

# Approximating 3D Facial Shape from Photographs Using Coupled Statistical Models

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

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

**Abstract.** In this paper we focus on the problem of developing a coupled statistical model that can be used to recover surface height from frontal photographs of faces. The idea is to couple intensity and height by jointly modeling their combined variations. We perform Principal Component Analysis (PCA) on the shape coefficients for both intensity and height training data in order to construct the coupled statistical model. Using the best-fit coefficients of an intensity image, height information can be implicitly recovered through the coupled statistical model. Experiments show that the method can generate good approximations of the facial surface shape from out-of-training photographs of faces.

## 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 [1–3]. 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 [4–7] 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 [7] and Castelán and Hancock [6] use the location of singular points to enforce convexity on the recovered surface. Zhao and Chellappa [4] have introduced a geometric constraint which exploited the approximate bilateral symmetry of faces.

On the other hand, Atick et al. [8] 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

---

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

expensive parameter search which attempts to minimise the error between the rendered surface and the observed intensity. Dovgard and Basri [9] combined the statistical constraint of Atick et al. and the geometric constraint of Zhao and Chellappa into a single shape-from-shading framework. However, asymmetry in real face images results in errors in the recovered surfaces. Blanz and Vetter [10] 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 [11], which simultaneously models 2D shape and texture.

## 2 Principal Component Analysis

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

### 2.1 The Intensity Model

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

$$\mathbf{I}' = [(\mathbf{i}_1 - \bar{\mathbf{i}}) | (\mathbf{i}_2 - \bar{\mathbf{i}}) | \dots | (\mathbf{i}_K - \bar{\mathbf{i}})] = [\mathbf{i}'_1 | \mathbf{i}'_2 | \dots | \mathbf{i}'_K]. \quad (1)$$

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

$$\mathbf{V}^i = \sum_{k=1}^K \mathbf{i}'_k \mathbf{i}'_k{}^T = \mathbf{I}' \mathbf{I}'^T. \quad (2)$$

Unfortunately, due to its size ( $MN \times MN$ ), computing the eigenvalues and eigenvectors of  $\mathbf{V}^i$  becomes intractable for large sets of data. However, the numerically efficient

method proposed in [13] can be used to overcome these difficulties. According to this method there are only  $K - 1$  non zero eigenvalues from  $\mathbf{V}^i$  and these can be computed from the  $K \times K$  *sampled* covariance matrix  $\widehat{\mathbf{V}}^i = \mathbf{I}'^T \mathbf{I}'$ . The eigen-vector equations for the explicit and sampled covariance matrices,  $\mathbf{V}^i$  and  $\widehat{\mathbf{V}}^i$ , are

$$\mathbf{V}^i \mathbf{u}_k^i = \lambda_k^i \mathbf{u}_k^i \quad \text{and} \quad \widehat{\mathbf{V}}^i \hat{\mathbf{u}}_k^i = \hat{\lambda}_k^i \hat{\mathbf{u}}_k^i, \quad (3)$$

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

$$\widehat{\mathbf{V}}^i \hat{\mathbf{u}}_k^i = \hat{\lambda}_k^i \hat{\mathbf{u}}_k^i, \quad (4)$$

$$\mathbf{I}'^T \mathbf{I}' \hat{\mathbf{u}}_k^i = \hat{\lambda}_k^i \hat{\mathbf{u}}_k^i, \quad (5)$$

$$\mathbf{I}' \mathbf{I}'^T \mathbf{I}' \hat{\mathbf{u}}_k^i = \hat{\lambda}_k^i \mathbf{I}' \hat{\mathbf{u}}_k^i, \quad (6)$$

$$\mathbf{V}^i (\mathbf{I}' \hat{\mathbf{u}}_k^i) = \hat{\lambda}_k^i (\mathbf{I}' \hat{\mathbf{u}}_k^i). \quad (7)$$

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

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

$$\mathbf{M}^i = \mathbf{I}' \widehat{\mathbf{U}}^i, \quad (8)$$

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

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

$$\mathbf{b}^i = \mathbf{M}^{iT} (\hat{\mathbf{i}} - \bar{\mathbf{i}}). \quad (9)$$

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

$$\hat{\mathbf{i}} \approx \bar{\mathbf{i}} + \mathbf{M}^i \mathbf{b}^i. \quad (10)$$

