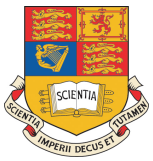


Imperial College London

MEng Individual Report

A Study Of Statistical Methods For Facial Shape-from-shading

Patrick Snape
pts08@ic.ac.uk

Supervisor: S. Zafeiriou
Second Marker: M. Pantic

June 18, 2012

Abstract

Shape-from-shading addresses the problem of recovering a 3D surface from a single 2D image. In this project, we concentrate on reconstructing faces from a single 2D image by incorporating a statistical model in to a Shape-from-shading algorithm. We build on the work presented by W.A.P Smith and investigate the reconstructive power of computing principal components from directed data. We also present a novel reconstruction technique that recovers a set of texture and shape coefficients in order to reconstruct an input image.

The novel reconstruction technique presented produces results that are similar in visual appearance to W.A.P Smith's reconstructions. We also found that projecting the normals in to a Euclidean space improved the ability of the statistical model to recover facial shape. Using unprojected normals, however, still produced acceptable results.

Acknowledgements

First and foremost, I would like to thank Stefanos Zafeiriou for supervising my project. He was always very helpful and willing to put in some long evenings to try and help me understand the intricacies of Matlab. My thanks also go out to Joan Alabort Medina and Ioannis Marras for putting in the effort to help me.

Finally I would like to thank my family for supporting me through university. Particularly my Dad for always being there to bounce ideas off of and my Mum for her unwavering belief in my ability.

Contents

Chapter 1

Introduction

Modelling human faces has been of interest to researchers in computer graphics for many years[18]. Since then increased computing power has allowed researchers to create more accurate facial models. These models are created in a number of ways and vary in their degree of accuracy. Powerful hardware such as laser scanners[14] have been used to generate high quality 3D models of faces, yet are very expensive to operate. Recent interest has largely been in the area of 3D reconstruction from 2D images. Since 2D images are inexpensive to produce they represent a cost-effective way of generating accurate 3D models.

Morphable models,[4] for example, provide a robust method of modelling from a single image. Morphable models use a database of 3D laser-scanned faces. They attempt to express novel faces as a linear combination of a statistical model form from the laser scans. Recovering facial shape using this technique produces robust performance under widely varying illumination conditions and poses. Shape-from-shading approaches, such as those proposed by [12] can also produce good looking results using just a single image and reference face shape. They also avoid local minima as image parameters are computed directly from the input image and not by solving a minimization problem.

This project investigates the feasability of improving Shape-from-shading by introducing a statistical model. Shape-from-shading looks to recover the surface normals from a single reference image. Surface normals, however, represent directional data and thus do not lie within a Euclidean space. Standard statistical techniques such as Principal Component Analysis (PCA) require a Euclidean space in order to be calculated. We investigate existing techniques to perform statistical analysis on non-Euclidean data, concentrating on the work outlined by William Smith [22, 23]. Our contribution includes a new method of reconstructing normals through an optimization loop that leverages the linear transformation model used by Principal Component Analysis. It looks to construct a set of normal and

texture coefficients that represent a close match to the input image.

1.1 Motivation and Objectives

Reconstructing 3D models of faces from 2D images has many applications in areas such as face recognition[11, 29] and facial expression analysis[16, 17]. One of the major challenges in face recognition is the occurence of a large variation in lighting and pose. 3D models have the distinct advantage that they can be re-rendered under new illumination conditions and in novel poses. Surface normals, in particular, represent a very descriptive representation of a face, as they are invariant under changing illumination. In practise it is extremely challenging to produce consistent results given unconstrained input data. For this reason it is interesting to investigate new techniques for recovering accurate surface normals from images.

1.2 Contributions

This project has provided the following notable contributions:

- A set of Matlab script to perform both the Azimuthal Equidistant Projection and Principal Geodesic Analysis on a face data set. These techniques are distinct from the area of facial reconstruction and could in fact be used on any directional data.
- An implementation of William Smith's reconstruction algorithms described in both [22] and [23]. These provide accurate reconstructions of faces from a single front-facing reference image under a directed light.
- A new reconstruction method based on recovering a set of shape coefficients to reconstruct the image from a statistical model. This reconstruction method, whilst not robust, yields acceptable results from a range of images.
- A comparison of the robustness of using surface normals in a statistical model. This includes building statistical models directly on the surface normals, data projected using the Azimuthal Equdistant Projection and performing Principal Geodesic Analysis.

Chapter 2

Background

In this Chapter we explain the state-of-the-art techniques that currently exist for reconstructing faces from 2D images. In particular we concentrate on analysis-by-synthesis and shape-from-shading. Analysis-by-synthesis is used within the context of morphable models[4] and shape-from-shading is used by a number of different reconstruction techniques[22, 12]. We also explain statistical techniques that are used within the context of statistical face reconstruction, namely PCA.

2.1 Analysis-by-synthesis

Analysis-by-synthesis refers to a technique that was first described within the context of signal processing[3]. The technique aims to synthesize signals from a known generator and compare them with signals that are to be analyzed. It then computes the error between the synthesized and analyzed signals. Multiple new signals are generated until one is produced that minimizes the measure of error.

2.1.1 Morphable 3D Face Model

Analysis-by-synthesis has been extended in to the realm of 3D face reconstruction by the seminal work of Blanz and Vetter[4]. They propose the morphable model to synthesize both novel faces as well as to fit their model to existing 2D images. The morphable model is constructed from a set of laser scans from the heads of 200 young adults (100 male, 100 female). In order for the scans to be used as a single dataset they first had to be brought in to correspondence. Point-to-point correspondence is necessary to allow the information from each scan to be parameterized within a common space. Once in correspondence the scans have the same number of vertices and each vertex can be matched to a corresponding vertex in all the other scans in the dataset. This correspondence is calculated using an Optic Flow algorithm

and a reference face (details can be found in [4, 24]).

Once all faces are in complete correspondence they can be parameterized by a common discrete reference space called (u, v) space. In order to provide continuity within this discrete space all vertices are treated as a triangle list and barycentric coordinates are used to interpolate within the triangles.

To represent the face within the morphable model both the shape and texture information are needed. The shape is normally represented as a shape-vector and the texture information as a texture-vector:

$$\boldsymbol{S} = (X_1, Y_1, Z_1, X_2, \ldots, Y_N, Z_N)^T \qquad \boldsymbol{T} = (R_1, G_1, B_1, R_2, \ldots, G_N, B_N)^T$$

where N is the number of vertices in the face scan. It is important to note that the number of vertices and the number of RGB values match exactly and each RGB corresponds to the colour of the texture of the vertex in that position.

Each face within the dataset can be represented by it's shape-vector S_i and texture-vector T_i. Since all faces are in correspondence, novel faces can be expressed as a linear combination of the shape and texture vectors:

$$\boldsymbol{S} = \sum_{i=1}^{m} \alpha_i \cdot S_i \qquad \boldsymbol{T} = \sum_{i=1}^{m} \beta_i \cdot T_i$$

$$\sum_{i=1}^{m} \alpha_i = \sum_{i=1}^{m} \beta_i = 1$$

where m is the total number of face scans and α_i and β_i are the shape texture coefficients of the ith scan. Also, in order to enforce plausible novel faces, the coefficients must sum to 1. Novel faces are generated by varying the parameters of α and β for each scan in the linear combination.

Principal Component Analysis (PCA) is then performed in order to reduce the dimensionality of the model. A more formal description of PCA can be found in Section 2.1.2.

2.1.2 Principal Component Analysis

Principal Component Analysis (PCA) can provide us with a more intuitive way of representing our data. It allows us to build a statistical model of faces whereby a large range of all possible faces can be described. Formally PCA is used to provide a basis to a set of data where the eigenvector that corresponds to the largest eigenvalue highlights the direction of highest variance. This eigenvector corresponds to a line that passes through the mean and gives the least squares distances of the points in the dataset from the line. The set of points in the dataset must lie within a Euclidean space. A Euclidean space is required to allow the calculation of the least squares distances.

This can be useful in reducing the dimensionality of the dataset. Removing the lowest principal components amounts to losing the least amount of information possible. PCA is performed seperately on both the shape and texture vectors - but the technique is identical.

In order to calculate the principal components of the shape vector the average shape vector must first be computed:

$$\bar{S} = \frac{1}{m}\sum_{i=1}^{m} S_i$$

where m is the number of face scans. This is then subtracted from each of the face scans to give us the shape differences:

$$\Delta S = S_i - \bar{S}$$

The covariance matrix C is computed as:

$$C = \frac{1}{m}\Delta S \Delta S^T$$

C can be expressed in terms of it's eigenvalues and corresponding eigenvectors (calculated using SVD[7]) in the form:

$$C = \frac{1}{m} U \Lambda^2 U^T$$

where the columns of U are the eigenvectors and the corresponding eigenvalues are given by the diagonal elements of $\frac{\Lambda^2}{m}$.

Due to the fact that the number of face scans is almost always going to be less than the number of vertices per scan, it can be assumed that $m \ll N$. This means that there are only $m-1$ non-zero eigenvalues and so only the first $m-1$ columns of U contain meaningful eigenvectors.

Given the above formulation of the principal components, the morphable model can now be described as a linear combination of the average shape vector and the weighted principal components:

$$S = \bar{S} + \sum_{i=1}^{m-1} \alpha_i \cdot s_i \qquad\qquad T = \bar{T} + \sum_{i=1}^{m-1} \beta_i \cdot t_i \tag{2.1}$$

where s_i and t_i are the principal components for the shape and texture vectors.

The PCA representation also holds the advantage that the probability of any coefficient (α or β) can be modelled by a multivariate normal distribution:

$$\begin{aligned} p(\vec{\alpha}) &\sim \exp\left[-\frac{1}{2}\sum_{i=1}^{m-1}(\alpha_i/\sigma_{i,S})^2\right] \\ p(\vec{\beta}) &\sim \exp\left[-\frac{1}{2}\sum_{i=1}^{m-1}(\beta_i/\sigma_{i,T})^2\right] \end{aligned} \tag{2.2}$$

where $\sigma_{i,S}$ and $\sigma_{i,T}$ are the eigenvalues from the shape and texture covariance matrices. $\vec{\alpha}$ and $\vec{\beta}$ represent vectors of all the coefficients eg. $\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_m)^T$ and $\vec{\beta} = (\beta_1, \beta_2, \ldots, \beta_m)^T$.

2.1.3 Matching a Morphable Model to an image

The main construct involved with matching a morphable model to an image is the analysis-by-synthesis loop. The loop attempts to optimize the α and β coefficients as well as a set of rendering parameters, $\vec{\rho}$. In every iteration a texture mapped 3D face is created and rendered from the current model parameters. It is then updated according to the error calculated in terms of the ℓ_2 norm.

$$E_I = \sum_{x,y} \|\boldsymbol{I}_{input}(x,y) - \boldsymbol{I}_{model}(x,y)\|^2$$

where $\boldsymbol{I}_{model}$ represents the RGB components rendered using perspective projection and phong illumination during the rendering phase.

In order to impose the synthesis of face-like surfaces the shape and texture vectors are limited to the vector space of the database. Solutions are restricted to this space by a tradeoff between reducing the Euclidean distance and prior probability. Utilizing the probabilities of the coefficients of $\vec{\alpha}$ and $\vec{\beta}$ shown in Equation 2.2 and using an ad-hoc estimate for $\vec{\rho}$, Bayes decision theory can be utilized to solve the problem. $\vec{\alpha}$, $\vec{\beta}$ and $\vec{\rho}$ can be used to fully reconstruct I_{model}, however, I_{input} may suffer from noise. Therefore, likelihood of I_{input} is modelled as:

$$p(I_{input}|\vec{\alpha}, \vec{\beta}, \vec{\rho}) \sim \exp\left[-\frac{1}{\sigma_{S,N}^2} \cdot E_I\right]$$

where $\sigma_{S,N}$ represents the standard deviation of the noise present in I_{input} modelled as a Gaussian distribution. Maximum posterior probability is achieved by minimizing the cost function[4]:

$$E = \frac{1}{\sigma_N^2} E_I + \sum_{j=1}^{m-1} \frac{\alpha_j^2}{\sigma_{S,j}^2} + \sum_{j=1}^{m-1} \frac{\beta_j^2}{\sigma_{T,j}^2} + \sum_j \frac{(\rho_j - \bar{\rho_j})^2}{\sigma_{\rho,j}^2} \tag{2.3}$$

A stochastic gradient descent algorithm is then performed based on a random selection of surface points. Parameters are updated using analytical derivates of the cost function (2.3):

$$\alpha_j \mapsto \alpha_j - \lambda_j \cdot \frac{\partial E}{\partial \alpha_j} \tag{2.4}$$

with suitable factors λ_j.

2.2 Shape-From-Shading

Shape-from-shading is the problem of using shading information from a single image in order to recover shape information. In a greyscale image the varying intensity at each pixel is due to four properties within the image; the shape of the surface, the illumination, the reflectance properties and the projection. Therefore, shape information can be recovered through an inversion of the previous factors. This is generally described as generating estimates from the Irradiance Equation for a Lambertian surface. Lambert's cosine law states that the radiance observed from a surface is proportional to the cosine of the angle between the viewing direction and the surface normal. A surface that obeys this law is said to be Lambertian and has the same radiance when observed from any viewing angle.

Given the intensity of a pixel, I_p, the local surface normal, $n_p = (p, q, -1)$, the albedo ρ, and a vector, s, describing the direction of a single distant light source, then according to Lambert's cosine law:

$$I_p = \rho(n_p \cdot s) \tag{2.5}$$

In general, shape-from-shading is ill-posed and calculating a solution requires some form of prior knowledge about the lighting conditions, albedo and boundary conditions (maximum depth of the surface). Shape-from-Shading is often solved using a constrained optimization framework [9, 8, 10] that looks to minimize a cost function using calculus of variations.

2.3 Photometric Stereo

Photometric Stereo is a technique used to recover a three-dimensional surface from a two-dimensional image. It was first proposed by Woodham in 1980 [25]. The technique assumes specific reflectivity properties about the surface and uses these assumptions to recover illumination invariant properties of the surface. Specifically it assumes Lambertian reflectance and looks to recover the surface normals and albedo. As shown in Equation 2.5 there are 3 unknowns, p, q and ρ. Photometric stereo looks to overcome the issue of recovering these parameters by producing a system of equations. Taking 3 or more images from the same position, but under different lighting conditions, will yield a system of equations from which the unknowns can be recovered. Whilst, under certain conditions, the unknowns can be recovered using just two images, 3 or more images ensures that a solution can be arrived at. However, specularities and areas of poor illumination can affect the accuracy of the reconstruction.

Adding extra images, usually a fourth image, can allow the reconstruction to identify areas of specularity and treat them as deviations from Lambert's law. In this project we use the 4-source technique proposed by Barsky

and Petrou [2]. Their proposal is to use a fourth image in order to identify areas of specularity and shadow and then to use the shading information from the colour images to aid in reconstruction. Normally the colour information in an image is redundant for photometric stereo as the photometric equations in the RGB channels are linearly dependant. However, the colour can be leveraged to perform Shape-from-Colour where only a single colour image under complex illumination is used. In the case of Barsky and Petrou they use the 4-source technique proposed by Coleman and Jain [15] and then they use a local spectral difference in order to identify highlights.

2.4 Recovering A Surface From Normals

In order to be able to reconstruct a shape from the field of normals there must exist some form of projection from the normals to a set of coordinates in 3D space. The optimum way of performing this reconstruction is to build a depth map from the normals. A depth map views the image as a grid whereby each data point is the height above the (x, y) plane. In the traditional shape-from-shading problem the normals are computed as opposed to directly recovering the depth map. This has been shown by Horn and Brooks[10] to be a simpler reconstruction than attempting to recover the depth map directly. Therefore, it is possible to attempt to recover the surface information by first recovering the components of the normal vector $(p, q, -1)$. In this case p and q represent first-order non-linear partial differential equations in x and y

$$p = \frac{\partial z}{\partial x} \qquad q = \frac{\partial z}{\partial y}$$

In order to provide consistency on both p and q a smoothness constraint is enforced on the surface constructed by integrating p and q. This surface is constrained by the following assumption:

$$z_{xy}(x, y) = z_{yx}(x, y)$$

which states that the surface height at any point is independant from the path of integration.

Frankot and Chellappa propose a technique for enforcing this integration constraint by using Fourier transforms[6]. In this project a single pass of the reconstruction is used as opposed to the iterative technique proposed in [6].

The Fourier transform of a grid from $\frac{-\pi}{2}$ to $\frac{\pi}{2}$ is taken and then centered to yield the frequency domain representation of the depth map. An inverse Fast-Fourier transform is performed and the data shifted to give the reconstructed depth map. Mathematically this amounts to solving the following minimization problem

$$\bar{C}(\omega) = \frac{-j\omega_x \hat{C}_x(\omega) - j\omega_y \hat{C}_y(\omega)}{\omega_x^2 + \omega_y^2}$$

where $\bar{C}(\omega)$ represents the vector of coefficients of the Fourier series expansion of the depth map. These coefficients are used in the reconstruction that follows from the following representation of the surface

$$\bar{z}(x, y) = \sum \bar{C}(\omega) exp\{j\omega \cdot (x, y)\}$$

where ω is a two-dimensional index in the range $(2\pi N, 2\pi N)$ for an $N \times N$ image.

