Fitting 3D Cartesian Models to Faces Using Irradiance and Integrability Constraints

Mario Castelán* and Edwin R. Hancock

Dept. of Computer Science, University of York,
York YO10 5DD, UK
{mario, erh}@cs.york.ac.uk

Abstract. This paper makes two contributions. First, we present an experimental analysis of three different ways of constructing three-dimensional statistical models of faces using Cartesian coordinates, namely, height, surface gradient and one based on Fourier domain basis functions. Second, we test the ability of each of the models for dealing with information provided by shape-from-shading by introducing a simple non-exhaustive parameter adjustment procedure subject to integrability and irradiance constraints. Experiments show that the surface gradient based representation is more robust to noise than the alternative Cartesian representations.

1 Introduction

In their work on eigenfaces, Turk and Pentland were among the first to explore the use of principal components analysis for performing face recognition [6]. This method can be rendered efficient using the technique described by Sirovich et al. [5] which shows how to avoid explicit covariance matrix computation for large sets of two-dimensional images of objects of the same class. The method involves cropping n examples of images of the same class (i.e. faces) and converting them into vectors $\mathbf{x}$ of size $m \times 1$. The mean example, $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$, is subtracted from each of the vectors $\mathbf{x}$ to form the matrix $\mathbf{X}_{m \times n}$. A set of *temporal* eigenvectors $\tilde{\Psi}_{n \times n}$ is calculated from the symmetric and non-negative outer product matrix $\mathbf{L}_{n \times n} = \mathbf{X}\mathbf{X}^T$. These eigenvectors are ordered according to the increasing magnitude of their corresponding eigenvalues. Subsequently, the *spatial* eigenvectors, or eigenmodes, $\Psi_{m \times n} = \mathbf{X}\tilde{\Psi}$ are generated. Every example can be recovered as a linear combination of a set of shape coefficients α with the eigenmodes Ψ , as $E = \bar{\mathbf{x}} + \sum_{i=1}^l \alpha_i \Psi_i$ where $1 \leq l \leq n$ is the chosen number of modes. By varying the shape coefficients α within ± 3 standard deviations, the resulting image is guaranteed to be a valid example of the class represented by the eigenmodes.

In 2D a set of aligned intensity images of faces is used to construct the eigenmodes, and the image data is usually encoded as a Cartesian long-vector by concatenating the rows or columns of the image. However, if a 3D model is to

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

Fig. 1. In the left-hand part of the Figure, a Cylindrical coordinate depth map (top left corner), height (bottom left corner), surface gradient (middle column) and Fourier domain (right-most column) Cartesian representations are shown as intensity plots. The right-hand part of the Figure shows, as height map plots, surface gradient (top row), Fourier basis (middle row) and height (bottom row) representations.

be constructed from range data then there exist alternative ways for representing the data. One of the most commonly used approaches is to adopt a representation that uses cylindrical coordinates. Using cylindrical coordinates, the surface of a human face (or head) can be parameterized by the function $r(\theta, \ell)$, where r is the radius and θ and ℓ are the height and angular coordinates respectively. This representation has been adopted since it captures the linear relations between basis heads, but can lead to ambiguity since different parameter sets can give rise to the same head-shape.

An alternative is to use a Cartesian representation, in which each surface point p is specified by its (x, y, z) co-ordinates, where the z -axis is in the direction of the viewer. The Cartesian coordinates are related to the cylindrical coordinates through $(x, y, z) = (x_0 + r(\theta, \ell)\sin\theta, y_0 + \ell, z_0 + r(\theta, \ell)\cos\theta)$, for some shift (x_0, y_0, z_0) relating the position of the origin in the two coordinate systems. Cartesian coordinates are not normally used since they introduce high z -variance (height variability) in areas with high spatial variance (i.e. face features such as nose, mouth and eyes). A face depth map in cylindrical coordinates could be thought as an unwrapped version of a depth map expressed in Cartesian coordinates. In Figure 1(a), a cylindrical coordinate depth map (top left corner) is shown together with its corresponding Cartesian coordinate depth map (bottom left corner).