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

## 2.2 The Surface Height Model

To explain how the surface height model was constructed, let us assume that each of the  $K$  surfaces in the training set may be represented by long vectors of height values  $\mathbf{h}$ . The mean height vector  $\bar{\mathbf{h}}$  is given by

$$\bar{\mathbf{h}} = \frac{1}{K} \sum_{k=1}^K \mathbf{h}_k. \quad (11)$$

In a similar manner to Equation 1, we form the  $MN \times K$  matrix of centered long vectors  $\mathbf{H}' = [(\mathbf{h}_1 - \bar{\mathbf{h}}) | (\mathbf{h}_2 - \bar{\mathbf{h}}) | \dots | (\mathbf{h}_K - \bar{\mathbf{h}})]$ . We calculate the eigenvectors  $\hat{\mathbf{u}}_k^h$  of the matrix  $\mathbf{H}'^T \mathbf{H}'$  and construct the height statistical model (as in Equation 8)

$$\mathbf{M}^h = \mathbf{H}' \hat{\mathbf{U}}^h, \quad (12)$$

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

$$\mathbf{b}^h = \mathbf{M}^{hT} (\mathbf{h} - \bar{\mathbf{h}}). \quad (13)$$

Note that the intensity model and the height model exhibit different modes of variation. This means that the information encoded by the intensity shape parameters,  $\mathbf{b}^i$ , is of limited use in directly recovering surface shape from intensity images. This problem has been circumvented by minimizing the distance between rendered views from recovered surfaces and input images, as in the work of Atick [8] and Vetter and Banz [10]. Unfortunately, the minimization of this distance is badly affected by the presence of local minima. This means that exhaustive search methods must be used, and this sacrifices efficiency. We overcome this problem by using a coupled statistical model to relate 2D intensity variations and variations in surface shape. Once fitted to data, the coupled model allows us to infer the shape-parameters from the best-fit intensity parameters, rather than using the distance between input images and rendered views of the recovered surfaces.

### 3 The Coupled Model

To construct the coupled model, each training example (i.e. pair of intensity image and corresponding aligned range image) can be summarized by the parameter vectors  $\mathbf{b}^i$  and  $\mathbf{b}^h$ . In both models, we assume that the lower eigenmodes represent small scale noise variation. Hence, if the  $k_{th}$  eigenvalue for the intensity model is  $\lambda_k^i$ , we need only  $S$  eigenmodes to retain  $Perc$  percent of the model variance. We choose  $S$  so that  $\sum_{k=1}^S \lambda_k^i \geq \frac{Perc}{100} \sum_{k=1}^K \lambda_k^i$ . Similarly, for the 3D models we retain  $T$  eigenmodes to capture  $Perc$  percent of the variance.

#### 3.1 Eigenmode Concatenation

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

$$\mathbf{b}_k^c = \begin{pmatrix} \mathbf{W} \mathbf{b}_k^i \\ \mathbf{b}_k^h \end{pmatrix} = \begin{pmatrix} \mathbf{W} \mathbf{M}^{iT} (\mathbf{t}_k^i - \bar{\mathbf{i}}) \\ \mathbf{M}^{hT} (\mathbf{t}_k^h - \bar{\mathbf{h}}) \end{pmatrix}, \quad (14)$$

where  $\mathbf{W}$  is a diagonal matrix of weights for each intensity model parameter, allowing for the different relative weighting of the intensity and surface models. As the elements of  $\mathbf{b}^i$  and  $\mathbf{b}^h$  represent different classes of data (grayscale and surface shape), they can not be compared directly. We follow Cootes and Taylor [11] and set  $\mathbf{W} = r\mathbf{M}^i$ , where  $r^2$  is the ratio of the total shape variance to the total intensity variance. The coupled model data matrix is  $(\mathbf{b}_1^c | \mathbf{b}_2^c | \dots | \mathbf{b}_K^c)$ .

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

$$\mathbf{b}^c = \mathbf{C}\mathbf{c} = \begin{pmatrix} \mathbf{C}^i \\ \mathbf{C}^h \end{pmatrix} \mathbf{c}, \quad (15)$$

where  $\mathbf{C}$  are the eigenvectors and  $\mathbf{c}$  is a vector of coupled parameters controlling the intensity and surface shape models simultaneously. The matrix  $\mathbf{C}^i$  has  $S$  rows, and represents the first  $S$  eigenvectors, corresponding to the intensity subspace of the model. The matrix  $\mathbf{C}^h$  has  $T$  rows, and represents the final  $T$  eigenvectors, corresponding to the surface shape subspace of the model.

