Statistical Symmetric Shape from Shading for 3D Structure Recovery of Faces

Roman Dovgard and Ronen Basri

Dept. of Applied Mathematics and Computer Science,
Weizmann Institute of Science,
Rehovot 76100, Israel
{romad,ronen}@wisdom.weizmann.ac.il

Abstract. In this paper, we aim to recover the 3D shape of a human face using a single image. We use a combination of symmetric shape from shading by Zhao and Chellappa and statistical approach for facial shape reconstruction by Atick, Griffin and Redlich. Given a single frontal image of a human face under a known directional illumination from a side, we represent the solution as a linear combination of basis shapes and recover the coefficients using a symmetry constraint on a facial shape and albedo. By solving a single least-squares system of equations, our algorithm provides a closed-form solution which satisfies both symmetry and statistical constraints in the best possible way. Our procedure takes only a few seconds, accounts for varying facial albedo, and is simpler than the previous methods. In the special case of horizontal illuminant direction, our algorithm runs even as fast as matrix-vector multiplication.

1 Introduction

The problem of estimating the shape of an object from its shading was first introduced by Horn [1]. He defined the mapping between the shading and surface shape in terms of the reflectance function $I_{x,y} = R(p,q)$ where $I_{x,y}$ denotes image intensity, $p \doteq z_x$ and $q \doteq z_y$, z being the depth of the object and (x,y) are projected spatial coordinates of the 3D object. In this paper we will assume orthographic projection and Lambertian reflectance model, thus obtaining the following *brightness constraint*:

$$I_{x,y} \propto \rho_{x,y} \frac{1 - pl - qk}{\sqrt{p^2 + q^2 + 1}\sqrt{l^2 + k^2 + 1}}, \tag{1}$$

where $\frac{(l,k,1)}{||(l,k,1)||}$ is the illuminant direction (we have here proportion, instead of equality, because of the light source intensity). The task of a shape from shading algorithm is to estimate the unknowns of Eq. (1), which are the surface albedos $\rho_{x,y}$, and the surface depths $z_{x,y}$. With only image intensities known, estimating both the depths and the albedos is ill-posed. A common practice is to assume a constant surface albedo, but in a survey [2] it is concluded that depth estimates for real images come out to be very poor with this simplistic assumption.

In this paper, we aim to recover the 3D shape of a human face using a single image. In this case there are some other constraints, that can be imposed on the unknown variables $\rho_{x,y}$ and $z_{x,y}$ in Eq. (1). Shimshoni et al. [3] and Zhao and Chelloppa [4, 5] presented shape from shading approaches for symmetric objects, which are applicable to human faces. In [3], geometric and photometric stereo are combined to obtain a 3D reconstruction of quasi-frontal face images. In [4, 5] *symmetry constraints* on depth and albedo are used to obtain another, *albedo-free brightness constraint*. Using this constraint and Eq. (1) they find a depth value at every point. Lee and Rosenfeld [6] presented an approximate albedo estimation method for scene segmentation. Using this albedo estimation method, Tsai and Shah [7] presented a shape from shading algorithm, which starts with segmenting a scene into a piece-wise constant albedo regions, and then applies some shape from shading algorithm on each region independently. Nandy and Ben-Arie [8, 9] assume constant facial albedo and recover facial depth using neural networks. All the methods mentioned above assume either constant or piece-wise constant albedo, which is not a realistic assumption for real 3D objects such as human faces. Finally, Ononye and Smith [10] used color images for 3D recovery and obtained good results for simple objects. Results which they showed on *unrestricted* facial images are not so accurate, however, probably because of the lack of independence of the R, G, B channels in these images.

Atick et al. [11] and Vetter et al. [12–15] provided statistical shape from shading algorithms, which attempt to reconstruct the 3D shape of a human face using statistical knowledge about the shapes and albedos of human faces in general. In [11], a constant facial albedo is assumed and a linear constraint on the shape is imposed. The authors of [12–15], went a step further and have dropped the constant albedo assumption, imposing linear constraint on both texture (albedo map) and shape of the face. Because facial texture is not as smooth in general as the facial shape, imposing linear constraint on the texture requires a special preprocessing stage to align the database facial images to better match each other. Both approaches use certain optimization methods to find the coefficients of the linear combinations present in the linear constraints they are using, thus providing no closed-form solution, and consuming significant computational time.

We present an algorithm which accounts for varying facial albedo. Our method provides a closed-form solution to the problem by solving a single least-squares system of equations, obtained by combining *albedo-free brightness* and *class linearity* constraints. Our approach requires a restrictive setup: frontal face view, known directional illumination (that can be estimated for example by Pentland's method [16]) and Lambertian assumption about the face. We also get some inaccuracies in the reconstructed faces because human faces are not perfectly symmetric [17]. However, this is the first algorithm for 3D face reconstruction from a single image, which provides a closed-form solution within a few seconds.

