4

# Face Image Synthesis and Interpretation Using 3D Illumination-Based AAM Models

Salvador E. Ayala-Raggi, Leopoldo Altamirano-Robles and Janeth Cruz-Enriquez

*Instituto Nacional de Astrofísica Óptica y Electrónica*
*México*

## 1. Introduction

One of the more exciting and unsolved problems in computer vision nowadays is automatic, fast and full interpretation of face images under variable conditions of lighting and pose. Interpretation is the inference of knowledge from an image. This knowledge covers relevant information, such as $3D$ shape and albedo, both related to the identity, but also information about physical factors which affect appearance of faces, such as pose and lighting. Interpretation of faces not only should be limited to retrieve the aforementioned pieces of information, but also, it should be capable of synthesizing novel facial images in which some of these pieces of information have been modified. This kind of interpretation can be achieved by using the paradigm known as analysis by synthesis, see Figure 1. Ideally, an approach based on analysis by synthesis, should consist of a generative facial parametric model that codes all the sources of appearance variation separately and independently, and an optimization algorithm which systematically varies the model parameters until the synthetic image produced by the model is as similar as possible to the test image, also called *input image*. A full interpretation approach should include the recovery of $3D$ shape, $3D$ pose, albedo and lighting from a single face image which exhibits any possible combination of these sources of variation.

Active appearance models, or simply $AAMs$ (Cootes et al. (2001); Edwards et al. (1998); Matthews & Baker (2004)), with respect to other approaches, represent a fast alternative to perform face interpretation using the *analysis by synthesis* paradigm. Texture and shape, are attributes modeled by $AAMs$ by using statistic tools such as *principal components analysis* or shortly $PCA$. However, the apparent texture of a face is an implicit combination of lighting and albedo. The separation process of these two attributes is not an easy task within the context of sparse models, like $AAMs$. $AAMs$ use a sparse set of vertices which outline the shape. Texture is interpolated over that shape. In fact, a detailed dense set of surface normals, which is not available in $AAMs$, is required to perform the separation of lighting and albedo. On the other hand, texture and shape variation among human faces is relatively small when uniform lighting is considered. $AAMs$ take advantage of this fact by supposing a constant relationship between changes of appearance and the variation of the model parameters producing those changes. This approximately constant relationship is a constant gradient which is used for performing fast fitting to input images. However, for most purposes, lighting is not uniform, and a proper separation of albedo and lighting becomes necessary. In a similar way as is texture variation in uniform lighting, albedo variation among human

faces is small. In contrast to albedo, lighting is not necessarily constrained to a small variation interval. In fact, lighting affects appearance more than identity and pose, and presents many degrees of freedom (see Ramamoorthi & Hanrahan (2001) and Basri et al. (2003)). During a fitting process, an initial model is gradually modified in each iteration until it match the input image. Therefore, if the illumination of the input image is too different from the illumination of the initial model, the ratio of appearance variation with respect to the parameters variation can not be the same during all the iterations of the fitting process. For instance, if we have a model with a pronounced left illumination, and a model with uniform illumination, the change of appearance caused by an increase on one of the model parameters, for example the parameter of scale, is not the same in both cases. This ratio of appearance variation with respect to the model parameters is in fact a Jacobian whose value changes in each iteration. Therefore, if we want to fit an $AAM$ to a face with any kind of lighting, a constant Jacobian is not the solution. On the other hand, recomputing the Jacobian in each iteration is an expensive computational task Cootes et al. (2001),Matthews & Baker (2004).
In this chapter, we introduce an innovative $3D$ extension of $AAM$s based on an illumination model. By using interpolation, we incorporate a dense set of surface normals to our sparse $3D$ $AAM$ model. In this way, we can model lighting within the process of synthesizing faces, and also within the optimization process used for fitting the face model to an input image. We propose a fitting method based on an inexpensive way for updating the Jacobian in accordance to the illumination parameters recalculated in each iteration. Our method is able to encode separately four of the more relevant sources of appearance variation: $3D$ shape, albedo, $3D$ pose and lighting. This approach estimates $3D$ shape, $3D$ pose, albedo, and illumination simultaneously during each iteration. Since our model uses analysis by synthesis, it has an inherent ability of adaptation to the input image. Adaptation is a desirable characteristic because it provides the possibility of designing person-independent face interpretation systems. Experimental results show that the proposed approach not only can be extended to face recognition, but also demonstrate its ability for fitting to novel faces and performing interpretation. We implement a novel way to cope with an important source of appearance variation which affects significatively face images: *lighting*. We anticipate that this approach can be extended to face recognition under difficult conditions of lighting and can be generalized to the analysis and recovery of other types of sources of appearance variation such as age, gender, expression, etc., where lighting interferes seriously in the analysis process.

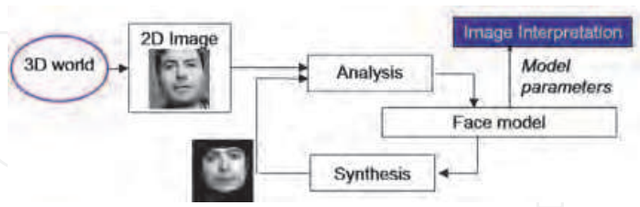


Fig. 1. Schematized flow of the analysis by synthesis approach.

Particularly, face interpretation has been faced through two paradigms: $3DMM$s Blanz et al. (1999; 2003); Romdhani et al. (2005; 2006) and $AAM$s Cootes et al. (1998; 2001); Dornaika et al. (2003); Edwards et al. (1998); Kahraman et al. (2007); Legallou et al. (2006); Matthews & Baker (2004); Sattar et al. (2007); Xiao et al. (2004). $3DMM$s cover a wide range of information recovery but are slow and cannot model properly every type of lighting. On the other hand, $AAM$s are fast but cannot model lighting and $3D$ information simultaneously. $AAM$ models have been used for fast $2D$ face alignment under variable conditions of lighting Huang et al. (2004); Kahraman et al. (2007); Legallou et al. (2006), but not for estimation of $3D$ pose,

3$D$ shape, albedo and illumination under non-uniform lighting conditions, which is still a challenging problem. In contrast, some authors Dornaika et al. (2003); Sattar et al. (2007); Xiao et al. (2004) have proposed 3$D$ $AAM$s for estimating 3$D$ pose and shape but do not include illumination. Finally, authors who reported lighting modeling for face recognition, do not propose methods for estimation of pose, shape, albedo and lighting simultaneously. This chapter describes a proposal for a complete 3$D$ approach for an automatic and fast recovery of 3$D$ shape, 3$D$ pose, albedo and lighting of a face under non-uniform lighting and variable pose. This recovery is performed by fitting a parametric 3$D$ Active Appearance Model based on the 9D subspace illumination model. Once we have finished the fitting process of the model to an input image, we obtain a compact set of parameters of shape, albedo, pose and lighting which describe the appearance of the original image. Because lighting parameters are not in a limited range, for faces with a pronounced non-uniform illumination, it is not possible to successfully use a constant Jacobian during all the fitting process as is done in original 2$D$ $AAM$ models Cootes et al. (2001). Instead of that, during the fitting stage, our algorithm uses the estimated lighting parameters, obtained in preceding iterations, for updating the Jacobian and the reference mean model on each iteration. The proposed method is called 3$D$ Illumination-Based Active Appearance Models Ayala-Raggi et al. (2008), Ayala-Raggi et al. (2009) and is suitable for face alignment, pose estimation and synthesis of novel views (novel poses and lighting) of aligned faces. In this chapter, we explain the method, measuring its capability to recover 3$D$ shape and albedo, and showing its capability to fit faces not included within the training set. Our experimental results, performed with real face images, show that the method could be extended to lighting-pose invariant face recognition.

## 2. Modeling lighting

Human face can be considered approximately as a convex surface with *Lambertian* reflectance Basri et al. (2003),Ramamoorthi & Hanrahan (2001). In Basri et al. (2003), Basri et al., propose using spherical harmonic functions to model lighting for face recognition. Spherical Harmonics are a set of functions which form an orthonormal basis which is able to represent all possible continuous functions defined in the sphere. The image of a face, illuminated by any lighting function can be expressed as a linear combination of harmonic reflectances (face images illuminated by harmonic lights),

$$I_i = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} I_{n,m} b_{n,m}(\mathbf{x}_i) \tag{1}$$

where $b_{n,m}$ are the set of harmonic reflectances and $\mathbf{x}_i$ is the i-th pixel of the object, in this case the face surface. In Basri et al. (2003), Basri et al. showed that the precision to approximate any function of light if we take a second order approximation $(n = 0, 1, 2)$ is at least 97.96%. From Equation (1) we see that this precision is achieved with only 9 harmonic images, and Equation (1) can be expressed in matrix notation as

$$\mathbf{I} = \mathbf{BL} \tag{2}$$

where $\mathbf{B}$ is a matrix with 9 columns. Each column is a harmonic image, and $L$ is a column vector containing 9 arbitrary parameters.