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

$$\mathbf{t}^i = \bar{\mathbf{i}} + \mathbf{M}^i \mathbf{W}^{-1} \mathbf{C}^i \mathbf{c}. \quad (16)$$

$$\mathbf{t}^h = \bar{\mathbf{h}} + \mathbf{M}^h \mathbf{C}^h \mathbf{c}. \quad (17)$$

For compactness we write:  $\mathbf{Q}^i = \mathbf{W}^{-1} \mathbf{C}^i$ . We retained 95% of the variance for each of the models.

### 3.2 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  $\mathbf{b}^i$  and the recovered parameters  $\mathbf{Q}^i \mathbf{c}$ . In doing so, we implicitly recover the surface shape represented by the coupled model parameters.

Suppose that  $\mathbf{t}^i$  is a centered vector of length  $M \times N$  that represents an intensity image of a face. Its best fit parameter vector,  $\mathbf{b}^i$ , is calculated using Equation 9. We fit the model to data seeking the vector  $\mathbf{c}$  of length  $S + T$  that satisfies the condition

$$\mathbf{c} = \arg \min_{\mathbf{c}} \{(\mathbf{b}^i - \mathbf{Q}^i \mathbf{c})^T (\mathbf{b}^i - \mathbf{Q}^i \mathbf{c})\} \quad (18)$$

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

$$\mathbf{t}^h \approx \bar{\mathbf{h}} + \mathbf{M}^h \mathbf{C}^h \mathbf{c} \quad (19)$$

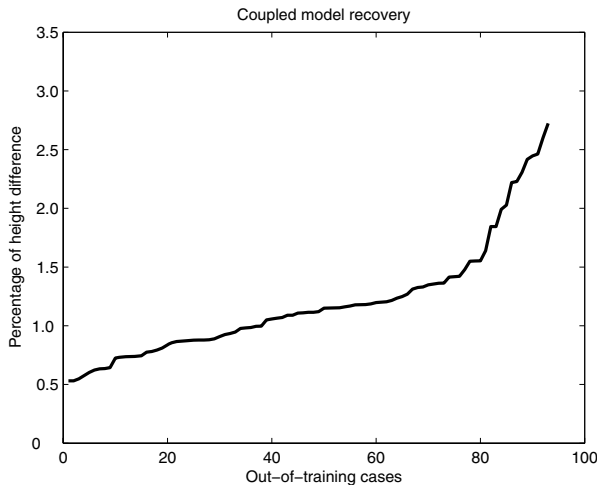
We used a Matlab implementation of the quasi-Newton minimization procedure to solve Equation 18. The fit was constrained such that each coupled parameter lies within  $\pm 3$  standard deviations from the mean. One input image took around a couple of seconds to converge to the best solution.

## 4 Experiments

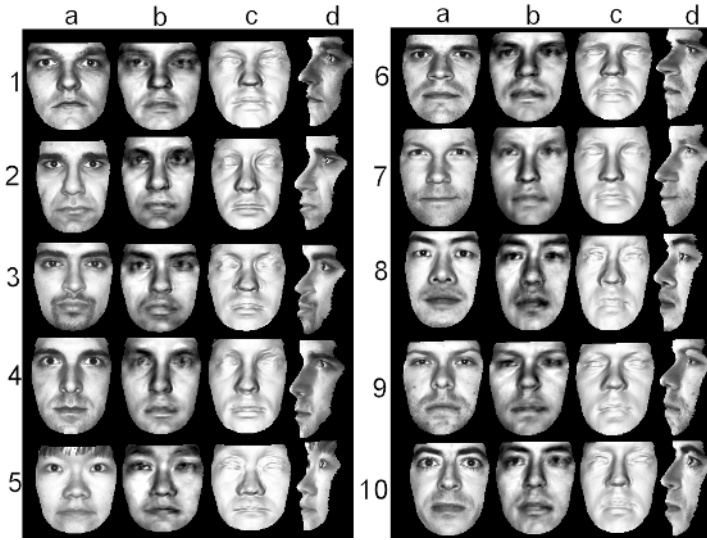
In this section we report experiments focused on using out-of-training-sample images to evaluate the ability of the coupled model to recover accurate surface information. The face database used for building the models was provided by the Max-Planck Institute for Biological Cybernetics in Tuebingen, Germany [14]. This database was constructed using Laser scans of heads of young adults, and provides head structure data in a cylindrical representation. For constructing the 3D based models, 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 constructed our models using 90 examples. We used 90 out-of-training-sample examples for surface reconstruction tests. We calculated the fractional height difference error  $\|Ground\_truth - Recovered\_surface\|/Ground\_truth$  as an average over the 90 surfaces and over all points on the surfaces. For the purposes of analysis, we ordered the out-of-training-samples examples according to their distance from the mean intensity image  $\bar{\mathbf{i}}$ . We used the sum of the first ten values of  $b^i$  (to account for at least 50% of the variability), i.e.,  $\sum_{j=1}^{10} b_j^i$  as a similarity measure. We test how well the coupled model performs using the 90 out-of-training-sample intensity images as input. We compare the recovered surfaces with the ground truth surface height data. In Figure 1 we plot the fractional height difference. The results were ordered according to ascending error for ease of visual comparison. The average surface recovery error was 1.19%. We also calculated the average error from every out-of-training example to the mean height shape, which was 1.71%.

