An Automated Method for Human Face Modeling and Relighting with Application to Face Recognition

George Toderici, Georgios Passalis, Theoharis Theoharis, and Ioannis A. Kakadiaris
Computational Biomedicine Lab
Department of Computer Science, University of Houston
4800 Calhoun Road
Houston, TX, 77004, USA
http://www.cbl.uh.edu/URxD

Abstract

In this paper, we present a novel method for human face modeling and its application to face relighting and recognition. An annotated face model is fitted onto the raw 3D data using a subdivision-based deformable model framework. The fitted face model is subsequently converted to a geometry image representation. This results in regularly sampled, registered and annotated geometry data. The albedo of the skin is retrieved by using an analytical skin reflectance model that removes the lighting (shadows, diffuse and specular) from the texture. Additional provisions are made such that if the input contains over-saturated specular highlights, an inpainting method with texture synthesis is used as a post-processing step in order to estimate the texture. The method is fully automatic and uses as input only the 3D geometry and texture data of the face, as acquired by commercial 3D scanners. No measurement or calibration of the lighting environment is required. The method's fully automatic nature and its minimum input requirements make it applicable to both computer vision applications (e.g., face recognition) and computer graphics applications (i.e., relighting, face synthesis and facial expressions transfer). Moreover, it allows the utilization of existing 3D facial databases. We present very encouraging results on a challenging dataset.

1. Introduction

The human face, as the most distinctive and descriptive human feature, has been widely researched in both the computer vision and computer graphics domains. With the proliferation of 3D scanners, 3D facial data are used in biometrics, motion pictures, games and medical applications.

However, the use of real facial data introduces a number of challenges. The texture acquired by most commercial 3D scanners is affected by the lighting conditions. In applications such as biometrics or face relighting the skin albedo is required. Therefore, the contribution of environmental lighting must be removed from the texture. Moreover, commercial 3D scanners produce 3D data that have artifacts and non-uniform sampling. Manual cropping is required to remove data that do not belong to the face. Finally, there is no common point of reference between facial scans even of the same person, as these data are unregistered.

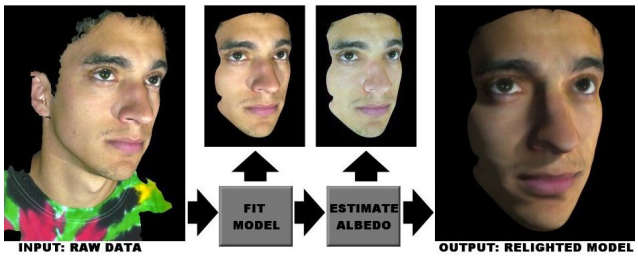


Figure 1: Overview of the method (left to right): Input raw data → Fitted model → Estimated albedo → Relighting example.

In this paper, we propose a novel automated method that tackles all of the above challenges and requires as input only 3D facial data and the corresponding texture (Fig. 1). Using a subdivision-based deformable model framework we fit an Annotated Face Model (AFM) to the raw data and export a geometry image [16]. This representation holds both geometry and texture information in the same parameter space. An Analytical Skin Reflectance Model (ASRM) is then applied to the texture to remove the lighting, thus acquiring the albedo of the face. It is then shown that the derived processed data can be used directly in a variety of applications, including Face Recognition.

Compared to other approaches, the proposed method offers a unique combination of features: **1) Fully automatic**. No user intervention is required in any step of the method. **2) Minimum input requirements**. Only 3D data and texture are used, no light calibration is needed. **3) Compact representation**. The facial geometry is regularly sampled, registered and annotated. The extracted geometry image can be used directly in modeling applications. **4) Arbitrary lighting**. Lighting is removed from the texture. The albedo can be used directly in rendering applications. **5) Shadows**. The albedo can be extracted even when self-shadowing and multiple lights are present. **6) Highly efficient**. The method utilizes the GPU for lighting calculations thus achieving high efficiency. **7) Widely applicable.**. The method can be applied to any facial dataset either from commercial 3D

scanners or existing databases, thus allowing its utilization in Face Recognition.

2. Previous Work

Modeling of the human face and skin is a widely researched topic. An in-depth overview of the issues related to modeling the skin is given by Igarashi *et al*. [12]. Existing approaches can be divided into two main categories: data-driven and model-driven. Data-driven approaches require additional information to a 3D facial scan. They measure the reflectance function using specialized hardware (such as a light dome or polarizing filters). Model-based approaches use an approximation of the reflectance model of the skin, usually in the form of a Bidirectional Reflectance Distribution Function (BRDF) or a more complex bidirectional surface scattering reflectance distribution function (BSSRDF).

Data-driven approaches include that of Debevec *et al*. [5] who pioneered the human skin reflectance capture *in vivo*, and introduced the use of a light dome for this purpose. Weyrich *et al*. [43] also proposed the use of a light dome along with a subsurface scattering measurement tool in order to model the human skin reflectance for the facial region. For an analytical form of the BRDF, they fit a Torrance-Sparrow or a Blinn-Phong local reflectance model to the data. For the BSSRDF, the fitting process uses the model proposed by Jensen *et al*. [14]. Even though these approaches present truly photorealistic results, they require a large amount of information as input. In most applications this additional information is not available.

Approaches with smaller information requirements that utilize polarization filters for obtaining the albedo include that of Nayar *et al*. [26]. The technique requires two images of the same scene, one taken without any filters, and one taken using a polarized filter. The difference of the two images provides the specular reflection of the object. The major drawback is that the images must be taken from the same viewpoint, and special hardware (the filters) must be employed. Zickler *et al*. [44] extended the above technique by incorporating the constraint that the shape must be known in advance. By using the extra information, they extract the BRDF specific to the input and the albedo.