2.5 Other Existing Work

The original shape-from-shading algorithm was proposed by Horn [8]. It has been investigated more recently by the likes of Zhang et al. [28] and is now regarded as a mature field.

More recently there have been a number of approaches to reconstructing faces from 2D images using shape-from-shading. Kemelmacher and Basri have proposed a single image[12] reconstruction using just a reference model that generates a good approximation of facial shape. They assume that the input image is roughly aligned to their reference model and approach the reconstruction by solving for lighting, depth, and albedo separately. They represent reflectance using spherical harmonics and use the reference model to help overcome boundary conditions and infer albedo.

Kemelmacher and Seitz also propose a novel method involving using many input images from a large multi-pose dataset [13]. They concentrate on pose normalization and look to build on photometric stereo techniques by leveraging the change in illumuination over the large input set.

There are also a number of reconstruction techniques that involve leveraging a large set of 3D laser scanned faces. Most notably morphable models [4] uses this technique. Morphable models have also recently been extended through the works of Zhang et al. [27] to include more robust lighting models.

Chapter 3

Project

In this chapter we discuss the work undertaken within the project. The project consists of two major sections. The first is an investigation in to the techniques available to project non-Euclidean data so that statistical analysis can be performed on it. Two techiques have been covered; the Azimuthal Equidistant Projection and Principal Geodesic Analysis. We also provide an implementation of statistical shape-from-shading as proposed by William Smith[22, 23].

The second part of the project is a novel reconstruction technique. This new algorithm looks to recover a set of coefficients that can be used in conjunction with a statistical model in order to reconstruct an input image. This method of reconstruction is not robust but yields similar results to Smith's reconstructions on front-facing images.

3.1 PCA on directional data

A normal is defined as the unit vector that is perpendicular to a particular point on a surface. Therefore, by definition, a set of surface normals represents a directional dataset. We can can view normals as points residing on the surface of a unit sphere. This projection on to a unit sphere is often referred to as a needlemap.

The linear combination of unit vectors is itself not a unit vector and thus cannot be treated in a linear fashion. For this reason it is not possible to perform Principal Component Anlysis on a set of surface normals. The major issue is that of determining an accurate distance between two points within the data set. Straight line distance is not an accurate measure as the real distance covered is on the surface of the sphere; the so called geodesic distance. Therefore, in order to perform statistical anlysis on the surface normals, we must project them in to an appropriate Euclidean space. Two techniques have been proposed by Smith[22, 23] in order to overcome this issue; the azimuthal equidistant project and principal geodesic an.

3.1.1 Azimuthal Equidistant Projection

The Azimuthal Equidistant Projection (AEP) is a commonly used cartographic projection. It is often used in order to produce polar-centric maps such that any distance measured along a line of longitude is an accurate geodesic distance. This is particularly useful for building statistical representations as it allows us to measure straight line distances that correspond to geodesic distances between points.

Given a set of k training images we can construct the AEP by first calculating the surface normals. We can construct the mean-surface normal direction at a particular pixel (i, j) by taking each normal from the training set at that pixel and computing an average.

$$\hat{n}(i,j) = \frac{\frac{1}{K}\sum_{k=1}^{K} n_k(i,j)}{\|\frac{1}{K}\sum_{k=1}^{K} n_k(i,j)\|}$$

We can then represent a surface normal through it's elevation angle $\theta_k(i, j)$ and azimuth angle $\phi_k(i, j)$

$$\theta_k(i,j) = \frac{\pi}{2} - \arcsin n_k(i,j)_z$$
$$\phi_k(i,j) = \arctan \frac{n_k(i,j)_y}{n_k(i,j)_x}$$

We also construct the average azimuth and elevation angles as

$$\hat{\theta}(i,j) = \frac{\pi}{2} - \arcsin \hat{n}(i,j)_z$$
$$\hat{\phi}(i,j) = \arctan \frac{\hat{n}(i,j)_y}{\hat{n}(i,j)_x}$$

In order to construct the AEP we begin by constructing the tangent plane to the unit sphere at the location of the mean-surface normal. A local coordinate system is used with the x-axis aligned to the local circle of latitude on the unit sphere. The center of the coordinate system is the point of contact between the tangent plane and the unit sphere. We can then calculate a new projected point for every normal $n_k(i, j)$ we will call $(x_k(i, j), y_k(i, j)$. The equations for the unit sphere to the tangent plane coordinate space are

$$
\begin{aligned}
\cos(c) &= \sin(\hat{\theta})\sin(\theta_k) + \cos(\hat{\theta})\cos(\theta_k)\cos(\phi_k - \hat{\phi}) \\
k' &= \frac{c}{\sin(c)} \\
x_k(i,j) &= k'\cos(\theta_k)\sin(\phi_k - \hat{\phi}) \\
y_k(i,j) &= k'(\cos(\hat{\theta})\sin(\phi_k) - \sin(\hat{\theta})\cos(\theta_k)\cos(\phi_k - \hat{\phi})
\end{aligned}
$$

The inverse equations - from tangent coordinates back to unit sphere space are

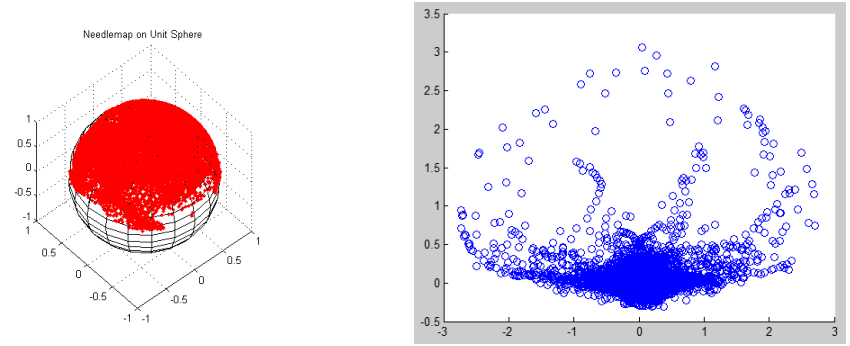
$$
\psi = \begin{cases} \frac{x_k \sin(c)}{c\cos(\hat{\theta})\cos(c) - y_k \sin(\hat{\theta})\sin(c)} & \text{for } \hat{\theta} \neq \pm\frac{\pi}{2} \\ -\frac{x_k}{y_k} & \text{for } \hat{\theta} = \frac{\pi}{2} \\ \frac{x_k}{y_k} & \text{for } \hat{\theta} = -\frac{\pi}{2} \end{cases}
$$

$$
c = \sqrt{x_k(i,j)^2 + y_k(i,j)^2}
$$

with projected angles

$$
\begin{aligned}
\theta_k(i,j) &= \arcsin(\cos(c)\sin(\hat{\theta}) - (1/c)y_k \sin(c)\cos(\hat{\theta})) \\
\phi_k(i,j) &= \hat{\phi} + \arctan(\psi)
\end{aligned}
$$

Figure 3.1 shows the outcome of rendering a set of surface normals on the unit sphere. It also shows the projected space constructed by applying the AEP on the normals.



(a) Projection of the needlemap to the unit sphere

(b) Projection to the tangent plane at the mean point

Figure 3.1: AEP projections

3.1.2 Principal Geodesic Analysis

Principal Geodesic Analysis (PGA) was a technique first proposed by Fletcher et al.[5] for use in tensor MRI data. It looks to build statistics on the variability of data on a manifold. It is seen as a generalization of Principal Component Analysis on Euclidean spaces to non-Euclidean manifolds. In the case of our needlemap we look to build a statistical model by viewing the normals as lying on a spherical manifold. More specifically we require a Reimannian manifold as we need to be able to measure distances between points on the manifold.

Manifolds

A manifold is a topological space that is locally Euclidean. That is to say that for any point P that lies on the manifold there is a neighbourhood around P that lies in the same Euclidean space. A sphere can be seen as a 2-manifold as it is possible to 'flatten' out local regions on the sphere on to planes. This process of projecting manifolds to a Euclidean space is known as charting. A common case of this is the idea of building 2D planar maps of the surface of the spherical Earth. Any manifold can be completely described by a series of charts, called an atlas.

A differentiable manifold is a manifold where each chart is locally similar enough to a linear space to allow calculus to be performed. If the transition from one chart to another is differentiable then any computations done in one chart are valid in any other differentiable chart. A Reimannian manifold is a smooth differentiable manifold that allows the computation of distances and angles on the manifold. This is formally defined as each tangent space being equipped with an inner product function that varies smoothly from point to point. In the case of a 2-sphere the tangent space can be thought of as the plane that intersects a point on the surface of the sphere with a normal perpendicular to the surface. The inner product for the 2-sphere would be the dot product. This is a particularly useful result for the analysis of normals as it allows us to accurately measure distances between points on the sphere. These distances are required to allow us to accurately perform PCA on the dataset.

The Log and Exponential Map

Two important operators for the 2-sphere manifold are the Log and Exponential maps. Given a point on the surface of a sphere and the normal n at that point we can define a plane tangent to the sphere at n. If we then have a vector $\vec{v}$ that points to another point on the tangent plane then we define the exponential map, Exp_n, as the point on the sphere that is distance $\|\vec{v}\|$ along the geodesic in the direction of $\vec{v}$ from n.

The log map, Log_n is the inverse of the exponential map. Given a point on the surface of the sphere it returns the corresponding point on the tangent plane at n.

The Reimannian distance between two points, also known as the geodesic distance, is denoted as $d(n1, n2)$. In the case of the 2-sphere the geodesic distance can be defined in terms of the log map

$$d(n1, n2) = \|Log_{n1}(n2)\| \tag{3.1}$$

Figure 3.2 below shows a visualization of the exponential map. Here a single normal represented by the point $(0, 1, 0)$ is projected by a vector of $(0.5, 0, 0)$. The blue line around the line of latitude shows the geodesic distance between the points. Figure 3.3 shows a visualization of the log map. Here we see the tangential plane at the north pole and a point on the surface of the sphere projected back on the tangent plane.

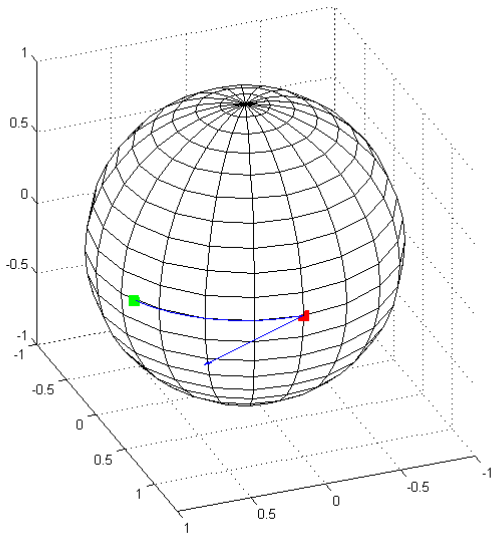


Figure 3.2: Red point, p, is the point $(0, 1, 0)$, the vector, $\vec{v}$, is $(0.5, 0, 0)$ and the green point is $Exp_p(\vec{v})$

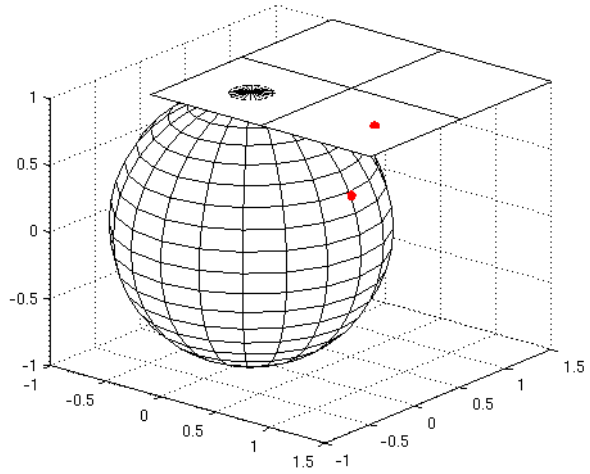


Figure 3.3: Tangent plane at n, the north pole $(0, 0, 1)$. Shows the point on the sphere, p $(0.909, 0, 0.431)$, projected back on the the tangent plane by $Log_n(p)$

Practical Implementation of the Log and Exponential Maps

In order to actually calculate the Log and Exp we use the formulation specified by Smith and Patel in [19]. This uses a stereographic projection in order to project every point on the sphere on to a plane defined in terms of a particular base point. We can define the Log map of a point x on the surface of the sphere at a base point b as follows:

1. Define a point $-b$ that is opposite the base point b
2. Calculate the tangent vector v' that is the vector that intersects the point $-b$, x and the tangent plane at b
3. Resize the tangent vector v' to have the correct magnitude such that $\|v\| = Log(b, x)$

Mathematically the Log map can be defined as

$$Log_b(x) = b + \frac{\theta(v' - b)}{\|v' - b\|}$$

where b is the base point (point of projection) and

$$
\begin{aligned}
v' &= \frac{2b + x}{\|b + x\| \cos\alpha} - b \\
\alpha &= \arccos\left(\frac{4 + \|b + x\|^2 - \|b - x\|^2}{4\|b + x\|}\right) \\
\theta &= \arccos(b \cdot x)
\end{aligned}
$$

Similarly the exponential map can be calculated by

$$
Exp_b(v) = \sqrt{2(1 + \cos\theta)} \left(\frac{v' + b}{\|v' + b\|}\right) - b
$$

where

$$
\begin{aligned}
\theta &= \|b - v\| \\
v' &= b + \frac{(v - b)2\tan\alpha}{\|v - b\|} \\
\alpha &= \arccos\left(\frac{4 + 2(1 + \cos\theta) + (2\sin(\theta/2))^2}{4\sqrt{2(1 + \cos\theta)}}\right)
\end{aligned}
$$

Figure 3.4 below gives a diagram of the equations above. Here we can clearly see the base point b at the top of the circle. We then define the vector v' that intersects $-b$, x and the tangent line at b. The point marked v is the resized vector that represents the true geodesic distance between b and x. It also shows the angle θ used in the exponential map reconstruction.

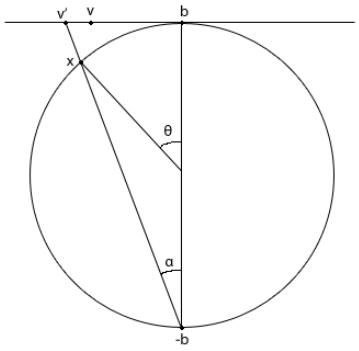


Figure 3.4: Cross section showing the steregraphic projection from a base point $b = (0, 1, 0)$

Spherical Medians

The classic Euclidean mean for a sphere is defined as

$$\bar{n}(i,j) = \frac{\frac{1}{N}\sum_{n=1}^{N} n_k(i,j)}{\|\frac{1}{N}\sum_{n=1}^{N} n_k(i,j)\|}$$

which is also often referred to as the centre of mass. This is the Euclidean centre of the set of points normalized back on to a point on the sphere. This lends itself to calculating the extrinsic mean as it looks to minimize the Euclidean distance on the sphere.

A better representation for a sphere would be to treat each data point as lying on the sphere and use their geodesic distance from one another. As a 2-sphere is a Reimannian manifold this amounts to being the Reimannian distance calculated using the inner product. In the case of a 2-sphere this is the *arccosine* of the dot product between the points. However, as shown earlier in Equation 3.1 we use the log map in order to calculate the distance between two points on a sphere. Formally this is defined as performing a gradient descent to minimize the equation

$$\mu_p = Exp_p(\frac{1}{K}\sum Log_p(n_i)) \tag{3.2}$$

taken from Pennec's paper [21]

Computing Principal Geodesics

Now that we have descriptions of calculating distances on spherical manifolds we can perform statistical analysis on the surface normals. In standard PCA each eigenvector calculated gives a straight line. In PGA this straight line is with replaced by a geodesic curve that lies on a geodesic manifold. In the case of the 2-sphere this amounts to a great circle on the sphere. In PCA data is projected into a lower-dimensional subspace in a linear way. In PGA a different kind of projection operator is required. The projection must be on to the nearest geodesic on the sphere. However, due to the nature of manifolds we can approximate this projection linearly in the tangent space at each point. Formally the projection operator is defined as

$$\pi_G(n_0) = \arg\min_{n \in G} d(n_0, n)^2$$

This is then approximated in the tangent space as

$$Log_\mu(\pi_G(n_0)) \approx \sum_{i=1}^{K} v_i \cdot Log_\mu(n_0)$$

Where v_i are part of an orthogonal basis of vectors in the tangent plane.

