

Estimation of Facial Angular Information Using a Complex-Number-Based Statistical Model

Mario Castelan* and Edwin R. Hancock

Department of Computer Science, University of York, York YO1 5DD, UK

Abstract. In this paper we explore the use of complex numbers as means of representing angular statistics for surface normal data. Our aim is to use the representation to construct a statistical model that can be used to describe the variations in fields of surface normals. We focus on the problem of representing facial shape. The fields of surface normals used to train the model are furnished by range images. We compare the complex representation with one based on angles, and demonstrate the advantages of the new method. Once trained, we illustrate how the model can be fitted to brightness images by searching for the set of parameters that both satisfy Lambert's law and minimize the integrability error.

1 Introduction

The problem of acquiring surface models of faces is an important one with potentially significant applications in biometrics, computer games and production graphics. There are many ways in which surface models can be acquired, and these include the use of range-scanners, stereoscopic cameras and structured light sensors. However, one of the most appealing methods is to use shape-from-shading (SFS), since this is a non-invasive process which mimics the capabilities of the human vision system. Shape-from-shading aims to recover surface orientation, and hence surface height by solving the image radiance equation. In general, though, SFS is an under-constrained problem since the two degrees of freedom for surface orientation (slant and tilt), must be recovered from a single measured intensity value. In contrast to the human visual system[3], it seems that computer vision systems encounter more difficulty in estimating the tilt of a surface from a single image than its slant (see Figure 1).

One way to overcome the problems with general purpose SFS is to draw on a domain specific model that can be used to constrain the directions of the surface normals. This approach has proved to be particularly effective in the analysis of faces. Attick et al.[1] were the among the first to build 3D statistical shape models of faces for use in conjunction with SFS. Working with cylindrical coordinates they develop an eigenmode model (referred to as eigenheads) for surface height. In fitting the model to data, they impose an image irradiance constraint using the shape coefficients of the model. Later, Vetter et al.[9] decoupled the effects of

* Supported by National Council of Science and Technology (CONACYT), Mexico, under grant No. 141485.

texture and shape on facial appearance. Assuming that full facial correspondence information is to hand, they perform PCA separately on the texture and shape components. Thus they develop a statistical model that can be fitted to image brightness data. These methods deliver accurate and photo-realistic results, but at the expense of considerable computational overheads and simplicity of implementation. This is largely due to the brute force search method used to adjust the model parameters to fit the input image data. Recently, Dogvard and Basri[2] have combined statistical models with symmetry constraints. The model relies on a Cartesian representation of the surface height. Surface gradient is expressed in terms of a set of deformation coefficients, and this allows the symmetric SFS equation to be transformed into a linear system of equations. The linear system can be efficiently solved and used to estimate surface height. Their results show that accuracy is sacrificed for a gain in computational efficiency.

The aim in this paper is to explore whether angular representations can be used to construct statistical models that can be used in conjunction with shape-from-shading. This is a natural approach since it is surface orientation and not surface height that is responsible to the perceived image brightness. Angular data is more difficult to model than Cartesian data since angles wrap around. Hence, small differences in distance on a sphere can correspond to large differences in angles. The classical example here is a short walk across one of the poles of a sphere, when large differences in longitude correspond to small differences travelled. In shape from shading the surface normal is determined by the azimuth and zenith angles. When the surface is illuminated in the direction of the viewer and if the surface reflectance is Lambertian, then the arc-cosine of the zenith angle is determined by the normalized image brightness. The azimuth angle, on the other hand, must be determined using additional constraints provided by smoothness or the occluding boundary. The aim in this paper is to develop a statistical model that can be used to model the distribution of azimuthal direction in faces. To overcome the problems with the representation of angular data, we use complex numbers. Our idea is to encode the azimuth angles as complex numbers and to capture their distribution by adapting the Sirovich snapshot method to deal with complex eigenvectors. We show how the model can be trained using range images and fitted to brightness images using constraints on surface normal direction provided by Lambert's law and surface integrability.

2 Statistical Information for Angular Data

In this paper, we aim to construct a statistical model for the angular variation in surface normal direction. We train our model on surface normals extracted from range images. We aim to recover surface normals from image brightness data by fitting the model to facial images using constraints provided by Lambert's law and integrability.