### 2.1 Forcing the lighting model to be positive

By using Equation (2), we could obtain not physically realizable images if we take arbitrary linear combinations of the harmonic images. In fact, any arbitrary combination could produce

an image with negative values. The harmonic images themselves have negative values, and as we know, light intensity is always positive. Therefore, different combinations of lightings must produce positive intensity values too. In Basri et al. (2003), the authors showed that the soft harmonic space spanned by the harmonic images can be discretized by using a sufficiently populated set of point light sources (delta functions) uniformly distributed around the sphere. Thus, Equation (2) can be modified as

$$\mathbf{I} = \mathbf{B}\mathbf{H}^T\mathbf{L} \tag{3}$$

where $\mathbf{L}$ is a column vector of arbitrary lighting parameters and $\mathbf{B}$ is a matrix with columns formed by the nine harmonic images. $\mathbf{H}$ is a matrix whose columns contain samples of the harmonic functions, whereas its rows contain the transform of the delta functions corresponding to the discrete number of point light sources.
This is a mathematical way of making discrete the smooth harmonic subspace by sampling the harmonic reflectance images. The more densely populated with deltas is $\mathbf{H}$, the better is the approach to the original space of the 9 harmonics. In order to obtain a good approximation to the original harmonic space, we should use a large set of point lights uniformly distributed around the sphere. However, in Lee et al. (2001), Lee et al., found an important result about how to approximate the illumination cone of lighting (see Georghiades et al. (1998)) with a small number of deltas. Only nine light point sources strategically distributed are necessary for approximating any reflectance on a face. Thus, $\mathbf{H}$ will be a constant $9 \times 9$ matrix.
In fact, the basis images can be obtained from two possible ways, the first one is the explained here, by using the compact notation through the spherical harmonics reflectances, and the second one is to explicitly render each one of the basis images, obtained from computing the intensity of each point by using the Lambert's law. This intensity can be computed if we know the surface normal, the albedo and the corresponding vector of the point light source.

## 3. Face synthesis using a $3D$ illumination-based active appearance model ($3D$-IAAM)

In this section, we describe an original method for face image synthesis based on the $3D-IAAM$ model proposed in this chapter. Our face synthesizer is capable of creating face images with arbitrary $3D$ pose, identity and illumination.

### 3.1 Construction of a bootstrap set of surfaces and albedo maps

In order to construct parametric models of shape and albedo, we need a bootstrap set of $3D$ face surfaces of different individuals, and their corresponding $2D$ albedo maps. This set of surfaces and albedo maps will be used to train models of $3D$ shape and $2D$ albedo, respectively.

#### 3.1.1 Recovery of the face surface for each training identity

A bootstrap set of face surfaces can be obtained under well controlled laboratory conditions by using a set of distant directional lights which illuminate the face one at the time but all working during a short period of time, in such a way that there is not movement from one image to the next.
Surfaces can be recovered by using a technique known as *photometric stereo* Forsyth & Ponce (2002); Horn et al. (1978); Silver (1980); Woodham (1989). By using $M$ ($M > 3$) different images per individual, each one illuminated by a different point light source, it is possible to simultaneously estimate the surface normals map and the albedo map of a face. This is

accomplished by using minimum squares for solving a linear system of $M$ equations, each one expressing the pixel intensity as a function of the direction of the incident light (Lambert's cosine law) for each pixel. From surface normals maps, it is possible to reconstruct the surface of each face by using *shapelets* Kovesi (2005). This is done by correlating the surface normals with those of a bank of *shapelet* basis functions. The correlation results are summed to produce the reconstruction. The summation of shapelet basis functions results in an implicit integration of the surface while enforcing surface continuity.
On the other hand, a mean surface normals map, computed from the set of surface normals maps, is used as a deformable template for building basis reflectance images during the fitting stage.

### 3.2 Constructing the models of shape and albedo

In order to obtain a parametric $3D$ shape model, first of all, we have to capture the more significative modes of shape variation. This can be accomplished by using a statistical method such as $PCA$ (*principal component analysis*) applied to a set of training faces with different identity. We can place $3D$ landmarks over the surface of $N$ training faces. To be sure that we are only modeling variations in shape and not in pose, we have to align the $3D$ shape models first, by using an iterative algorithm based on Procrustes analysis (see Figure 2).

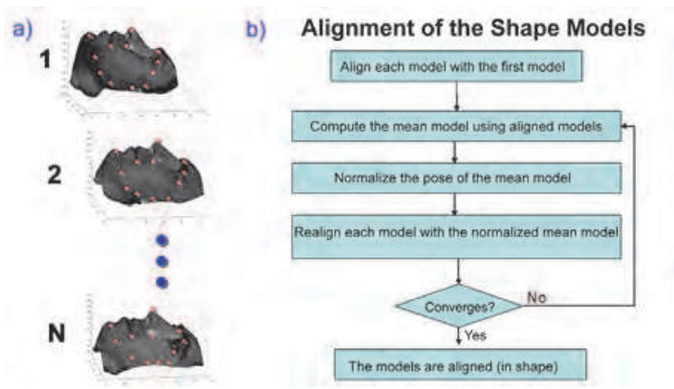


Fig. 2. The shape models (each one defined as the set of landmarks over a particular face surface) (a) must be aligned by using Procrustes Analysis (Ross (2004)) (b) before performing the statistical study of shape variation.

Then we apply $PCA$ to the set in order to obtain the principal modes of variation of $3D$ shape. We can generate an arbitrary model using the following expression

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{Q}_s \mathbf{c} \tag{4}$$

where $\bar{\mathbf{s}}$ is the mean shape model and $\mathbf{Q}_s$ is a matrix which contains the basis shapes (also known as *eigenshapes*) and $\mathbf{c}$ is a vector with arbitrary shape parameters. Similarly, we apply $PCA$ to the set of shape-normalized $2D$ albedos maps. Before applying $PCA$, the albedos map of each training face must be shape-normalized (using the bidimensional projection of the mean shape frame) as is shown in Figure 3.
A triangulation is designed to warp original images into the mean shape frame. Finally, any shape-normalized albedo image can be generated with

$$\lambda = \bar{\lambda} + \mathbf{Q}_\lambda \mathbf{a} \tag{5}$$

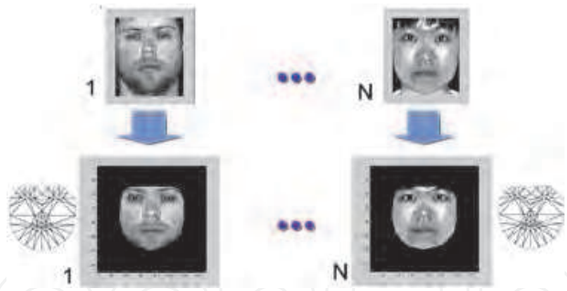


Fig. 3. Normalizing in shape the albedo images by warping the original albedo images into the 2*D* projection of the mean shape. Top: Original albedo images. Bottom: shape-normalized albedo images.

where $\bar{\lambda}$ is the mean albedo image, $\mathbf{Q}_\lambda$ is a matrix which contains principal albedo variation modes and **a** is a vector of arbitrary parameters.

### 3.3 Synthesizing faces with novel appearances

By using Equation (5), it is possible to synthesize an arbitrary albedo image $\lambda$ and then warp it to the 2*D* projection of an arbitrary frontal shape generated with Equation (4). This new face is not illuminated yet. In the same process of warping the albedo image to the new shape, it is also possible to carry out a 2*D* warping from the 2*D* mean map of surface normals (calculated during the training stage) to the same new shape **s**. So far, we have a new albedo image and a new map of surface normals, both of them shaped according to the new generated shape. With these two maps (albedos and normals), we can construct 9 basis reflectance images as is described in Section 2 by using Equation 3. Any illumination can be generated by a linear combination of these basis images. In order to give a 3*D* pose to the model, we use the 3*D* landmarks of the new generated 3*D* shape. By applying a rigid body transformation $(\mathbf{T}, \mathbf{R}, s)$ to these landmarks we give any pose and size to the created face.

If we suppose that the distance from the camera to the face is considerably greater than the depth of the face itself, then it is reasonable to use a simple orthographic projection model. Orthographic projection is the projection of a 3*D* object onto a plane by a set of parallel rays orthogonal to the image plane.

Finally, we warp the frontal face to the 2*D* orthographic projection of the transformed 3*D* shape. Figure 4 illustrates the synthesis process.

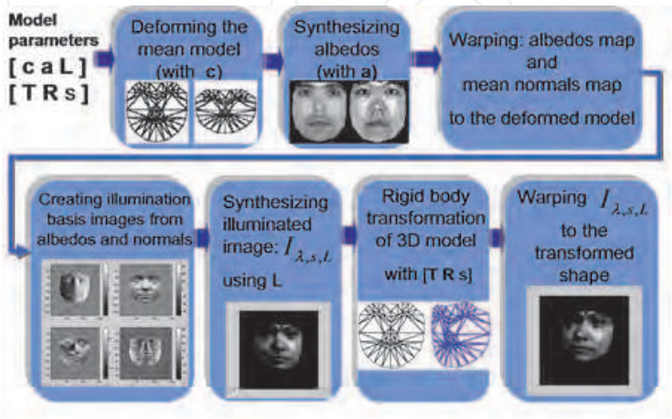


Fig. 4. Face synthesis process.

## 4. Face alignment using the $3D$-IAAM model

The original $2D$ $AAM$ approach for face alignment presented in Cootes et al. (2001), consists of an iterative algorithm which minimizes the residual obtained by comparing a shape-normalized region (taken from the target image) with a reference mean-shape model which evolves in texture in each iteration. This method supposes a constant relationship between residuals and the additive increments to the model parameters. This approximation uses a constant Jacobian during all the fitting process, and works well when lighting is uniform because texture variation is small and residuals are always computed in the same reference frame, see Cootes et al. (2001). Since we know, in contrast to texture in human faces, lighting variation is not limited. Therefore, if the initial reference model is substantially different in lighting to that in the input image, it is not possible to consider a constant Jacobian for all the fitting process. Here, we propose an iterative fitting algorithm capable of correcting the Jacobian in each iteration by using the current estimation of lighting, which in turn, is used to update the reference model too.