The minimization problem of computing all of the principal directions is equivalent to the standard PCA analysis of the vectors $Log_\mu(n_i)$ in the tangent plane. Formally the process of computing the principal geodesics can be described as

Algorithm 1 Calculate Principal Geodesic Analysis

Input: $x_1, ..., x_N \in S^2$
Output: Principal Directions, $v_k \in T_\mu S^2$
μ = Intrinsic mean of $\{x_i\}$ (Equation 3.2)
$u_i = Log_\mu(x_i)$
$S = (1/N) \sum_{i=1}^{N} u_i u_i^T$
$\{v_k\}$ = Eigenvectors of S

3.1.3 The Hard Irradiance Constraint

In shape-from-shading we assume a Lambertian reflectance model. Due to this assumption we know that the surface normal is constrained to fall somewhere on a cone whose axis is the light source direction and whose opening angle is the inverse cosine of the image intensity. By maintaing this hard irradiance constraint we can build more accurate models by fitting the statistical model to the best-fit surface normals and then rotating the normals back on to the nearest on-cone position.

Worthington and Hancock [26] describe a relatively simple method of rotating a normal back to the closest on-cone position. It involves using the generalized formula for rotating a vector about the axis (u, v, w) in 3D space by a particular angle ϕ given by the matrix:

$$\begin{bmatrix} \cos\phi + u^2c' & uvc' - w\sin\phi & uwc' + v\sin\phi \\ vuc' + w\sin\phi & \cos\phi + v^2c' & vwc' - u\sin\phi \\ wuc' - v\sin\phi & wvc' + u\sin\phi & \cos\phi + w^2c' \end{bmatrix} \tag{3.3}$$

where $c' = 1 - \cos\phi$.
In the case of the hard irradiance constraint we rotate the matrix about the angle ϕ:

$$\begin{aligned} \phi &= \theta(x, y) - \arccos(n^{(t)}(x, y) \cdot s) \\ (u, v, w) &= n^{(t)}(x, y) \times s \end{aligned}$$

where $n^{(t)}(x, y)$ represents an off-cone surface normal at iteration t, s denotes the direction of the light source from the surface and $\theta(x, y)$ represents the intensity at pixel (i, j).

This method is usually initialized by placing the surface normals on the reflectance cones such that they are aligned with the direction opposite to

the local image gradient. However, since we are using a statistical model, a better initialization is to align them on the cone at the point closest to the local average normal direction. Figure 3.5 gives a diagram of the on-cone rotation performed by the rotation matrix given in Equation 3.3. We can clearly see the off-cone normal n^k being rotated back on to the reflectance cone by the angle ϕ.

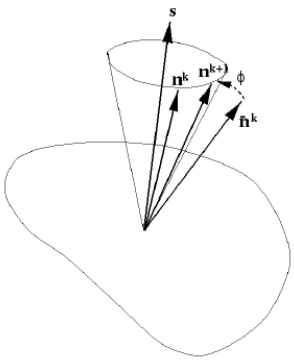


Figure 3.5: Taken from [26]. It shows the notion of an off-cone surface normal $\bar{n}^k$ being rotated by an angle ϕ back to the nearest on-cone position

3.1.4 Implementing Azimuthal Equidistant Projection

In order to test the robustness of the Azimuthal Equidistant Projection we implemented Smith's algorithm in Matlab. The most challenging aspect of the reconstruction was ensuring that the calculated azimuth and elevation angles resided within the correct quadrant. Due to the periodic nature of the *sin* and *cosine* functions it was necessary to shift the elevation angle to ensure it was in the range $[-\pi/2, \pi/2]$.

Smith's algorithm centers around iteratively reconstructing the field of surface normals using the statistical model and then imposing the hard-irradiance constraint. A summarization of the algorithm is as follows:

```
for i=1:3
    n = npp;

    % Loop until convergence
    v0 = spherical2azimuthal(n, mu);

    % vector of best-fit parameters
    b = U' * v0;
    % transformed coordinates
    vprime = U * b;

    nprime = azimuthal2spherical(vprime, mu);
```

```
    % Normalize
    nprime = reshape2colvector(nprime);
    nprime = reshape(bsxfun(@rdivide, nprime, colnorm(nprime)), [], 1);

    npp = OnConeRotation(theta, nprime, s);
end
```

where *OnConeRotation* performs the rotation described by the matrix 3.3.

Smith's original paper suggests testing for convergence using the cost metric $\sum \arccos(n'(i,j) \cdot n''(i,j)) < \epsilon$. However, there is no formal proof for convergence of this method and thus a fixed number of iterations was used.

In order to initialize the normals each normal was set to lie on the local average normal direction:

$$n^{(0)}(i,j) = \begin{pmatrix} \sin\theta(i,j)\cos\hat{\phi}(i,j) \\ \sin\theta(i,j)\sin\hat{\phi}(i,j) \\ \cos\theta(i,j) \end{pmatrix}$$

3.1.5 Implementing Principal Geodesic Analysis

To compare the reconstructive ability of Principal Geodesic Analysis against the Azimuthal Equidistant Projection we implemented an equivalent iterative reconstruction algorithm using PGA. As in the AEP algorithm it involves finding a set of best-fit parameters and then enforcing the hard irradiance constraint and rotating the normals back on to the relfectance cone.

In order to calculate PGA a set of spherical medians must be found. Finding a set of spherical medians is an optimization problem that amounts to performing a gradient descent to minimize the Equation 3.2. This algorithm is performed over every single normal in the training set in order to find the spherical median for each point across all the training images. Once the set of spherical medians is found PCA can be performed on the training set by projecting every normal to the tangent plane defined by the spherical median. Formally we construct a matrix, D, defined as

$$D = [v^1, ..., v^K]$$
$$v^k = [Log_{\mu_1}(n_p^k), ..., Log_{\mu_N}(n_N^k)]^T$$

PCA is then performed on the matrix D and, similar to the AEP reconstruction, we can define a reconstruction algorithm as follows

```
for i = 1:3
    % Reshape the column vector to a 3xN matrix
    ncol = reshape2colvector(n);
    v0 = ncol;
```

```
    % Project using the log map to the tangent planes
    % defined at each spherical median
    for p = 1:size(ncol, 2)
        v0(:, p) = logmap(mus(:, p), ncol(:, p));
    end
    % Reshape back to a column vector
    v0 = reshape(v0, [], 1);

    % Vector of best-fit parameters
    b = U' * v0;
    % Transformed coordinates
    vprime = U * b;

    % Reshape the column vector to a 3xN matrix
    vcol = reshape2colvector(vprime);
    nprime = vcol;
    % Reconstruct the normals from the projected
    % parameters
    for p = 1:size(vcol, 2)
        nprime(:, p) = expmap(mus(:, p), vcol(:, p));
    end
    % Reshape back to a column vector
    nprime = reshape(nprime, [], 1);

    % Normalize
    nprime = reshape2colvector(nprime);
    nprime = reshape(bsxfun(@rdivide, nprime, colnorm(nprime)), [], 1);

    % Satisfy the hard irradiance constraint
    npp = OnConeRotation(theta, nprime, s);
end
```

As with the AEP reconstruction we apply this for a fixed number of iterations. The output is a set of normals that respect the hard irradiance constraint.

3.2 Novel Reconstruction Method

For our reconstruction method we borrow from ideas used in morphable model reconstructions. Given a texture and shape model of faces, we can generate a reconstruction of a given input image by finding the best-fitting coefficients. We do this by first noting that the intensity of any given pixel, $I(x)$ can be represented as given in Equation 2.5. We can, therefore, look to reconstruct a face by minimizing the equation

$$\sum_x \|I(x) - \rho_x n_x \cdot s\|^2$$

where x represents every pixel in the image and s represents the light vector. In terms of the statistical model we find two sets of coefficients, a and c, that minimize the equation

$$\sum_x \|I(x) - \left(\sum_{i=1}^{P} a_i b_i^x + \mu_t^x\right)\left(\sum_{i=1}^{P} c_i \tilde{n}^x \cdot s + \mu_n^x \cdot s\right)\|^2 \tag{3.4}$$

where b_i^x are the principal components of the texture, $\tilde{n}$ the principal components of the shape, μ_t^x is the average texture vector, μ_n^x is the average normal vector, a the texture coefficients and c the shape coefficients.

In order to solve for the texture coefficients, a, we can rearrange the above equation as follows

$$\begin{aligned}
&\sum_x \left(I(x) - w_x\left(\sum_{i=1}^{P} a_i b_i^x + \mu_t^x\right)\right)^2 \\
&= \sum_x \Big(I(x)^2 - 2w_x I(x)\sum_{i=1}^{P} a_i b_i^x - 2w_x I(x)\mu_t^x + w_x^2\left(\sum_{i=1}^{P} a_i b_i^x\right)^2 + \\
&\qquad 2w_x^2\mu_t^x\sum_{i=1}^{P} a_i b_i^x + {\mu_t^x}^2 w_x^2\Big) \\
&= \sum_x I(x)^2 - 2\sum_{i=1}^{P} a_i \sum_x w_x I(x) b_i^x - 2\sum_x w_x I(x)\mu_t^x + \\
&\qquad \sum_x w_x^2 \sum_{i=1}^{P}\sum_{j=1}^{P} a_i a_j b_i^x b_j^x + 2\sum_{i=1}^{P} a_i \sum_x w_x^2 \mu_t^x b_i^x + \sum_x {\mu_t^x}^2 w_x^2
\end{aligned}$$

where $w_x = \sum_{i=1}^{P} c_i \tilde{n} \cdot s + \mu_n^x \cdot s$.
We can then re-arrange the above result in to the matrix form below

$$\sum_x I(x)^2 - 2a^T \sum_x k_x + a^T \sum_x M_x a + \sum_x {\mu_t^x}^2 w_x^2$$

where

$$k_x = \begin{bmatrix} w_x b_i^x \left(I(x) - w_x \mu_t^x\right) \\ \vdots \\ w_x b_P^x \left(I(x) - w_x \mu_t^x\right) \end{bmatrix} \qquad M_x = b_x b_x^T \qquad b_x = \begin{bmatrix} w_x b_i^x \\ \vdots \\ w_x b_P^x \end{bmatrix}$$

In order to minimize this formulation we can take the zero-crossing of the derivative. Therefore we calculate the gradient with respect to a

$$\begin{aligned} \nabla a &= -2 \sum_x k_x + 2 \sum_x M_x a = 0 \\ a &= \left[\sum_x M_x \right]^{-1} \sum_x k_x \end{aligned} \tag{3.5}$$

which we can solve in order to give a set of best-fit texture coefficients. It thus follows that we can provide a similar rearrangement to calculate the shape coefficients c. We begin by redefining Equation 3.4 as

$$\sum_x \| I(x) - \left(\sum_{i=1}^{P} c_i q_i^x \rho_x + \tilde{q}_x \right) \|^2$$

where $q_i^x = \tilde{n} \cdot s$, $\tilde{q}_x = \mu_n^x \cdot s$ and $\rho_x = \sum_{i=1}^{P} a_i b_i^x$.
Similarly to before we can produce a rearrangment as follows

$$\begin{aligned} &\sum_x \left(I(x) - \rho_x \left(\sum_{i=1}^{P} c_i q_i^x + \tilde{q}_x \right) \right)^2 \\ &= \sum_x \Big(I(x)^2 - 2\rho_x I(x) \sum_{i=1}^{P} c_i q_i^x - 2\rho_x I(x) \tilde{q}_x + \rho_x^2 \left(\sum_{i=1}^{P} c_i q_i^x \right)^2 + \\ &\qquad 2\rho_x^2 \tilde{q}_x \sum_{i=1}^{P} c_i q_i^x + \tilde{q}_x^2 \rho_x^2 \Big) \\ &= \sum_x I(x)^2 - 2 \sum_{i=1}^{P} c_i \sum_x \rho_x I(x) q_i^x - 2 \sum_x \rho_x I(x) \tilde{q}_x + \\ &\qquad \sum_x \rho_x^2 \sum_{i=1}^{P} \sum_{j=1}^{P} c_i c_j q_i^x q_j^x + 2 \sum_{i=1}^{P} c_i \sum_x \rho_x^2 \tilde{q}_x q_i^x + \sum_x \tilde{q}_x^2 \rho_x^2 \end{aligned}$$

where

$$\tilde{k}_x = \begin{bmatrix} \rho_x q_i^x \left(I(x) - \rho_x \tilde{q}_x\right) \\ \vdots \\ \rho_x q_P^x \left(I(x) - \rho_x \tilde{q}_x\right) \end{bmatrix} \qquad \tilde{M}_x = l_x l_x^T \qquad l_x = \begin{bmatrix} \rho_x q_i^x \\ \vdots \\ \rho_x q_P^x \end{bmatrix}$$

We can then solve the derivative of the above rearrangment to yield an equation for c

$$\begin{aligned} \nabla c &= -2\sum_x \tilde{k}_x + 2\sum_x \tilde{M}_x c = 0 \\ c &= \left[\sum_x \tilde{M}_x\right]^{-1} \sum_x \tilde{k}_x \end{aligned} \tag{3.6}$$

Iterating the solutions of both Equation 3.6 and Equation 3.5 will yield a set of best-fit coefficients to reconstruct a given image. We can initialize a and c to any random values and then terminate the algorithm when the cost metric from Equation 3.4 converges. This algorithm is a reasonably intuitive reconstruction of the parameters given a statistical model. It is important to note that at every iteration the normals must be normalized in order to yield an accurate result. This is achieved rearranging w_x above to

$$w_x = \sum_{i=1}^{P} \left(\frac{c_i \tilde{n} + \mu_n^x}{\|c_i \tilde{n} + \mu_n^x\|} \right) \cdot s$$

Chapter 4

Experiments and Results

In this chapter we describe the experimental process used in order to evaluate the performance of the statistical shape-from-shading techniques and present our results.

4.1 Experimental Process

To generate the results presented in Sections 4.3 and 4.4 the following procedure was followed:

1. **Generate a training set** as described in Section 4.2. This set of images was used as the basis of the statistical model.

2. **Align the training set.** This was to ensure correspondance across the faces. Without correspondance, the principal components calculated will not describe the facial features that carry the most information. A simple affine transformation was perfomed in order to achieve alignment.

3. **Perform statistical analysis on the training set** to build a model. PCA was performed on three different data sets: the unprojected normals, the AEP projected normals and the PGA projected normals.

4. **Align the testing dataset.** Two seperate sample datasets were created; from a database of photometric stereo images and from the morphable model itself. The morphable model sample involved generating a random set of coefficients in order to create a new texture and shape. This texture and shape were then rendered using four novel lighting conditions, including shadowing.

5. **Perform the reconstructions.** Four reconstructions were performed from an input image. In order to measure the accuracy of the reconstruction a ground truth set of normals were produced. For the morphable model this was calculated by computing the normals from the

shape vector. A surface was produced from the photometric stereo database by performing photometric stereo as described by Barksy and Petrou[2].

6. **Compare the reconstructions** by computing the angular error and the height differences between the surfaces formed through integration of the normals.

4.2 Generating A Statistical Model

In order to build the statistical model we used the 3D morphable model data set provided by the University of Basel[20]. 199 faces were randomly generated from the data set using the linear reconstruction provided by the morphable model. These faces were restricted to be within a single standard deviation from the mean in order to ensure that true faces and not caricatures were generated. The morphable model provides shape and texture vectors but not any renderings. Therefore, using a simple perspective projection and a single directed light, texture renderings were produced for each face. The underlying shape model provided by the morphable model was also triangulated to provide a normal per vertex. The normal was calculated by taking the average normal of each triangle in which the vertex participated.

Once the faces were generated they were aligned to a template face in order to give correspondance between the points. The morphable model provides landmarks on the face, which were aligned to a pre-calculated facial template using a simple affine transformation. The four landmarks chosen were the corner of each of the eyes, the tip of the nose and the bottom of the chin. The texture and normals were both aligned so that a statistical model could be generated from them. Figure 4.1 shows a representation of what the underlying template face looks like. The red box gives the area that each face was restricted to. Figure 4.2a and Figure 4.2b show an example of a manually aligned texture and an aligned training set texture.

