A Coupled Statistical Model for Face Shape Recovery

Mario Castelán*, William A.P. Smith, and Edwin R. Hancock

Department of Computer Science, University of York, York Y01 5DD, UK

Abstract. We focus on the problem of developing a coupled statistical model that can be used to recover surface height from brightness images of faces. The idea is to couple intensity and height by jointly modeling their combined variations. The models are constructed by performing Principal Component Analysis (PCA) on the shape coefficients for both intensity and height training data. By fitting the model to intensity data, the height information is implicitly recovered from the coupled shape parameters. Experiments show that the methods generate accurate surfaces from out-of training intensity images.

1 Introduction

One of the simplest approaches to facial shape recovery using shape-from-shading is to extract a field of surface normals and then recover the surface height function by integrating the surface normals [4,8,14]. Unfortunately, there are a number of obstacles that are encountered when this simple strategy is applied to real-world data. The most important of these is that when integrated, the concave/convex ambiguities in the needle-map can lead to the distortion of the topography of the reconstructed face. One of the most serious instances of this problem is that the nose can become imploded.

In general, shape-from-shading is an under-constrained problem since a surface normal has two degrees of freedom corresponding to the elevation and azimuth angles on the unit sphere which can not be recovered from a single brightness measurement. Domain specific constraints have been used to overcome this problem. Several authors [15,11,5,10] have shown that, at the expense of generality, the accuracy of recovered shape information can be greatly enhanced by restricting a shape-from-shading algorithm to a particular class of objects. For instance, both Prados and Faugeras [10] and Castelán and Hancock [5] use the location of singular points to enforce convexity on the recovered surface. Zhao and Chellappa [15] have introduced a geometric constraint which exploited the approximate bilateral symmetry of faces.

On the other hand, Atick et al. [1] proposed a statistical shape-from-shading framework based on a low dimensional parametrization of facial surfaces. Principal components analysis was used to derive a set of 'eigenheads' which compactly captures 3D facial shape. Unfortunately, it is surface orientation and not height which is conveyed by image intensity. Therefore, fitting the model to an image equates to a computationally expensive parameter search which attempts to minimise the error between the rendered surface and the observed intensity. Dovgard and Basri [7] combined the statistical constraint of Atick et al. and the geometric constraint of Zhao and Chellappa into a single

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

D.-Y. Yeung et al. (Eds.): SSPR&SPR 2006, LNCS 4109, pp. 898–906, 2006.

shape-from-shading framework. However, asymmetry in real face images results in errors in the recovered surfaces. Blanz and Vetter [3] decoupled surface texture from shape and performed PCA on the two components separately. Their framework could be used regardless of pose and illumination changes, but linear combinations of shape and texture had to be formed separately for the eyes, nose, mouth and the surrounding area. In addition, expensive alignment and parameter fitting procedures had to be carried out. The results delivered by fitting this morphable model proved to be accurate enough to generate photo-realistic views from an input image, though sacrificing efficiency and simplicity.

The aim in this paper is to explore how coupled statistical models can be used to overcome these difficulties. We couple height surface with intensity, developing a coupled statistical model that jointly describes variations in image brightness and height data over the surface of a face. The coupled model is inspired by the active appearance model developed by Cootes, Edwards and Taylor [6], which simultaneously models 2D shape and texture.

2 Principal Component Analysis

In this section we describe how the intensity and 3D data are represented, and how eigenspace models are constructed for these data. Here we follow the approach adopted by Turk and Pentland who were among the first to explore the use of principal components analysis for performing face recognition [13]. Further, we make use of the technique described by Sirovich et al. [12] to render the method efficient.

2.1 Generating an Intensity Model

The image data is vectorized by stacking the image columns to form long column vectors $\mathbf{p}$. If the K training images contain M columns and N rows, then the pixel with column index j_c and row index j_r corresponds to the element indexed $j = (N-1)j_c + j_r$ of the long column vector. The long column vectors are centered by computing the mean $\mathbf{m}_p = \frac{1}{K}\sum_{k=1}^{K}\mathbf{p}_k$.