### 4.1 Overview of the iterative fitting process

Once we have created the models of shape and albedo, we can use them in the face alignment process. The alignment process consists of an iterative algorithm which captures a region within the input image, performs a normalization of this region according to the current set of model parameters and compares this normalized image with a reference model. The comparison is always performed into a fixed reference shape. The reference model evolves only in lighting in each iteration. The resulting residual from that comparison is used in conjunction with a Jacobian for calculating suitable increments to be added to the current model parameters. During the following iteration the new set of model parameters are used again to capture and normalize a new region within the input image, and so on. At the beginning of the alignment process, a set of initial model parameters is defined by the user. Commonly, shape, albedo and rotation parameters are initialized with zero, illumination parameters are initialized to a medium illumination, and translation and scale parameters are initialized to a rough value near to the real $2D$ position and size of the face. In other words, initial parameters are initialized in such a way that they would produce a frontal mean face placed over the face in the input image.

On the other hand, at the end of the alignment process, the final set of model parameters should be capable of synthesizing a face image similar to the original in the input image by using the synthesis process described in section 3.3. The normalization process over the input image is composed by a pose normalization, a shape normalization and an albedo normalization, all described in the following subsections.

### 4.2 Pose and shape normalization

In each iteration the model parameters of $3D$ shape and $3D$ pose determine a $3D$ structure whose orthographic $2D$ projection is used to define a region within the input image. This region can be mapped to a reference shape-normalized frame.

By using the rigid body transformation parameters $(\mathbf{T}, \mathbf{R}, s)$ and the shape parameters $\mathbf{c}$, a region in the image is sampled and warped to the $2D$ mean shape frame. This new shape-normalized image is denoted as $\mathbf{I}_{shape\,aligned}$.

### 4.3 Albedo normalization

A novel contribution of this work is a method for normalizing albedo when we have an estimate of lighting and albedo parameters. In fact, at the beginning of the fitting process,

albedo parameters have a zero value, then the normalization will produce the same image before normalization, see Equation 13. In contrast, as the albedo and illumination parameters get closer to the ideal values for synthesizing a face equal to the original, then normalization will produce an image more similar to a face with mean albedo illuminated by the actual lighting present in the original image. The image normalized in pose, shape and albedo, can be compared with a reference mean-shape mean-albedo face which evolves in lighting each iteration. The residual obtained from this comparison will give us the possibility to use a gradient matrix, or simply a Jacobian which is almost constant and is easily updated by using the estimated illumination parameters.

#### 4.3.1 Albedo normalization by using a current estimation of parameters of albedo and illumination

In Section 2 we have showed that every illumination over a face can be synthesized by using the following expression

$$\mathbf{I} = \mathbf{B}\mathbf{H}_{9PL}^{T}\mathbf{L} \tag{6}$$

as explained before, $\mathbf{B}\mathbf{H}_{9PL}^{T}$ represents a matrix with nine columns each one being a real and positive basis reflectance image. In order to compact the notation, we can denote that matrix as

$$\beta_{9PL} = \mathbf{B}\mathbf{H}_{9PL}^{T} \tag{7}$$

then Equation 6 can be rewritten as

$$\mathbf{I}_{illuminated\,face} = \beta_{9PL}\mathbf{L} = ([\lambda..\lambda] \cdot \mathbf{\Phi})\mathbf{L} \tag{8}$$

where $\lambda$ is the albedos map represented as a column vector repeated in order to form a matrix with the same dimensions as the basis reflectances matrix without albedo, represented by $\mathbf{\Phi}$. These two matrices are multiplied in an element-wise fashion (Hadamard product). Then, $\mathbf{I}_{illuminated\,face}$ can be rewritten as

$$\mathbf{I}_{illuminated\,face} = \lambda \cdot (\mathbf{\Phi}\mathbf{L}) \tag{9}$$

Now, suppose that the fitting algorithm has successfully recovered the shape and pose parameters corresponding to the input image. In that situation, the process of pose and shape normalization explained in the preceding section would produce a frontal shape-normalized face.

On the other hand, if we would know the correct illumination parameters $\mathbf{L}$ of that face, we could solve for the albedo by manipulating Equation 9 and using $\mathbf{I}_{shape\,aligned}$ instead of $\mathbf{I}_{illuminated\,face}$,

$$\hat{\lambda} = \frac{(\mathbf{I}_{shape\,aligned})}{(\mathbf{\Phi}\hat{\mathbf{L}})} \tag{10}$$

where the division denotes an element-wise division.

Suppose now, that we have a correct estimation of the albedo parameters ($\mathbf{a}$). Then, by using $\hat{\lambda}$ and the albedo parameters ($\mathbf{a}$) we can derive an approximated mean albedo by using Equation 5,

$$\tilde{\lambda} \approx \hat{\lambda} - \mathbf{Q}_{\lambda}\mathbf{a} \tag{11}$$

Finally, we can normalize the image in albedo by using $\tilde{\lambda}$,

$$\mathbf{I}_{aligned} = (\tilde{\lambda}) \cdot (\mathbf{\Phi}\hat{\mathbf{L}}) \tag{12}$$

where $\hat{\mathbf{L}}$ is a vector containing the current estimated illumination parameters. We can rewrite Equation 12 as

$$\mathbf{I}_{aligned} = [I_{shape\,aligned}./(\mathbf{\Phi}\hat{\mathbf{L}}) - \mathbf{Q}_{\lambda}\mathbf{a}] \cdot (\mathbf{\Phi}\hat{\mathbf{L}}) \tag{13}$$

The residuals vector can be calculated as

$$\mathbf{r} = \mathbf{I}_{aligned} - \bar{\lambda} \cdot (\mathbf{\Phi}\hat{\mathbf{L}}) \tag{14}$$

The energy of this residual image is a quantity to minimize by the iterative optimization algorithm

$$\|\mathbf{r}\|^2 = \|\mathbf{I}_{aligned} - \bar{\lambda} \cdot (\mathbf{\Phi}\hat{\mathbf{L}})\|^2 \tag{15}$$

where $[\bar{\lambda} \cdot (\mathbf{\Phi}\hat{\mathbf{L}})]$ represents the reference model with mean shape, mean pose, mean albedo, but illumination determined by the last estimated lighting parameters $\hat{\mathbf{L}}$. The process for obtaining residuals in each iteration is shown in Figure 5, where the reference model $[\bar{\lambda} \cdot (\mathbf{\Phi}\hat{\mathbf{L}})]$ is denoted by $\bar{\mathbf{f}}$.

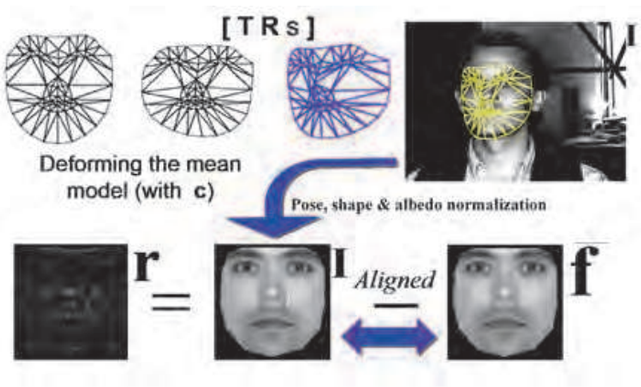


Fig. 5. Estimation of residuals during a step of the fitting process. The mean shape is deformed by using the current parameters **c** and **ø** (top). Then, the region within the 2*D* projection of this new structure is warped from the test image to the reference mean shape frame (in the bottom and in the middle) in order to apply the process of albedo normalization. The resulting image called $I_{Aligned}$ is compared with a reference model in order to obtain a residual image.

In order to work with a more compact notation, we can view the pose-shape and albedo normalization as an inverse transformation to the 3*D*-IAAM synthesis process. Therefore, we can denote that process as

$$\mathbf{I}_{aligned} = T_{\mathbf{p}}^{-1}(\mathbf{I}_{input}) \tag{16}$$

where $\mathbf{I}_{input}$ represents the input image and $\mathbf{p}$ is a vector containing the model parameters $\mathbf{p} = (\mathbf{T}^T, \mathbf{R}^T, s, \mathbf{c}^T, \mathbf{L}^T, \mathbf{a}^T)^T$. The initial parameters for the start of a fitting process are denoted as

$$\mathbf{p}_0 = (\mathbf{T}_0^T, \mathbf{R}_0^T, s_0, \mathbf{c}_0^T, \mathbf{L}_0^T, \mathbf{a}_0^T)^T \tag{17}$$

where $\mathbf{T}_0^T$ is the initial position vector $(x_0, y_0)$ given by the user. $\mathbf{R}_0^T = (0,0,0)$ is the initial rotation vector, and $s_0$ the initial scale factor (commonly equal with 1). $\mathbf{c}_0^T = (0,0,0,0,...)$ is the initial shape parameters vector. $\mathbf{L}_0^T = (L0_1, L0_2, L0_3, L0_4, L0_5, L0_6, L0_7, L0_8, L0_9)$ is the initial illumination parameters vector. Finally, $\mathbf{a}_0^T = (0,0,0,0,...)$ is the initial albedo parameters vector.