To construct the statistical model, we represent the surface normal data extracted from the range data as long-vectors. The range data is vectorized by stacking the image columns. If the range images contain M columns and N rows,

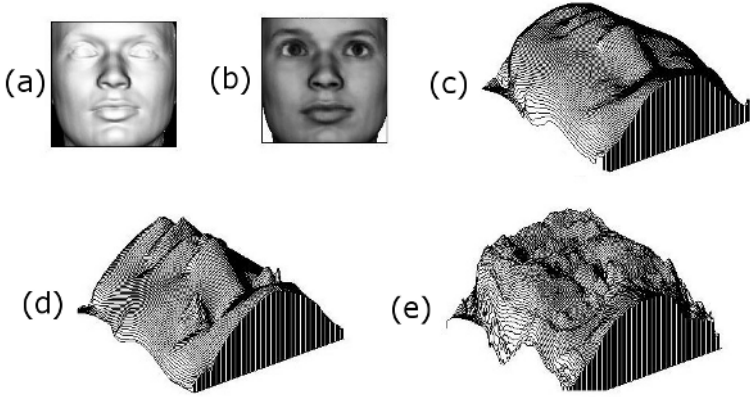


Fig. 1. Effect of incorrectly calculated azimuth and zenith angle in face shape recovery: (a) orthogonal Lambertian (constant albedo) image, (b) true irradiance (non-constant albedo) image, (c) ground-truth surface, (d) surface preserving true azimuth angle but with its zenith angle estimated through SFS and (e) surface preserving true zenith angle, but with its azimuth angle estimated through SFS. Note how the effect of wrongly estimated tilt angle cause a severe deterioration on the recovered surface.

then the pixel with column index j_c and row index j_r corresponds to the element indexed $j = (N-1)j_c + j_r$ of the long-vector. Let $n_j^k = \frac{1}{\sqrt{(p_j^k)^2 + (q_j^k)^2 + 1}}(p_j^k, q_j^k, 1)^T$ be the surface normal at the pixel indexed j of the k th training image. Here $p = \frac{\partial h}{\partial x}$ and $q = \frac{\partial h}{\partial y}$ are the partial derivatives of the surface height h in the x and y directions. The zenith and azimuth angles of the surface normal are respectively $\theta_j^k = \arctan \sqrt{(p_j^k)^2 + (q_j^k)^2}$ and $\phi_j^k = \arctan \frac{q_j^k}{p_j^k}$. Here we use the four quadrant arc-tangent function and therefore $-\pi \leq \phi_j^k \leq \pi$. Unfortunately the angles can not be used to construct statistical models. The reason for this is that statistical calculations performed on angular data can be biased by the angle *cut* point (see Figure 2). To illustrate this problem consider two points on a unit circle placed just above and just below the cut-line. Although the two points are close to one another on the unit circle, when the difference in angles is computed then this may be close to 2π .

Our idea in this paper is to overcome this problem by working with a complex number representation of the azimuth angles of the surface normal. We encode the azimuth angle using the complex number

$$z_j^k = \exp(i\phi_j^k) = \cos \phi_j^k + i \sin \phi_j^k, \quad (1)$$

where $i = \sqrt{-1}$. The azimuth angle is hence given by the real (Re) and imaginary (Im) components of the complex number, i.e.

$$\phi_j^k = \arctan \frac{\text{Im } z_j^k}{\text{Re } z_j^k}. \quad (2)$$

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

$$\hat{z}_j = \frac{1}{K} \sum_{k=1}^K z_j^k. \tag{3}$$

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

$$\hat{\phi}_j = \arctan \frac{\text{Im } \hat{z}_j}{\text{Re } \hat{z}_j} \quad \text{and} \quad \hat{r}_j = \sqrt{(\text{Im } \hat{z}_j)^2 + (\text{Re } \hat{z}_j)^2}. \tag{4}$$