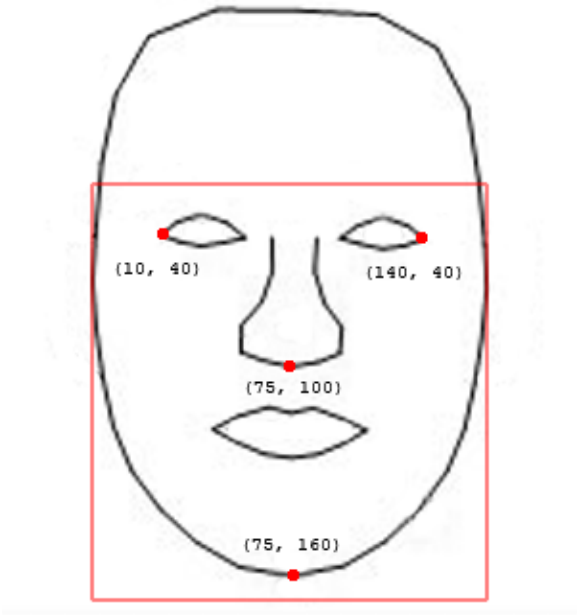
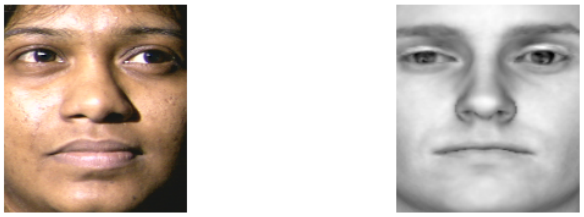


Figure 4.1: The template face used to align the training set with



(a) Manually aligned face (b) Aligned training image

Figure 4.2: Aligned Images

Once aligned the images were concatenated row-wise, in to a single column vector, so that Principal Component Analysis could be performed. A matrix was then created such that each column represents a single image in the training set. The average column vector was calculated and then subtracted from each individual column in order to yield the mean-centered dataset. Three seperate statistical models were then produced. One model consisted of the principal components of the unprojected surface normals. Whilst performing PCA on directional data should not yield an accurate model, it was interesting to see the results produced. The other two mod-

els consisted of; PCA being performed on the data as transformed by the Azimuthal Equidistant Projection and the third after performing Principal Geodesic Analysis.

4.3 Photometric Stereo Database Images

An existing set of photometric stereo images was used to provide a ground truth model. We used the photometric stereo algorithm presented by Barsky and Petrou [2] in order to reconstruct a set of normals. We consider the normals computed by photometric stereo as ground truth due to their accuracy. The accuracy of the normals provided by photometric stereo is described in Section 2.3.

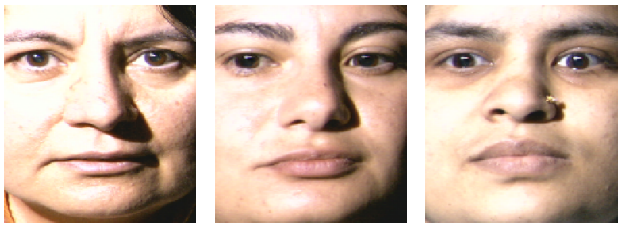
Seven people from the set were chosen and used as examples below. Each of the four images of each person was aligned as described above and photometric stereo was performed. Then, one of the images was chosen and each of the Shape-from-shading techniques described in Section 4.2 were performed on it. We have arranged the data below such that each subsection represents the images generated for each of the seven people chosen. Within each subsection, we present a few different rendered views, including a color coded image of the generated normals. This image gives a good indication as to how well the underlying set of reconstructed normals fits the texture provided. We also include an image that represents the angular error between the reconstruction and the ground truth. The angular error was calculated using the equation

$$|\arctan \frac{\|\vec{g} \times \vec{e}\|}{\vec{g} \cdot \vec{e}}|$$

where $\vec{g}$ is the ground truth normal and $\vec{e}$ is the estimated normal. This is performed for every pixel. The brighter the pixel produced, the larger the error and thus the further the estimate has deviated from the ground truth. Therefore, a perfect match would produce a black image.

Each of the images used for reconstruction is provided in Figure 4.3 below. Below each image is a coded name for the subject. For the following subsections there will be 5 images per row. The first image in every row is the ground truth. The other images, from left to right are: the novel reconstruction, Smith's reconstruction directly on the normals, the AEP reconstruction and the PGA reconstruction.

(a) Bej (b) Bln (c) Fav (d) Mut



(e) Pet (f) Rob (g) Srb

Figure 4.3: Aligned color images from the photometric stereo set

4.3.1 Subject 1 - Bej

In this section we present the reconstruction of the person we refer to as Bej. All the reconstructions were produced from one of the four images provided by the photometric set. The sample image is given in Figure 4.3a. Bej represents a particularly difficult reconstruction as she is wearing a head scarf which causes a white frame around her face.

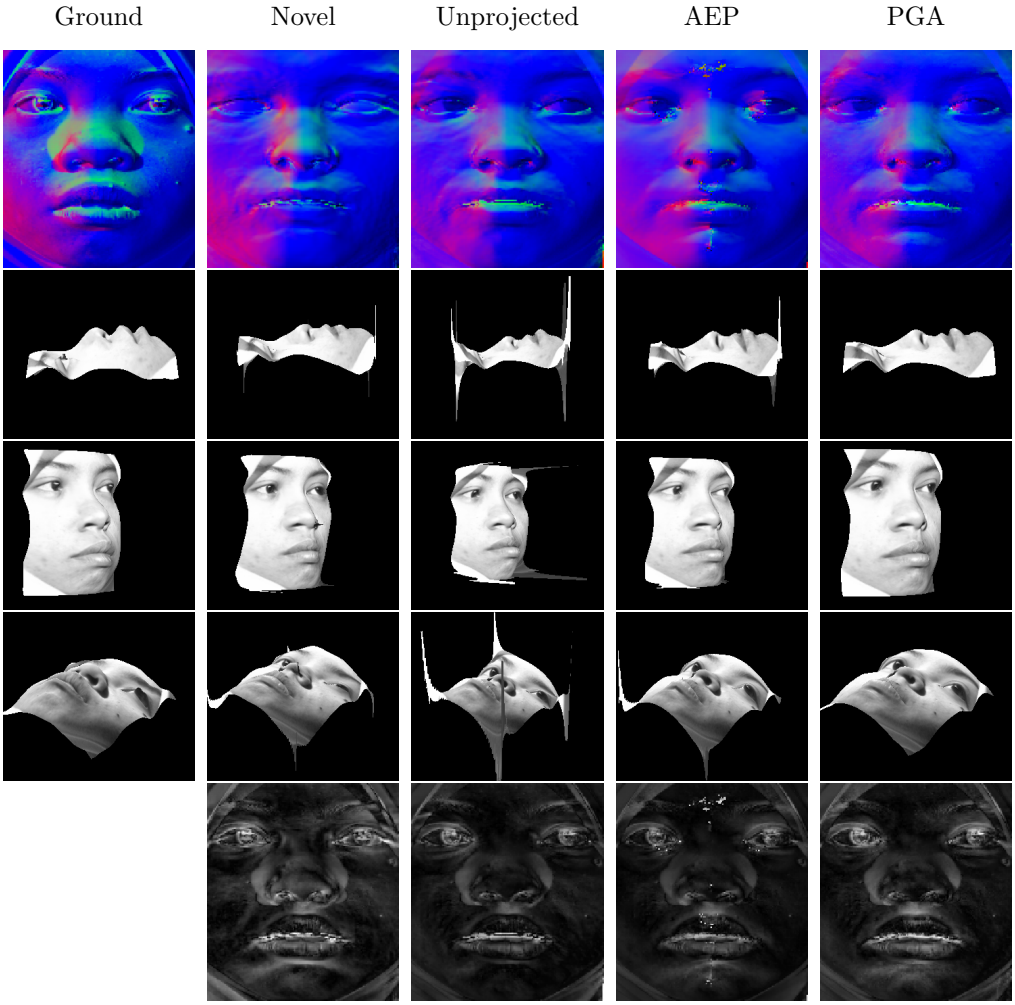


Figure 4.4: The reconstructions of subject bej. The first row is the normal map. The second row is a profile rendering. The third row is a side view. The fourth row is a rendering from below the chin. The final row is the angular error between the reconstruction and the ground truth.

4.3.2 Subject 2 - Bln

In this section we present the reconstruction of the person we refer to as Bln. All the reconstructions were produced from one of the four images provided by the photometric set. The sample image is given in Figure 4.3b.

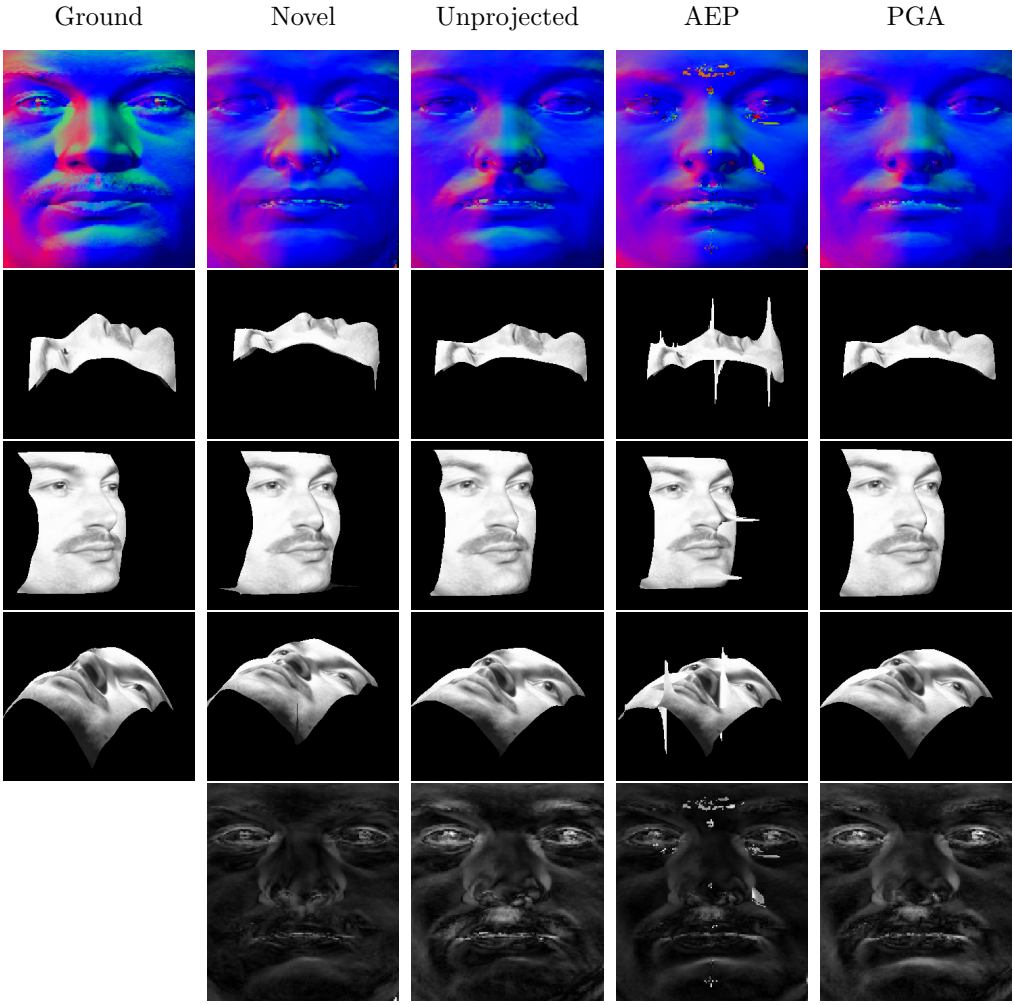


Figure 4.5: The reconstructions of subject bln. The first row is the normal map. The second row is a profile rendering. The third row is a side view. The fourth row is a rendering from below the chin. The final row is the angular error between the reconstruction and the ground truth.

4.3.3 Subject 3 - Fav

In this section we present the reconstruction of the the person we refer to as Fav. All the reconstructions were produced from one of the four images provided by the photometric set. The sample image is given in Figure 4.3c.

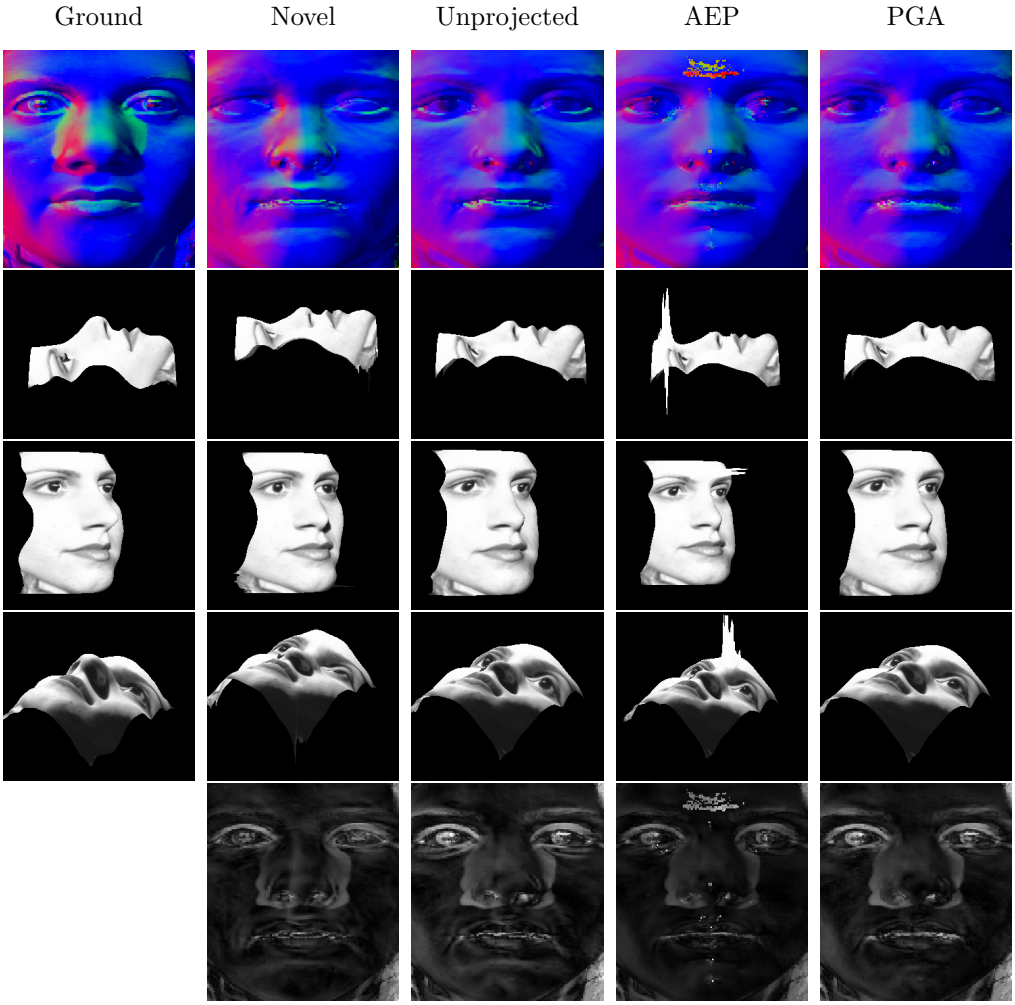


Figure 4.6: The reconstructions of subject fav. The first row is the normal map. The second row is a profile rendering. The third row is a side view. The fourth row is a rendering from below the chin. The final row is the angular error between the reconstruction and the ground truth.

4.3.4 Subject 4 - Mut

In this section we present the reconstruction of the person we refer to as Mut. All the reconstructions were produced from one of the four images provided by the photometric set. The sample image is given in Figure 4.3d.

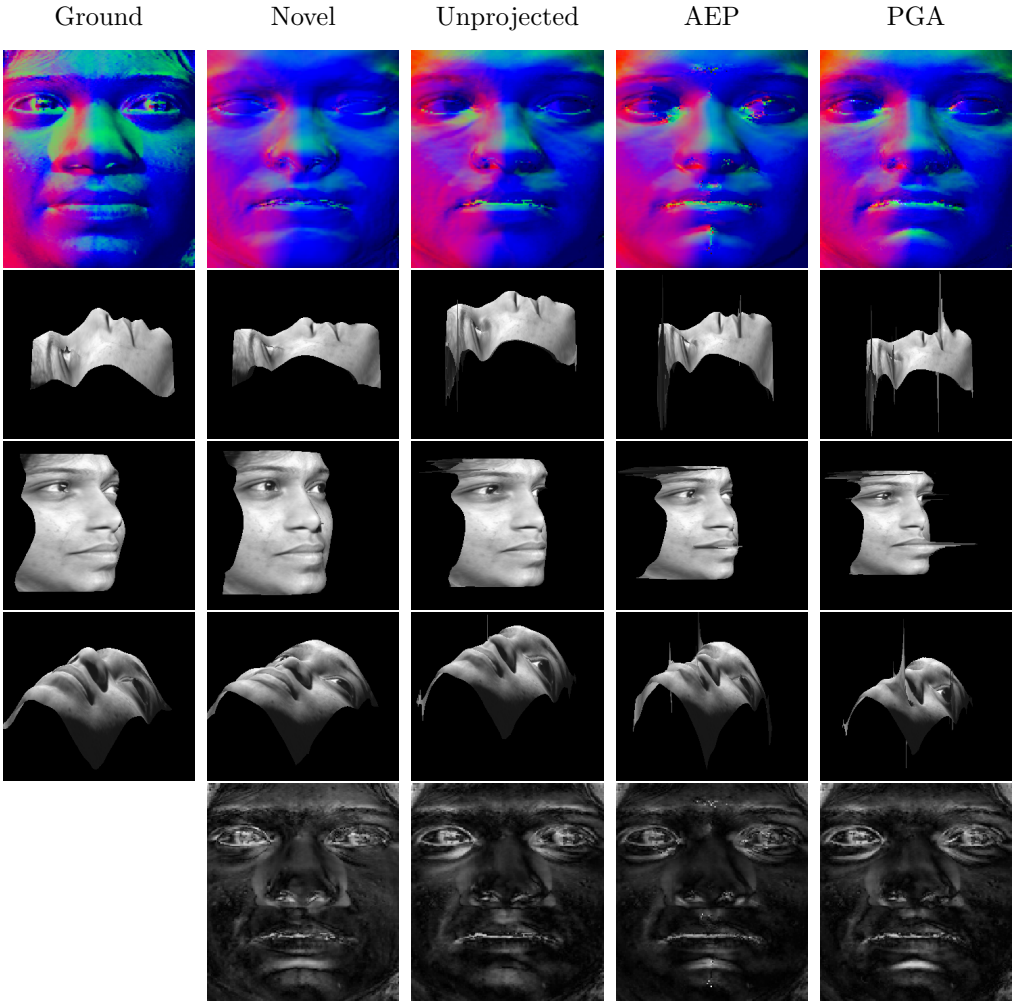


Figure 4.7: The reconstructions of subject mut. The first row is the normal map. The second row is a profile rendering. The third row is a side view. The fourth row is a rendering from below the chin. The final row is the angular error between the reconstruction and the ground truth.