Using a PDE approach, Mallick *et al*. [22] recovered the specular component from images, without the use of polarization filters. The results presented are convincing, but the approach has high computational complexity and does not take into consideration the actual surface of the face.

Ikeuchi and Sato [13] proposed to use the Torrance-Sparrow BRDF to model the diffuse and specular components of the light illuminating the face. They model the diffuse component using the Lambertian formulation. The parameters are estimated using an iterative least squares fitting, and the distinction between specular, diffuse, and shadow pixels is accomplished using a threshold, which must be set manually.

Model-based approaches include that of Paris *et al*. [30] who proposed a lightweight approach based on a simple Phong BRDF that uses graphics hardware. Compared to our method, their approach has three limitations: it requires a light probe of the environment, annotation must be performed manually, and it does not handle self-shadowing. In this paper, we show that shadows can be successfully removed without any artifacts using an efficient approach with even lower input information requirements.

Blanz and Vetter [3] employed a morphable model technique in order to acquire the geometry of faces from 2D images. The albedo is captured in the fitting process. The main drawback, however, is that they need to manually initialize the morphable model for each input image. In contrast to our method, this is a statistical approach. The descriptiveness of statistical approaches depends heavily on the variety and quality of the training set and the creation of such sets is not a trivial task. In [7] they focus their method in computing the reflectance of the face and they also demonstrate BRDF transfer. Even though this work does not require a 3D model, they take several 2D images with constrained lighting conditions. The actual amount of information used is comparable to that of the proposed approach.

Smith and Hancock [38] presented an approach for albedo estimation from 2D images which uses a 3D morphable model that is fitted to the input image. The normals of the fitted model are then used for the computation of the shading, assuming a Lambertian reflectance model. The main limitation of their method is the assumption that the subject is illuminated by a single light source which is placed very close to the viewer [39]. Georghiades [9] also presented a method where only 2D images are used in order to compute the skin BRDF and 3D shape. This work is limited to greyscale images and does not handle self-shadowing.

The method proposed in this paper does not challenge data-driven approaches in terms of photorealism. When minimal input data are available, only model-based approaches are applicable. The proposed method, having significantly fewer constraints and limitations than previous approaches, widens the applicability of such methods.

3. Methods

In most practical applications only the 3D data and texture of a face are available. These are either captured by a commercial 3D scanner or found in existing facial databases. Therefore, a method that aims to be widely used for face modeling and relighting cannot rely on additional information (e.g., light probes, reflectance measurements). Moreover, for applications where a large number of faces must be processed (such as Face Recognition), full automation is extremely important. The proposed method provides both. The aim is to derive high quality geometry and texture from the minimal input data that can be directly used in various applications (Sections 4 and 5). We first register and fit the Annotated Face Model (AFM) using a deformable model framework [16]. Then, the geometry is converted into a geometry image representation that automatically provides annotation, registration and regular sampling. Our contribution, which extends previous work, is removing the lighting from the texture using an Analytical Skin Reflectance Model (ASRM), thus acquiring the albedo. Note that the proposed method is completely different from the morphable model approach of Blanz [3]. In

that approach, the morphable model is statistical, and the fitting process is optimizing the model's parameters. In our method, the deformable model is strictly geometric, and the fitting process deforms the model's surface iteratively.

3.1. Annotated Face Model

Using the work of Kakadiaris *et al*. [16], the AFM defines the control points of a subdivision surface and it is annotated into different areas (e.g, mouth, nose, eyes). These facial areas have different properties associated with them which are used by our method. For example, the mouth area, is considered less rigid than the nose area in the fitting step, whereas in the albedo estimation step, the mouth area has a decreased specular coefficient compared to the nose area.

A continuous global UV parametrization has been applied to the AFM. The specific parametrization is essentially a mapping form R^3 to R^2 and allows the conversion of the AFM from a polygonal representation to an equivalent geometry image representation. Geometry images [10, 15, 17, 34] are regularly sampled 2D images that have three channels, encoding geometric information (x, y and z coordinates of a vertex in R^3). In this paper, the number of channels in the geometry image is greater than three, as apart from geometric information we also encode texture and annotation. Also, for practical purposes, in all applications we used a resolution of 512×512.

3.1.1 Registration

Facial data acquired from 3D scanners have arbitrary orientation. Before the AFM is fitted to these data, both must have the same orientation. To this end, we employ the registration strategy of [16] that is two-phase *rigid* registration step. The first phase, that provides a rough registration, utilizes the Iterative Closest Point (ICP) algorithm [2] and it does not require manual initialization. The second phase, which provides a finer registration, utilizes the registration algorithm presented by Papaioannou *et al*. [29]. It computes depth images of both the AFM and the raw data and uses a non-linear optimizer [18, 36] to minimize the discrete sum of differences of the depth images' derivatives. Simulated Annealing minimizes the following objective function:

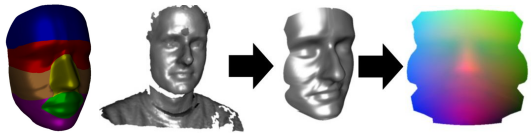
$$\varepsilon_\omega = \sum_{i=1}^{S}\sum_{j=1}^{S} |D_x^1(i,j) - D_x^2(i,j)| + |D_y^1(i,j) - D_y^2(i,j)|$$

where D_y^1, D_y^1, D_x^2, D_y^2 are the X and Y derivatives of the depth images of the AFM and the raw data respectively, while S is their spatial resolution.