From the centered vectors an $MN \times K$ data matrix $\mathbf{P} = (\mathbf{p}_1 - \mathbf{m}_p | \mathbf{p}_2 - \mathbf{m}_p | ... | \mathbf{p}_K - \mathbf{m}_p)$ is constructed, whose covariance matrix is $\mathbf{\Sigma}_p = \frac{1}{K}\mathbf{P}\mathbf{P}^T$. Unfortunately, since it is of size $MN \times MN$ the computation of the eigenvalues and eigenvectors of $\mathbf{\Sigma}_p$ becomes computationally impossible for large sets of data. However, the numerically efficient method proposed in [12] can be used to overcome these difficulties. This involves computing the eigen-decomposition of the $K \times K$ matrix $\frac{1}{K}\mathbf{P}^T\mathbf{P} = \mathbf{U}_p\mathbf{\Lambda}_p\mathbf{U}_p^T$, where the ordered eigenvalue matrix $\mathbf{\Lambda}_p$ and *temporal* eigenvector matrix $\mathbf{U}_p$ are both real. The *spatial* eigenvectors (or eigenfaces) of the covariance matrix $\mathbf{\Sigma}_p = \frac{1}{K}\mathbf{P}\mathbf{P}^T$ are given in terms of the eigenvectors of $\mathbf{P}^T\mathbf{P}$ by $\widetilde{\mathbf{P}} = \mathbf{P}\mathbf{U}_p$.

We deform the mean long-vector of image intensities in the directions defined by the eigenvalue matrix $\widetilde{\mathbf{P}}$. If we truncate $\widetilde{\mathbf{P}}$ after the L leading eigenvectors then the deformed long vector is $\mathbf{p}^* = \mathbf{m}_p + \widetilde{\mathbf{P}}\mathbf{b}_p$, where $\mathbf{b}_p = [b_{p_1}, b_{p_2}, \cdots, b_{p_L}]^T$ is a column vector of real valued parameters of length L. Suppose that $\mathbf{p}^o$ is a centered long-vector of measurements to which we wish to fit the statistical model. We seek the parameter

vector $\mathbf{b}_p^*$ that minimizes the squared error. The solution to this least-squares estimation problem is

$$\mathbf{b}_p^* = \widetilde{\mathbf{P}}^T \mathbf{p}^o. \tag{1}$$

In order to be valid examples of the class represented by the training set, the values of the coefficients $\mathbf{b}_p^*$ should be constrained to fall in the interval $\mathbf{b}_p \in [-3\sqrt{\mathbf{\Lambda}_p}e, +3\sqrt{\mathbf{\Lambda}_p}e]$, where $e = [1, 1, \cdots, 1]^T$ is the all-ones vector.

2.2 Generating a Height Model

We aim to train a surface height model corresponding to the image intensity data using range images. However, for range data there are alternative representations, One of the most commonly used alternatives is a representation that uses cylindrical coordinates [1,2]. 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. Unfortunately, it can lead to ambiguity since different data can be fitted to the same head-model.

An alternative, which overcomes this problem, is to use a Cartesian representation [7], in which each surface point 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 $(x, y, z) = (x_0 + R(\theta, \ell)\sin\theta, y_0 + \ell, z_0 + R(\theta, \ell)\cos\theta)$, where (x_0, y_0, z_0) is a reference shift. In this paper we will use the Cartesian form.

Each of the K range images (which are registered with the intensity images) in the training set may be represented by long vectors of height values $\mathbf{h}$ in the same way as the intensity data. The mean height vector $\mathbf{m}_h$ is given by $\mathbf{m}_h = \frac{1}{K}\sum_{k=1}^{K} \mathbf{h}_k$. We form the $MN \times K$ matrix of centered long vectors $\mathbf{H} = (\mathbf{h}_1 - \mathbf{m}_h | \mathbf{h}_2 - \mathbf{m}_h | ... | \mathbf{h}_K - \mathbf{m}_h)$. We can perform PCA to extract the set of spatial modes of variations of $\mathbf{H}$, $\widetilde{\mathbf{H}} = \mathbf{H}\mathbf{U}_h$. In the same manner, a centered long vector of height values $\mathbf{h}^o$ can be projected onto the eigenheads and represented using the vector of model coefficients $\mathbf{b}_h^* = \widetilde{\mathbf{H}}^T \mathbf{h}^o$.

3 Coupling Surface Height with Intensity

We now show how the intensity and the height models described above can be combined into a single coupled model. Each training sample can be summarized by the parameter vectors $\mathbf{b}_p$ and $\mathbf{b}_h$, representing the intensity and height of the sample respectively. In both models, we may consider small scale variation as noise. Hence, if the i_{th} eigenvalue for the intensity model is λ_p^i (where $\boldsymbol{\lambda}_p$ is the diagonal vector of the eigenvalue matrix $\mathbf{\Lambda}_p$), we need only S eigenmodes to retain p percent of the model variance. We choose S using $\sum_{i=1}^{S} \lambda_p^i \geq \frac{p}{100}\sum_{i=1}^{K} \lambda_p^i$. Similarly, for the height model we keep T eigenmodes to capture p percent of the variance.

