Face Shape Recovery from a Single Image View

Mario Castelán

Submitted for the degree of Philosophy Doctor

Department of Computer Science

THE UNIVERSITY of York

October, 2006

Abstract

The problem of acquiring surface models of faces is an important one with potentially significant applications in biometrics, computer games and production graphics. For such task, the use of shape-from-shading (SFS) is appealing since it is a non-invasive method that mimics the capabilities of the human visual system. In this thesis, our interest lies on the recovery of facial shape from single image views. We make four novel contributions to this area.

We commence by describing an algorithm for ensuring data-closeness and integrability in Shape-from-Shading. The combination of these constraints is aimed to overcome the problem of high dependency on the image irradiance.

Next, we focus on developing a practical scheme for face analysis using SFS. We describe a local-shape based method for imposing a novel convexity constraint. We show how to modify the orientations in the surface gradient field using critical points on the surface and local shape indicators.

Then, we explore the use of statistical models that can be used in conjunction with SFS to reconstruct facial shape. We describe four different ways of constructing the 3D statistical models of faces using Cartesian representations: the surface height, the surface gradient, the surface normal azimuthal angle and finally a model based on Fourier domain basis functions. The models can be fitted to input images using a data-driven procedure which satisfies the image irradiance equation as a hard constraint and is also integrable.

Finally, we propose a coupled statistical model that can be used to recover facial shape from brightness images of faces. We jointly capture variations in intensity and surface shape. The model is constructed by performing principal components analysis (PCA) on sets of parameters describing the contents of the intensity images and the facial shape representations. By fitting the coupled model to intensity data, facial shape is implicitly recovered from the shape parameters.

Contents

List of Figures

Acknowledgement

To commence, I would like to thank my supervisor Professor Edwin R. Hancock, for his support, advice and patience during the course of this work.

Also, I would like to thank Consejo Nacional de Ciencia y Tecnologa (CONACYT), for the financial support provided to pursue my degree at the University of York.

Thanks also go to my colleagues and friends at the Department of Computer Science, with special regards to William A.P. Smith, for his constructive suggestions and helpful discussion.

I should thank as well my examiners, Richard Wilson and Mark Nixon, for the invaluable comments to improve this thesis.

I also wish to thank my family and friends back home, who have been with me all these years despite the distance.

I also mention my friends in the United Kingdom, with special regards to Natalie Ford, Oleg Lisagor, Bernadette Martínez-Hernández, Barry Miller, Richard Catarino, Max Salazar-Lechuga and Esther Vergara-Díaz, who became my family in Britain.

My deepest gratitude I must reserve for my mother Tomasa Castelán, for her constant understanding, support and love.

Declaration

I declare that 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 the next page.

List of Publications

The following is a list of publications that have been produced during the course of my Ph.D. research.

2003

- Castelán, M. and Hancock, E.R., "Imposing Integrability in Geometric Shape-from-Shading", LNCS Vol. 2905. Proceedings of Progress in Pattern Recognition, Speech and Image Analysis: CIARP 03, pp. 193-203.

2004

- Castelán, M. and Hancock, E.R., "Combining Data-Closeness and Fourier Domain Integrability Constraints in Shape-from-Shading", IEEE 17th International Conference on Pattern Recognition, ICPR'04, Volume 3, pp. 115-118.

- Castelán, M. and Hancock, E.R., "Acquiring Height Maps of Faces from a Single Image", IEEE 3D Data Processing, Visualization, and Transmission, 2nd International Symposium on, 3DPVT, pp. 183-190.

- Castelán, M. and Hancock, E.R., "Improving Height Recovery from a Single Image of a Face Using Local Shape Indicators", LNCS Vol. 3211. Proceedings of Image Analysis and Recognition: International Conference, ICIAR 04, pp. 613-620.

- Castelán, M. and Hancock, E.R., "A Method for Re-illuminating Faces from a Single Image", LNCS Vol. 3287. In Progress in Pattern Recognition, Image Analysis and Applications: 9th Iberoamerican Congress on Pattern Recognition, CIARP 04, pp. 267-274.

2005

- Castelán, M. and Hancock, E.R., “Improved Face Shape Recovery and Re-illumination using Convexity Constraints”, LNCS Vol. 3617. 13th International Conference on Image Analysis and Processing, ICIAP 05, pp. 487-494.

- Castelán, M. and Hancock, E.R., “Fitting 3D Cartesian Models to Faces using Irradiance and Integrability Constraints”, LNCS Vol. 3617. 13th International Conference on Image Analysis and Processing, ICIAP 05, pp. 876-883.

- Castelán, M. and Hancock, E.R., “A Comparison of Cartesian Representations for the Principal Component Analysis of three- dimensional Data of Human Faces”, IEEE International Conference on Image Processing, ICIP 05, pp. 1154-1157.

- Castelán, M. and Hancock, E.R., “Estimation of Facial Angular Information Using a Complex-number-based Statistical Model”, LNCS Vol. 3773. 10th Iberoamerican Congress on Pattern Recognition, CIARP 05, pp. 327-338.

2006

- Castelán, M. and Hancock, E.R., “Acquiring Height Data from a Single Image of a Face Using Local Shape Indicators”, Computer Vision and Image Understanding, CVIU 103, Elsevier Press, pp. 64-79.

- Castelán, M., Smith, W.A.P. and Hancock, E.R., “A Full Coupled Statistical Model for Face Shape Recovery”, LNCS Vol. 4109. Proceedings of the International Workshop on Statistical Techniques in Pattern Recognition, SSPR 06, pp. 898-906.

- Castelán, M. and Hancock, E.R., “A Facial Statistical Model from Complex Numbers”, IEEE International Conference on Pattern Recognition, ICPR 06, pp. 235-238.

- Castelán, M. and Hancock, E.R., “A Simple Coupled Statistical Model for 3D Face Shape Recovery”, IEEE International Conference on Pattern Recognition, ICPR 06, pp. 231-234.

- Castelán, M. and Hancock, E.R., “Using Cartesian Models of Faces with a Data-driven and Integrable Fitting Framework”, LNCS Vol. 4141. Proceedings of Image Analysis and Recognition: International Conference ICIAR 06.

- Castelán, M., Smith, W.A.P. and Hancock, E.R., “Approximating 3D Face Shape from Photographs”, LNCS, (to appear in) 11th Iberoamerican Congress on Pattern Recognition, CIARP 06.

- Castelán, M., Smith, W.A.P. and Hancock, E.R., “A Coupled Statistical Model for Face Shape Recovery from Brightness Images”, (to appear in) IEEE Transactions on Image Processing.

Chapter 1

Introduction

This thesis is about the recovery of three dimensional surface shape from intensity images of faces. The intensities of an image (shading) provide a very important source of information for the perception and understanding of 3D objects in computer vision (Marr, 1982; Horn and Brooks M, 1986). There is a wide range of psychophysical experiments that support the idea that the human visual system can extract shape from shading (Koenderink and Van Doorn, 1992a; Erens et al., 1993a; Erens et al., 1993b; Koenderink et al., 1996).

After the publication of Marr's book *Vision* (Marr, 1982), most attention in the literature was paid to the recovery of an object's three dimensional shape as the main goal for a computer vision system. This is due to the theory stated by Marr and Nishihara (Marr and Nishihara, 1978) concerning the way human brains store the perceived world. They proposed a coordinate system centered on the perceived object. Their theory about perception is object-centered. In computational terms, this means storing three-dimensional models of objects that can be later manipulated for the task of recognition (Sutherland, 1979). The impact of the object-centered approach was to focus on methods for recovering the three dimensional shape of objects from image cues, namely, shape-from-X techniques,

i.e. shape-from-shading, shape-from-texture and shape-from-motion (Trucco and Verri, 1998).

As far as it relates to face shape recovery, the problem of acquiring surface models of faces is an important one with potentially significant applications in biometrics, computer games and production graphics. Some of the ways to acquire surface models include the use of range-scanners (Jarvis, 1983; Besl, 1989; Blais, 2003), stereoscopic cameras (Starks, 1995) and structured light sensors (Batlle et al., 1998; Scharstein and Szeliski, 2003). However, one of the most appealing ways is to use Shape-from-Shading (SFS) (Horn and Brooks, 1989), since this is a non-invasive process which seems to mimic the capabilities of the human vision system.

Ideally, an automated SFS system takes as input the intensity image of an object and returns as output the shape of that object (see Figure 1.1). However, the history of SFS has shown that the original problem setting has to be constrained in order to obtain feasible results (Zhang et al., 1999). When the problem is specialized in face shape recovery, the bilateral symmetry of faces (Shimshoni et al., 2003; Zhao and Chellapa, 2000) as well as critical points over the intensity image (Prados et al., 2006) have been explored. The use of these constraints has improved the quality of the results obtained. On the other hand, accuracy can be achieved through the use detailed constraints concerning the recovery of facial shape. For example, statistical methods have been explored for the objective of face shape recovery (Atick et al., 1996; Blanz and Vetter, 2003) from single image views. These approaches require a database of accurate suitable information belonging to objects of the same class, i.e. a database of three-dimensional faces.

Among the topics that have been given particular importance due to the impact they have in the field of SFS we can mention: illumination models, pose estimation, estimation of the reflectance properties, integration techniques and estimation of the illumination direction. In this thesis, we do not directly address these topics. Rather, we focus research

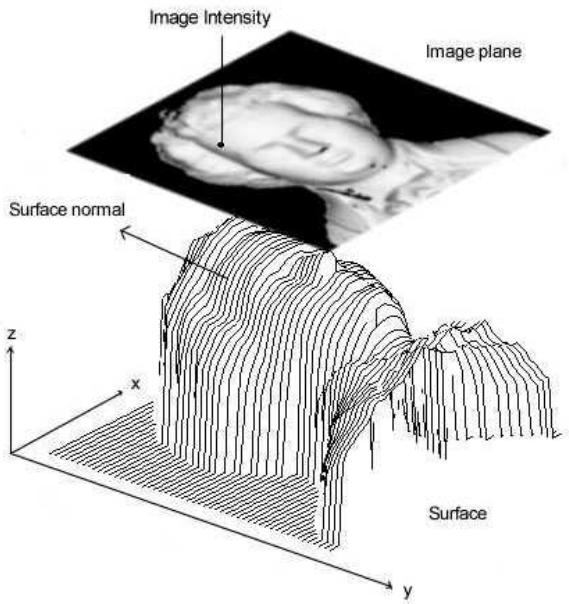

Figure 1.1: SFS is an inverse rendering problem. A 2D intensity image is taken as an input. The underlying 3D surface producing irradiance of the image is the expected output.

efforts towards the problem of recovering facial shape from a single image.

We make four novel contributions to the topic. The first two contributions deal with traditional SFS approaches. First, we combine integrability and image irradiance constraints for needle map stabilization. Second, we propose a new convexity constraint that addresses the problem of imploded facial features in needle maps of faces.

The last two contributions are related to the problem of using statistical models of faces for face shape recovery. Here we explore the use of Cartesian representations for creating statistical models of faces that can be fitted through image irradiance and integrability constraints. Finally, we relate intensity and 3D surface shape into a coupled model for face shape recovery. Here the main idea is to use the shape coefficients of the intensity based model in order to directly obtain 3D surface shape from intensity images of faces.

A summary of the contributions of the thesis is given in the next section.

1.1 Motivation

The recovery of facial shape through SFS techniques has been a topic of research in the computer vision community (Zhao and Chellapa, 2000; Shimshoni et al., 2003; Atick et al., 1996; Dovgard and Basri, 2004; Blanz and Vetter, 2003). Several problems have been encountered depending on the approach used to solve them. It is clear that, from a single image view, a precise height map is difficult to obtain by integration of the field of surface normals delivered by traditional SFS, due to local errors in the direction of the surface normals. The most important problem is that when integrated, the concave/convex ambiguities in the needle-map can lead to the distortion of the topography of the reconstructed face. One of the most serious instances of this problem is that the nose can become imploded (see Figure 1.2). However, this does not imply that the entire gradient field is in error. In fact, some regions on the image do provide directional information that is sufficiently faithful for qualitatively good surface reconstruction. A convexity-enforcement constraint can be designed to help the imploded facial regions arise, resulting in overall improvement for the recovered structure of the face.

On the other hand, the use of statistical models has also proved to be useful for facial shape recovery from brightness images. In the two-dimensional domain, variations in facial appearance can be captured using the eigenfaces technique (Turk and Pentland, 1991). Here a set of aligned facial intensity images are used to construct the eigenmodes. The image data is usually encoded as a Cartesian long-vector by concatenating the rows or columns of the image. However, if a 3D model is to be constructed in an analogous manner from range data, then there exist alternative ways for representing the training data. One of the simplest and most commonly used approaches is to adopt a cylindrical coordinate representation (Atick et al., 1996; Blanz and Vetter, 2003). This representation is used since it captures the linear relations between basis heads. Unfortunately, it can lead to ambiguity since different data can be fitted to the same head-model. An alternative is to

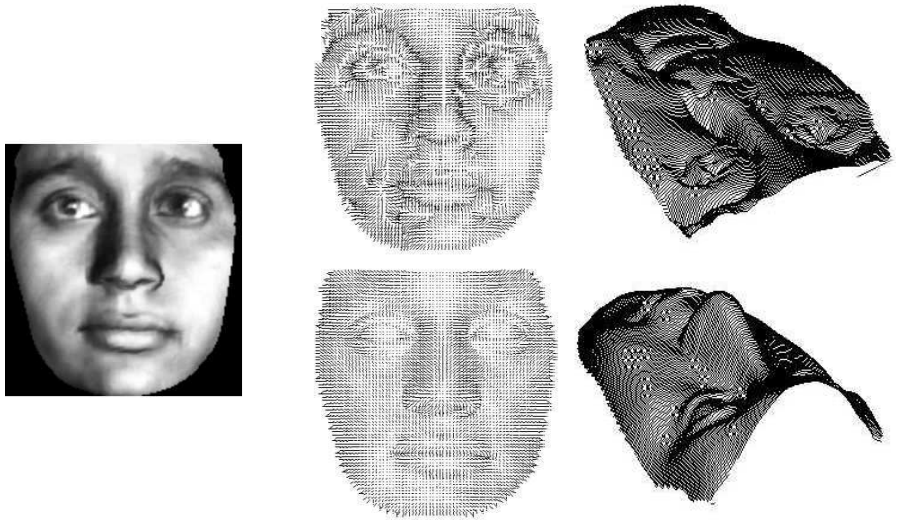

Figure 1.2: The problem of imploded facial features in SFS. The first column shows an intensity image of a face. The next two columns present needle maps and height maps, respectively. The upper row shows results obtained using the Worthington and Hancock SFS algorithm (Worthington and Hancock, 1999). The bottom row shows ground truth data. Note the problem of imploded facial features over the upper row figures.

use a Cartesian representation, i.e. height maps (Dovgard and Basri, 2004), in which each surface point is specified by its (x, y, z) coordinates, where the z-axis is in the direction of the viewer. A general drawback of Cartesian coordinates is the high spatial variance caused by face features such as nose, mouth and eyes.

Height maps, however, are not the only way for representing 3D information in Cartesian coordinates. Alternative encodings can be drawn from 2.5D information such as the partial derivatives of a surface. Although the 2.5D representation is less appealing since it must be integrated to recover a surface, because of the image irradiance equation the 2.5D representation is closer to the raw image brightness data than a height surface.

Although statistical approaches have the drawback of requiring a database of faces to train the model, the accuracy of the recovered shapes is enough to generate novel views far from the range of the viewer's direction. To this end, a minimization scheme has to be carried out in order to find the optimum set of parameters for the model. This problem

has been addressed by minimizing the distance between rendered views from recovered surfaces and input images (Atick et al., 1996; Blanz and Vetter, 2003). Unfortunately, the minimization of this distance is badly affected by the presence of local minima. This means that exhaustive search methods must be used, which sacrifices efficiency and simplicity. This problem suggests that research directions ought to be focused on exploring alternative ways to relax this minimization problem.

1.2 Contribution

The contributions of this thesis can be divided into two groups: the ones related to traditional SFS and the ones related to statistical SFS. As far as traditional SFS is concerned, the contributions are:

1 **The development of an SFS scheme that combines integrability and data closeness.**

Here the combination of data-closeness and integrability constraints aims at overcoming the problem of high dependency on image irradiance. Data closeness is ensured by constraining surface normals to fall on an irradiance cone, whose axis points in the light source direction and whose apex angle varies with iteration number. Integrability is ensured by projecting the non-integrable set of surface normals to the nearest integrable one by globally minimizing the distance between them in the Fourier domain. As a result, the new method recovers needle maps that are smooth and integrable as well as improving height surface stability.

2 **The inclusion of a new constraint that enforces convexity in facial SFS.**

Here we aim at developing a practical scheme for face analysis using SFS. Traditional SFS methods have a tendency to recover surfaces in which convex features such as the nose are imploded. This is a result of the fact that subtle changes in the elements of the field of surface normals can cause significant changes in the corresponding integrated surface. To overcome this problem, we show how to modify the orientations in the surface gradient field using critical points on the surface and local shape indicators. This results in greatly improved height reconstructions and more realistic surface re-illuminations.

The contributions that are related to statistical SFS are:

3 **The analysis of alternative Cartesian representations for building statistical models of faces.**

Here we describe four different ways of constructing the required three-dimensional statistical models of faces using Cartesian representations. The representations studied are the surface height, the surface gradient, the surface normal azimuthal angle and, finally, a model based on Fourier domain basis functions. The models can be fitted to image brightness data using an alternative non-exhaustive parameter adjustment procedure. This procedure ensures that the recovered surface satisfies the image irradiance equation as a hard constraint and is also integrable.

4 **The development of coupled statistical models for face shape recovery.**

Here we aim at generating accurate shape from out-of-training-sample intensity images in an efficient way. To this end, we focus on the problem of developing a coupled statistical model that can be used to recover facial shape from brightness images of faces. We capture variations in both intensity and surface shape

using a coupled statistical model (Cootes et al., 1998). This model is constructed by performing principal components analysis on sets of parameters describing the contents of the intensity images and the facial shape representations. By fitting the coupled model to intensity data, facial shape is implicitly recovered from the shape parameters.

1.3 Thesis overview

A review of the literature relevant to this thesis is presented in Chapter 2. This review covers classical approaches in SFS as well as recent advances in facial shape recovery from a single image view.

Chapter 3 describes an iterative algorithm that combines data closeness and integrability in SFS. The aim of this method is to relax the problem of overreliance in the image irradiance equation by introducing an integrability condition into the geometric SFS framework.

In Chapter 4, we introduce a procedure to enforce convexity on SFS-acquired needle maps. This can be used to recover shape structure from frontal images of faces. To do this, we use local shape indicators to re-direct the field of surface normals.

An exploration of statistical models of faces based on Cartesian coordinates is presented in Chapter 5. Here we experiment with directional and height data in order to build statistical models of faces. We develop data-driven procedure to fit the model on brightness data. We also explore how the integrability constraint can be integrated into the fitting method.

In Chapter 6, we combine statistical models of intensity and surface shape into a coupled statistical model. In this way, we jointly capture variations in intensity and the surface shape representations. The model is constructed by performing principal compo-

nents analysis on sets of parameters describing the contents of the intensity images and the facial shape representations. By fitting the coupled model to intensity data, facial shape is implicitly recovered from the shape parameters.

Finally, Chapter 7 highlights directions for future work and comments on ways that the methods presented here can be extended and improved.

Chapter 2

Literature Review

In SFS, a field of surface normals and hence the height-map of the viewed surface is reconstructed from a single image. The SFS process was identified by Marr as key in the computation of the 2.5D sketch (Marr, 1982), and was studied in depth by Horn (Horn and Brooks, 1989). The topic has also been the focus of recent research in the psychophysics literature. Over the years, a number of SFS methods have been developed by various researchers (Zhang et al., 1999). The general conclusion is that the solution of the SFS problem is somewhat elusive when the input image departs from Lambert's law, which is the case of the majority of real world imagery.

Most recent SFS methods have made efforts towards constraining the problem to a specific domain. This is the case when exploiting symmetry (Zhao and Chellapa, 2000; Shimshoni et al., 2003) or using critical points (Prados et al., 2006) for facial SFS. However, only when more detailed constraints are applied, as in statistical SFS (Atick et al., 1996; Blanz and Vetter, 2003) or photometric stereo (Georghiades et al., 2001; Forsythe and Ponce, 2001), do the results show the necessary accuracy demanded by model-based face recognition systems. Photometric stereo requires at least three images of the same object illuminated from different points while the statistical SFS uses a database of accu-

rate training information for objects of the same class, i.e. a database of 3D faces.

The literature review in this chapter covers recent advances in the development of SFS. We start with a brief definition of the SFS problem in Section 2.1. We present a concise survey on psychophysical support for SFS in Section 2.2. An overview of classic computational SFS methods is presented in Section 2.3. Finally, in Section 2.5 we review the state-of-the-art on SFS for facial shape recovery.

2.1 Shape-from-shading

In brief, SFS aims to solve the image irradiance equation, $E(x, y) = R(p, q, \mathbf{s})$, where E is the image brightness value of the pixel with position (x, y), and R is a function referred to as *the reflectance map* (Horn, 1997). The reflectance map uses the surface gradients $p = \frac{\partial Z(x,y)}{\partial x}$ and $q = \frac{\partial Z(x,y)}{\partial y}$ together with the light source direction vector $\mathbf{s}$ to compute a brightness estimate which can be compared with the observed brightness, using measure of error.

A Lambertian surface exhibits matte or diffuse reflectance. Though most SFS methods assume Lambertian reflectance, there are also more sophisticated reflectance models which deal with the issue of specularity (Beckmann and Spizzichino, 1963; Torrance and Sparrow, 1967). For a survey of reflectance models, see (Schlick, 1994). Recently, some efforts for handling highlights in SFS schemes have been proposed (Ragheb and Hancock, 2003).

If the surface normal at the image location (x, y) is $\mathbf{n} = (p, q, -1)^T$, then under the Lambertian reflectance model, with a single light source direction, no inter-reflections and constant albedo, the image irradiance equation becomes

$$E(x, y) = \mathbf{n} \cdot \mathbf{s}. \tag{2.1}$$

In other words, the SFS problem is the one of recovering the surface that, after interaction with the environment (illumination conditions, objects' reflectance properties, inter-reflections), produces the radiances perceived by human eyes as intensities. In general, though, SFS is an under-constrained problem since the two degrees of freedom for surface orientation (slant and tilt) must be recovered from a single measured brightness value. In the following section, we provide a review of psychophysical experiments that support the use of SFS in living systems.

2.2 Psychophysical support for SFS

There is considerable body of research on psychophysics in relation to SFS. This is mainly motivated since the way living systems work can assist the design of artificial systems.

At the onset in the development of SFS, it was assumed that the human visual system is able to perform accurate surface recovery from shading surfaces (Horn, 1986). However, a series of experiments started by Mingolla and Todd (Mingolla and T., 1986) as well as Stevens (Stevens and Brookes, 1987) and continued by Koenderink and his coworkers (Koenderink and Van Doorn, 1992a) revealed some interesting features of human perception pertinent to SFS.

These experiments largely consisted in using gauge figures and monochrome views of piecewise-smooth objects. Humans were asked to determine if the figures were tangent to the surface in the picture in order to obtain samples of gradient information as perceived by humans. The information obtained from the experiments was transformed to the slant and tilt angles of the surface normals estimated by the subjects. With gradient estimates at hand, a surface integration process was carried out to recover the shape perceived by the subjects.

The conclusion from (Koenderink and Van Doorn, 1992a) was that the human visual

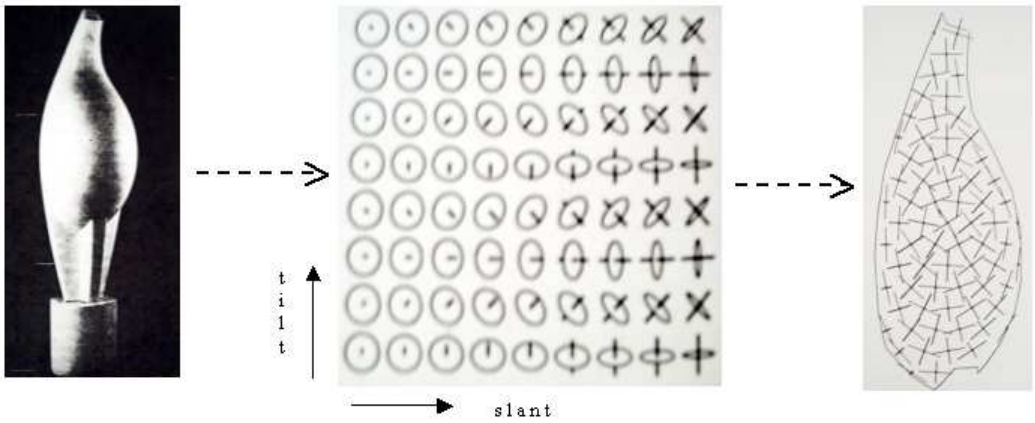

Figure 2.1: Example of the tests realized in Koenderinks experiments (Koenderink and Van Doorn, 1992a). The image in the left corresponds to the picture to be gauged. The image in the middle shows example of the gauge figure used in the experiments. The image in the right shows the final sampled set of normals as perceived by one of the subjects in the experiments.

system is not able to accurately recover quantitative height information from shading, though qualitative data seemed to be consistent with the stimuli. This can be explained due to the considerable degree of variation for the slant angle estimates among different subjects, while the tilt estimation remained consistent with the image. Also, it was deduced that human observers perceive the silhouette (i.e contours, local surface orientations) of objects rather than accurate depths.

Ramachandran's research supported these suggestions and demonstrated that the human visual system recovers depth information using not only shading, but also outlines, features and prior knowledge about the observed objects (Ramachandran, 1988). Also in (Barrow and Tanenbaum, 1993) experiments showed that human SFS is affected by stereoscopic processing, since the line drawing of the shading pattern seemed to play a central role in its interpretation. Further experiments demonstrated that the human visual system cannot recover the shape of objects when only shading information is present, since humans are not capable of classifying such objects into basic groups of shapes, i.e.

cylinder-like, sphere-like, saddle-like, ellipsoid-like (Erens et al., 1993b). The shape index, a measure proposed in (Koenderink and Van Doorn, 1992b), was used for evaluating subject performance. Alternative experiments carried out by (Koenderink et al., 1996) showed the importance of edges over shading. By gradually adding edge cues to objects similar to the ones used in (Erens et al., 1993b), it was shown that the subjects were able to better classify objects than if shading alone was used.

Of the results presented in this brief review of psychophysical experiments, the most important ones suggest that, though the human visual system is not capable of recovering exact three-dimensional information solely on the basis of shading, it does recover broadly similar surfaces to the ones presented as stimuli. This is the case not only for simple convex surfaces (Koenderink and Van Doorn, 1992a), but also for more complicated surfaces (Koenderink et al., 1996).

If we compare the behavior of the human visual system to the way a computer vision system functions, we cannot expect an artificial SFS program to deliver accurate height information, but rather an approximation of the surface that generates the input image. However, the performance can be improved using photometric stereo (Forsythe and Ponce, 2001; Georghiades et al., 2001) and statistical SFS (Atick et al., 1996; Blanz and Vetter, 1999; Dovgard and Basri, 2004).

2.3 Traditional SFS approaches

Numerous methods for SFS have been proposed in the literature. The most common classification of the methods is based on how they specify the problem. In the following subsections we use a classification similar to the one presented in (Zhang et al., 1999).

Global minimization methods

In global minimization approaches, the problem is minimizing cost functionals (Ikeuchi and Horn, 1981; Brooks and Horn, 1985; Frankot and Chellappa, 1988; Horn, 1989; Zheng and Chellapa, 1991). The complete set of image intensities contribute to the surface reconstruction. The minimization of the functional is performed using *the calculus of variations*. In fact, between 1981 and 1991, SFS was considered as a variational problem and was frequently posed as that of minimizing cost functionals that capture constraints on the gradient field (Horn and Brooks M, 1986). In practice the method uses iterative schemes in the discrete domain.

Usually, the cost functionals encapsulated the criteria of surface smoothness, integrability, compliance with the image irradiance and unit normal. If isolated, these functionals are not sufficient to constrain the problem. Therefore, a combination of functionals is often needed. Once a suitable set of constraints is selected, the next step is to choose the method for minimizing them.

The main drawback of the cost functionals associated with the variational approach is its dependency on the correct choice of the weighting factor for each constraint. If the combination of functionals is placed into a regularization framework, as proposed by (Brooks and Horn, 1985), then the method tends to oversmooth the recovered needle map, thus losing surface detail. Also, solving the differential equations can be difficult if the problem is not discretized correctly.

In (Worthington and Hancock, 1999), a method for more accurate needle map calculation was presented. Here the image irradiance equation was treated as a hard constraint by forcing every recovered surface normal to lie on its nearest position on the reflectance cone. The needle maps delivered by this geometric framework have proved to be useful for computing topographic characteristics that can be used for 3D object recognition from 2D brightness images (Worthington and Hancock, 2001).

Propagation methods

Propagation methods commence with points of known height surface from which shape information is extended across the whole surface. The method of characteristic strips (Horn, 1970) is the classic propagation method. A characteristic strip is a line in the image from which depth and orientation can be calculated, provided that these are known at the starting point of the line. Horn solved this problem by using a system of five ordinary differential equations.

Later, a solution for the shading differential equations governing surface shading appeared in (Rouy and Tourin, 1992). Here the use of viscosity solutions and Hamilton-Jacobi-Bellman equations was proposed. Additionally, Oliensis and Dupuis formulated SFS as an optimal control problem, which can be solved using numerical methods, provided that information about the location and the height of singular points on the surface is known (Dupuis and Oliensis, 1994).

Alternatively, Bichsel and Pentland proposed a two-step method (Bichsel and Pentland, 1992), which is a simplified version of that of Oliensis and Depuis. First, using directional derivatives, they determined possible surface slopes in various directions of the image (using an 8-neighborhood), by rotating the image to align the tilt angle of the light source direction with one of the directions of the 8-neighborhood. The subsequent reconstruction process iteratively computes a new height value by checking which of the eight calculated slopes most strongly propagates the surface towards the illumination direction. The maximal slope is used to update the height estimated. The final recovered surface is rotated back to its original pose.

More recently, Prados and Faugeras (Prados et al., 2006) proposed a new method for the Lambertian SFS problem based on the notion of Crandall-Lions viscosity solution. This method requires the knowledge of the solution (the surface to be reconstructed) only on some part of the boundary or of the singular set (the set of the points at maximal

intensity). Their work draws on the notion of viscosity solutions and the work of Dupuis and Oliensis dealing with classical solutions and value functions.

Local methods

Local SFS approaches assume local geometry for every point of surface of the object under study. These methods only calculate surface orientation, which can be integrated later to obtain a corresponding height map. In (Lee and Rosenfeld, 1985), a method for the recovery of surface orientation as well as the estimation of the light source direction was proposed. This is based on the work of Pentland (Pentland, 1982). They approximate local surface regions by spherical patches. For such task, a transformation of system coordinates is needed. First, they calculate the slant and tilt of the surface based on the light source direction. Second, they project these directions back to the viewer coordinate system. The method does not seem to function well when the assumption of local spherical surface is violated.

2.4 Surface integration from gradient fields

In SFS and photometric stereo, surface reconstruction from an estimated gradient field is required. Unfortunately, the gradient field is not always integrable, i.e. it has non-zero curl. This is why the surface integration, or height-from-gradient problem has received considerable attention in its own right. Integration techniques can be classified as being local or global. Local techniques are based on curve integrals and can use different integration paths and neighborhoods (Wu and Li, 1988; Bors et al., 2003; Robles-Kelly and Hancock, 2004). Global methods deal with the minimization of error functionals (Horn, 1989; Frankot and Chellappa, 1988). Global approaches tend to be more robust to noise and recover smoother surfaces. An interesting analysis concerning gradient fields integra-

tion methods is given in (Klette and Schluens, 1996).

2.4.1 Local integration methods

Local integration methods define paths of integration along which they propagate height information. One of the first local integration methods in the literature is the two-point technique (Coleman and Jain, 1982), which specifies a cross-like path starting in the middle of the gradient field. The surface height can be recovered by considering the surface normal vectors at the two adjacent points of a given locality, computing the average tangent through the given point, and interpolating the height and the surface normals. An extension of this method using eight adjacent points is described in (Healey and Jain, 1984).

In (Wu and Li, 1988), a similar local integration method has been developed. They start with initial height values, which are propagated according to a local approximation rule (e.g., based on the 4- neighborhood) using the given gradient data. This operation can be repeated using different scan algorithms. The resulting height values are determined by averaging operations. However, initial height values have to be provided.

Alternatively, a geometric height recovery approach for synthetic aperture radar (SAR) images of terrain has been proposed in (Bors et al., 2003). Here known and unknown height values are first localized and then the unknown height values are calculated using a gradient updating algorithm. A more sophisticated approach for the integration of gradient fields is proposed in (Robles-Kelly and Hancock, 2004). Here spectral-graph theory is used to find an integration path using the leading eigenvector of the transition matrix. The needle map is characterized by a this matrix, which is computed from the sectional curvature for different locations on the surface. The path of integration is finally located using a graph seriation method. A severe restriction for this method is that it becomes highly computationally demanding in resources for large images.