Attick et al.[1] were among the first to build 3D statistical shape models of faces. They developed a shape-from-shading method which recovered shape-coefficients that satisfy the image irradiance equation¹. The eigenmode surfaces,

¹ The image irradiance equation[4] states that the measured brightness of the image is proportional to the radiance at the corresponding point on the surface. If $n(x, y)$ is the normal vector to the surface of an object at the position (x, y) and s is the light source direction vector, then the perceived brightness can be defined as $I = n \cdot s$. This is the simplest reflectance model and does not take into account albedo and inter-reflection effects.

which they called eigenheads, were parameterized in cylindrical coordinates. Later, Vetter et al.[7] decoupled texture from shape, and performed PCA on the two components separately. Using facial feature correspondences, they developed a model that could be fitted to image data. Although these methods sacrifice efficiency and simplicity since parameter tuning is performed using exhaustive search, they deliver accurate photo-realistic results. Recently, Dogvard and Basri[2] constructed a Cartesian model by combining statistical methods and symmetry constraints. Here the aim was to express the surface gradient in terms of the set of deformation coefficients. This allows shape-from-shading to be transformed into a linear system of equations, that can be simply solved for the shape coefficients and used to reconstruct the height function for the face, subject to symmetry constraints. Although accuracy is sacrificed, the method is computationally efficient.

Encouraged by Dogvard and Basri's work, in this paper our aim is to explore the use of alternative Cartesian representations for constructing 3D statistical models of faces. We also present an intuitive and straightforward method for shape coefficient adjustment which draws on the geometric shape-from-shading method of Worthington and Hancock [8]. In this way we avoid the use of costly heuristic search methods.

The paper is organized as follows: In Section 2 we give a brief explanation about the different Cartesian representations used. Section 3 describes how each of the models was constructed. The experimental results obtained using a data-driven shape coefficients search approach are presented in section 4. Finally, we present conclusions and suggest some possible lines for future work.

2 The Cartesian Representations

We have explored the use of three alternative Cartesian coordinate representations to construct the statistical models. These are height, surface gradient and a combination of these two, which we refer to as the Fourier domain method.

The *Height* representation is based on the surface height function $Z(x, y)$. An example is shown in Figure 1(a), leftmost column, bottom row, where intensity is used to represent the height map of a face.

The *Surface Gradient* representation is based on the directional partial derivatives of the height function $p = \frac{\partial Z(x, y)}{\partial x}$ and $q = \frac{\partial Z(x, y)}{\partial y}$. The set of first partial derivatives of a surface is also known as the gradient space, and it is related to the surface normal through $n = (p, q, -1)$. An example is shown in the middle column of Figure 1(a): x -partial (top) and y -partial (bottom) where intensity is used to represent the surface gradients. From the height function $Z(x, y)$, the first partial derivatives can be calculated through central, forward or backward finite differences.

The *Fourier Basis* representation draws on the concept of Fourier domain integrability for surface height recovery from surface gradient introduced by Frankot and Chellappa [3]. An integrable surface Z (i.e. $\frac{\partial^2 Z(x, y)}{\partial x \partial y} = \frac{\partial^2 Z(x, y)}{\partial y \partial x}$), can be defined by the basis expansion

$$\tilde{Z}(x, y) = \sum_{\omega \in \Omega} \tilde{C}(\omega) \phi(x, y, \omega) \quad (1)$$

