work

Throughout this thesis we have drawn attention to areas which warrant further research and routes by which the results presented may be improved upon. We conclude the thesis by providing clear directions in which we intend to take the work in future.

7.3.1 Combining Surface Normal Constraints with a Morphable Model

State of the art results in face reconstruction and recognition from single images have been achieved by fitting a morphable model to images in an analysis-by-synthesis framework and using the fitted model parameters to perform recognition [15]. More recently, the spherical harmonic basis morphable model has extended this approach in order to account for arbitrarily complex illumination in the input image [164]. We believe the most fruitful avenue for extending our work would be to combine these state of the art techniques with the methods developed in this thesis. The idea would be to combine the benefits of the two approaches in the hope of exceeding the performance of the techniques taken individually. For example, we would aim to improve the speed of convergence of morphable model fitting through updating the model parameters based on surface normal information extracted directly from the input image. In addition, the accuracy of local, atypical shape features (such as the wrinkles shown above) would be better recovered using local irradiance constraints on the surface normals.

In combining the two approaches, we would also address many of the weaknesses of the work presented in this thesis, for example the lack of pose invariance and assumptions made about illumination.

7.3.2 Constraints on Reflectance Properties

We have already highlighted that the lack of constraint on the recovered albedo values is a weakness of our work. We would like to investigate in future enforcing a linear statistical constraint on the recovered albedo map. Likewise, the method we proposed for recovering the radiance function in Chapter 6 allowed the skin reflectance properties to vary arbitrarily. It is clear that this approach would benefit from additional constraints. These could perhaps be provided by a biophysical skin reflectance model in which the pa-

rameters of the model have biological meaning and, hence, a sensible range of allowable values can be inferred.

7.3.3 Face Recognition and Classification

Much of the work presented in this thesis is ripe for exploitation in more complex face recognition systems. Our consideration of recognition has been limited. Where we have considered it, this has been based on the most simple of linear classifiers (nearest neighbour in a PCA-derived space) or using existing techniques (linear subspace recognition). However, clearly much of the shape and skin reflectance information we extract from single images could prove useful to accurate pose and illumination insensitive recognition. For example, there may be useful biometric information contained in the radiance function estimates or albedo maps. How these different sources of identity information would be combined is also worthy of consideration.

The same argument applies to classification tasks such as gender or age discrimination. There is strong evidence from psychological research that shape information plays a strong role in face classification tasks. There is therefore a significant motivation for attempting to use the shape information we can recover from single images for the purposes of classification.

Appendix A

List of Publications

The research documented in this thesis has resulted in the following publications to date.

Journal

- W.A.P. Smith and E.R. Hancock, "Recovering Facial Shape using a Statistical Model of Surface Normal Direction," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 12, pp. 1914-1930, 2006.
- W.A.P. Smith and E.R. Hancock, "Estimating Facial Albedo from a Single Image," *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 20, No. 6, pp. 955-970, 2006.
- M. Castelán, W.A.P. Smith and E.R. Hancock, "A Coupled Statistical Model for Face Shape Recovery from Brightness Images," *IEEE Trans. on Image Processing*, Vol. 16, No. 4, pp. 1139-1151, 2007.
- M.P. Ewbank, W.A.P. Smith, E.R. Hancock and T.J. Andrews, "The M170 Reflects a Viewpoint-Dependent Representation for Both Familiar and Unfamiliar Faces," *Cerebral Cortex*, to appear, 2007.

Conference

- W.A.P. Smith and E.R. Hancock, "Face Recognition using 2.5D Shape Information," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 2, pp. 1407-1414, 2006.

- W.A.P. Smith and E.R. Hancock, "Recovering Facial Shape in the Presence of Cast Shadows," *Proc. British Machine Vision Conference*, 2006.

- W.A.P. Smith and E.R. Hancock, "Estimating Cast Shadows using SFS and Class-based Surface Completion," *Proc. International Conference on Pattern Recognition*, Vol. 4, pp. 86-90, 2006.

- Y. Li, W.A.P. Smith and E.R. Hancock, "Face Recognition using Patch-based Spin Images," *Proc. International Conference on Pattern Recognition*, Vol. 1, pp. 408-411, 2006.

- W.A.P. Smith and E.R. Hancock, "Facial Shadow Removal," *Proc. SSPR*, pp. 569-577, 2006.

- M. Castelan, W.A.P. Smith and E.R. Hancock, "A Coupled Statistical Model for Face Shape Recovery," *Proc. SSPR*, pp. 898-906, 2006.

- J. Wu, W.A.P. Smith and E.R. Hancock, "Gender Classification using Principal Geodesic Analysis and Gaussian Mixture Models," *Proc. CIARP*, pp. 58-67, 2006.

- M. Castelan, W.A.P. Smith and E.R. Hancock, "Approximating 3D Facial Shape from Photographs Using Coupled Statistical Models," *Proc. CIARP*, pp. 89-98, 2006.

- Y. Li, W.A.P. Smith and E.R. Hancock, "Face Recognition with Region Division and Spin Images," *Proc. CIARP*, pp. 109-117, 2006.

