Human Face Shape Analysis under Spherical Harmonics Illumination Considering Self Occlusion

Jasenko Zivanov jasenko.zivanov@unibas.ch
Andreas Forster forster.andreas@gmail.com
Sandro Schönborn sandro.schoenborn@unibas.ch
Thomas Vetter thomas.vetter@unibas.ch
Department of Mathematics and Computer Science
University of Basel
4056 Basel, Switzerland

Abstract

In this paper we present a novel method of considering self-occlusion when fitting faces under global illumination. The effects of self-occlusion under spherical harmonics (SH) illumination are predicted through the use of a a morphable model of faces. A linear function is precomputed which maps morphable model coordinates directly to surface irradiance values.

We demonstrate that considering self-occlusion yields a clear improvement over an otherwise identical reconstruction where self-occlusion is neglected. Our method succeeds at reconstructing facial attributes that are visible primarily as a consequence of self-occlusion, such as the depth of the eye sockets and the shape of the nose.

1. Introduction

We aim to estimate the shape of a face from a single photograph under unknown arbitrary illumination and an unknown albedo while considering the effects of self-occlusion. Since this is an ill posed problem, we apply a statistical model of human faces to solve the arising ambiguities. In those models, a specific shape or albedo map is represented by a small number of real-valued coefficients. The small number of those coefficients allows us to estimate them in a very stable way.

The illumination is modelled as a continuous function on a sphere which describes the amount of light incoming from any given spatial direction. Since faces seen in photographs are often illuminated from many different angles, this allows us to express a wide range of illumination conditions encountered in the real world. We represent the illumination function as a linear combination of real spherical harmonics (SH) basis functions. This provides us with another low dimensional model, this time one of incoming illumination.

Because under continuous SH illumination light can irradiate any vertex from any direction, computing the effects of self-occlusion becomes computationally very demanding. The novelty of our approach is that we propose a linear approximation of the effects of self-occlusion that is coupled with our morphable model. This makes the problem tractable on today's computing machines and it allows us to consider global self-occlusion in our shape extraction.

Our illumination model is limited to soft and distant lighting, i.e. the illumination function is smooth over the incoming direction and the amount of light incoming from any given direction is the same across the entire face.

1.1. Previous Work

A first simple shape prior for facial image analysis was proposed by Vetter and Blanz in 1999. They define a low-dimensional generative statistical model of human faces and use it to perform shape from shading constrained by that model. Their method assumes a minimal single source illumination model and it consists of an analysis by synthesis optimization loop. In 2005, Romdhani [1] improved upon that estimation technique by introducing a separate specular highlight term and an edge correspondence term to the cost function. In 2009, Knothe [2] proposed a novel morphable model, the global-to-local model (G2L), which aims for a better spatial separation of the components of the model.

On the computer graphics side, Ramamoorthi [3] was the first to propose using real-valued spherical harmonics (SH) functions to model the illumination in a scene in his seminal paper from 2000. That approach was extended by Sloan and Kauz [4] in 2003, when they presented a method of precomputing the effects of self occlusion under SH lighting. In 2006, Zhang and Samaras [5] proposed an application of SH lighting to the machine vision context, namely to the problem of face recognition.

In 2011, Kemelmacher-Schlizerman and Basri [6] proposed an application of SH illumination to facial shape from shading. They use only one single mean shape in place of

an entire morphable model. That mean shape is deformed in order to explain the shading observed in the image. Their method does not consider the effects of self occlusion.

In 2012, Elhabian *et al*. [7] presented another shape from shading algorithm based on SH illumination which uses a more realistic Torrance-Sparrow reflection model, also without considering self-occlusion.

Most recently, Aldrian and Smith [8] have proposed a method for the inverse rendering of faces that relies on a linear model of facial self-occlusion. They do not aim to reconstruct the shape from the observed shading. Their model assumes a constant incoming light distribution which corresponds to considering only the first SH basis function.

1.2. Contribution

In this paper, we propose a method of fitting a morphable model of human faces to a single image under omnidirectional SH illumination while considering self-occlusion. The use of a morphable model allows us to precompute the effects of self-occlusion and thus consider them in the fitting process.

Without self-occlusion, the radiance of a surface point depends only on the surface normal at that point, even under omnidirectional illumination. If self-occlusion is considered, however, the evaluation of the radiance suddenly becomes computationally very expensive, because for each surface point P and every incoming light direction ω, the entire facial geometry needs to be looked at, and it needs to be determined whether direction ω is occluded at point P. That amount of computation makes determining self-occlusion intractable in any iterative shape fitting scheme.