The organization of the paper is as follows. In Sect. 2, we describe in detail previous work that is relevant to our approach, mainly symmetric and statistical shape from shading algorithms. Later, we present our algorithm in Sect. 3, and give experimental results in Sect. 4. Finally, we draw conclusions in Sect. 5.

2 Previous Work

2.1 Symmetric Shape from Shading

Zhao and Chellappa [4, 5] introduced symmetric shape from shading algorithm which uses the following *symmetry constraint*:

$$\rho_{x,y} = \rho_{-x,y}, \quad z_{x,y} = z_{-x,y} \tag{2}$$

to recover the shape of bilaterally symmetric objects, e.g. human faces. From *brightness* and *symmetry constraints* together the following equations follow:

$$I_{-x,y} \propto \rho_{x,y} \frac{1 + pl - qk}{\sqrt{p^2 + q^2 + 1}\sqrt{l^2 + k^2 + 1}} \tag{3}$$

$$\frac{I_{-x,y} - I_{x,y}}{I_{-x,y} + I_{x,y}} = \frac{pl}{1 - qk} \tag{4}$$

Denoting $D_{x,y} \doteq I_{-x,y} - I_{x,y}$ and $S_{x,y} \doteq I_{-x,y} + I_{x,y}$ we obtain the following *albedo free brightness constraint*:

$$Slp + Dkq = D. \tag{5}$$

Using Eq. (5), Zhao and Chellappa write p as a function of q, and substitute it into Eq. (1), obtaining equation in two unknowns, q and the albedo. They approximate the albedo by a piece-wise constant function and solve the equation for q. After recovering q and then p, the surface depth z can be recovered by any surface integration method, e.g., the one used in [18].

Yilmaz and Shah [19] tried to abandon the albedo piece-wise constancy assumption by solving Eq. (5) directly. They wrote Eq. (5) as an equation in $z_{x,y}$, instead of p and q, and tried to solve it iteratively. A linear partial differential equation

$$a_{x,y} z_x + b_{x,y} z_y = c_{x,y}, \tag{6}$$

of the same type in z, in different context, appears in the linear shape from shading method of Pentland [20], and was used also in [21].

Pentland tried to solve Eq. (6) by taking the Fourier transform of both sides, obtaining

$$A_{u,v}(-iu)Z_{u,v} + B_{u,v}(-iv)Z_{u,v} = C_{u,v}, \tag{7}$$

where A, B, C and Z stand for the Fourier transforms of a, b, c and z, respectively. It was stated in [20], that $Z_{u,v}$ can be computed by rearranging terms in Eq. (7) and taking the inverse Fourier transform. However, rearranging terms in Eq. (7) results in $Z_{u,v} = iC_{u,v}/(A_{u,v}u + B_{u,v}v)$. This equation is undefined when $A_{u,v}u + B_{u,v}v$ vanishes, and thus it leaves ambiguities in $Z_{u,v}$, and therefore also in $z_{x,y}$.

As was noted in [22], but was noticed neither in [19] nor in [20, 21], Eq. (6), being a linear partial differential equation in z, can be solved only up to a *one dimensional ambiguity* (hidden in the initial conditions), for example via the characteristic strip method [23]. Hence, Eq. (5) cannot be solved alone, without leaving large ambiguities in the solution.

2.2 Statistical Shape from Shading

Atick et al. [11] took a collection of 200 heads where each head was scanned using a CyberWare Laser scanner. Each surface was represented in cylindrical coordinates as $r^t(\theta, h), 1 \leq t \leq 200$ with 512 units of resolution for θ and 256 units of resolution for h. After cropping the data to a 256×200 (angular $\times$ height) grid around the nose, it was used to build *eigenhead decomposition* as explained below. They took $r_0(\theta, h)$ to be the average of 200 heads and then performed principal component analysis [24] obtaining the eigenfunctions Ψ_i, such that any other human head could be represented by a linear combination $r(\theta, h) = r_0(\theta, h) + \sum_{i=1}^{199} \beta_i \Psi_i(\theta, h)$. After that they applied conjugate gradient optimization method to find the coefficients β_i, such that the resulting $r(\theta, h)$ will satisfy the brightness constraint in Eq. (1), assuming constant albedo.

In [12–15], the authors went a step further and have dropped the constant albedo assumption, imposing linear constraint on both texture (albedo map) and shape of the face. In order to achieve shape and texture coordinate alignment between the basis faces, they parameterized each basis face by a fixed collection of 3D vertices (X_j, Y_j, Z_j), called a point distribution model, with associated color (R_j, B_j, G_j) for each vertex. Enumeration of the vertices was the same for each basis face. They modelled the texture of a human face by eigenfaces [25] and its shape by eigenheads, as described above. They recovered sets of texture and shape coefficients via complex multi-scale optimization techniques. These papers also treated non-Lambertian reflectance, which we do not treat here.