- W.A.P. Smith and E.R. Hancock, "Recovering Facial Shape and Albedo using a Statistical Model of Surface Normal Direction," *Proc. International Conference on Computer Vision*, pp. 588-595, 2005.

- W.A.P. Smith and E.R. Hancock, "Estimating the Albedo Map of a Face from a Single Image," *Proc. International Conference on Image Processing*, pp. 780-783, 2005.

- W.A.P. Smith and E.R. Hancock, "Recovering Facial Shape using a Statistical Surface Normal Model," *Proc. International Conference on Image Processing*, pp. 113-116, 2005.

- W.A.P. Smith and E.R. Hancock, "A Model-based Method for Face Shape Recovery," *Proc. IbPRIA*, pp. 268-276, 2005.

- W.A.P. Smith and E.R. Hancock, "Face Recognition using a Surface Normal Model," *Proc. ICIAP*, pp. 423-430, 2005.

- W.A.P. Smith and E.R. Hancock, "Modelling Surface Normal Distribution using the Azimuthal Equidistant Projection," *Proc. The Eleventh IMA Conference on Mathematics of Surfaces*, pp. 381-394, 2005.

- W.A.P. Smith and E.R. Hancock, "Coupled Statistical Face Reconstruction," *Proc. CAIP*, pp. 153-161, 2005.

- W.A.P. Smith and E.R. Hancock, "Single Image Estimation of Facial Albedo Maps," *Proc. BV&AI*, pp. 517-526, 2005.

- W.A.P. Smith, A. Robles-Kelly and E.R. Hancock, "Reflectance Correction for Perspiring Faces," *Proc. International Conference on Image Processing*, pp. 1389-1392, 2004.

- W.A.P. Smith, A. Robles-Kelly and E.R. Hancock, “Skin Reflectance Modelling for Face Recognition,” *Proc. International Conference on Pattern Recognition*, pp. 210-213, 2004.

- W.A.P. Smith, A. Robles-Kelly and E.R. Hancock, “Single image facial view synthesis using SFS,” *Proc. British Machine Vision Conference*, 2004.

- W.A.P. Smith, A. Robles-Kelly and E.R. Hancock, “Facial View Synthesis from a Single Image using Shape from Shading,” *Proc. International Symposium on 3D Data Processing, Visualisation and Transmission*, pp. 510-517, 2004.

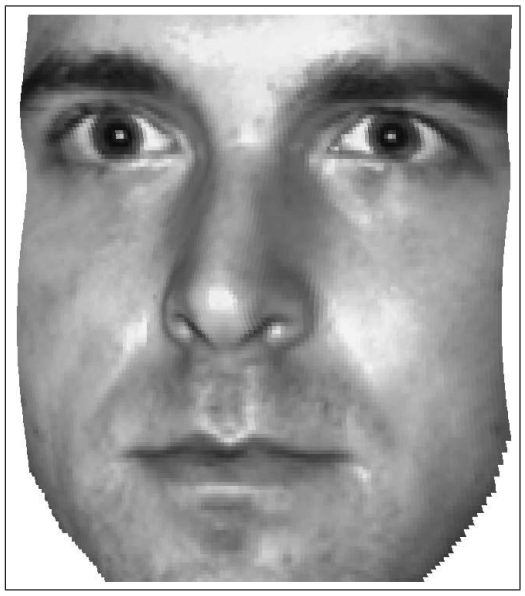
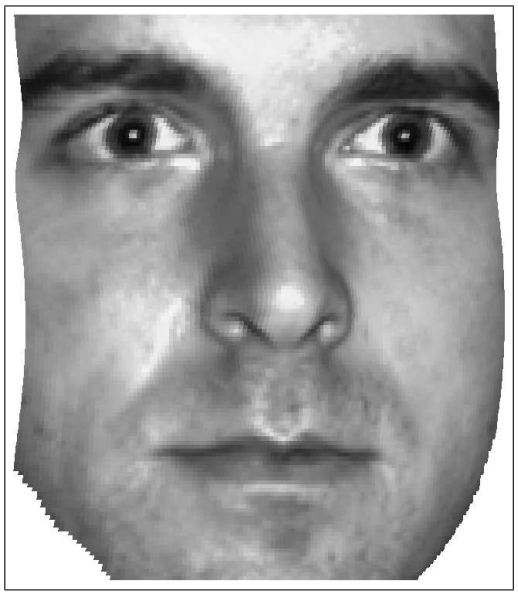
Appendix B