where ω is a two dimensional index belonging to the domain Ω , and $\phi(x, y, \omega)$ is a set of basis functions which are not necessarily mutually orthogonal (i.e. the Fourier transform). The first partial derivatives of $\tilde{Z}$ can also be expressed in terms of this set of basis functions and will share the same set of coefficients $\tilde{C}_x(\omega) = \tilde{C}_y(\omega)$. If a set of first partial derivatives is not integrable then their corresponding transform coefficients will differ from each other (i.e. $\hat{C}_x(\omega) \neq \hat{C}_y(\omega)$). The distance between the non-integrable and the integrable partial derivatives can be minimized in the transform domain by making $\hat{C}_x(\omega) = \hat{C}_y(\omega) = \tilde{C}(\omega)$. It is been shown that $\tilde{C}(\omega) = \tilde{C}_1(\omega) + \tilde{C}_2(\omega)$, with

$$\tilde{C}_1(\omega) = \frac{P_x(\omega)\hat{C}_1(\omega)}{P_x(\omega) + P_y(\omega)}, \tilde{C}_2(\omega) = \frac{P_y(\omega)\hat{C}_2(\omega)}{P_x(\omega) + P_y(\omega)} \quad (2)$$

where $P_x(\omega)$ and $P_y(\omega)$ are $\int \int \|\phi_x(x, y, \omega)\|^2 dx dy$ and $\int \int \|\phi_y(x, y, \omega)\|^2 dx dy$ respectively. As the surface $\tilde{Z}$ is the inverse transform of $\tilde{C}(\omega)$, then $\tilde{Z}_1$ and $\tilde{Z}_2$ are the inverse transforms of $\tilde{C}_1(\omega)$ and $\tilde{C}_2(\omega)$ respectively, and $\tilde{Z} = \tilde{Z}_1 + \tilde{Z}_2$.

In the rightmost column of Figure 1(a) we show as intensity plots, the contributions to the height maps from $\tilde{Z}_1$ (top) and $\tilde{Z}_2$ (bottom).

In a complementary manner, depth maps corresponding to each representation are shown in Figure 1(b). From left to right we show: the x and y components of the surface gradient representation (top row), the x and y components of the Fourier representation (middle row) and height representation (bottom row).

3 Construction of the Models

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 [7], this database was constructed using Laser scans (*CyberwareTM*) of 200 heads of young adults, and provides head structure data in a cylindrical representation. For constructing the height based model, we converted the cylindrical coordinates to Cartesian coordinates and solved for the values of $Z(x, y)$. We were also provided with the ground-truth surface normals for the faces, which we used to calculate the surface gradient and Fourier domain eigenmodes.

For model construction, instead of using intensity images to generate the eigenmodes, as with eigenfaces, we used the eigenmodes obtained from height, surface gradient and Fourier domain, which we will refer to as Ψ^{hgt} , Ψ^{grd} and Ψ^{FB} respectively.

For Ψ^{hgt} , only one set of eigenmodes is required. We used 150 examples of face height data. This model is constructed so as to reproduce the face-height data variations as a linear combination of Ψ^{hgt} ,

$$E^{hgt} = \bar{\mathbf{x}}^{hgt} + \sum_{i=1}^l \alpha_i \Psi_i^{hgt} \quad (3)$$

To train Ψ^{grad} , we used 150 examples of ground truth face surface gradient data. It is important to note that this model is composed of two sets of eigenmodes, constructed using the first and second partial derivatives of the training data. This model re-produces surface gradient data variations as a linear combination of the set $\Psi^{grad} = \{\Psi^{grad^P}, \Psi^{grad^Q}\}$, with the two components:

$$E^{grad^P} = \bar{\mathbf{x}}^{grad^P} + \sum_{i=1}^l \alpha_i^P \Psi_i^{grad^P} \quad \text{and} \quad E^{grad^Q} = \bar{\mathbf{x}}^{grad^Q} + \sum_{i=1}^l \alpha_i^{grad^Q} \Psi_i^{grad^Q} \quad (4)$$

In the case of Ψ^{grad} , rather than directly applying the method to height data, we used surface gradient based information to construct the eigenmodes and calculated estimates of the first and second partial derivatives, p and q . We performed integration on the the gradient field to recover surface height data.