Both statistical shape from shading approaches described above do not use the simple Cartesian $(x, y) \mapsto z(x, y)$ parameterization for the eigenhead surfaces. While the parameterizations described above are more appropriate for capturing linear relations among the basis heads, they have the drawback of projecting the same head vertex onto different image locations, depending on the shape coefficients. Thus, image intensity depends on the shape coefficients. To make this dependence linear, the authors in [11] used Taylor expansion of the image I, with approximation error consequences. In [12–15], this dependence is taken into account in each iteration, thus slowing down the convergence speed.

3 Statistical Symmetric Shape from Shading

We use in this paper a database of 138 human heads from the USF Human-ID 3D Face Database [26] scanned with a CyberWare Laser scanner. Every head, originally represented in the database in cylindrical coordinates, is resampled to a Cartesian grid of resolution 142×125. Then, we have used a threshold on the standard deviation $z^{\mathtt{std}}$ of the facial depths, in order to mask out pixels (x, y) which are not present in all the basis faces, or alternatively are unreliable (see Fig. 1 (a) for the masked standard deviation map). After that we perform PCA on the first 130 heads, obtaining the 130 eigenheads z^i (first five of which are shown in Fig. 1 (b)-(f)), and keeping the remaining eight heads for testing purposes. We then constrain the shape of a face to be reconstructed, $z_{x,y}$, to the

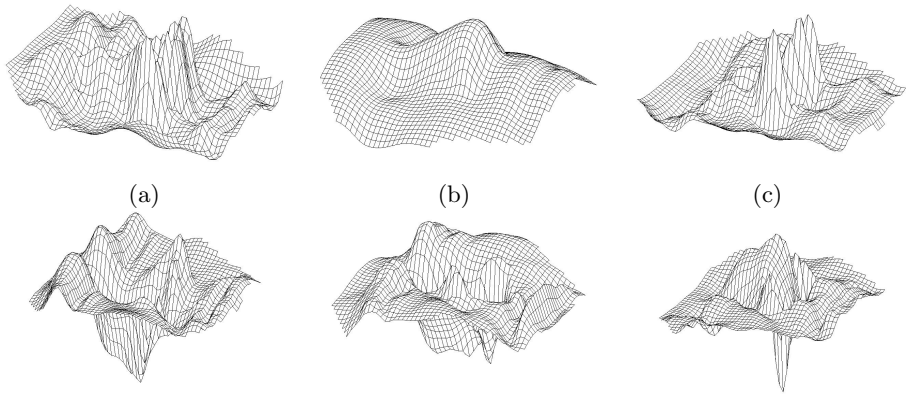


Fig. 1. (a) Standard deviation surface $z^{\mathtt{std}}$. Note high values of $z^{\mathtt{std}}$ around the nose for example: this is problem of Cartesian representation - high z variance in areas with high spatial variance, like area around the nose. (b)-(f) The five most significant eigenheads $z^1, \ldots, z^5$, out of the 130 eigenheads.

form:

$$z_{x,y} = \sum_{i=1}^{130} \alpha_i z^i_{x,y}, \tag{8}$$

for some choice of coefficients $\{\alpha_i\}$.

Since our *face space constraint* (8) is written in a Cartesian form, we can take derivatives w.r.t. x and y of both sides to obtain face space constraints on p, q:

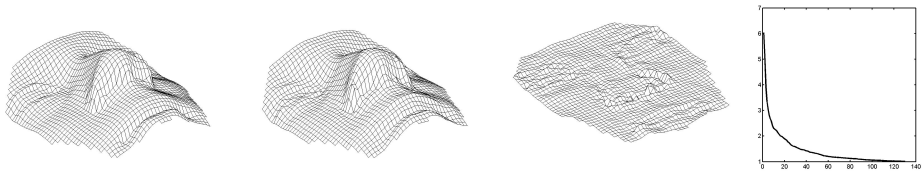
$$p = \sum_{i=1}^{130} \alpha_i p^i, \quad q = \sum_{i=1}^{130} \alpha_i q^i, \tag{9}$$

where $p^i \doteq z^i_x$ and $q^i \doteq z^i_y$. The two equations above, together with the albedo free brightness constraint (5) result in the following equation chain:

$$D = Slp + Dkq = Sl \sum_{i=1}^{130} \alpha_i p^i + Dk \sum_{i=1}^{130} \alpha_i q^i = \sum_{i=1}^{130} (Slp^i + Dkq^i)\alpha_i. \tag{10}$$

This equation is linear in the only unknowns $\alpha_i, 1 \leq i \leq 130$. We find a least-squares solution, and then recover z using Eq. (8). One can speed up calculations by using less than 130 eigenheads in the face space constraint (8).