In general, local integration approaches suffer the drawback of unstable error propagation if the gradient field is noisy. The locality of the computations propagates errors along the integration path. As a result, these methods strongly depend on data accuracy.

2.4.2 Global integration methods

Global integration methods minimize an error functional. The classic global integration method is the Frankot and Chellappa's algorithm (Frankot and Chellappa, 1988). They enforce integrability by orthogonally projecting the non-integrable field onto a vector subspace spanning the set of integrable slopes. However, their method is dependent on the choice of basic functions.

Alternatively, the integrability constraint and the surface curvature and area constraints have been combined into a single functional, which is then minimized (Wei and Klette, 2002). This enforces changes in the height map to be more regular. To solve the minimization problem, Fourier domain basis functions are employed. Nonetheless, associated weighting parameters had to be carefully chosen in order to avoid over-smoothing the integrated surface.

Recently, Agrawal and Chellappa have proposed an algebraic approach to enforce integrability in the discrete domain (Agrawal et al., 2005). They formulate enforcing integrability as the solution of a single linear system over the image. As this system is generally undetermined, they provide conditions under which the system can be solved and a method to obtain those conditions based on graph theory. Their approach is non iterative and possesses the property of local error confinement.

2.5 Face shape recovery using SFS

As noted earlier, in contrast to the human visual system, it seems that computer vision systems encounter difficulty in estimating the tilt of a surface from a single image than its slant (Koenderink and Van Doorn, 1992a). When propagated iteratively, these errors can result in poor recovery of the topography of the recovered surface. For instance, because of the concave-convex ambiguity (Gregory, 1997), there may be regions where the sign of the curvature becomes reversed. Moreover, since only a single image is to hand, there is no additional evidence that can be used to correct the resulting shape-errors. For face analysis, the use of SFS has proved to be an elusive task, since the concave-convex ambiguity can result in the inversion of important features such as the nose. To overcome this problem, domain specific constraints have been used.

2.5.1 Constraining the problem

Symmetry is an important cue for the recovery of surface shape from images of faces. Zhao and Chellappa (Zhao and Chellapa, 2001) have exploited the bilateral symmetry of faces in SFS. Their framework deals with Lambertian surfaces with unknown and potentially varying albedo. The problem of source-from-shading is also addressed. Orthographic projection and frontal views are also assumed. Two image irradiance equations are used, one is the standard equation used in SFS and the other is a self-ratio image irradiance equation. The latter equation relates to the self-ratio image which is defined as the ratio of two halves of the input image to light source and surface shape. This symmetric scheme is proved to have a unique global solution. The method works well with synthetic surfaces but does not perform well when tested on images of faces with natural albedo variation. The method has nonetheless been used for effective face recognition (Zhao and Chellapa, 2000).

Similarly, Shimshoni et al. (Shimshoni et al., 2003) have presented a shape reconstruction method for bilaterally symmetric surfaces from a single image. The basic idea is that an image taken from a general, non frontal view point, under non-frontal illumination can be regarded as a pair of images. Each image of the pair is one half of the object, taken from different viewing positions and with different lighting directions. Geometric and photometric information can be used in order to obtain a dense correspondence map between pairs of symmetric points, for the purposes of dense shape recovery. Unknown lighting and viewing parameters are also recovered by the process. Lambertian surfaces, unknown constant albedo and weak perspective projection are assumed. This method gives results whose quality depends on whether input images are frontal or non-frontal views. Better results are obtained from non-frontal-view examples.

A method for the integration of non-linear holonomic constraints in physics-based deformable models has been introduced in (Samaras and Metaxas, 2003), where the use of Lagrange multipliers and a Baumgarte stabilizer allow for the robust integration of these constraints. This unifying approach can be used for the problem of shape and illumination direction estimation from shading. The method can be applied to both Lambertian and non-Lambertian images and does not require knowledge of the illumination direction. Although the authors mostly realize experiments over the standard set of SFS images presented in (Zhang et al., 1999), they also use one example applied to the recovery of facial shape from a close-to-frontal intensity image.

The SFS approach of Prados and Faugeras (Prados et al., 2006) has also been applied to facial shape recovery. Since their SFS scheme relies on the existence of a unique critical point, the input images has to be acquired under carefully controled conditions. A single camera with a basic flash in a dark room is used to assure that the distance of the face to the camera and the focal length are sufficiently small. The effect of this setting is a resulting photograph with the brightest point at a single image location. Their SFS

scheme has proved to work well with symmetric and non-symmetric surfaces, regardless of facial pose. It gives qualitatively good facial reconstructions.

Photometric stereo has also proved to be effective in the recovery of facial shape for recognition purposes. A generative appearance-based method for recognizing human faces under variations in lighting and view point has been described in (Georghiades et al., 2001). The method exploits the fact that the set of images of an object in fixed pose and under all possible illumination conditions is a convex cone in the space of images. Using at least six training images of each face taken with different lighting directions, the shape and albedo of the face can be reconstructed.

2.5.2 The statistical approaches

In the two-dimensional domain, variations in facial appearance can be captured using the eigenface technique (Turk and Pentland, 1991). Here a set of aligned facial intensity images are used to construct the eigenmodes. The image data is usually encoded as a Cartesian long-vector by concatenating the rows or columns of the image. Turk and Pentland were among the first to explore the use of principal components analysis for face recognition (Turk and Pentland, 1991). They used the technique described by Kirby and Sirovich. (Kirby and Sirovich, 1990) to render the method efficient.

Cootes and Taylor (Cootes et al., 1998) develop a novel method of interpreting images using an Active Appearance Model (AAM) that can be used for locating deformable objects in images. An AAM contains a statistical model of the shape and grey-level appearance of the object of interest which can generalize to almost any valid example. During a training phase, the relationship between model parameter displacements and the residual errors induced between a training image and a synthesized model example is learnt. A new image is matched by measuring the current residuals and the model is then used to predict changes to the current parameters.

As far as the 3D domain is concerned, Atick et al. (Atick et al., 1996) were the first to propose a statistical SFS framework based on a low dimensional parametrization of facial surfaces. Principal components analysis was used to derive a set of 'eigenheads' which compactly captures 3D facial shape. They minimized the error between the rendered surface and the observed intensity, defining an irradiance constraint on the low parametric space. Unfortunately, it is surface orientation and not depth which is conveyed by image intensity. Therefore, fitting the model to an image equates to a computationally expensive parameter search.

Blanz and Vetter extended the work of Attick et al. and decoupled surface texture from shape by performing principal component analysis on the two components separately (Blanz and Vetter, 1999). Using full facial feature correspondences in cylindrical coordinates, they develop a model that could be fitted to image data. Their model can be used to make reasonable estimates of the full 3D shape and texture of a face even when only a single picture is available. When applying the method to several images of a person, the reconstructions approach the quality of those obtained with laser range scanners. They have also developed a morphable model for face recognition (Blanz and Vetter, 2003). Their framework can be used regardless of pose and illumination changes, but linear combinations of shape and texture have to be formed separately for the eyes, nose, mouth and the surrounding area. As a starting point in the fitting process, an initial 3D shape has to be aligned to the input image. The alignment procedure starts with a manually assigned set of fiducial points, from which pose and illumination parameters are calculated. The fitting procedure attempts to minimize the error between the input image and the rendered reconstruction, which is recovered through a vector of concatenated parameters. These parameters represent facial shape and texture as well as calibration data such as pose angles, 3D translation, focal length, ambient and directed light intensities, color contrast and offsets of color channels. The results delivered by fitting this morphable

model have proved to be accurate enough to generate photo-realistic views from an input image, though sacrificing efficiency and simplicity. In fact, the approach of Blantz and Vetter have become a benchmark for current face shape recovery systems.

Nandy and Ben-Arie have proposed a different approach to solve the problem attempting to learn the relationship between 3D shape and image intensity for a number of face parts (Nandy and Ben-Arie, 1999). They explore the idea of shape-from-recognition, i.e. the idea that pre-recognized face parts can constrain the space of possible solutions for the image irradiance equation, allowing the recovery of the 3D structure of a specific part. Using back-propagation neural networks, the principal component coefficients of the Lambertian images are mapped to a set of principal component coefficients that represent depth information. The recovery of a complete face is performed by merging face parts to minimize the squared error.

More recently, Dogvard and Basri (Dovgard and Basri, 2004) have combined the statistical constraint of Atick et al. (Atick et al., 1996) and the symmetry constraint of Zhao and Chellappa (Zhao and Chellapa, 2000) into a single SFS framework. Here the aim was to express the surface gradient in terms of a set of deformation coefficients. This allows shape-from-shading to be transformed into a linear system of equations. This system can be simply solved for the shape coefficients and used to reconstruct the height function for the face. Although it uses a statistical model, the method is efficient. However, facial asymmetry produces significant error in the recovered surfaces.

Finally, Smith and Hancock have shown how statistical models can be constructed in the surface normal domain using the azimuthal equidistant projection (Smith and Hancock, 2005a). Problems with the representation of angular data are overcome by transforming the surface normals to Cartesian points. The modes of facial shape variation are captured using a point distribution model. The model can be simply fitted to image brightness data using geometric constraints on the direction of the surface normals that

result from Lambert's law.

2.6 Remarks

From the review of traditional SFS approaches for facial shape recovery from a single image, it is clear that the recovered surfaces are seriously affected by factors such as inaccurately calculated illumination direction and departures from Lambert's law. Both of these factors are difficult to correct when only a single image is to hand. Moreover, the restrictions imposed by most SFS schemes on the gradient field (smoothness, irradiance, integrability, unit length) are insufficient to overcome the biases introduced by the problems mentioned above.

There are a number of interesting differences between the methods reported in Section 2.5. Only Prados and Faugeras have proposed a generic approach and make SFS a well-posed problem. Zhao and Chellappa and Shimshoni et al. have constrained the problem to the domain of faces. The success of these methods will depend on the correct setting of illumination, pose and albedo, and the resulting facial reconstructions are only qualitatively accurate. Nonetheless, recognition can also be realized using alternative features such as local shape (Worthington and Hancock, 2001) and the quality of this information can be considerably increased by improved SFS methods.

Alternatively, more detailed constraints can be defined using 3D statistical models, which have proved to deliver the most effective results for recovering shape from a single image of a face. However, most of these methods rely on expensive heuristic searches for parameter fitting which sacrifice efficiency and ease of implementation.

In this thesis, we contribute to the literature in two ways. First, we combine and propose new constraints for the SFS problem. Second, we contribute in the field of statistical shape models of faces by analyzing alternative representations and exploring the

link between facial shape and intensity using coupled models.

To avoid the problem of the high dependency on the image irradiance constraints, we incorporate the integrability condition into the geometric SFS framework (Worthington and Hancock, 1999) in Chapter 3. To correct wrongly calculated directions in surface normals of the facial needle map, in Chapter 4 we propose a new convexity constraint based on local shape curvature. In Chapter 5, we analyze and compare different Cartesian representations for the construction of facial statistical models. Finally, in Chapter 6 we explore the relation between these Cartesian representations and facial intensity by using coupled statistical models. By coupling facial shape with facial intensity we can recover height data from the best fit coefficients of the intensity model.

Chapter 3

Ensuring Data-Closeness and Integrability in SFS

3.1 Introduction

As it has been explained in the literature review in Chapter 2, the image irradiance equation is under-constrained, since the family of surface normals fall on a reflectance cone whose apex angle is equal to arc cosine of the normalized image brightness, and whose axis points in the light source direction. Several constraints have been used to overcome the under-constrained nature of the Lambertian SFS problem. However, their main drawback is that they have a tendency to over-smooth the recovered surface slopes and result in poor data-closeness. The net result is a loss of fine surface detail.

These problems may be overcome by constraining the surface normals to lie on the reflectance cone and allowing them to rotate about the light source direction subject to curvature consistency constraints (Worthington and Hancock, 1999). Unfortunately, the needle maps delivered by the method are not guaranteed to satisfy the integrability constraint, which means that the recovered partial derivatives are not independent on the path

of integration (i.e. the height function may not be recoverable). Besides, these needle maps also suffer the drawback of high dependency on the image intensities, making the method prone to noisy data such as specularities, roughness and overshadowed areas.

Some SFS methods calculate directly height information, others deliver gradient data while sometimes combining the calculation of height and gradient data is considered. If the method only calculates gradient information, a necessary postprocessing integration step might be required in order to get a height map. There are a number of ways in which a surface may be recovered from a field of surface normals. One of the most elegant approaches is that described by Frankot and Chellappa (Frankot and Chellappa, 1988) which shows how the surface may be reconstructed subject to integrability constraints by performing a Fourier analysis of the field of surface normals.

The work described in this chapter aims at developing an SFS scheme that can be used to recover integrable needle maps subject to hard constraints on Lambertian reflectance as well as relaxing the image intensity dependance driven by such constraints.

In order to demonstrate how the two techniques can be combined, we will briefly explain the geometric approach for SFS (Section 3.2)as well as the algorithm proposed by Frankot and Chellappa for enforcing integrability in SFS (Section 3.3). The combination of these schemes is described in Section 3.4, while experiments and conclusions are given in Sections 3.5 and 3.6, respectively.

3.2 Geometric approach for SFS

Shape-from-shading aims to solve the image irradiance equation, $E(x, y) = R(p, q, \mathbf{s})$, where E is the brightness value of the pixel with position (x, y), R is a function referred to as *the reflectance map* (Horn, 1997) and $\mathbf{s}$ is the light source direction vector.

The reflectance map uses the surface gradients $p = \frac{\partial Z(x,y)}{\partial x}$ and $q = \frac{\partial Z(x,y)}{\partial y}$ together

with the light source direction vector $\mathbf{s}$ to compute a brightness estimate which can be compared with the observed one using a measure of error. If the surface normal at the image location (x, y) is $\mathbf{n} = (p, q, -1)^T$, then under the Lambertian reflectance model and assuming a normalized surface normal, the image irradiance equation becomes

$$E(x, y) = \mathbf{n} \cdot \mathbf{s}. \tag{3.1}$$

Worthington and Hancock (Worthington and Hancock, 1999) have developed an SFS method in which the image irradiance equation is treated as a hard constraint by demanding that the recovered surface normals lie on the reflectance cone whose axis is the light source direction and whose opening angle is the inverse cosine of the normalized image brightness. Compliance with Lambert's law is effected by rotating an estimated smoothed surface normal onto the nearest location on the local irradiance cone. The rotated on-cone surface normal is given by

$$\mathbf{n}'' = \mathbf{\Psi}\mathbf{n}' \tag{3.2}$$

where $\mathbf{\Psi}$ is a rotation matrix computed from the cone apex angle and the angle between the smoothed surface normal direction $\mathbf{n}'$ and the light source direction $\mathbf{s}$. To restore the surface normal to the irradiance cone, it must be rotated by an angle

$$\psi = \arccos(\text{irradiance}) - \cos^{-1}\left(\frac{\mathbf{n}' \cdot \mathbf{s}}{\|\mathbf{n}'\| \cdot \|\mathbf{s}\|}\right) \tag{3.3}$$

about the axis $(x, y, z)^T = \mathbf{n}' \times \mathbf{s}$. Hence, the rotation matrix is

$$\mathbf{\Psi} = \begin{pmatrix} c + x^2c' & -zs + xyc' & ys + xzc' \\ zs + xyc' & c + y^2c' & -xs + yzc' \\ -ys + xzc' & xs + yzc' & c + z^2c' \end{pmatrix} \tag{3.4}$$

where $c = \cos(\psi)$, $c^{'} = 1 - c$ and $s = \sin(\psi)$.

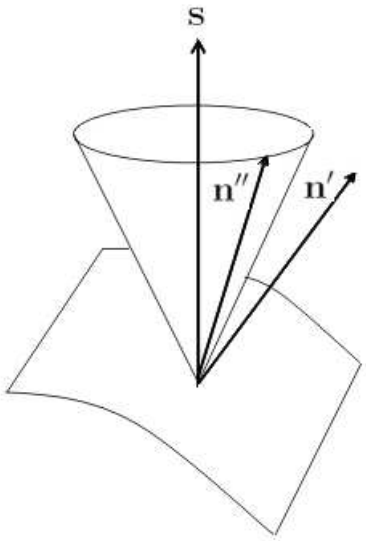

Figure 3.1: The geometric shape-from-shading approach. The surface normal $\mathbf{n}'$ is rotated by the matrix $\mathbf{\Psi}$ to obtain $\mathbf{n}''$. The resulting normal $\mathbf{n}''$ is the projection of $\mathbf{n}$ onto the reflectance cone, whose axis is the light source direction $\mathbf{s}$ and whose opening angle is the inverse cosine of the normalized image brightness.

Roughly, the geometric SFS algorithm is described as follows:

1. Calculate an initial field of surface normals (i.e using gradient initialization).

2. Smooth the obtained surface normal field. The authors of the method proposed a robust regularization approach to iteratively smooth the surface normal field (Worthington and Hancock, 1998).

3. For each pixel, calculate the rotation matrix $\mathbf{\Psi}$ and use it to rotate each of the smoothed normals (Equation 3.2). Fix the projected normal field as the current one and return to the previous step.

Note that the smoothed surface normal $\mathbf{n}'$ does not necessarily lie on the irradiance

cone, therefore, by rotating it using the matrix $\mathbf{\Psi}$, compliance with the image irradiance equation is achieved. In Figure 3.1 we illustrate the geometric SFS projection.

3.3 The Frankot and Chellappa integration method

In (Frankot and Chellappa, 1988) a method to project a gradient field to the nearest integrable solution was described. They suggested to use a set of integrable basis functions to represent the surface slopes so as to minimize the distance between an ideally integrable gradient field and a non integrable one.

Suppose that $Z(x, y)$ denotes the reconstructed height at the image location with coordinates (x, y). The integrability condition for the surface demands that the height function does not depend on the integration path. This in turn means that the surface must satisfy the condition

$$Z_{xy} = Z_{yx}, \tag{3.5}$$

where $Z_{xy} = \frac{\partial^2 Z(x,y)}{\partial x \partial y}$ and $Z_{yx} = \frac{\partial^2 Z(x,y)}{\partial y \partial x}$.

This condition can also be regarded as a smoothness constraint, since the partial derivatives of the surface need to be continuous in order that they can be integrable.

An integrable surface Z, can be represented by the basis expansion

$$\tilde{Z}(x, y) = \sum_{\omega \in \Omega} \tilde{C}(\omega)\varphi(x, y, \omega) \tag{3.6}$$

where $\varphi(x, y, \omega)$ is a set of basis functions which are not necessarily mutually orthogonal (i.e. Fourier transform), $\omega = (\omega_x, \omega_y)$ is a two dimensional index and Ω is a finite set of indexes. If each $\varphi(\omega)$ satisfies 3.5, then so does Z. The first partial derivatives of $\tilde{Z}$ can

also be expressed in terms of this set of basis functions thus

$$\frac{\partial \tilde{Z}(x,y)}{\partial x} = \sum_{\omega \in \Omega} \tilde{C}(\omega)\varphi_x(x,y,\omega) \quad \text{and} \quad \frac{\partial \tilde{Z}(x,y)}{\partial y} = \sum_{\omega \in \Omega} \tilde{C}(\omega)\varphi_y(x,y,\omega), \tag{3.7}$$

where $\varphi_x = (\partial\varphi/\partial x)$ and $\varphi_y = (\partial\varphi/\partial y)$. Since these are the first partial derivatives of an integrable surface, they share the same set of coefficients $\tilde{C}(\omega)$. In the same way, the possibly non integrable gradient field can be represented as

$$\frac{\partial \hat{Z}(x,y)}{\partial x} = \sum_{\omega \in \Omega} \hat{C}_1(\omega)\varphi_x(x,y,\omega) \quad \text{and} \quad \frac{\partial \hat{Z}(x,y)}{\partial y} = \sum_{\omega \in \Omega} \hat{C}_2(\omega)\varphi_y(x,y,\omega) \tag{3.8}$$

Since this set of first partial derivatives is not integrable, their corresponding transform coefficients will differ from each other (i.e. $\hat{C}_1(\omega) \neq \hat{C}_2(\omega)$). The distance between the non-integrable and the integrable partial derivatives can be minimized in the transform domain by making $\hat{C}_1(\omega) = \hat{C}_2(\omega) = \tilde{C}(\omega)$. The goal then is to find the set of coefficients that minimize the quantity

$$d\Big\{(\hat{Z}_x, \hat{Z}_y), (\tilde{Z}_x, \tilde{Z}_y)\Big\} = \int\int \Big\|\tilde{Z}_x - \hat{Z}_x\Big\|^2 + \Big\|\tilde{Z}_y - \hat{Z}_y\Big\|^2 dxdy \tag{3.9}$$

where the subindexes x and y denote first partial derivatives.

As Frankot and Chellapa proved, the set of coefficients $\tilde{C}(\omega)$ minimizing the error given in the above equation is

$$\tilde{C}(\omega) = \frac{P_x(\omega)\hat{C}_1(\omega) + P_y(\omega)\hat{C}_2(\omega)}{P_x(\omega) + P_y(\omega)}, \tag{3.10}$$

where $P_x(\omega) = \int\int \|\varphi_x(x,y,\omega)\|^2 dxdy$ and $P_y(\omega) = \int\int \|\varphi_y(x,y,\omega)\|^2 dxdy$. By projecting the set of coefficients $\tilde{C}(\omega)$ back from the transform domain into the spatial do-

main, a height map corresponding to the nearest integrable surface $\tilde{Z}(x, y)$ can be obtained from the input gradient field.

Equation 3.10 can be separated into two equations

$$\tilde{C}^x(\omega) = \frac{P_x(\omega)\hat{C}_1(\omega)}{P_x(\omega) + P_y(\omega)} \quad \text{and} \quad \tilde{C}^y(\omega) = \frac{P_y(\omega)\hat{C}_2(\omega)}{P_x(\omega) + P_y(\omega)} \tag{3.11}$$

where $\tilde{C}(\omega) = \tilde{C}^x(\omega) + \tilde{C}^y(\omega)$. As the surface $\tilde{Z}$ is the inverse transform of $\tilde{C}(\omega)$, then $\tilde{Z}^x$ and $\tilde{Z}^y$ are the inverse transforms of $\tilde{C}^x(\omega)$ and $\tilde{C}^y(\omega)$ respectively, and

$$\tilde{Z} = \tilde{Z}^x + \tilde{Z}^y \tag{3.12}$$

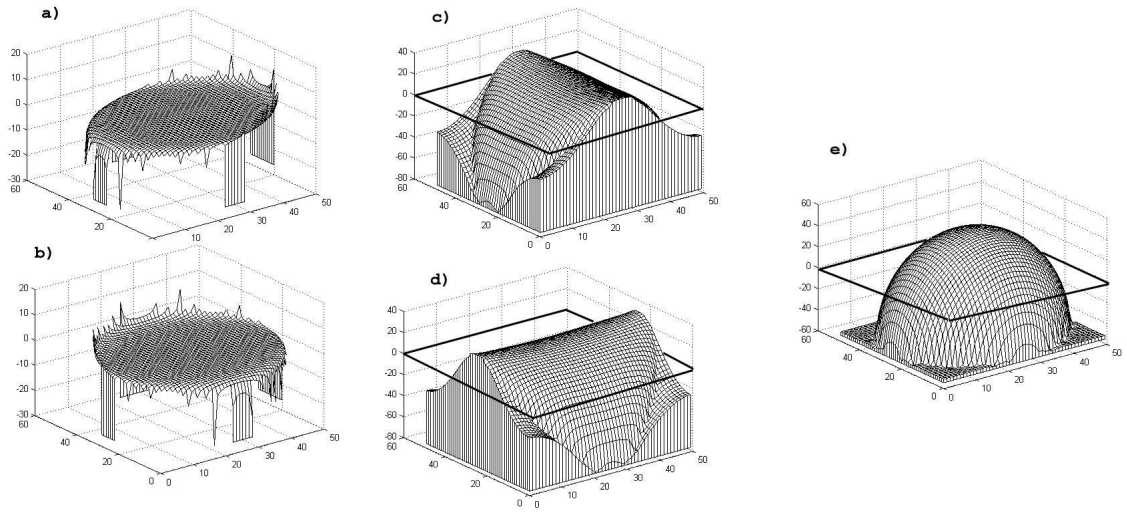

Figure 3.2: Different stages in the integration process of the partial derivatives of a sphere. In (a) and (b) we present the partial derivatives Z_x and Z_y respectively, while in (c) and (d) we present the two recovered Fourier surfaces $\tilde{Z}^x$ and $\tilde{Z}^y$. The final integrated surface is shown in (e). Note that we used a finite difference method to approximate the derivatives. The steep changes at the boundary region in (a) and (b) are a consequence of the difference with the background. Some height values are located in the bounding box and, as a consequence, the partial derivatives are approximated with high error.

In this fashion, we can recover the surface $\tilde{Z}$ up to an unknown scaling factor. In Figure 3.2 we show the different stages in the integration process of the partial derivatives of a sphere. In (a) and (b) we show the partial derivatives Z_x and Z_y respectively. In (c) and (d) we show the two recovered Fourier surfaces $\tilde{Z}^x$ and $\tilde{Z}^y$. The final integrated surface is shown in (e). We highlight the separation between positive and negative values with a bounding rectangle. Note that these values cancel out each other for the two surfaces in (b). After summing $\tilde{Z}^x$ and $\tilde{Z}^y$, the result is the sphere shown in (c).

3.4 Introducing the integrability condition in the geometric approach for SFS

Although the geometric framework for SFS ensures that the image irradiance equation is satisfied as a hard constraint, this makes the method prone to error propagation if the input image departs from Lambert's law. Real world images are contaminated by inaccurate data such as highlights and shadowed areas. Since the image irradiance dictates the apex angle of the reflectance cone, non-Lambertian images contain locations where this cone is not correctly set. A way to relax this problem can be lying the set of surface normals on less unstable and more smoothed and continuous surfaces.

The idea underpinning this section is to calculate the nearest integrable surface from a needle map using Frankot and Chellappa's method. The apex angle of the reflectance cone can be obtained on this surface with each iteration. In this way, we ensure that the surface normals will lie on reflectance cones whose apex angles correspond to integrable surfaces.

The algorithm can be summarized as follows (see figure 3.3):

1. Calculate an initial estimate of surface normals $\mathbf{n}$.

2. From $\mathbf{n}$ obtain the nearest integrable surface $\tilde{Z}$ by minimizing Equation 3.11. From the integrated surface calculate a new set of surface normals $\tilde{\mathbf{n}}$.

3. Get the apex angle α of the irradiance cone using the values of $\tilde{Z}$, that is to say, $\alpha = cos^{-1}(\tilde{Z})$.

4. Smooth $\tilde{\mathbf{n}}$ to obtain $\tilde{\mathbf{n}}'$.

5. Calculate $\tilde{\mathbf{n}}''$, by rotating $\tilde{\mathbf{n}}'$, using Equation 3.2.

6. Make $\mathbf{n} = \tilde{\mathbf{n}}''$ and return to step 2. Repeat until a desired number of iterations has been reached.

In the above algorithm, the rotation matrix does not remain static through the iterative process, since the changes in the apex angle of the reflectance cone depend on the recovered surface after each iteration. As a consequence, the new scheme satisfies a combination of integrability and data-closeness constraints.

Note how the image irradiance equation is still treated as a hard constraint, since at each iteration the surface normals are projected back to lie on the reflectance cone. However, the hardness of this constraint is relaxed when making the reflectance cone to be based on continuous surfaces calculated after each iteration instead of making them lie on the irradiance of the image along the whole iterative process.

It is also important to mention that due to the projection of the surface normals to the reflectance cone after each iteration, the z-component of the normal $\tilde{\mathbf{n}}''$ will always correspond to the calculated height surface of the final gradient field when using the Frankot and Chellappa height recovery method. By contrast, in the original method the z-component will always be the normalized input intensity image. Therefore, besides calculating surface gradients, the new algorithm also calculates height information.

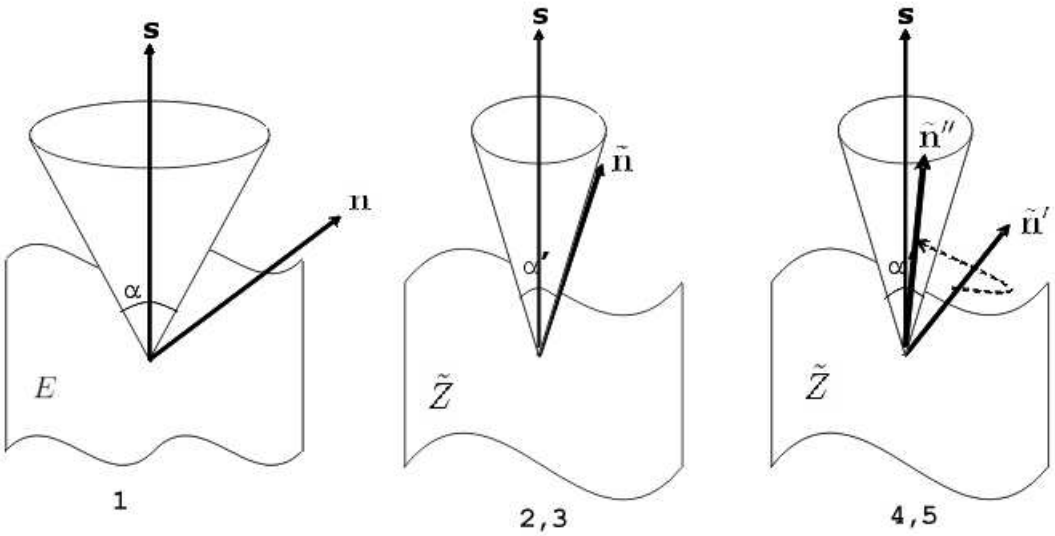

Figure 3.3: Graphic representation of the behavior of the new algorithm. The numbers under the pictures correspond to each step. In step 1, a surface normal estimation $\mathbf{n}$ is carried out. This surface normal is not suppose to lie on the reflectance cone. In steps 2 and 3, a new integrable set of surface normals $\tilde{\mathbf{n}}$ together with its corresponding surface $\tilde{Z}$ are calculated using a global minimization approach in the Fourier domain. Note how the surface changes from E (image intensity) to $\tilde{Z}$ (nearest integrable surface). Also, the apex angle of the reflectance cone changes. Step 4 is a regularizing step in order to calculate the smoothed surface normal $\tilde{\mathbf{n}}'$, which is projected back to the updated cone to get the final set of surface normals $\tilde{\mathbf{n}}''$, in steps 5. Both the width of the apex angle and the surface are just randomly represented in the graphic, also the updating behavior of the surface normals is supposed.

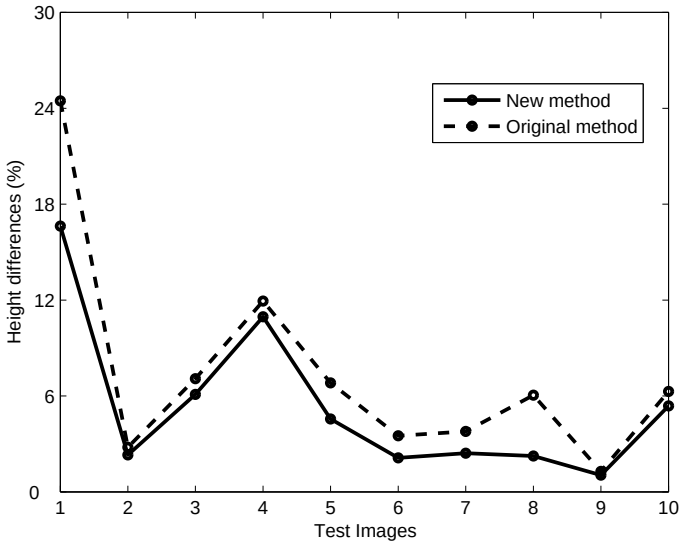

Figure 3.4: Plot of the squared height differences between the recovered surface and the ground-truth. Dotted line: original algorithm, solid line: new algorithm.

3.5 Experiments

The algorithm was tested on real world images. The evaluation criteria was based on the squared height difference and degree of gradient consistency (i.e. the percentage of pixels of every image whose differences $Z_{xy} - Z_{yx}$ are less than or equal to a certain threshold[1]). In our experiments we have compared the results obtained with the original geometric approach and the new integrable-geometric approach.

We used twenty-eight real world images. Ten of these with corresponding height data, taken from the range database of the Signal Analysis and Machine Perception Laboratory, Department of Electrical Engineering, the Ohio State University. The rest of the images were taken from the Coil database (Nene et al., 1996). For all the tests, the light source direction was assumed to be [0,0,1].

Figure 3.4 shows the results for the squared height differences. The original approach

[1]For all the experiments this threshold was set to 0.1.