Note that the cartesian coordinates of the points of \hat{z}_j on the complex plane are defined by the average of the cosines (x-axis) and sines (y-axis) of all of the observations ϕ_j^k of the training set, therefore

$$\text{Re } \hat{z}_j = \hat{r}_j \cos \hat{\phi}_j = \frac{1}{K} \sum_{k=1}^K \cos \phi_j^k \quad \text{and} \quad \text{Im } \hat{z}_j = \hat{r}_j \sin \hat{\phi}_j = \frac{1}{K} \sum_{k=1}^K \sin \phi_j^k. \tag{5}$$

Unfortunately, although this allows us to overcome the problems of representing the azimuth angle statistics, it yields complex numbers that no longer have unit modulus. In fact r_j can fluctuate between 0 and 1. However, r_j is an important measure of the concentration of the azimuth angles in the training data. If the directions of the azimuth angles in the training set are strongly clustered, then r_j will tend to be 1. If, on the other hand, they are scattered then r_j will tend to 0.

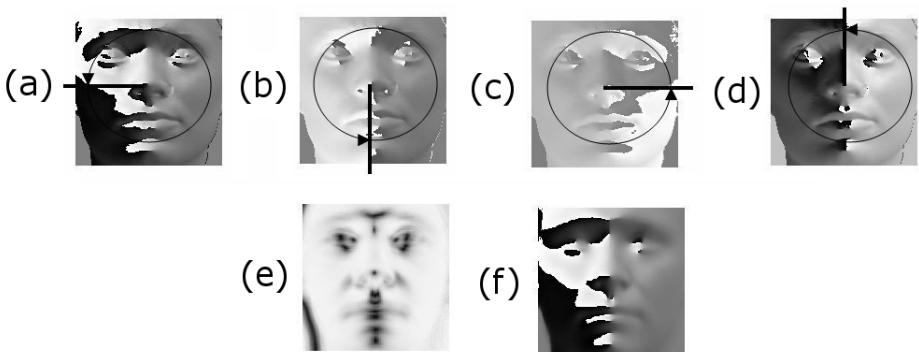


Fig. 2. In the top row, different arguments for one training set example z^k are shown as intensity maps. From left to right, $(-\pi, \pi]$ (a), $(-\frac{\pi}{2}, \frac{3\pi}{2}]$ (b), $(0, 2\pi]$ (c) and $(-\frac{3\pi}{2}, \frac{\pi}{2}]$ (d). The mean direction $\hat{\phi}$ (e) and the mean resultant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

Although the mean resultant length \hat{r}_j is a very important measure of dispersion, for purposes of comparison with data on the line we should consider measures of dispersion based on circular data, like the sample circular variance $v_j = 1 - \hat{r}_j$, $0 \leq v_j \leq 1$. Following [5], if $1 - \cos(\alpha_1 - \alpha_2)$ is a measure of distance between two angles α_1 and α_2 , then the dispersion of the angles $\phi_j^1, \phi_j^2, \dots, \phi_j^K$ about a given angle β is

$$D(\beta) = \frac{1}{K} \sum_{k=1}^K \{1 - \cos(\phi_j^k - \beta)\}. \quad (6)$$

For any set of angular data, the dispersion of its mean direction over the set is equal to its circular variance, i.e., $D(\hat{\phi}_j) = v_j = 1 - r_j$. In Figure 2(e) and (f), the mean arguments $\hat{\phi}$ and moduli of the center of mass \hat{z} are shown as intensity maps.

3 Construction of the Statistical Models

In their work on eigenfaces, Turk and Pentland were among the the first to explore the use of principal components analysis for performing face recognition [8]. This method can be rendered efficient using the technique described by Sirovich et al.[7] which shows how to avoid explicit covariance matrix computation for large sets of two-dimensional images of objects of the same class.