The choice of Cartesian coordinates was necessary for obtaining face space constraints on p and q in Eq. (9). Although Cartesian parameterization is less appropriate for eigenhead decomposition than cylindrical or point distribution model parameterizations, it provides eigenhead decomposition with sufficient



(a) (b) (c) (d)

Fig. 2. Testing of *generalization* ability of our eigenhead decomposition. (a) One of the eight out-of-sample heads not used to construct the face space. (b) Projection of surface in (a) onto the 130-dimensional face space. (c) Error surface. Note some correlation with standard deviation surface from Fig. 1. (d) Plot of dependance of the reconstruction error on the number of modes used in the representation. The error is defined as $|z^{\texttt{actual}} - z^{\texttt{estimated}}| \cdot \cos\sigma^{\texttt{slant}}(z^{\texttt{actual}})$ averaged over all points on the surface and over eight out-of-sample heads, and is displayed as a percentage from the whole dynamic range of z. Error is normalized by the cosine of the slant angle, to account for the fact that the actual distance between two surfaces $z^{\texttt{actual}}$ and $z^{\texttt{estimated}}$ is approximately the distance along the z-axis times the cosine of the slant angle of the ground truth surface.

generalization ability as is shown in Fig. 2 - in reconstruction example (a)-(c) we see relatively unnoticeable error, and in (d) we see a rather fast decay of the *generalization error* when the number of basis heads increases. Dividing the generalization error with the first eigenhead only (which is just the average head) by the generalization error with 130 eigenheads we obtain the *generalization quality* $\frac{||z^{\texttt{actual}} - z^m||}{||z^{\texttt{actual}} - z^{\texttt{estimated}}||} = 5.97$ (here z^m stands for the average of the first 130 heads, and $z^{\texttt{actual}}$ with $z^{\texttt{estimated}}$ stand for the true and estimated depths, respectively), while [11] achieves with cylindrical coordinates a generalization quality of about 10. We provide in Sect. 3.3 a recipe for improving the generalization ability of the model, albeit without giving empirical evidence for it. The generalization errors we have right now are insignificant, comparing to errors in the reconstructions themselves, so that they do not play a major role in the accuracy of the results.

3.1 Special Case of $k = 0$

In the case of $k = 0$, Eq. (10) has a simpler form,

$$\frac{D}{Sl} = \sum_{i=1}^{130} p^i \alpha_i. \tag{11}$$

Setting $\{p^i\}$ to be the principal components of the x derivatives of the facial depths, and z^i accordingly to be the linear combinations of the original depths with the same coefficients which are used in p^i, one could solve for α's by setting $\alpha_i = \left\langle \frac{D}{Sl}, p^i \right\rangle$ (this is ≈ 150 times faster than in our general method; in fact, this is as fast as a matrix-vector multiplication), and for z via Eq. (8). This could be done up to a scaling factor, even without knowledge of light source direction.

Because different heads, which were used to build the face space, have face parts with different y coordinates, the face space has some vertical ambiguities. This means that for a given face in the face space, also some face with y shifts of a face components is also in the face space. In contrast to the general case, where ambiguity is driven by PDE's characteristic curves, and therefore is pretty random, in the case of $k = 0$ ambiguity is in vertical direction and thus is going in resonance with ambiguity in the face space. So, in the case of $k = 0$, final solution also will contain vertical ambiguities of the type $z_{x,y} = z^0_{x,y} + A(y)$ (where $A(y)$ is the ambiguity itself). We will show empirically, that due to vertical ambiguities in the the face space, results turn out to be quite inaccurate in this special case.

3.2 Extending Solution's Spatial Support

All basis heads used in build-up of the face space have different x, y support, and spatial support of the eigenheads is basically limited to the intersection of their supports. Therefore spatial support of the reconstructed face area is also limited. To overcome this shortcoming of our algorithm, one can fit surface parameterized via a point distribution model (PDM) [12–15] to our solution surface $z = z_{x,y}$. This fitting uses our partial solution surface for 3D reconstruction of the whole face, and is straightforward, as opposed to fitting of PDM to a 2D image.

3.3 Improving Generalization Ability

In the Cartesian version of the face space (8) eigenheads do not match each other perfectly. For example noses of different people have different sizes in both x, y and z directions. A linear combination of two noses with different x, y support produces something which is not a nose. Suppose now that we have a certain face with certain x, y support for its nose. In order to get this nose we need basis faces with noses of similar x, y support to this nose. This means that only a few basic faces will be used in a linear combination to produce this particular nose. This observation explains why a Cartesian version of face space has the highest generalization error among all face space representations: other representations have a better ability to match the supports of different face parts such as noses.

In order to overcome the drawbacks mentioned above, of the Cartesian principle component decomposition, we suggest another Cartesian decomposition for the shape space of human faces. Using NMF [27] or ICA [28, 29], it is possible to decompose the shape of a human face as a sum of its basic parts $z^i, 1 \leq i \leq 130$. In contrast to the principal component analysis, each z^i will have a compact x, y support. Therefore, from each z^i several z^i_j can be derived, which are just slightly shifted and/or scaled copies of the original z^i, in the x, y plane, in random directions. Using these shifted (and/or scaled) copies of original z^i, $1 \leq i \leq 130$, a broader class of face shapes can be obtained by a linear combination.

An alternative method for improving the generalization quality of the face space is to perform a spatial alignment of DB faces via image warping using for example manually selected landmarks [30]. For a progressive alignment it is possible to use eigenfeatures, which can be computed from the aligned DB faces.