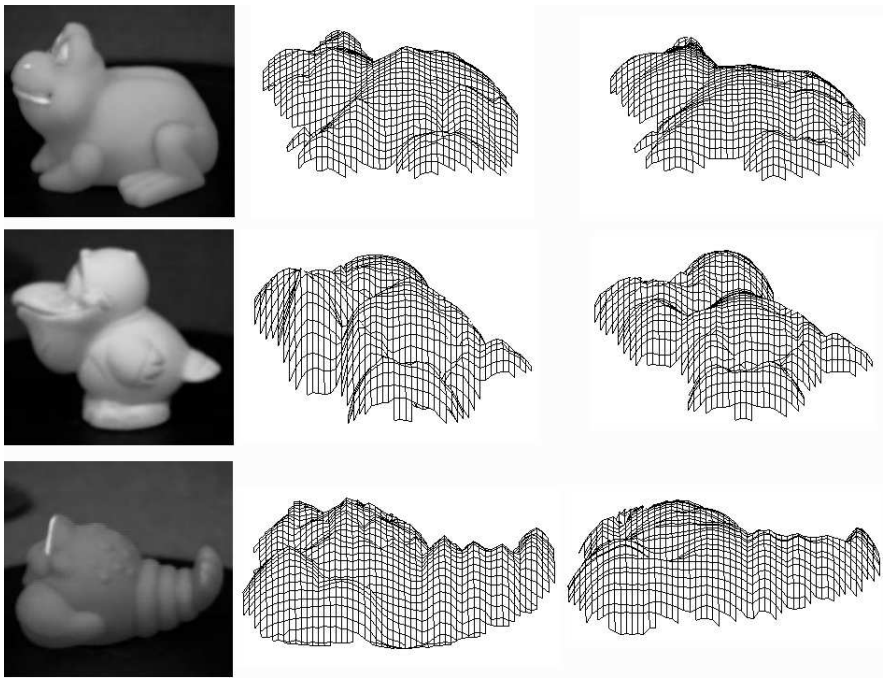

Figure 3.5: Individual height recovery analysis: (frog, bird and lobster). Left column: input image; middle column: original method; right column: new method.

is represented by the dotted line, while the new one is represented by the solid line. The plot reveals that the new method minimizes the error in a better way than the original method. This improvement is not significant though.

In a further analysis of the results, Figure 3.5 shows 3D plots of the recovered heights for each method. The left-most column of the figure shows a subset of input images, the subsequent columns represent the recovered height maps for the original method (middle column) and combined method (right-most column). We can observe that the new algorithm seems to stabilize the surface, avoiding some of the sudden changes present in the recovered surface for the original method. Specifically, in the cases of the frog and the bird, the recovered surface appears to be smoother, with none of the spurious peaks in the height map which result from the use of the original method. Although the recovered surfaces for each method are rather similar, the smoothing effect enforced by the integrability

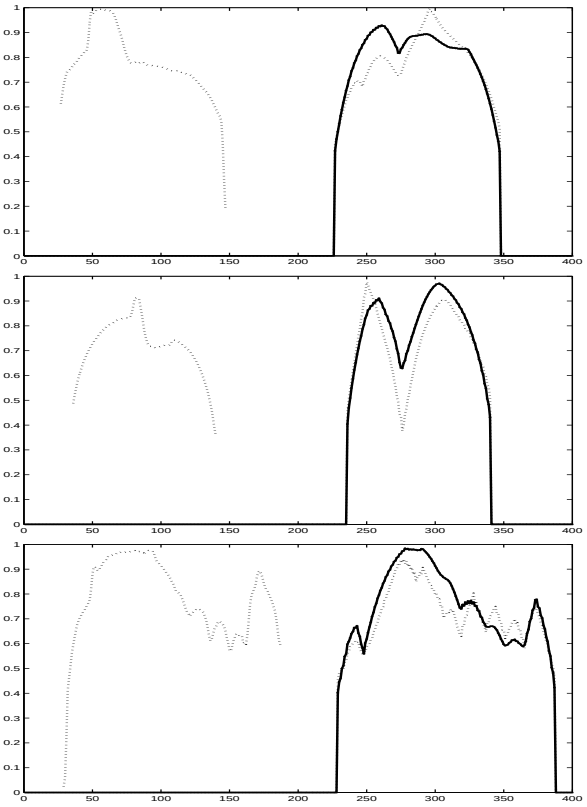

Figure 3.6: Cross section plots of the frog, bird and lobster. The single dotted plot on the left side corresponds to the ground truth surface for each case. The interpolated plots on the right side correspond to the recovered surface for the original method (dotted) and the new method (solid).

condition can explain the slightly better results for the new method, shown in Figure 3.4.

In Figure 3.6, the cross sections of the ground truth and recovered surfaces for each methods are shown (at row Y = 100, exactly the middle row of the images). From top to bottom, frog, bird and lobster. The single dotted plot on the left side corresponds to the ground truth surface for each case. The interpolated plots on the right side correspond to the recovered surface for the original method (dotted) and the new method (solid). By analyzing the interpolated plots we can notice how the new method tends to stabilize the recovered surface. In all of the cases, the high peaks on the surfaces seem to be regularized and more continuous, which can be interpreted as a consequence of the integrability

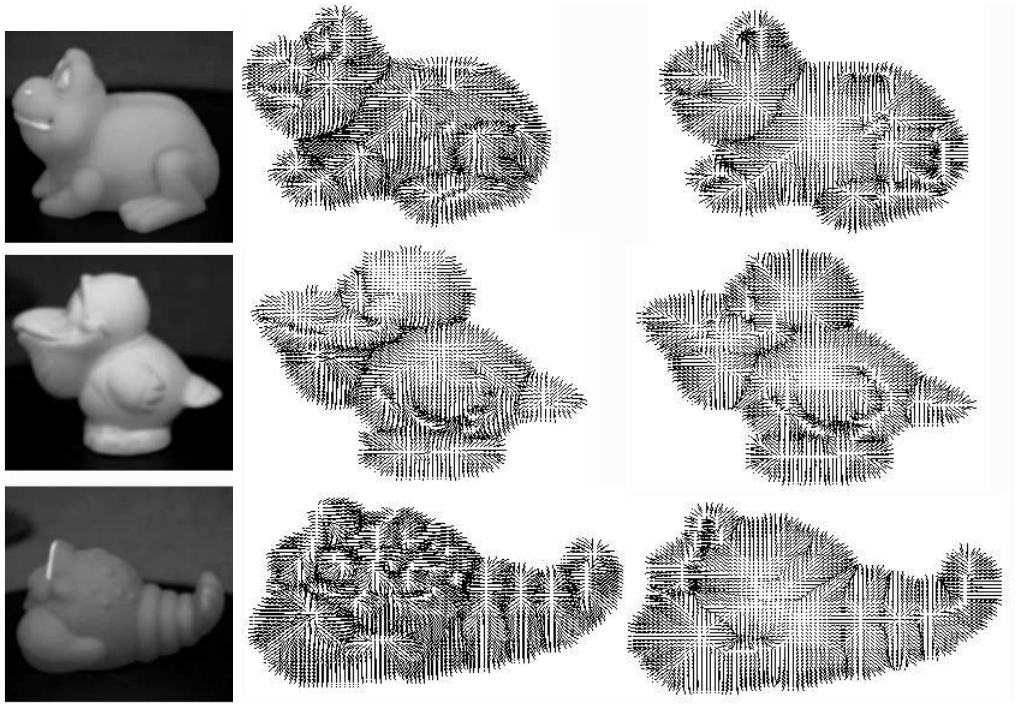

Figure 3.7: Recovered needle maps for each method. Left-most column: input images, Middle column: original algorithm. Right-most column: new algorithm.

constraint. We can also note that, despite the high difference among the recovered surfaces and the ground truth, the new algorithm seems to deliver more similar heights than the original one, this effect is more evident for the case of the lobster (third row).

Figure 3.7 shows the recovered needle maps for each method. A visual examination of the results suggests that the new method delivers needle maps that are both smoother and also contain fine topographic detail. The inclusion of integrability constraint leads to a less dependency on the image irradiance equation, therefore avoiding biasing the surface normals to highlighted points, as seen, for example, on the bird's head case.

The results of the experiments for degree of gradient consistency are summarized in Figure 3.8. The figure shows that the combined algorithm (solid line) gives better results than the original one (dotted lined), as the percentage of gradient consistency is always greater and more stable for the new approach, since at least 95% of the pixels, in all the

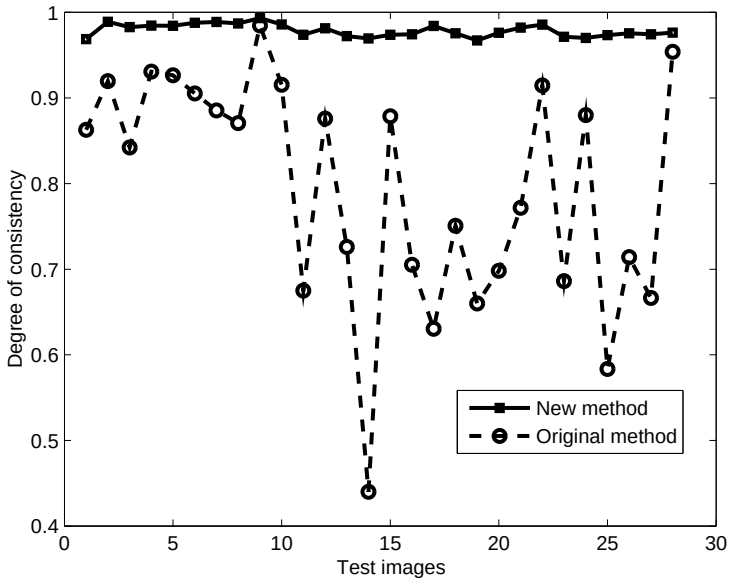

Figure 3.8: Plot of the gradient consistency degree tests. Dotted line: original algorithm, solid line: new algorithm.

cases, observe integrability. For the original algorithm, the unstable behavior shown by the dotted line is evident. This suggests that the new method, as expected, is enforcing integrability in the original method.

Figure 3.9 shows a gradient consistency comparison for the case of the lobster. The original and new methods are represented by the left and right columns respectively. The regions in black correspond to those pixels violating the integrability condition according to different thresholds (the lower the threshold, the harder the degree of integrability). From top to bottom, the applied thresholds were 0.0, 0.001, 0.01 and 0.1. The figure reveals that the new method's needle map has more stable consistency.

Finally, we experiment with input images of faces. We used a subset of frontal images of the Yale B database (Georghiades et al., 2001). In Figure 3.10 we present iso-contour plots of the recovered surfaces for two individuals, i.e. regions in the images are color coded according to height, darker colors correspond to higher values. From left to right

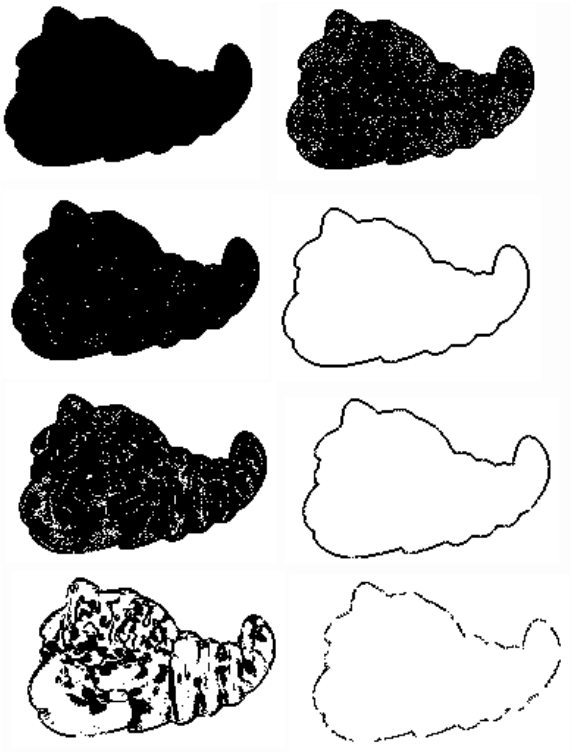

Figure 3.9: Visual comparison for the degree of consistency tests. Left column: original method, right column: combined method. From top to bottom, threshold used: 0.0,0.001,0.01 and 0.1. The regions in black correspond to those pixels not satisfying the integrability condition according to the applied thresholds.

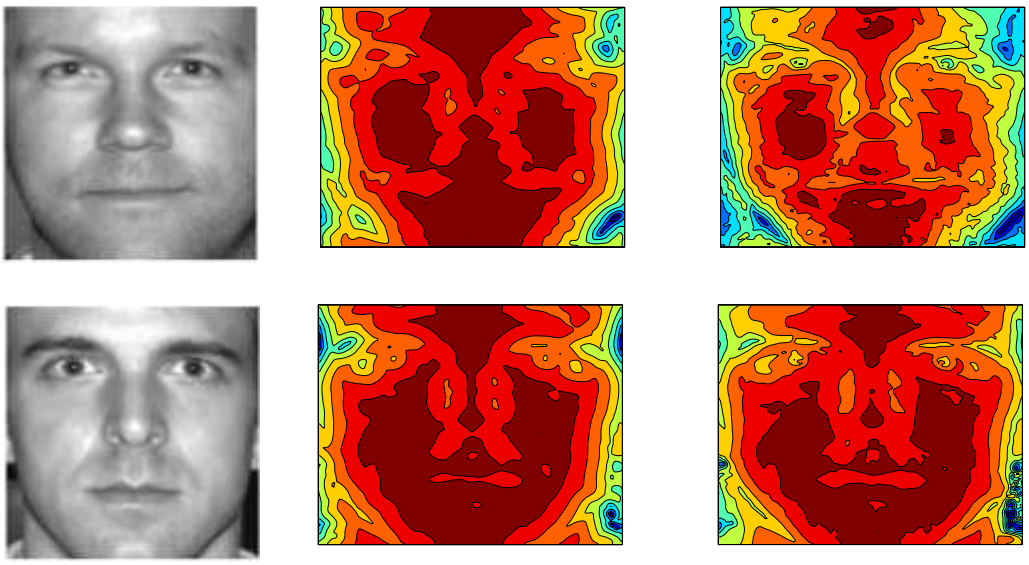

Figure 3.10: Height recovery for two individuals of the Yale B database. We present iso-contour plots of the recovered surfaces. Left column: input image; middle column: new method; right column: original method.

the columns show the input images, the results corresponding to the new method and the results corresponding to the original method. We can verify that the imposing integrability results in a benefit on the recovered surface (middle column), specially in the first row, where regions tend to be much more symmetric than in the original method (right-most column). This effect is not that noticeable in the second column.

In Figure 3.11 we present height maps of the recovered surfaces for two individuals. From left to right the columns show the input images, the results corresponding to the new method and the results corresponding to the original method. Again, some of the spurious peaks present in the original method's height maps seem to be stabilized by the integrability condition introduced by the new method. Nonetheless, it is clear that face shape recovery cannot be achieved using either approach, since regions around the nose and mouth appear imploded. In order to overcome this problem, in the next chapter of this thesis we propose a method for enforcing convexity on gradient fields of faces.

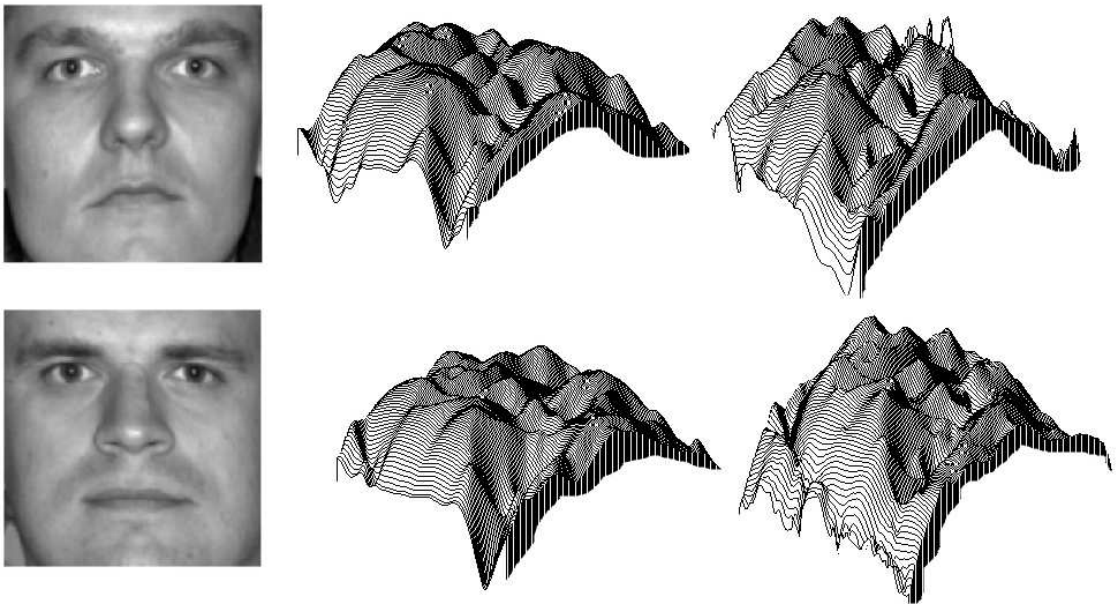

Figure 3.11: Height recovery for two individuals of the Yale B database. We present height maps of the recovered surfaces. Left column: input image; middle column: new method; right column: original method.

3.6 Conclusions

In this chapter we have demonstrated how to impose integrability constraints on a geometric approach for SFS. We follow Frankot and Chellappa and impose the constraints in the Fourier domain. Experiments reveal that the resulting method exhibits improved robustness and gradient consistency. However, although the height difference statistics do not reveal any systematic improvement in algorithm performance, both the recovered height surfaces and the needle maps delivered by the new algorithm appear to be better behaved and also preserve fine surface detail. It is important to comment that in this new method the calculation of surface orientations is less constrained by the irradiance of the image, as the rotation matrix changes through the iterative process. This is a way of relaxing the original method's problem of hard constraints on data-closeness with the image irradiance equation.

Chapter 4

Local Shape Indicators for Face Shape Recovery

4.1 Introduction

This chapter describes work aimed at developing a practical scheme for face analysis using shape-from-shading. As explained in the literature review, existing methods have a tendency to recover surfaces in which convex features such as the nose are imploded. This is a result of the fact that subtle changes in the elements of the field of surface normals can cause significant changes in the corresponding integrated surface. To overcome this problem, in this chapter we describe a local-shape based method for imposing convexity constraints. We show how to modify the orientations in the surface gradient field using critical points on the surface and local shape indicators. The critical point is located on the tip of the nose and it is assigned manually. The method is applied to both surface height recovery and face re-illumination. Experiments show that altering the field of surface normals so as to impose convexity results in greatly improved height reconstructions and more realistic re-illuminations.

The outline of this chapter is as follows. Section 4.2 introduces the local shape indicators which we use to characterize the surface topography. In Section 4.3, we describe our method for reassigning the surface gradient orientations. We provide experiments to evaluate the method on human faces in Section 6.3. Finally, in Section 4.5 we present some conclusions and identify directions for future work.

4.2 Local shape indicators

Every solid shape can be approximated locally by a collection of quadric surface patches. A quadric surface can be expressed by

$$P(u, v) = \frac{1}{2}(\kappa_1 u^2, \kappa_2 v^2),$$

where κ_1 and κ_2 are the principal curvatures with directions u and v respectively.

The principal curvatures (minimum and maximum curvatures of the surface patch) are given by the two eigenvalues of the local Hessian matrix, i.e., the matrix of the second derivatives. The principal curvatures may be estimated using the surface normal directions to compute the Hessian matrix, which can be deduced from local changes in the surface normal directions. Knowing the curvatures, the surrounding of each surface can be classified as convex, concave or saddle-type. The classification relies on the sign of the two curvatures.

Curvature-based information has been widely used in shape analysis, especially for surface segmentation and 3D object recognition. A local shape indicator is a scalar that conveys information concerning the local topography of a surface using its principal curvatures.

Local shape indicators are usually coupled. For instance, the HK classification (Besl and Jain, 1986) uses the Gaussian and mean curvatures

$$H = \frac{(\kappa_1 + \kappa_2)}{2} \text{ and } K = \kappa_1 * \kappa_2$$

respectively. By distinguishing between the cases in which H and K are individually negative, zero or positive, it is possible on the basis of the joint behavior to assign topographic labels to points on a surface. A different and slightly more convenient set of attributes is the curvedness/shape-index representation developed by Koenderink and Von Doorn (Koenderink and Van Doorn, 1992b). Here the principal curvatures are used to compute the shape index

$$S = -\frac{2}{\pi} * arctan(\frac{\kappa_1 + \kappa_2}{\kappa_1 - \kappa_2})$$

(for $\kappa_1 \geq \kappa_2$), and the curvedness,

$$C = \sqrt{\kappa_1^2 + \kappa_2^2}.$$

The shape index is an angular variable that relates to the local surface topography. It varies continuously from -1 to $+1$ as the surface changes through cup, rut, saddle-rut, saddle, saddle-ridge, ridge and dome, and cup again. The curvedness relates to the degree of curvature of the surface.

The curvedness is a convenient indicator of potential surface discontinuity. The reason for this is the higher the curvedness, the more likely the presence of a rapid variation in height. A surface with low curvedness corresponds to a highly continuous one, while a complicated surface will give rise to high curvedness at many locations. For instance, in the case of faces, the curvedness is large for features such as the boundaries of the nose, mouth and eyes.

For our experiments, we utilize the local descriptors of shape-index and curvedness to characterize the regions on a gradient field where a change of orientation should be performed to enforce convexity.

4.3 Using local shape indicators to redirect SFS gradient fields

Inevitably, any surface gradient field delivered by SFS will be inaccurate due to noise or albedo changes, and these in turn cause variations in the intensities of the input image. SFS works well for objects that are uniformly concave or convex. However, if the object under study is more complex, with both concave and convex regions, then SFS can fail. In these situations although the recovered surface normal direction is consistent with the measured image brightness, the recovered surface does not reflect the structure of the object under study. In particular, there may be inversions of the sign of the surface curvature with convex regions appearing concave and vice-versa. However, in the case of faces (and many other objects) the surface under study is largely convex.

Based on this above observation, in this chapter we present a method for enforcing the convexity of the integrated surface. We use the location of the global height maximum on the surface to enforce the condition.

Formally stated, suppose that Z is a smooth surface immersed in $\mathbb{R}^3$. Let p be a critical point of Z and U_p a neighborhood of p. Suppose that Z is locally concave over U_p. Then, the new surface $\widetilde{Z}$ constructed from Z by reversing the sign of all its partial derivatives, Z_x and Z_y, is locally convex within U_p. Moreover, a local maximum on $\widetilde{Z}$ will be located at that point where the function ceases increasing and starts decreasing[1]. Suppose that all the partial derivatives of $\widetilde{Z}$ with respect to x, $\widetilde{Z}_x$, have a negative sign before reaching the position of the critical point p along the x axis and have a positive sign after reaching it. Suppose also that the same occurs for $\widetilde{Z}_y$. Under these conditions, then the critical point p on U_p will be the position of the global maximum[2] of $\widetilde{Z}$.

The basic idea underlying this chapter is to enforce the condition that the integrated

[1]Of course, $\widetilde{Z}$ will present many local maxima for a face-like surface.
[2]It might be a maximum or a minimum depending on the integration method.

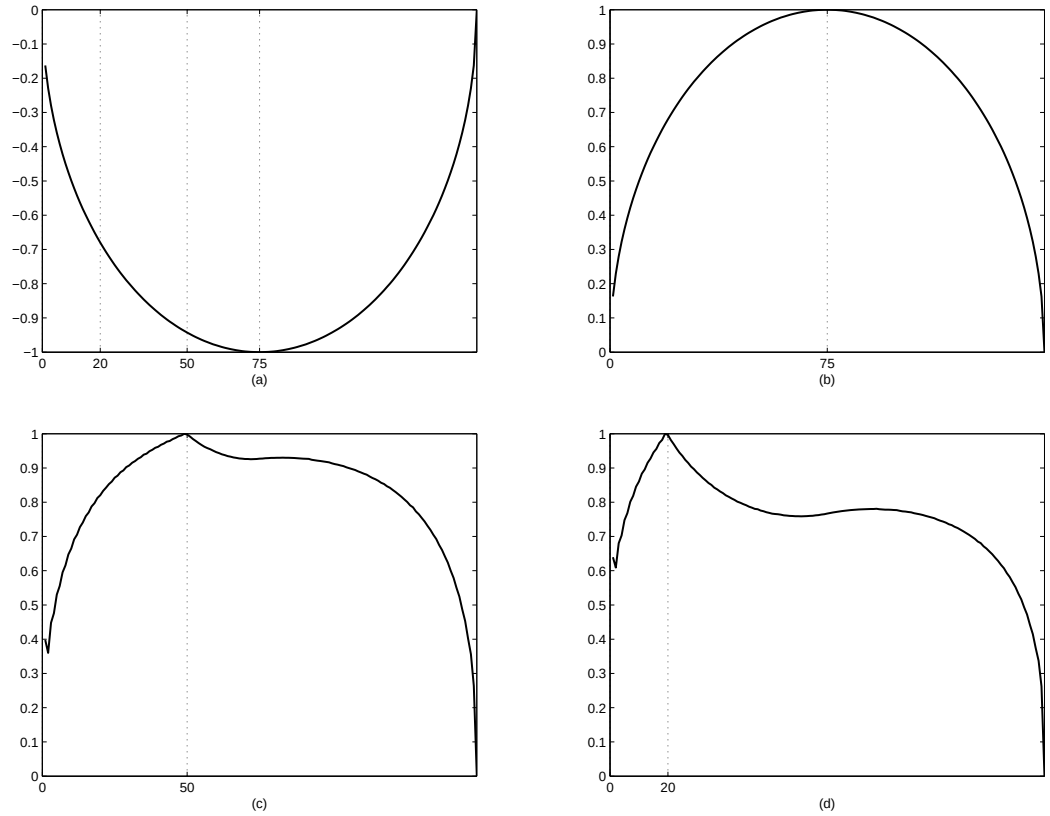

Figure 4.1: Applying the method to the derivatives of a concave sphere. The radius of the sphere is 75 units. We show transverse plots of the original concave surface (top left corner) and the recovered surface height after changing the sign of the derivatives. The global height maximum coordinates (a, b) are set to (75,75), (50,75) and (20,75) respectively for the panels labeled (b), (c) and (d). The thresholds τ_x and τ_y are set to zero.

surface has a global height maximum. For the face analysis problem, we select this point to be at the tip of the nose. By choosing such a point we can divide the surface into positively and negatively signed areas of the needle map. To enforce this condition we apply the simple rule:

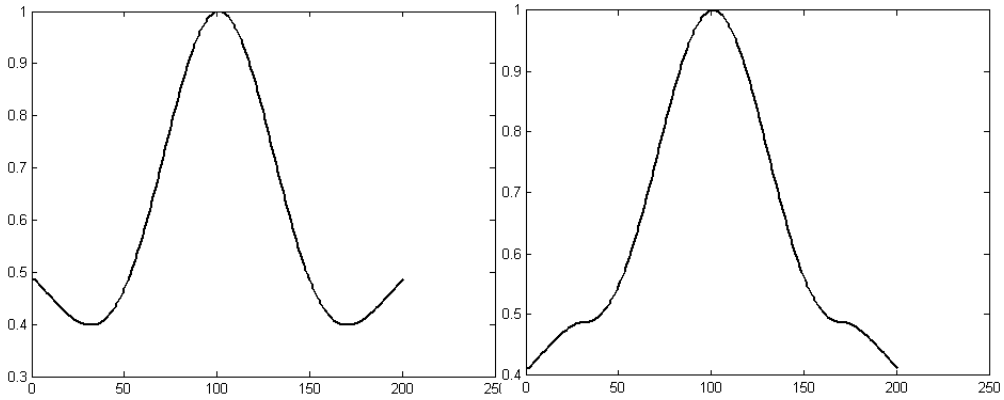

Figure 4.2: Applying the method to the Mexican hat function. From left to right we show the original surface and recovered surface after applying the method taking as the global height maximum the center of the surface with $\tau_x = \tau_y = 0$. Note how the concave parts of the Mexican hat become convex.

$$\breve{Z}_x(x,y) = \begin{cases} abs(\hat{Z}_x(x,y)) & \text{if } x \leq a \text{ and } C(x,y) \geq \tau_x \\ -abs(\hat{Z}_x(x,y)) & \text{if } x > a \text{ and } C(x,y) \geq \tau_x \\ \hat{Z}_x(x,y) & \text{otherwise} \end{cases}$$

$$\breve{Z}_y(x,y) = \begin{cases} abs(\hat{Z}_y(x,y)) & \text{if } y \leq b \text{ and } C(x,y) \geq \tau_y \\ abs(\hat{Z}_y(x,y)) & \text{if } y > b \text{ and } C(x,y) \geq \tau_y \\ \hat{Z}_y(x,y) & \text{otherwise} \end{cases}$$

where $\breve{Z}_x$ and $\breve{Z}_y$ are the updated gradients, and $\hat{Z}_x$ and $\hat{Z}_y$ are the original gradients. The desired global height maximum is located at the point with co-ordinates (a, b) (the tip of the nose) and this is assigned manually. To decide whether the element of the gradient field at the location (x, y) will be altered or not, we compare the curvedness indicator,

$C(x, y)$ to the thresholds τ_x and τ_y. Similarly, the shape-index (SI) can also be utilized to redirect the surface normals, with the rule:

$$\breve{Z}_x(x,y) = \begin{cases} abs(\hat{Z}_x(x,y)) & \text{if } x \leq a \text{ and } SI(x,y) \leq \tau_x \\ -abs(\hat{Z}_x(x,y)) & \text{if } x > a \text{ and } SI(x,y) \leq \tau_x \\ \hat{Z}_x(x,y) & \text{otherwise} \end{cases}$$

$$\breve{Z}_y(x,y) = \begin{cases} abs(\hat{Z}_y(x,y)) & \text{if } y \leq b \text{ and } SI(x,y) \leq \tau_y \\ abs(\hat{Z}_y(x,y)) & \text{if } y > b \text{ and } SI(x,y) \leq \tau_y \\ \hat{Z}_y(x,y) & \text{otherwise} \end{cases}$$

In this case, $SI(x, y)$ is the shape-index which is compared to the thresholds τ_x and τ_y for deciding whether the element of the gradient field at the location (x, y) will be altered or not.

The following diagrams show the behavior of the method when applied to the derivatives of a sphere and a Mexican hat. Both experiments were realized using the curvedness indicator.

To illustrate the global height maximum enforcement procedure, Figure 4.1 shows the results of applying the method to the derivatives of a concave sphere with radius 75 units, which is shown at the top left corner of the figure. We show transverse plots of the recovered surface height in the direction of the x axis. The global height maximum coordinates (a, b) are set to (75,75), (50,75) and (20,75) respectively for (b), (c) and (d). The thresholds τ_x and τ_y are both set to zero.

The effect on the convexity of the surface is clearer in Figure 4.2, where the method

is applied to the Mexican hat function. Transverse sections of the recovered surface are shown, and from left to right they show the original surface and recovered surface after applying the method taking as the global height maximum the center of the surface with $\tau_x = \tau_y = 0$. Note how the concave parts of the Mexican hat become convex.

It is evident that the peak-enforcement procedure will segment the recovered surface into four quadrants. As a result the curvedness of the recovered surface will be reduced. This is not desirable for surface height recovery from a face. The net effect will be to make the surface structure pyramidal. This problem is overcome by using the thresholding procedure to either force the normals to change direction, or to allow them to remain unchanged. This procedure is rather heuristic, and different thresholds apply to different images of faces. Although we have discovered that a fluctuation between 0.2 and 0.3 is generally successful for most of the cases, it is recommended to commence with the hardest case (all the derivatives change) and gradually modify the thresholds until the best shape is generated, as suggested in Figure 4.5.

It is worth commenting that the signs of a needle map are not the only attributes which can alter the resulting integrated surface. Modifying the influence of the z component of the surface normal leads to an alteration which affects the gradient and therefore forces some regions of the height map to be either flatter or more curved after performing the global integration.

Note that we propose a directional correction which can be used in conjunction with already established SFS methods. We focus on face shape recovery for the frontal pose (since the critical point to consider is the tip of the nose). However, other assumptions such as light source direction, albedo, boundary conditions and inter-reflections are incorporated by the SFS method used to compute the derivatives. In this chapter we use the geometric framework (Worthington and Hancock, 1999) outlined in Section 3.2. Although this framework assumes constant albedo, we test our experiments with variable

albedo images of faces.

In the following section some experiments will be presented in order to illustrate these points on an application involving face reconstruction using SFS.

4.4 Experiments

This section is organized into two parts. We commence by describing experiments focussed on height recovery, and then proceed to describe results obtained by re-illuminating the recovered surfaces.

In first part of the study, which focusses on height recovery, we show examples of the effect of the gradient re-direction process on the recovered surfaces. We also present an analysis of the errors produced by the method on 50 synthetic test images from the Max-Plank database. Results of using the method over real world images are also presented.

Figure 4.3: Face for analysis.

In the second part of the study, which focusses on surface re-illumination, we show the results of using the recovered surface gradients to synthesize new face images. Here, we investigate the effect of moving the light source direction. We compare the results with ground-truth.