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

$$\mathbf{b}_c^k = \begin{pmatrix} \mathbf{W}\mathbf{b}_p^k \\ \mathbf{b}_h^k \end{pmatrix} = \begin{pmatrix} \mathbf{W}\widetilde{\mathbf{P}}^T(\mathbf{p}_k - \mathbf{m}_p) \\ \widetilde{\mathbf{H}}^T(\mathbf{h}_k - \mathbf{m}_h) \end{pmatrix}, \tag{2}$$

where $\mathbf{W}$ is a diagonal matrix of weights for each intensity model parameter, allowing for the different relative weighting of the intensity and height models. As the elements of $\mathbf{b}_p$ and $\mathbf{b}_h$ represent different classes of data (grayscale and height), they can not be compared directly. We follow Cootes and Taylor [6] and set $\mathbf{W} = r\mathbf{I}$, where r^2 is the ratio of the total height variance to the total intensity variance. The coupled model data matrix is $\mathbf{C} = (\mathbf{b}_c^1|\mathbf{b}_c^2|...|\mathbf{b}_c^K)$. We apply a final PCA to this data to obtain the coupled model:

$$\mathbf{b}_c = \widetilde{\mathbf{C}}\mathbf{c} = \begin{pmatrix} \widetilde{\mathbf{C}}_p \\ \widetilde{\mathbf{C}}_h \end{pmatrix} \mathbf{c}, \tag{3}$$

where $\widetilde{\mathbf{C}}$ are the eigenvectors and $\mathbf{c}$ is a vector of coupled parameters controlling the intensity and height models simultaneously. The matrix $\widetilde{\mathbf{C}}_p$ has S rows, and represents the fist S eigenvectors, corresponding to the intensity subspace of the model. The matrix $\widetilde{\mathbf{C}}_h$ has T rows, and represents the final T eigenvectors, corresponding to the height subspace of the model.

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

$$\mathbf{p} = \mathbf{m}_p + \widetilde{\mathbf{P}}\mathbf{W}^{-1}\widetilde{\mathbf{C}}_p\mathbf{c}. \tag{4}$$

$$\mathbf{h} = \mathbf{m}_h + \widetilde{\mathbf{H}}\widetilde{\mathbf{C}}_h\mathbf{c}. \tag{5}$$

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

A plot of cumulative variance versus number of eigenmodes is shown in Figure 1. The height, intensity and coupled models are represented by the dashed, solid and dotted lines respectively. It is evident that fewer eigenmodes are required to capture variance in facial height than in facial intensity. This is because the intensity model has to deal with

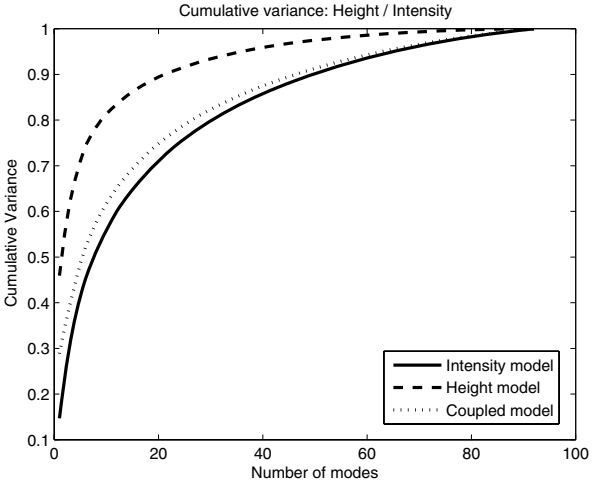

Fig. 1. Plot of cumulative variance versus number of eigenmodes used for intensity model (solid line), height model (dashed line) and coupled model (dotted line)

changes in shape and illumination, while the height model only deals with changes in shape. We retained 65 dimensions of the height model and 85 dimensions of the intensity model (each accounting for 95% of the variance). For the coupled model we retained 80 modes.

3.1 Fitting the Model to Intensity Data

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

Suppose that $\mathbf{p}^o$ is a centered vector of length MN that represents an intensity image of a face. Its best fit parameter vector, $\mathbf{b}_p^*$, is calculated through Equation 1. We fit the model to data seeking the vector $\mathbf{c}^*$ of length $S+T$ that satisfies the condition

$$\mathbf{c}^* = \arg\min_{\mathbf{c}} \{(\mathbf{b}_p^* - \mathbf{Q}_p\mathbf{c})^T(\mathbf{b}_p^* - \mathbf{Q}_p\mathbf{c})\} \tag{6}$$

The corresponding best fit vector of height values is given by

$$\mathbf{h} = \mathbf{m}_h + \widetilde{\mathbf{H}}\widetilde{\mathbf{C}}_h\mathbf{c}^* \tag{7}$$

We used a Matlab implementation of a quasi-Newton minimization procedure to solve Equation 7, constrained such that each coupled parameter lies within ± 3 standard deviations from the mean. One input image took around 5 seconds to converge to the best solution.