Next, we turn our attention to real world images. The first set of images used is drawn from the Yale B database [15] and is disjoint from the data used to train the statistical models. In the images, the faces are in frontal pose and were illuminated by



**Fig. 1.** Plot of the fractional height difference between ground-truth and recovered surface when using the 90 out-of-training intensity images as input

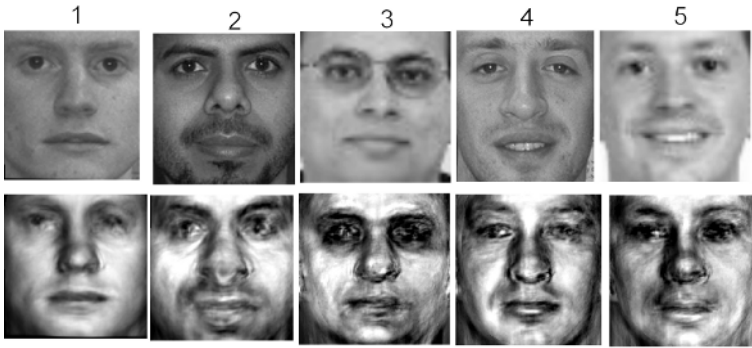


**Fig. 2.** Experiments with frontal images of 10 individuals of the Yale B database. The figure is divided into two panels, each of which contains five of the ten subjects in the database. In the figure, the rows are labeled with numbers and present the different subjects. The input image, intensity best-fit recovery, frontal illumination of the recovered height and profile view with warped input image are shown column-wise, for each panel.

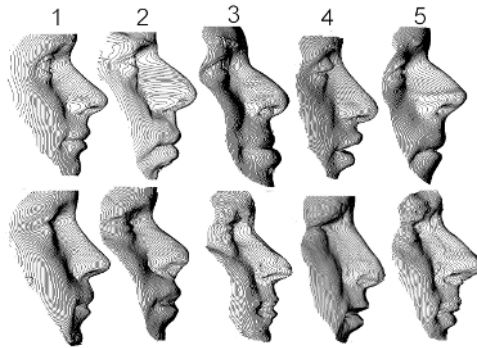
a point light source situated approximately in the viewer direction. We aligned each image with the mean intensity shape so that the eyes, nose tip and mouth center were in the same position. The surface recovery results using the coupled model are shown in Figure 2. The figure is divided into two panels, with five subjects on the left and five on the right. The result for the different subjects are shown in different rows. From left-to-right in each row we show the input image, the best-fit recovered intensity image, a frontal illumination of the recovered surface height and a profile view of texture mapping the input image onto the recovered surface. There are a number of features to note from the figure. First, the reconstructed images agree well with the input. Second, the overall shape of the profile view is subjectively convincing.

In Figure 3 we present another set of real world images. In the figure, the columns are labeled with numbers to identify the different subjects. The photographs were taken under uncontrolled lighting conditions. Also, the resolution of the photographs of subjects 3 and 5 was half the size of the one used for training. Note that we did not perform alignment operations to the mean intensity image. The top row of the figure shows the input image, while the bottom row shows the intensity best fit recovery. The intensity best fit recovery results are noisier than those obtained using the Yale database. Only subject 1 achieved a visually convincing reconstruction. The noise can be explained as consequence of not performing alignment operations as well as lower quality resolution.