4 Experiments

We have tested our algorithm on the Yale face database B [31]. This database contains frontal images of ten people from varying illumination angles. Alignment in the x, y plane between the faces in the Yale database and the 3D models is achieved using spatial coordinates of the centers of the eyes. We use eye coordinates of people from the Yale database, which are available. Also we marked the eye centers for our z^m and used it for the alignment with the images.

For every person in the Yale face database B, there are 63 frontal facial views with different known point illuminations stored. Also for each person its ambient image is stored. As in this paper we do not deal with the ambient component we subtract the ambient image from each one of the 63 images, prior to using them. As we need ground truth depths for these faces for testing purposes, and it is not given via a laser scanner for example, we first tried to apply photometric stereo on these images, in order to compute the depths. However, because the 63 light sources have different unknown intensities, performing photometric stereo is impossible. Hence, we have taken a different strategy for "ground truth" depth computation.

We took two images, taken under different illumination conditions, for each one of the ten faces present in the database. A frontal image F with zero azimuth Az and elevation El, and image I with azimuth angle $Az = 20°$ and elevation angle $El = 10°$. Substituting $l = -\tan 20°$ and $k = \tan 10°/\cos 20°$ into Eq. (1), we obtain a Lambertian equation for the image I. Doing the same with $l = 0$ and $k = 0$ we obtain a Lambertian equation for the image F. Dividing these two equations, we obtain the following equation, with $l = -\tan 20°$ and $k = \tan 10°/\cos 20°$ (we have here equality up to scaling factor λ caused by a difference of light source intensities used to produce the images I and F):

$$\lambda \frac{I_{x,y}}{F_{x,y}} = 1 - pl - qk. \tag{12}$$

Eqs. (12) and (9) together enable us to get the "ground truth" depth, given the scaling factor λ (which we will show how to calculate, later on):

$$F_{x,y} - \lambda I_{x,y} = \sum_{i=1}^{130} (F_{x,y} l p^i + F_{x,y} k q^i)\alpha_i. \tag{13}$$

Of coarse such a "ground truth" is less accurate than one obtained with photometric stereo, because it is based on two, rather than on all 63 images, and because it uses the face space decomposition, which introduces its own generalization error. Also small differences between the images I and F cause small errors in the resulting "ground truth". Still this "ground truth" has major advantage over results obtained by our symmetric shape from shading algorithm, which is that it does not use an inaccurate symmetry assumption about the faces [17].

Our algorithm estimates the depths of each one of the ten faces by solving Eq. (10). Then we take the estimated α_i's and plug them into Eq. (13). We find

Table 1. Quality and computational complexity of our algorithm, applied to all the ten Yale faces. Ten columns correspond to the ten faces. The first row contains asymmetry estimates of the faces (we subtract the lowest element). The second row contains the quality estimates, measured via an inverse normalized distance between the estimated and actual depths. The third row contains the quality estimates with statistical Cartesian shape from shading (assuming constant albedo as in the work of Atick et al. [11]). The last row contains running times (in seconds) of our algorithm on all ten faces

Asymmetry	0.53	1.08	0	0.52	1.36	0.87	0.75	0.68	0.65	0.38
Quality	1.81	1.78	3.39	1.92	1.91	1.29	1.96	2.59	1.45	2.94
Const Albedo	1.27	1.24	1.76	1.41	1.57	1.10	2.15	2.5	1.5	2.35
Running Time	1.76	1.76	1.76	1.76	2.21	1.76	1.75	1.76	1.75	1.76

the best λ satisfying this equation and use it in the "ground truth" calculation. This mini-algorithm for λ calculation is based on the fact that our estimation, and the ground truth are supposed to be close, and therefore our estimation can be used to reduce ambiguity in the "ground truth" solution. As λ has some error, we need to perform a small additional alignment between the "ground truth" and the estimated depths. We scale the "ground truth" solution at the end, so that it will have the same mean as the estimated depth.

In Table 1 we provide asymmetry estimates for all ten Yale DB faces along with quality and computational complexity estimates of the results. Asymmetry is measured via a Frobenius distance between normalized (to mean gray level 1) frontally illuminated face F and its reflect R. Quality is measured by the fraction $||z^{\texttt{actual}} - z^{m}||/||z^{\texttt{actual}} - z^{\texttt{estimated}}||$, and computational complexity measured by a running time, in MATLAB, on a Pentium 4 1600MHz computer. Correlation coefficient between the facial asymmetry and resulting quality estimates is -0.65, indicating a relatively strong anti-correlation between quality and asymmetry.

Results with Cartesian face space, replacing the symmetry with the constant albedo assumption (thus simulating the work of Atick et al. [11]), are found via iterative optimization on the α_i's according to Eq. (1), initialized by our solution. For each database image we chose its best scale to fit into the Lambertian equation, thus making the statistical shape from shading results as good as possible for a faithful comparison. Our results are slightly better on most faces (Table 1).