Note that ICP is less sensitive to initial conditions while SA offers more invariance to facial expressions and finer registration. After this step, there is a correspondence between the main facial features of the AFM and the raw data.

3.1.2 Deformable Model Fitting

In order to fit the AFM to the raw data, a subdivision based deformable model framework [16] is used. When the deformation concludes, the AFM acquires the shape of the raw data. This establishes a dense correspondence between the AFM's surface and the raw data's vertices. Additionally, since the deformation has not violated the properties of the original AFM, the deformed AFM can be converted to a geometry image. The extracted geometry image encodes the geometric information of the raw data (Fig. 2). Subsequently, the texture is converted to the same representation, by assigning to each pixel the UV coordinates of the closest point of the raw surface. Note that the deformable model framework discards data not belonging to the face and successfully handles artifacts (Fig. 2) without any special preprocessing.



(a) (b)

Figure 2: (a) Annotated Face Model (b) Raw data → Fitted AFM → Resulting geometry image.

The fitting framework [16], is an implementation of the deformable models [24] using subdivision surfaces [23]. We selected the Loop subdivision scheme [21] since it produces a limit surface with C^2 continuity while only 1-neighborhood area information is needed for each vertex. The AFM is used as the subdivision surface's control mesh, thus determining the degrees of freedom, while the limit surface is used to solve the following equation:

$$\mathbf{M}_q \frac{d^2\overrightarrow{q}}{dt^2} + \mathbf{D}_q \frac{d\overrightarrow{q}}{dt} + \mathbf{K}_q \overrightarrow{q} = f_q$$

where $\overrightarrow{q}$ is the control points vector, $\mathbf{M}_q$ is the mass matrix, $\mathbf{D}_q$ is the damping matrix, $\mathbf{K}_q$ is the stiffness matrix, and f_q are the external forces. The equation is solved iteratively. During this process the AFM gradually acquires the shape of the raw data.

3.2. Albedo Estimation

We used data acquired by two of our 3dMD^{TM} scanners and data from the Face Recognition Grand Challenge 2.1 database [31]. The 3dMD^{TM} systems we used were configured either with one flash, or with two flashes for testing the feasibility of the algorithm in the presence of two simultaneous shadows (Fig. 3). These flashes tend to be the major source of illumination. Therefore, we model the lighting conditions of the acquisition using one or two point lights, but the model can easily be extended for arbitrary point lights. We assume that the flashes produce pure white light.

3.2.1 Analytical Skin Reflectance Model

We chose to use a hybrid BRDF to model skin reflectance. We did not utilize a BSSRDF model since the input data did not have enough resolution in order to estimate a subsurface scattering component. The Analytical Skin Reflectance Model uses the Oren-Nayar BRDF to model the diffuse component and the Phong BRDF to model the specular component.

The Oren-Nayar [28] diffuse component at a surface point, where the incident light angles are denoted by (θ_i, ϕ_i)

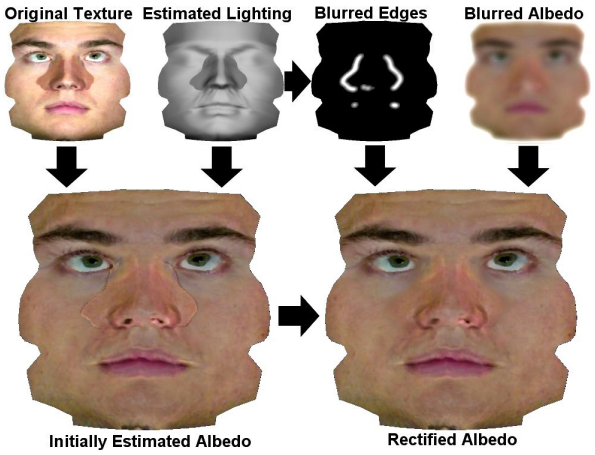


Figure 3: Overview of albedo estimation.

and the reflected light angles by (θ_r, ϕ_r), is given by the following equation:

$$I'_d = \frac{\rho}{\pi} E \cos\theta_i \left(A + B \max\left[0, \cos\left(\phi_r - \phi_i\right)\right] \sin\alpha \tan\beta\right)$$

where $A = 1.0 - 0.5\frac{\sigma^2}{\sigma^2+0.33}$, $B = 0.45\frac{\sigma^2}{\sigma^2+0.09}$, σ denotes the surface roughness parameter, ρ is a parameter related to the microfacet distribution, $\alpha = \max(\theta_i, \theta_r)$ and $\beta = \min(\theta_i, \theta_r)$. Note that ρ is usually omitted, as it can be included in the diffuse color of the surface.

The Oren-Nayar BRDF does not take into account the specular reflections caused by the oily layer of the skin. To accommodate this we use the BRDF proposed by Phong [32]. The intensity of the specular reflection at a surface point is: $I_s = E \cos^n \phi$, where ϕ is the angle between the view vector and the reflected light and n is a parameter that controls the size of the highlight. Note that each facial area has different specular properties, therefore we utilize a specular map based on the annotation of the AFM.

The above equations describe an ASRM for a single point light. In order to model multiple point lights, the contribution of each light's ASRM must be summed.