The face database used for our experiments was provided by the Max-Planck Institute for Biological Cybernetics in Tuebingen, Germany. As described in (Blanz and Vetter,

Figure 4.4: Illumination and height analysis. In the left-hand columns of the figure we show the result obtained using the un-modified needle map, while the right-hand columns show the result obtained with the modified field of surface normals with $\tau_x = 0.3$ and $\tau_y = 0.4$. Each panel shows re-illuminations obtained when the light source direction is $(0, 0, 1)^T$ followed by recovered height-maps as intensity plots.

1999), this database was constructed using a Cyberware laser scanner. The range images are of the heads of young adults represented in a cylindrical coordinate system. We have converted the cylindrical coordinates into Cartesian coordinates and recovered the associated height values. We were also provided with synthetic textures corresponding to each face and these were used to render the range images. We used frontal pose of the recovered surface illuminated by a light source at infinity and parallel to the viewer direction. We also utilized real world face images to complement our experiments. We refer to images from the Max-Planck database unless otherwise stated.

Note that we used the geometric SFS framework described in Section 3.2 to get the initial needle maps to be later modified using local shape. We used the framework for face-analysis since it has been demonstrated to recover a field of surface normals that preserves fine topographic detail. The output of the algorithm is used as an initial estimate of the field of surface normals. We show how to use shape-indicators computed from this field of surface normals to correct for convex-concave surface inversions, and hence to improve the quality of the recovered height-map. For gradient field integration we utilized the Frankot and Chellappa global integration algorithm explained in Section 3.3

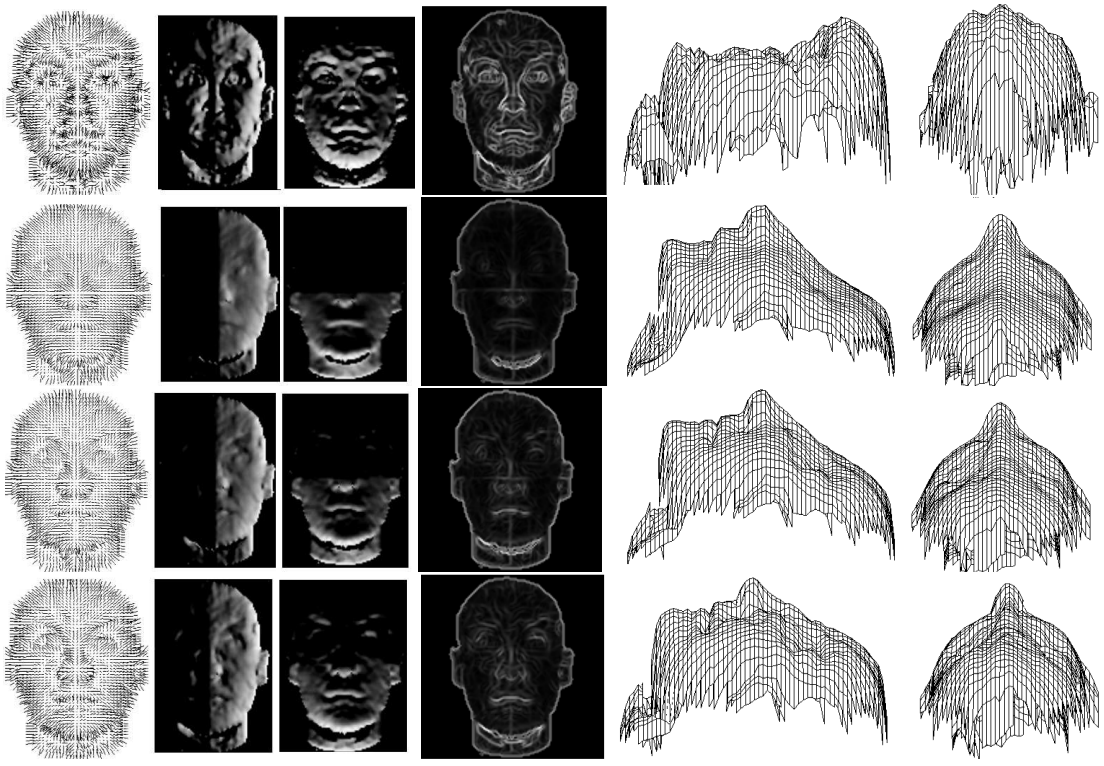

Figure 4.5: Curvedness, surface gradient and surface recovery analysis. The first column shows the field of surface normals. The second and third columns show the result of re-illuminating the surface normals with light-source directions $(1, 0, 0)^T$ and $(0, -1, 0)^T$. The last three columns show the curvedness map, together with profile (side) and top-down views of the surface wire-frames. In the top row we show the original case. Here the implosion of the nose is clear. The subsequent rows are for $\tau_x = \tau_y = 0$, $\tau_x = \tau_y = 0.2$ and finally $\tau_x = 0.3$ and $\tau_y = 0.4$, respectively.

4.4.1 Height recovery

The first series of tests were carried out on the image shown in Figure 4.3, and serves as an illustration of the method described in this chapter. To compute the surface gradients from the raw image brightness we followed the procedure described in Section 3.2. As noted previously, this construction ensures that the image irradiance equation is satisfied as a hard constraint. For the surface integration step we used the global method proposed by Frankot and Chellappa (Frankot and Chellappa, 1988) and discussed in Section 4. This method recovers surface height using the inverse Fourier transform of the field of surface

normal directions.

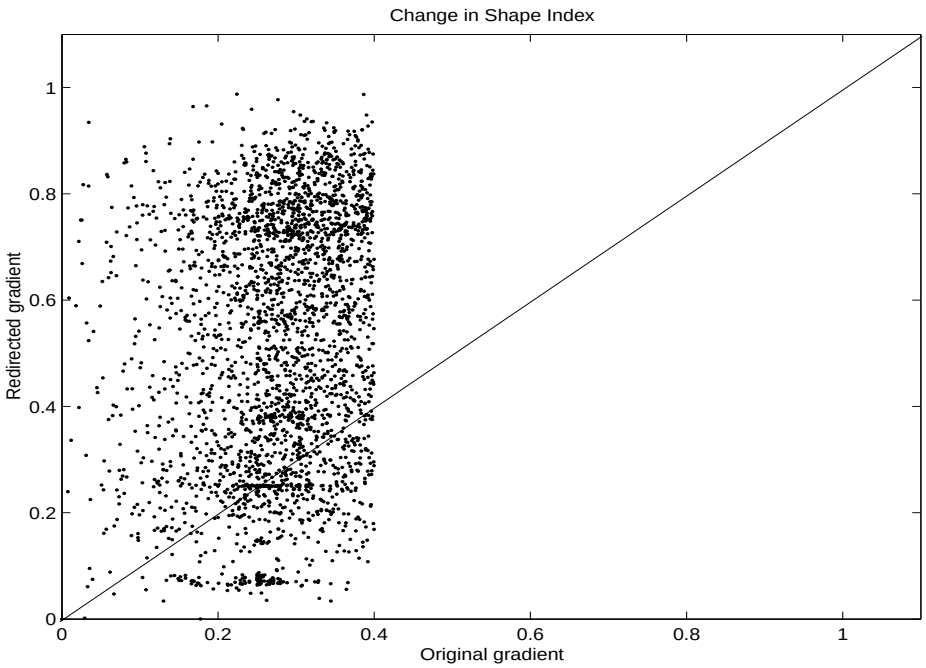

Figure 4.6: Convexity enforcement using shape index. Scatter plot comparison between the original and modified shape-indexes (the shape-index scale was normalized from 0 to 1). The x-axis corresponds to the shape index of the input field of normals. The y-axis corresponds to the shape index of that input needle map after redirecting it using $\tau_x = \tau_y = 0.4$ case

In Figure 4.5, we illustrate the effects of using our method for redirecting the field of surface normals. The first column shows the needle map. The second and third columns show the result of re-illuminating the surface normals with light-source directions $(1, 0, 0)^T$ and $(0, -1, 0)^T$. The top row shows the result obtained with the initial field of surface normals. Notice how the re-illuminations suggest that the nose is imploded, since the region surrounding it is shadowed. The subsequent rows show the recovered surface illuminations after applying the method with $\tau_x = \tau_y = 0$, $\tau_x = \tau_y = 0.2$ and finally $\tau_x = 0.3$ and $\tau_y = 0.4$, respectively. For the hardest case (second row), where no data is filtered by the thresholds, the four-quadrant effect is very marked. However, the effect diminishes when the thresholds are increased. The harder the threshold, the

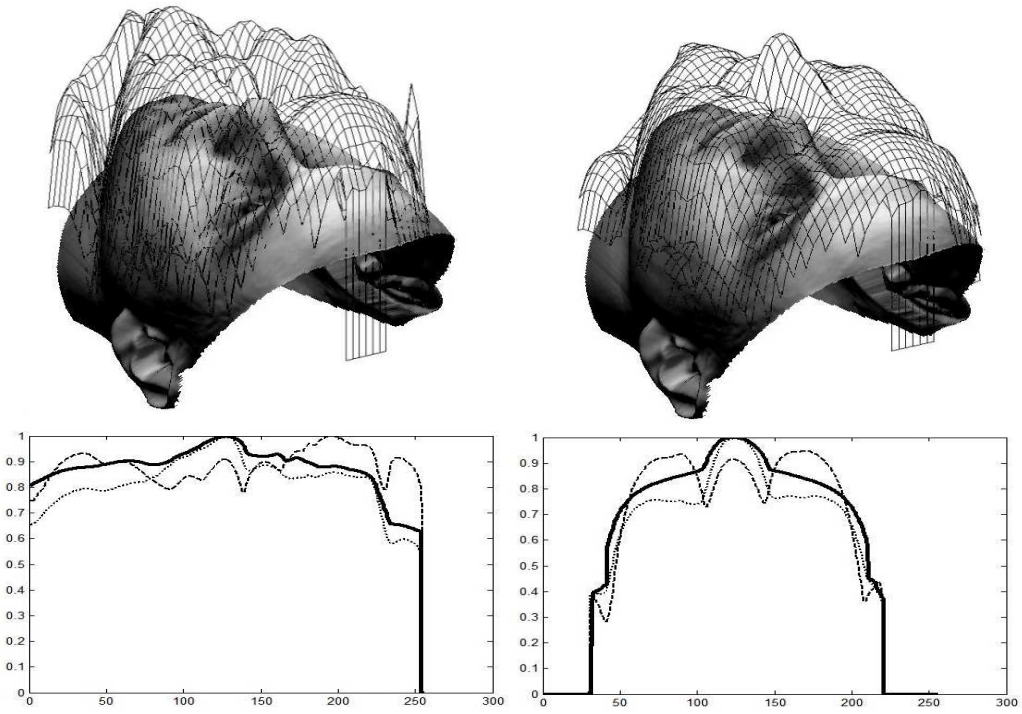

Figure 4.7: Height maps and cross plots comparison. We compare with ground truth the recovered height maps obtained with (right column) and without (left column) redirection of the normal field. The top row figure shows the recovered height maps superimposed on the image from which they were generated. In the bottom row we show longitudinal and transverse sections of the recovered height maps. The left-hand diagram corresponds to the longitudinal section along the x-axis and the right-hand diagram shows the transverse section along the y-axis. In both plots, the thick solid line represents the ground-truth surface, the dotted line illustrates the recovered surface using our method, and the dashed line shows the height data integrated using the original field of surface normals.

stronger the influence of the face features. It is important to note the differences between the needle maps appearing in the top and bottom rows of the figure. The change of signs in the surface normals suggests that the new surface normals relate more strongly to the underlying shape of the face. This is a consequence of inverting the originally concave regions (the ones around the nose, eyes and mouth) to become convex ones, causing the imploded facial areas to "pop back" so that a better face shape can be recovered.

We take this analysis one step further in the last three columns of Figure 4.5 where

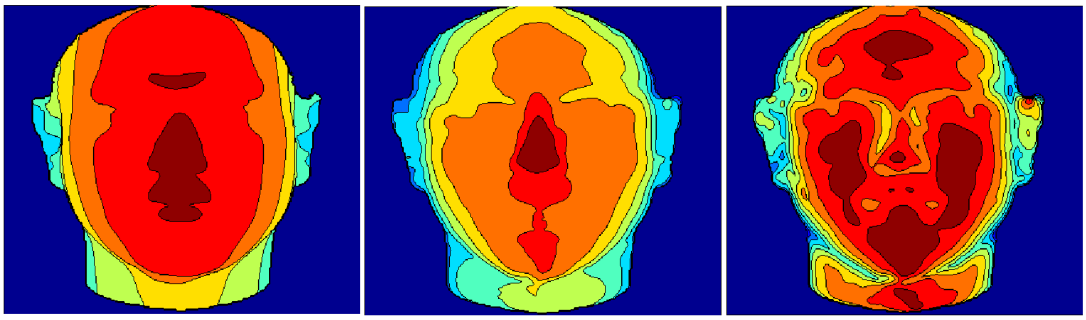

Figure 4.8: Iso-contour comparison. Regions in the images are color coded according to height. Discarding the background, the darkest regions correspond to highest values. The leftmost panel corresponds to the ground-truth surface, the middle panel presents the case when our method was applied, while the rightmost figure represents the recovered surface without gradient redirection.

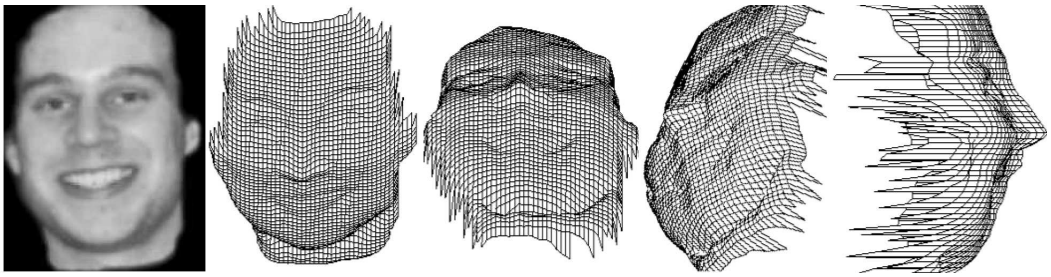

Figure 4.9: Recovered surface for a smiling face.

from left to right we show the curvedness map, together with profile (side) and top-down views of the surface wire-frames. Note in the top row how the curvedness is increased, as well as the poor quality of the recovered surface in terms resemblance to a face. The curvedness is of course minimized when no threshold is taken into account (second row), but the four-quadrant effect again becomes obvious. By incrementing τ_x and τ_y the fine features of the face seem to be preserved and the overall structure of the face is still sound. By choosing an appropriate threshold we are able to enhance the salient features of a particular face while maintaining the overall face composition. It is worth commenting on the manual assignment procedure for selecting the location of the critical point. The

procedure consists of clicking on the tip of the nose using a standard computer mouse. The resulting height maps do not greatly vary from one another if the critical point is moved to a location inside the area surrounding the tip of the nose. However, the method is subjective and further improvement is needed to automatically select the location of the critical point.

The correction of the directions of the normals is clearer in Figure 4.4. In the first two columns we show the result obtained using the un-modified needle map, while the last two columns show the result obtained with the modified field of surface normals with $\tau_x = 0.3$ and $\tau_y = 0.4$. Each panel shows re-illuminations obtained when the light source direction is $(0, 0, 1)^T$ followed by recovered height-maps as intensity plots. There are several features that deserve comment. For instance, the intensities around the nose produce the appearance of a convex surface, making it appear more natural.

As far as the reconstruction using shape-index is concerned, the height data recovered is very similar to that obtained using curvedness. Figure 4.6 shows a scatter plot comparison between the original and modified shape-indexes[3]: the x-axis corresponds to the original case while the y-axis corresponds to the $\tau_x = \tau_y = 0.4$ case. This diagram presents only those pixels in the original gradient field with a shape index lower or equal to 0.4 (x-axis), therefore we can analyze the new value of such pixels in the redirected field of surface normals (y-axis). Note how the majority of the points are distributed above the line $x = y$, which shows how the original shape index turned into a grater one, suggesting that the concave regions changed to convex. Such pixels belong to the regions surrounding the nose, mouth and eyes. The small cloud of points below the line, representing the pixels where the shape index remained lower than the threshold is mainly related to the pixels surrounding the face boundary.

We found that for most experiments good qualitative results were obtained with τ_x

[3]The shape-index scale was normalized from 0 to 1.

and τ_y varying between 0.3 and 0.4. A further analysis is shown by Figure 4.7. Here we compare with ground truth the recovered height maps obtained with (right column) and without (left column) redirection of the normal field. The top row figure shows the recovered height maps superimposed on the image from which they were generated. Note how the imploded features (left image) of the face, such as nose and mouth, become visible after the method is applied (right image). In the bottom row we show longitudinal and transverse sections of the recovered height maps. The left-hand diagram corresponds to the longitudinal section along the x-axis and the right-hand diagram shows the transverse section along the y-axis. In both plots, the thick solid line represents the ground-truth surface, the dotted line illustrates the recovered surface using our method, and the dashed line shows the height data integrated using the original field of surface normals. There is a clear improvement of the recovered height map once the re-direction operation is applied to the surface normals. Although some errors still remain, the rectification of the implosion problem is sufficient to reveal plausible facial shape.

An iso-contour comparison is shown in Figure 4.8. Here, regions in the images are color coded according to height. Discarding the background, the darkest regions correspond to highest values. The leftmost panel corresponds to the ground-truth surface, the middle panel presents the case when our method was applied, while the rightmost figure represents the recovered surface without gradient redirection. Note the similarities between the ground-truth and the recovered surface after changing the gradient orientations. It is evident that, in the height maps generated by both the ground-truth and re-directed gradients, the peak region is located in the area of the nose, with some discrepancies in the mouth and chin areas though. The un-modified gradient generated height contains equal peak regions over the entire image, i.e., cheeks, chin, front and a tiny one on the tip of the nose.

The images shown in Figure 4.10 provide an absolute height difference analysis. Here

we realize comparisons using profile images. The left-most column shows the raw input image. The middle and right-most columns show superimposed plots of the recovered height maps after applying our method on the profile views (i.e. side) of the images from which the original gradient fields were calculated. From the superimposed surfaces, it is clear that the major differences in the recovered height maps and the ground-truth surfaces are near the nose, the cheeks and the mouth area.

To provide a detailed analysis of our method, a more exhaustive set of tests was carried out on fifty images of faces from the database. The average percentage of height difference comparison plot[4] is shown in Figure 4.11. We calculated the percentage of height difference error $\|Ground\_truth - Recovered\_surface\|/Ground\_truth$ as an average over all points of the 50 surfaces. The diagram shows a scatter plot comparing the average percentage of height difference from the original gradient integrated surface (x-axis) against the redirected gradient integrated surface (y-axis). Observe how when the original gradient is used, the error is concentrated between 8% and 12%. The error, however, when the redirected gradient is used, is concentrated between 2% and 4%.

Figure 4.12 illustrates how pre-processing operations might improve the recovery of height maps from images. The left-most panel shows the calculated height map of the image of the face in Figure 4.16 after removing the eyes and eyebrows by setting their pixel brightness values to be that of the average over the skin. The right-most image shows the calculated height map on the un-edited face image. Note how this pre-processing step generates more accurate height maps in the eye-region when our re-direction method is used.

However, some regions present a higher degree of error for the new surfaces. This is illustrated by the analysis given by Figure 4.13. The first column shows the longitudi-

[4]For these experiments, the ground-truth surface was generated by integrating the known ground-truth gradient from each image, using the Frankot and Chellappa method. This was done so that all the surfaces were generated on the same basis for comparison purposes. This reduces the biases involved in the integration method.

nal sections along the face for the recovered surfaces both before (dotted line) and after (dashed line) redirection of surface normals. The recovered height map after integration of the ground-truth normals is shown as a solid line. The second and third columns respectively show the iso-contour representations of the absolute height differences before and after re-directing the surface normals. From these pictures it is clear that the absolute height difference is reduced after changing the direction of the surface normals. However, there are still errors and these occur mainly in the proximity of the chin and eyes areas. These problems can be attributed to changes of albedo (eye area) and instabilities produced by the boundaries of the chin and neck . It is also interesting to note how the separation between the lips tends to disappear after redirection. This can be explained as a consequence of erroneously enforcing convexity.

Finally, Figure 4.9 presents some wire-frame views of the recovered surface obtained from an image of a smiling face (leftmost). Here we have used a real world (non-synthetic) image which was taken with a digital camera. This is a challenging example since the face is in a non-frontal pose. The overall structure of the face was well recovered, however there is again some error in the area of the chin and eyes.

4.4.2 Re-illumination

Dealing with variation in illumination direction is a topic of central concern in face recognition. The reason for this is that light-source effects are responsible for more variability in the appearance of face images than changes in identity (Moses et al., 1994). In this section we investigate how the surfaces recovered using our method can be used for synthesizing new facial images under different lighting conditions. Here we use a simple Lambertian re-illumination model using albedo maps derived from the input images.

Figure 4.14 presents re-illumination experiments for two example faces. The first column corresponds to the input image. The remaining columns show the generated re-

illuminations after applying our method. For the second and third columns, the light source vector is nearly parallel to the x axis in both negative and positive directions, while nearly parallel to the y axis for the fourth and fifth columns.

A more exhaustive analysis is shown in Figure 4.16. For both sets of images, the top row represents the re-illumination results obtained using the ground-truth normals, the middle row shows those obtained using the unaltered gradient field and the bottom row those obtained using the re-directed surface normals. From left to right, the light source direction makes an angle of -45, -25, 25 and $+45$ degrees to the image normal in the horizontal (x) direction for the upper set of images. In the lower set of images, the light source is moved in the same manner in the vertical (y) direction. It is interesting to note the similarities between the ground-truth and re-directed gradient re-illuminations. Although the recovered surface does not accurately represent the shape of the image from which it was acquired, the overall shape is sufficiently accurate to create realistic re-illuminations provided that the light source is not moved by more than 45^o. The results are best when the light source is moved in the horizontal direction. This is a consequence of the vertical symmetry of human faces. On the other hand, the re-illumination results for the un-modified gradient fields show artifacts of implosion in the area around the nose and mouth. This becomes more severe when the light source moves further away from the viewer direction.

A second analysis is shown in figure 4.15, where an unprocessed image of a face (single image in between two rows) was used for the experiments. This image is a real world one, taken by a digital camera. The first row presents the results for the modified gradient field while second shows those obtained with the original gradient field. Note how in the top row the quality of the re-illuminations is improved. This contrasts with the imploded features presented in the second row. It is important to note, however, that errors appear in some areas of the face, i.e. those surrounding the mouth. This can be explained

as the consequence of the change in reflectance properties on the lips. Of course, the eyes and the mustache area also present different reflectance properties. This suggests that more attention should be paid to the reflectance model used for re-illumination.

4.5 Conclusions

We have presented a method for correcting a gradient field of a face. The aims in doing this are twofold. First, we wish to generate a height map with a global maximum located at a critical point located at the tip of the nose. Second, we aim to force the recovered surface to be convex in accordance with evidence provided by local shape indicators. We have proved that the simple idea of modifying the surface normal directions so as to restore the convexity of imploded features using the constraints derived from the location of a point of global maximum height seems to work well with the recovery of face surfaces. After integration, the recovered shape preserves most of the salient facial features, including the nose lips and eye-sockets. As the accuracy of the reconstruction will depend on the gradient data, pre-processing steps for correcting intensities (i.e. removing specularities and areas of albedo variation such as the eye areas) would probably improve the quality of the results, if a more accurate height map is required.

Although the method reported here is effective in correcting feature implosion when the surface gradient is computed using a geometric SFS approach, there clearly remains scope for further improvement in computing accurate surface gradients. The improved height recovery from single images of faces also suggests the possibility of generating coarse extrapolation for rendering novel views with few degrees of rotation. Considering that our only input information is a single frontal image of a face, the raw height maps resulting from our algorithm can help as a starting point for other refinements aimed at generating more precise information. Of course, the gradient maximum constraint is

natural for faces, since the tip of the nose is a global height maximum. However, the constraint could be used for more general surfaces in a local manner for surface height recovery and where there are local regions of implosion.

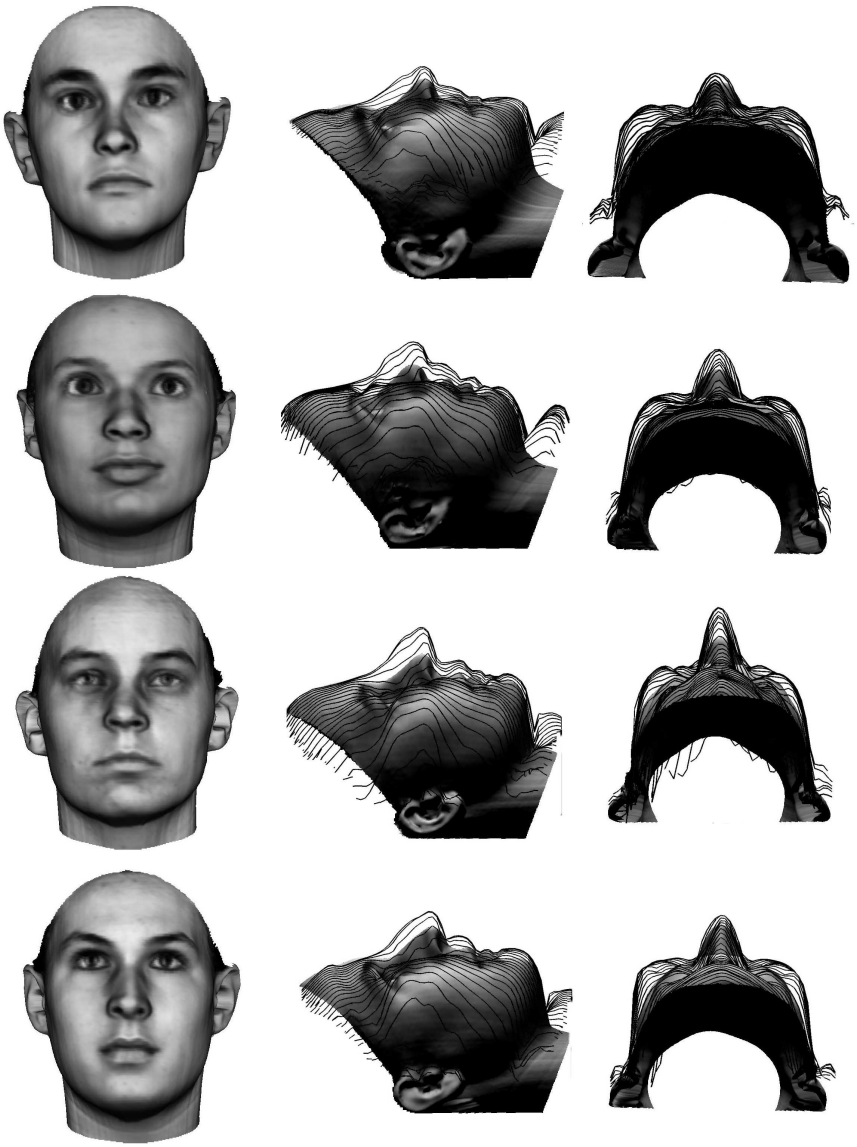

Figure 4.10: Height map analysis. The left-most column shows the raw image. The middle and right-most columns show superimposed plots of the recovered height maps after applying our method on the profile views (i.e. side) of the images from which the original gradient fields were calculated.

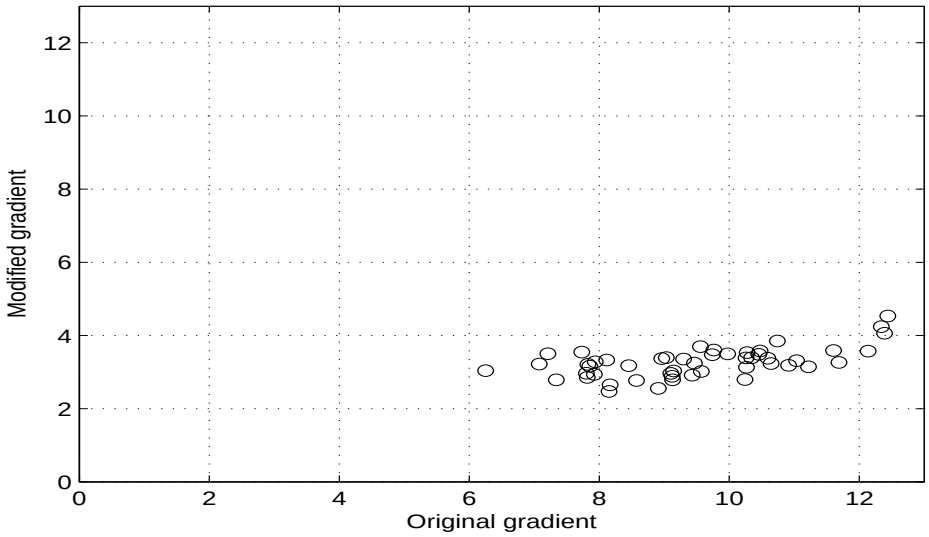

Figure 4.11: Average percentage of height difference. The diagram shows a scatter plot comparing the average percentage of height difference from the original gradient integrated surface (x-axis) against the redirected gradient integrated surface (y-axis).

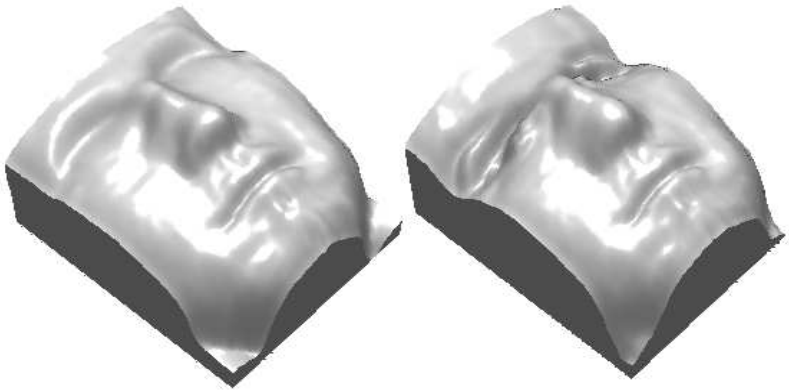

Figure 4.12: Results on an edited image. The left-most panel shows the calculated height map of the image of the face in Figure 4.16 after removing the eyes and eyebrows by setting their pixel brightness values to be that of the average over the skin. The right-most image shows the calculated height map on the un-edited face image.

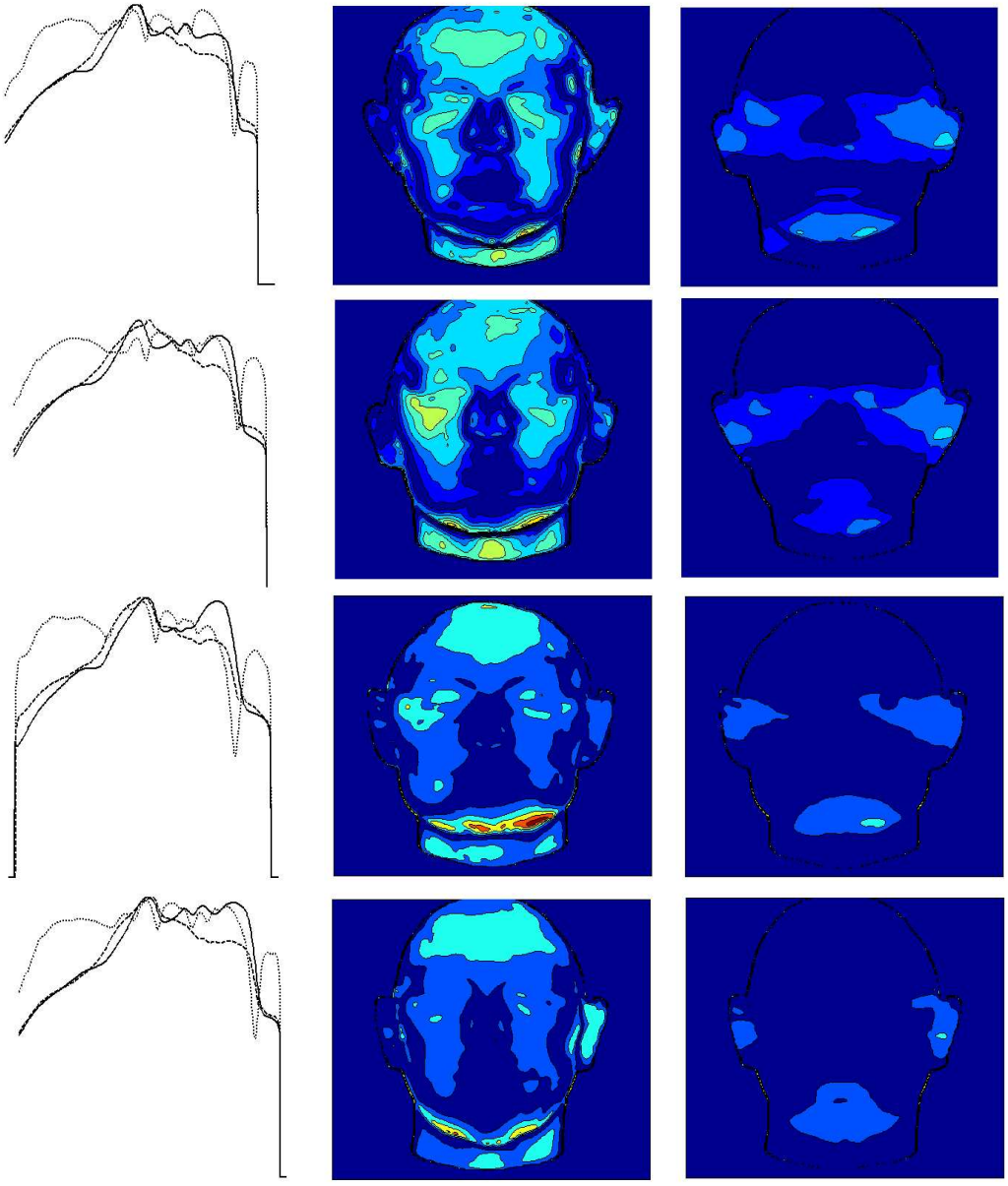

Figure 4.13: Height difference cross sections and iso-contour plots. The first column shows the longitudinal sections along the face for the recovered surfaces both before (dotted line) and after (dashed line) redirection of surface normals. The recovered height map after integration of the ground-truth normals is shown as a solid line. The second and third columns respectively show the iso-contour representations of the absolute height differences before and after re-directing the surface normals.