4 Experiments

In this section we report experiments focused on out-of-training characterization for the coupled model. The face database used for building the models was provided by the Max-Planck Institute for Biological Cybernetics in Tuebingen, Germany [2]. This database was constructed using Laser scans ($Cyberware^{TM}$) of 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 height values. We were also provided with the intensity maps for each 3D face.

We used 93 out-of-training cases. We calculated the fractional height difference error $\|Ground\_truth - Recovered\_surface\|/Ground\_truth$ as an average over the 93 surfaces and over all points. For the purposes of analysis, we ordered the out-of-training cases so that the first examples were those close in appearance to the mean shape $\mathbf{m}_p$. We used the sum of the first ten values of $\mathbf{b}_p$ (to account for at least 50% of the variability), i.e., $\sum_{i=1}^{10} b_{p_i}$ as a similarity measure.

In Figure 2 we show surface recovery results for three cases. The figure is divided into five columns. The different rows are for the three different subjects. In the first column, the three panels show the input image together with its frontal and profile views. The second panel contains the recovered best-fit intensity image. The third and

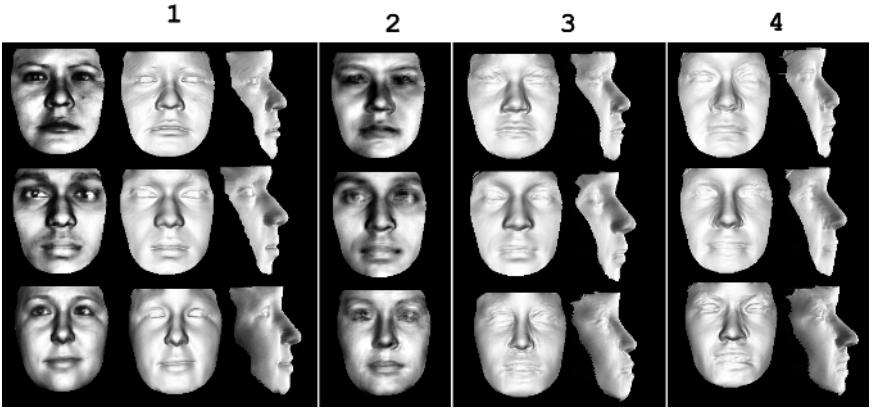

Fig. 2. Surface recovery results for three cases. The figure is divided in four columns. The first column shows the input image together with its frontal and profile views. The second column presents the best-fit intensity recovery. The next two panels present frontal and profile views from the intensity-height coupled model (third column) and the single height model (fourth column).

fourth columns contain panels which show frontal and profile views. The results of applying the full model are shown in the third column and those of applying the single height model are shown in the fourth column (i.e. the height data for the surface in panel 1 was used as an input for the single model $\widetilde{\mathbf{H}}$). The first two rows present cases where a percentage of height error around 1% while the last row shows cases with percentage of error bigger then 2%. As expected, the results shown in columns 3 seem to match

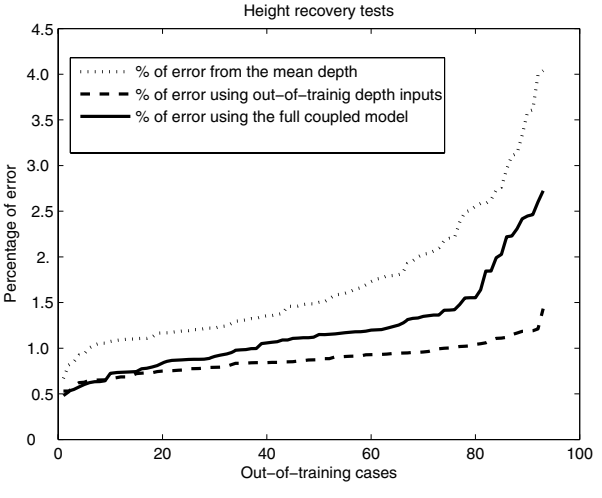

Fig. 3. Percentage of height difference for the 93 cases, ordered in an ascending way. The solid lines shows the full coupled model performance. The dashed line presents the error using a single height model (using height information as input), while the dot-dashed line shows the error from the mean height shape.

the best-fit intensity in column 2 rather than the original data in panel 1. However, even in those cases that differ significantly from the mean (last row), there is a good resemblance to the original data. This may be a consequence of basing surface recovery on the best-fit parameters directly from an intensity image. From the recovered faces, one can infer that in column 3, the surface recovery was led by the appearance of an input image. On the other hand, a visual analysis of the profiles in column 4 suggests that surface recovery was determined by an input height map.