The three best results of our algorithm, with quality at least 2.5 (faces 3,8 and 10 in the Yale ordering), are depicted in Fig. 3 (along with their statistical SFS counterparts). Also, in the first three rows of Fig. 4, we show textured faces (with texture being the image of frontally illuminated face), rendered with our "ground truth", estimated and average depths. In the first three rows of Fig. 5, we render these faces as if they were shot using frontal illumination, by taking images with $Az = 20^\circ$ and $El = 10^\circ$ and cancelling out side illumination effect by dividing them by $1 - lp - kq$, where p and q are recovered by our algorithm from the image I and are given directly (without using z) by Eq. (9).

In the last row of Fig. 4, we show results for the face number 6 in the Yale database, which has the worst reconstruction quality (see Table 1). In the last row

Table 2. Quality of results of our algorithm for the case that $k = 0$, on all the ten Yale faces. The ten columns correspond to the ten faces. We performed quality estimation for reconstructions from images with $Az = -10^\circ$ and $Az = -25^\circ$. Prior to estimating the quality, we shifted and stretched the estimated depths, so that they will have the same mean and variance as the "ground truth" depths. We estimated quality with and without, row normalization, in which each depth row to be estimated was normalized to have the same mean as its counterpart row in the "ground truth" depth

Without normalization (true quality)										
$Az = -10^\circ$	0.70	0.96	0.88	1.25	0.76	0.90	1.06	1.63	0.81	1.20
$Az = -25^\circ$	0.86	1.24	1.14	1.21	0.91	0.68	1.29	1.81	0.92	1.50

With normalization (additional test)										
$Az = -10^\circ$	1.34	1.63	1.58	1.69	1.52	1.11	1.46	2.92	1.54	2.07
$Az = -25^\circ$	1.70	2.20	2.27	1.93	2.12	0.90	2.08	3.14	1.88	2.27

of Fig. 5, we show renderings for this face. One can note a significant asymmetry of the face, which explains the rather bad reconstruction results in Fig. 4. We provide, in the additional material, results of the algorithm on *all* ten Yale faces.

We attribute inaccurate results of our algorithm, on many faces, to facial asymmetry. Results in Fig. 5 can be compared with similar results by Zhao and Chellappa [4, 5] (see Figs. 14 and 15 in [4]). Note that both are affected by facial asymmetry. Using some illumination invariant feature matcher [32], features on two sides of a face, could be matched based solely on the albedo, and warp the face to one with symmetric shape and texture, but a warped illumination (with less impact on errors, due to smoothness of the illumination). However, we doubt whether this approach is feasible, because of the matching errors.

4.1 Results in the case of $k = 0$

In the special case of zero light source elevation angle El, we took two images with different light source azimuth angle Az, for each one of the ten faces present in the database. One image with $El = -10^\circ$ and the other with $El = -25^\circ$. Our algorithm estimates depths of each one of the ten faces by solving Eq. (11). We shift and stretch the estimated depths, so that they will have the same mean and variance as the "ground truth" depths. In first part of Table 2, we provide quality estimates of the results of our special case algorithm, on all the ten faces.

We have done further alignment between the estimated and "ground truth" depths. We have normalized all the rows of the estimated depths to have the same mean as their counterpart rows in the "ground truth" depths. Thereafter we have measured the quality estimates of the results, and presented them in the second part of Table 2. One can note a significant increase in estimates, relatively to the first part of Table 2, which is an indication of a significant 1D ambiguity which is left in the solution of Eq. (11). Quality estimates in the second part of Table 2 are comparable with those of our main results in Table 1.

5 Conclusions

In this paper we have presented a successful combination of two previous facial shape reconstruction approaches - one which uses symmetry and one which uses statistics of human faces. Although our setup in this paper is rather restrictive and results are inaccurate on many faces, still our approach has a major advantages over the previous methods - it is very simple, provides a closed-from solution, accounts for facial nonuniform albedo and has extremely low computational complexity. The main disadvantage of the algorithm is inaccurate results on some faces caused by asymmetry of these faces. On most faces, however, we obtain reconstructions of sufficient quality for creation of realistically looking new synthetic views (see new geometry synthesis in Fig. 4 and new illumination synthesis in Fig. 5). In general, synthesizing views with new illumination does not require very accurate depth information, so that our algorithm can be considered appropriate for this application because of its simplicity and efficiency.

Acknowledgements

We are grateful to Prof. Sudeep Sarkar, University of South Florida, for allowing us to use the USF DARPA HumanID 3D Face Database for this research. Research was supported in part by the European Community grant number IST-2000-26001 and by the Israel Science Foundation grants number 266/02. The vision group at the Weizmann Inst. is supported in part by the Moross Laboratory for Vision Research and Robotics.

