

# Implementation of 3D Face Construction from a Face Image

Abdullah Taha Arslan<sup>1</sup>, Erol Seke<sup>1</sup>

<sup>1</sup> Eskişehir Osmangazi University  
{atarslan, eseke}@ogu.edu.tr

## Abstract

Shape reconstruction of an object from a two-dimensional image is an ill-posed problem in its nature where additional information is needed. Lately, researchers suggested an approach that a reference shape of similar objects is to be used during the reconstruction. This method introduces a promising alternative to regular shape from shading algorithms. Moreover, in face reconstruction problems, spherical harmonic expansion of the reflectance simplifies the solution further. In this paper, an iterative reconstruction algorithm is tested in order to determine optimal values of some parameters. Test results confirm that the algorithm is somewhat sensitive to searched parameters.

## 1. Introduction

Image processing approaches in three dimensions have been an increasingly popular research topic lately. In human face applications, it has been shown that 3D images bring several opportunities to overcome some problems related to conventional 2D images. It is not an easy task to directly obtain a 3D digital image of an individual at the moment. Therefore, many researchers focus on 2D to 3D reconstruction methods. One such obvious method is photometric stereo, where at least two images of the scene taken under different lighting conditions are needed in order to recover the depth information. With a morphable face model, the depth information of a given input image can be estimated iteratively on condition that a large database of such models are in hand [1]. Computational costs of such methods are making the approach unattractive in time-sensitive situations.

One another well-known set of methods, called shape from shading, first analyzed by B. K. P. Horn needs only a single input image [2]. Due to the ill-posed nature of this problem, the end results of vast majority of these applications are not satisfactory [3]. A recent approach utilizes the fact that group of objects in general and human faces in particular share common characteristics is proposed in [4]. Given an input face image to recover the depth map from, one another depth map of a reference face image is all that's needed. It has been shown simultaneously in [5] and [6] that Lambertian surfaces act like a low-pass filter and first and second-order spherical harmonics capture the vast majority of light that is reflected from the surface, hence simplifying the overall algorithm. In this paper, the findings from the work conducted in a similar fashion in [4] are presented. In Section 2, the problem at the hand is formulated. Our experimental setup and end results are discussed in Section 3. Conclusions and a roadmap for future work are presented in Section 4.

## 2. The Fundamentals of the Problem

Human faces reflect light approximately similar to Lambertian model in non-shiny situations. Under the assumption of orthogonal projection and a point source lighting, image brightness  $I$  is given by

$$I(x, y) = \rho(x, y)R(n(x, y)) = \rho(x, y)n(x, y) \cdot s \quad (1)$$

where  $\rho$ ,  $s$  and  $n$  represent the surface albedo, light source vector and surface normal, respectively, for a point at  $(x, y)$  on the image. Furthermore, radiance map  $R$  can be expanded in aforementioned spherical harmonics as

$$R(x, y) \approx \sum_{n=0}^N \sum_{m=-n}^n l_{nm} \alpha_n Y_{nm}(x, y). \quad (2)$$

$\alpha$  coefficients contribute as a low-pass filter for the equation and are set as  $\alpha_0 = \pi$  and  $\alpha_1 = \pi/\sqrt{3}$ .  $l_{nm}$  are the harmonics coefficients for the light and for the first order expansion there are 4 of them. Details regarding this expansion are investigated in [6] and [4]. For our purposes, for the first-order expansion if we let  $n = 1$ ,

$$Y(n(x, y)) = (c_0, c_1 n_x, c_1 n_y, c_1 n_z)^T \quad (3)$$

where  $c_0 = \sqrt{4\pi}$  and  $c_1 = \sqrt{3/4\pi}$ . Overall algorithm depends on the reference image for the values when they are missing from the 2D image, i.e. values for the albedo ( $\rho_{ref}$ ), and the surface normals ( $n_{ref}$ ). Firstly, the light coefficients  $l$  are estimated by minimizing the following equation:

$$\sum_{x, y \in \Omega} I(x, y) - \rho_{ref} \cdot l^T \cdot Y(n_{ref}(x, y)). \quad (4)$$