We convert each image in the training set into a long-vector. Two encodings are investigated. In the first of these the long-vector has the measured azimuth angles as components. Hence, the j th component of the long-vector for the training image indexed k is $V_\phi^k(j) = \phi_j^k$. The second encoding involves using the complex number representation. Here the j th component of the long-vector for the training image indexed k is the complex number $V_z^k(j) = z_j^k$. We center the long-vectors by computing the mean

$$\hat{V} = \frac{1}{K} \sum_{k=1}^K V^k. \quad (7)$$

From the centered long-vectors we construct the data matrix

$$X = (\hat{V}^1 | \hat{V}^2 | \dots | \hat{V}^K). \quad (8)$$

In the case of the real-valued azimuth angle data the covariance matrix is $\Sigma_\phi = \frac{1}{K} X_\phi X_\phi^T$. For the complex representation the covariance matrix is $\Sigma_z = \frac{1}{K} X_z X_z^\dagger$, where \dagger denotes the transpose of the complex conjugate matrix. The resulting covariance matrices Σ_ϕ and Σ_z are respectively symmetric and Hermitian. We follow Atick et al.[1] and use the numerically efficient method of Sirovich [7] to compute the eigenvectors of Σ . For the real valued azimuth angle data this involves computing the eigen-decomposition of the matrices

$$Y_\phi = X_\phi^T X_\phi = U_\phi \Lambda_\phi U_\phi^T, \quad (9)$$

where the ordered eigenvalue matrix Λ_ϕ and eigenvector matrix U_ϕ are both real. Similarly, for the complex representation we compute the eigen-decomposition

$$Y_z = X_z^\dagger X_z = U_z \Lambda_z U_z^T. \tag{10}$$

In this case the ordered eigenvalue matrix Λ_z is real, but the elements of the eigenvector matrix U_z are complex. The eigenvectors of the matrices $X_\phi X_\phi^T$ and $X_z X_z^\dagger$ (or eigen-modes) are respectively the real matrix $\hat{U}_\phi = X_\phi U_\phi$ and the complex matrix $\hat{U}_z = X_z U_z$.

We deform mean long-vectors in the directions defined by the eigen-mode matrices. If \tilde{U}_L is the result of truncating U after the L leading eigenvectors then the deformed long vector is

$$V = \hat{V} + \sum_{l=1}^L \tilde{U}_l b_l, \tag{11}$$

where $b = [b_1, b_2, \dots, b_L]^T$ is a vector of real valued parameters of length L and \tilde{U}_l is the l th column of matrix \tilde{U} . Suppose that V_o is a centered long-vector of measurements to which we wish to fit the statistical model. We seek the parameter vector b that minimizes the squared error. The solution to this least-



Fig. 3. From left to right, the first six eigen-modes of \tilde{U}_z (a) and \tilde{U}_ϕ (b). The two first rows represent, respectively, $\hat{V}_z + \sqrt{3}\Lambda_z\hat{U}_z$ and $\hat{V}_z - \sqrt{3}\Lambda_z\hat{U}_z$. The variations $\hat{V}_\phi + \sqrt{3}\Lambda_\phi\hat{U}_\phi$ and $\hat{V}_\phi - \sqrt{3}\Lambda_\phi\hat{U}_\phi$ are shown in the two rows of (b).

squares estimation problem is $b^* = \tilde{U}^T V_o$. The best fit long-vector allowed by the model is $V_o^* = U U^T V_o$. In the case of both the real and complex representations, the parameter vectors are real.

In Figure 3 we compare the eigen-modes obtained using the real and complex representations. The rows of the figure show the first six eigen-modes. In the top two rows we show the five eigen-modes for Σ_z . The first row is the result of the displacement $\tilde{V}_z = \hat{V}_z + \sqrt{3}\Lambda_z \hat{U}_z$, and the bottom row the result of the displacement $\tilde{V}_z = \hat{V}_z - \sqrt{3}\Lambda_z \hat{U}_z$. We display the azimuth angles $\tilde{\phi}_z(j) = \arctan \frac{\text{Im}(\tilde{V}_z(j))}{\text{Re}(\tilde{V}_z(j))}$ at the pixel location indexed j . In the third and fourth rows, we repeat this analysis for the real-representation. Here we show the eigenmodes of Σ_ϕ . The third row shows the result of displacement $\tilde{V}_\phi = \hat{V}_\phi + \sqrt{3}\Lambda_\phi \hat{U}_\phi$, and the bottom row the result of the displacement $\tilde{V}_\phi = \hat{V}_\phi - \sqrt{3}\Lambda_\phi \hat{U}_\phi$. We display the azimuth angles $\tilde{\phi}_z(j) = \tilde{V}_\phi(j)$ at the pixel location indexed j . In general, both models seem to encapsulate the same facial features though the complex model \tilde{U}_z shows less noise than the real model \hat{U}_ϕ . These errors are most evident where $\hat{r}(j)$ is near zero. This suggests that the complex representation \tilde{U}_z is profiting of the inherent accuracy attached in the center of mass \hat{z}_j , which might be sacrificed by being projected onto the unit circle while calculating the mean direction $\hat{\phi}_j$.