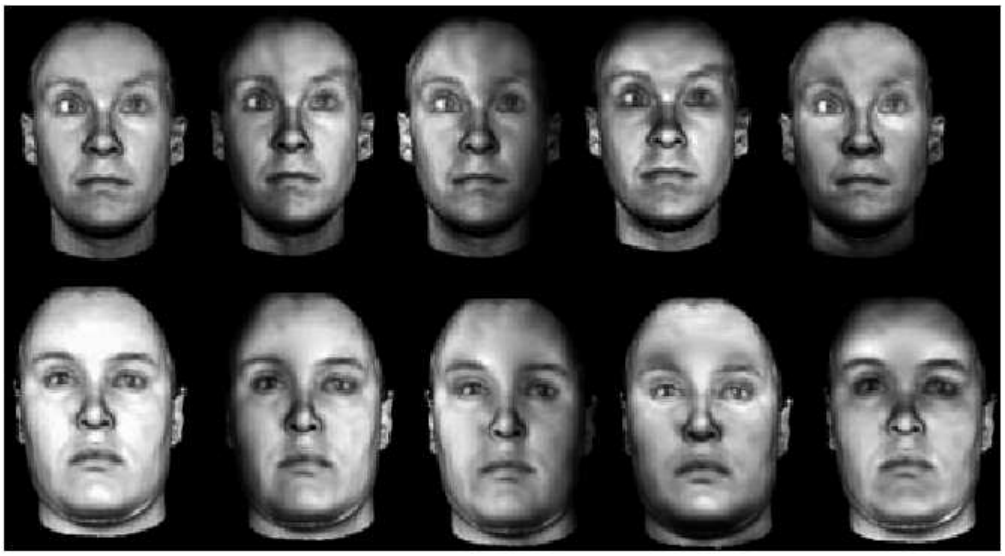

Figure 4.14: Re-illumination tests for two different images. The first column corresponds to the input image. The remaining columns show the generated re-illuminations after applying our method. For the second and third columns, the light source vector is nearly parallel to the x axis in both negative and positive directions, while nearly parallel to the y axis for the fourth and fifth columns.

Figure 4.15: Comparison of re-illuminations using an unprocessed image of a face (single image in between two rows). The first row presents the results for the modified gradient field while second shows those obtained with the original gradient field.

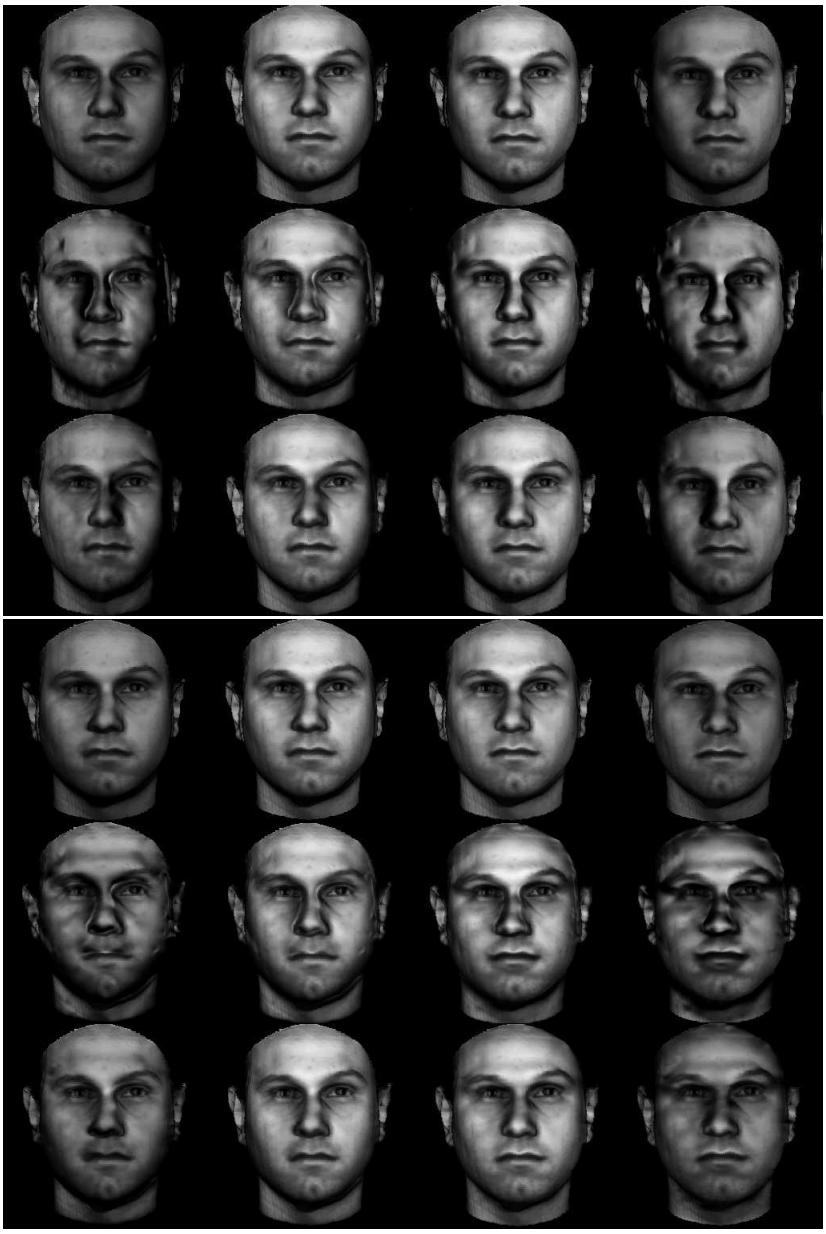

Figure 4.16: Comparison for re-illumination tests. For both sets of images, the top row represents the re-illumination results from the ground-truth normals, the middle row shows those obtained from the unaltered gradient field and the bottom row those obtained from the re-directed surface normals.

Chapter 5

Building Cartesian Coordinate-based Models of Faces

5.1 Introduction

In the two-dimensional domain, variations in facial appearance can be captured using the eigenfaces technique (Turk and Pentland, 1991). Here a set of aligned facial intensity images are used to construct the eigenmodes. The image data is usually encoded as a Cartesian long-vector by concatenating the rows or columns of the image. However, if a 3D model is to be constructed in an analogous manner from range data, then there exist alternative ways for representing the training data. One of the simplest and most commonly used approaches is to adopt a cylindrical coordinate representation. Using cylindrical coordinates, the surface of a human face (or head) can be parameterized by the function $rad(\mu, \ell)$, where rad is the radius and μ and ℓ are respectively the height and angular coordinates. This representation is used since it captures the linear relations between basis heads. Unfortunately, it can lead to ambiguity since different data can be fitted to the same head-model.

An alternative is to use a Cartesian representation, in which each surface point is specified by its (x, y, z) coordinates, where the z-axis is in the direction of the viewer. The Cartesian coordinates are related to the cylindrical coordinates through

$$(x, y, z) = (x_0 + r(\mu, \ell) \sin \mu, y_0 + \ell, z_0 + r(\mu, \ell) \cos \mu), \tag{5.1}$$

for some reference shift (x_0, y_0, z_0).

A face depth map in cylindrical coordinates can be thought of as an unwrapped version of a depth map expressed in Cartesian coordinates. In Figure 5.1, a cylindrical coordinate depth map (left) is shown together with its corresponding Cartesian coordinate depth map (right).

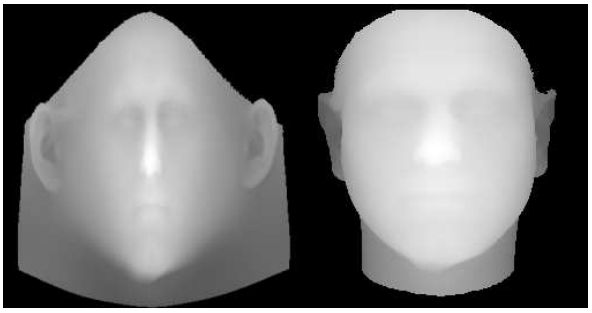

Figure 5.1: Cylindrical and Cartesian coordinate depth maps.

A general drawback of Cartesian coordinates is the high spatial variance caused by face features such as nose, mouth and eye. Moreover, this variance can be exaggerated if the training data is misaligned. If this is the case, additional z-variance (height variability) is introduced in areas with high spatial variance due to alignment error. As a result, Cartesian coordinates have not been used for generating 3D statistical models of faces. However, Cartesian coordinates were recently used by Dogvard and Basri (Dovgard and Basri, 2004) to construct a statistical model of faces using symmetry constraints. To overcome the problem of alignment errors they expressed the surface gradient in terms

of a set of deformation coefficients. This allows shape-from-shading to be transformed into a linear system of equations that can be simply solved for the shape coefficients, and then used to reconstruct the surface height function for the face. Although accuracy is sacrificed, the method is computationally efficient.

Height maps, however, are not the only way to representing 3D information in Cartesian coordinates. Alternative encodings can be drawn from 2.5D information such as the partial derivatives of a surface. Although the 2.5D representation is less appealing since it must be integrated to recover a surface, because of the image irradiance equation the 2.5D representation is closer to the raw image brightness data than a height surface.

In this chapter we explore and experiment with alternative Cartesian representations for constructing 3D statistical models of faces. We explore two different routes. The first of these is based on height, while the second is based on directional information. In the case of the directional models, we investigate how surface integrability can be enforced. Finally, we show how the models can be fitted to image brightness data using geometric constraints on surface normal direction provided by Lambert's law (Worthington and Hancock, 1999) subject to integrability (Frankot and Chellappa, 1988).

The chapter is organized as follows. In Section 5.2 we provide a brief explanation of the different Cartesian representations explored. The construction of the statistical models using each representation is explained in Section 5.3. The fitting procedure used to test the performance of the models is described in Section 5.4. Experimental results and analytic comparison of the models is given in Section 6.3. Finally, we present conclusions and suggest some possible lines for future work in Section 5.6.

5.2 The Cartesian representations

We have explored the use of four Cartesian coordinate representations to construct statistical models of facial shape. Two of these are based on directional information, while the remaining two are based on height information. We work under orthographic projections, i.e, the viewed surface is assumed to have been projected into 2D space of the image plane such that the direction of the projection axis is opposite to that of the viewer. Every visible point on the surface is then projected to the image plane.

5.2.1 Cartesian representations based on directional information

Information about a surface that is intermediate between a full 3D representation and a 2D projection onto a plane is often referred to as a 2.5D surface representation (Marr, 1982). Surface orientation is one of the most important 2.5D representations. For every visible point on a surface, there exists a corresponding orientation which is usually represented by either surface normal, surface gradient or the azimuth and zenith angles of the surface normal.

In contrast to height data, directional information cannot be used to generate novel views in a straightforward way. However, given the illumination direction and the surface albedo properties, then surface normal directional plays the central role the surface radiance generation process. This is of particular interest in face analysis since light-source effects are responsible for more variability in the appearance of facial images than changes in shape or identity (Moses et al., 1994).

In this chapter we explore two Cartesian representations based on directional information, namely the surface gradient and the surface normal azimuthal angle.

Surface gradient

The *Surface Gradient* representation is based on the directional partial derivatives of the height function $p = \frac{\partial Z(x,y)}{\partial x}$ and $q = \frac{\partial Z(x,y)}{\partial y}$. The set of first partial derivatives of a surface is also known as the gradient space. This is a 2D representation of the orientation of visible points on the surface.

Figure 5.2: Intensity plots of the surface gradients w.r.t x (left) and w.r.t. y (right) of a face.

In Figure 5.2 we show the gradient space of a face. The left and right hand panels respectively show the slope parameters p and q represented as intensity images. Here we have calculated the surface normals by fitting bicubic patches to the surface height function $Z(x, y)$.

Azimuth angle

Directional information can also be expressed using the zenith (slant) and azimuthal (tilt) angles of the surface normals. In terms of the slope parameters, the zenith angle is $\theta = \arctan\sqrt{p^2 + q^2}$ and the azimuth angle is $\phi = \arctan\frac{q}{p}$ (see Figure 5.3). Here we use the four quadrant arc-tangent function and therefore $-\pi \leq \phi \leq \pi$. In contrast to the human visual system (Erens et al., 1993b), it seems that computer vision systems encounter more difficulty in estimating the tilt of a surface from a single image than its slant (see Figure 5.4).

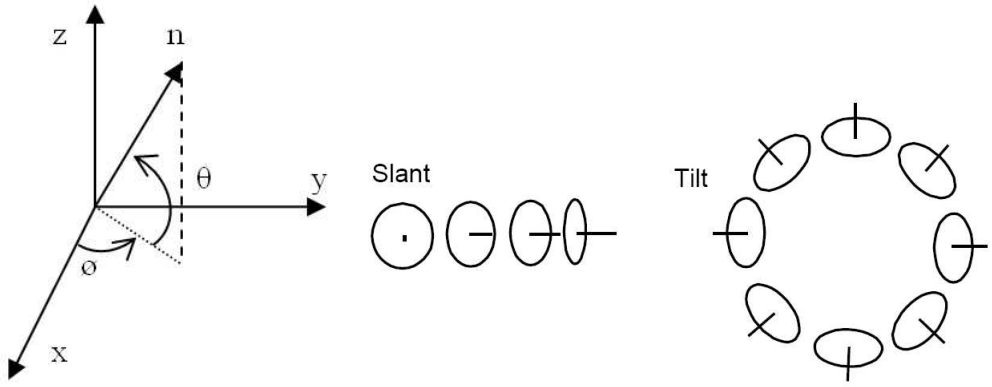

Figure 5.3: The azimuth (ϕ) and zenith (θ) angles of a surface normal (left) and the visual interpretation of the slant and the tilt (right).

Angular data is more difficult to model than Cartesian data. The reason for this is angle wrap around. Hence, small differences in distance on a sphere can correspond to large differences in the angles latitude or longitude. The classical example here is a short walk across one of the poles of a sphere, when large differences in longitude correspond to small differences traveled across the pole. In shape-from-shading, since the surface normal is determined by the azimuth and zenith angles, when the surface is illuminated in the direction of the viewer and when the surface reflectance is Lambertian, then the arc-cosine of the zenith angle is determined by the image brightness. The azimuth angle, on the other hand, must be determined using additional constraints provided by smoothness or the occluding boundary.

5.2.2 Cartesian representations based on height information

Surface height recovery through the integration of surface normal data can be prone to error (Wu and Li, 1988; Klette and Schluens, 1996). Hence the main advantage in the direct use of height over direction is that height does not need to be integrated and is not prone to these errors. Here we study one representation based on raw height data and one

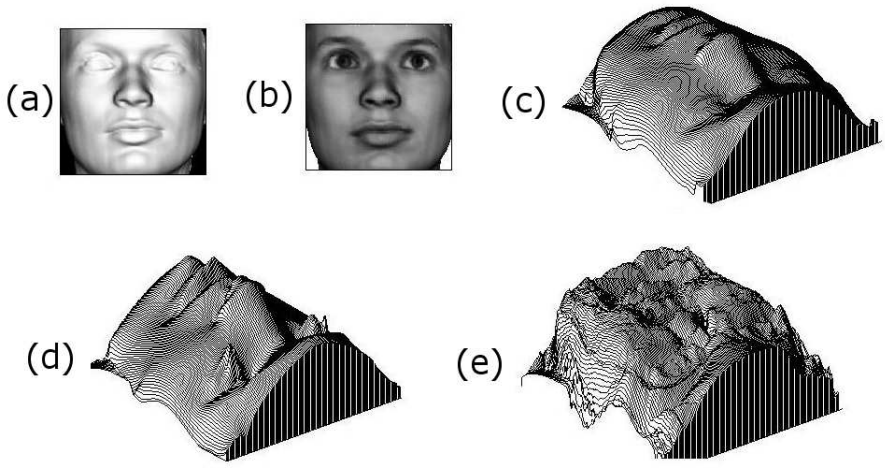

Figure 5.4: Effect of incorrectly calculated azimuth and zenith angle in face shape recovery: (a) orthogonal Lambertian (constant albedo) image, (b) true irradiance (non-constant albedo) image, (c) ground-truth surface, (d) surface preserving true azimuth angle but with its zenith angle estimated through SFS and (e) surface preserving true zenith angle, but with its azimuth angle estimated through SFS. Note how the effect of wrongly estimated tilt angle cause a severe deterioration on the recovered surface.

based on the representation of height using Fourier basis functions.

Height

The *Height* representation is based on the surface height function $Z(x, y)$, i.e. the relation between every visible point of a surface with a unique height value. In Figure 5.5 we illustrate the use of the height function. In panel (a) we show a close-to-profile view of the reconstruction of a face using cylindrical coordinates. Panel (b) shows the projection of the surface height function onto the image plane. In panel (c) we show the close-to-profile view from the Cartesian height map. Panel (d) shows the Cartesian height map with the boundaries of the face removed. Unlike the Cartesian representation, in the cylindrical reconstruction it is possible to define regions near the occluding boundary of the frontal view (ears and sides of the neck). Nonetheless, the salient facial features such as the eyes, nose and mouth can be easily represented in Cartesian coordinates, as shown in panel (d).

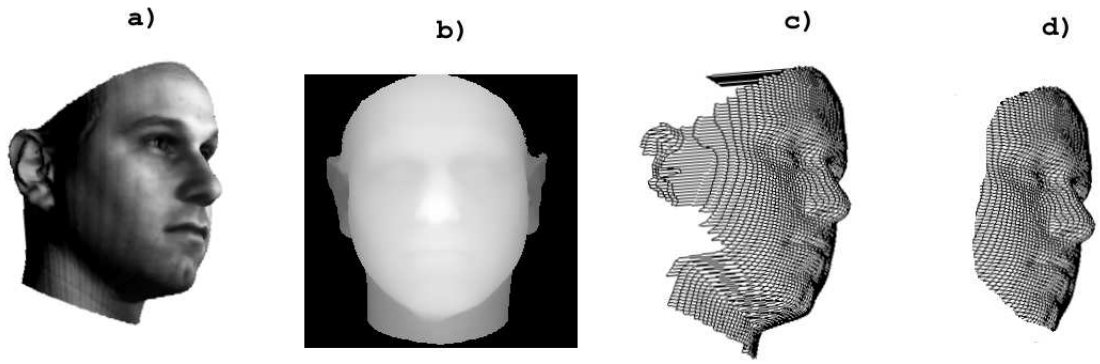

Figure 5.5: Cylindrical and Cartesian reconstructions. We show a close-to-profile view of the reconstruction of a face from cylindrical coordinates (a), its projection to the image plane as a height function (b), the same close-to-profile view from the Cartesian height map (c) and without the surrounding region close to the boundaries of the face.

Fourier basis

The *Fourier Basis* representation draws on the concept of Fourier domain integrability for surface height recovery from surface gradient. The representation builds on the Fourier basis introduced by Frankot and Chellappa (Frankot and Chellappa, 1988). This method has been previously explained in Section 3.3.

In Figure 5.6 we show the Fourier surfaces of a face. The first two figures correspond to the height map and surface plot of the x-component ($\tilde{Z}^x$). Similarly, the right-most figures represent the y-component ($\tilde{Z}^y$).

5.3 Principal component analysis

In this section we describe how eigenspace models are constructed for Cartesian data. Here we follow the approach adopted by Turk and Pentland who were among the first to explore the use of principal components analysis for face recognition (Turk and Pentland, 1991). We make use of the technique described by Kirby and Sirovich. (Kirby and

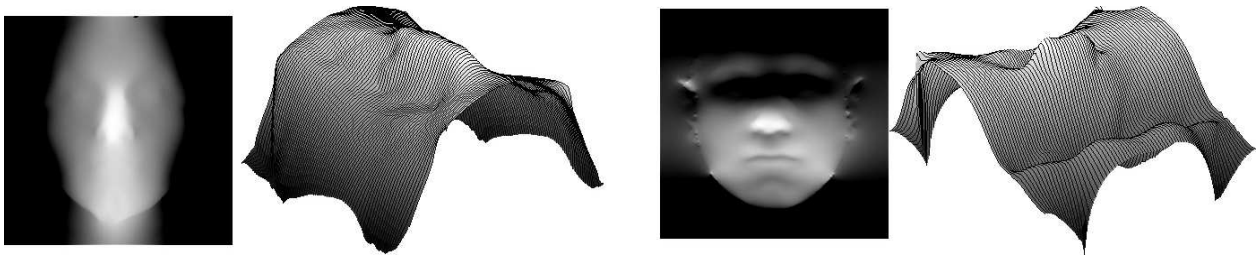

Figure 5.6: Fourier basis representation for faces. The first two figures correspond to the height map and surface plot of the x-component of the Fourier domain representation. Similarly, the right-most figures represent the y-component.

Sirovich, 1990) to render the method efficient.

5.3.1 The intensity model

The image data is vectorized by stacking the image columns to form long column vectors $\mathbf{i}$. If the K training images contain M columns and N rows, then the pixel with column index j_c and row index j_r corresponds to the element indexed $j = (N-1)j_c + j_r$ of the long column vector. The training set data-matrix, $\mathbf{I} = [\mathbf{i}_1|\mathbf{i}_2|\cdots|\mathbf{i}_K]$ is then formed by using the long vectors $\mathbf{i}_k$ as columns. The differences from the average face image, $\bar{\mathbf{i}}$ (the sample mean) are used to construct the centered training data matrix

$$\mathbf{I}' = [(\mathbf{i}_1 - \bar{\mathbf{i}})|(\mathbf{i}_2 - \bar{\mathbf{i}})|\cdots|(\mathbf{i}_K - \bar{\mathbf{i}})] = [\mathbf{i}'_1|\mathbf{i}'_2|\cdots|\mathbf{i}'_K]. \tag{5.2}$$

Principal Component Analysis (PCA) seeks a set of $K-1$ orthogonal vectors which, in a least squares sense, best describe the distribution of the columns of $\mathbf{I}'$. The solution to the least squares problem is found by calculating the eigenvectors of the *explicit* covariance matrix

$$\mathbf{\Sigma}^i = \sum_{k=1}^{K} \mathbf{i}'_k {\mathbf{i}'_k}^T = \mathbf{I}'{\mathbf{I}'}^T. \tag{5.3}$$

Unfortunately, due to its size ($MN \times MN$), computing the eigenvalues and eigenvec-

tors of $\mathbf{\Sigma}^i$ becomes intractable for large sets of data. However, the numerically efficient method proposed in (Kirby and Sirovich, 1990) can be used to overcome these difficulties. According to this method there are only $K-1$ non zero eigenvalues from $\mathbf{\Sigma}^i$ and these can be computed from the $K \times K$ *sampled* covariance matrix $\widehat{\mathbf{\Sigma}}^i = \mathbf{I'}^T\mathbf{I'}$. The eigen-vector equations for the explicit and sampled covariance matrices, $\mathbf{\Sigma}^i$ and $\widehat{\mathbf{\Sigma}}^i$, are

$$\mathbf{\Sigma}^i\mathbf{u}_k^i = \lambda_k\mathbf{u}_k^i \qquad \text{and} \qquad \widehat{\mathbf{\Sigma}}^i\hat{\mathbf{u}}_k^i = \hat{\lambda}_k^i\hat{\mathbf{u}}_k^i, \tag{5.4}$$

where $\mathbf{u}_k^i$, $\hat{\mathbf{u}}_k^i$ and λ_k^i, $\hat{\lambda}_k^i$ are the eigenvectors and eigenvalues of $\mathbf{\Sigma}^i$ and $\widehat{\mathbf{\Sigma}}^i$, respectively. To demonstrate the relationship between the two sets of eigenvectors, we note that

$$\begin{aligned}
\widehat{\mathbf{\Sigma}}^i\hat{\mathbf{u}}_k^i &= \hat{\lambda}_k^i\hat{\mathbf{u}}_k^i, && (5.5)\\
\mathbf{I'}^T\mathbf{I'}\hat{\mathbf{u}}_k^i &= \hat{\lambda}_k^i\hat{\mathbf{u}}_k^i, && (5.6)\\
\mathbf{I'}\mathbf{I'}^T\mathbf{I'}\hat{\mathbf{u}}_k^i &= \hat{\lambda}_k^i\mathbf{I'}\hat{\mathbf{u}}_k^i, && (5.7)\\
\mathbf{\Sigma}^i(\mathbf{I'}\hat{\mathbf{u}}_k^i) &= \hat{\lambda}_k^i(\mathbf{I'}\hat{\mathbf{u}}_k^i). && (5.8)
\end{aligned}$$

As a result $\mathbf{u}_k^i = \mathbf{I'}\hat{\mathbf{u}}_k^i$ and $\lambda_k^i = \hat{\lambda}_k^i$. This means that the eigenvectors of the explicit covariance matrix can be calculated by multiplying the centered training set by the eigenvectors of the sampled covariance matrix. Likewise, the non-zero eigenvalues of the explicit covariance matrix are equal to the eigenvalues of the sampled covariance matrix.

The eigenfaces are then the eigenvectors of $\mathbf{\Sigma}^i$ and are constructed by multiplying the centered training-set data-matrix $\mathbf{I'}$ by the eigenvectors of the sampled covariance matrix $\widehat{\mathbf{\Sigma}}^i$, i.e.

$$\mathbf{M}^i = \mathbf{I'}\widehat{\mathbf{U}}^i, \tag{5.9}$$

where $\widehat{\mathbf{U}}^i = [\hat{\mathbf{u}}_1^i|\hat{\mathbf{u}}_2^i|\cdots|\hat{\mathbf{u}}_K^i]$.

An out-of-training-sample face $\dot{\mathbf{i}}$ can be fitted to the eigenfaces $\mathbf{M}^i$ by calculating the parameter vector $\mathbf{b}^i = [b_1^i, b_2^i, \cdots, b_k^i]$ that minimizes the squared error. The solution to this least-squares estimation problem is

$$\mathbf{b}^i = \mathbf{M}^{i^T}(\dot{\mathbf{i}} - \bar{\mathbf{i}}). \tag{5.10}$$

The vector of parameters $\mathbf{b}^i$ measures the contributions from each eigenface to the recovered approximation of the out-of-training face $\dot{\mathbf{i}}$, and is given by

$$\dot{\mathbf{i}} \approx \bar{\mathbf{i}} + \mathbf{M}^i\mathbf{b}^i. \tag{5.11}$$

In order to be valid examples of the class represented by the training set, the values of the vector $\mathbf{b}^i$ should be constrained to fall in the interval $b_k \in [-3\sqrt{\lambda_k^i}, +3\sqrt{\lambda_k^i}]$.

5.3.2 The surface shape statistical models

The face database used for building the models was provided by the Max-Planck Institute for Biological Cybernetics in Tuebingen, Germany. As described in (Blanz and Vetter, 1999), this database was collected using laser scans of 200 heads of young adults. The data is stored in a cylindrical representation. For constructing the height based model, we converted the cylindrical coordinates to Cartesian coordinates and solved for the values of $Z(x, y)$. We were also provided with the ground truth surface normals for each of the faces. We used this ground truth data to construct the surface gradient, azimuthal angle and Fourier domain statistical models. We used 150 examples for constructing each of the four models. The remaining 50 out-of-training-sample example were used for performance tests.

To explain how the surface models were constructed, we commence with the height representation. Each of the K surfaces in the training set may be represented by long

vectors of height values $\mathbf{h}$. The mean height vector $\bar{\mathbf{h}}$ is given by

$$\bar{\mathbf{h}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{h}_k. \tag{5.12}$$

In a similar manner to Equation 5.2, we form the $MN \times K$ matrix of centered long vectors $\mathbf{H}' = [(\mathbf{h}_1 - \bar{\mathbf{h}})|(\mathbf{h}_2 - \bar{\mathbf{h}})|\cdots|(\mathbf{h}_K - \bar{\mathbf{h}})]$. We calculate the eigenvectors $\hat{\mathbf{u}}_k^h$ of the matrix $\mathbf{H'^T H'}$ and construct the height statistical model (as in Equation 5.9)

$$\mathbf{M}^h = \mathbf{H}'\widehat{\mathbf{U}}^h, \tag{5.13}$$

where $\widehat{\mathbf{U}}^h = [\hat{\mathbf{u}}_1^h|\hat{\mathbf{u}}_2^h|\cdots|\hat{\mathbf{u}}_K^h]$. An out-of-training-sample centered long-vector of height values, $\dot{\mathbf{h}} - \bar{\mathbf{h}}$, can be projected onto the model and represented using the vector of coefficients

$$\mathbf{b}^h = {\mathbf{M}^h}^T(\dot{\mathbf{h}} - \bar{\mathbf{h}}). \tag{5.14}$$

Let us now extend the above notation to the surface gradient and Fourier basis representations. We need two separate models for the x and y components of the representation. We use p and q to refer to the surface gradient w.r.t. x and w.r.t y, respectively. Likewise, we use f and g for the x and y components of the Fourier surface basis, i.e. $\tilde{Z}^x$ and $\tilde{Z}^y$ (see Equation 3.12). Using this notation, the statistical models for surface gradient and the Fourier surfaces basis are respectively

$$\mathbf{M}^p = \mathbf{P}'\widehat{\mathbf{U}}^p, \mathbf{M}^q = \mathbf{Q}'\widehat{\mathbf{U}}^q. \tag{5.15}$$

$$\mathbf{M}^f = \mathbf{F}'\widehat{\mathbf{U}}^f, \mathbf{M}^g = \mathbf{G}'\widehat{\mathbf{U}}^g. \tag{5.16}$$

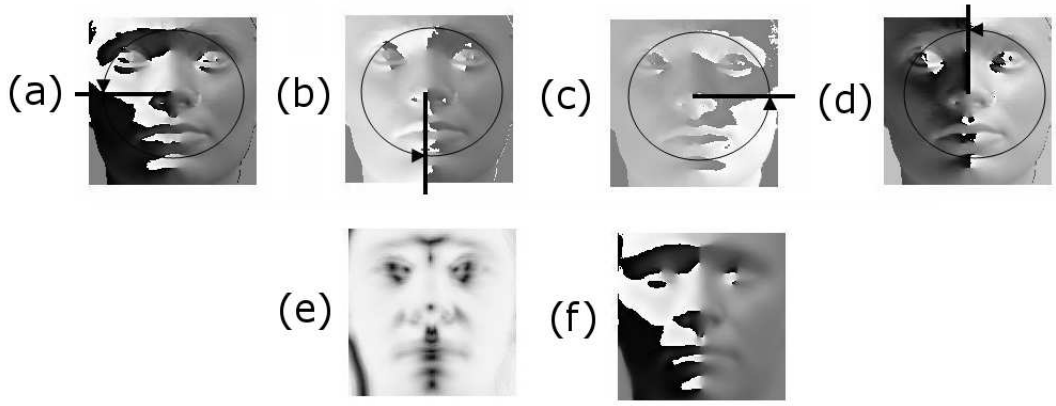

Figure 5.7: In the top row, different arguments for one training set example a^k are shown as intensity maps. From left to right, $(-\pi, \pi]$(a), $(-\frac{\pi}{2}, \frac{3\pi}{2}]$ (b), $(0, 2\pi]$ (c) and $(\frac{-3\pi}{2}, \frac{\pi}{2}]$ (d). The mean direction $\hat{\phi}$ (e) and the mean resultant length $\hat{r}$ (f) are presented in the bottom row, from left to right, as intensity plots. Note how $\hat{r}$ demonstrates that the directions of the angles are widely dispersed through the regions where the zenith angle is close to 0, i.e. tip of the nose, centers of the eyes and mouth, and forehead.

A statistical model for azimuthal angles

In the case of the azimuthal angle representation, the angles can not be used directly to construct statistical models. The reason for this is that statistical calculations performed on angular data can be biased by the angle *cut* point (see Figure 5.7). To illustrate this problem consider two points on a unit circle placed just above and just below the cut-line. Although the two points are close to one another on the unit circle, when the difference in angles is computed then this may be close to 2π.