We encode the self-occlusion information in the dimensions of the morphable model itself, and thereby avoid having to determine that information at runtime. This provides us with a linear function which maps shape model coordinates directly to surface radiance values under a given set of illumination coefficients. Inverting this linear function allows us to predict a shape that implies the self-occlusion effects observed in the input image.

Note that the actual radiance of a face under a given illumination is not a linear function of vertex displacements. We will demonstrate, however, that a linear function is a good approximation for the small amount by which corresponding vertices of a human face shift between different faces.

2. Theory

Shape The shapes of our faces are represented through a linear morphable model. The morphable model defines the position $s_v \in \mathbb{R}^3$ of each vertex v as follows,

$$s_v = s_v^\mu + \sum_{i=0}^{N} q_i^S \sigma_i^S s_v^i, \tag{2.1}$$

where $s_v^\mu \in \mathbb{R}^3$ is the mean position of vertex v and $s_v^i \in \mathbb{R}^3$ is its displacement as a result of each component i of the morphable model. The coefficients $q_i^S \in \mathbb{R}$ are the model coordinates of the head while the values $\sigma_i^S \in \mathbb{R}$ are the standard deviations of the corresponding model components. The superscript S denotes that we are referring to *shape* model coordinates and standard deviations, in order to distinguish them from the albedo coordinates and standard deviations which will be introduced later.

Shading We assume a global illumination function which is modulated individually for each vertex. We refer to the modulation coefficients as the *generalized vertex irradiance* which is predicted from model coordinates using a linear approximation, the *local irradiance model*.

Our illumination model is defined by an illumination function $L(\omega)$ which maps directions on a sphere $\omega \in \mathbb{S}^2$ to RGB light values $L \in \mathbb{RGB}$. The space $\mathbb{RGB}$ is equivalent to a triplet of positive real numbers. The illumination function $L(\omega)$ is projected into SH space,

$$L(\omega) = \sum_{l=0}^{B} \sum_{m=-l}^{l} \lambda_l^m Y_l^m(\omega) = \sum_{k=0}^{K} \lambda_k Y_k(\omega), \tag{2.2}$$

where Y_l^m are the real SH basis functions and $\lambda_l^m \in \mathbb{RGB}$ are the illumination coefficients which characterize a specific illumination environment. Throughout this project, we used three bands of SH ($B = 2$), which correspond to nine basis functions $Y_l^m(\omega)$. To improve legibility, we will refer to combinations of indices l and m as a combined index k throughout this paper.

The radiance R_v of a surface vertex v is equal to the integral over the light contributions from all unoccluded directions ω,

$$R_v(\omega_{out}) = \int_{\mathbb{S}^2} V_v(\omega) \max(0, \cos(\omega, n_v)) F(\omega, \omega_{out}) L(\omega) d\omega, \tag{2.3}$$

where $R_v(\omega_{out}) \in \mathbb{RGB}$ is the exitant radiance at v in direction $\omega_{out} \in \mathbb{S}^2$, $V_v(\omega) \in \{0, 1\}$ is the visibility function, which is one for visible directions ω and zero for occluded ones, $F \in \mathbb{RGB}$ is the bidirectional reflectance distribution function (BRDF) and L is our illumination function. This expression is equal to Kajiya's rendering equation [9], except that the effects of interreflection are neglected.

We assume Lambertian reflectiveness, so our BRDF is equal to a constant albedo $A_v \in \mathbb{RGB}$. By applying our definition of L (eq. 2.2), we can decompose the radiance as follows,

$$R_v = A_v \widehat{\rho}_v^\lambda, \tag{2.4}$$

$$\widehat{\rho}_v^\lambda = \sum_{k=0}^{K} \lambda_k \rho_{vk}, \tag{2.5}$$

$$\rho_{vk} = \int_{\mathbb{S}^2} V_v(\omega) \max(0, \cos(\omega, n_v)) Y_k(\omega) d\omega. \quad (2.6)$$

We refer to $\rho_{vk} \in \mathbb{R}$ as the *generalized vertex irradiances*, which measure the light contribution by each of the SH basis functions Y_k to each vertex v. They are independent of the illumination coefficients λ. The terms $\widehat{\rho}_v^{\lambda} \in \mathbb{RGB}$ are referred to as the *specific vertex irradiances*, and they describe the actual irradiance of vertex v under a specific illumination λ.