3.2.2 Parameter Estimation

In order to estimate the albedo we need to find the optimum parameters of the ASRM along with the position of the light(s). In our implementation we use OpenGL and Cg [27], a high-level GPU programming language. The ASRM is implemented as a Cg shader and for self-shadowing the shadow mapping technique of [6] is used. The implementation fully utilizes the GPU, thus achieving high efficiency (it takes less than a minute in a modern PC). Note that the textures used are represented in geometry image space.

The texture M_T is the result of the lighting applied on the unknown albedo M_A and is given by:

$$M_T = I_s + (I_d + I_a) \cdot M_A$$

where I_a is the ambient component. By solving this equation for the albedo, we get: $M_A = \frac{M_T - I_s}{I_d + I_a}$. In our implementation, we use the HSV color space, and we apply the albedo estimation only on the V component. We assume that in the input image there is enough information to estimate the hue and saturation. If that is not the case, we can handle this case in the post-processing step.

The optimum ASRM parameters are found using Simulated Annealing. The parameter vector has two parameters for the light position which consist of two angles on a sphere centered around the centroid of the object, two parameters for the diffuse component, and two for the specular component. Other parameters such as the ambient component, or the gamma correction of the texture have static values and are left outside the optimization. The objective function minimizes the intensity variation in M_A. To this end, M_A is divided into small areas and for each area we sum the differences between the individual pixels and the mean intensity value of that area. The value of the objective function is the normalized sum of the local area differences. An example of M_T, M_A as well as the estimated lighting is provided in Fig. 3(left).

3.2.3 Post-processing Rectification

The last step in the albedo estimation process is the post-processing rectification of the albedo. The purpose of this rectification is to remove artifacts caused by the lighting estimation. A visual inspection of the initially estimated albedo of Fig. 3(bottom-left) reveals artifacts on the outline of the shadow of the nose. This issue is not attributed to the fact that shadow mapping produces hard shadows. Even if soft shadows are used, certain artifacts can appear in areas where the light estimation has significant intensity variation (or discontinuities). These artifacts do not affect the result of the optimization but are very noticeable to the human eye.

To rectify this, we first detect these areas using an edge detector on the estimated lighting image. A Gaussian kernel is then applied to retrieve a blurred edge image M_E. The same Gaussian kernel is applied on the albedo to acquire a blurred albedo M_B. Both of these images are depicted in Fig. 3(top-right). The rectified albedo M_R is given by:

$$M_R(i,j) = (1 - M_E(i,j)) \cdot M_A(i,j) + M_E(i,j) \cdot M_B(i,j)$$

for every image pixel (i, j), where M_A is the initially estimated albedo. The rectified albedo uses information from the blurred albedo only in areas where artifacts are expected to appear (with discontinuities in the intensity). The artifacts are thus removed (Fig. 3, bottom-right). From a quantitative point of view the contribution of this step is minimal, however the qualitative contribution is significant. Without this step the albedo of a self-shadowed face is not plausible visually, a problem commonly encountered in other methods [30]. In their work Lensch *et al.* [20] also tackled the shadow outline problem, but with a different implementation, as they blurred the shadow map.

If the input image did not have enough resolution to be able to accommodate the full color gamut of the person being imaged (e.g., a strong flash was used and the person had an oily face, combined with a low dynamic range camera), we use an inpainting approach with texture synthesis [11]. We apply it on the areas of the albedo that have an

estimated specular intensity greater than a threshold (e.g., 0.1), and we use the neighboring pixels as the source for the texture synthesis (search radius of 40 pixels on 512x512 albedo images). Figure 4 depicts the application of the full pipeline. We also employ a mask defined in the UV space of the model (and thus we need to define this mask only once) which prevents the inpainting to be applied on areas of the face which we know in advance not to have specular components (e.g., eyes). This approach is similar to that of Tan *et al*. [40] in the sense that we also use inpainting, but the key difference is that we use the specular light intensity in order to estimate the regions which need to be inpainted. The annotation of the face in areas with different properties is an inherent advantage of the deformable model fitting.

4. Validation

In order to provide a quantitative validation of parameter fitting for the albedo, we performed two tests on synthetic cases for the Lambertian model: 1) we rendered a hemisphere with a known texture under various light conditions, and 2) we rendered a fitted AFM with the same texture we used for the hemisphere. We used these as input to our light estimation algorithm. The mean squared error for the albedo estimation (Fig. 8) of the hemisphere was 1.07%, while for the face it was 2.01%. In both cases, we used the same parameter limits as in our face recognition experiment.

Additionally, in Fig. 9 we present a subject as imaged under three illumination conditions. We computed that the mean standard deviation in the texture (a-b) was 11%. On the albedo, however the mean standard deviation was reduced to 4%.

Relighting: Processed datasets can be relit using any rendering algorithm. A dataset that was acquired using a two-flash optical scanner is depicted in Fig. 5(a). The raw texture had shadows in the area of the nose which were successfully removed from the albedo. This allows relighting under arbitrary lighting conditions, as depicted in Figs. 5(b,c). A dataset that was acquired using an one-flash optical scanner is depicted in Fig. 5(d). The raw texture had a strong specular component which was removed, as depicted in Figs. 5(e,f).

Texture swap: All processed datasets have co-registered geometry and texture data. This allows us to directly transfer the texture from one person to another. An example is shown in Fig. 6. The only limitation is that the two persons must have the same facial expression.

Face Synthesis: Data registration along with the geometry image representation also allows us to synthesize new faces through the linear interpolation of existing ones. Fig. 6 depicts a synthetic face (e) by blending between faces (a) and (b) with equal weights.