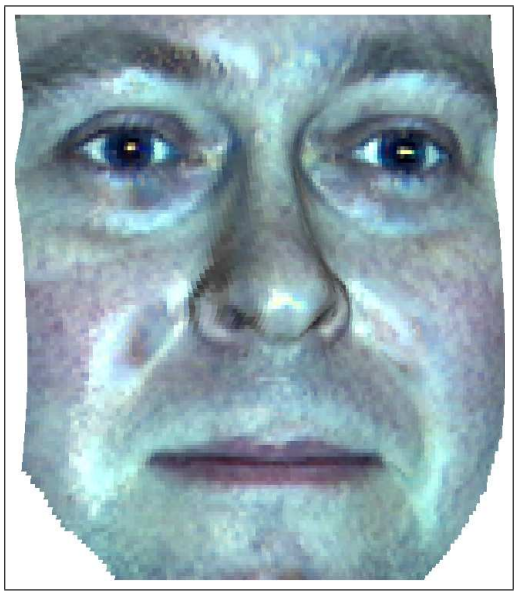
Stereo Pairs

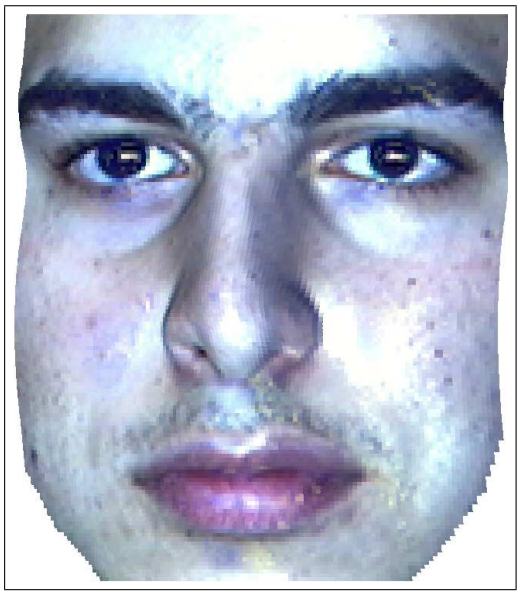
For the sake of interest and as an alternative means of displaying our results, we include a number of stereo pairs. These have been synthesised from the information recovered from one image using the techniques described in this thesis. Stereo pairs exploit the strong perceptual importance of the stereo cue to human vision in order to provide a convincing 3D effect.

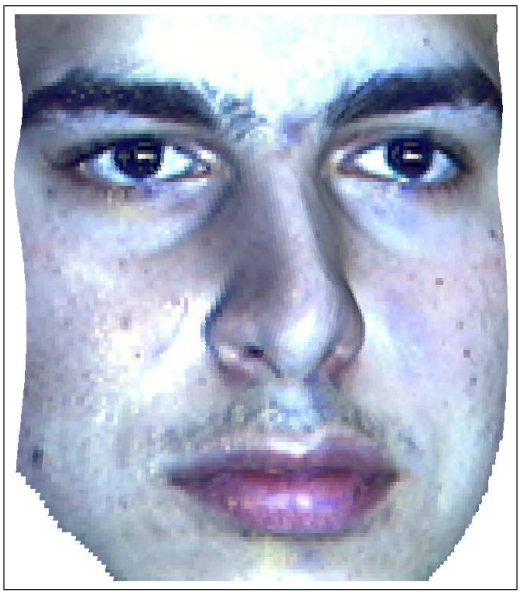
How to view a stereo pair

These are crossed stereo pairs, in which the idea is to view the left image with the right eye and the right image with the left eye. This requires going ‘cross eyed’. The easiest way to do this is to hold your finger about 10 centimetres in front of your face and focus on it. Now slowly remove your finger while viewing the page from a distance of about 30 centimetres. When the pair of images have been successfully fused, a third image should appear between the two in which the face is in a pose mid way between the outer images. This fused view provides a means to visually inspect the quality of the reconstructions. In particular, the nose and eye sockets should give a good sense of varying depth.







Bibliography

[1] USF HumanID 3D Face Database, Courtesy of Sudeep. Sarkar, University of South Florida, Tampa, FL.

[2] Bernard Achermann, Xiaoyi Jiang, and Horst Bunke. Face recognition using range images. In *Int. Conf. Virtual Systems and MultiMedia*, pages 129–136, 1997.

[3] A. Agrawal, R. Chellappa, and R. Raskar. An algebraic approach to surface reconstruction from gradient fields. In *Proc. ICCV*, volume 1, pages 174–181, 2005.

[4] E. Angelopoulou, R. Molana, and K. Daniilidis. Multispectral skin color modeling. In *Proc. CVPR*, pages 635–642, December 2001.

[5] J. J. Atick, P. A. Griffin, and A. N. Redlich. Statistical approach to SFS: Reconstruction of 3D face surfaces from single 2D images. *Neural Comp.*, 8(6):1321–1340, 1996.

[6] Joseph J. Atick, Paul A. Griffin, and A. Norman Redlich. The vocabulary of shape: principal shapes for probing perception and neural response. *Netw.-Comput. Neural Syst.*, 7:1–5, 1996.

[7] S. Avidan and A. Shashua. Novel view synthesis by cascading trilinear tensors. *IEEE Trans. Vis. Comp. Gr.*, 4(4), 1998.

[8] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(2):218–233, 2003.

[9] P. N. Belhumeur, J. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(7):711–720, 1997.

[10] Peter N. Belhumeur, David J. Kriegman, and Alan L. Yuille. The bas-relief ambiguity. *Int. J. Comput. Vision*, 35(1):33–44, 1999.