Once  $l$  coefficients are found, depth information can be recovered by minimizing

$$\int_{\Omega} (I - \rho_{ref} l^T Y(n))^2 + \lambda_1 (\Delta G * d_z)^2 dx dy. \quad (5)$$

In (5), the first and second part of the equation are the data term and the regulation or the smoothing part, respectively. If  $z(x, y)$  is the depth value for a pixel to be found and  $z_{ref}(x, y)$  is the reference depth value then  $d_z$  is the difference between these two. One can recover either a surface normal or the depth since finding one is equivalent to finding the other as they are related by:

$$n(x, y) = \frac{1}{\sqrt{p^2 + q^2 + 1}}(p, q, -1)^T \quad (6)$$

where

$$p = z(x + 1, y) - z(x, y) \quad (7)$$

and

$$q = z(x, y + 1) - z(x, y) \quad (8)$$

The regulation part is implemented as a set of equations of

$$\lambda_1(z(x, y) - G * z(x, y)) = \lambda_1(z_{ref}(x, y) - G * z_{ref}(x, y)) \quad (9)$$

$\lambda_1$  is a control scalar that determines how large the smoothing factor would be. A larger value is expected to let the recovered depth map be closer to the depth information of the reference face. In the last part, with  $l$  and the surface normals  $n$  in hand, the albedo of the 2D can be estimated in a similar way by minimizing the equation:

$$\int_{\Omega} (I - \rho l^T Y(n))^2 + \lambda_2(\Delta G * d_{\rho})^2 dx dy \quad (10)$$

Again, the second part of the equation is the regulating term and this time it satisfies the set of equations:

$$\lambda_2 \Delta G * \rho = \lambda_2 \Delta G * \rho_{ref}. \quad (11)$$

The algorithm is implemented using reference models from The Texas 3D Face Recognition Database [7]. The database consists of portrait face images of individuals and corresponding depth maps. The albedo map of the reference image can be estimated once the light source vector is known. Several algorithms for recovering the light source from known depth and a 2D image have been proposed. One straightforward and efficient method can be found in [8]. With known surface normals  $\mathbf{n}_{ij}$  light source vector for an image is given by

$$\mathbf{s}^k = \left[ \sum_{i,j \in \Omega} \mathbf{n}_{ij}^k \mathbf{n}_{ij}^{kT} \right]^{-1} \sum_{i,j \in \Omega} I_{ij} \mathbf{n}_{ij}^k \quad (12)$$



**Figure 1.** Albedo outputs from Texas database for 3 individuals. On each row, the images are in the order of the portrait, the shape and the recovered albedo map

Once  $\mathbf{s}$  is calculated,  $\rho$  can be found using (1) easily. Examples of output albedo maps are shown in Fig. 1.

### 3. Experiment

All experiments were conducted in MATLAB environment. In any step of the algorithm, an equation for each pixel contribute towards a set of linear system of  $Ax = b$  and the system is solved to recover the unknown values. For the depth and albedo estimation parts,  $A$  is a large sparse matrix that has as many columns as the number of pixels. Once the data term is constructed, the coefficients of the regulating term's Gaussian kernel are added to the corresponding pixel locations in the matrix. UMFPAK from SuiteSparse library ([9]) has also been tested but MATLAB's internal solver produced very similar results in a shorter time.

The average of 116 individuals' images were used for the reference face image (Fig. 2). Any input test image is aligned with the reference image by manually marking 4 fiducial points on the face. Alignment is conducted by a geometric transformation process between two images. Before starting the main process, all the input assets, the test image, the depth map and the portrait, as well as the estimated albedo map of the reference are normalized. After that, the normal vector for each pixel is calculated according to (6), (7) and (8).