Similarly, the Fourier domain based set of eigenmodes, Ψ^{FB} can also be divided into two sub-models $\Psi^{FB} = \{\Psi^{FB^X}, \Psi^{FB^Y}\}$

$$E^{FB^X} = \bar{\mathbf{x}}^{FB^X} + \sum_{i=1}^l \alpha_i^X \Psi_i^{FB^X} \quad \text{and} \quad E^{FB^Y} = \bar{\mathbf{x}}^{FB^Y} + \sum_{i=1}^l \alpha_i^{FB^Y} \Psi_i^{FB^Y} \quad (5)$$

To construct the models, we used 150 examples of ground truth face surface gradient data and converted them to the surfaces $\tilde{Z}_1$ and $\tilde{Z}_2$ (as described in Section 2) through Equation 2 using Frankot and Chellappa's Fourier domain integration method. Therefore Ψ^{FB} characterizes the Fourier-basis surface variation as a linear combination of its eigenmodes. Given new examples, we can calculate the separate variations in $\tilde{Z}_1$ and $\tilde{Z}_2$ and add them together with their respective mean shapes to generate new surfaces of faces.

The generalization of the models, or the ability of the models to capture the features of the database from which they were constructed, is illustrated in Figure 2 (left), where the percentage of generalization is shown as a function of the number of modes used. The required number of modes was calculated through the formulae $\sum_{i=1}^t \lambda_i \geq f_v V_T$, where λ_i are the eigenvalues of the matrix $\mathbf{L}$, V_T is the total variance (i.e. the sum of all the eigenvalues) and f_v defines the proportion of the total variation captured by the model. We can see that Ψ^{FB} (solid and dotted lines) generalizes in a similar manner to Ψ^{hgt} (long-dashed line), both achieve more than 90% using just 20 modes. By contrast, Ψ^{grad} (short-dashed and dot-dashed lines) required a considerably larger number of modes (at least 100) to achieve the 90% level. Interestingly, for both Ψ^{FB} and Ψ^{grad} the x -component of the model shows a slightly better generalization than the y -component of the model. This may be attributable to the left-to-right symmetry of human faces.

4 Coefficient Adjustment Using Data-Driven Search

The next series of experiments used as input the 50 prototype intensity images (orthogonally illuminated, without texture) of the out-of-sample faces not

Fig. 2. Generalization of the models (left): note how Ψ^{FB} (solid and dotted lines) and Ψ^{hgt} (long-dashed line) show a similar behavior (achieving more than 90% when using at least 20 modes), while Ψ^{grad} (short-dashed and dot-dashed lines) requires at least 100 modes to achieve the 90% level. On the other hand, the performance tests (right) show that Ψ^{grad} (reaching an error of around 0.1%) outperforms both Ψ^{hgt} (around 5% as minimum) and Ψ^{FB} (around 25% as minimum).

used in training. The surface normals for these images are extracted using the geometric SFS approach described by [8], which forces the surface normals to comply with the image irradiance equation (i.e Lambert’s reflectance law) as a hard constraint. We do this in the following iterative fashion:

1. Extract initial surface normals from the intensity data using SFS and transform the set of normals to one of the three Cartesian representations ² for each of the representations (height, gradient, Fourier domain).
2. Calculate the corresponding set of coefficients α_i and back-project them back to the representation domain.
3. Re-apply Geometric SFS to this back-projected data.
4. Enforce integrability on this set of normals using Frankot and Chellappa’s global integration method, described in Section 2 and return to step 1 until a desired number of iterations is reached.