4.3.5 Subject 5 - Pet

In this section we present the reconstruction of the person we refer to as Pet. All the reconstructions were produced from one of the four images provided by the photometric set. The sample image is given in Figure 4.3e.

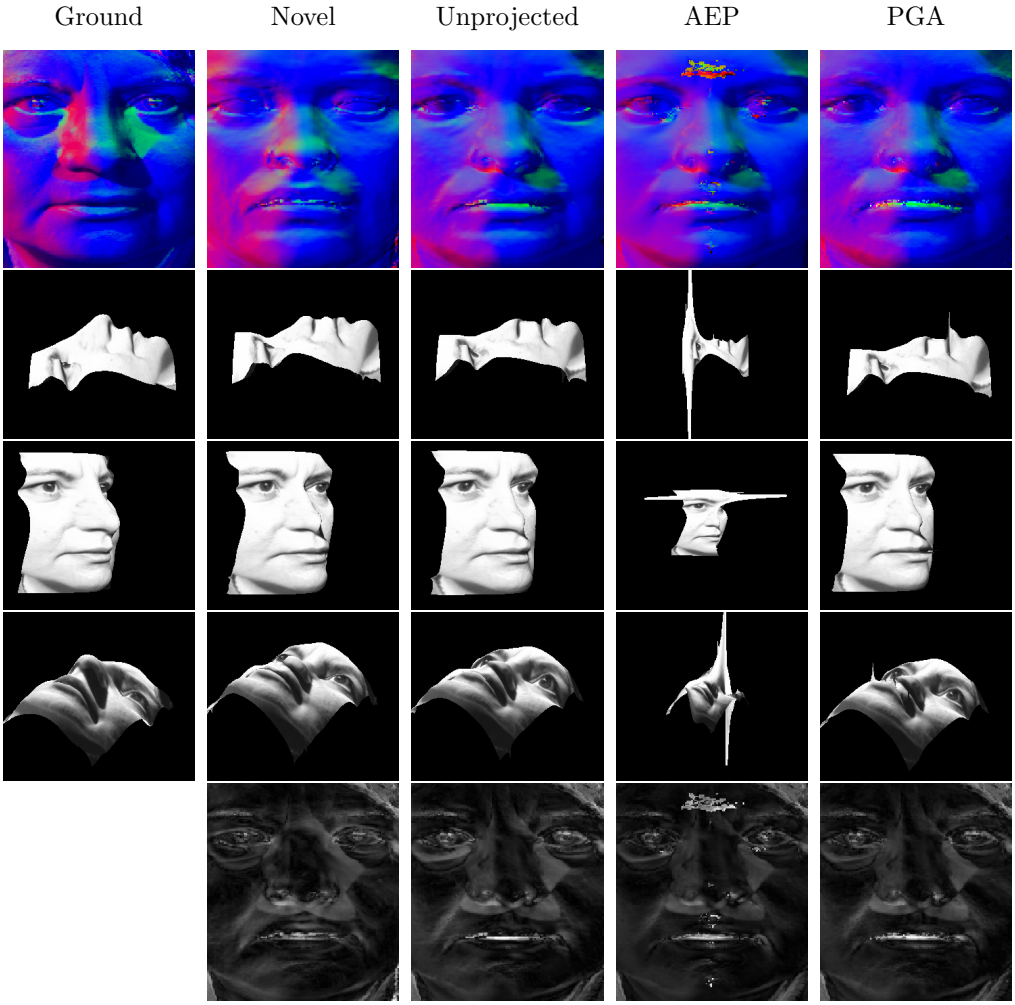


Figure 4.8: The reconstructions of subject pet. The first row is the normal map. The second row is a profile rendering. The third row is a side view. The fourth row is a rendering from below the chin. The final row is the angular error between the reconstruction and the ground truth.

4.3.6 Subject 6 - Rob

In this section we present the reconstruction of the person we refer to as Rob. All the reconstructions were produced from one of the four images provided by the photometric set. The sample image is given in Figure 4.3f.

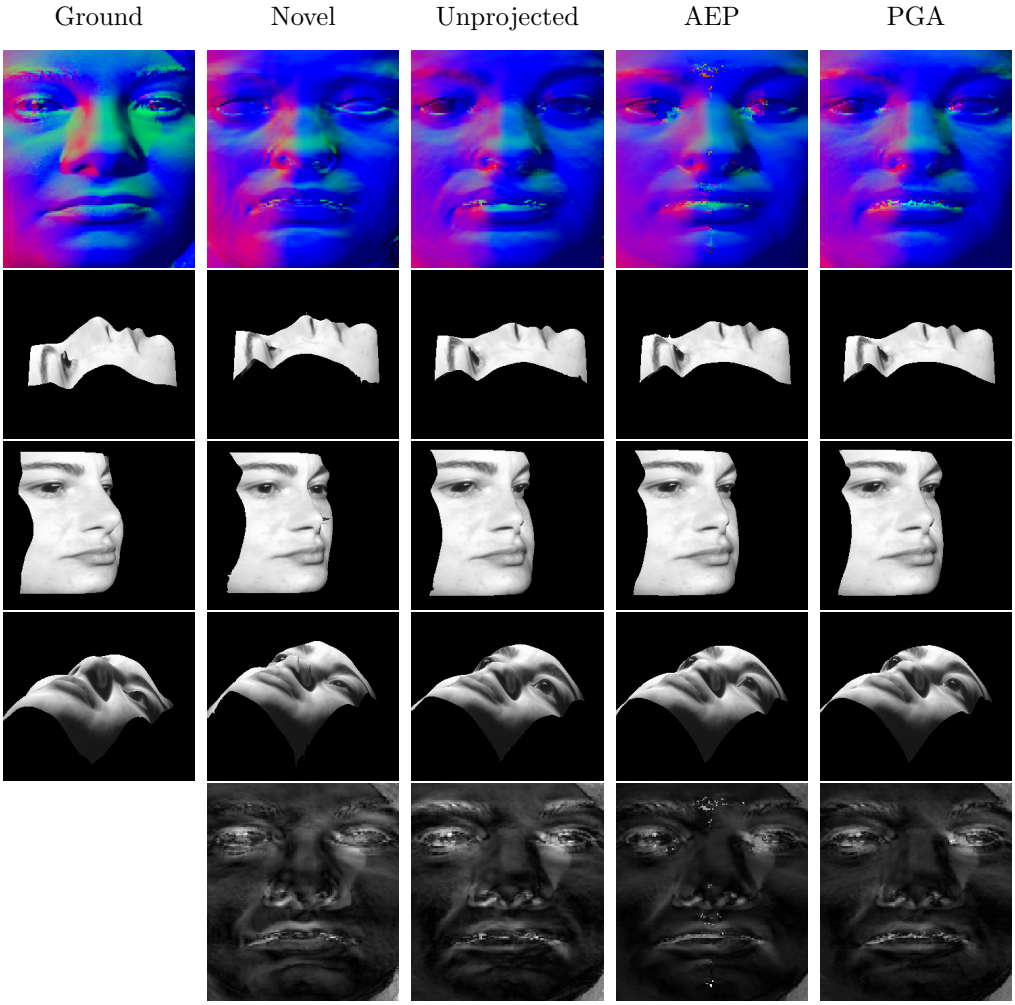


Figure 4.9: The reconstructions of subject rob. The first row is the normal map. The second row is a profile rendering. The third row is a side view. The fourth row is a rendering from below the chin. The final row is the angular error between the reconstruction and the ground truth.

4.3.7 Subject 7 - Srb

In this section we present the reconstruction of the person we refer to as Srb. All the reconstructions were produced from one of the four images provided by the photometric set. The sample image is given in Figure 4.3g.

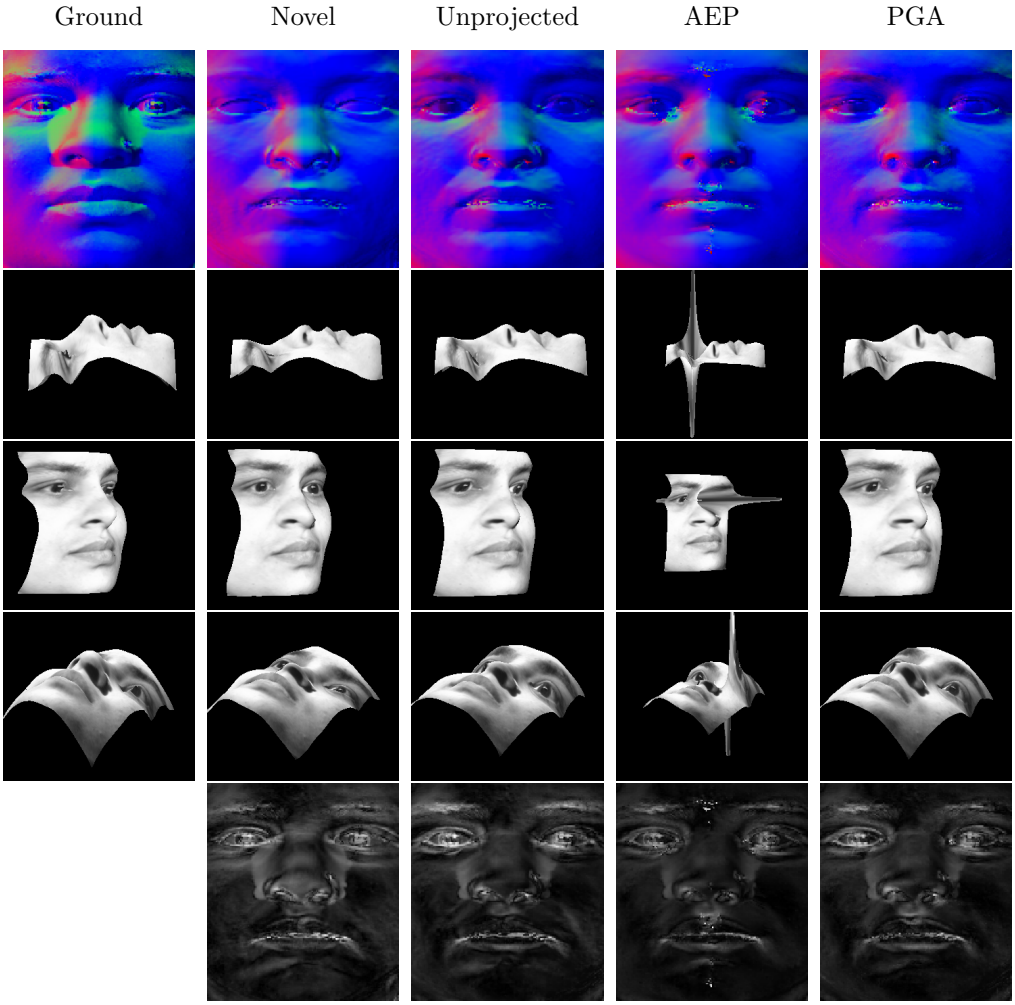


Figure 4.10: The reconstructions of subject srb. The first row is the normal map. The second row is a profile rendering. The third row is a side view. The fourth row is a rendering from below the chin. The final row is the angular error between the reconstruction and the ground truth.

4.4 Illuminated Morphable Model Images

In order to test the ability of the models to reconstruct images that are lit in challenging ways, we generated some images from the morphable model. Although these images are technically in the training set, as they are generated using the same set of eigenvectors as the constructed model, the lighting conditions still provide a challenge. The original training set was ambiently lit and so these generated faces provide a challenge due to specular highlights and self shadowing. Difficult light directions are demonstrated, including from below the chin and above the forehead. These are particularly chal-

lenging, as much of the face is in shadow. Due to the global nature of the reconstructions a lot of this error is propagated across the recovered shape. In the sections below, we present the face under four different lighting conditions. We give the different lighting conditions in terms of their spherical coordinates, θ and ϕ, where θ is the horizontal rotation and ϕ is the vertical rotation. As in Section 4.3, we present the results, from left to right: ground truth, novel reconstruction, Smith's reconstruction directly on the normals, the AEP reconstruction and the PGA reconstruction. The four images used for reconstruction are given in Figure 4.11 below.

(a) $\theta = 0, \phi = 0$ (b) $\theta = 45, \phi = 0$ (c) $\theta = 0, \phi = 45$ (d) $\theta = 0, \phi = -45$

Figure 4.11: Aligned images generated from the morphable model

4.4.1 Lit morphable model - Front

Here we present the morphable model lit from $\theta = 0$, $\phi = 0$. This should be a simple reconstruction as it is simply reconstructing a texture from within the eigenspace.

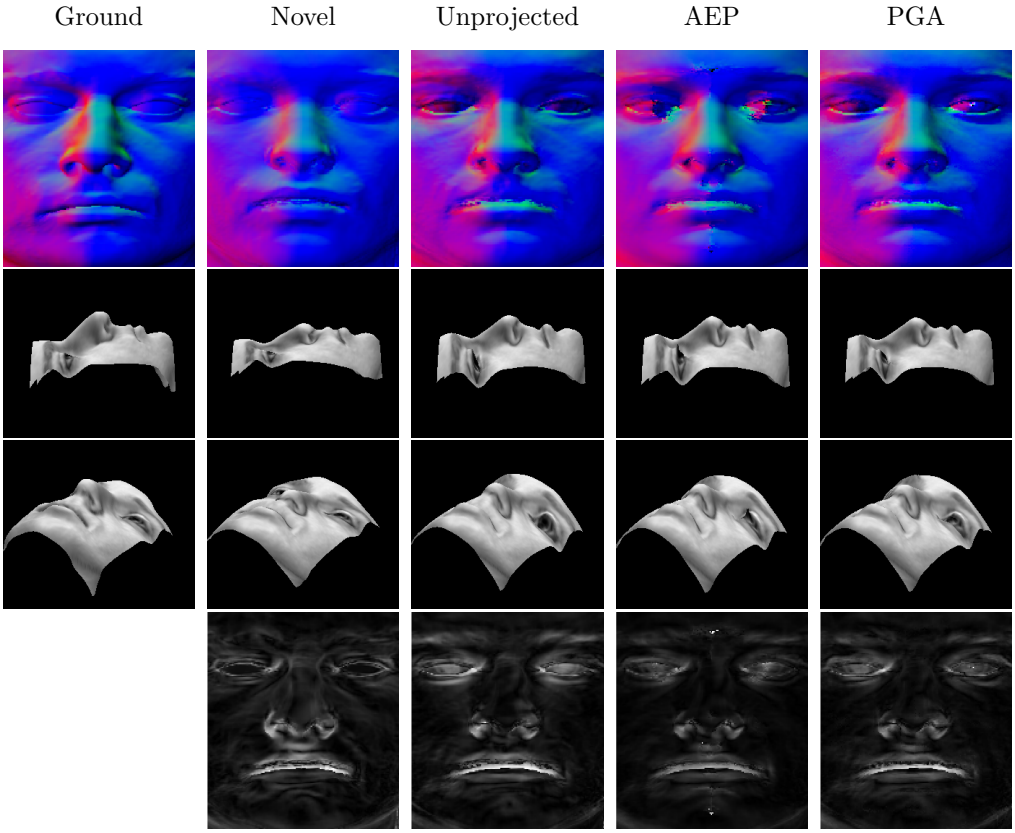


Figure 4.12: The first row is the normal map. The second row is a profile rendering. The third row is a rendering from below the chin. The final row is the angular error between the reconstruction and the ground truth.