### 4.4 Modeling the residuals

During the fit, according to the last estimated parameters, the pixels inside of a region in the image are sampled and transformed. So, the residuals image computed with (14) is a function of the model parameters $\mathbf{p}$, that is $\mathbf{r} = \mathbf{r}(\mathbf{p})$. The first order Taylor expansion of (14) gives $\mathbf{r}(\mathbf{p}+\delta\mathbf{p}) = \mathbf{r}(\mathbf{p}) + \frac{\delta\mathbf{r}}{\delta\mathbf{p}}\delta\mathbf{p}$, here, $\mathbf{p}^T = (\mathbf{T}^T|\mathbf{R}^T|s|\mathbf{c}^T|\mathbf{a}^T|\mathbf{L}^T)$, and the $ij-th$ element of the matrix $\frac{\delta\mathbf{r}}{\delta\mathbf{p}}$ is $\frac{\partial r_i}{\partial p_j}$. We desire to choose $\delta\mathbf{p}$ such that it minimize $||r(\mathbf{p}+\delta\mathbf{p})||^2$. Equating $\mathbf{r}(\mathbf{p}+\delta\mathbf{p})$ to zero leads to the solution

$$\delta\mathbf{p} = -\mathbf{J}^{-1}\mathbf{r}(\mathbf{p}) \tag{18}$$

and $\mathbf{J}^{-1}$ can be calculated by pseudo-inverting the Jacobian matrix (Moore-Penrose pseudo-inverse), or by using the normal equations:

$$\mathbf{J}^{-1} = (\frac{\delta\mathbf{r}^T}{\delta\mathbf{p}}\frac{\delta\mathbf{r}}{\delta\mathbf{p}})^{-1}\frac{\delta\mathbf{r}^T}{\delta\mathbf{p}} \tag{19}$$

where $\frac{\delta\mathbf{r}}{\delta\mathbf{p}}$ is actually a gradient matrix or Jacobian changing in each iteration. Recalculating it at every step is expensive. Cootes et al. in Cootes et al. (2001), assume it to be constant since it is being computed in a normalized reference frame. This assumption is valid when we are only considering variations of texture, and lighting is ignored because it is uniform. Since texture parameters do not present a large variation between training faces, then, it is possible to compute a weighted average of the residuals images for each displaced parameter in order to obtain an average constant Jacobian. In our case, we are dealing with non-uniform illumination, therefore we propose to construct an adaptive Jacobian as is explained later.

### 4.5 Iterative fitting algorithm

In Cootes et al. (2001), authors propose to utilize a precalculated constant Jacobian matrix which is used during all the fitting process. Each iteration, a sampled region of the image is compared with a reference face image normalized in shape which is updated only in texture according to the current estimated parameters. Ideally, this reference image constitutes a reference model evolving in texture which should be associated to a Jacobian evolving in texture too. However, in practice, a mean constant Jacobian, computed from the different textures found in the training set, is used. This constant Jacobian works well in uniform lighting conditions, because texture variation between training faces is relatively small. Nevertheless, using a constant Jacobian would produce bad alignments in both, the approach described in Cootes et al. (2001) and in our $3D$ approach Ayala-Raggi et al. (2008) when the lighting of the input face is considerably different from the lighting used during the training stage. In our $3D$ approach, an ideal procedure to achieve good convergence results, at a high computational cost, would be to recalculate completely the Jacobian each iteration. This operation could be performed each iteration by displacing the parameters of

the reference model. The parameters of albedo and illumination would be displaced from their current estimated values, and the $3D$ shape and pose parameters from their mean state values. All these parameter displacements would be used to synthesize displaced face images which, in turn, would be used for computing residuals by subtracting the images without displacement from the displaced images. Finally, residual images and their respective parameter displacements would be used to calculate the Jacobian. This process of synthesis of multiple images should be performed on-line during the fitting stage and certainly would be an extremely expensive operation.

In contrast, we propose a computationally inexpensive way to update the Jacobian by using the current illumination parameters. Each iteration, our optimization algorithm samples a region of the image and normalizes it in pose, shape and albedo. Albedo normalization is performed by using the current estimated illumination parameters. Thus, a comparison should be computed between this normalized image and the reference mean model (a model with mean shape and albedo) illuminated by using the same current illumination parameters. The estimated residuals and an updated Jacobian (*re-illuminated* by using the current estimated lighting) can be used to compute the new parameters displacements.

Updating the Jacobian with the current estimated illumination parameters is an easy and computationally inexpensive step, because we use the fact that lighting and albedo are separated vectors and they are independent of basis reflectance images, see Equation 9. In training time, we construct a set of displaced images that will be used during the fitting stage to update the Jacobian. We know that basis reflectances $\mathbf{\Phi}$ (without albedo) are not affected by albedo displacements, but they can be modified by pose and shape increments. Our model uses 33 parameters: 6 for pose, 9 for $3D$ shape, 9 for illumination, and 9 for albedo. We construct 15 ($6 + 9 = 15$) basis reflectance matrices $\mathbf{\Phi}_{p_i+\Delta p_i}$ by displacing, in a suitable quantity, each one of the 15 parameters of pose and shape. That is, by using face synthesis (through our model), we synthesize each reflectance image represented as a column within the matrix $\mathbf{\Phi}_{p_i+\Delta p_i}$ by giving the following synthesis parameters:

$$\mathbf{p} = (p_1, p_2, ..., p_i + \Delta p_i, ..., p_{15})^T \tag{20}$$

For instance, if $i = 8$, i.e. we are constructing the matrix for the second shape parameter, then the generating parameters $\mathbf{p}$ will be:

$$\mathbf{p} = (\mathbf{T}_0^T, \mathbf{R}_0^T, s_0, [0 \quad (0 + \Delta p_8) \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0])^T \tag{21}$$

In practice, we construct 30 basis reflectance matrices because we consider 15 positive displacements and 15 negative displacements. In a similar way, by displacing each parameter with a suitable increment $p_i + \Delta p_i$ (positive and negative), we obtain 30 albedo images for positive and negative increments in pose and shape parameters, and 18 albedo images for positive and negative increments in albedo parameters. These albedo images do not have information about lighting.

These 30 reflectance matrices and 48 albedo images are created during training time (*off-line*). During the alignment stage, we can create a Jacobian *on-line* according to the current parameters of illumination $\mathbf{L}$:

$$\frac{\delta \mathbf{r}}{\delta \mathbf{p}} = [\frac{\partial \mathbf{r}_1}{\partial \mathbf{p}_1} \cdots \frac{\partial \mathbf{r}_{33}}{\partial \mathbf{p}_{33}}] \tag{22}$$

where each column can be calculated as:

$$\frac{\partial \mathbf{r}_i}{\partial \mathbf{p}_i} = [\frac{\partial \mathbf{r}_i}{\partial \mathbf{p}_{i\,(\Delta+)}} + \frac{\partial \mathbf{r}_i}{\partial \mathbf{p}_{i\,(\Delta-)}}] \times \frac{1}{2} \tag{23}$$

with $i = 1, 2, ..., 33$. Here, $\frac{\partial \mathbf{r}_i}{\partial \mathbf{p}_{i\,(\Delta+)}}$ and $\frac{\partial \mathbf{r}_i}{\partial \mathbf{p}_{i\,(\Delta-)}}$ can be computed as:

$$\frac{\partial \mathbf{r}_i}{\partial \mathbf{p}_{i\,(\Delta+)}} = \frac{\lambda_{p_i+\Delta p_i} \cdot [\mathbf{\Phi}_{p_i+\Delta p_i}\mathbf{L}] - \lambda_0 \cdot [\mathbf{\Phi}_0\mathbf{L}]}{\Delta p_i} \tag{24}$$

$$\frac{\partial \mathbf{r}_i}{\partial \mathbf{p}_{i\,(\Delta-)}} = \frac{\lambda_{p_i-\Delta p_i} \cdot [\mathbf{\Phi}_{p_i-\Delta p_i}\mathbf{L}] - \lambda_0 \cdot [\mathbf{\Phi}_0\mathbf{L}]}{-\Delta p_i} \tag{25}$$

where $\lambda_0$ is the mean albedo, and $\mathbf{\Phi}_0$ is the matrix which columns are the mean basis reflectances (without albedo information). When $p_i$ corresponds to an albedo parameter, then we use $\mathbf{\Phi}_{p_i+\Delta p_i} = \mathbf{\Phi}_0$, since the reflectance matrices are not affected by albedo variations.
Because the Jacobian is constructed using the last estimated lighting parameters, we denote it as $\mathbf{J}(\hat{\mathbf{L}})$,

$$\mathbf{J}(\hat{\mathbf{L}}) = \frac{\delta \mathbf{r}}{\delta \mathbf{p}} \tag{26}$$

The iterative fitting algorithm is outlined in Figure 6.
Basically, the algorithm can be summarized as follows: When the fitting process begins, $I_{aligned}$ is an unprocessed region of the test image delimited only by the position of the initial model over the image. There is not shape or albedo normalization at this moment, so that the residual (step 2) will be computed between the region (without transformation) and the model in a similar way such as it is done in the $2D\ AAM$ algorithm Cootes et al. (2001). This first residual in combination with the Jacobian (which is a precalculated constant the first time) produces (such as it happens in Cootes et al. (2001)) an additive increment vector $\delta\mathbf{p}$ to be added to the initial parameters. $\delta\mathbf{p}$ is iteratively reduced by re-scaling it (step 15) until the energy of the residual is lower than its initial estimate. If this value does not decrease after a fixed number of reductions, the algorithm claims that convergence was not reached and stops. Otherwise, if the value is lower than the initial, then the new set of model parameters is used again to normalize a new region within the test image. The new residual in combination with a new Jacobian is used to compute a new set of increments to the parameters, and so on. Figure 7 illustrates two consecutive iterations of the fitting process.
On the other hand, Figure 8 shows the evolution of the model during the fitting process. Figure 8 is illustrative and shows only five representative iterations in both alignments. Actually, the algorithm takes an average of 14 iterations to reach convergence.
In practice, we have implemented this algorithm using a pyramid of two resolution levels. A multi-resolution approach overcomes to the single resolution method and improves the convergence of the algorithm, even if we place the initial model farther from the actual face. On the other hand, the columns within the Jacobian matrix which correspond to illumination parameters, are maintained fixed during the fitting process and they are precalculated from a mean state of uniform lighting.