[11] D. Beymer and T. Poggio. Face recognition from one example view. In *Proc. ICCV*, pages 500–507, 1995.

[12] V. Blanz, C. Basso, T. Poggio, and T. Vetter. Reanimating faces in images and video. In *Proce. EUROGRAPHICS*, pages 641–650, 2003.

[13] V. Blanz, P. Grother, and J. Phillipsand T. Vetter. Face recognition based on frontal views generated from non-frontal images. In *Proc. CVPR*, volume 2, pages 454–461, 2005.

[14] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *Proc. SIGGRAPH*, pages 187–194, 1999.

[15] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(9):1063–1074, 2003.

[16] W. W. Bledsoe. The model method in facial recognition. Technical Report PRI:15, Panoramic research Inc., Paolo Alto, CA, 1964.

[17] A. G. Bors, E. R. Hancock, and R. C. Wilson. Terrain analysis using radar shape-from-shading. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(8):974–992, 2003.

[18] A. Bronstein, M. Bronstein, E. Gordon, and R. Kimmel. 3D face recognition using geometric invariants. In *Proc. International Conference on Audio- and Video-Based Person Authentication*, 2003.

[19] A. Bronstein, M. Bronstein, and R. Kimmel. Three-dimensional face recognition. *Int. J. Comput. Vis.*, 64(1):5–30, 2005.

[20] Sir Thomas Browne. *Religio medici*. Crooke, London, 1645.

[21] V. Bruce and A. Young. *In the Eye of the Beholder: The Science of Face Perception*. Oxford University Press, Oxford, 2000.

[22] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(10):1042–1052, 1993.

[23] Mario Castelán and Edwin R. Hancock. Imposing integrability in geometric shape-from-shading. In *Proc. CIARP*, pages 196–203, 2003.

[24] Mario Castelán and Edwin R. Hancock. Acquiring height maps of faces from a single image. In *Proc. 3DPVT*, pages 183–190, 2004.

[25] Mario Castelán and Edwin R. Hancock. Acquiring height data from a single image of a face using local shape indicators. *Comput. Vis. Image Underst.*, 103(1):64–79, 2006.

[26] S. Chang, M. Rioux, and J. R. Domey. Face recognition with range images and intensity images. *Opt. Eng.*, 36(4):1106–1112, 1997.

[27] R. Chellappa, C. L. Wilson, and S. Sirohey. Human and machine recognition of faces, a survey. *Proc. of the IEEE*, 83(5):705–740, 1995.

[28] K. N. Choi, P. L. Worthington, and E. R. Hancock. Estimating facial pose using shape-from-shading. *Pattern Recognit. Lett.*, 23(5):533–548, 2002.

[29] E. Claridge, S. Cotton, P. Hall, and M. Moncrieff. From colour to tissue histology: physics based interpretation of images of pigmented skin lesions. *Med. Image Anal. J.*, 7:489–502, 2003.

[30] R. Clutterbuck and R. A. R. A. Johnston. Demonstrating how unfamiliar faces become familiar using a face matching task. *Eur. J. Cognit. Psychol.*, 17(1):97–116, 2005.

[31] A. M. Coombes. *Shape Classification: Towards a Mathematical Description of the Face*. PhD thesis, University of London, 1992.

[32] A. M. Coombes, R. Richards, A. Linney, E. Hanna, and V. Bruce. Shape-based description of the facial surface. In *IEE Colloquium on Machine Storage and Recognition of Faces*, volume 17, part 9, pages 1–4, 1992.

[33] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *Proc. ECCV*, pages 484–498, 1998.

[34] T. F. Cootes, C. J.Taylor, D. Cooper, and J. Graham. Training models of shape from sets of examples. In *Proc. BMVC*, pages 9–18, 1992.

[35] T. F. Cootes and C. J. Taylor. Statistical models of appearance for computer vision. Technical report, Division of Imaging Science and Biomedical Engineering, University of Manchester, 2001.

[36] T. F. Cootes, C. J. Taylor, D. Cooper, and J. Graham. Active shape models - their training and application. *Comput. Vis. Image Underst.*, 61:38–59, 1995.

[37] T. F. Cootes, K. N. Walker, and C. J. Taylor. View-based active appearance models. In *Proc. Int. Conf. on Face and Gesture Recognition*, pages 227–232, 2000.

[38] T. F. Cootes, G. V. Wheeler, K. N. Walker, and C. J. Taylor. Coupled-view active appearance models. In *Proc. BMVC*, volume 1, pages 52–61, 2000.

[39] N. P. Costen, T. F. Cootes, G. J. Edwards, and C. J. Taylor. Automatic extraction of the face identity-subspace. *Image Vis. Comput.*, 20:319–329, 2002.

[40] I. Craw and P. Cameron. Parameterising images for recognition and reconstruction. In *Proc. BMVC*, pages 367–370, 1991.

[41] A. R. Damasio, H. Damasio, and G. E. Van Hoesen. Prosopagnosia: Anatomic basis and behavioural mechanisms. *Neuropsychologia*, 2:237–246, 1982.