4.4.2 Lit morphable model - Left

Here we present the morphable model lit from $\theta = 45$, $\phi = 0$. This causes shadows to be cast by the nose and thus should affect negatively affect the reconstruction. However, due to the global nature of the reconstructions they should still yield reasonable results.

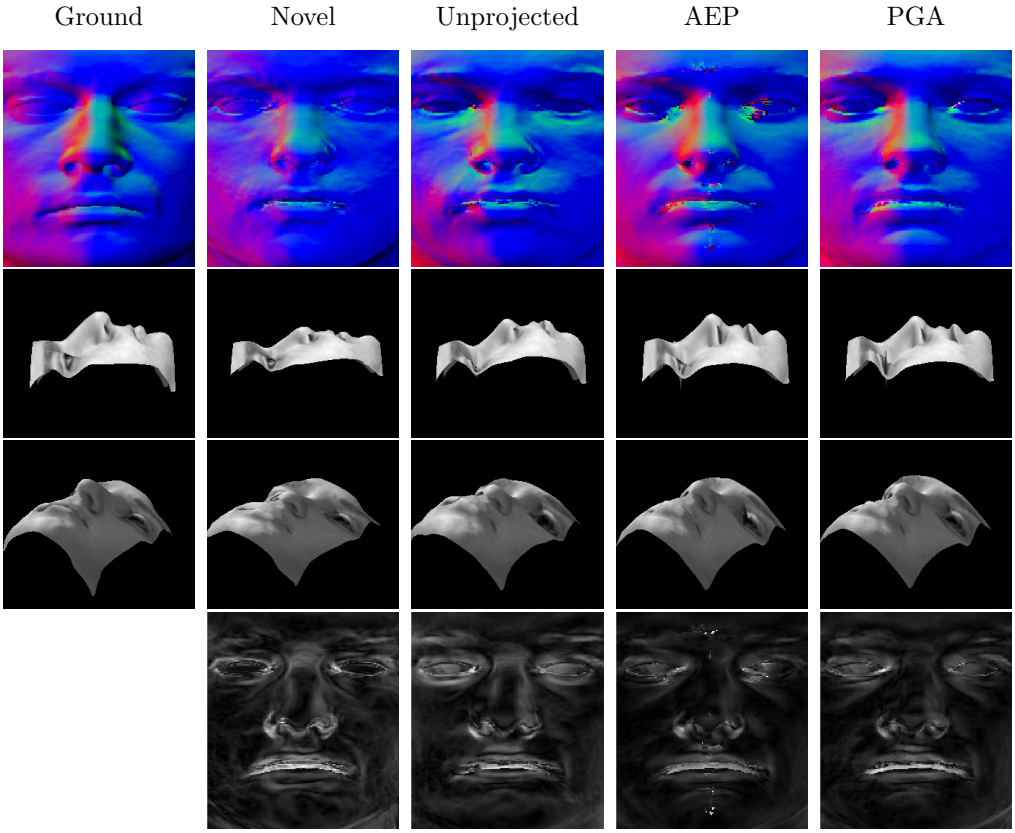


Figure 4.13: The first row is the normal map. The second row is a profile rendering. The third row is a rendering from below the chin. The final row is the angular error between the reconstruction and the ground truth.

4.4.3 Lit morphable model - Above

Here we present the morphable model lit from $\theta = 0$, $\phi = 45$. This puts a signifcant portion of the face in shadow and is a very difficult angle to reconstruct from. It will negatively affect the reconstruction below the nose, particularly the mouth and chin.

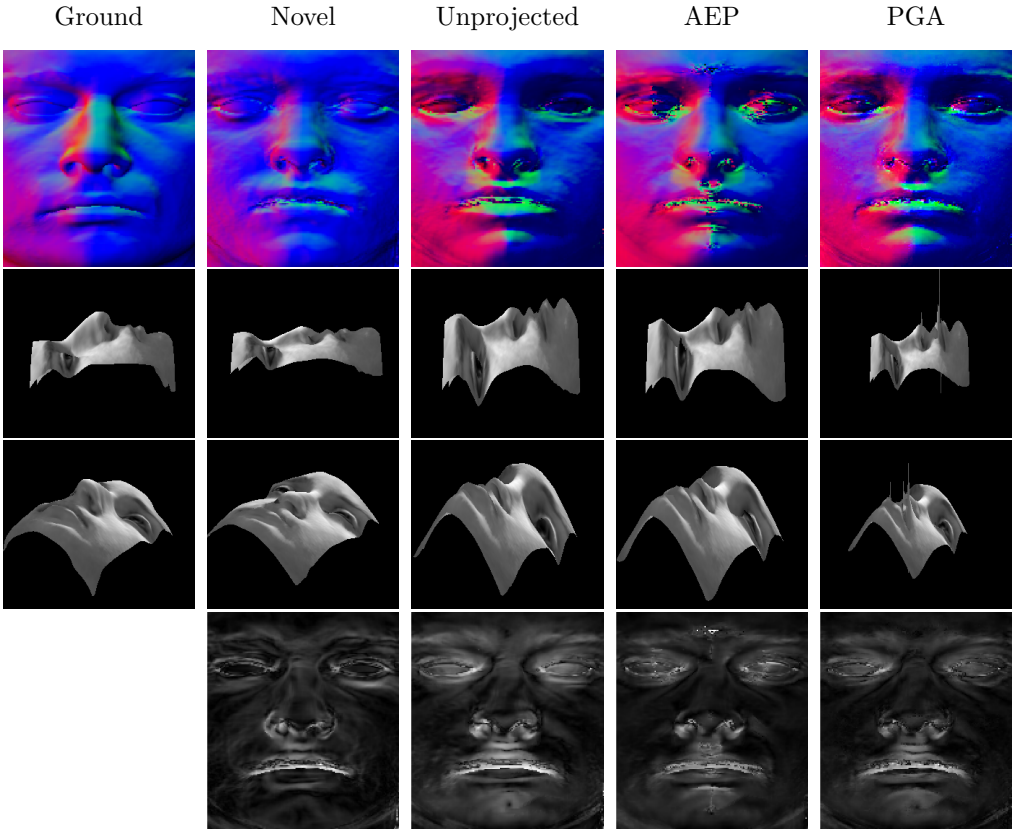


Figure 4.14: The first row is the normal map. The second row is a profile rendering. The third row is a rendering from below the chin. The final row is the angular error between the reconstruction and the ground truth.

4.4.4 Lit morphable model - Below

Here we present the morphable model lit from $\theta = 0$, $\phi = -45$. This is arguably the hardest angle to reconstruct from as the chin causes a large area of occlusion. For this reason a lot of the face is cast in shadow and areas such as the eyes become very innacurate. It also affected by the large shadow that the nose casts across the area of the face beneath the eyes.

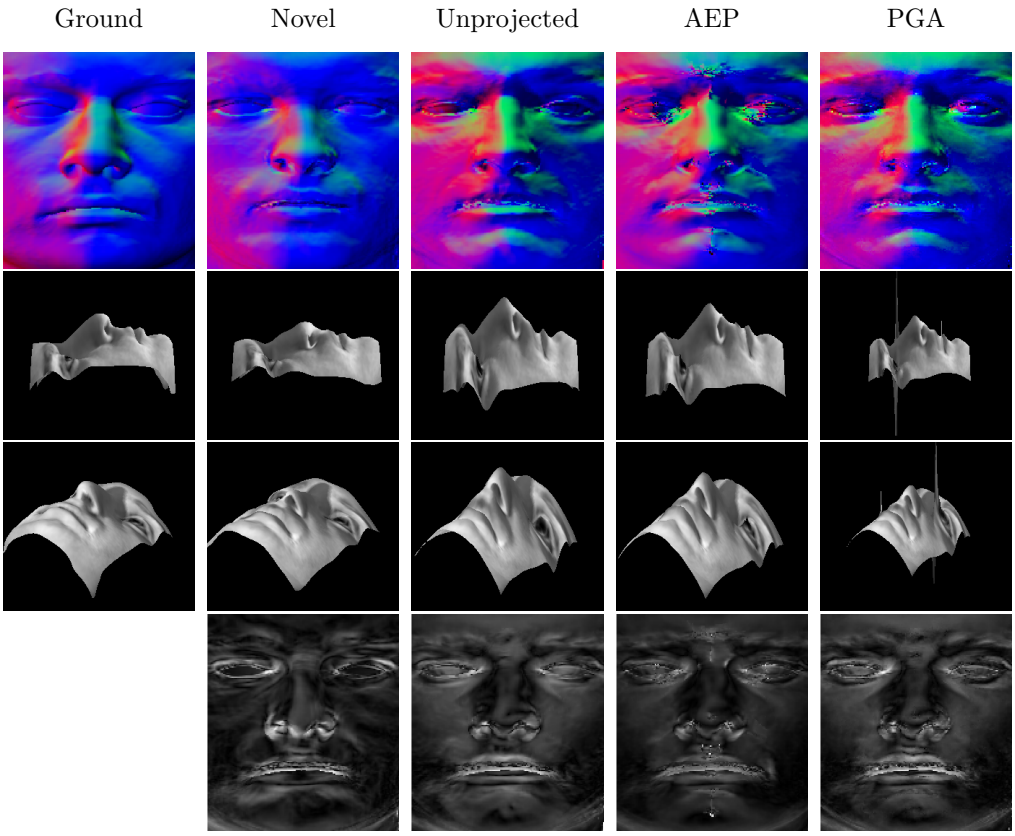


Figure 4.15: The first row is the normal map. The second row is a profile rendering. The third row is a rendering from below the chin. The final row is the angular error between the reconstruction and the ground truth.

Chapter 5

Evaluation

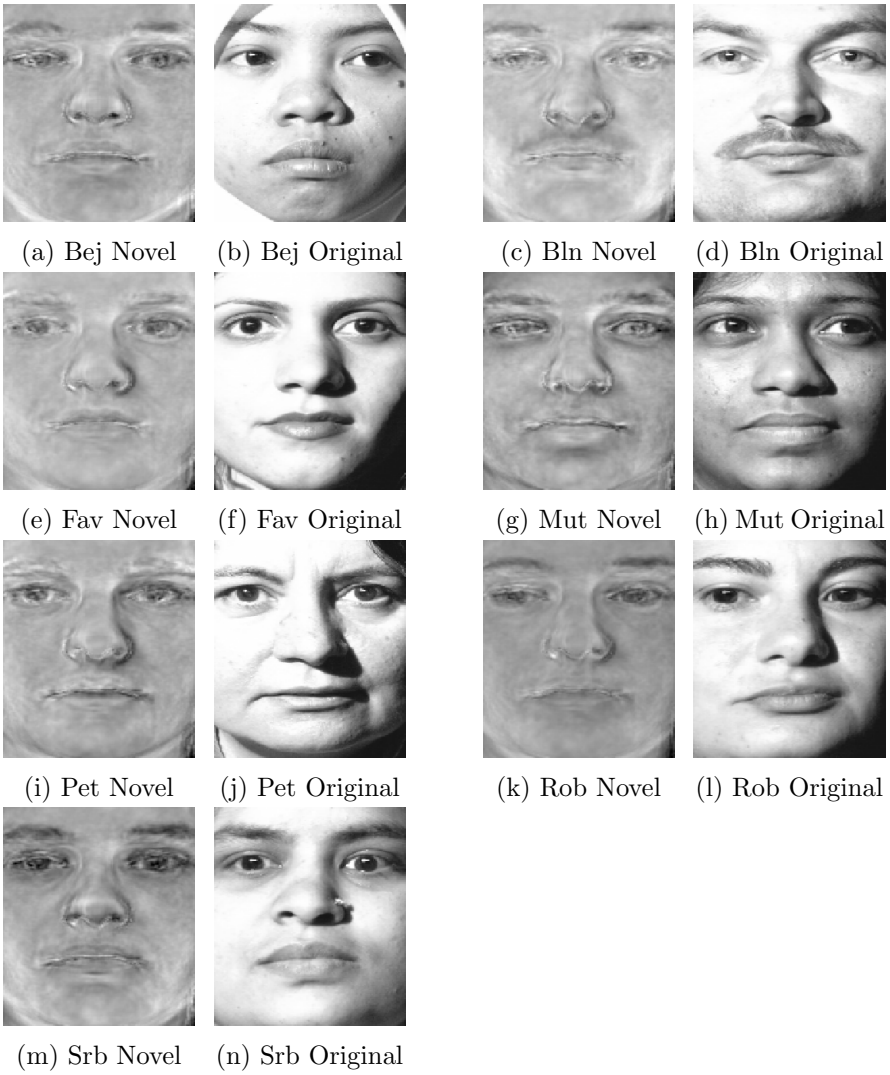
In this Chapter we evaluate the results provided in Chapter 4. We begin by presenting the textures reconstructed from each of the subjects considered. The novel reconstruction presented in Chapter 3 gives coefficients for reconstructing the texture as well as the shape. We therefore show the textures from which the normals were generated. We go on to present some analysis on the angular error calculated between the reconstructions and the ground truths. Finally, we compare and contrast the techniques performed by picking out specific reconstructions and evaluating them. The majority of the analysis presented below is qualitative and involves comparing the shapes of facial features, such as the mouth and nose.

5.1 Reconstructed Textures

In this Section we present the reconstructed textures from our novel technique. The overall results in terms of the texture match were acceptable. It should be noted, however, that the texture reconstructions were particularly poor when attempting to reconstruct from images that were in the training set. In this case the technique should produce an exact match, but it does not.

5.1.1 Photometric Stereo Database Reconstructed Textures

In Figure 5.1 we present the reconstructed textures for each of the photometric stereo images used in Section 4.3.



(a) Bej Novel (b) Bej Original (c) Bln Novel (d) Bln Original

(e) Fav Novel (f) Fav Original (g) Mut Novel (h) Mut Original

(i) Pet Novel (j) Pet Original (k) Rob Novel (l) Rob Original

(m) Srb Novel (n) Srb Original

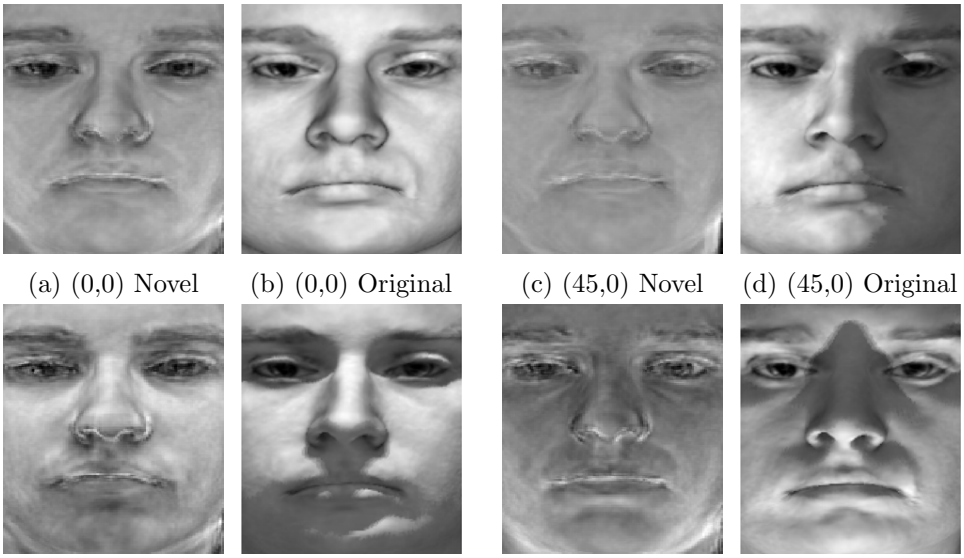
Figure 5.1: The reconstructed textures for each of the faces presented in Section 4.3.

Figure 5.1 clearly shows that the texture recovered has a resemblance to the original texture. Figure 5.1c is a particularly accurate reconstruction as we even have some shading on the top lip that is from the moustache. The images do not exactly match the textures, however, due to their reliance on reconstructing the normals. Since the reconstruction of the texture coefficients is reliant on the accuracy of the normals we are restricted as to how accurate the texture reconstruction will ever be. However, due to the fact that we are concentrating on recovering the surface shape and not the texture, we can simply use the original texture with our recovered shape. For textures such as Figure 5.1b our reconstruction is actually poor. We

haven't managed to capture the headscarf at all and thus, the texture quite stongly resembles the mean texture.

5.1.2 Morphable Model Reconstructed Textures

In Figure 5.2 we present the textures reconstructed from the morphable model under unique lighting conditions.



(a) (0,0) Novel (b) (0,0) Original (c) (45,0) Novel (d) (45,0) Original

(e) (0,45) Novel (f) (0,45) Original (g) (0,-45) Novel (h) (0,-45) Original

Figure 5.2: The reconstructed textures for each of the faces presented in Section 4.4. (0,0) indicates spherical coordinates $\theta = 0$, $\phi = 0$

In the case of Figure 5.2a above we have an extremely poor reconstruction. Given that this face was constructed using the morphable model we should be able to reproduce this texture exactly. This sample face is in the eigenspace of the original training set and thus, the textures should be identical. However, the normals are not constructed as accurately as would be expected and so, the texture is never truly accurate either.