References

1. Horn, B.K.P.: Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View. MIT AI-TR-232, 1970.
2. Zhang, R., Tsai, P.S., Cryer, J.E., Shah, M.: Shape from Shading: A Survey. IEEE Trans. on PAMI **21**(8), 690–706, 1999.
3. Shimshoni, I., Moses, Y., Lindenbaum, M.: Shape Reconstruction of 3D Bilaterally Symmetric Surfaces. IJCV **39**(2), 97–110, 2000.
4. Zhao, W., Chellappa, R.: Robust Face Recognition using Symmetric Shape-from-Shading. University of Maryland, CARTR-919, 1999.
5. Zhao, W., Chellappa, R.: Illumination-Insensitive Face Recognition using Symmetric Shape-from-Shading. In Proc. of CVPR, 286–293, 2000.
6. Lee, C.H., Rosenfeld, A.: Albedo estimation for scene segmentation. Pattern Recognition Letters **1**(3), 155–160, 1982.
7. Tsai, P.S., Shah, M.: Shape from Shading with Variable Albedo. Optical Engineering, 1212–1220, 1998.
8. Nandy, D., Ben-Arie, J.: Shape from Recognition and Learning: Recovery of 3-D Face Shapes. In Proc. of CVPR **2**, 2–7, 1999.
9. Nandy, D., Ben-Arie, J.: Shape from Recognition and Learning: Recovery of 3-D Face Shapes. IEEE Trans. On Image Processing **10**(2), 206–218, 2001.
10. Ononye, A.E., Smith, P.W.: Estimating the Shape of A Surface with Non-Constant Reflectance from a Single Color Image. In Proc. BMVC, 163–172, 2002.

11. Atick, J.J., Griffin, P.A., Redlich, A.N.: Statistical Approach to Shape from Shading: Reconstruction of Three-Dimensional Face Surfaces from Single Two-Dimensional Images. Neural Computation **8**(6), 1321–1340, 1996.
12. Blanz, V., Vetter, T.A.: A morphable model for the synthesis of 3d faces. In Proc. of SIGGRAPH, 187–194, 1999.
13. Romdhani, S., Blanz, V., Vetter, T.: Face Identification by Fitting a 3D Morphable Model using Linear Shape and Texture Error Functions. In Proc. of ECCV **4**, 3–19, 2002.
14. Blanz, V., Romdhani, S., Vetter, T.: Face Identification across different Poses and Illuminations with a 3D Morphable Model. In Proc. of the IEEE Int. Conf. on Automatic Face and Gesture Recognition, 202–207, 2002.
15. Romdhani, S., Vetter, T.: Efficient, Robust and Accurate Fitting of a 3D Morphable Model. In Proc. of ICCV, 59–66, 2003.
16. Pentland, A.P.: Finding the Illuminant Direction. JOSA **72**, 448–455, 1982.
17. Liu, Y., Schmidt, K., Cohn, J., Mitra, S.: Facial Asymmetry Quantification for Expression Invariant Human Identification. Computer Vision and Image Understanding **91**(1/2), 138–159, 2003.
18. Basri, R., Jacobs, D.W.: Photometric Stereo with General, Unknown Lighting. In Proc. of CVPR, 374-381, 2001.
19. Yilmaz, A., Shah, M.: Estimation of Arbitrary Albedo and Shape from Shading for Symmetric Objects. In Proc. of BMVC, 728-736, 2002.
20. Pentland, A.P.: Linear Shape from Shading. IJCV **4**(2), 153–162, 1990.
21. Cryer, J., Tsai, P., Shah, M.: Combining shape from shading and stereo using human vision model. University of Central Florida, CSTR-92-25, 1992.
22. Jacobs, D.W., Belhumeur, P.N., Basri, R.: Comparing images under variable illumination. In Proc. of CVPR, 610–617, 1998.
23. Zauderer, E.: Partial Differential Equations of Applied Mathematics, 2nd Edition, John Wiley and Sons, 1983.
24. Jollife, I.T.: Principal Component Analysis. Springer-Verlag, New York, 1986.
25. Turk, M., Pentland, A.: Eigenfaces for Recognition. Journal of Cognitive Neuroscience **3**(1), 71–86, 1991.
26. USF DARPA Human-ID 3D Face Database, Courtesy of Prof. Sudeep Sarkar, University of South Florida, Tampa, FL. http://marthon.csee.usf.edu/HumanID/
27. Lee, D.D., Seung, H.S.: Learning the Parts of Objects by Non-Negative Matrix Factorization. Nature **401**, 788–791, 1999.
28. Hyvaerinen, A., Oja, E.: Independent Component Analysis: Algorithms and Applications. Neural Networks **13**(4), 411–430, 2000.
29. Bartlett, M.S., Movellan, J.R., Sejnowski, T.J.: Face recognition by independent component analysis. IEEE Trans. on Neural Networks **13**(6), 1450–1464, 2002.
30. Cheng, C.M., Lai, S.H., Chang. K.Y.: A PDM Based Approach to Recovering 3D Face Pose and Structure from Video. In Proc. of Int. Conf. on Information Technology, 238–242, 2003.
31. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From Few To Many: Generative Models For Recognition Under Variable Pose and Illumination. In Proc. of IEEE Int. Conf. on Automatic Face and Gesture Recognition, 277–284, 2000.
32. Jin, H., Favaro, P., Soatto, S.: Real-Time Feature Tracking and Outlier Rejection with Changes in Illumination. In Proc. of ICCV, 684–689, 2001.