In Figure 3 we plot the fractional height difference for the 93 cases. The solid line shows the coupled model performance. The dashed line shows the error obtained using a single height model (using height information as input), while the dot-dashed line shows the error from the mean height shape. The results were ordered in an ascending way for the purposes of comparison. The average error for the simple height model, coupled model and from the mean height are respectively 0.08%, 1.19% and 1.71%. Observe that many out-of-training examples whose intensities cannot be accurately recovered will generate less accurate height maps. However, considering that we are comparing two kinds of inputs (height and intensity), we can say that the coupled model delivers encouraging results.

Finally, we illustrate the utility of the coupled model with real world face images. These are drawn from the Yale B database [9] and are disjoint from the data used to train the statistical model. In the images the faces are in the frontal pose and were illuminated by a point light source situated approximately in the viewer direction. We aligned each

Fig. 4. Height recovery results using five examples from the Yale B database. From left to right: input image, intensity best-fit recovery, frontal illumination of the recovered height and profile and close-to-profile views with warped albedo-free and input images.

image with the mean intensity shape so that the eyes, nose tip and mouth center were in the same position. The surface recovery results are shown in Figure 4, where we present, from left to right: input image, intensity best-fit recovery, frontal illumination of the recovered height and profile and close-to-profile views with warped albedo-free and input image. Notice that even when the best-fit recovered intensity image is of lower quality than those in Figure 2(2), the surface reconstructions from the best-fit intensity parameters are sufficiently good to render novel views.

5 Conclusions

We have explored a way for coupling intensity and height information to construct statistical models of facial shape. Our coupled model strongly links the best-fit coefficients for intensity and height into a single statistical model. To recover the parameters of the coupled model, and hence reconstruct height, requires an optimization method. In this way best-fit intensity parameters can be calculated directly from an input image, and then used to recover height through the optimization search. The process only take some few seconds to converge to a minimum. The coupled model proved to be good enough to generate accurate surfaces from real world intensity imagery in an efficient way. Future research will focus on exploring the effect of these approaches to alternative representations including surface gradient, azimuth and zenith angles, and surface normals.

References

1. G. P. Atick, J. and N. Redlich. Statistical approach to shape from shading: Reconstruction of three-dimensional face surfaces from single two-dimensional images. *Neural Computation*, 8:1321–1340, 1996.
2. V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 187–194, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co.
3. V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(9):1063–1074, 2003.
4. H. E. Bors, A.G. and R. Wilson. Terrain analysis using radar shape-from-shading. *IEEE Trans, on Pattern Analysis and Machine Intelligence*, 25(5), 2003.
5. M. Castelán and E. Hancock. Acquiring height maps of faces from a single image. In *Proc. IEEE 3DPVT*, pages 183–190, 2004.
6. E. G. Cootes, T.F. and C. Taylor. Active appearance models. In *Proc. European Conference in Computer Vision*, pages 484–498, 1998.
7. R. Dovgard and R. Basri. Statistical symmetric shape from shading for 3d structure recovery of faces. In *Proc. European Conference on Computer Vision*, pages 99–113, May 2004.
8. R. Frankot and R. Chellapa. A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10:438–451, 1988.
9. B. D. Georghiades, A. and D. Kriegman. From few to many: Illumination cone models fro face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 634–660, 2001.

10. E. Prados and O. Faugeras. Unifying approaches and removing unrealistic assumptions in shape from shading: Mathematics can help. In *Proc. European Conference on Computer Vision*, pages 141–154, May 2004.
11. D. Samaras and D. Metaxas. Illumination constraints in deformable models for shape and light direction estimation. *IEEE Trans. PAMI*, 25(2):247–264, 2003.
12. L. Sirovich and R. Everson. Management and analysis of large scientific datasets. *The International Journal of Supercomputer Applications*, 6(1):50–68, 1992.
13. M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–591, 1991.
14. Z. Wu and L. Li. A line integration based method for depth recovery from surface normals. *CVGIP*, 43(1):53–66, 1988.
15. W. Zhao and R. Chellapa. Illumination-insensitive face recognition using symmetric shape-from-shading. In *Proc. Conference on Computer Vision and Pattern Recognition*, pages 286–293, 2000.