1. Project the sampled region from the input image $\mathbf{I}_{input}$ to the mean-shape model frame by applying pose-shape-albedo normalization $\mathbf{I}_{aligned} = T_{\mathbf{p}_0}^{-1}(\mathbf{I}_{input})$ with parameters $\mathbf{p} = \mathbf{p}_0$.
2. Compute the residual, $\mathbf{r} = \mathbf{I}_{aligned} - \bar{\lambda} \cdot (\mathbf{\Phi}\mathbf{L}_0)$
3. Compute the predicted displacements, $\delta\mathbf{p} = -[\mathbf{J}_0]^{-1}\mathbf{r}(\mathbf{p})$. Where $\mathbf{J}_0$ is a Jacobian computed in the training stage by taking little displacements of the parameters from their initial values $\mathbf{p}_0$. $[\mathbf{J}_0]^{-1}$ is the Moore-Penrose pseudoinverse matrix of the Jacobian.
4. Take only the new estimate of illumination parameters and put the other parameters in their initial values ignoring the estimates, $\mathbf{p}_0 = (\mathbf{T}_0, \mathbf{R}_0, s_0, \mathbf{c}_0, \hat{\mathbf{L}}, \mathbf{a}_0)$
5. Set $\mathbf{p} = \mathbf{p}_0$.
6. Project the sampled region from the input image $\mathbf{I}_{input}$ to the mean-shape model frame by applying pose-shape-albedo normalization $\mathbf{I}_{aligned} = T_{\mathbf{p}}^{-1}(\mathbf{I}_{input})$
7. Compute the residual, $\mathbf{r} = \mathbf{I}_{aligned} - \bar{\lambda} \cdot (\mathbf{\Phi}\hat{\mathbf{L}})$
8. Compute the current error, $E = ||\mathbf{r}||^2$
9. Compute the predicted displacements, $\delta\mathbf{p} = -\mathbf{J}^{-1}\mathbf{r}(\mathbf{p})$. Here $\mathbf{J}^{-1} = [\mathbf{J}(\hat{\mathbf{L}})]^{-1}$. Jacobian $\mathbf{J}(\hat{\mathbf{L}})$ is assembled by using the precomputed images of basis reflectance and albedo in combination with the estimated parameters $\hat{\mathbf{L}}$ computed in last iteration, see Equations 24 and 25.
10. Update the model parameters $\mathbf{p} \longrightarrow \mathbf{p} + k\delta\mathbf{p}$, where initially $k = 1$.
11. Using the new parameters, calculate the new face structure $\mathbf{X}$ and the new mean-shape reference model $\bar{\lambda} \cdot (\mathbf{\Phi}\hat{\mathbf{L}})$.
12. Compute $\mathbf{I}_{aligned} = T_{\mathbf{p}}^{-1}(\mathbf{I}_{input})$
13. Calculate a new residual $\mathbf{r} = \mathbf{I}_{aligned} - \bar{\lambda} \cdot (\mathbf{\Phi}\hat{\mathbf{L}})$
14. If $||\mathbf{r}||^2 < Threshold$ then terminate else go to the next step
15. If $||\mathbf{r}||^2 < E$, then accept the new estimate, make $k = 1$ and go to step 8; otherwise go to step 10 and try at $k = 0.5, k = 0.25$, etc.. (In practice, after 7 attempts of reducing $k$, if $||\mathbf{r}||^2 \geq E$ then the fitting process is finished.)

Fig. 6. Fitting algorithm.

## 5. Experimental results

### 5.1 Individuals used

We evaluated our approach on two different datasets. The first one was called set $A$ and is composed by the 10 identities contained in the Yale B Database. Each subject in the database is photographed in six different poses. For each pose many different illuminations are available. A second dataset, that we call set $B$ is composed by 20 individuals. This second dataset is composed by the 10 identities from Yale B Database plus other 10 identities randomly selected from the extended Yale B database (which contains 28 identities from different ethnicity).

### 5.2 Setup for experiments

The test set for this experiments was composed by 60 real images (with a size of $320 \times 240$ pixels) taken from Yale database B in the following manner: all images have the pose number 6 which presents a similar angle in azimuth to the left and elevation up. This pose has an angle of 24 degrees from the camera axis. We choose 6 different illuminations for using with each one of the identities. Each illumination is generated by a single point light source, and its

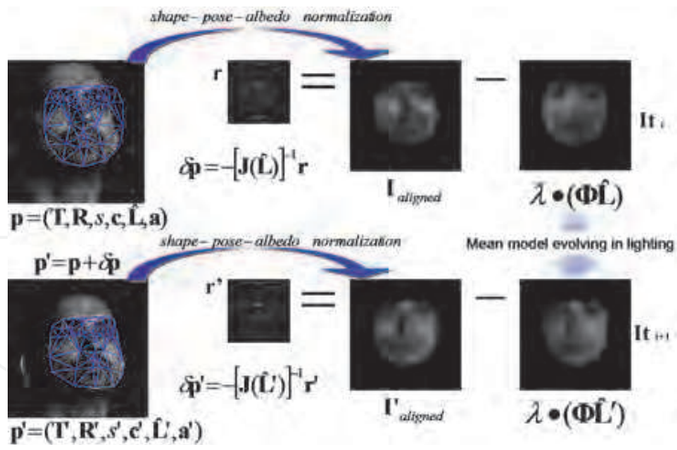


Fig. 7. Two consecutive iterations of the fitting process. During the iteration $It_i$ a region in the test image is captured and normalized according to the current parameters $\mathbf{p}$ producing the image $I_{aligned}$. A residual $\mathbf{r}$ is calculated by comparing $I_{aligned}$ with a reference shape-normalized model illuminated by the current parameter $\hat{\mathbf{L}}$. An additive increment vector $\delta\mathbf{p}$ is computed. $\delta\mathbf{p}$ is iteratively reduced by re-scaling it (step 15) until the energy of the residual is lower than its initial estimate. When this event occurs, the new set of model parameters $\mathbf{p}'$ is used again to normalize a new region within the test image. The new residual $\mathbf{r}'$ in combination with a new Jacobian $\mathbf{J}(\hat{\mathbf{L}}')$ is used to compute a new set of increments to the parameters, and so on.

Fig. 8. Evolution of the synthetic face produced by the model during the fitting process, from initialization to convergence.

direction is specified with an azimuth angle and an elevation angle with respect to the camera axis, see table 1.

| L1 | L2 | L3 | L4 | L5 | L6 |
|---|---|---|---|---|---|
| $A+50E+00$ | $A+35E+15$ | $A+10E+00$ | $A-10E+00$ | $A-35E+15$ | $A-50E+00$ |

Table 1. Illuminations used for experiments.

The initial conditions of the model at the beginning of the fitting process were manually setup only in translation and scale. The rest of the parameters: rotation, $3D$ shape, illumination and albedo were always initialized in their mean state for all the alignments, i.e., rotation: $\mathbf{R}_0^T = [0,0,0]$, $3D$ shape: $\mathbf{c}_0^T = [0,0,0,0,0,0,0,0,0]$, albedo: $\mathbf{a}_0^T = [0,0,0,0,0,0,0,0,0]$, and illumination: $\mathbf{L}_0^T = [0.6,0.6,0.6,0.4,0.4,0.9,0.9,0.4,0.4]$ (this configuration of the intensity of the light sources produces a *mean lighting* which illuminates uniformly the face).
In all the alignments, the translation and scale parameters were initialized with the output values of a manual pose estimator which uses three landmarks manually placed on the two external eye corners and on the tip of the nose. The output of this manual estimator are rigid

body parameters $(\mathbf{T}, \mathbf{R}, s)$ computed by using $3D$ geometry. From those parameters, we only used the translation and scale values in order to initialize the fitting process.
Our fitting algorithm is a local optimization and can fall into local minima, particularly if the initial model is placed far from the face to fit. We observed that the algorithm converges if we give an initial translation value with a maximum difference of $\pm 10$ pixels far from the ideal initial position. Therefore, the algorithm tolerates up to certain degree of imprecision in the initial position of the model over the test image.
Over the test set (the 60 images) we performed 180 alignments distributed within the following groups:

1. Group 1: 60 alignments using the fitting algorithm programmed with 4 computations of the adaptive Jacobian. That is, the algorithm has been allowed to recompute the Jacobian only during the first 4 consecutive iterations.
2. Group 2: 60 alignments using the fitting algorithm programmed with 2 computations of the adaptive Jacobian. That is, the algorithm has been allowed to recompute the Jacobian only during the first 2 consecutive iterations.
3. Group 3: 60 alignments using the fitting algorithm programmed with a constant Jacobian.