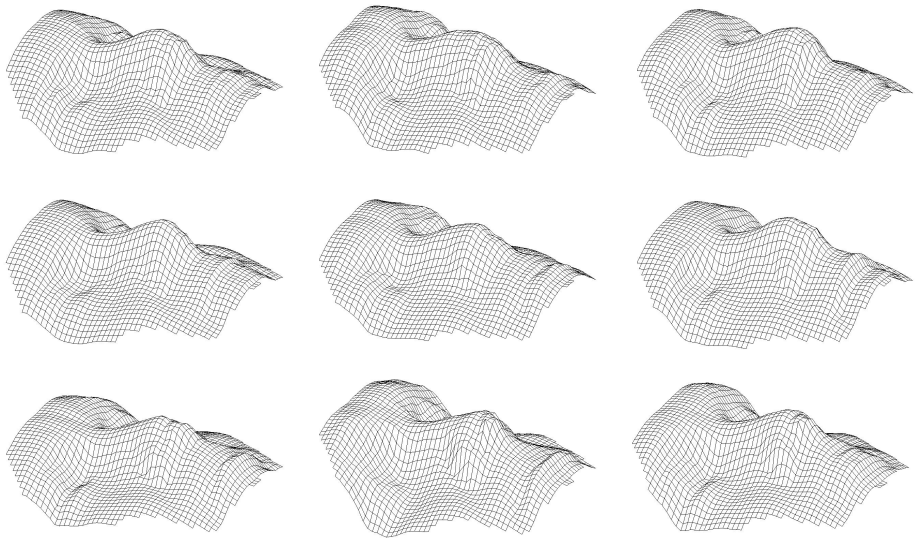


Fig. 3. Three columns correspond to three different faces from the Yale face database B. First row contains meshes of the faces with "ground truth" depths. Second row contains the meshes reconstructed by our algorithm from images with lighting $Az = 20^\circ$ and $El = 10^\circ$. Third row contains reconstructions from statistical Cartesian SFS algorithm.

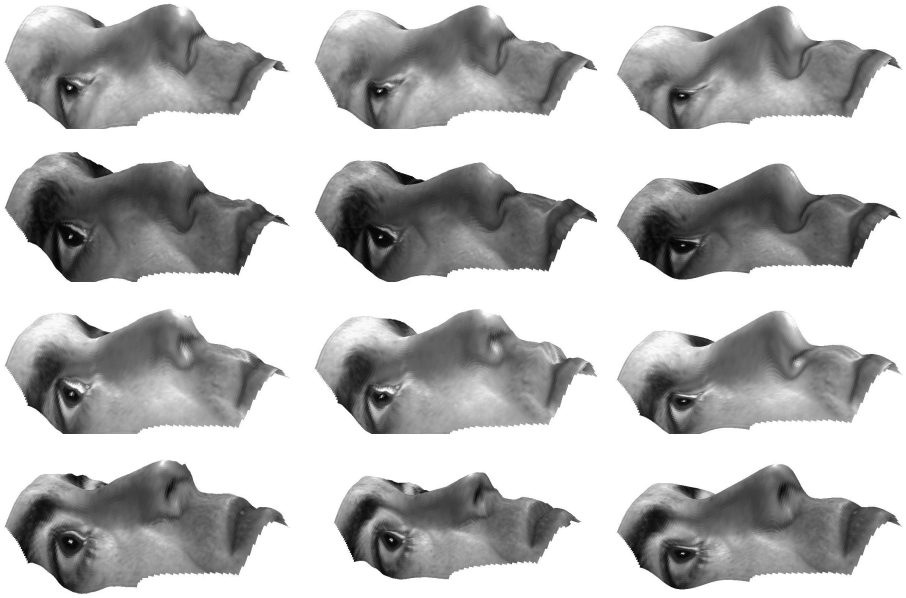


Fig. 4. Rows correspond to different Yale faces. First, second and third columns contain renderings of the faces with "ground truth", reconstructed and average depths, respectively. For the first three faces, texture matches rather well both "ground truth" and estimated depths, but poorly the average depth (mainly in the area between nose and mouth), indicating good shape estimation of our algorithm, at least for these faces.

Fig. 5. The four different rows correspond to four different faces from the Yale face database B. The first column contains renderings of faces with side illumination $Az = 20^\circ$ and $El = 10^\circ$. The second column contains images rendered from images in column 1 using the depth recovered by our algorithm. The faces in the second column should be similar to the frontally illuminated faces in column 3 (one should ignore shadows present in the rendered images, because such a simple cancellation scheme is not supposed to cancel them out). Finally, the last column contains frontally illuminated faces from column 3, flipped around their vertical axis. By comparing two last columns, we can see noticeable facial asymmetry, even in the case of the three best faces. For the fourth face asymmetry is rather significant, specially depth asymmetry near the nose, causing rather big errors in the reconstructed depth, as can be seen in Fig. 4.