**Figure 2.** Average of 116 individuals in Texas face database: Gray-scale portrait image and the shape

### 4. Results and Conclusion

An image for each of the 116 individuals from the database were selected for ground-truth testing since a corresponding depth map is at the hand. Average error for each of these were calculated in a similar fashion given in [4]. The mean and standard deviation of error of 116 reconstructed images were found to be 8.76 and 4.49 percent respectively. Three outputs from the database can be seen in Fig. 4. Three independent faces can be seen in Fig. 5 for visual testing.

Regions of a human face that exhibit rapid changes are the least successfully reconstructed parts. These include the nose tip and around, eyes and the mouth regions. The input image needs to be in neutral position and be recorded while the person facing the camera for a more realistic result. The boundaries are also areas of interest that contribute to the overall error. Misalignment between test and reference images significantly degrades the outcomes. The regulation or the smoothing factor  $\lambda_1$  is another important parameter that defines how closer the result is to the reference. This outcome can be seen in Fig. 3. Smaller values make the characteristics of test image be more prominent albeit significant degradations. An optimal value for the trade-off can be determined with an algorithmic approach, which is planned to be studied. One another crucial parameter is the Gaussian kernel size and its sigma value, that should be investigated further.

The mean reference model does not yield satisfactory results for all input images. The most suitable candidate for a



**Figure 3.** Regulation or smoothing coefficient  $\lambda_1$  affects the algorithm significantly. Larger values force the recovered shape to resemble more like the reference while smaller values preserve individual characteristics with more degradation. In this figure, from top-left to bottom-right, reconstructed shapes for the same input test image using the same reference depth are shown for  $\lambda_1$  values 3, 5, 10, 15, 20, 25 and finally the reference shape.

reference depth map is obviously the most resembling one to the test image. This may be achieved with a pool of reference models to be analyzed with PCA or a similar method and with a automatic selection for the test image. Another possible approach that may improve overall algorithm performance is a locally adaptive selection and iteration process that flows differently on smooth and complex parts of the input image.

For future works, boundary conditions will be investigated further. The second order spherical harmonics expansion are to be compared with these results. A question of whether another input image of the same individual in stereo can be incorporated into the system also arises.

## 5. References

[1] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1999, pp. 187–194.

[2] B. K. Horn, *Obtaining shape from shading information*. MIT press, 1989.

[3] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah, "Shape-from-shading: a survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21, no. 8, pp. 690–706, 1999.

[4] I. Kemelmacher-Shlizerman and R. Basri, "3d face reconstruction from a single image using a single reference face shape," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 2, pp. 394–405, 2011.

[5] R. Ramamoorthi and P. Hanrahan, "On the relationship between radiance and irradiance: Determining the illumina-



**Figure 4.** End results of the algorithm. The images are from Texas database and in the order of the portrait, the ground truth shape, the recovered shape without texture and with texture mapped.



**Figure 5.** Three general examples of arbitrary individuals

tion from images of a convex lambertian object," *J. Optical Society of America A*, vol. 18, no. 10, pp. 2448–2459, 2001.

[6] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 2, pp. 218–233, 2003.

[7] S. Gupta, K. R. Castleman, M. K. Markey, and A. C. Bovik, "Texas 3d face recognition database," in *Image Analysis & Interpretation (SSIAI), 2010 IEEE Southwest Symposium on*. IEEE, 2010, pp. 97–100.

[8] M. J. Brooks and B. K. Horn, "Shape and source from shading," in *Proceedings of the 9th international joint conference on Artificial intelligence-Volume 2*. Morgan Kaufmann Publishers Inc., 1985, pp. 932–936.

[9] T. A. Davis, "Algorithm 832: Umfpack v4. 3—an unsymmetric-pattern multifrontal method," *ACM Transactions on Mathematical Software (TOMS)*, vol. 30, no. 2, pp. 196–199, 2004.