[42] Kristin J. Dana, Bram van Ginneken, Shree K. Nayar, and Jan J. Koenderink. Reflectance and texture of real-world surfaces. *ACM Trans. Graphic.*, 18(1):1–34, 1999.

[43] J. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(1):1148–1161, 1993.

[44] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, and W. Sarokin. Acquiring the reflectance field of a human face. In *Proc. SIGGRAPH*, 2000.

[45] R. Desimone. Face-selective cells in the temporal cortex of monkeys. *J. Cognitive Neurosci.*, 3(1), 1998.

[46] M. P. Do Carmo. *Differential Geometry of Curves and Surfaces*. Prentice-Hall, New Jersey, 1976.

[47] R. Dovgard and R. Basri. Statistical symmetric shape from shading for 3D structure recovery of faces. In *Proc. ECCV*, volume 2, pages 99–113, 2004.

[48] R.O. Dror. *Surface reflectance recognition and real-world illumination statistics*. PhD thesis, Massachusetts Institute of Technology, 2002.

[49] P. Dupuis and J. Oliensis. An optimal control formulation and related numerical methods for a problem in shape reconstruction. *Ann. Appl. Probab.*, 4(2):287–346, 1994.

[50] J-D. Durou, M. Falcone, and M. Sagona. A survey of numerical methods for shape from shading. Technical Report 2004-2-R, IRIT, 2004.

[51] R. Epstein, P. Hallinan, and A. Yuille. 5 ± 2 eigenimages suffice: An empirical investigation of low-dimensional lighting models. In *Proc. IEEE Workshop Physics-based Vision*, pages 108–116, 1995.

[52] N. I. Fisher. Spherical medians. *J. R. Statist. Soc. B*, 47(2):342–348, 1985.

[53] R. A. Fisher. The use of multiple measures in taxonomic problems. *Ann. Eugen.*, 7:179–188, 1936.

[54] R. A. Fisher. Dispersion on a sphere. *Proc. Roy. Soc., Ser. A*, 217:295–305, 1953.

[55] R.W. Fleming, R.O. Dror, and E.H. Adelson. How do humans determine reflectance properties under unknown illumination? In *Proc. Workshop on Identifying Objects Across Variations in Lighting at CVPR*, Hawaii, December 2001.

[56] P. T. Fletcher, S. Joshi, C. Lu, and S. M. Pizer. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. Med. Imaging*, 23(8):995–1005, 2004.

[57] D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, New Jersey, 2002.

[58] *Frankenstein*. Directed by James Whale. Universal Pictures, 1931.

[59] R. T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 10(4):439–451, 1988.

[60] P. Fua and Y. G. Leclerc. Using 3-dimensional meshes to combine image-based and geometry-based constraints. In *Proc. ECCV*, pages 281–291, 1994.

[61] I. Gauthier, M. J. Tarr, A. W. Anderson, P. Skudlarski, and J. C. Gore. Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat. Neurosci.*, 2(6):568–573, 1999.

[62] A. Georghiades. Recovering 3-d shape and reflectance from a small number of photographs. In *Eurographics Symposium on Rendering*, pages 230–240, 2003.

[63] A. Georghiades, D. Kriegman, and P. Belhumeur. Illumination cones for recognition under variable lighting: Faces. In *Proc. CVPR*, pages 52–59, 1998.

[64] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(6):643–660, 2001.

[65] G. G. Gordon. *Face Recognition from Depth and Curvature*. PhD thesis, Harvard University, Division of Applied Sciences, 1991.

[66] G. G. Gordon. Face recognition based on depth and curvature features. In *Proc. CVPR*, pages 108–115, 1992.

[67] R. L. Gregory. Knowledge in perception and illusion. *Phil. Trans. R. Soc. Lond. B*, 352:1121–1128, 1997.

[68] P. Hallinan. A low-dimensional representation of human faces for arbitrary lighting conditions. In *Proc. CVPR*, pages 995–999, 1994.

[69] P. W. Hallinan, G. Gordon, A. L. Yuille, P. Giblin, and D. Mumford. *Two- and Three-Dimensional Patterns of the Face*. A K Peters, Massachusetts, 1999.

[70] P. J. B. Hancock, V. Bruce, and A. M. Burton. Testing principal component representations for faces. In J. A. Bullinaria, D. W. Glasspool, and G. Houghton, editors, *4th Neural Computation and Psychology Workshop: Connectionist representations*, pages 84–97, London, 1997. Springer.

[71] P. J. B. Hancock, V. Bruce, and A. M. Burton. Recognition of unfamiliar faces. *Trends Cog. Sci.*, 4(9):330–337, 2000.

[72] P. Hanrahan and W. Krueger. Reflection from layered surfaces due to subsurface scattering. In *Proc. Computer graphics and interactive techniques*, pages 165–174, 1993.

[73] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang. Face recognition using laplacianfaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(3):328–340, 2005.