We overcome the above problem by working with a complex number representation of the azimuth angles for the surface normals. Since we will be dealing with Lambertian images of faces, it is not necessary to construct a statistical model for the surface normal zenith angles since they be directly recovered from the arc-cosine of the measured image brightness. For the j_{th} pixel of the k_{th} training example in the database, we encode the

azimuth angle using the complex number

$$a_j^k = \exp(i\phi_j^k) = \cos\phi_j^k + i\sin\phi_j^k, \tag{5.17}$$

where $i = \sqrt{-1}$. The azimuth angle can be recovered from the real (Re) and imaginary (Im) components of the complex number using

$$\phi_j^k = \arctan\frac{\text{Im } a_j^k}{\text{Re } a_j^k}. \tag{5.18}$$

The azimuth angle ϕ_j^k is therefore the principal argument (a unique angle value from $-\pi$ to π) of a_j^k. At the image location indexed j, the mean complex number (center of mass) over the training set is given by

$$\hat{a}_j = \frac{1}{K}\sum_{k=1}^{K} a_j^k. \tag{5.19}$$

The azimuth angle associated with this complex number (mean direction) and its moduli are, respectively

$$\hat{\phi}_j = \arctan\frac{\text{Im } \hat{a}_j}{\text{Re } \hat{a}_j} \qquad \text{and} \qquad \hat{r}_j = \sqrt{(\text{Im } \hat{a}_j)^2 + (\text{Re } \hat{a}_j)^2}. \tag{5.20}$$

The Cartesian coordinates of the points of $\hat{a}_j$ on the complex plane are defined by the average of the cosines (x-axis) and sines (y-axis) of all of the observations ϕ_j^k of the training set. As a result

$$\text{Re } \hat{a}_j = \hat{r}_j\cos\hat{\phi}_j = \frac{1}{K}\sum_{k=1}^{K}\cos\phi_j^k \quad \text{and} \quad \text{Im } \hat{a}_j = \hat{r}_j\sin\hat{\phi}_j = \frac{1}{K}\sum_{k=1}^{K}\sin\phi_j^k. \tag{5.21}$$

Unfortunately, although this allows us to overcome the problems of representing the azimuth angle statistics, it yields complex numbers that no longer have unit modulus.

In fact r_j can fluctuate between 0 and 1. However, r_j is an important measure of the concentration of the azimuth angles in the training data. If the directions of the azimuth angles in the training set are strongly clustered, then r_j will tend to be 1. If, on the other hand, they are scattered then r_j will tend to 0.

Although the mean resultant length $\hat{r}_j$ is an important measure of dispersion, for the purposes of comparison with data on the line we should consider measures of dispersion based on circular data. A useful measure is the sample circular variance $v_j = 1 - \hat{r}_j$, $0 \leq v_j \leq 1$. Following (Mardia, 1972), if $1 - \cos(\alpha_1 - \alpha_2)$ is a measure of distance between two angles α_1 and α_2, then the dispersion of the angles $\phi_j^1, \phi_j^2, \ldots, \phi_j^K$ about a given angle β is

$$D(\beta) = \frac{1}{K} \sum_{k=1}^{K} \{1 - \cos(\phi_j^k - \beta)\}. \tag{5.22}$$

For any set of angular data, the dispersion of its mean direction over the set is equal to its circular variance, i.e., $D(\hat{\phi}_j) = v_j = 1 - r_j$. In Figure 5.7(e) and (f), the mean arguments $\hat{\phi}$ and the moduli of the center of mass $\hat{a}$ are shown as intensity maps.

Repeating the construction outlined in Equations 5.15 and 5.16, the complex-number based statistical model is

$$\mathbf{M}^a = \mathbf{A}' \widehat{\mathbf{U}}^a, \tag{5.23}$$

where $\mathbf{A}'$ is the centered complex training set data-matrix and $\widehat{\mathbf{U}}^a$ is the eigenvector matrix for the sampled covariance matrix. Note that the complex azimuth angle representation leads to a Hermitian covariance matrix, $\mathbf{A}'^{\dagger}\mathbf{A}'$, where † denotes the transpose of the complex conjugate matrix. Since $\mathbf{A}'^{\dagger}\mathbf{A}'$ is Hermitian, its eigenvector matrix $\widehat{\mathbf{U}}^a$ is complex, while its eigenvalues λ_k^a are real.

5.4 The parameter fitting procedure

In this section we explain the method used to fit the parameters of the models to image brightness data so that the irradiance equation is satisfied. This algorithm is similar to that proposed by Smith and Hancock (Smith and Hancock, 2005b) and draws ideas from the geometric shape-from-shading framework of Worthington and Hancock (Worthington and Hancock, 1999). However, here we add an integrability enforcement step to the parameter fitting procedure. This is done using the method of Frankot and Chellappa (Frankot and Chellappa, 1988) which we have outlined in Section 3.3.

According to the geometric approach to SFS developed by Worthington and Hancock (Worthington and Hancock, 1999), the image irradiance equation is treated as a hard constraint. Lambert's law is enforced by demanding the recovered surface normals to fall on the reflectance cone whose axis is the light source direction and whose opening angle is the inverse cosine of the normalized image brightness. Compliance with Lambert's law is effected by rotating an estimated surface normal onto the nearest location on the local irradiance cone.

We have used this technique to fit the Cartesian models to brightness images of faces. If we have a field of initial surface normals $\mathbf{n}$ estimated from from the brightness data[1], then the iterative steps in the fitting process are defined as follows:

1. Transform the field of normals $\mathbf{n}$ into each of the four Cartesian representations. From the surface gradient, the complex azimuth angle can be obtained using Equation 5.17. Likewise, the height and Fourier basis representations are obtained by integrating the field of surface normals using Equation 3.12.

2. For each representation, subtract the mean shape and calculate the corresponding

[1] We used a standard image gradient initialization.

set of shape parameters using one of the following

$$\mathbf{b}^h = \mathbf{M}^{h^T}(\dot{\mathbf{h}} - \bar{\mathbf{h}}), \tag{5.24}$$

$$\mathbf{b}^a = \mathbf{M}^{a^T}(\dot{\mathbf{a}} - \bar{\mathbf{a}}), \tag{5.25}$$

$$\mathbf{b}^p = \mathbf{M}^{p^T}(\dot{\mathbf{p}} - \bar{\mathbf{p}}) \quad \text{and} \quad \mathbf{b}^q = \mathbf{M}^{q^T}(\dot{\mathbf{q}} - \bar{\mathbf{q}}), \tag{5.26}$$

$$\mathbf{b}^f = \mathbf{M}^{f^T}(\dot{\mathbf{f}} - \bar{\mathbf{f}}) \quad \text{and} \quad \mathbf{b}^g = \mathbf{M}^{g^T}(\dot{\mathbf{g}} - \bar{\mathbf{g}}), \tag{5.27}$$

depending on the relevant representation.

3. Recover the surface shape from the best-fit parameters using one of the following

$$\dot{\mathbf{h}} \approx \bar{\mathbf{h}} + \mathbf{M}^h\mathbf{b}^h, \tag{5.28}$$

$$\dot{\mathbf{a}} \approx \bar{\mathbf{a}} + \mathbf{M}^a\mathbf{b}^a, \tag{5.29}$$

$$\dot{\mathbf{p}} \approx \bar{\mathbf{p}} + \mathbf{M}^p\mathbf{b}^p \quad \text{and} \quad \dot{\mathbf{q}} \approx \bar{\mathbf{q}} + \mathbf{M}^q\mathbf{b}^q, \tag{5.30}$$

$$\dot{\mathbf{f}} \approx \bar{\mathbf{f}} + \mathbf{M}^f\mathbf{b}^f \quad \text{and} \quad \dot{\mathbf{g}} \approx \bar{\mathbf{g}} + \mathbf{M}^g\mathbf{b}^g. \tag{5.31}$$

4. Apply the integrability constraint. This is done by generating a surface from the best fit parameters of the surface gradient and azimuthal angle representations, $\dot{\mathbf{p}}$, $\dot{\mathbf{q}}$ and $\dot{\mathbf{a}}$.

5. From the reconstructed surfaces we calculate a field of surface normals. This is done by performing a bicubic patch fit to the surface height data. We enforce the irradiance constraint by rotating the recovered surface normals onto the irradiance cone using Equation 3.2. We then return to step 1.

Instead of searching for valid linear eigenmode combinations that minimize the bright-

ness error using exhaustive search, the parameter fitting procedure attempts to minimize the brightness error using simple geometric operations that satisfy the irradiance constraint provided by Lambert's law. The method hence provides an intuitive and straightforward way for adjusting the shape coefficients to image brightness data.

5.5 Experiments

Our experimental evaluation is divided into two parts. First, we provide an experimental comparison for the four Cartesian representations. Second, we focus on the performance of each representation when fitted to image brightness data.

5.5.1 Comparing the models

The generalization of the models, or their ability to capture the features of the database from which they were built, is illustrated in Figure 5.8. We show the generalization as a function of the number of modes used. The required number of modes was calculated through the formulae $\sum_{l=1}^{t} \lambda_L \geq f_v V_T$, where λ_l are the eigenvalues of the sampled covariance matrix, V_T is the total variance (i.e. the sum of all the eigenvalues) and f_v defines the proportion of the total variation to be conserved by the model. Both $(\mathbf{M}^f, \mathbf{M}^g)$ (dotted lines) and $\mathbf{M}^h$ (dashed-dotted line) achieve more than 90% when using at least 20 modes. By contrast, $\mathbf{M}^a$ (dashed line) and $(\mathbf{M}^p, \mathbf{M}^q)$ (solid lines) required a considerably larger number of modes to achieve the 90% level. Interestingly, for both $(\mathbf{M}^f, \mathbf{M}^g)$ and $(\mathbf{M}^p, \mathbf{M}^q)$ the x-related model shows a slightly better generalization than the y-related model. This may be attributable to the left-to-right symmetry of human faces. It is important to remark that the similarities between the retained variances shown by $\mathbf{M}^h$ and $(\mathbf{M}^f, \mathbf{M}^g)$ can be regarded as a consequence of both models being based on height information, one (the height surface) being the sum of two component surfaces (the Fourier

basis surfaces).

Figure 5.9 shows the percentage of error for out-of-training cases when the number of eigenmodes is varied. The plot shows the height difference error $\|H^{actual} - H^{eigenmode}\|/H^{actual}$ computed as a function of the number of modes used in the representation, for the 50 out-of-sample faces. For the models $(\mathbf{M}^p, \mathbf{M}^q)$ and $\mathbf{M}^a$, the Frankot and Chellappa's integrability method(Frankot and Chellappa, 1988) was used to recover surface from surface gradient. For recovering the surfaces from the model $(\mathbf{M}^f, \mathbf{M}^g)$, Equation 3.12 was used. As we assume Lambertian reflectance, the zenith angle was calculated directly from the albedo-free images and then be used in conjunction with the recovered azimuth angle to obtain a surface normal estimate and hence recover a height map through surface integration. The figure shows the average over the 50 surfaces and over all the points on the surface. Here $(\mathbf{M}^f, \mathbf{M}^g)$ (dotted line) achieves the lowest percentage of error (up to around 0.3%), followed by $\mathbf{M}^h$ (around 0.5%), $\mathbf{M}^a$ (around 0.8%), and $(\mathbf{M}^p, \mathbf{M}^q)$ (solid line, around 0.9%) with full number of eigenmodes used. This behavior could be explained from the results in Figure 5.8, where it is clear that both $(\mathbf{M}^p, \mathbf{M}^q)$ and $\mathbf{M}^a$ require a larger number of modes for characterizing most of the model features. On the other hand, from its original error (zero modes: the mean shape), 1.91%, a steep error change occurs after using 10 modes for the three models. It is important to mention that, visually the effect of errors less than 1% is close to being negligible.

In Figure 5.10 we show the first six modes of variation for the models described above. The different rows in the figure correspond to different eigenmodes. For each eigenmode, we have two rows in the figure. The top and bottom row in each pair show the result of varying the eigenmode by ± 3 standard deviations from the mean. The columns of the figure have been divided into four groups of images, labeled with the letters (a), (b), (c) and (d). Each group gives the results obtained with one of the representation used for constructing the statistical models. Group (a) shows the variations for the intensity model

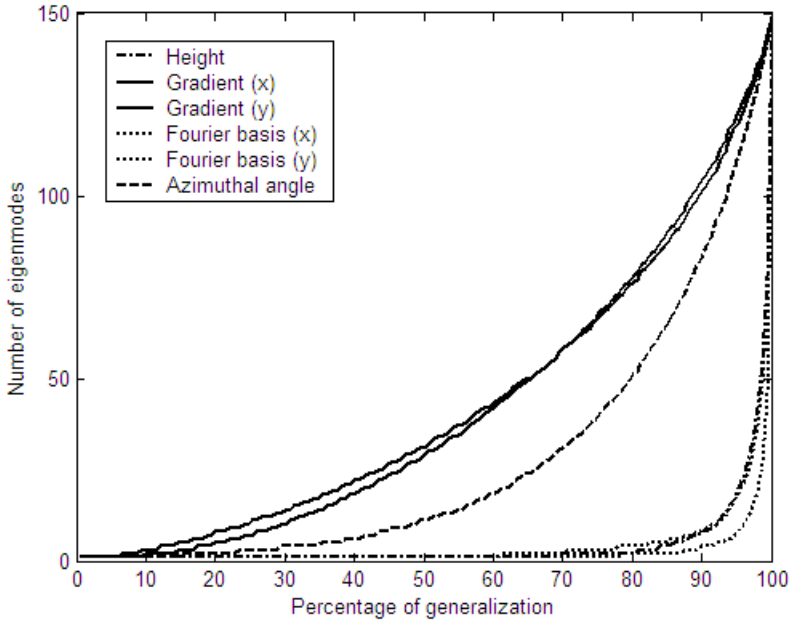

Figure 5.8: Generalization of the models. The figure presents the retained variance of each model as a function of the number of modes used. Both models $\mathbf{M}^h$ and $(\mathbf{M}^f, \mathbf{M}^g)$ tend to encapsulate most of the variance of the training set in the first few modes, while $(\mathbf{M}^p, \mathbf{M}^q)$ and $\mathbf{M}^a$ clearly need a bigger number of modes to achieve at least a %90 of the generalization.

$\mathbf{M}^i$. In group (b), from left to right, we show the modes for $\mathbf{M}^p$, $\mathbf{M}^q$ and the frontal illumination of their integrated surface. The frontal illumination of variations in $\mathbf{M}^h$ are shown in group (c). Finally, frontal illuminations of the surface variations corresponding to $\mathbf{M}^f$ and $\mathbf{M}^g$ are shown in group (d). The right-most illumination in group (d) is the result of summing the two Fourier surfaces (i.e. using Equation 3.12).

In Figure 5.11 we compare the eigen-modes obtained using models for the azimuthal angle. Here we compare the results of using a model based on complex numbers and one based on real numbers. The columns of the figure show the first six eigen-modes. In the top two rows we show the results obtained by using the complex model, and in the bottom two rows the results for the real model. For each model the two rows show the result of varying the different eigenmodes by ± 3 standard deviations from the mean. In general, both models seem to encapsulate the same facial features, though the complex

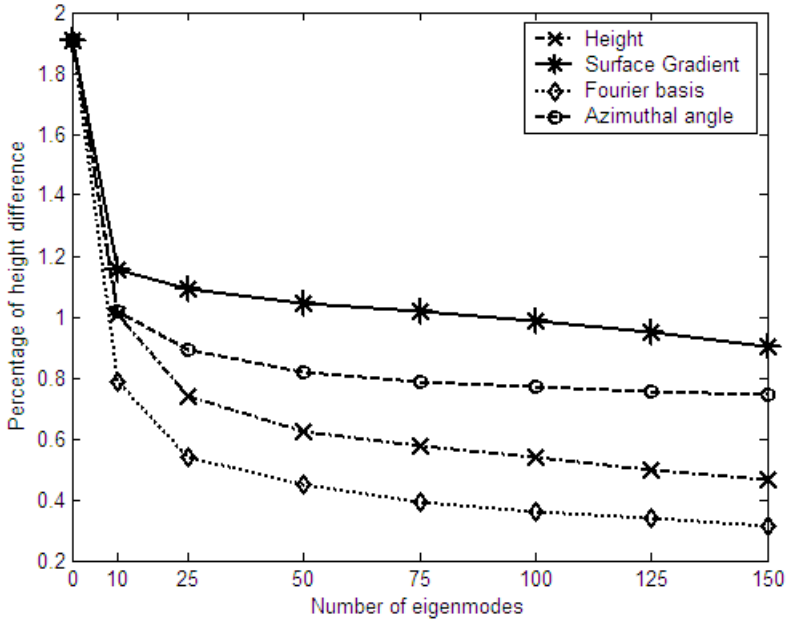

Figure 5.9: Out-of-training best-fit tests. The figure shows the percentage of error for out-of-training cases while varying the number of modes.

model shows less noise than the real model. These errors are most evident where $\hat{r}_j$ is near zero. This suggests that the complex representation gains advantage from using the center of mass $\hat{a}_j$, which might be sacrificed by being projected onto the unit circle while calculating the mean direction $\hat{\phi}_j$.

We compare the performance of the real and complex azimuth angle models by fitting them to out-of-sample data (i.e. data not used in training). These results are shown in Figure 5.12. In the top row the model has been fitted to a male subject and the bottom row shows the result of fitting the model to a female subject. We show two panels of results. In the left panel we present the ground truth data, the result of fitting the complex model and the result of fitting the real model. The main feature to note from the panel is that the complex model achieves more accuracy for regions where the zenith angle is small. In the rightmost panel we show the absolute angular difference[2] averaged over 50 out-of-

[2]The angular difference between the angles α_1 and α_2, in radians, can be defined as $\pi - \|\pi - \|\alpha_1 - \alpha_2\|\|$.

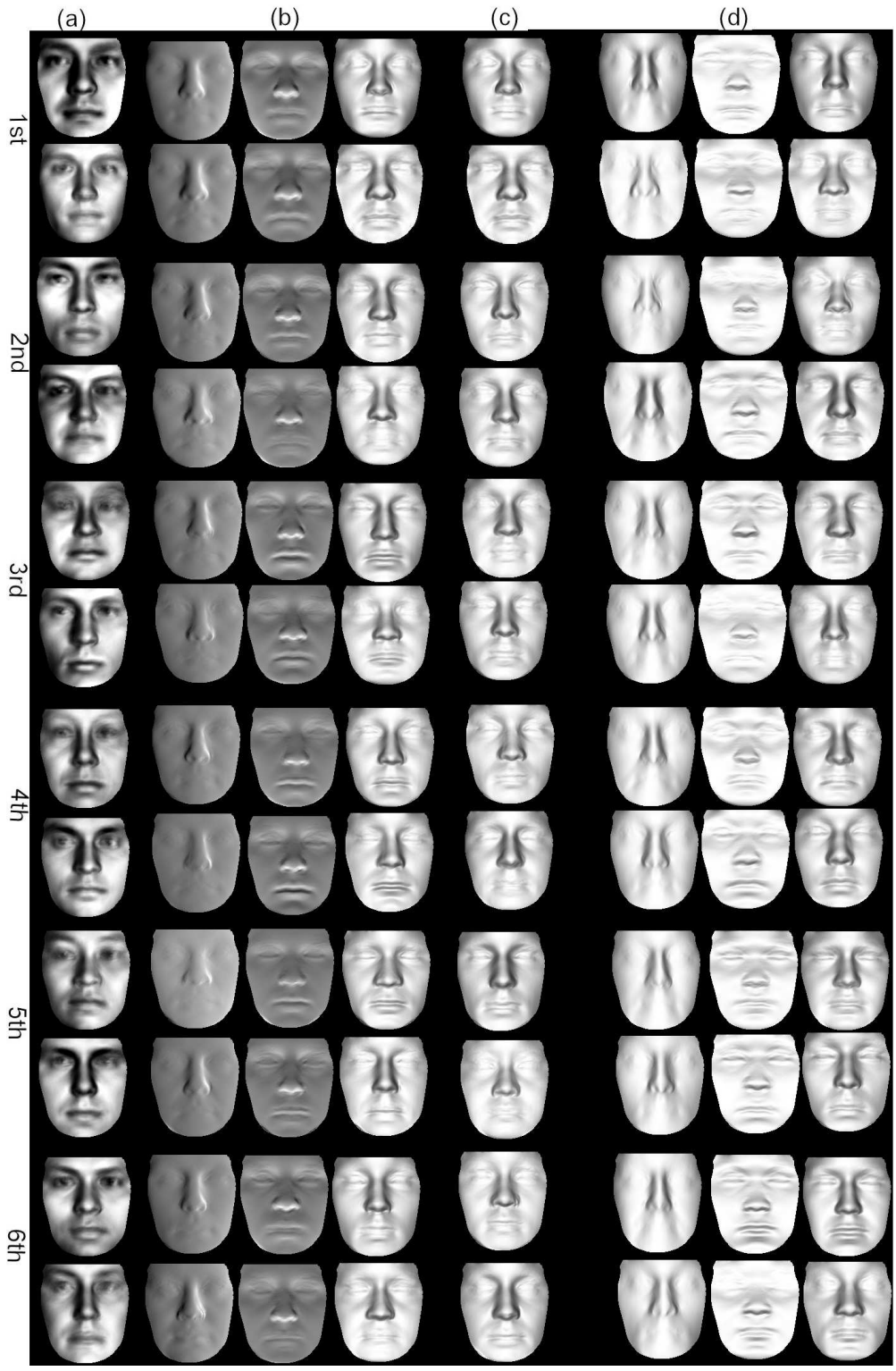

Figure 5.10: First six modes of variation for the intensity (a), surface gradient (b), height (c) and Fourier basis (d) models.

Figure 5.11: From left to right, the first six eigen-modes of the complex azimuth model (a) and the real azimuth model (b). The two first rows represent, respectively, +3 and -3 standard deviations from the mean. Likewise, these variations are shown in the two rows of (b).

training examples as a function of the number of eigenmodes used for the complex and real models. From the plot it is clear that the complex model outperforms the real model. The behavior of both models is similar, and the gap between the lines can be explained as a consequence of the poorly recovered regions for the real-valued model.

In Figure 5.13 we present the results obtained by varying the number of eigenmodes for an out-of-training case. From the example shown in the top row of the figure, we calculated shape parameters and projected them back to each representation domain by adding the corresponding mean shape. We repeated this process, varying the number of eigenmodes from 10 to 150 modes. This experiment is illustrated in the columns of the figure, which show frontal re-illuminations with the different number of eigenmodes. A final recovered surface is shown in the right-most column. The different rows in the figure are for the models $\mathbf{M}^h$, $(\mathbf{M}^p, \mathbf{M}^q)$, $(\mathbf{M}^f, \mathbf{M}^g)$ and $\mathbf{M}^a$. Only for the height and Fourier

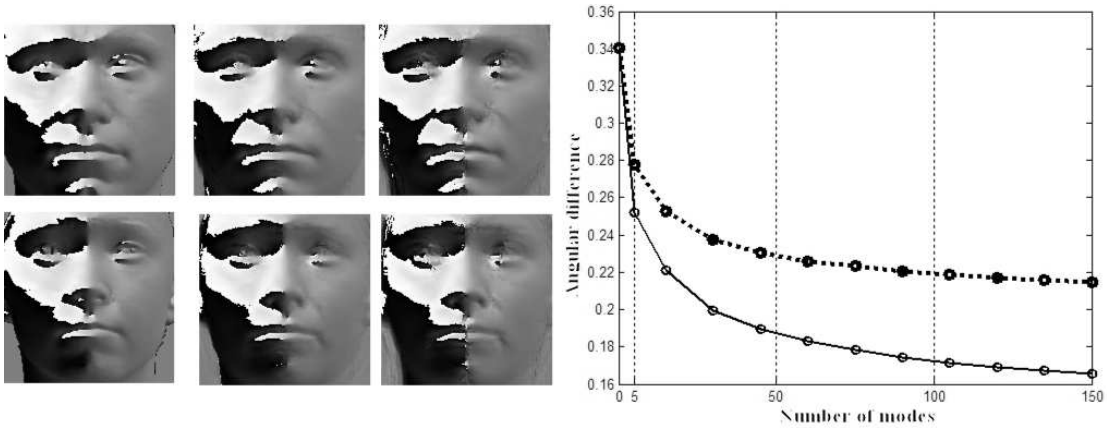

Figure 5.12: Out-of-sample recovery analysis. From left to right: the first three columns show the ground truth azimuth angle, recovered azimuth angle using the complex and real models. The rightmost diagram shows the angular difference averaged over 50 out-of-sample data as a function of number of eigenmodes used for the complex (solid line) and the real (dashed line) models.

basis models does setting the number of modes to more than 25 introduce spatial artifacts. This is noticeable over facial regions such as the nose, mouth and eye-sockets. The first 25 eigenmodes seem to be sufficient to encapsulate the most salient facial features. The remaining eigenmodes appear to capture noise rather than fine facial details, as expected from the results shown by Figure 5.8.

The models behavior on both $(\mathbf{M}^p, \mathbf{M}^q)$ and $\mathbf{M}^a$ shows a rather different behavior. Here apparently little change occurs while increasing the number of eigenmodes. These results support those in Figure 5.9 and the smooth variation in the descent of the solid and dashed lines in Figure 5.8. Note how the surfaces generated by $(\mathbf{M}^p, \mathbf{M}^q)$ and $\mathbf{M}^a$ suggest a smoothed version of those recovered by $\mathbf{M}^h$ and $(\mathbf{M}^f, \mathbf{M}^g)$.

5.5.2 Fitting the model from brightness images

Equation 5.10 shows how to determine the shape coefficients from a long-vector of data in the appropriate representation. However here we would like to fit the different models

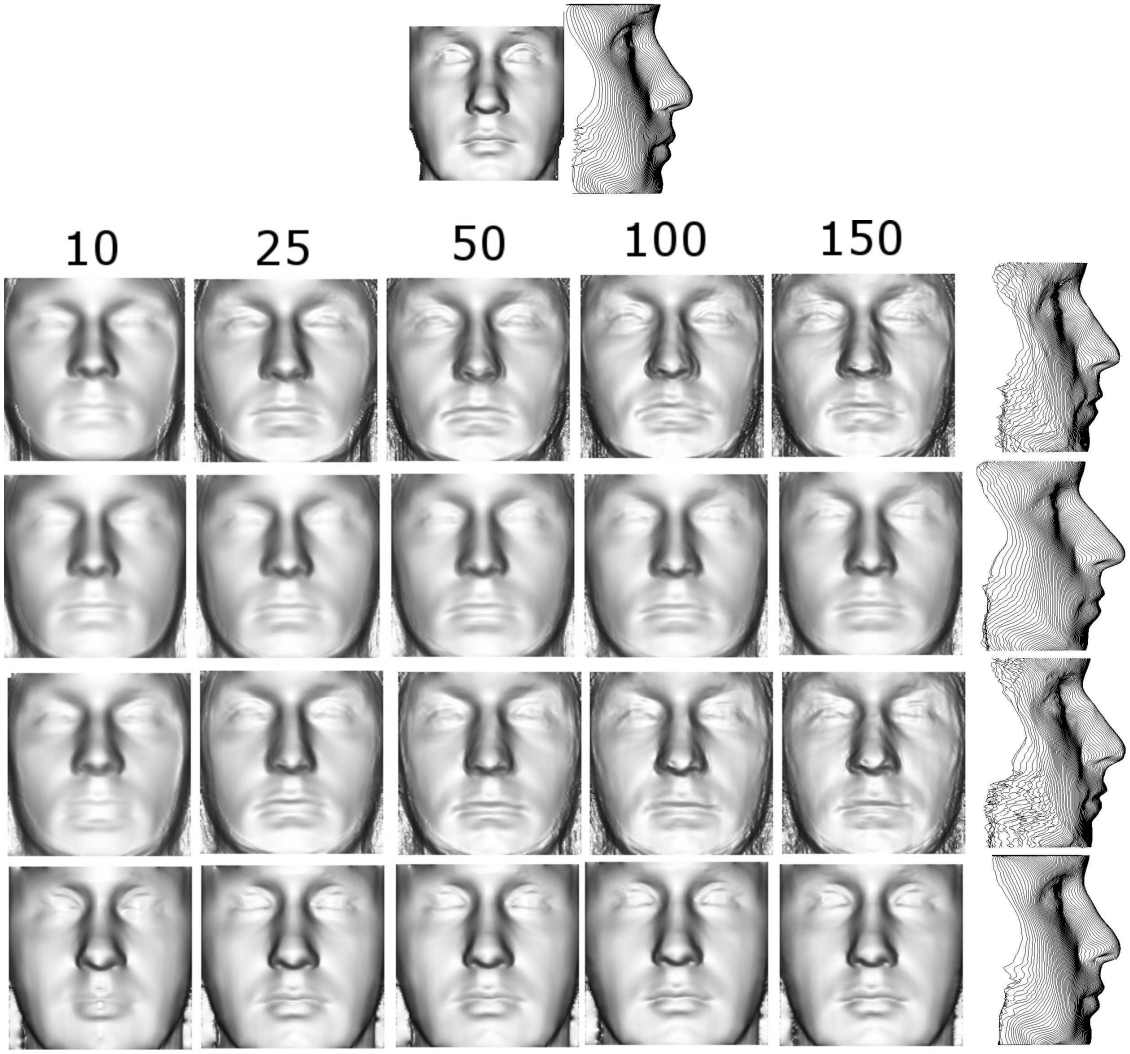

Figure 5.13: Varying the number of eigenmodes for an out-of-training case. The top row shows frontal illuminations and ground truth profile of the example used for the experiment. The following rows present the results using $\mathbf{M}^h$, $(\mathbf{M}^p, \mathbf{M}^q)$, $(\mathbf{M}^f, \mathbf{M}^g)$ and $\mathbf{M}^a$, respectively. From the first to the fifth column we present frontal re-illuminations and the final recovered surface is shown at the right-most column. The number of modes used (from 10 to 150) is indicated at the top of the panel. Note how the final recovered surface using $(\mathbf{M}^p, \mathbf{M}^q)$ and $\mathbf{M}^a$ suggest to be smoother than the best fits using $\mathbf{M}^h$ and $\widetilde{\mathbf{F}}$.

to long-vectors elicited from image data using shape-from-shading. When the iterative fitting procedure described in Section 5.4 is applied, then it never converges to a feasible solution for the models based on height information, $\mathbf{M}^h$ and $(\mathbf{M}^f, \mathbf{M}^g)$. However, the surface gradient and complex azimuth angle based models, $(\mathbf{M}^p, \mathbf{M}^q)$ and $\mathbf{M}^a$ work better.

To overcome this problem we redefined the starting state for both $\mathbf{M}^h$ and $(\mathbf{M}^f, \mathbf{M}^g)$, applying geometric SFS over the surface normals of the mean height surface. In this manner we forced the mean surface normals to satisfy the irradiance cone constraints. The resulting height maps using this initialization resulted in an error of around 1.4%.

Figure 5.14 (top) shows the percentage of error as a function of the number of iterations for the fitting procedure for each of the four representations. From the diagram it is clear that only the complex azimuth angle (dashed line) and, more noticeably, the surface gradient (solid line) based model behave well with the fitting procedure. On the other hand, the height based model (dash-dotted line) improves until it reaches the third iteration. From this point the error increases. This effect is more pronounced for the Fourier basis model (dotted line), which improves only in the first iteration. Note how the initial state is different for the models derived from directional information, i.e. $(\mathbf{M}^p, \mathbf{M}^q)$ and $\mathbf{M}^a$, and the ones based on height data, $\mathbf{M}^h$ and $(\mathbf{M}^f, \mathbf{M}^g)$. The average error for the initial state for both $\mathbf{M}^h$ and $(\mathbf{M}^f, \mathbf{M}^g)$ is around 1.4%. For clarity, we omitted the initial state of $(\mathbf{M}^p, \mathbf{M}^q)$ and $\mathbf{M}^a$ on the diagram, since it was of about 15% (approximately ten times less accurate than that of $\mathbf{M}^h$ and $(\mathbf{M}^f, \mathbf{M}^g)$).