Facial Expressions Transfer: Finally, transferring of facial expressions is possible as shown in Fig. 7. By using neutral expression datasets for two persons (a,c) and one with facial expression for the first person (b) we can transfer this expression to the second person (d). This is achieved by adding to the neutral dataset of the second person the difference between the two datasets of the first person.

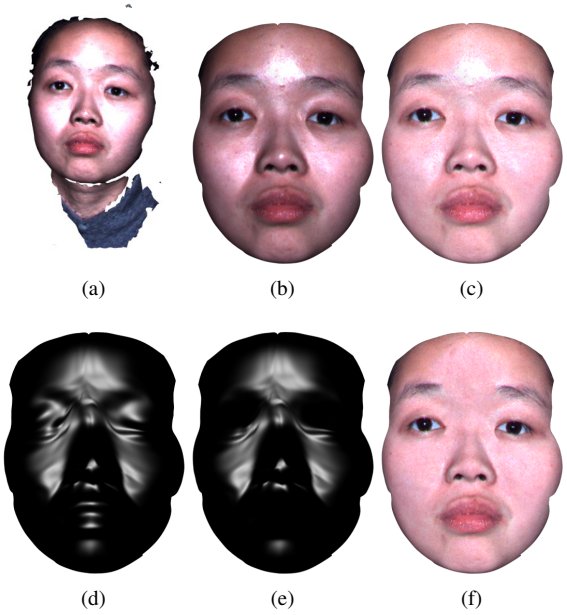


Figure 4: Handling of over-saturated images: (a) Input mesh acquired from a 1-pod active stereo system; (b) Fitted model; (c) Estimated albedo; (d) Corresponding estimated specular component; (e) Specular components restricted to the areas known to have specular components in the UV space of the model (e.g., we do not want to have a high specular response in the eye region or the mouth); (f) Inpainting with texture synthesis [11] of the masked specular component and blending with the estimated albedo.

5. Application to Face Recognition

Our method's automatic nature allows its application to the Face Recognition domain. In this paper, we do not propose a novel Face Recognition approach; this is beyond the scope of this paper. Our goal is to show that our method can improve the accuracy of existing face recognition approaches. For this, we constructed a database of 3D and 2D facial data. We utilize the 3D data only to estimate the albedo. We then use only the albedo to perform 2D Face Recognition. In order to compare the 2D images we utilize a pyramid wavelet transform. This approach is not the state-of-the-art in 2D Face Recognition but it can be used to evaluate the importance of the illumination removal as well as the differences between various BRDFs. An excellent survey of the 2D/3D face recognition literature is provided by Bowyer *et al*. [4].

Database A: The database we constructed for the Face Recognition experiment consists of 637 datasets from 229 subjects. Each dataset includes texture and geometry information. A 3dMD^{TM} optical scanner with a single flash was used. The acquisition conditions were not controlled, thus there is a wide range of variation both in the 2D textures and the 3D meshes. Moreover, the database contains datasets with a wide array of facial expressions. The first dataset of each subject was used as *gallery* while the rest were used as *probes*. For all subjects in the database the minimum number of probes is 1 and the maximum is 4. Even though this

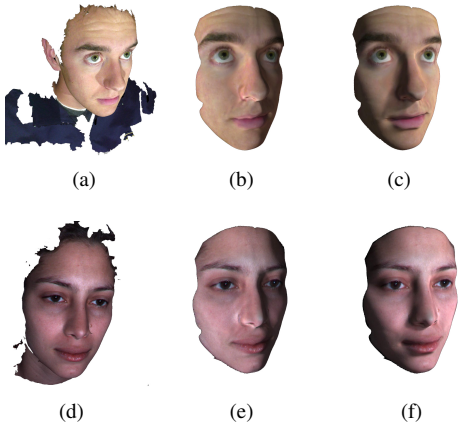


Figure 5: Relighting for two different persons: (a,d) Raw geometry and texture; (b,c,e,f) Processed geometry and texture under varying lighting conditions.

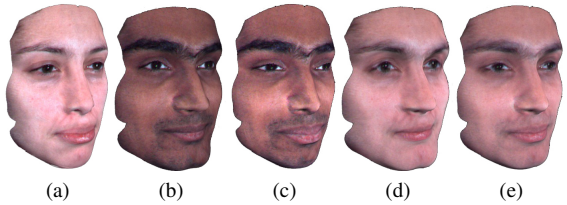


Figure 6: Texture swap between real faces (a) and (b) resulting in (c) and (d). Face synthesis of real faces (a) and (b) resulting in synthetic face (e).

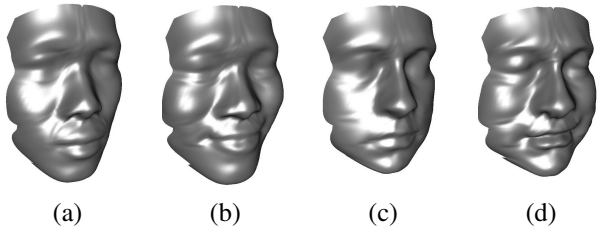


Figure 7: (a,b) Neutral and smiling datasets of first person; and (c,d) Neutral and synthesized smiling datasets of the second person.

database is limited in terms of size, we consider it challenging due to the uncontrolled acquisition conditions and the poor quality of most 3D meshes.