[74] T. Heap and D. Hogg. Extending the point distribution model using polar coordinates. *Image Vis. Comput.*, 14:589–599, 1996.

[75] T. Heseltine, N. Pears, and J. Austin. Evaluation of image pre-processing techniques for eigenface based face recognition. In *The 2nd International Conf. on Image and Graphics, SPIE vol. 4875*, pages 677–685, 2002.

[76] T. Heseltine, N. Pears, and J. Austin. Three-dimensional face recognition: A fishersurface approach. In *Proc. ICIAR*, pages 684–691, 2004.

[77] G. W. Hill. Algorithm 571: Statistics for von Mises' and Fisher's distributions of directions: $I_1(x)/I_0(x)$, $I_{1.5}(x)/I_{0.5}(x)$ and their inverses. *ACM Trans. Math. Software*, 7(2):233–238, 1981.

[78] H. Hill and V. Bruce. A comparison between the hollow-face and 'hollow-potato' illusions. *Perception*, 23:1335–1337, 1994.

[79] H. Hill and V. Bruce. Effects of lighting on the perception of facial surfaces. *J. Exp. Psychol. Human*, 22(4):986–1004, 1996.

[80] H. Hill, P. G. Schyns, and S. Akamatsu. Information and viewpoint dependance in face recognition. *Cognition*, 62:201–222, 1997.

[81] B. K. P. Horn. *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View*. PhD thesis, Massachusetts Institute of Technology, 1970.

[82] B. K. P. Horn. Hill shading and the reflectance map. *Proc. of the IEEE*, 69(1):1447, 1981.

[83] B. K. P. Horn. Height and gradient from shading. *Int. J. Comput. Vis.*, 5(1):37–75, 1990.

[84] B. K. P. Horn and M. J. Brooks. The variational approach to shape from shading. *Comput. Vis. Graph. Image Process.*, 33(2):174–208, 1986.

[85] B. K. P. Horn, R. S. Szeliski, and A. L. Yuille. Impossible shaded images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(2):166–170, 1993.

[86] P. Huber. *Robust Statistics*. Wiley, Chichester, 1981.

[87] T. J. Hutton, B. F. Buxton, P. Hammond, and H. W. W. Potts. Estimating average growth trajectories in shape-space using kernel smoothing. *IEEE Trans. Med. Imaging*, 22(6):747–753, 2003.

[88] H.W. Jensen, S.R. Marschner, M. Levoy, and P. Hanrahan. A practical model for subsurface light transport. In *Proc. SIGGRAPH*, pages 511–518, August 2001.

[89] A. Johnston, H. Hill, and N. Carman. Recognising faces: effects of lighting direction, inversion, and brightness reversal. *Perception*, 21:365–375, 1992.

[90] I.T. Jolliffe. *Principal Component Analysis*. Springer, New York, 2nd edition, 2002.

[91] J.T. Kajiya and T.L. Kay. Rendering fur with three dimensional textures. *Comput. Graphics*, 23(3):271280, 1989.

[92] Takeo Kanade. *Picture Processing System by Computer Complex and Recognition of Human Faces*. PhD thesis, Kyoto University, 1973.

[93] N. G. Kanwisher, J. McDermott, and M. M. Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.*, 17(11):4302–4311, 1997.

[94] M.D. Kelly. Visual identification of people by computer. Technical Report AI-130, Stanford AI Project, Stanford, CA, 1970.

[95] D. A. Kleffner and V. S. Ramachandran. On the perception of shape from shading. *Percept Psychophys.*, 52(1):18–36, 1992.

[96] J. Koenderink and S. Pont. The secret of velvety skin. *Mach. Vis. Appl.*, 14:260–268, 2003.

[97] J. H. Lambert. *Photometria sive de Mensura et Gratibus Luminus, Colorum et Umbrae*. Eberhard Klett, Augsburg, 1760.

[98] A. Lanitis, C. J. Taylor, and T. F. Cootes. Towards automatic simulation of ageing effects on face images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):442–455, 2002.

[99] C. H. Lee and A. Rosenfeld. Improved methods of estimating shape from shading using the light source coordinate system. *Artif. Intell.*, 26:125–143, 1985.

[100] K. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(5):1–15, 2005.

[101] K. M. Lee and C. C. J. Kuo. Shape from shading with a linear triangular element surface model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(8):815–822, 1993.

[102] K. M. Lee and C. C. J. Kuo. Shape from shading with a generalized reflectance map model. *Comput. Vis. Image Underst.*, 67(2):143–160, 1997.

[103] C. R. Leslie, editor. *Memoirs of the life of John Constable*. London, 1843.

[104] S. Z. Li and A. K. Jain, editors. *Handbook of Face Recognition*. Springer, New York, 2005.

[105] C. H. Liu, C. A. Colin, and A. Chaudhuri. Does face recognition rely on encoding of 3-d surface? examining the role of shape-from-shading and shape-from-stereo. *Perception*, 29:729–743, 2000.

[106] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar. *Handbook of Fingerprint Recognition*. Springer, New York, 2005.