4 Fitting the Model to Brightness Data

In brief SFS aims to solve the image irradiance equation, $I(x, y) = R(p, q, \mathbf{s})$, where I is the intensity value of the pixel with position (x, y) , R is a function referred to as *the reflectance map* [6]. The reflectance map uses the surface gradients $p = \frac{\partial Z(x, y)}{\partial x}$ and $q = \frac{\partial Z(x, y)}{\partial y}$ together with the light source direction vector \mathbf{s} to compute a brightness estimate which can be compared with the observed one using a measure of error. If the surface normal at the location (x, y) is $\mathbf{n} = [p, q, -1]$, then under Lambertian reflectance model, the image irradiance equation becomes $I(x, y) = \mathbf{n} \cdot \mathbf{s}$. Surface information can also be decoupled in azimuth (tilt) and zenith (slant) angles (φ and ϑ respectively), related to the surface normal by $\mathbf{n} = [\cos \varphi \sin \vartheta, \sin \varphi \sin \vartheta, \cos \vartheta]$.

Let I_j be the normalized image brightness at the pixel indexed j . From Lambert's law, the zenith angle for this pixel is $\vartheta_j = \arccos I_j$. Let $\tilde{\phi}_j = \arctan \frac{\text{Im} \tilde{V}(j)}{\text{Re} \tilde{V}(j)}$ be the azimuth angle at the pixel j obtained by fitting the complex model $\tilde{V}_z = \hat{V}_z + \tilde{U}_z b$. From the surface normal at the pixel j , $\mathbf{n}_j = [\cos \phi_j \sin \vartheta_j, \sin \phi_j \sin \vartheta_j, \cos \vartheta_j]^T$, we compute a numerical estimate of the Hessian matrix using first-differences. We are interested in the off-diagonal elements of this matrix $H_{xy}(j) = \frac{\partial n_x^z}{\partial y}$ and $H_{yx}(j) = \frac{\partial n_y^z}{\partial x}$.

Our aim is to fit the complex-model to brightness data so as to minimize the integrability error for the recovered field of surface normals. The error is defined to be

$$\text{Err}(b) = \sum_{j=1}^{MN} |H_{xy}(j) - H_{yx}(j)|. \quad (12)$$

Our algorithm for finding the best-fit parameter vector b is one based on search and involves varying its elements over equally divided intervals between $-\sqrt{3}\Lambda_z e$ and $+\sqrt{3}\Lambda_z e$ where $e = (1, 1, \dots, 1)^T$ is the all-ones vector of length L . Briefly, the algorithm is as follows: We initially zero all the components of the parameter vector b . Commencing from the first component, i.e. the one corresponding to the largest eigenvalue of Σ_z , we vary this component in S steps between $-3\sqrt{\Lambda_z(1)}$ and $+3\sqrt{\Lambda_z(1)}$ until $\text{Err}(b)$ is minimized. For each parameter setting, we recompute the field of surface normals so that the integrability error can be calculated. When the best-fit value of $b(1)$ is found it is fixed, and then we repeat the procedure for each of the remaining components of b in turn.

5 Experiments

In this section we present experiments with our statistical model for surface normal data. We commence by showing how the model can be trained on range data, and then fitted to out-of-sample range images. The second strand to the study is to show how the model can be fitted to brightness images to recover surface normals subject to Lambertian reflectivity and integrability constraints.

The face database used for constructing the surface models was provided by the Max-Planck Institute for Biological Cybernetics in Tuebingen, Germany. As described in [9], this database was constructed using Laser scans (*CyberwareTM*) of 200 heads of young adults, and provides head structure data in a cylindrical representation as well as ground-truth surface gradient for each of the face examples. We used $K = 150$ examples of size $M \times N = 150 \times 150$ pixels.