In the following, we will drop the subscript v when we refer to sets of quantities for all the vertices in the mesh, and the subscript k when we refer to the values associated with all nine SH basis functions.

3. Setup

Local Irradiance Model We use a linear local irradiance model $\Phi(q^S)$ to predict the generalized vertex irradiances ρ from morphable model coordinates $q^S \in \mathbb{R}^N$. $\Phi(q^S)$ is inverted in our algorithm to find optimal morphable model coordinates. It is defined as follows,

$$\begin{aligned} \Phi : \mathbb{R}^N &\to \mathbb{R}^{KN_v} \\ q^S &\mapsto \rho_\Phi^\mu + \sum_{i=0}^{N} q_i^S \rho_\Phi^i \approx \rho(q^S). \end{aligned} \quad (3.1)$$

The mean generalized irradiance ρ^μ is defined as the generalized irradiance of the mean shape s^μ. The irradiance deformations ρ^i are defined as the irradiances of the unit shapes in the linear shape model minus ρ^μ. A unit shape is a shape generated by equation 2.1 where all entries of q^S are zero, except for one which is equal to one.

The computation of the generalized irradiance of a shape is performed by evaluating the integral in equation 2.6 using Monte Carlo integration. This Monte Carlo integration consists of rendering many orthographic shadow maps of the head from many viewing directions ω in order to obtain the values of the visibility function $V_v(\omega)$ and then adding up the resulting values of the integrand.

At runtime, once a set of illumination coefficients λ_i has been estimated, the local irradiance model Φ is projected into that illumination, yielding the specific local irradiance model Φ^λ. This is done by applying equation 2.5 to ρ_Φ^μ and all ρ_Φ^i.

Albedo Model Second, we require an illumination-free albedo model. For that purpose we have determined the albedo maps of 200 heads of known shape by computing their generalized irradiance through Monte Carlo integration. We have then estimated the illumination coefficients λ and then divided the observed vertex radiances by their implied irradiances. Finally, we have constructed a PCA model of those 200 albedo maps. That model is very similar to the shape model. It is defined as follows,

$$A_v = A_v^\mu + \sum_{i=0}^{N_A} q_i^A \sigma_i^A A_v^i, \quad (3.2)$$

where A_v^μ is the mean albedo of vertex v and A_v^i are the principal components of the albedo model for v.

Initial Shape and Pose Our method is initialized by performing a landmark fit in order to estimate the correspondence between the vertices of the morphable model and the pixels of the image. That landmark fit consists of finding the pose of the face and an initial shape that maps a number of given landmarks to their known 2D positions in the image. The fit is performed using the L-BFGS optimization algorithm as described by Knothe [2]. No illumination model is used during this intialization step.

The pose is given by a 4×4 view matrix W, such that $(\widetilde{s}_x, \widetilde{s}_y)$ is the image position corresponding to world space position s, with,

$$\widetilde{s}_x = x_0 + \frac{\widehat{s}_x}{\widehat{s}_w}, \quad \widetilde{s}_y = y_0 + \frac{\widehat{s}_y}{\widehat{s}_w}, \quad \widehat{s} = W \begin{pmatrix} s_x \\ s_y \\ s_z \\ 1 \end{pmatrix}. \quad (3.3)$$

4. Method

Our method aims to estimate a new set of shape coordinates $q^{S\prime}$ which better explain the shading in the input image than the initial shape coordinates q^S. For that purpose, we also require estimates of the illumination coefficients λ_l^m and albedo coefficients q_i^A.

Our Algorithm consists of the following three steps:

1. Illumination Fit
2. Albedo Fit
3. Shape Fit

All three steps are linear least-squares problems. They are solved by performing an alternating least squares optimization. The illumination fit is repeated after the albedo fit since it is far quicker than the albedo fit or the shape fit. We have found through our experiments that only negligible changes to the shape take place after only three iterations of the above four steps.

As initial values, we assume the mean albedo A^μ and the generalized irradiance corresponding to the initial shape, i.e. $\Phi(q^S)$.

4.1. Illumination

We begin by estimating the illumination coefficients. This is done by solving a linear system of equations in a least squares sense.