The varied lighting conditions do not seem to have had a particularly large impact on the reconstructive power of the model. Whilst Figure 5.2c does lack the proper shading, the overall structure of the face still appears to be accurate.

5.2 Shape Reconstruction

In this Section we present the reconstructed shapes from each of the techniques performed. We also compare the total angular error and height difference between the different reconstructions.

5.2.1 Angular Error

In each of the angular error images given in Sections 4.3 and 4.4, every pixel represents the deviation between the ground truth normals and the reconstruction. In order to produce a quantitative measure to compare how accurate the techniques were in comparison to one another, we summed every pixel. Whilst some areas, shown as white in the images, show large deviations in the normals, the sum of the normals gives a better overall indication as to the accuracy of the reconstruction.

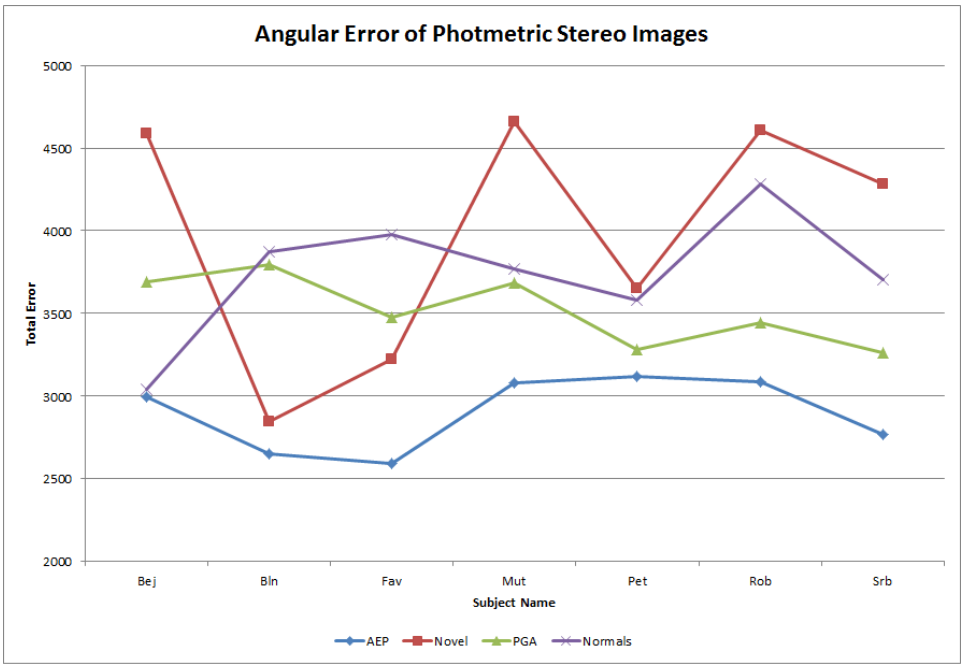


Figure 5.3: Each image in the photometric dataset is analysed. There are four series that represent each of the reconstruction techniques respectively. The series labelled normals represents Smith's reconstruction without a projection applied to the normals.

Figure 5.3 shows the total error for each of the seven images reconstructed in Section 4.3. It clearly shows that the azimuthal equidistant projection yields the most accurate reconstructions. It provided the least total error across all seven of the images tested. Principal geodesic analysis was very consistent across all the images tested and was more accurate than the unprojected normals in the majority of the reconstructions. Unsurprisingly, our novel reconstruction and Smith's unprojected reconstruction were consistently worse than the projected techniques. This supports the

evidence that performing PCA on directed data yields inaccurate principal components.

The data above is misleading, however, as from visual inspection we can see that the recovered height maps often have large artifacts in the AEP reconstruction. This is particularly noticeable in the reconstruction of Fav in Figure 4.3c. Here we can clearly see that the AEP reconstruction has produced large artifacts due to the fact that the normals recovered are poor. In Figure 5.4, we can see the reason for the poor reconstruction. The area highlighted by the red rectangle shows the misaligned normals. It also clearly demonstrates that some of the normals are pointing in the incorrect direction, which will cause artefacts after integration.

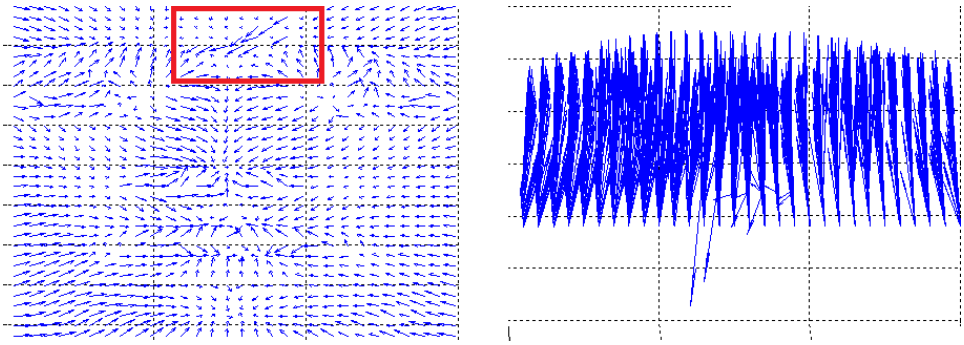


Figure 5.4: Demonstrates the incorrect normals that cause the artefacts seen in the AEP reconstruction for Fav.

In fact, given another of the Fav images, we can see that Smith's method of rotating back on to the reflectance cone yields very poor results. This is largely to do with the two black patches at the bottom corners of the image, which are in shadow. Smith's rotation method causes a global error that propagates and causes a very poor reconstruction. The dark patches are shown in Figure 5.5.

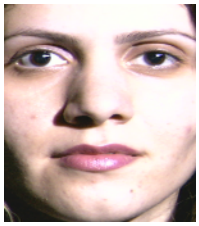


Figure 5.5: The black patches at the corner of this figure causes global errors in Smith's reconstructions.

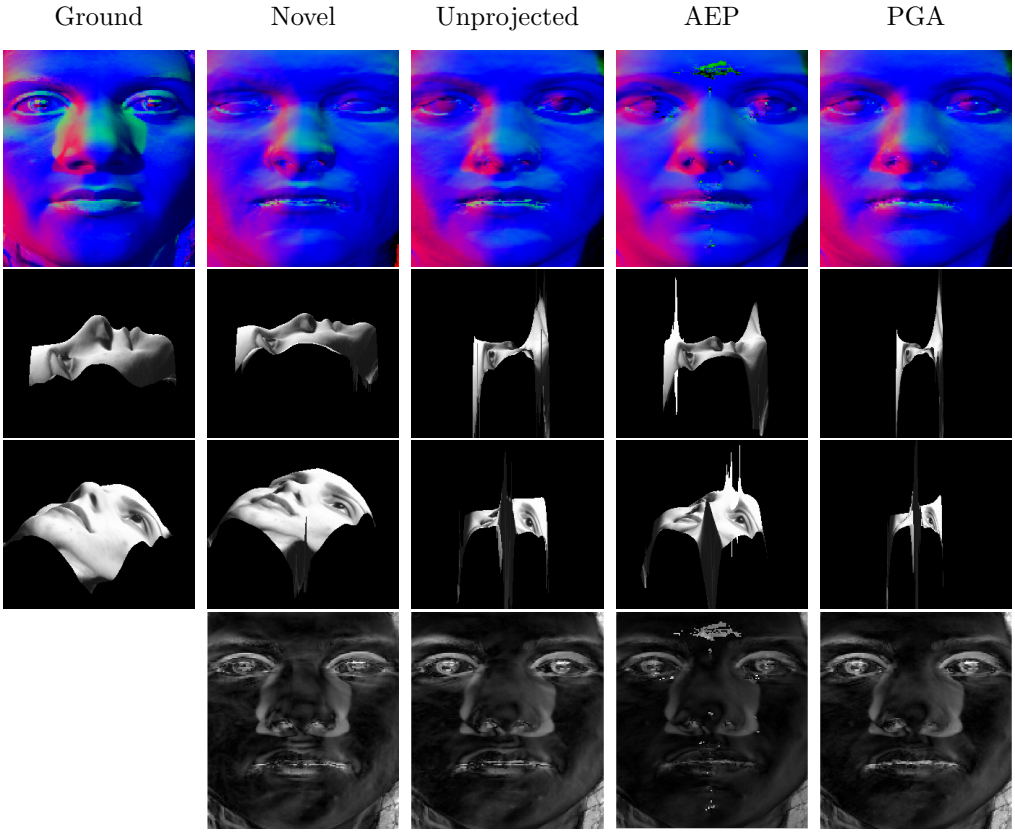


Figure 5.6: The first row is the normal map. The second row is a profile rendering. The third row is a rendering from below the chin. The final row is the angular error between the reconstruction and the ground truth.

5.2.2 Height Difference

A more accurate measure of the quality of the reconstruction may be the difference between the ground truth and reconstruction height maps. In Figure 5.7, we can clearly see that the AEP reconstruction deviates a lot more than suggested by Figure 5.3. However, despite the AEP reconstruction suffering from the impact of a few particularly bad normals, it does minimise the angular error. The reason for the AEP reconstruction causing such poor normals is likely due to the structure of the projection. The azimuthal equidistant projection relies on calculating both θ and ϕ using trigonometric techniques. These techniques can propagate error due to floating point errors and because of the periodic nature of both the *sine* and *cosine* functions. In order to correctly calculate the projection it was necessary to restrict the range of θ between $[-\pi/2, \pi/2]$.

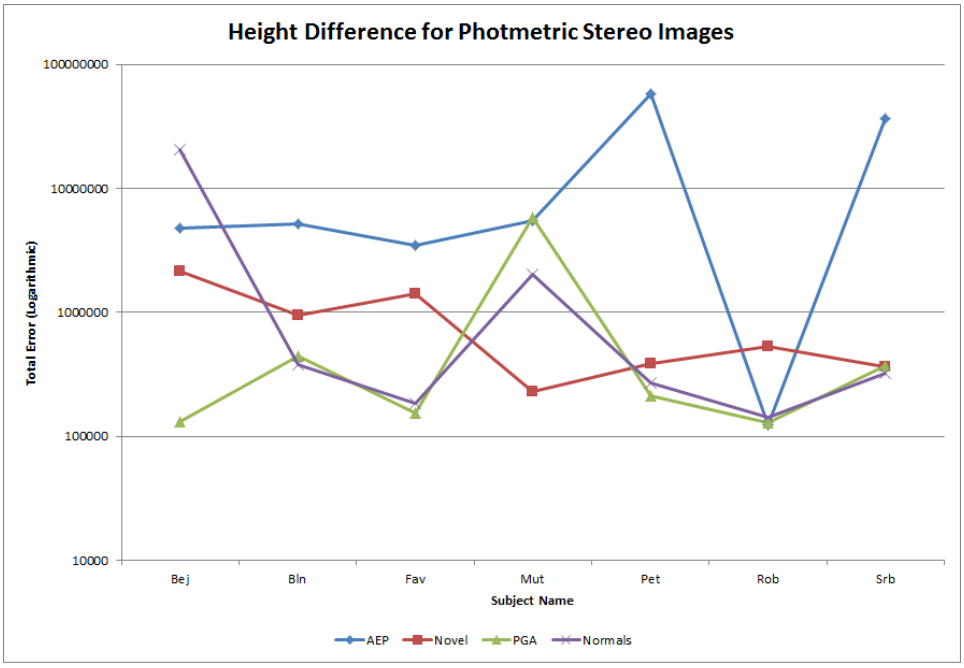


Figure 5.7: The absolute difference between each reconstruction technique and the ground truth. There are four series that represent each of the reconstruction techniques respectively. The series labelled normals represents Smith's reconstruction without a projection applied to the normals.

We also provide the angular error and height difference metrics for the morphable model reconstructions given in Section 4.4. Figure 5.8 gives the height difference and Figure 5.9 gives the angular error. In this case the height difference was fairly even across all the reconstruction techniques, except for PGA. PGA seems particurly poor at reconstructing the images lit from above and below. Smith does mention that the global nature of his reconstruction technique can cause issues under poor lighting conditions. In fact in [23] he suggests using a robust statistic in order to minimize the affect of shadows and low albedo regions.

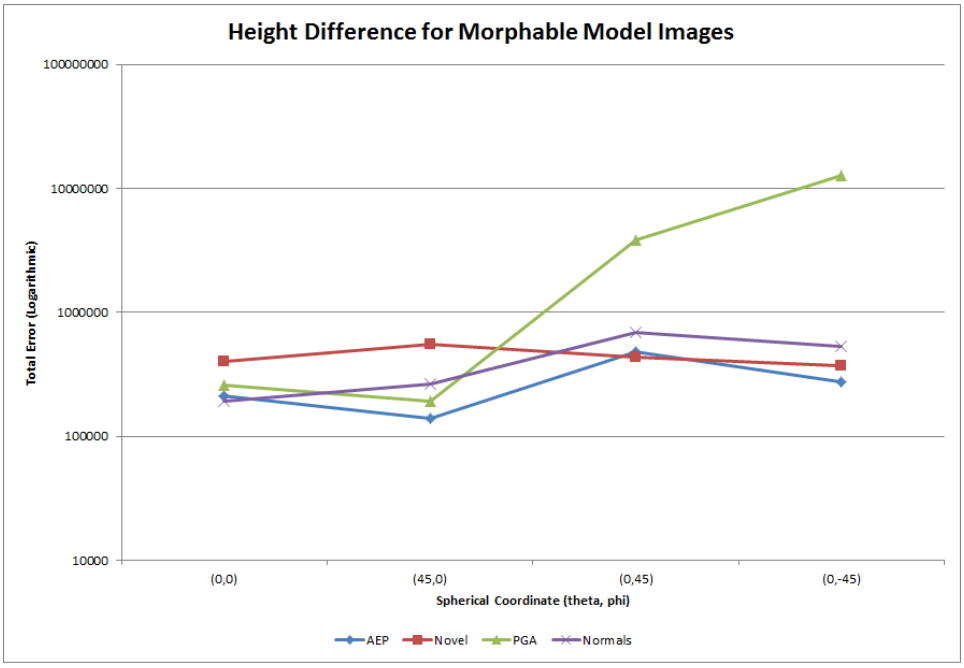


Figure 5.8: The absolute difference between each reconstruction technique and the ground truth. There are four series that represent each of the reconstruction techniques respectively. The series labelled normals represents Smith's reconstruction without a projection applied to the normals.

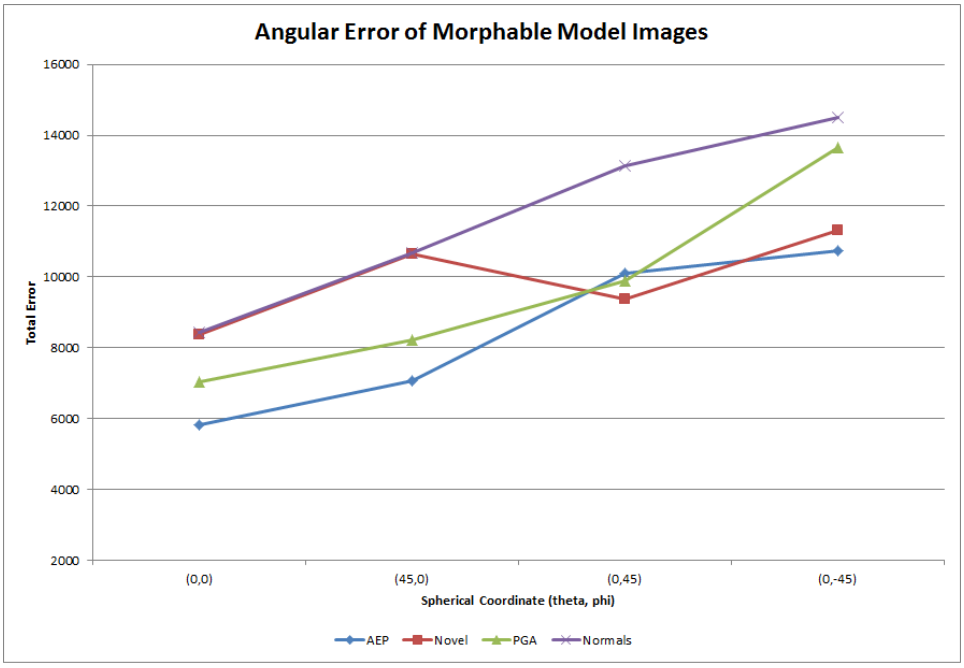


Figure 5.9: Each image in the morphable model reconstruction is analysed. There are four series that represent each of the reconstruction techniques respectively. The series labelled normals represents Smith's reconstruction without a projection applied to the normals.