The results of fitting the model to out-of-sample range data (i.e. data not used in training) are shown in Figure 4. In the top row we show the result of

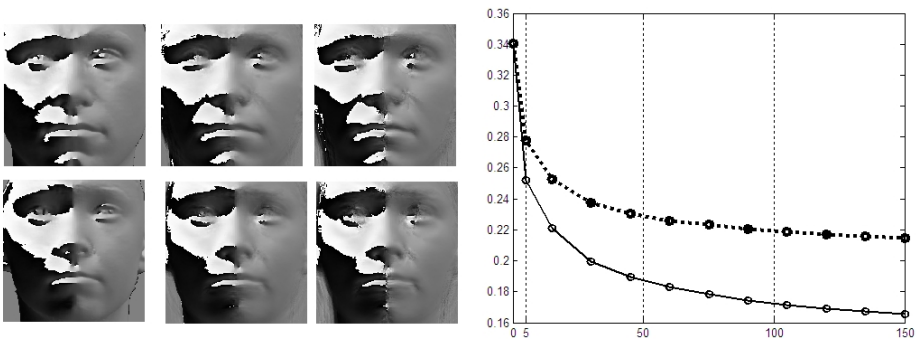


Fig. 4. Out-of-sample recovery analysis. From left to right: the first three columns show the ground truth azimuth angle, recovered azimuth angle using \tilde{U}_z and recovered azimuth angle using \tilde{U}_ϕ . The rightmost diagram shows the angular difference averaged over 50 out-of-sample data as a function of number of eigenmodes used for \tilde{U}_z (solid line) and \tilde{U}_ϕ (dashed line).