In their few first iterations both $\mathbf{M}^h$ and $(\mathbf{M}^f, \mathbf{M}^g)$ seem to slightly outperform $(\mathbf{M}^p, \mathbf{M}^q)$ and $\mathbf{M}^a$ in its final iteration. This does not mean that the method gives better surface fits, but that the method of coupling the model to intensity data does not give an accurate measure of error. In other words, models that are based on height information are not suitable for the fitting technique adopted here. However, the approach is a natural one for the

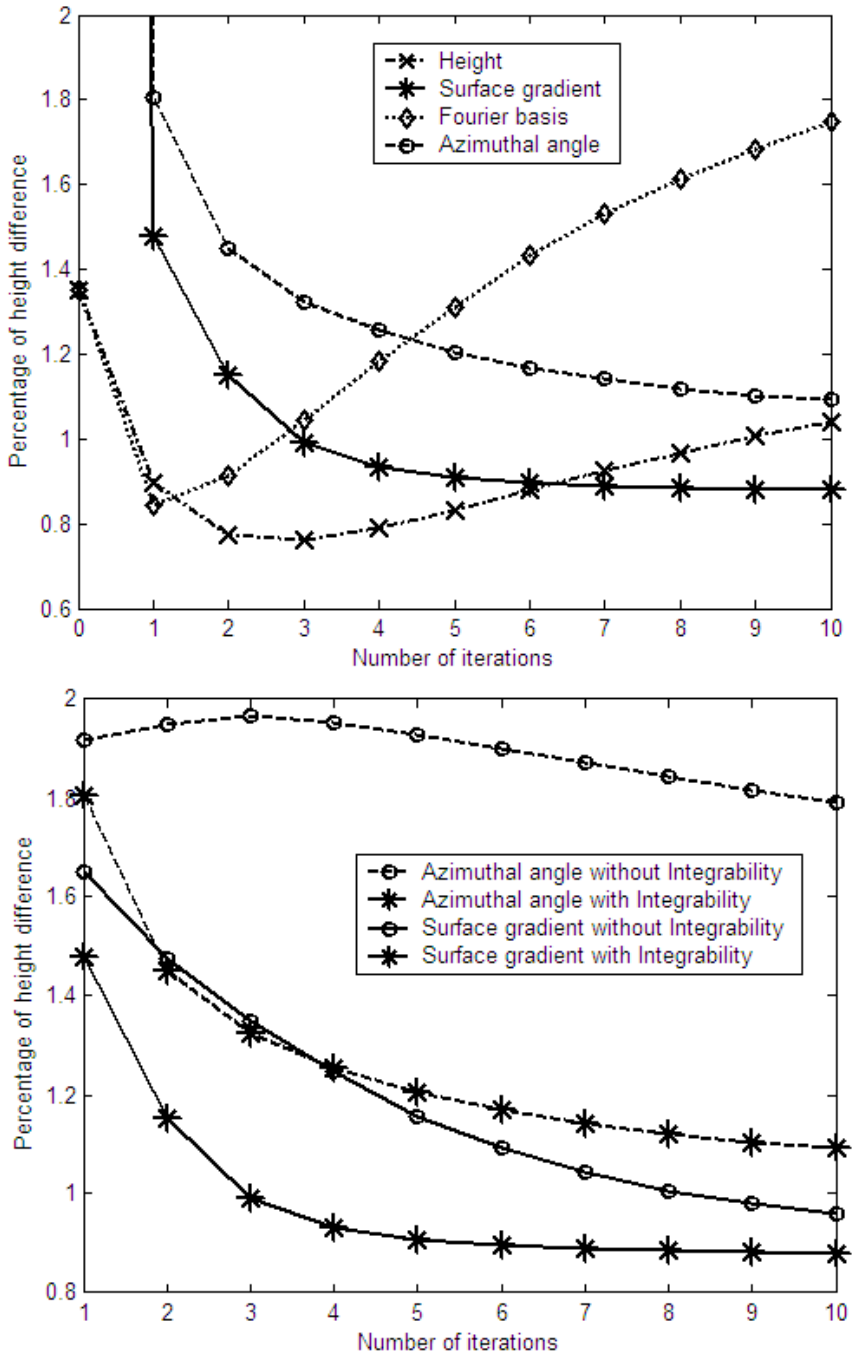

Figure 5.14: The above figure presents the percentage of error as a function of number of iterations. Note how only the models based on directional information, $(\mathbf{M}^p, \mathbf{M}^q)$ and $\mathbf{M}^a$ seem to be favored by the fitting procedure. On bottom figure shows the behavior of the fitting procedure with (asterisk markers) and without (circular markers) the integrability enforcement step are shown for $(\mathbf{M}^p, \mathbf{M}^q)$ (solid line) and $\mathbf{M}^a$ (dashed line). The results suggest that enforcing integrability benefits the fitting method.

models $(\mathbf{M}^p, \mathbf{M}^q)$ and $\mathbf{M}^a$ which are based on surface orientation and not surface height, since it is surface orientation that is responsible to the perceived image brightness.

We also tested $(\mathbf{M}^p, \mathbf{M}^q)$ and $\mathbf{M}^a$ starting with the same initial state as $\mathbf{M}^h$ and $(\mathbf{M}^f, \mathbf{M}^g)$, but the final result did not show a significant difference to the one presented in Figure 5.14 (top).

In Figure 5.14 (bottom), we show the behavior of the orientation-based models $(\mathbf{M}^p, \mathbf{M}^q)$ (solid line) and $\mathbf{M}^a$ (dashed line) omitting the integrability enforcement step in the fitting procedure. The asterisks indicate integrability enforcement, whilst the circles indicate omission of the integrability step. The diagram shows poor results when integrability is not enforced. This is most marked in the case of the complex azimuth angle representation, where the error goes from 1.09% to almost 2%.

A more detailed analysis is presented in Figure 5.15. The figure is divided into two panels. Each panel presents the results for an out-of-training example, whose frontal re-illumination and ground-truth profile view are shown at the top row of the figure. For each panel, the columns contain the recovered surface after a given number of iterations (from 0 to 10). The results obtained using $\mathbf{M}^h$, $(\mathbf{M}^p, \mathbf{M}^q)$ and $(\mathbf{M}^f, \mathbf{M}^g)$ and $\mathbf{M}^a$ are shown from second to bottom row. Note how the initial state is much poorer for $(\mathbf{M}^p, \mathbf{M}^q)$ and $\mathbf{M}^a$, and how after each iteration the integrated surfaces moves closer to the ground truth.

Finally, we present experiments with a number of real world face images. These images are drawn from the Yale B database (Georghiades et al., 2001) and are disjoint from the data used to train the statistical model. In the images, the faces are in the frontal pose and were illuminated by a point light source situated approximately in the viewer direction. We aligned each image with the mean intensity shape so that the eyes, nose tip and mouth center were in the same position. We then performed a Lambertian correction (Smith et al., 2004) over the aligned images. The surface recovery results after twenty iterations are shown in Figure 5.16. From left to right we show the corrected

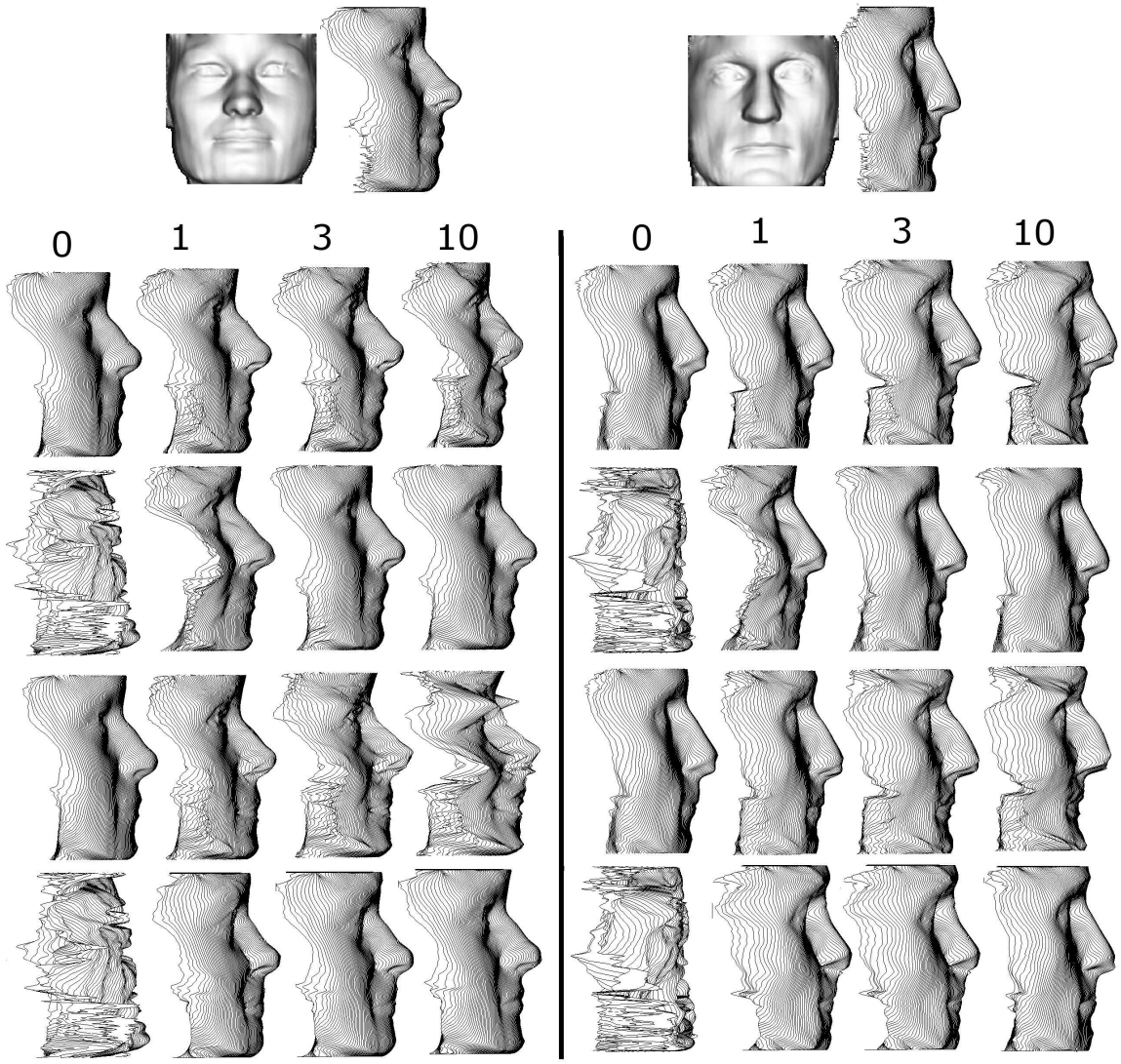

Figure 5.15: Individual analysis for the fitting procedure. The figure is divided into two panels. Each panel contains the results of applying the fitting procedure to the brightness images shown at the top row of the figure. In the top row, the ground-truth profile view is shown along with the brightness input image. The columns of each panel present the recovered surface after certain number of iterations (from 0 to 10). From second to bottom row we present the results for $\mathbf{M}^h$, $(\mathbf{M}^p, \mathbf{M}^q)$, $(\mathbf{M}^f, \mathbf{M}^g)$ and $\mathbf{M}^a$.

Lambertian input image, frontal re-illumination of the recovered surface followed by two rotated views. We used the surface gradient representation with enforced integrability. Although the method struggled to recover the shape of the eye sockets, the overall structure of the face is well reconstructed, On the other hand, the eyebrow location, nose length and width of the face clearly match those of the input images, even when there is facial hair present.

5.6 Conclusions

We have presented an analysis of four Cartesian representations for constructing three-dimensional statistical models of faces. We also showed how to fit the models to brightness images of faces using irradiance equation and integrability constraints. All of the models work well when used to fit data in the same form. However, for fitting to image brightness data to recover facial shape, only the models based on directional representations, i.e. surface gradient and complex azimuthal angle proved robust. Hence we have demonstrated that 3D statistical models of faces based on Cartesian representations of orientation data can work accurately without special heuristics. As future work we are planning to explore the behavior of the Cartesian representations with alternative methods for shape coefficient adjustment as well as dealing with albedo changes.

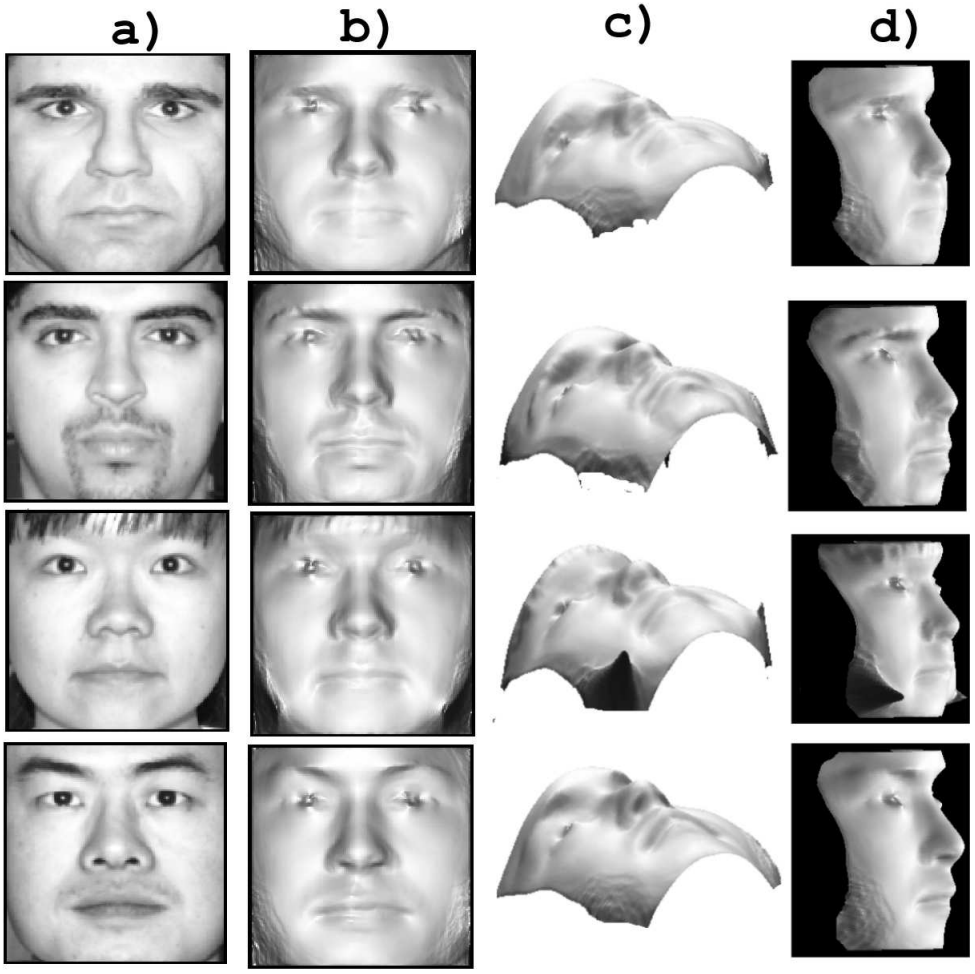

Figure 5.16: Applying the method on four real images of faces. From left to right we show the corrected Lambertian input image, frontal re-illumination of the recovered surface and two views of the recovered surface.

Chapter 6

Coupled Statistical Models of Intensity and Shape

6.1 Introduction

In this chapter, our aim is to explore whether the alternative representations for surface shape described in Chapter 5 can be coupled to variations in image intensity using a coupled statistical model. The method described in the last chapter performed well under Lambertian conditions, however, errors were spread if varying albedo images were used. Here we aim to overcome this problem by using a statistical model to learn the link between image brightness variations and surface shape.

The coupled model is inspired by the active appearance model developed by Cootes, Edwards and Taylor (Cootes et al., 1998), which simultaneously models 2D shape and texture. Here however, we model joint variations in image brightness and surface shape. We explore three different surface shape representations. These are the surface height function, the surface gradient and a Fourier basis representation. The model is trained using corresponding pairs of range images and brightness images. From the range images, we

extract the shape-attributes. We construct separate eigenspaces for the image brightness variation and the surface shape variation from the covariance matrices of the training data. The coupled model links the two eigenspaces. By fitting the image brightness model to input images, we are able to recover the corresponding parameters of surface shape. From the surface model parameters we recover facial shape.

Let us analyze again Figure 5.10, where the different eigenmodes for each surface shape representation are rendered. An important feature to note is that different models tend to encapsulate different shape characteristics. It is interesting to note that the intensity model and the surface models exhibit different modes of variation. This means that the information encoded by the intensity shape parameters, $\mathbf{b}^i$, is of limited use in directly recovering surface shape from intensity images. This problem has been circumvented by minimizing the distance between rendered views from recovered surfaces and input images, as in the work of Atick (Atick et al., 1996) and Blanz and Vetter (Blanz and Vetter, 2003). Unfortunately, the minimization of this distance is badly affected by the presence of local minima. This means that exhaustive search methods must be used, and this sacrifices efficiency. We overcome this problem by using a coupled statistical model to relate 2D intensity variations and variations in surface shape. Once fitted to data, the coupled model allows us to infer the shape-parameters from the best-fit intensity parameters, rather than using the distance between input images and rendered views of the recovered surfaces.

6.2 The coupled models

In this section we focus on how to relate the modes of variation of the intensity and surface shape models. To this end, we develop a coupled statistical model that links the coefficients of the intensity and shape-models for a set of training data. Here the training

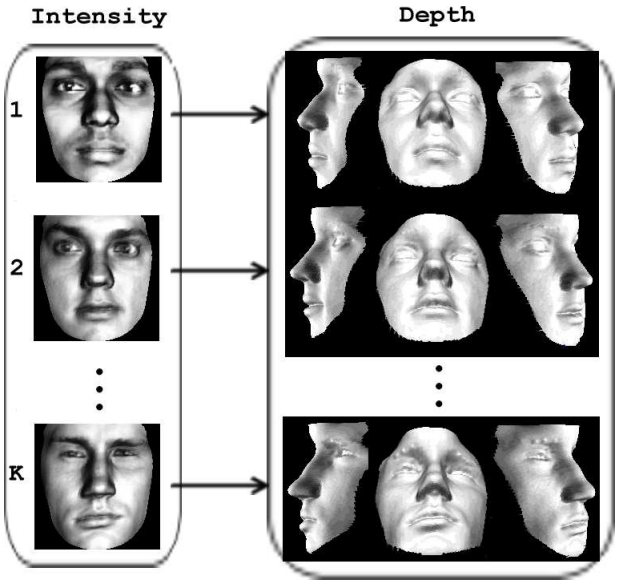

Figure 6.1: We used a database of intensities and a database of height information. As shown in the figure, there is a direct (pixel to pixel) relation between the members of each database. From the height database we generated the Surface Gradient and Fourier basis databases.

data consists of aligned pairs of intensity and surface shape images of the faces of different subjects. By fitting the intensity model to out-of-training-sample images of faces, we can use the coupled model to recover the corresponding surface shape parameters.

To construct the coupled model, we draw on the intensity and surface representations described in Sections 5.3.1 and 5.3.2 respectively. Each training example i.e. pair of intensity image and surface representation extracted from a corresponding aligned range image (see Figure 6.1). can be summarized by the parameter vectors $\mathbf{b}^i$ and $\mathbf{b}^x$, where x represents the surface model (either height, surface gradient or Fourier basis) of the sample. In both models, we assume that the lower eigenmodes represent small scale noise variation. Hence, if the k_{th} eigenvalue for the intensity model is λ_k^i, we need only S eigenmodes to retain $Perc$ percent of the model variance. We choose S so that $\sum_{k=1}^{S} \lambda_k^i \geq \frac{Perc}{100} \sum_{k=1}^{K} \lambda_k^i$. Similarly, for the 3D models we retain T eigenmodes to capture $Perc$ percent of the variance.

It is important to mention that, unlike the experiments presented in Chapter 5, in this chapter we use a mask to cut off the regions of the face near the neck and ears (i.e. the ones with the highest spatial variance). This binary mask consists of calculating the total variance of the height training set and setting the high variance pixels to zero. The mask is then applied to every example in the training sets. The effect of the masking functions is to avoid the high spatial variability introduced by the surrounding regions of the face in each training sample.

6.2.1 Eigenmode concatenation

For the k_{th} training sample we can generate the concatenated parameter vector of length $S + T$:

$$\mathbf{b}_k^c = \begin{pmatrix} \mathbf{W}\mathbf{b}_k^i \\ \mathbf{b}_k^x \end{pmatrix} = \begin{pmatrix} \mathbf{W}(\mathbf{M}^{i^T}(\mathbf{t}_k^i - \bar{\mathbf{i}})) \\ \mathbf{M}^{x^T}(\mathbf{t}_k^x - \bar{\mathbf{x}}) \end{pmatrix}, \tag{6.1}$$

where $\mathbf{W}$ is a diagonal matrix of weights for each intensity model parameter, allowing for the different relative weighting of the intensity and surface models. As the elements of $\mathbf{b}^i$ and $\mathbf{b}^x$ represent different classes of data (grayscale and surface shape), they can not be compared directly. We follow Cootes and Taylor (Cootes et al., 1998) and set $\mathbf{W} = rI$, where r^2 is the ratio of the total shape variance to the total intensity variance and I is the identity matrix . The coupled model data matrix is $(\mathbf{b}_1^c|\mathbf{b}_2^c|...|\mathbf{b}_K^c)$.

By applying PCA to the concatenated intensity-shape parameter vectors, we obtain the coupled model:

$$\mathbf{b}^c = \mathbf{C}\mathbf{c} = \begin{pmatrix} \mathbf{C}^i \\ \mathbf{C}^x \end{pmatrix} \mathbf{c}, \tag{6.2}$$

where $\mathbf{C}$ are the eigenvectors and $\mathbf{c}$ is a vector of coupled parameters controlling the inten-

sity and surface shape models simultaneously. The matrix $\mathbf{C}^i$ has S rows, and represents the first S eigenvectors, corresponding to the intensity subspace of the model. The matrix $\mathbf{C}^x$ has T rows, and represents the final T eigenvectors, corresponding to the surface shape subspace of the model.

We may express the vectors of projected intensity and 3D values directly in terms of the parameter vector $\mathbf{c}$:

$$\mathbf{t}^i = \bar{\mathbf{i}} + \mathbf{M}^i\mathbf{W}^{-1}\mathbf{C}^i\mathbf{c}. \tag{6.3}$$

$$\mathbf{t}^x = \bar{\mathbf{x}} + \mathbf{M}^x\mathbf{C}^x\mathbf{c}. \tag{6.4}$$

For compactness we write: $\mathbf{Q}^i = \mathbf{W}^{-1}\mathbf{C}^i$.

A plot of cumulative variance versus number of eigenmodes used is shown in Figure 6.2. The surface shape, intensity and coupled models are represented by the dashed, solid and dotted lines respectively. It is evident that fewer eigenmodes are required to the capture variance in facial depth (i.e. the height and Fourier basis models) than in facial intensity. This is because the intensity model has to deal with variations caused by changes in both facial shape and illumination. The depth-based models, on the other hand, need only accommodate changes in facial shape. We retained 95% of the variance for each of the models.

6.2.2 Fitting the models to intensity data

Fitting the model to intensity data involves estimating the parameter vector $\mathbf{c}$ from input images of faces. To do this we seek the coupled model parameters which minimize the error between the best fit parameters $\mathbf{b}^i$ and the recovered parameters $\mathbf{Q}^i\mathbf{c}$. In doing so, we implicitly recover the surface shape represented by the coupled model parameters.

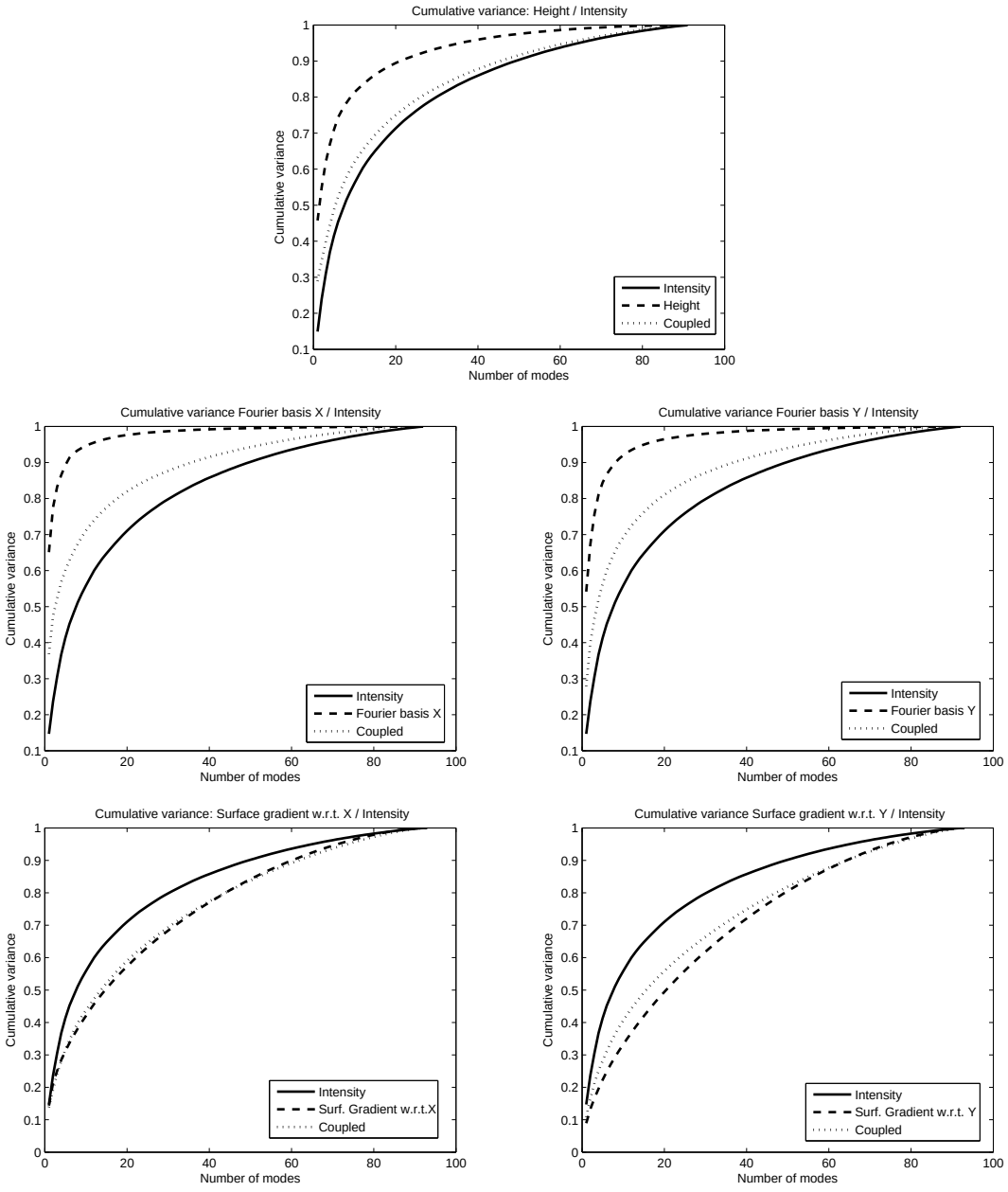

Figure 6.2: Plot of cumulative variance versus number of eigenmodes used for each surface shape model (dashed line), intensity model (solid line) and full coupled models (dotted line).

Suppose that $\mathbf{t}^i$ is a centered vector of length $M \times N$ that represents an intensity image of a face. Its best fit parameter vector, $\mathbf{b}^i$, is calculated using Equation 5.10. We fit the model to data seeking the vector $\mathbf{c}$ of length $S + T$ that satisfies the condition

$$\mathbf{c} = \arg\min_{\mathbf{c}} \{(\mathbf{b}^i - \mathbf{Q}^i\mathbf{c})^T(\mathbf{b}^i - \mathbf{Q}^i\mathbf{c})\} \tag{6.5}$$

The corresponding best fit vector of surface shape values is given by

$$\mathbf{t}^x \approx \bar{\mathbf{x}} + \mathbf{M}^x\mathbf{C}^x\mathbf{c} \tag{6.6}$$

We used a Matlab implementation of the quasi-Newton minimization procedure to solve Equation 6.5. The fit was constrained such that each coupled parameter lies within ± 3 standard deviations from the mean. One input image took around a couple of seconds to converge to the best solution.

6.3 Experiments

In this section we report experiments focused on using out-of-training-sample images to evaluate the ability of the coupled model to recover accurate surface information.

We constructed our models using 90 examples. We used 90 out-of-training-sample examples for surface reconstruction tests. We calculated the fractional height difference error $\|Ground\_truth - Recovered\_surface\|/Ground\_truth$ as an average over the 90 surfaces and over all points on the surfaces. For the purposes of analysis, we ordered the out-of-training-samples examples according to their distance from the mean intensity image $\bar{\mathbf{i}}$. We used the sum of the first ten values of $\mathbf{b}^i$ (to account for at least 50% of the variability), i.e., $\sum_{j=1}^{10} b_j^i$ as a similarity measure.

We commence by analyzing the shape recovery results obtained using the coupled

Figure 6.3: Surface recovery results for four cases using the height-and-intensity coupled model. The figure is divided in four panels. The first panel shows the input image together with frontal ground-truth re-illumination and profile view. The second panel presents the best-fit intensity recovery. The third panel presents results on surface recovery using the coupled model. We present frontal re-illumination and profile view followed by the intensity map of the height difference between ground-truth and recovered surface. The fourth column presents the intensity map of the height difference between ground-truth and the best-fit surface from height input (i.e. the surface in panel 1 was used as an input for the single model $\mathbf{M}^h$). Note that the height different plots have been normalized for presentation purposes (the brightest pixels represent a 10% difference).

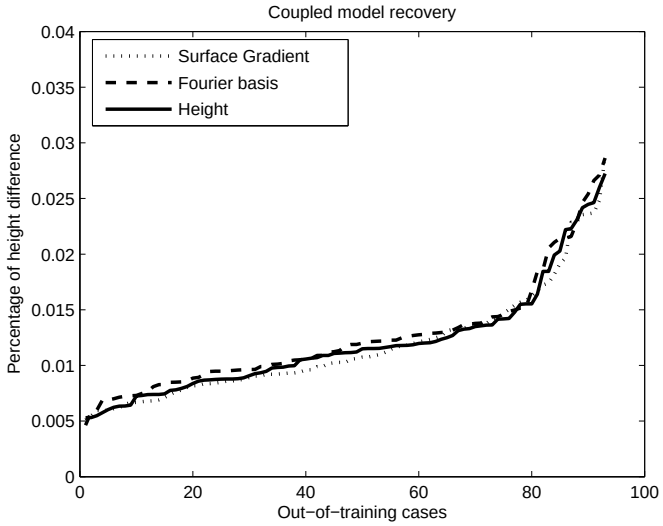

Figure 6.4: Plot of the fractional height difference between ground-truth and recovered surface when using the 90 out-of-training intensity images as input. The results were ordered in an ascending way for the purposes of comparison. The solid, dashed and dotted lines represent the height, Fourier basis and surface gradient coupled models.

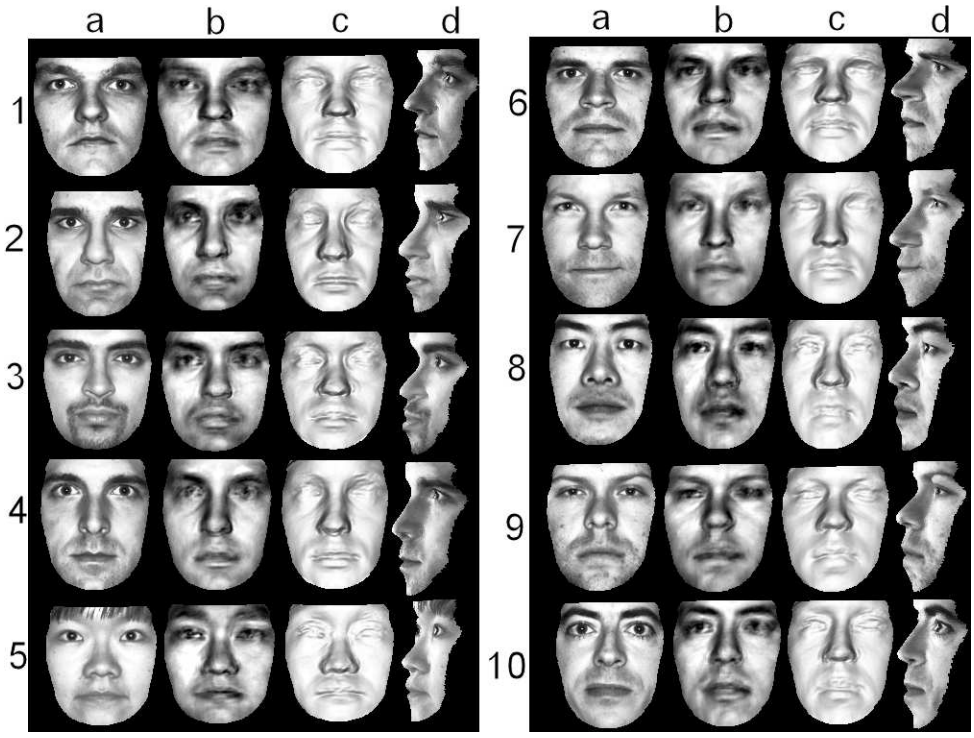

Figure 6.5: Experiments with frontal images of 10 individuals of the Yale B database. The figure is divided into two panels, each of which contains five of the ten subjects in the database. In the figure, the rows are labeled with numbers and present the different subjects. The input image, intensity best-fit recovery, frontal illumination of the recovered height and profile view with warped input image are shown column-wise, for each panel. We present the recovered surfaces using the intensity and height coupled model.