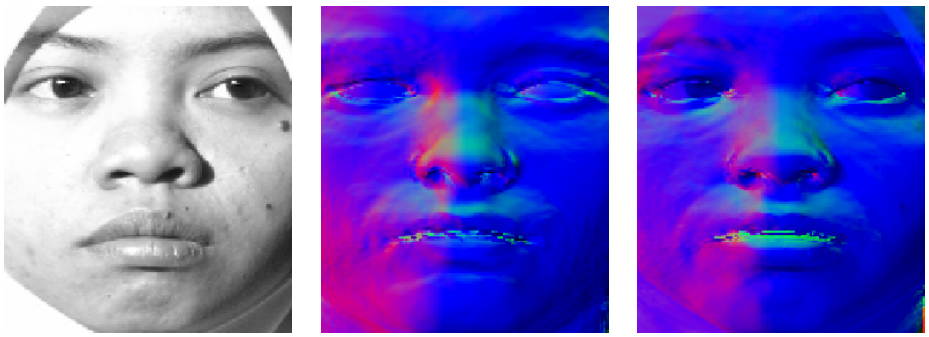
The novel reconstruction does reasonably well under these conditions, likely due to the fact that it is reconstrucing from it's own texture and not the original texture. Therefore, these areas of self shadow do not appear on the reconstructed texture and so, do not affect the reconstruction of the normals.

5.3 A Visual Comparison

One of the major differences between the novel reconstruction method and Smith's reconstruction method is the shape of the normal map. The first row in each of the sets of images provided in Sections 4.3 and 4.4 shows this difference. The colour scheme is created by mapping each component of a normal (x, y, z) to a colour in RGB space, (r, g, b). Therefore, accurate reconstructions should have similarly coloured regions, as this would suggest that the normals are pointing in a similar direction. Figure 5.10 clearly shows that Smith's normal maps look more like the reference face than that of the novel reconstruction. This has to do with the on cone rotation

performed in Smith's reconstruction. Since Smith looks to make his normals closely match the orientation of the reflectivity cone, he achieves a normal map that more accurately mimics the face. However, in areas with specular highlights or low albedo, the reflectivity cone calculated will be inaccurate. This will cause the on cone rotation to produce poor normals that cause issues during surface reconstruction.

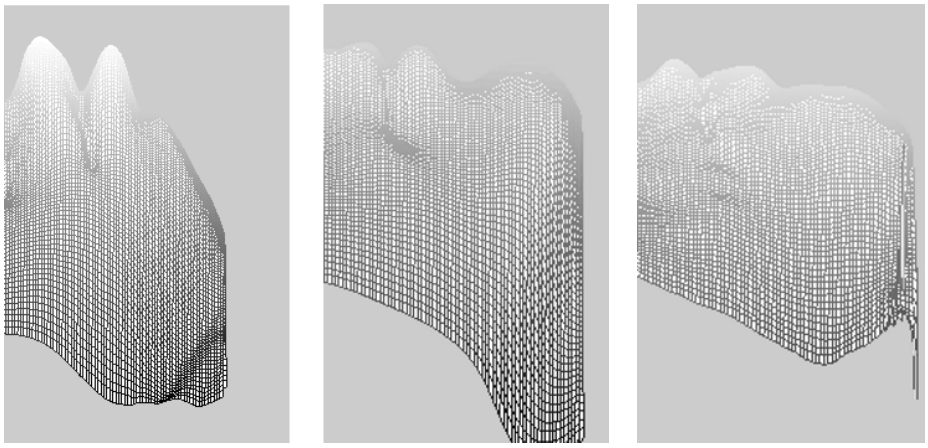
When two normals in close proximity do not agree on the surface direction, integrating them does not yield a smooth surface. This is most notable in Smith's reconstructions of Pet, shown in Figure 4.8, and Srb, shown in Figure 4.10. The novel reconstruction is less constrained and is more likely to generate a face like surface. Unfortunately, this means that the novel implementation often generates faces that don't appear to deviate from the mean by very much. This is due to the fact that if the texture it reconstructs is not accurate, then the normals won't be either. This is particularly apparent in Subsection 4.4.4 where the morphable model is lit from below. Here the novel reconstruction doesn't appear to have matched the sample at all. The nose and the mouth do not match the shape of the ground truth and the nose in particular looks to be very poor. In comparison to Smith's reconstructions, however, it could be argued that our novel implementation is more face like.



(a) Sample (b) Novel (c) Smith

Figure 5.10: The novel reconstruction compared against Smith's reconstruction. Smith's reconstruction more accurately matches the original texture.

The novel reconstruction of Bej shows that it does not always reproduce the mean face. The chin from the ground truth of Bej is very flat, much flatter than the mean shape of the training set. In Figure 5.11 we show this difference between the ground truth, the mean shape and the novel reconstruction. We can clearly see that the novel method has attempted to reconstruct the chin of Bej, despite it deviating highly from the training set.

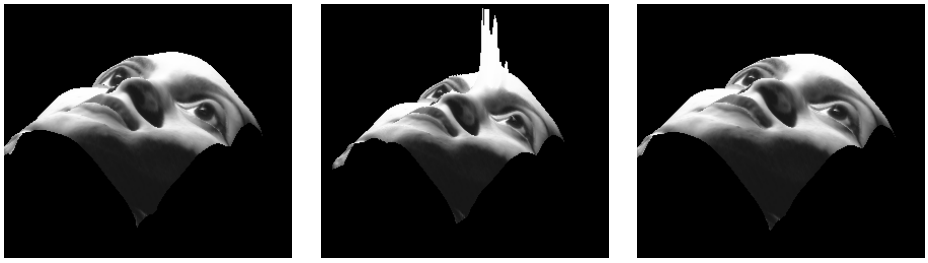


(a) Bej (b) Mean (c) Novel

Figure 5.11: Illustrating the attempt of the novel technique to reconstruct subject Bej's chin. From left to right we have the ground truth of Bej's chin, the mean chin of the training set and the novel reconstruction of Bej's chin.

5.3.1 A Comparison of Projected and Unprojected Normals

Here we visually compare the results gained from performing the azimuthal equidistant projection and principal geodesic analysis against Smith's technique on the unprojected normals. Unfortunately, we are unable to compare against the novel technique as the novel technique was only performed directly upon the normals. Despite the fact that the angular error appears to be minimized by using the projections, the reconstructions don't appear to agree. Visually, they are very similar when comparing between AEP, PGA and the unprojected data. This is likely due to the fact that they are all constrained by the hard irradiance constraint. The similarity between the three techniques is easily demonstrated by a side-by-side comparison of Fav's reconstruction. In Figure 5.12 we can see that all three of the reconstructions are very similar.



(a) Unprojected (b) AEP (c) PGA

Figure 5.12: All three reconstructions are visually very similar, despite the angular error being minimized in the AEP reconstruction.

The similarity evident in Figure 5.12 suggests that the projections may be unnecessary. The quality of the unprojected reconstruction is quite surprising when compared to the projected results. If the visual quality of the result is more important than the surface accuracy, the unprojected reconstruction may be sufficient.

Chapter 6

Conclusion

We have presented a novel implementation of shape from shading that borrows from the morphable model construction. We have shown that it is capable of reconstructing a variety of faces, including under challenging lighting conditions. We have also demonstrated the expressive power of the our training set and its ability to recover both shape and texture coefficients that form a reasonable appoximation of the input image. We also demonstrated the stability of our algorithm in comparison to William Smith's algorithm, particularly under difficult lighting conditions. Accuracy of the algorithm could likely be improved by performing an optical flow algorithm as described in Blanz and Vetter's work[4]. The low number of feature points chosen for image alignment, coupled with the simple affine transformation performed, meant that unique face shapes did not match the training set particularly well.

We also presented a set of Matlab functions that enable both azimuthal equidistant projection and principal geodesic analysis to be performed. These functions are useful in any domain that wishes to build a statistical model on directed data. We also present Matlab code to perform William Smith's reconstruction as described in his papers [22, 23]. In providing an implementation of Smith's work we also made a comparison between the reconstructive power of projected versus unprojected normals. We found that projected normals outperform unprojected normals in minimizing the angular error but tended to be less stable. This is likely due to compounding of errors in taking *sine* and *cosine* of floating point numbers. Visually, however, the projected and unprojected reconstructions are difficult to tell apart. It could, therefore, be argued that performing the projection is unnecessary, if the visual quality of the shape is all that mattered. Upon reflection, comparison between the reconstructions would have been enhanced by the addition of a projection operator to the novel technique. Without this second method to compare against, it is difficult to judge the reconstructive power of projected normals against unprojected normals.

The project has been largely successful in that two reconstruction methods were produced and it was possible to compare the different projection methods. In terms of reconstructive power, Smith's method still appears to be superior to the novel method. However, the novel method does reconstruct visually pleasing faces from input images. It is also reasonably similar to Smith's method in terms of overall error, particuarly the difference in height maps. This is largely due to the stability of the novel implementation in comparison to Smith's reconstruction.

6.1 Future Work

There are a number of extensions that could potentially improve the results that were generated in this investigation. However, the underlying framework provided would give a solid base upon which to implement the improvements.

- In order to give a more thorough comparison of projected normals against unprojected normals, it would be interesting to add a projection function to the novel reconstruction. This would require reformulating the update loop to take into account some projection function when reconstructing the normals. The texture coefficient would remain unchanged. This comparison would help confirm the validity of projecting the directed data before performing PCA.

- The novel reconstruction algorithm presented relies on a least squares distance to calculate the error. This form of cost function is not robust and better results would be provided if it was replaced with a robust cost function. A robust function is stable in the presence of outliers. This is particularly applicable to the domain of facial reconstruction as faces are prone to areas of poor lighting or self-cast shadows. In these areas the normals produced will be noisy and distort the shape of the reconstructed face. If we replaced the least squares distance with an M-estimator (Maximum likelihood estimator) we would achieve better results in these poorly lit regions. An M-estimator replaces the squared residuals by a kernel function that limits the effects of large residuals on the result.

- The current techniques investigated, as well as the novel reconstruction, only work on front-facing images. It would be a useful extension of this work to attempt to add a model of motion to the reconstruction. This would allow pose invariant reconstructions and is particularly useful in the area of unconstrained image reconstructions. Given the large number of unconstrained photosets that exist, this could be very useful in areas such as face recognition.

- Lambertian reflectance is asssumed by the techniques investigated in this project. Lambertian reflectance, however, is not a particularly accurate lighting model for faces. Areas of low-albedo around the eyes and mouth, facial hair for example, do not exhibit Lambertian reflectance. It would be a worthy extension of our model to investigate the feasability of using spherical harmonics in order to model the lighting conditions. Previous work by Ahmed and Farag[1] have investigated using statistical models involving spherical harmonics. However, they investigate the use of height maps and not the surface normals themselves. It would, therefore, be novel work to investigate the use of spherical harmonics with surface normals.

Bibliography

[1] Ahmed, A., and A. Farag (2007), A new statistical model combining shape and spherical harmonics illumination for face reconstruction, in *Advances in Visual Computing*, *Lecture Notes in Computer Science*, vol. 4841, edited by G. Bebis, R. Boyle, B. Parvin, D. Koracin, N. Paragios, S.-M. Tanveer, T. Ju, Z. Liu, S. Coquillart, C. Cruz-Neira, T. Mller, and T. Malzbender, pp. 531–541, Springer Berlin / Heidelberg.

[2] Barsky, S., and M. Petrou (2003), The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *25*(10), 1239 – 1252, doi:10.1109/TPAMI.2003.1233898.

[3] Bell, C. (1961), Reduction of speech spectra by analysis-by-synthesis techniques, *Acoustical Society of America Journal*, *33*, 1725, doi:10.1121/1.1908556.

[4] Blanz, V., and T. Vetter (1999), A morphable model for the synthesis of 3d faces, in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '99, pp. 187–194, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, doi:http://dx.doi.org/10.1145/311535.311556.

[5] Fletcher, P., C. Lu, S. Pizer, and S. Joshi (2004), Principal geodesic analysis for the study of nonlinear statistics of shape, *Medical Imaging, IEEE Transactions on*, *23*(8), 995 –1005, doi:10.1109/TMI.2004.831793.

[6] Frankot, R., and R. Chellappa (1988), A method for enforcing integrability in shape from shading algorithms, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *10*(4), 439 –451, doi:10.1109/34.3909.

[7] Golub, G., and C. Reinsch (1970), Singular value decomposition and least squares solutions, *Numerische Mathematik*, *14*, 403–420, 10.1007/BF02163027.

[8] "Horn, B. K. P. ("1989"), *"Shape from shading"*, "The MIT Press", "Cambridge Mass. [etc.]".

[9] Horn, B. K. P. (1990), Height and gradient from shading, *International Journal of Computer Vision*, *5*(1), 37–75.

[10] Horn, B. K. P., and M. J. Brooks (1986), The variational approach to shape from shading, *Computer Vision, Graphics, and Image Processing*, *33*(2), 174–208.

[11] Huang, J., B. Heisele, and V. Blanz (2003), Component-based face recognition with 3d morphable models, in *Audio- and Video-Based Biometric Person Authentication*, *Lecture Notes in Computer Science*, vol. 2688, edited by J. Kittler and M. Nixon, pp. 1055–1055, Springer Berlin, Heidelberg.

[12] Kemelmacher-Shlizerman, I., and R. Basri (2011), 3d face reconstruction from a single image using a single reference face shape, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *33*(2), 394 –405, doi:10.1109/TPAMI.2010.63.

[13] Kemelmacher-Shlizerman, I., and S. Seitz (2011), Face reconstruction in the wild, in *Computer Vision (ICCV), 2011 IEEE International Conference on*, pp. 1746 –1753, doi:10.1109/ICCV.2011.6126439.

[14] Lee, Y., D. Terzopoulos, and K. Waters (1995), Realistic modeling for facial animation, in *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, SIGGRAPH '95, pp. 55–62, ACM, New York, NY, USA, doi:http://doi.acm.org/10.1145/218380.218407.

[15] North Coleman, E., Jr., and R. Jain (1992), Shape recovery, chap. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry, pp. 180–199, Jones and Bartlett Publishers, Inc., USA.

[16] Pantic, M., and L. Rothkrantz (2000), Automatic analysis of facial expressions: the state of the art, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *22*(12), 1424 –1445, doi:10.1109/34.895976.

[17] Pantic, M., and L. Rothkrantz (2004), Facial action recognition for facial expression analysis from static face images, *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, *34*(3), 1449 –1461, doi:10.1109/TSMCB.2004.825931.

[18] Parke, F. I. (1972), Computer generated animation of faces, in *in ACM National Conference*, pp. 451–457.

[19] Patel, A., and W. Smith (2010), Exploring the identity manifold: Constrained operations in face space, in *Computer Vision ECCV 2010*, *Lecture Notes in Computer Science*, vol. 6316, edited by K. Daniilidis, P. Maragos, and N. Paragios, pp. 112–125, Springer Berlin, Heidelberg.

[20] Paysan, P., R. Knothe, B. Amberg, S. Romdhani, and T. Vetter (2009), A 3d face model for pose and illumination invariant face recognition, in *Proceedings of the 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, AVSS 2009, pp. 296–301, IEEE Computer Society, Washington, DC, USA.

[21] Pennec, X., X. Pennec, and P. Epidaure (2004), Probabilities and statistics on riemannian manifolds: A geometric approach, in *In IEEE Workshop on Nonlinear Signal and Image Processing*, pp. 194–198.

[22] Smith, W., and E. Hancock (2006), Recovering facial shape using a statistical model of surface normal direction, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *28*(12), 1914 –1930, doi: 10.1109/TPAMI.2006.251.

[23] Smith, W. A. P., and E. R. Hancock (2007), Facial shape-from-shading using principal geodesic analysis and robust statistics, in *Proceedings of the 12th IMA international conference on Mathematics of surfaces XII*, pp. 412–426, Springer-Verlag, Berlin, Heidelberg.

[24] Vetter, T., and V. Blanz (1998), Estimating coloured 3d face models from single images: An example based approach, in *In Proceedings, European Conference on Computer Vision*, pp. 499–513, Springer.

[25] Woodham, R. J. (1980), Photometric method for determining surface orientation from multiple images, *Optical Engineering*, *19*(1), 139–144.

[26] Worthington, P., and E. Hancock (1999), New constraints on data-closeness and needle map consistency for shape-from-shading, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *21*(12), 1250 –1267, doi:10.1109/34.817406.

[27] Zhang, L., S. Wang, and D. Samaras (2005), Face synthesis and recognition from a single image under arbitrary unknown lighting using a spherical harmonic basis morphable model, in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, pp. 209 – 216 vol. 2, doi:10.1109/CVPR.2005.152.

[28] Zhang, R., P.-S. Tsai, J. Cryer, and M. Shah (1999), Shape-from-shading: a survey, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *21*(8), 690 –706, doi:10.1109/34.784284.

[29] Zhao, W. Y., and R. Chellappa (2000), Illumination-insensitive face recognition using symmetric shape-from-shading, in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, vol. 1, pp. 286 –293 vol.1, doi:10.1109/CVPR.2000.855831.