[107] K. V. Mardia and P. E. Jupp. *Directional Statistics*. John Wiley and Sons Ltd, 2000.

[108] D. Marr. *Vision*. W. H. Freeman and Co., San Francisco, 1982.

[109] S.R. Marschner, S.H. Westin, E.P.F. Lafortune, K.E. Torrance, and D.P. Greenberg. Reflectance measurements of human skin. Technical Report PCG-99-2, Program of Computer Graphics, Cornell University, 1999.

[110] I. Matthews and S. Baker. Active appearance models revisited. *Int. J. Comput. Vis.*, 60(2):135–164, 2004.

[111] A. M. Megreya and A. M. Burton. Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, 34:865–876, 2006.

[112] T. Moore. *Odes of Anacreon*. John Stockdale, London, 1800.

[113] J. Morton and M. H. Johnson. Conspec and conlearn: a two-process theory of infant face recognition. *Psychol. Rev.*, (98):164–181, 1991.

[114] Y. Moses, Y. Adini, and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):721–732, 1997.

[115] D. Nandy and J. Ben-Arie. Shape from recognition: A novel approach for 3-D face shape recovery. *IEEE Trans. Image Process.*, 10(2):206–217, 2001.

[116] S. Nickell, M. Hermann, M. Essenpreis, T. J. Farrell, U. Krämer, and M. S. Patterson. Anisotropy of light propagation in human skin. *Phys. Med. Biol.*, 45:2873–2886, 2000.

[117] M. Oren and S. K. Nayar. Generalization of the lambertian model and implications for machine vision. *Int. J. Comput. Vis.*, 14(3):227251, 1995.

[118] A. J. O'Toole, T. Vetter, N. F. Toje, and H. H. Bülthoff. Sex classification is better with three-dimensional head structure than with texture. *Perception*, 26:75–84, 1997.

[119] Ovid. *Ars Amatoria*. 2 BC.

[120] P. S. Penev and J. J. Atick. Local feature analysis: A general statistical theory for object representation. *Netw.-Comput. Neural Syst.*, 7(3):477–500, 1996.

[121] X. Pennec. Probabilities and statistics on Riemannian manifolds: basic tools for geometric measurements. In *Proc. IEEE Workshop on Nonlinear Signal and Image Processing*, 1999.

[122] X. Pennec. Probabilities and statistics on riemannian manifolds: A geometric approach. Technical Report RR-5093, INRIA, 2004.

[123] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *Proc. CVPR*, 1994.

[124] A. P. Pentland. Finding the illuminant direction. *J. Opt. Soc. Am.*, 72(4):448–455, 1982.

[125] A. P. Pentland. Shape information from shading: a theory about human perception. In *Proc. ICCV*, pages 404–413, 1988.

[126] P. Jonathon Phillips, Hyeonjoon Moon, Syed A. Rizvi, and Patrick J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(10):1090–1104, 2000.

[127] P.J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone. Face recognition vendor test 2002. Evaluation Report IR 6965, National Institute of Standards and Technology, 2003.

[128] B. T. Phong. Illumination for computer generated images. *Commun. ACM*, 18(6):311317, 1975.

[129] G. Poirier. Human skin modelling and rendering for computer graphics. Master's thesis, University of Waterloo, 2004.

[130] E. Prados and O. Faugeras. A rigorous and realistic shape from shading method and some of its applications. Technical Report RR-5133, INRIA, 2004.

[131] E. Prados and O. D. Faugeras. Unifying approaches and removing unrealistic assumptions in shape from shading: Mathematics can help. In *Proc. ECCV*, pages 141–154, 2004.

[132] M. Proust. *À l'ombre des jeunes filles en fleurs*. 1919.

[133] H. Ragheb and E. R. Hancock. A probabilistic framework for specular shape-from-shading. *Pattern Recognit.*, 36(2):407–427, 2003.

[134] T. Rindfleisch. Photometric method for lunar topography. *Photogramm. Eng.*, 32:262–276, 1966.

[135] A. Robles-Kelly and E. R. Hancock. Estimating the surface radiance function from single images. *Graph. Models*, 67(6):518–548, 2005.

[136] D. Scharstein. *View Synthesis Using Stereo Vision. Lecture Notes in Computer Science (LNCS), volume 1583*. Springer Verlag, 1999.

[137] A. Schopenhaur. *Parerga und Paralipomena: Kleine philosophische Schriften*. Hayn, Berlin, 1851.

[138] J. W. Shepherd, G. M. Davies, and H. D. Ellis. Studies of cue saliency. In G. M. Davies, H. D. Ellis, and J. W. Shepherd, editors, *Perceiving and Remembering Faces*, London, UK, 1981. Academic Press.

[139] L. Sirovich. Turbulence and the dynamics of coherent structures. *Quart. Appl. Math.*, XLV(3):561–590, 1987.

[140] W. A. P. Smith and E. R. Hancock. Face recognition using shape-from-shading. In *Proc. BMVC*, volume 2, pages 597–606, 2002.

[141] W.A.P. Smith, A. Robles-Kelly, and E. R. Hancock. Reflectance correction for perspiring faces. In *Proc. ICIP*, pages 1389–1392, 2004.