Figure 9 shows the alignments belonging to Group 1 (4 computations of the Jacobian) for each one of the 6 illuminations for identity 7.

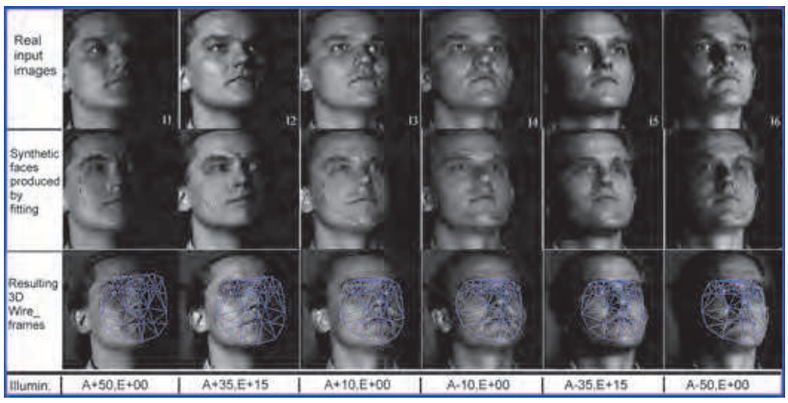


Fig. 9. Alignments for identity 7 with each one of the 6 different illuminations.

### 5.3 Recovery of $3D$ shape and Albedo and measuring its quality through identification

In order to measure the quality of the recovered $3D$ shape and albedo, we have considered that this quality is encoded into the recovered shape and albedo parameters. These estimated shape and albedo parameters are directly related to identity. Therefore, it is reasonable to compare them with those stored within a gallery containing parameters of all the training identities. In fact, $PCA$ allows the computation of the generative parameters for each training identity when the models of shape and albedo are created (see Section 3.2).
As a previous step before performing the comparison between estimated and stored parameters, they have to be re-scaled by dividing them by their respective standard deviations. Then, we measure the distance between the recovered parameters and the original parameters from the gallery. An appropriate distance measure in this case is the cosine of the angle between both vectors. This metric has the advantage of being insensitive to the norm of both vectors. In fact, that norm does not modify the perceived identity (see Romdhani (2005)). This operation was performed separately for vectors of albedo and for vectors of shape.

If we denote as $\sphericalangle(\hat{\mathbf{a}}, \mathbf{a}_i)$ the angle between the vector $\hat{\mathbf{a}}$ (estimated albedo parameters) and the vector $\mathbf{a}_i$ (stored albedo parameters for identity $i$), then the cosine can be computed with the following expression for albedo:

$$\Omega_i^a = cos(\sphericalangle(\hat{\mathbf{a}}, \mathbf{a}_i)) = \frac{\hat{\mathbf{a}}^T \mathbf{a}_i}{\sqrt{(\hat{\mathbf{a}}^T \hat{\mathbf{a}})(\mathbf{a}_i^T \mathbf{a}_i)}} \tag{27}$$

and the following expression for $3D$ shape parameters:

$$\Omega_i^s = cos(\sphericalangle(\hat{\mathbf{c}}, \mathbf{c}_i)) = \frac{\hat{\mathbf{c}}^T \mathbf{c}_i}{\sqrt{(\hat{\mathbf{c}}^T \hat{\mathbf{c}})(\mathbf{c}_i^T \mathbf{c}_i)}} \tag{28}$$

where $\hat{\mathbf{c}}$ are the estimated shape parameters vector, and $i$ is an index which indicates the identity of the parameters vector stored in the gallery. Because the cosine function might be negative, $\Omega_i^s$ and $\Omega_i^a$ are equated to zero in such a case. This positive cosine function works fine because we are interested on detecting only small angles related with the presence of high similarity between faces.

In order to perform the identification, we have to combine these two results (cosines for shape and cosines for albedo) to obtain a single identification result. An appropriate approach to combine both cosines is to convert them in likelihood values.

Using the known probability property which states that the sum of all likelihoods must be 1, we can normalize the computed cosines for albedo:

$$IL_i^a = \frac{\Omega_i^a}{(\Omega_1^a + \Omega_2^a + \Omega_3^a + \cdots + \Omega_{10}^a)} \tag{29}$$

and normalize the computed cosines for shape:

$$IL_i^s = \frac{\Omega_i^s}{(\Omega_1^s + \Omega_2^s + \Omega_3^s + \cdots + \Omega_{10}^s)} \tag{30}$$

where $IL_i^a$ (with $i = 1, 2, ..., 10$) represents the identity likelihood of the estimated albedo for each one of the ten identities stored in the gallery. Similarly, $IL_i^s$ (with $i = 1, 2, ..., 10$) represents the identity likelihood of the estimated shape for each one of the ten identities stored in the gallery.

Now, we can combine both likelihoods using a weighted sum. By experimentation, we found that weights with better identification rates are 0.6 for albedo, and 0.4 for shape. This experimental result can be explained by the following fact: $3D$ shape information of the original face is lost when the $2D$ image is formed. In fact, our fitting approach has to infer a probable shape. On the other hand, albedo which can be considered as $2D$ is recovered with more accuracy. In our experimental results we saw, that in some cases, the values of the cosine measured between the estimated shape vector and the shape vectors from the gallery, were very similar. These similar values of the cosine can produce confusion in the decision of the identity based only on the shape. Therefore, we considered that using probability functions instead of the cosine values is a more appropriate way to obtain a correct decision of the identity.

The conditional likelihoods $IL_i^a$ and $IL_i^s$, for albedo and shape respectively, can be combined to obtain a single likelihood $IL_i$:

$$IL_i = 0.6(IL_i^a) + 0.4(IL_i^s) \tag{31}$$

For instance, if the face of the test image corresponds to the identity $i = 2$, then, we would expect a higher value for $IL_2$ (theoretically $IL_2$ = 1) with respect to the values for $IL_i$ with $i = 1, 3, 4, 5, 6, 7, 8, 9, 10$ (theoretically 0).
In order to show the probability of the algorithm to select the correct identity under a specific illumination, we have used (from Group 1) ten alignments for test images of individuals $i = 1, 2, 3, ..., 10$ under the same illumination. Then, for each alignment a single value $IL_i$ (with $i$ being the test identity) was computed. For each illumination an average of the $IL_i$ values was computed and plotted in Figure 10 (b). The little vertical segments represent the associated standard deviation. We see that the mean $IL$ is greater when lighting is frontal to the face (illumination 5). Figure 10 (c) shows the identification rate for each illumination. The identification rate for each illumination is computed by summing the number of correct identifications and dividing this result by the total number of alignments for that specific illumination. In this graph we have plotted the identification rate computed for Group 1, Group 2, and Group 3 of alignments.

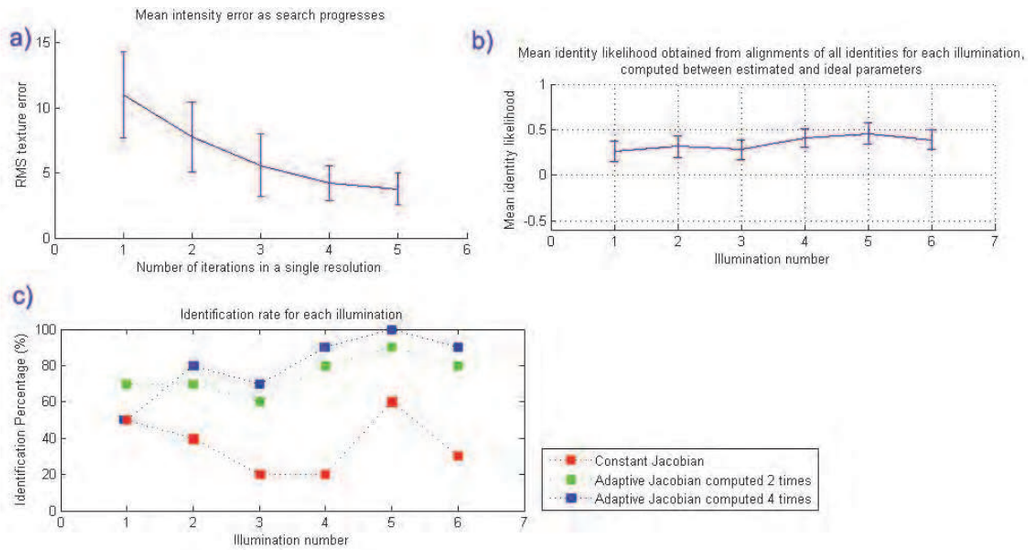


Fig. 10. a) Evolution of RMS error in intensity difference. b) Average (over the 10 identities) of the identity likelihood measured between estimated and ideal parameters. c) Identification rates for each one of the six illuminations.