The artifacts present in the bottom row of Figure 3 occur especially around the mouth and nose regions of the face. This effect may be sufficient to distort the recovered height



**Fig. 3.** Experiments with photographs taken under uncontrolled lighting conditions. Top row: input images. Bottom row: best fit recovery from the intensity model.



**Fig. 4.** Surface reconstruction comparison. Top row: ground truth surface. Bottom row: recovered surface using the coupled model.

surface, as we can verify in Figure 4, where we show profile views of the ground truth surfaces (top row) together with profile views of the recovered surfaces using the coupled model (bottom row). The ground truth surfaces of the subjects were acquired with a Cyberware laser scanner. Note that the input photographs are not aligned to the ground truth surfaces, therefore they exhibit slightly different poses and expressions. For this reason we did not perform height difference tests for the experiments on this set of images. A visual examination of both the recovered and the ground truth profiles reveals interesting features. First, the surface recovery of subject 1 presents the best agreement with the ground truth. This is a consequence of the good quality of its best fit intensity recovery. Second, the overall reconstructed facial shape is similar to the ground truth, specially in the nose and mouth shape. The area around the eyes seems to be the most difficult to recover.

Finally, in Figure 5 we present re-illumination experiments on the recovered surfaces for the five subjects. The light source direction makes an angle of  $-45$  and  $+45$  degrees to the image normal in the horizontal ( $x$ ) direction for rows 1 and 2, respectively. In





**Fig. 5.** Experiments on re-illuminating the recovered surfaces. The different columns show different subjects. The light source direction makes an angle of  $-45$  and  $+45$  degrees in the horizontal ( $x$ ) direction for rows 1 and 2 respectively. Similarly, the light source is moved in the vertical ( $y$ ) direction, in rows 3 and 4.

rows 3 and 4, the light source is moved in the same manner in the vertical ( $y$ ) direction. From the figure we can note that, despite the instabilities around the nose, mouth and eyes, the re-illuminations seem to be in accordance with the overall face structure.

## 5 Conclusions

We have proposed the use of coupled statistical models of 3D face shape and intensity in order to obtain facial surface approximation from photographs. The coupled model strongly links the best-fit coefficients for intensity and height data into a single statistical model. To recover the parameters of the coupled model, and hence reconstruct height shape, requires an optimization method whose objective function relies on the best-fit intensity parameters. Depending on the quality of the best-fit intensity recovery, the coupled model can be good enough to generate accurate surfaces from intensity imagery in an efficient way.

## References

1. H. E. Bors, A.G. and R. Wilson, "Terrain analysis using radar shape-from-shading," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, 2003.
2. R. Frankot and R. Chellapa, "A method for enforcing integrability in shape from shading algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, pp. 438–451, 1988.

3. Z. Wu and L. Li, "A line integration based method for depth recovery from surface normals," *CVGIP*, vol. 43, no. 1, pp. 53–66, 1988.
4. W. Zhao and R. Chellapa, "Illumination-insensitive face recognition using symmetric shape-from-shading," in *Proc. Conference on Computer Vision and Pattern Recognition*, 2000, pp. 286–293.
5. D. Samaras and D. Metaxas, "Incorporating illumination constraints in deformable models for shape and light direction estimation," *IEEE Trans. PAMI*, vol. 25, no. 2, pp. 247–264, 2003.
6. M. Castelán and E. Hancock, "Acquiring height maps of faces from a single image," in *Proc. IEEE 3DPVT*, 2004, pp. 183–190.
7. C. F. Prados, E. and O. Faugeras, "A unifying and rigorous shape from shading method adapted to realistic data and applications," *Journal of Mathematical Imaging and Vision*, 2006.
8. 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*, vol. 8, pp. 1321–1340, 1996.
9. R. Dovgand and R. Basri, "Statistical symmetric shape from shading for 3d structure recovery of faces," in *Proc. European Conference on Computer Vision*, May 2004, pp. 99–113.
10. V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, 2003.
11. E. G. Cootes, T.F. and C. Taylor, "Active appearance models," in *Proc. European Conference in Computer Vision*, 1998, pp. 484–498.
12. M. Turk and A. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1991, pp. 586–591.
13. L. Sirovich and R. Everson, "Management and analysis of large scientific datasets," *The International Journal of Supercomputer Applications*, vol. 6, no. 1, pp. 50–68, 1992.
14. 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*. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1999, pp. 187–194.
15. B. D. Georghades, A. and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 634–660, 2001.