[142] J. P. Snyder. *Map Projections–A Working Manual, U.S.G.S. Professional Paper 1395*. United States Government Printing Office, Washington D.C., 1987.

[143] Y. Su, W. Wang, K. Xu, and C. Jiang. The optical properties of skin. *Proc. Soc. Photo-Opt. Inst.*, 4916:299–304, 2002.

[144] J. L. Swerdlow. Unmasking skin. *Natl. Geogr.*, pages 112–134, November 2002.

[145] Y. Tian, J. F. Cohn, and T. Kanade. Facial expression analysis. In S. Z. Li and A. K. Jain, editors, *Handbook of face recognition*, pages 247–276. Springer, New York, 2005.

[146] K. Torrance and E. Sparrow. Theory for off-specular reflection from roughened surfaces. *J. Opt. Soc. Am.*, 57(9):1105–1114, 1967.

[147] N. F. Troje and D. Kersten. Viewpoint-dependent recognition of familiar faces. *Perception*, 28(4), 1999.

[148] P. S. Tsai and M. Shah. Shape from shading using linear approximation. *Image Vis. Comput.*, 12(8):487–498, 1994.

[149] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proc. CVPR*, pages 586–591, 1991.

[150] J. van Diggelen. A photometric investigation of the slopes and heights of the ranges of hills in the Maria of the moon. *Bull. Astron. Inst. Netherlands*, 11, July 1951.

[151] M. J. C. van Gemert, S. L. Jacques, H. J. C. M. Sterenborg, and W. M. Star. Skin optics. *IEEE Trans. Bio-Med. Eng.*, 36(12):1146–1154, 1989.

[152] J. van Wyhe. *Phrenology and the Origins of Victorian Scientific Naturalism*. Ashgate, Aldershot, 2004.

[153] J. W. von Goethe. *Zur Farbenlehre*. Cotta, Tuebingen, 1810.

[154] T. Wei and R. Klette. Height from gradient using surface curvature and area constraints. In *Proc. ICVGIP*, pages 204–210, 2002.

[155] Laurenz Wiskott, Jean-Marc Fellous, Norbert Krüger, and Christoph von der Malsburg. Face recognition by elastic bunch graph matching. In *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, pages 355–396, 1999.

[156] P. L. Worthington and E. R. Hancock. New constraints on data-closeness and needle map consistency for shape-from-shading. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(12):1250–1267, 1999.

[157] P. L. Worthington and E. R. Hancock. Surface topography using shape-from-shading. *Pattern Recognit.*, 34(4):823–840, 2001.

[158] Y. Wu, K. L. Chan, and L. Wang. Face recognition based on discriminative manifold learning. In *Proc. ICPR*, volume 4, pages 171–174, 2004.

[159] Z. Wu and L. Li. A line integration based method for depth recovery from surface normals. *Comput. Vis. Graph. Image Process.*, 43(1):5366, 1988.

[160] J. Xiao, S. Baker, I. Matthews, and T. Kanade. Real-time combined 2D+3D active appearance models. In *Proc. CVPR*, pages 535–542, 2004.

[161] Y. Yan and J. Zhang. Rotation-invariant 3d reconstruction for face recognition. In *Proc. ICIP*, volume 1, pages 156–160, 1998.

[162] M. H. Yang. Face recognition using extended isomap. In *Proc. ICIP*, pages 117–120, 2002.

[163] S. Yeh, O. S. Khalil, C. Hanna, S. Kantor, X. Wu, T. Jeng, and R. A. Bolt. Temperature dependence of optical properties of in vivo human skin. *Proc. Soc. Photo-Opt. Inst.*, 4250:455–461, 2001.

[164] L. Zhang, S. Wang, and D. Samaras. Face synthesis and recognition under arbitrary unknown lighting using a spherical harmonic basis morphable model. In *Proc. CVPR*, pages 209–216, 2005.

[165] R. Zhang, P. S. Tsai, J. E. Cryer, and M. Shah. Shape-from-shading: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21(8):690–706, 1999.

[166] W. Y. Zhao and R. Chellappa. Illumination-insensitive face recognition using symmetric SFS. In *Proc. CVPR*, pages 286–293, 2000.

[167] W. Y. Zhao and R. Chellappa. SFS based view synthesis for robust face recognition. In *4th IEEE Conf. AFGR*, pages 285–292, 2000.

[168] W. Y. Zhao and R. Chellappa. Symmetric shape-from-shading using self-ratio image. *Int. J. Comput. Vision*, 45:55–75, 2001.

[169] W.Y. Zhao and R. Chellappa, editors. *Face Processing: Advanced Modeling and Methods*. Academic Press, U.S.A., 2005.

[170] W.Y. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.

[171] W.Y. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips. Face recognition: A literature survey. Technical Report CAR-TR-948, UMD CfAR, 2000.

[172] Q. Zheng and R. Chellappa. Estimation of illuminant direction, albedo, and shape from shading. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(7):680–702, 1991.