In the case of fitting with four computations of an adaptive Jacobian we see the worst identification rate (50%) with the illumination number 1, and the best identification rate (100%) using the illumination number 5 which is nearly frontal to the face. A similar relation among identification rates for all the six lightings is conserved for the case of fitting with two computations of the adaptive Jacobian (plot in the middle). The phenomenon is repeated again for the case of fitting with a constant Jacobian (plot in the bottom). Anyway, we can see an important improvement on the quality of the reconstructions when the adaptive Jacobian is computed more times.
In a similar way, we have evaluated the fitting algorithm now trained with the 20 identities from set $B$.
For this test we used 6 images (with a size of $320 \times 240$ pixels) with the same pose and different lighting for each one of the 20 individuals from the set $B$. Hence, our test set is composed by 120 real images. Again, all images have the pose number 6 which presents a similar angle in azimuth to the left and elevation up. This pose has an angle of 24 degrees from the camera

axis. We choose the same 6 different illuminations for using with each one of the identities. See table 1.
Over the test set (the 120 images) we performed 360 alignments distributed within the following groups:

1. Group 1B: 120 alignments using the fitting algorithm programmed with 4 computations of the adaptive Jacobian.
2. Group 2B: 120 alignments using the fitting algorithm programmed with 2 computations of the adaptive Jacobian.
3. Group 3B: 120 alignments using the fitting algorithm programmed with a constant Jacobian.

Figure 11 shows the identification rate for each illumination. In this graph we have plotted the identification rate computed for Group 1B, Group 2B, and Group 3B of alignments.

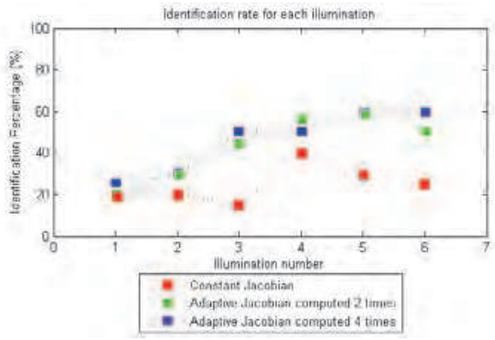


Fig. 11. Identification rates for each one of the six illuminations

In the case of fitting with four computations of an adaptive Jacobian we see the worst identification rate (25%) with the illumination number 1, and the best identification rate (60%) with the illumination number 5 and 6 which are nearly frontal to the face. In a similar way as in the case of experiments for set $A$, a similar relation among identification rates for all the six lightings is conserved for the case of fitting with two computations of the adaptive Jacobian (plot in the middle). Again, we can see an important improvement on the quality of the reconstructions when the adaptive Jacobian is computed more times.
In this test we used a training set of 20 individuals. In a similar way as in all experiments, model parameters have been limited to 9 shape parameters and 9 albedo parameters.
We used Principal Component Analysis for reducing the dimensionality of shape and albedo variation. Model parameters of shape and albedo are weights of a weighted sum of eigenvectors, see 4 and 5. Eigenvectors of shape or albedo represent variation modes (20 modes) and they are sorted according to their associated variances, from the higher to the lower value of these variances. Each variance associated to each eigenvector represents the relevance of the eigenvector into the weighted sum. The greater the variance, the more relevant the eigenvector (variation mode).
In order to reduce the dimensionality of the training set and using the same number of parameters, we have taking into account only the first 9 relevant eigenvectors of shape and albedo. We can compute the percentage of total variance that can be represented by the model using only 9 parameters of shape as

$$\Xi_{\sigma^2} = \frac{\sum_{i=1}^{9} \sigma_i^2}{\sum_{i=1}^{20} \sigma_i^2} \times (100) = 83.5\% \tag{32}$$

where $\Xi_{\sigma^2}$ is the shape representation capability of the model relative to the training set. In a similar way, since we only used the first 9 eigenvectors of albedo, we can compute the percentage of total variance that can be represented by the model using only 9 parameters of albedo as

$$\Xi_{\eta^2} = \frac{\sum_{i=1}^{9} \eta_i}{\sum_{i=1}^{20} \eta_i} \times (100) = 87.1\% \tag{33}$$

where $\Xi_{\eta^2}$ is the albedo representation capability of the model relative to the training set. These percentages are lower than those corresponding to the experiments using 10 training individuals where the number of model parameters of shape and albedo is also 9 and 9 respectively, where model parameters cover 100% of the total variance. Therefore, we can conclude that using a bigger set of training faces while keeping a fixed number of model parameters decrements the ability of representing the 100% of shape and albedo variation contained into the training set. In turn, that conclusion explains the lower identification rates observed on Figure 11 with respect to those observed on Figure 10 (c).
Figure 12 illustrates the difference in fitting with a constant Jacobian in contrast to fit with an adaptive one. Here we show reconstructions for two different lightings.

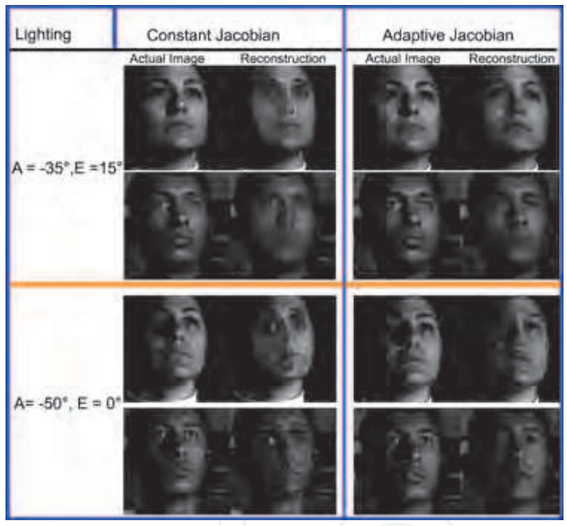


Fig. 12. Reconstructions of two individuals from set $B$ under two different lightings. The reconstructions obtained with the fitting algorithm which uses an adaptive Jacobian are visually better than those obtained from using a fitting algorithm which uses a constant Jacobian.

### 5.4 Face alignment of faces not included into the training set: Fitting novel faces

The $3D - IAAM$ model trained with the set of 20 individuals has been tested for fitting to novel faces not contained within the training set. Again, we used 33 model parameters: 6 for 3D pose, 9 for 3D shape, 9 for illumination, and 9 for albedo.
We selected 5 individuals not contained within the training set and captured in 3 poses each one (-24ž, 0, and +24ž with respect to the camera axis). For all the images, the 3D pose only varies in azimuth: -25, 0, and 25 degrees with respect to the camera axis. Figures 13, 14, 15, 16, and 17 show alignments for novel faces take from the extended Yale B database and originally numbered as 18,25,35,36. The fifth face belongs to the author of this work.


Fig. 13. Alignments for the face number 18 from the extended Yale B database. This face is not included in the training set. The recovered 3D pose angles are specified in degrees

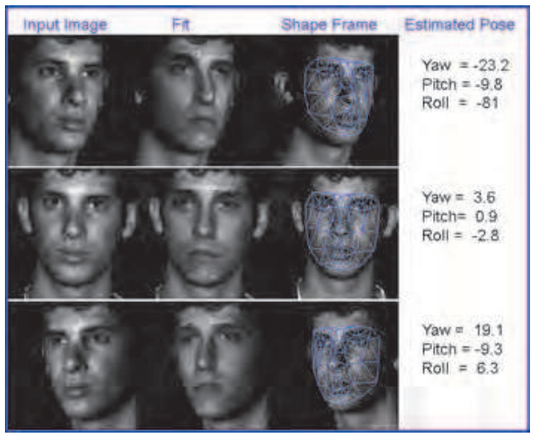


Fig. 14. Alignments for the face number 25 from the extended Yale B database. This face is not included in the training set. The recovered 3D pose angles are specified in degrees

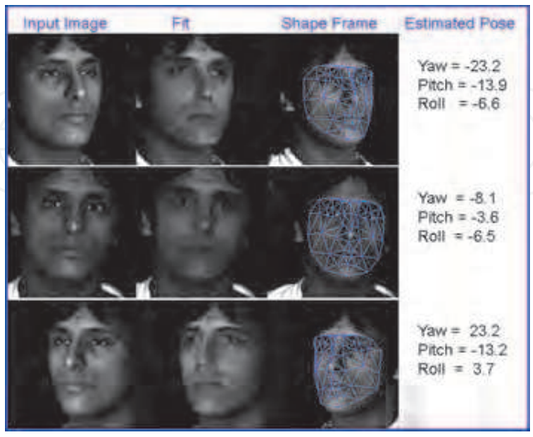


Fig. 15. Alignments for the face number 35 from the extended Yale B database. This face is not included in the training set. The recovered 3D pose angles are specified in degrees

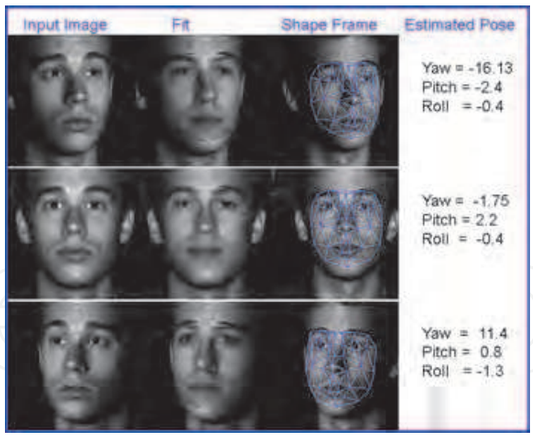


Fig. 16. Alignments for the face number 36 from the extended Yale B database. This face is not included in the training set. The recovered 3D pose angles are specified in degrees

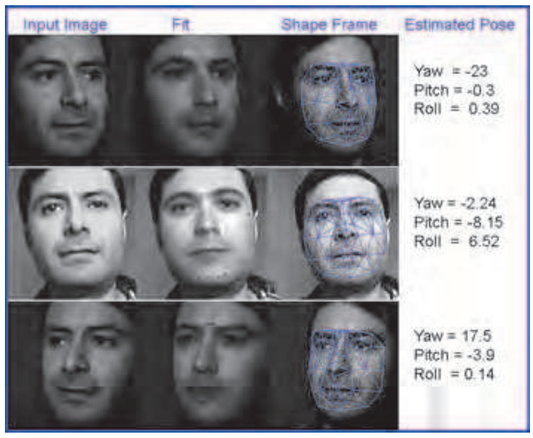


Fig. 17. Alignments for the author's face. This face is not included in the training set. The recovered 3D pose angles are specified in degrees

The experiments performed over faces not included within the training set give us with signs about the ability of the method for adapting to novel faces, and also demonstrate that it is possible to estimate relevant information about a new face. That information is provided at the end of the fitting process, and is delivered to us through the model parameters. We think that a possible generalization of the method consisting on the capability of fitting any human face may be feasible. The solution will be based on making a careful and systematic selection of the training faces according to desired characteristics. That could be another research problem.