The illumination coefficients λ_k are determined by solving the following set of equations in a least squares sense for all valid vertices v,

$$\bar{R}_v = A_v^{\mu} \sum_{k=0}^{K} \lambda_k \rho_{vk}, \tag{4.1}$$

where $\bar{R}_v$ are the vertex radiances observed in the image. A vertex is considered valid for the illumination fit if it is visible and if its variance in both albedo and generalized irradiance is below certain threshold values σ_t^A and σ_t^{ρ}. Those variances have been determined from the same set of 200 example heads also used for the computation of the albedo model and the linear local irradiance model Φ.

4.2. Albedo

Once an initial estimate of the illumination has been determined, we proceed to approximate the albedo. This is done using our albedo model. Although the albedo model is defined for all vertices in the face, the albedo fit is performed multiple times, once for each facial segment. The segments define the following facial areas: skin, eyes, eyebrows, nostrils and lips. Breaking up the albedo model in this way increases its expressiveness, since the albedo maps of those segments can then be fitted independently of each other.

The albedo estimation consists of solving the following linear least squares problem for the albedo model coefficients $q_i^{A,g}$ in each segment g,

$$w_v^g(\bar{R}_v - \mu_v^A \rho_v^{\lambda}) = w_v^g \rho_v^{\lambda} \sum_{i=0}^{N_A} q_i^{A,g} \sigma_i^A u_{v,i}^A, \tag{4.2}$$

$$0 = \tau^A q_i^A, \tag{4.3}$$

where the weight $w_v^g \in \mathbb{R}$ describes the amount to which vertex v belongs to segment g. The sum of w_v^g over all segments is equal to 1 at each vertex. The first term provides an equation in $\mathbb{RGB}$ for each valid vertex v and it aims to explain the colors observed in the image. A vertex is considered valid for the albedo fit if its variance in generalized irradiance lies below the treshold variance σ_t^{ρ}. This eliminates vertices of high geometric uncertainty. The second term is a regularization term and it provides N^A equations, one for each albedo model dimension, and it is intended to enforce a solution which is considered likely by the model. The number τ^A is a regularization constant and it allows for a trade-off between reconstruction precision and model prior.

The estimated coefficients are used to evaluate the vertex albedo values $A_v^g \in \mathbb{RGB}$ for each segment using equation 3.2. They are then interpolated using the segment weights w_v^g to obtain the final vertex albedo values A_v.

4.3. Shape

Here, we invert the local irradiance model $\Phi(q^S)$ defined in equation 3.1 in order to obtain a set of shape coordinate displacements δq_i so that the displaced shape space coordinates $q^{S\prime} = q^S + \delta q$ correspond to a mesh that better explains the shading in the input image. At the same time, we also need to preserve the model-to-image correspondence established by the initial landmark fit. This is accomplished by requiring the image plane shift of certain vertices to be close to zero through an additional correspondence term.

The shape fit consists of solving the following linear system of equations for δq_i,

$$\bar{R}_v - R_v^0 = \sum_{i=0}^{N} \rho_{\Phi v}^i \delta q_i, \tag{4.4}$$

$$0 = \tau^C w_v^C \sum_{i=0}^{N} \bar{W} \sigma_i^S s_v^i \delta q_i, \tag{4.5}$$

$$\tau^S q_i^0 = -\tau^S \delta q_i. \tag{4.6}$$

The first term (4.4) represents one equation in $\mathbb{RGB}$ for each vertex, or three equations per vertex in $\mathbb{R}$. Its purpose is to find a shape space shift δq that better explains the observed shading. The left hand side describes the color change required for vertex v, while the right hand side describes the color change as a function of δq_i. R_v^0 is the vertex radiance implied by the current shape under the estimated illumination and albedo.

The second term (4.5) is referred to as the correspondence term and it represents two further equations in $\mathbb{R}$ for each vertex. It aims to limit the amount by which vertex v shifts in the image plane as a result of δq_i. Since the lateral positions of all vertices are provided only by the initial correspondence fit, we need to take care not to distort that initial correspondence. The collective correspondence weight τ^C allows us to form a tradeoff between shading reconstruction accuracy and correspondence preservation. The individual weights w_v^C are equal to one for all landmark vertices and zero for all others. Since the correspondence has only been established at the landmarks, all other vertices are allowed to move laterally in the image plane. The matrix $\bar{W}$ is the upper left 2×3 submatrix of the view matrix W, as defined in equation 3.3.

The final term represents a set of N regularization equations, and it serves to keep the resulting shape space coordinates $q_i^1 = q_i^0 + \delta q_i$ close to zero, in order to ensure a resulting shape that is close to the mean and thus considered likely by the morphable model.