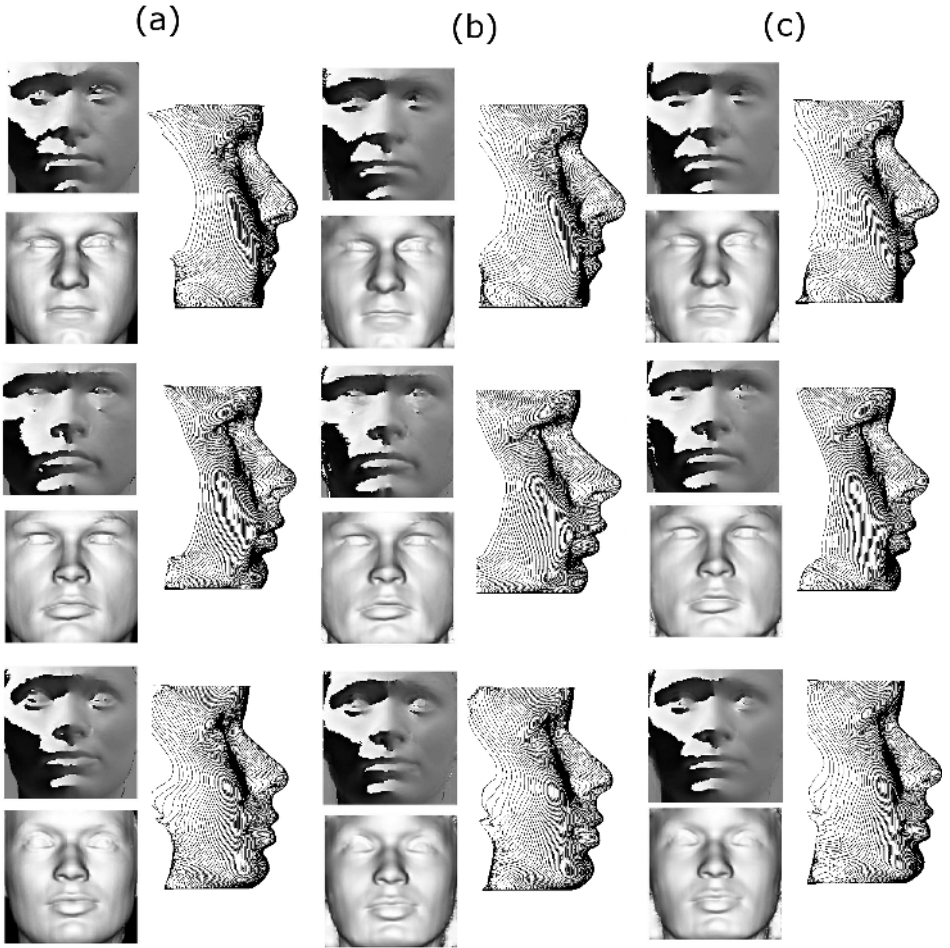


Fig. 5. Fitting the parameters to brightness data. We present three row-wise blocks representing different subjects. The results are organized in individual three-elements panels containing the following attributes: azimuthal angle (at the top), integrated surface (big surface at the right) and frontal re-illumination from integrated surface (at the bottom). Three column-wise blocks show true azimuth angles (a), best fit from true azimuth angles (b), fit from a Lambertian image using integrability constraints (d).

fitting the model to a male subject and the bottom row shows the result of fitting the model to a female subject. We show two panels of results. In the left panel we present the ground truth data, the result of fitting the complex model and the result of fitting the real model. The main feature to note from the panel is that the complex model achieves more accuracy on regions where the zenith angle is small, which in the estimation from the real model can be appreciated as small perturbations. In the rightmost diagram we show the absolute angular

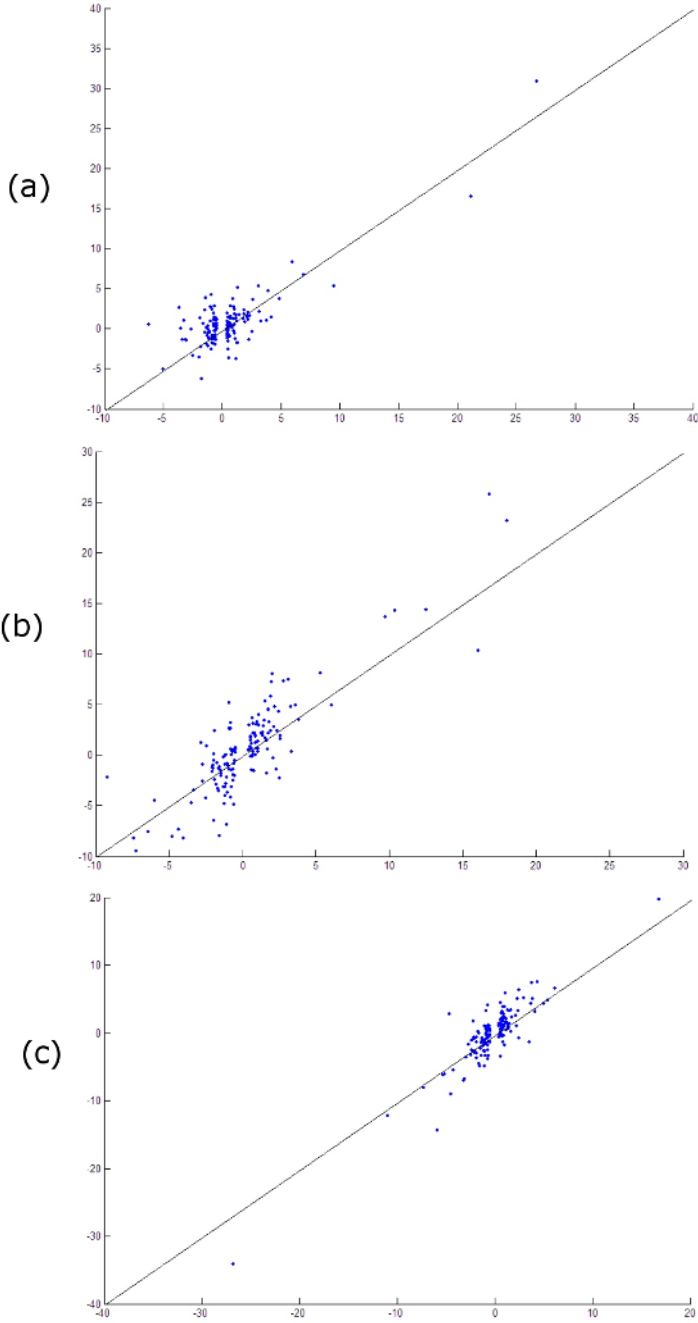


Fig. 6. The scatter plots show the relation between the best-fit parameters (y-axis) and the integrability-based adjustment (x-axis). We show three examples corresponding to the ones presented in Figure 5.