Database B: This database was created to evaluate the performance of the albedo estimation. We used a 3dMD^{TM} dual-pod system, to which we added a Canon Rebel XTi SLR camera in order to be able to capture varying lighting conditions. We enrolled 10 subjects and we acquired 10 images under 3 distinct illumination conditions (directional light coming from the front, 30°, and 60° orientation) in order to simulate the acquisition conditions of a laser scanner (they do not tend to use a flash like the 3dMD^{TM} system does). Fig. 9 depicts three datasets which are representative of this database. We used the first image with frontal illumination as the gallery for each subject.

Distance Metric: We first apply a wavelet transform to

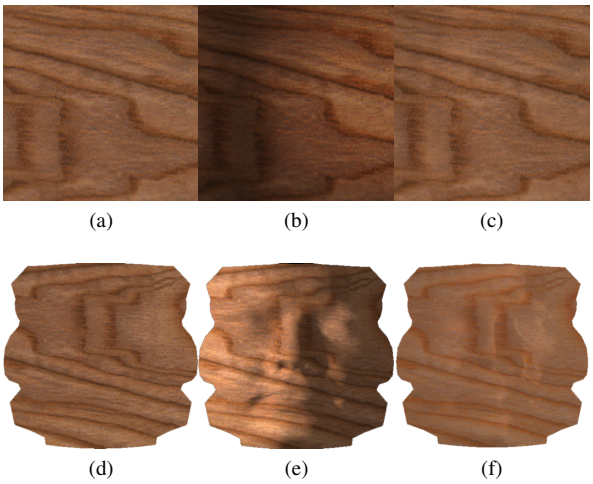


Figure 8: Albedo estimation validation using a hemisphere and a face rendered with a wood texture and lit by one Lambertian light. Albedo used for the hemisphere (a) and face (d). "Baked" wooden texture corresponding to the hemisphere (b) and face (e). Estimated albedo for hemisphere (c) and face (f).

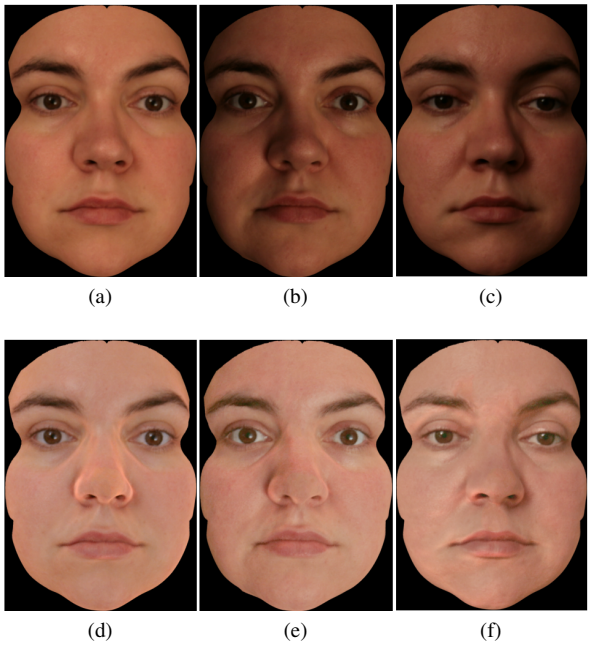


Figure 9: Comparison between the textures and estimated albedo for three illumination conditions. (a-c) Texture rendered on fitted AFM, and (d-f) Estimated albedo.

the albedo images. This transform decomposes the images using the complex version [33] of the steerable pyramid transform [37], a linear multi-scale, multi-orientation image decomposition algorithm. The image is first divided into highpass and lowpass subbands. The lowpass subband is then fed into a set of steerable bandpass filters, which produce a set of oriented subbands and a lower-pass subband.

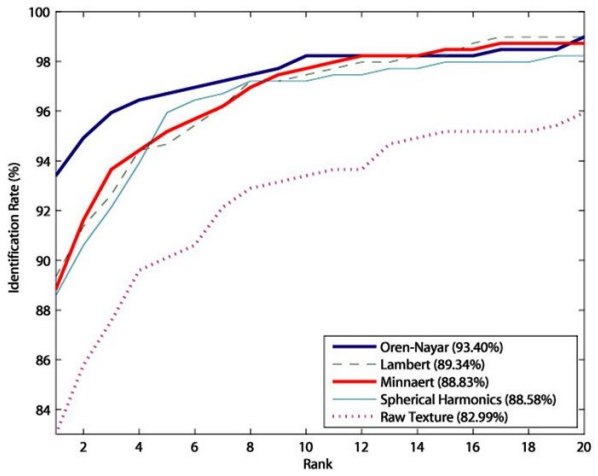


Figure 10: CMC graphs for various BRDFs on Database A.

This lower-pass subband is subsampled by 2 and recursively fed to the same set of steerable bandpass filters. Such pyramid wavelet representation is translation-invariant and rotation-invariant. To maintain reasonable image resolution and computational complexity our algorithm applies a 3-scale, 10-orientation complex steerable pyramid transform to decompose each component of the image. Only the oriented subbands at the farthest scale are stored, allowing us to compare the subband coefficients of the two images directly without the overhead of reconstruction. These coefficients are compared using the Complex Wavelet SSIM (CW-SSIM) index algorithm, a translational insensitive image similarity measure inspired by the structural similarity (SSIM) index algorithm [41]. CW-SSIM iteratively measures the similarity indices between the two sliding windows placed in the same positions of the two images and uses the weighted sum as a final similarity score. This score is measured by a variation of the CW-SSIM index equation originally proposed by Wang and Simoncelli [42].