5. Experiments

In the following, we will describe a number of experiments performed to evaluate our method.

The first part, section 5.1, serves to validate the approximation of facial irradiance as a linear function of morphable model coordinates. In the second part, section 5.2, we test the ability of the complete algorithm to recover the shape from images 5.2.

5.1. Irradiance Mapping Test

We first prove the validity of our local irradiance model $\Phi(q^S)$. For that purpose we work on a set of 60 test faces sampled from the morphable model according to its statistical distribution.

Let $q^S \in \mathbb{R}^N$ a set of coefficients corresponding to one test face. The vertex positions $s_v \in \mathbb{R}^3$ of all heads have been computed as shown in equation 2.1. For each head, its explicit generalized irradiance ρ^{ref} has been determined using Monte Carlo integration.

The purpose of this first experiment is to prove that the generalized irradiance ρ can be approximated as a linear function of shape coordinates q^S using our mapping function ϕ. The irradiance estimates ρ^{est} of all 60 test faces have been determined using ϕ and compared to the explicit irradiances ρ^{ref} of the heads. We have defined the following reconstruction error:

$$e(\rho) = \sum_{i=0}^{N_v} |\rho_v - \rho_v^{\text{ref}}|, \tag{5.1}$$

where $|\cdot|$ is the euclidean distance in $\mathbb{R}^K$, the space of SH coefficients, and N_v is the number of vertices in the mesh.

The reconstruction error $e(\rho^{\text{est}})$ has then been compared to the error $e(\rho_\phi^\mu)$ implied by using the mean irradiance ρ_ϕ^μ as an estimate. The error of the mean is on average 6.86 times greater than the error of the reconstruction. A histogram of the results is shown in fig. 1.

5.2. Shape Recovery

For the following experiment, we have generated renderings of 17 human heads under four light probes as illumination environments. The reflectance has been assumed Lambertian and self-occlusion has been considered by means of Monte Carlo integration. The 2D positions of the same set of anchor vertices as above have then been used as landmarks to determine pose and an initial set of morphable model coefficients. The shape corresponding to those coefficients serves as the input to our algorithm.

We have also constructed two versions of the irradiance function Φ, one that considers self-occlusion and one that does not. The algorithm has been applied to all input heads, once with the original Φ function, and once with a version where self-occlusion has been neglected.

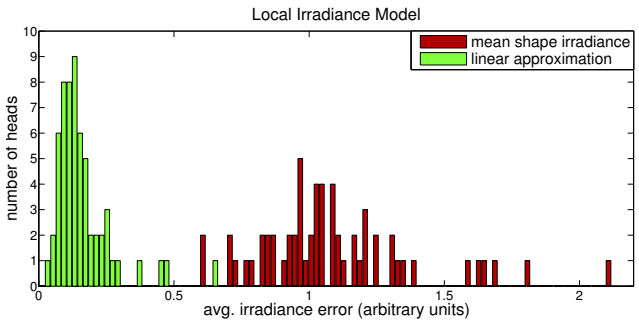

Figure 1: Histograms of the generalized irradiance error over 60 heads. We compare the error of our linear irradiance model (green) to the error when assuming the irradiance of the mean head (red). Note that our linear model provides a far better approximation than the mean irradiance. The irradiance can not be measured in any meaningful units, since it consists of arbitrary $\mathbb{R}^K$ values.

The following parameters have been used: $\tau^A = 20$, $\tau^C = 0.1$, $\tau^S = 0.5$ and $K = 9$. The colors have been defined in the range $[0, 1]$, with $(0, 0, 0)$ being black and $(1, 1, 1)$ being white. Distances were measured in μm. The face mesh was composed of $N_v = 33,334$ vertices. The albedo model consisted of $N_A = 200$ components and the shape model of $N = 300$ components.

The resulting shapes have then been compared to the ground truth shapes in both the depth maps and the normal maps as seen by the camera. For the depth comparison, we have centered each depth map at its mean depth, since the absolute depth can not be inferred precisely from a photograph. For the depth comparison, we have compared the mean absolute values of the per-pixel depth differences of both centered depth maps. For the orientation comparison, we have measured the mean absolute value of the angle between the normals at corresponding pixels of the normal maps. The comparison was limited to skin pixels, i.e. the eyes, nostrils, mouths and eyebrows have been neglected.

	without self-occlusion	with self-occlusion
depth (mm)	2.823	2.542
angle (°)	9.621	8.643

Table 1: Mean reconstruction errors. The values have been averaged over all relevant vertices of all 17 faces under all four illuminations.