Hence, instead of searching for valid linear combinations of the set of coefficients α_i , we are exploring a rather intuitive and straightforward way for adjusting their values. In this way we are also testing the flexibility of each of the models, since the input data is far from a valid model example. Figure 2 (bottom row) shows the fractional error as a function of the number of iterations. The diagram shows that Ψ^{grad} (reaching an error of around 0.1%) clearly outperforms both Ψ^{hgt} (around 5% as minimum) and Ψ^{FB} (around 25% as minimum). These results suggest that Ψ^{grad} is more robust to poor initial surface normal estimates

² It is important to remark that SFS information will greatly differ from the data used in the former set of experiments (though out-of-sample, the 50 faces did represent 3D face shape information), since neither integrated height nor surface gradient nor Fourier basis information derived from SFS will present the accuracy of the database examples.

Fig. 3. Scatter plot diagrams, 10th iteration against true data. From left to right: Ψ^{hgt} , Ψ^{grd^P} , Ψ^{grd^Q} , Ψ^{FB^X} , Ψ^{FB^Y} . Note how the surface gradient based representation is the only one that seems to be benefited from the geometric SFS adjustment.

than the alternative representations, which might need a different initialization and probably a heuristic search for the values of their coefficients. Building the eigenmodes with surface gradients works significantly better. This corroborates the choice of Sirovich's method in conjunction with image-like bases to construct models (surface gradient is closer to intensity images than height).

The former observation is reinforced by the scatter plot diagrams shown in Figure 3. The x-axis represents the shape coefficients extracted from the surface normals extracted from intensity data using 10 iterations of SFS, while the y-axis represents the shape-coefficients extracted from the surface normals estimated from the ground-truth range data. From left to right the plots show the results obtained with Ψ^{hgt} , Ψ^{grd^P} , Ψ^{grd^Q} , Ψ^{FB^X} , Ψ^{FB^Y} . In the second and third diagrams, it is clear that the data-driven adjustment influences the correction of the initial set of shape coefficients for Ψ^{grd} , since the points in the scatter plots produce a diagonal regression-line. This regression-line is better defined for the coefficients α^{grd^P} (second image) than for α^{grd^Q} (third image), and this again suggests the models are better suited to capture left-to-right facial features (symmetry). For the remaining cases, α^{hgt} (first diagram) and α^{FB} (last two diagrams), there seems to be no relation between the geometric SFS coefficient adjustment and the true data-based shape coefficients.

Fig. 4. Profile views of the ground truth surfaces (top row, odd columns) and gradient-based model recovered surfaces (bottom row, odd columns) after the tenth iteration. Frontal re-illumination for each surface are shown in the even columns.

Finally, Figure 4 shows profile views of the ground truth surface (top row, odd columns) and the surface recovered using the gradient-based model after the tenth iteration (bottom row, odd columns). Frontal re-illuminations corresponding to each surface are also shown in the even columns. The difference between the ground truth and the recovered surface using Ψ^{grd} is almost imperceptible, confirming the robustness of Ψ^{grd} .

5 Conclusions

We have presented an analysis of three Cartesian representations for constructing three-dimensional statistical models of faces. These models were tested on SFS information, but only the surface gradient-based representation showed the necessary robustness to avoid the implementation of a specialized and time-consuming shape coefficient search technique, as required by cylindrical representations. When used to fit the model to intensity data using the image irradiance equation to constrain the surface normal direction, and also imposing integrability constraints on the recovered surface, then this representation leads to a natural way of refining the shape-coefficients. Hence we have demonstrated that 3D statistical models of faces based on Cartesian representations can work accurately without special search methods. As future work we are planning to explore the behavior of the Cartesian representations with alternative methods for shape coefficient adjustment.

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.
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. 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.
4. B.K.P. Horn. (1997), Understanding Image Intensities. *Artificial Intelligence*, Vol. 8, pp. 201-231.
5. 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.
6. 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.
7. Blanz, V. and Vetter, T. (1999), A Morphable model for the synthesis of 3D Faces, *Proceedings of SIGGRAPH '99*, pp. 187 - 194.
8. Worthington, P. L. and Hancock, E. R. (1999), New Constraints on Data-closeness and Needle Map Consistency for Shape-from-shading, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 12, pp. 1250-1267.