Identification Experiment: Using the gallery/probe division of our database, we performed an identification experiment. The performance is measured using a Cumulative Match Characteristic (CMC) curve. Note that only the 2D data are compared using the distance metric. The 3D data were not used, since the experiment's purpose is to highlight the importance of illumination removal in 2D Face Recognition. Neither Database A, nor Database B contain oversaturated images, so the synthesis step was not used in our experiments.

We measured the performance using the albedo estimated with our BRDF and using the raw texture. Additionally, we measured the performance using albedos estimated with other BRDFs, like the Lambertian [19], the Minnaert [25], and Spherical Harmonics [35]. In all BRDFs we used exactly the same Phong specular component. Note that the Spherical Harmonics BRDF required significantly more parameters to optimize compared to the other BRDFs. A total of 27 parameters were needed even though we used only the first two degrees of the Spherical Harmonics.

In Fig. 10 it is shown that the albedo estimated with the Oren-Nayar BRDF resulted in the highest rank-one recognition rate of 93.4% on Database A. Using albedos estimated with other BRDFs yielded lower recognition rates; 89.34%, 88.83% and 88.58% for the Lambertian, Minnaert, and Spherical Harmonics, respectively. The use of the raw texture resulted in a recognition rate of 82.99%. Even though these results depend on the database used, the approximately 10% performance increase between the best albedo and the raw texture shows that Face Recognition can be significantly benefited by the proposed method. We also investigated the performance of PCA, and PCA+LDA [1] after Multiscale Retinex (MSR) [8] on this database[1]. However, due to the small number of images available for training, the results were extremely poor: 41.44% for PCA+LDA, and 40.31% for PCA. We used Database B as training.

On Database B, our face recognition approach achieved 95.72% rank-1 recognition rate when using the textures. When applied on albedos estimated with the Lambertian and Oren-Nayar formulations, the rank-1 recognition rate was 99.14%, while Minnaert obtained 97.86%. On the raw textures, PCA achieved 70.34%, while PCA+LDA obtained 76.04%. When using MSR, the PCA results improved to 74.35%, while PCA+LDA improved to 85.04%. The training set we used for PCA and PCA+LDA was Database A. We also ran MSR and our CW-SSIM implementation. This yielded 96.58% rank-1 recognition.

6. Conclusions

A novel method for human face modeling and relighting was presented. It has wide applicability as it requires only 3D geometry and texture. It does not need light measurement information and is fully automatic. It was shown that the raw geometry data can be converted to a geometry image representation which makes them suitable for modeling applications. Additionally, the albedo can be extracted from the raw texture and used in relighting applications. Compared to other approaches, the proposed method has fewer limitations, such as the ability to handle self-shadowing.

Finally, we show that our method can be utilized in Face Recognition, through the use of the estimated albedo. The Face Recognition experiment demonstrates the successful illumination removal from the texture. In the same experiment it is shown that the Oren-Nayar BRDF we use performs better than other BRDFs.

Future work will be directed towards modeling the face in more detail. Certain facial features such as the eyes and mouth need separate geometry and reflectance models. Additionally, the effect of facial hair (e.g., eyebrows, beard) on the reflectance model has to be investigated. Finally, the combination of more than one facial scans of the same person can be used to increase the quality of the geometry and reflectance models.

References

[1] *The CSU face identification evaluation system.* http://www.cs.colostate.edu/evalfacerec/

[1] We determined the optimal parameters for MSR experimentally. Additionally, we tested the various distance metrics available in the CSU suite and chose the best for each experiment.

algorithms5.html, 2003. Version 5.0. 7

[2] P. Besl and N. McKay. A method for registration of 3-D shapes. *IEEE T PATTERN ANAL*, 14(2):239–256, 1992. 3

[3] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *ACM SIGGRAPH*, pages 187–194, 1999. 2

[4] K. Bowyer, K. Chang, and P. J. Flynn. A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. *Comput. Vis. Image*, 101(1):1–15, Jan. 2006. 5

[5] P. Debevec, T. Hawkins, C. Tchou, H. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. In *ACM SIGGRAPH*, pages 145–156, 2000. 2

[6] C. Everitt, A. Rege, and C. Cebenoyan. Hardware shadow mapping. White paper, NVIDIA Corp., 2001. 4

[7] M. Fuchs, V. Blanz, H. Lensch, and H.-P. Seidel. Reflectance from images: a model-based approach for human faces. *IEEE Transactions on Visualization and Computer Graphics*, 11(3):296–305, May/June 2005. 2

[8] B. Funt, K. Barnard, M. Brockington, and V. Cardei. Luminance-based multi-scale retinex. In *Proc. AIC Color*, volume I, pages 330–333, 1997. 7

[9] A. S. Georghiades. Recovering 3-D shape and reflectance from a small number of photographs. In *Eurographics Symposium on Rendering: 14th Eurographics Workshop on Rendering*, pages 230–240, June 2003. 2

[10] X. Gu, S. Gortler, and H. Hoppe. Geometry images. In *ACM SIGGRAPH*, pages 355–361, 2002. 3

[11] P. Harrison. A non-hierarchical procedure for re-synthesis of complex textures. In *Int. Conf. in Central Europe on Comp. Graphics, Visualization and Computer Vision (WSCG)*, pages 190–197, 2001. 4, 5

[12] T. Igarashi, K. Nishino, and S. Nayar. The appearance of human skin. Tech. Report, Columbia University, 2005. 2

[13] K. Ikeuchi and K. Sato. Determining reflectance properties of an object using range and brightness images. *IEEE T PATTERN ANAL*, 13(11):1139 – 1153, Nov. 1991. 2