difference¹ averaged over 50 out-of-sample examples as a function of the number of eigen-modes used for \tilde{U}_z and \tilde{U}_ϕ . From the diagram it is clear that the complex model \tilde{U}_z outperforms the real model. The behavior of both models is similar, and the gap between the lines can be explained as a consequence of the badly recovered regions by the real model.

The result of applying the fitting algorithm outlined in Section 4 to three Lambertian images are shown in Figure 5. We used $L = 150$ parameters and $S = 10$ equally spaced values for making rough estimates of the parameter vector b . The results for each subject are organized into three row-wise blocks. Each block presents: (a) the results obtained from the ground truth surface normals, (b) the results obtained from the best fit to the ground truth azimuth angles, and (c) the results obtained by fitting to brightness information using the lambertian reflectance model and the integrability constraint. Three-elements Individual panels show the following attributes: the estimated azimuth angles (top), profile view of the surface reconstructed by integrating the recovered field of surface normals using the Frankot and Chellappa integration method[4] (bigger surface at the right), and the result of re-illuminating the reconstructed surface in the frontal viewer direction (bottom). We Note that there seems to be no significant differences between the results shown for the best adjustment and the ones shown for the integrability-constrained adjustment.

In Figure 6, the relationship between the best fit parameters and the parameters estimated by the integrability-constrained algorithm are shown as scatter plots, for the three examples explained above. Note how the scatters show a tendency to form a line, revealing a good correlation with the best-fit parameters.

6 Conclusions

We have explored the use of complex numbers as means of representing angular statistics for surface normal data. The reason for doing this is the difficulties encountered in representing surface normal information when attempting to fit statistical shape models to face images using shape-from-shading. We show how the complex valued model can be trained on surface normals delivered by range data and fitted to image brightness data. The fitting to brightness data is effected so as to satisfy Lambert's law and minimize an integrability error. For future work we are planning to experiment with textured images as well as adding different attributes for developing constraints (i.e. irradiance, variable albedo, curvature).

References

1. Atick, J., Griffin, P. and Redlich, N. (1996), Statistical Approach to Shape from Shading: Reconstruction of Three-Dimensional Face Surfaces from Single Two-Dimensional Images, *Neural Computation*, Vol. 8, pp. 1321-1340.

¹ The angular difference between the angles α and β , in radians, can be defined as $\pi - \|\pi - \|\alpha - \beta\|\|$.

2. Dovgird, R. and Basri, R. (2004), Statistical symmetric shape from shading for 3D structure recovery of faces, *European Conf. on Computer Vision (ECCV 04)*, Prague, May 2004.
3. Erens, R.G.F., Kappers, A.M.L. and Koenderink, J.J. (1993), Perception of Local Shape from Shading. *Perception and Psychophysics*, Vol. 54, No. 2, pp. 145 - 156.
4. Frankot, R.T. and Chellapa, R. (1988), A Method for Enforcing Integrability in Shape from Shading Algorithms, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 10, No. 4, pp. 438 - 451.
5. Mardia, K. V. (1972), Statistics of Directional Data, *Academic Press London and New York*.
6. Horn, B.K.P. (1977), Understanding Image Intensities, *Artificial Intelligence*, Vol. 8, pp. 201-231.
7. Sirovich, L. and Everson, Richard. (1992), Management and Analysis of Large Scientific Datasets, *The International Journal of Supercomputer Applications*, Vol. 6, No. 1, pp. 50 - 68.
8. Turk, M.A. and Pentland, A.P. (1991), Face Recognition Using Eigenfaces, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586 - 591.
9. Blanz, V. and Vetter, T. (1999), A Morphable model for the synthesis of 3D Faces, *Proceedings of SIGGRAPH '99*, pp. 187 - 194.