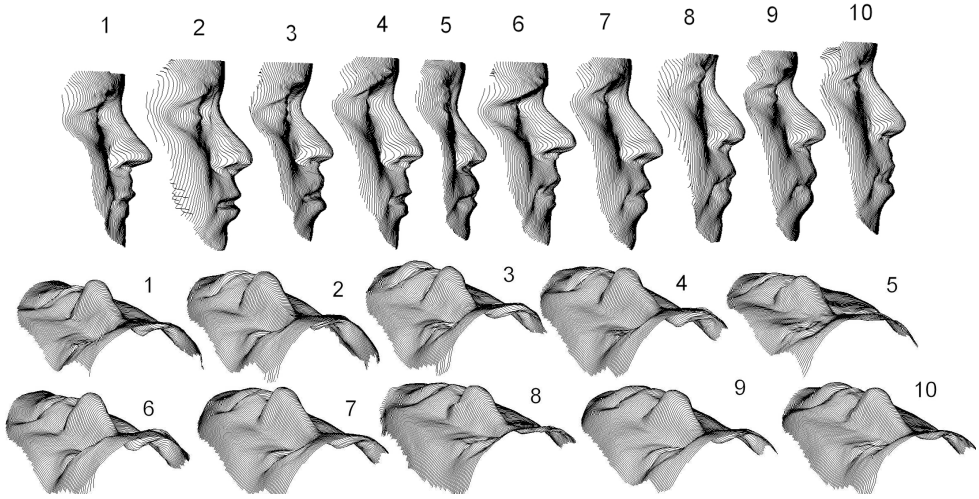

Figure 6.6: Two views of the recovered surfaces of the ten examples of the Yale database. These views have been rendered with no warped texture. The number accompanying the surfaces corresponds to the input images presented in figure 6.5.

Figure 6.7: Experiments with frontal images of individuals of the CMUPie database of faces. Different rows represent different subjects. The first three columns of the figure present input image, intensity best-fit recovery and frontal illumination of the recovered height. The rest of the columns show actual views of the individuals along with similar rendered novel views.

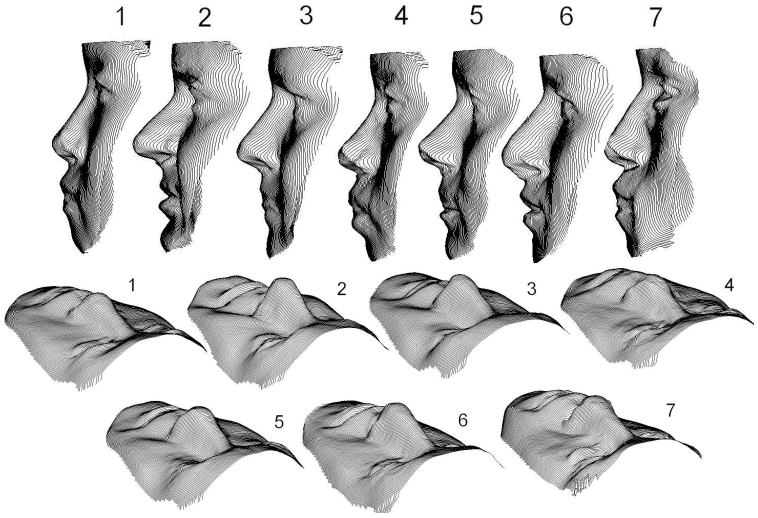

Figure 6.8: Two views of the recovered surfaces of the seven examples of the CMU Pie database. These views have been rendered with no warped texture. The number accompanying the surfaces corresponds to the input images presented in figure 6.7.

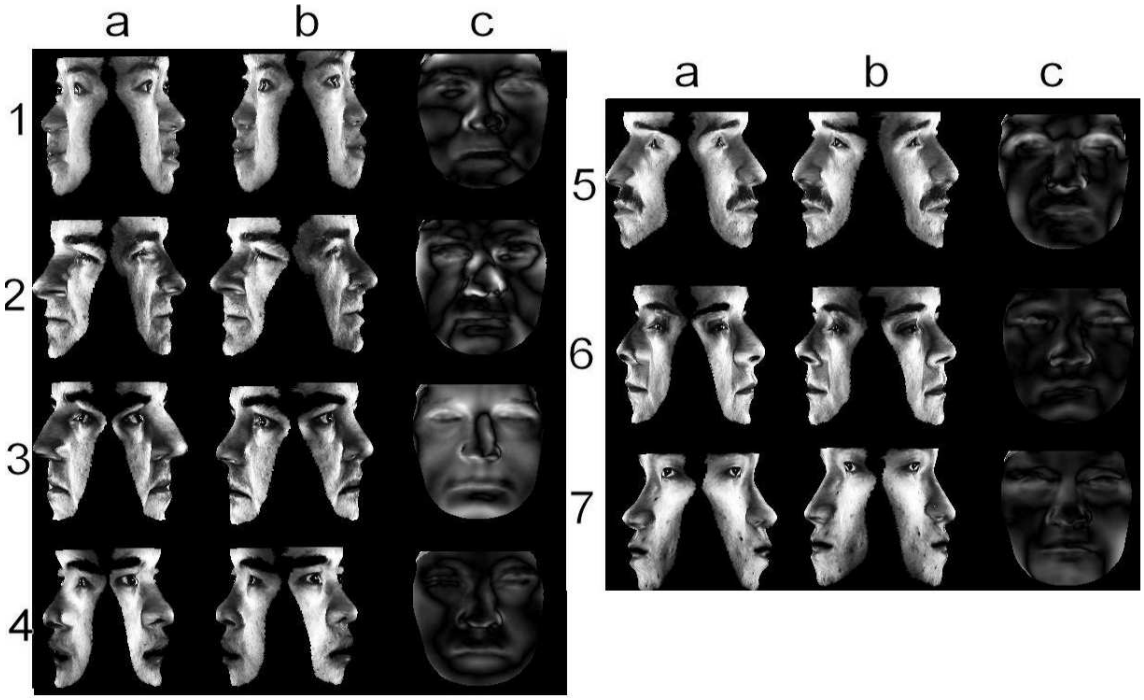

Figure 6.9: Novel view synthesis analysis. The figure is divided into two panels. For each panel, different rows represent different subjects of the CMUPie database. In the first two columns of both panels we show the close-to-profile novel views generated using the nearest in-training surface and the recovered surface using the coupled model. The third column presents the height difference error between these two surfaces as an intensity image.

model with the surface height representation. In Figure 6.3 we show surface recovery results for four examples. The different rows are for different individuals. As we go from top to bottom, the intensity images are increasingly different from the mean intensity model. In each row the results are organized into 4 groups. In the first group, the three columns show the input image together with its ground-truth re-illumination and profile view. The second group contains the recovered best-fit intensity image. The third group shows the results obtained using the coupled model. Here, the columns show the frontal re-illumination, a profile view and an intensity map of the fractional error between ground-truth and recovered surface height. The fourth group shows the error-map if the height data for the surface in group 1 was used as an input for the single model $\mathbf{M}^h$. As expected, the results shown in the third group (i.e. the coupled model) seem to match the best-fit intensity image in the second group rather than the original data in the first group. However, even for the input images that differ significantly from the mean intensity model, there is a good resemblance to the original data. On the other hand, the error plots in group 3 suggest a degree of similarity with the input image. This feature is not as strong in group 4. This may be a consequence of basing surface recovery on the best-fit parameters directly from an intensity image.

Next, we turn now our attention to the quantitative performance of each of the coupled models. We test how well the different representations perform using the 90 out-of-training-sample intensity images as input. We compare the recovered surfaces with the ground truth surface height data. In Figure 6.4 we plot the fractional height difference. The results were ordered according to ascending error for ease of visual comparison. The solid, dashed and dotted lines represent the height, Fourier basis and surface gradient coupled models. There seems to be no significant difference between the behavior of the models. The average surface recovery error was 1.194%, 1.249% and 1.168% for the height, Fourier basis and surface gradient coupled models. We also calculated the average

error from every out-of-training example to the mean height shape, which was 1.71%.

Finally, we turn out our attention to real world images. The first set of images used is drawn from the Yale B database (Georghiades et al., 2001) and is disjoint from the data used to train the statistical models. In the images, the faces are in frontal pose and were illuminated by a point light source situated approximately in the viewer direction. We aligned each image with the mean intensity shape so that the eyes, nose tip and mouth center were in the same position. The surface recovery results using the coupled model for the height representations are shown in Figure 6.5. The figure is divided into two panels, with five subjects on the left and five on the right. The result for the different subjects are shown in different rows. From left-to-right in each row we show the input image, the best-fit recovered intensity image, a frontal illumination of the recovered surface height and a profile view of texture mapping the input input image onto the recovered surface. There are a number of features to note from the figure. First, the reconstructed images agree well with the input. Second, the overall shape of the profile view is subjectively convincing.

In Figure 6.6, we show two views of the recovered surfaces for each of the ten Yale subjects. The number attached to the surface views correspond to that used in Figure 6.5. There are a number of features to note from these surface views. First, the most noticeable problems with the recovery of fine surface detail occur in the proximity of the mouth and the eyes. This is a consequence basing the surface recovery process on the best-fit parameters from the intensity model. The quality of the reconstruction will depend on the fidelity with which the intensity best-fit parameters are able to characterize an input image. Note that even when the best-fit recovered intensity image is of lower quality than those in Figure 6.3, the surface reconstructions from the best-fit intensity parameters are sufficiently good to render novel facial views.

In Figure 6.7 we present a second set of real world experiments for subjects from the

CMUPie database (Sim et al., 2003). In this set of experiments we explore the robustness to inaccurate alignment. We did not perform alignment to the mean intensity image. In the figure, the rows are labeled with numbers to identify the different subjects. The first three columns of the figure show the input image, the image reconstructed from the best-fit intensity model, and the frontal illumination of the recovered facial height. The remainder of the columns show actual views of the individuals (left) together with views of the reconstructed surface texture mapped with the input image. As consequence of not performing alignment operations, the surface recovery results seem to be slightly noisier than those obtained using the Yale database. This occurs especially around the mouth and nose regions of the face. This effect may be sufficient to distort the close-to-profile synthetic views. This is particularly noticeable for subjects 6 and 7, where noise in the proximity of the nose area generates incorrect nose profiles.

In Figure 6.8 we present two views of the recovered surfaces for the CMU Pie database subjects. These surfaces exhibit more variations in nose, chin and mouth shape than those recovered from the Yale database.

Finally, we calculated the height difference between the examples in the training database and the recovered surfaces for the CMUPie database. For each recovered surface, we located the surface from the training sample that minimized the height difference. We then texture-mapped the input image onto the surface to generated novel close-to-profile views. This experiment aims to demonstrate that new surface shapes are being generated by the best-fit intensity model parameters. In other words, we aim to demonstrate that the coupled model is not simply a table look-up procedure for sets of surface parameters. The results of this experiment are shown in Figure 6.9. The different rows are for different subjects in the CMUPie database. In the first column of both panels we show the close-to-profile views generated using the nearest in-training surface and the second column shows that obtained using the surface recovered using the coupled model. The

third column shows the height difference error between these two surfaces displayed as an intensity image. These error plots show that the main differences are located in the nose and eye regions.

6.3.1 Discussion

There are some additional observations to make concerning the experiments with the Yale and CMUPie databases. It has been previously noted that the accuracy with which the coupled models recover 3D shape depends on how well the best-fit intensity parameters are able to reconstruct the input intensity image. We can examine this effect by considering the quality of the best-fit intensity reconstructions in column (b) of Figures 6.5 and 6.7. If we compare these reconstructions with their corresponding input images in column (a) of the figures it is clear that the Yale database examples give better reconstructions than those in the CMUPie data-base. In particular, in the former case, there are less instabilities around the eyes, nose and mouth. This can be explained by a number of different factors. For example, unlike the Yale images, the CMUPie images were not aligned to the mean intensity image. Also, the light source direction in the Yale images is frontal, and all areas of the face are well illuminated. In the case of the CMUPie images, some areas of the face are in shadow. The quality of the reconstructions can be regarded as an effect of the linear nature of Equation 5.11. Clearly, as the best-fit intensity parameters are used to recover 3D shape through the coupled models, information encoded in these parameters is "inherited" to their 3D counterparts. This information may include errors due to misalignment and poor illumination settings. The resulting 3D shape improves in accuracy in accordance with the best-fit intensity recovery. Hence, for successful 3D shape recovery using coupled models, the input images must be well illuminated and aligned.

It has also been pointed-out (Atick et al., 1996) that the Cartesian representations used to construct the models give rise to more spatial variability than cylindrical coordinates.

However, the errors on the estimated surfaces are more strongly related to the illumination and alignment factors mentioned above than to the representations themselves. Suppose that to avoid problems with the Cartesian representations, the 3D shape models had been constructed under cylindrical representations. In this case, the correlation between the Cartesian intensity and the cylindrical 3D shape would probably not be sufficiently strong to approximate 3D shape from intensity images using coupled models. To exploit correlation in cylindrical coordinates, both intensity and 3D shape models would have to be expressed in the appropriate coordinate system. Moreover, the Cartesian input images would need to be transformed into cylindrical coordinates and that would introduce further alignment errors.

Finally, it is worth commenting in more detail on the differences between the method studied in this paper and Blanz and Vetter's morphable model (Blanz and Vetter, 1999), which has become a benchmark in the field of face shape recovery. The two approaches are compared in Table 6.1. At the conceptual level, they compare and contrast in the following ways. Firstly, Blanz and Vetter's method aims to generate photo-realistic renderings of faces which are compared to the input image through an exhaustive fitting process that minimizes a relatively complex error criterion. Pose and illumination information is also encoded in the model. This makes their framework robust to changes in illumination and pose. Also, separate models for 3D shape and intensity are created in cylindrical coordinates. The examples used to training the model have full correspondence with each other. The generation of photo-realistic views is greatly improved by the construction of independent models for different facial regions in the proximity of the eyes, nose and mouth. On the other hand, with our coupled approach we attempt to model the relationship between Cartesian intensity images and 3D shape by recovering the height surface of a face directly from the information encoded by the best-fit intensity parameters. This means that the coupled model only works on frontal or nearly frontal views and is sensi-

Comparison of Face Reconstruction Methods		
Constraints on input image	**Blanz and Vetter**	**Coupled statistical models**
Illumination:	Unknown, single point	Unknown, single point (frontal or nearly frontal)
Pose:	Arbitrary	Frontal
Minimum number of images:	1	1
Albedo:	Linear Statistical Constraint	Linear Statistical Constraint
Alignment:	Manually initialized with sparse feature points, fitting improves	Not required, but manual alignment improves results
Approximate shape recovery time:	4.5 minutes	2 - 5 seconds

Table 6.1: Comparison between the morphable model of Blanz and Vetter and the coupled statistical model.

tive to changes in illumination. However, the computations involved in the recovery of 3D shape are straightforward and this offers the possibility of approximating face shape in a computationally efficient way. The results presented in the experimental section suggest that there exists a correlation between the Cartesian intensity image and 3D shape that can be further studied to obtain improved reconstruction using coupled models. Generating independent models for different face regions and exploring the outcome of using different basis functions for PCA (such as 2D-PCA (Yang et al., 2004)) are possible routes to achieve improvements in future work.

6.4 Conclusions

We have explored a way for coupling intensity and 3D shape to construct statistical models of facial shape that can be used to recover shape from intensity images of faces. We have explored the performance of the statistical models over three representations: the surface height function, the surface gradient and a Fourier basis representation. For each

representation, the coupled model strongly links the best-fit coefficients for intensity and 3D shape data into a single statistical model. To recover the parameters of the coupled model, and hence reconstruct 3D shape, requires an optimization method whose objective function relies on the best-fit intensity parameters. The coupled models proved to be good enough to generate accurate surfaces from real world intensity imagery in an efficient way.

Chapter 7

Conclusions and Future Work

In this chapter, the main contributions of this thesis in the field of face shape recovery from a single image view are summarized. These are the development of a novel convexity constraint using local shape indicators, the construction of statistical models of faces based on Cartesian coordinates, and the definition of coupled statistical models of 3D shape and intensity. Suggestions for future work in the area will also be made in this chapter.

7.1 Summary of contributions

The objective of this thesis has been to propose and explore ways to solve the problem of face shape recovery from a single image view. To this end, we have addressed the problem of improving traditional SFS schemes (Chapters 3 and 4), and utilizing statistical models of faces (Chapters 5 and 6).

In Chapter 3 we combined two constraints that can be used in iterative SFS methods. We proposed an extension of a geometric SFS algorithm that treats the image irradiance equation as a hard constraint. Our approach overcomes the problem of high dependency on the image irradiance equation, which is a weakness of the geometric SFS approach.

This dependency can be relaxed through integrability constraints. Also, we have shown how the spurious peaks present in integrated height maps delivered by the original method can be stabilized if the integrability condition is introduced.

It is important to note that the inclusion of the integrability constraint has normally been used as a regularizing tool for needle maps in iterative SFS algorithms (Frankot and Chellappa, 1988). Instead, we use integrability to regularize the image irradiance, and the effect of this is to allow the opening angle of the reflectance cone to vary with each iteration. Although the algorithm improves the recovery of the shape of piece-wise smooth objects, it fails when applied to images of faces. Nonetheless, the results described in Chapter 3 support those in Chapter 5, where we show that integrability can improve the recovery of surfaces from single images of faces.

In Chapter 4, we have shown that problems in facial SFS can be overcome through the modification of the surface normal direction using convexity constraints. Here local shape indicators can be used to decide whether regions on a surface are concave or convex. The modification of the direction of the surface overcomes the problem of facial features implosion.

There are some differences between our proposed convexity constraint and other attempts at facial SFS. For example, Zhao and Chellappa's method failed when tested on varying albedo images (Zhao and Chellapa, 2000). The robustness of Prados and Faugeras can recover height regardless of facial pose (Prados et al., 2006), but input images must be such that the brightest point is encountered at the optical center of the camera. On the other hand, our method works on frontal pose images with varying albedo.

In Chapter 5 we provide a comparative study of the behavior of four Cartesian representations for constructing statistical models of faces. The representations studied are the surface height, the surface gradient, the surface normal azimuthal angle and finally a model based on Fourier domain basis functions. We analyze how well the models per-

formed if a parameter fitting procedure (subject to irradiance and integrability conditions) is used to calculate the models parameters. The results show that, unlike the models based on height data, those based on directional data work well with the fitting procedure. This feature can be explained by the smooth transitions between the models eigenmodes, i.e. the variability of the training data is not concentrated in just some few components (which is the case of the models based on height information). Moreover, we propose the use of complex numbers for building the azimuthal angle based models. This representation offers the advantages over models based on real numbers, i.e. degrees or radians.

In Chapter 6, we show how to learn the statistical relation between the image irradiance and three of the above mentioned Cartesian representations. To this end, we jointly capture variations in intensity and the surface shape representations using coupled statistical models. In order to build the coupled models, we perform PCA on sets of parameters describing the contents of the intensity images and the facial shape representations. The best fit coefficients from the intensity based model can be used to minimize an objective function to calculate the couple models coefficients. Surface shape can be then directly recovered from the coupled model coefficients. Our experiments demonstrated that the coupled models are able to generate good shape approximations from out-of-training-sample intensity images.

Interestingly, and in contrast to the results obtained in Chapter 5, there seemed to be no major difference in the performance of the coupled models based on height data and the coupled models based on directional data. Note that in both Chapters 5 and 6 we used fitting procedures that are guided by the image irradiance equation. However, the coupled model allows us to use the image irradiance in an encoded form represented by the intensity models best fit parameters. This means that errors are not propagated locally across the surface after each iteration. This is in contrast to the iterative method used in Chapter 5 where only the surface gradient and azimuth angle based models proved

to be robust to these errors. On the other hand, the errors contained in the intensity model best fit parameters are transferred to the estimated shape model parameters. This results in some instabilities in the recovered facial structure, mainly due to misalignment, uncontrolled lighting conditions and poor resolution of the input images.

It is important to note that the coupled models described in Chapter 6 were inspired by the active appearance model developed by Cootes, Edwards and Taylor (Cootes et al., 1998), which simultaneously models 2D shape and texture. Coupling intensity and surface shape in this fashion had not been explored in the literature before. The importance of coupled statistical models of intensity and surface shape lies on the direct usability of the intensity model best-fit parameters. We have shown that these parameters can assist in addressing the problem of expensive computational operations when fitting the model to intensity data.

7.2 Future work

There are several obvious shortcomings of the methods proposed in this thesis which require further research effort.

A problem with the convexity constraint of Chapter 4 is how to choose appropriate values for the thresholds. This drawback may be addressed by the minimization of some cost functional, for example, the sum of curvedness across the needle-map, to choose an optimal value for the threshold. Also, the convexity constraint has only been used as a corrective step for needle maps that have been previously calculated. Experiments are needed to analyze how this constraint can be introduced into an iterative SFS framework for the recovery of needle maps.

The Cartesian models described in Chapter 5 suffer from the drawback that they propagate errors when the images studied exhibit non-Lambertian reflectance. To overcome

this problem, we could explore ways of applying the constraint to the reflectance map of Horn (Horn, 1997). We could also apply methods for removing specular points (Ragheb and Hancock, 2003) in order to improve the quality of the input images.

In Chapter 6, although quantitatively good results are achieved by coupling intensity and 3D surface shape, the main drawback of the method is an overreliance on the best fit parameters of the intensity-based model. Even though the coupled models have been tested over both aligned and non-aligned images, it is necessary to conduct a more thorough analysis on the outcome of different alignment procedures applied on the input intensity images. Also, experiments on the sensitivity of the method to effects such as illumination changes, image resolution and facial expression can be used to understand the quality of the output reconstructed face. These experiments would also help to determine whether an input image needs pre-processing operations before applying the coupled models.

Another important line of investigation is to perform experiments on the robustness of both the coupled and the single Cartesian models built on alternative basis functions such as 2D-PCA (Yang et al., 2004). These experiments would help to understand the behavior of the models when they are constructed on inputs other than intensity eigenfaces. Moreover, it would be interesting to investigate the effect of coupled models of intensity and 3D surface shape on separate facial features. In other words, to generate coupled models independently for eyes, mouth and nose, as done previously by (Nandy and Ben-Arie, 1999) and (Blanz and Vetter, 2003).

Furthermore, it would be interesting to study the relationship between image intensity and its Lambertian component, i.e. the albedo-free frontal reillumination of a 3D face. Using coupled statistical models might offer a way of recovering facial albedo from non-Lambertian faces. This seems an obvious approach since the coupled components are both based on 2D images. This is similar to the work of (Tang and Wang., 2004), who

showed how to couple images of faces and sketches. Knowledge of the facial albedo of an image would result in improved usability of the Cartesian representations using the data-driven procedure described in Chapter 5.

Finally, it would be worthwhile to investigate how the proposed methods in this thesis can be used for face recognition purposes. In practice, we could perform experiments where changes in pose and illumination are signified in order to take advantage of the recovered facial shapes.

Bibliography

Agrawal, A., Raskar, R., and Chellappa, R. (2005). An algebraic approach to surface reconstruction from gradient fields. In *IEEE International Conference in Computer Vision*.

Atick, J., Griffin, P., and Redlich, N. (1996). Statistical approach to shape from shading: Reconstruction of three-dimensional face surfaces from single two-dimensional images. *Neural Computation*, 8:1321–1340.

Barrow, H. and Tanenbaum, H. K. (1993). Retrospective on interpreting line drawings as three-dimensional surfaces. *Artificial Intelligence*, 59:71–80.

Batlle, J., Mouaddib, E., and Salvi, J. (1998). Recent progress in coded structured light as a technique to solve the correspondence problem: a survey. *Patter Recognition*, 31:963–982.

Beckmann, P. and Spizzichino, A. (1963). *The Scattering of Electromagnetic Waves from Rough Surfaces*. Pergamon Press.

Besl, P. (1989). Active optical range imaging sensors. In *Advances in Machine Vision*, pages 1–63.

Besl, P. and Jain, R. C. (1986). Invariant surface characteristics for 3d object recognition in range images. *Comput. Vision Graph. Image Process.*, 33(1):33–80.

Bichsel, M. and Pentland, A. (1992). A simple algorithm for shape from shading. In *IEEE Proceedings of Computer Vision and Pattern Recognition*, pages 459–465.

Blais, F. (2003). A review of 20 years of range sensor development. In *Proc. Videometrics VII, SPIE*, pages 62–76.

Blanz, V. and Vetter, T. (1999). A morphable model for the synthesis of 3d faces. In *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 187–194, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

Blanz, V. and Vetter, T. (2003). Face recognition based on fitting a 3d morphable model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(9):1063–1074.

Bors, A., Hancock, E., and Wilson, R. (2003). Terrain analysis using radar shape-from-shading. *IEEE Trans, on Pattern Analysis and Machine Intelligence*, 25(5).

Brooks, M. and Horn, B. (1985). Shape and source from shading. In *Proceedings of International Joint Conference in Artificial Intelligence*, pages 932–936.

Coleman, J. and Jain, R. (1982). Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. In *CGIP*, volume 18, pages 309–328.

Cootes, T., Edwards, G., and Taylor, C. (1998). Active appearance models. In *Proc. European Conference in Computer Vision*, pages 484–498.

Dovgard, R. and Basri, R. (2004). Statistical symmetric shape from shading for 3d structure recovery of faces. In *Proc. European Conference on Computer Vision*, pages 99–113.

Dupuis, P. and Oliensis, J. (1994). An optimal control formulation and related numerical

methods for a problem in shape reconstruction. In *The Annals of Applied Probability*, volume 4, pages 287–346.

Erens, R., Kappers, A., and Koenderink, J. (1993a). Estimating local shape from shading in the presence of global shading. *Perception and Psychophysics*, 54(3):334–342.

Erens, R., Kappers, A., and Koenderink, J. (1993b). Perception of local shape from shading. *Perception and Psychophysics*, 54(2):145–156.

Forsythe, D. and Ponce, J. (2001). *Computer Vision: a Modern Approach*. Prentice-Hall.

Frankot, R. and Chellappa, R. (1988). A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10:438–451.

Georghiades, A., Belhumeur, D., and Kriegman, D. (2001). From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 634–660.

Gregory, R. (1997). Knowledge in perception and illusion. *Transactions of the Royal Society of London. B 352*, pages 1121–1128.

Healey, G. and Jain, R. (1984). Depth recovery from surface normals. In *ICPR '84, pages = 894–896*.

Horn, B. (1970). *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View, PhD Thesis, MIT.*

Horn, B. (1986). *Robot Vision.* MIT Press, Cambridge, MA.

Horn, B. (1989). Height and gradient from shading. *International Journal of Computer Vision*, 5(1):37–75.

Horn, B. (1997). Understanding image intensities. *Artificial Intelligence*, 8:201–231.

Horn, B. and Brooks, M. (1989). *Shape from Shading*. MIT Press, Cambridge, MA.

Horn, B. and Brooks M, J. (1986). The variational approach to shape-from-shading. *Computer Vision and Image Processing*, 33:174–208.

Ikeuchi, K. and Horn, B. (1981). Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, 17:141–184.

Jarvis, R. (1983). A perspective on range finding techiques fro computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(2):122–139.

Kirby, M. and Sirovich, L. (1990). Appliction of the karhumen-loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108.

Klette, R. and Schluens, K. (1996). Height Data from Gradient Maps. In *Proc. SPIE Vol. 2908, p. 204-215, Machine Vision Applications, Architectures, and Systems Integration V, Susan S. Solomon; Bruce G. Batchelor; Frederick M. Waltz; Eds.*, pages 204–215.

Koenderink, J. and Van Doorn, A. (1992a). Surface perception in pictures. *Perception and Psychophysics*, 52(5):487–496.

Koenderink, J. and Van Doorn, A. (1992b). Surface shape and curvature scales. *Image en Vision Computing*, 10:557–565.

Koenderink, J., Van Doorn, A., Christou, C., and Lappin, J. (1996). Perturbation study of shading in pictures. *Perception*, 25(9):1009–1026.

Lee, C. and Rosenfeld, A. (1985). Improved methods of estimating shape from shading using light source coordinate system. *Artificial Intelligence*, 26:125–143.

Mardia, K. V. (1972). *Statistics of Directional Data*. Academic Press London and New York.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of the Visual information*. Freeman.

Marr, D. and Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three dimensional shapes. In *Proc. Royal Society of London, B*, volume 200.

Mingolla, E. and T., T. J. (1986). Perception of solid shape from shading. *Biological Cybernetics*, 53:137–151.

Moses, Y., Adini, Y., and Ullman, S. (1994). Face recognition: the problem of compensating for changes in illumination direction. In *Proc. European Conference on Computer Vision*, pages 286–296.

Nandy, D. and Ben-Arie, J. (1999). Shape from recognition and learning: Recovery of 3-d face shapes. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2.

Nene, S. A., Nayar, S. K., and Murase, H. (1996). Columbia object image library: Coil-100. *Technical Report CUCS-006-96, Columbia University*.

Pentland, A. (1982). The visual inference of shape: Computation from local features. *PhD Dissertation, Dep. Pshychology, MIT, Cambridge, MA.*

Prados, E., Camilli, F., and Faugeras, O. (2006). A unifying and rigorous shape from shading method adapted to realistic data and applications. *Journal of Mathematical Imaging and Vision*.

Ragheb, H. and Hancock, E. (2003). A probabilistic framework for specular shape-from-shading. *Pattern Recognition*, 36:407–427.

Ramachandran, V. (1988). Perceiving shape from shading. *Scientific American*, 129:76–83.

Robles-Kelly, A. and Hancock, E. R. (2004). A graph-spectral approach to shape-from-shading. *IEEE Transactions on Image Processing*, 13(7):912–926.

Rouy, E. and Tourin, A. (1992). A viscosity solutions approach to shape from shading. *SIAM Journal of Numerical Analysis*, 29(3):867–884.

Samaras, D. and Metaxas, D. (2003). Incorporating illumination constraints in deformable models for shape and light direction estimation. *IEEE Trans. PAMI*, 25(2):247–264.

Scharstein, D. and Szeliski, R. (2003). High-accuracy stereo depth maps using structured light. In *IEEE Proceedings on International Conference on Computer Vision and Pattern Recognition*.

Schlick, C. (1994). A survey of shading and reflectance models. *Computer Graphics Forum*, 13(2):121–131.

Shimshoni, I., Mose, Y., and Lindenbaum, M. (2003). Shape reconstruction of 3d bilaterally symmetryc surfaces. *The International Journal of Computer Vision*, 2:1–15.

Sim, T., S., B., and Bsat, M. (2003). The cmu pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618.

Smith, W. and Hancock, E. (2005a). Face recognition using a surface normal model. In *Proc. International Conference on Image Analysis and Processing*, pages 423–430.

Smith, W. and Hancock, E. R. (2005b). Recovering facial shape and albedo using a statistical model of surface normal direction. In *Proc. Tenth IEEE International Conference on Computer Vision (ICCV'05)*, pages 588–595.

Smith, W., Robles-Kelly, A., and Hancock, E. R. (2004). Reflectance correction for perspiring faces. In *Proc. ICIP*, pages 1389–1392.

Starks, M. (1995). Stereoscopic imaging technology: Review of patents and literature. *International Journal of Virtual Reality*, 1(2):2–25.

Stevens, K. and Brookes, A. (1987). Probing depth in monocular images. *Biological Cybernetics*, 56:355–366.

Sutherland, N. (1979). The representation of three-dimensional objects. *Nature*, (278):395–398.

Tang, X. and Wang., X. (2004). Face sketch recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):50–57.

Torrance, K. and Sparrow, E. (1967). Theory for off-specular reflection from roughened surfaces. *Journal of the Optican Society of America*, 57(9):1105–1114.

Trucco, E. and Verri, A. (1998). *Introductory Techniques for 3-D Computer Vision*. Prentice-Hall, San Francisco.

Turk, M. and Pentland, A. (1991). Face recognition using eigenfaces. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–591.

Wei, T. and Klette, R. (2002). Height from gradient using surface curvature and area constraints. In *Proc. ICVGIP*.

Worthington, P. L. and Hancock, E. R. (1998). Needle map recovery using robust regularizers. *Image and Vision Computing*, 17(8):545–559.

Worthington, P. L. and Hancock, E. R. (1999). New constraints on data-closeness and needle map consistency for shape-from-shading. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(12):1250–1267.

Worthington, P. L. and Hancock, E. R. (2001). Object recognition using shape-from-shading. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(5):535–542.

Wu, Z. and Li, L. (1988). A line integration based method for depth recovery from surface normals. *CVGIP*, 43(1):53–66.

Yang, J., Zhang, D., Frangi, A., and Yang, J. (2004). Two-dimensional pca: a new approach to appearance-based face representation and recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(1):131–137.

Zhang, R., Tsai, P., Cryer, J., and Shah, M. (1999). Shape from shading: A survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(8):690–706.

Zhao, W. and Chellapa, R. (2000). Illumination-insensitive face recognition using symmetric shape-from-shading. In *Proc. Conference on Computer Vision and Pattern Recognition*, pages 286–293.

Zhao, W. and Chellapa, R. (2001). Symmetric shape-from-shading using self-ratio image. *International Journal of Computer Vision*, 45(1):55–75.

Zheng, Q. and Chellapa, R. (1991). Estimation of illumination direction, albedo and shape from shading. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(7):680–702.