[14] H. Jensen, S. Marschner, M. Levoy, and P. Hanrahan. A practical model for subsurface light transport. In *ACM SIGGRAPH*, pages 511 – 518, 2001. 2

[15] I. Kakadiaris, M. Papadakis, L. Shen, D. Kouri, and D. Hoffman. m-HDAF multiresolution deformable models. In *Proc. 14th International Conference on Digital Signal Processing*, pages 505–508, Santorini, Greece, July 1-3 2002. 3

[16] I. Kakadiaris, G. Passalis, , G. Toderici, Y. Lu, N. Karambatziakis, N. Murtuza, and T. Theoharis. 3D face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE T PATTERN ANAL*, 29(4):640–649, Apr. 2007. 1, 2, 3

[17] I. Kakadiaris, L. Shen, M. Papadakis, D. Kouri, and D. Hoffman. g-HDAF multiresolution deformable models for shape modeling and reconstruction. In *Proc. British Machine Vision Conference*, pages 303–312, Cardiff, United Kingdom, Sept. 2-5 2002. 3

[18] S. Kirkpatrick, C. Gelatt, and M. Vecchi. Optimization by simulated annealing. *Science*, 22(4598):671–680, 1983. 3

[19] J. Lambert. *Photometria sive de mensure de gratibus luminis, colorum umbrae*. Eberhard Klett, 1760. 7

[20] H. Lensch, J. Kautz, M. Goesele, W. Heidrich, and H.-P. Seidel. Image-based reconstruction of spatial appearance and geometric detail. *ACM Transactions on Graphics*, 22(2):234–257, Apr. 2003. 4

[21] C. Loop. Smooth subdivision surfaces based on triangles. Master's thesis, Dept. of Mathematics, U. of Utah, 1987. 3

[22] S. Mallick, T. Zickler, D. Kriegman, and P. Belhumeur. Specularity removal in images and videos: A PDE approach. In *Proc. ECCV*, pages 550–563, Graz, Austria, 2006. 2

[23] C. Mandal. *A Dynamic Framework For Subdivision Surfaces*. PhD thesis, University of Florida, 1998. 3

[24] D. Metaxas and I. A. Kakadiaris. Elastically adaptive deformable models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(10):1310–1321, 2002. 3

[25] M. Minnaert. The reciprocity principle in lunar photometry. *Astrophysical Journal*, 1941. 7

[26] S. Nayar, X. Fang, and T. Boult. Separation of reflection components using color and polarization. *INT J Comput Vision*, 21(3):163–186, 1997. 2

[27] NVIDIA Corp. *Cg Language Spec. 1.5*, May 2004. 4

[28] M. Oren and S. Nayar. Generalization of Lambert's reflectance model. In *ACM SIGGRAPH*, pages 239–246, 1994. 3

[29] G. Papaioannou, E. Karabassi, and T. Theoharis. Reconstruction of three-dimensional objects through matching of their parts. *IEEE T PATTERN ANAL*, 24(1):114–124, 2002. 3

[30] S. Paris, F. Sillion, and L. Quan. Lightweight face relighting. In *Pacific Graphics*, pages 41–50, 2003. 2, 4

[31] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the Face Recognition Grand Challenge. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 947–954, San Diego, CA, 2005. 3

[32] B. Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311317, 1975. 4

[33] J. Portilla and E. Simoncelli. A parametric texture model based on joint statistic of complex wavelet coefficients. *INT J Comput Vision*, 40. 6

[34] E. Praun and H. Hoppe. Spherical parametrization and remeshing. In *ACM SIGGRAPH*, pages 340–349, 2003. 3

[35] R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment maps. In *ACM SIGGRAPH*, pages 497–500, 2001. 7

[36] P. Siarry, G. Berthiau, F. Durbin, and J. Haussy. Enhanced simulated annealing for globally minimizing functions of many-continuous variables. *ACM T on Mathematical Software*, 23(2):209–228, 1997. 3

[37] E. Simoncelli, W. Freeman, E. Adelson, and D. Heeger. Shiftable multi-scale transforms. *IEEE T INFORM THEORY*, 38:587–607, 1992. 6

[38] W. Smith and E. Hancock. Estimating the albedo map of the face from a single image. In *IEEE ICIP*, volume 3, pages 780–783, 2005. 2

[39] W. Smith, A. Robles-Kelly, and E. Hancock. Reflectance correction for perspiring faces. In *IEEE ICIP*, volume 2, pages 1389–1392, 2004. 2

[40] P. Tan, S. Lin, L. Quan, and H.-Y. Shum. Highlight removal by illumination-constrained inpainting. In *Proc. ICCV*, volume 01, pages 164–169, 2003. 5

[41] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE T IMAGE PROCESS*, 13(4):600–612, 2004. 7

[42] Z. Wang and E. Simoncelli. Translation insensitive image similarity in complex wavelet domain. In *IEEE ICASSP*, volume 2, pages 573–576, 2005. 7

[43] T. Weyrich, W. Matusik, H. Pfister, B. Bickel, C. Donner, C. Tu, J. McAndless, J. Lee, A. Ngan, H. Jensen, and M. Gross. Analysis of human faces using a measurement-based skin reflectance model. In *ACM SIGGRAPH*, pages 1013–1024, 2006. 2

[44] T. Zickler, R. Ramamoorthi, S. Enrique, and P. Belhumeur. Reflectance sharing: predicting appearance from a sparse set of images of a known shape. *IEEE T PATTERN ANAL*, 28(8):1287–1302, 2006. 2