We have found that for synthetic images, the surface orientation of the resulting shapes exhibits a clear improvement when self-occlusion is considered. The depth error is also reduced, albeit to a lesser degree. The results are displayed in fig. 2 and table 1.

We have also reconstructed a number of faces from real

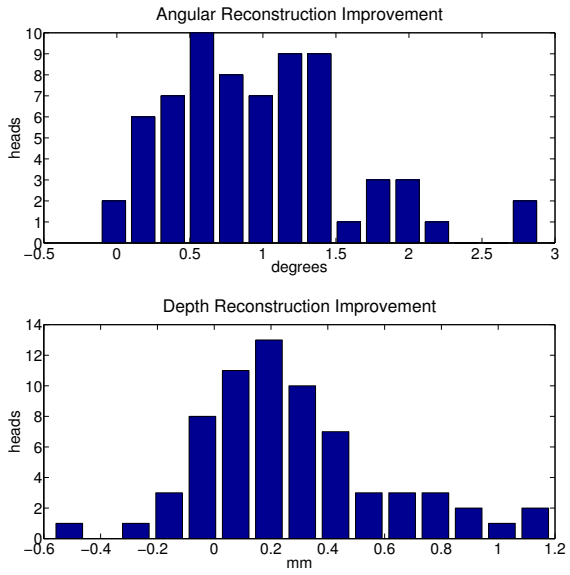

Figure 2: Histograms of reconstruction improvement when self-occlusion is considered.

photographs from the AFLW database [10]. The results are shown in fig. 4.

6. Summary and Conclusions

We have demonstrated a novel method of extracting the shape of a face from a single photograph under unknown illumination and albedo that considers self-occlusion. Our method is able to predict the effects of self-occlusion using self-shadowing information encoded in the dimensions of a morphable model of faces. It consists of a sequence of linear least-squares problems each of which can be solved within less than one minute on contemporary hardware, even though the effects of self-occlusion are considered. We have shown that our method yields a significant improvement over a reconstruction that does not consider self-occlusion.

Future work in this area will need to find a way of combining our method of considering self-occlusion with more expressive, non-Lambertian reflection models. This will probably also mean that the linearity of our approach will have to be sacrificed, since the high-frequency nature of specular reflections makes them difficult to represent as global functions of shape.

References

[1] S. Romdhani and T. Vetter, "Estimating 3d shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior", in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 2, pp. 986–993.

[2] R. Knothe, *A global-to-local model for the representation of human faces*, PhD thesis, University of Basel, 2009, `http://edoc.unibas.ch/diss/DissB_8817`.

[3] R. Ramamoorthi and P. Hanrahan, "An efficient representation for irradiance environment maps", in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, 2001, pp. 497–500.

[4] P.P. Sloan, J. Kautz, and J. Snyder, "Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments", in *ACM Transactions on Graphics (TOG)*. ACM, 2002, vol. 21, pp. 527–536.

[5] L. Zhang and D. Samaras, "Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 3, pp. 351–363, 2006.

[6] Ira Kemelmacher-Shlizerman and Ronen Basri, "3d face reconstruction from a single image using a single reference face shape", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 394–405, 2011.

[7] S. Elhabian, E. Mostafa, H. Rara, and A. Farag, "Non-lambertian model-based facial shape recovery from single image under unknown general illumination", in *Proceedings of the Ninth Conference on Computer and Robot Vision (CRV)*, 2012.

[8] O. Aldrian and W. Smith, "Inverse rendering of faces on a cloudy day", *Computer Vision–ECCV 2012*, pp. 201–214, 2012.

[9] J.T. Kajiya, "The rendering equation", *ACM SIGGRAPH Computer Graphics*, vol. 20, no. 4, pp. 143–150, 1986.

[10] M. Kostinger, P. Wohlhart, P.M. Roth, and H. Bischof, "Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization", in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2144–2151.

Figure 3: Faces extracted from synthetic images demonstrating the effect of considering self-occlusion. Input image (left), reconstructions without and with considering self-occlusion (middle) and the angular errors of the two reconstructions (right).

Figure 4: Faces extracted from real photographs from the AFLW database. Input image (left) and reconstruction (right).