## 6. Conclusions

Shape and albedo are estimated more accurately when an adaptive Jacobian is used. The adaptive Jacobian is a way to express the appearance variation produced by parameters variation as a function of the lighting parameters computed in each iteration. Hence, the adaptive Jacobian works better than the constant one when the initial model is different (in lighting) from the test image, as it actually happens in the most of cases. The improvement provided by the use of an adaptive Jacobian was confirmed when we obtained better estimations of shape and albedo whenever we were increasing the times that this Jacobian was

computed. On the other hand, we determined that the computational time used in calculating the Jacobian is linear with respect to the number of times that this Jacobian is computed. In contrast, the improvement in the recovery of the parameters was not significative when the Jacobian was computed more than two times. Therefore, we conclude that four computations of the Jacobian is sufficient to obtain acceptable reconstructions. On the other hand, the capability of the fitting algorithm to reconstruct novel faces not contained within the training set was demonstrated in this chapter. Finally, our proposed interpretation approach not only provides information from a face image, also it is capable of creating new information by reconstructing unseen novel views of a recovered face. This work has addressed the problem of automatic and fast interpretation of a face which exhibits any pose and any lighting. Modern approaches have important limitations regarding processing speed, fully automatic operation, 3D, lighting invariant and simultaneous handling of multiple appearance variation sources. We introduced a novel and fast method for automatic interpretation of face images from a single image. Pose, shape, albedo, and lighting are sources of appearance variation which modify the face image simultaneously. For that reason, trying to estimate only one of these factors without considering the others would produce inaccurate estimates. In order to avoid an inaccurate estimation of each one of these sources of appearance variation, our fitting method estimates simultaneously, in each iteration, the appropriate increments for parameters of 3D shape, 3D pose, albedo and lighting. At the end of the fitting process our proposed algorithm provides us with a compact set of parameters of 3D pose, 3D shape, albedo and lighting which describe the test image. The fitting algorithm is based on *a priori* knowledge of the relationship between the appearance variation (of the model) and the parameters. The appearance variation of the model is produced by changes in pose, shape, albedo and lighting. This appearance variation maintains a non-linear relationship with respect to the model parameters. However, in the case of pose, shape, and albedo, the appearance variation range is sufficiently small so that we can approximate this non-linear relationship with a linear relationship which can be easily learned. On the other hand, the range of appearance variation produced by changes in lighting is unlimited. Then, it is not possible to approximate the appearance variation with respect to the lighting parameters with a simple linear relationship. Fortunately, we found a way to separate lighting from the other sources of appearance variation, in such a way that we can learn a linear relationship between a set of parameters (pose, shape, and albedo) and the appearance variation caused by these parameters. This learned linear relationship is completely independent from lighting. By incorporating a particular lighting to this linear relationship in each iteration of the fitting process, it is possible to reconstruct a new relationship between the full appearance variation and the changes of all the model parameters, i.e. pose, shape, albedo and lighting. This new relationship is represented in our fitting algorithm by the adaptive Jacobian which is reconstructed in each iteration according to the current estimated lighting parameters. Our results, both quantitative and qualitative, show that the method is able to align a 3D deformable model not only in shape but also in albedo, pose and lighting simultaneously. The identification results lead us to think that our approach could be extended to automatic face recognition under arbitrary pose and non-uniform illumination. Besides, the model can synthesize unseen face images of people not used to train the model

## 7. Future work

In our approach, the process of creating synthetic faces is used during the synthesis of the basis reflectance images created during the training time. This set of resulting images is utilized later for the on-line construction of the Jacobian during the test stage. We could improve

the accuracy in the synthesis of lighting by refining the mapping of the normals from the mean model to the new deformed model. Presently, this mapping is purely 2D, but because our shape model is 3D, normals can be reoriented according to the new 3D position of each triangular facet. A more accurate representation of lighting should improve the recovery of 3D shape and albedo, and therefore the identification rate. We think that our method can also be optimized in fitting speed by reducing the times that the Jacobian is updated. According to the initial estimated lighting it would be possible to establish a criterium to determine the minimum necessary number of Jacobian updates, while is preserved an acceptable alignment. Also, a robust face recognition scheme can be implemented if we increase the number of identities for training, in such a way, that they have the enough kinds of extreme variations in shape and albedo for modeling all intermediate possibilities. There are many interesting avenues of feature work. With a careful and systematic selection of the faces for the training set, our method can be extended to a generic person-independent automatic 3D face interpretation system, useful for face recognition in difficult conditions of lighting and pose. Combined with other methods for identification, this kind of generic approach could be a suitable part of a complete biometric system for identity recognition.

## 8. References


Ayala-Raggi, S., Altamirano-Robles, L., Cruz-Enriquez, J. (2008). Towards an Illumination-Based 3D Active Appearance Model for Fast Face Alignment, *CIARP 2008*, pp. 568-575

Ayala-Raggi, S., Altamirano-Robles, L., Cruz-Enriquez, J. (2009). Recovering 3D Shape and Albedo from a Face Image under Arbitrary Lighting and Pose by Using a 3D Illumination-Based AAM Model, *ICIAR 2009*, Halifax, vol. 5627, pp. 584-593

Basri, R., Jacobs, D.W. (2003). Lambertian Reflectance and Linear Subspaces, In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 218-233

Blanz, V., Vetter, T. (1999). A Morphable Model for the Synthesis of 3D Faces, *Siggraph 1999*, pp. 187-194

Blanz, V., Vetter, T. (2003). Face Recognition Based on Fitting a 3D Morphable Model, In: *IEEE Transaction Pattern Analysis and Machine Intelligence*, vol. 25, pp. 1063–1074

Cootes, T.F., Edwards, G.J., Taylor, C.J. (1998). Active Appearance Models, *ECCV 1998*. LNCS, vol. 1407, pp. 484–498. Springer, Freiburg

Cootes, T.F., Edwards, G.J., Taylor, C.J. (2001). Active Appearance Models, In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 681-685

Dornaika, F., Ahlberg, J. (2003). Fast And Reliable Active Appearance Model Search For 3d Face Tracking, *Proceedings of Mirage 2003*, pp. 10-11. INRIA Rocquencourt, France

Forsyth, D. A., Ponce, J. (2002). Computer Vision: A Modern Approach, *US ed Ed. Prentice Hall*

Edwards, G. J., Taylor, C. J., Cootes T. F. (1998). Interpreting face images using active appearance models, *Proc. 3rd IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 300Ű305

Georghiades, A. S., Kriegman, D.J., Belhumeur, P.N. (1998). Illumination Cones for Recognition under Variable Lighting: Faces, *IEEE CVPR 1998*, pp. 52

Horn, B.K.P., Woodham R.J., Silver (1978). Determining shape and Reflectance Using Multiple Images, In: *A.I. Laboratory Memo 490*, MIT. Cambridge, Mass. (August 1978)

Huang, Y., Lin, S., Li, S.Z., Lu, H., Shum, H.Y. (2004). Face Alignment Under Variable Illumination, *Proceedings of the FGR 2004*, pp. 85-90

Kahraman, F., Gökmen, M., Darkner, S., Larsen, R. (2007). An Active Illumination and Appearance (AIA) Model for Face Alignment, *CVPR 2007*

Kovesi, P. (2005). Shapelets Correlated with Surface Normals Produce Surfaces, *ICCV 2005*, pp. 994-1001

Lee, K.C., Ho, J., Kriegman, D.J. (2001). Nine Points of Light: Acquiring Subspaces for Face Recognition under Variable Lighting, *CVPR 2001*, pp. 519-526

Le Gallou, S., Breton, G., García, C., Séguier, R. (2006). Distance Maps: A Robust Illumination Preprocessing for Active Appearance Models, *VISAPP'06*, vol. 2, pp. 35-40

Matthews, I., Baker, S. (2004). Active Appearance Models Revisited, In: *International Journal on Computer Vision*, vol. 60, pp. 135-164

Ramamoorthi, R., Hanrahan, P. (2001). An Efficient Representation for Irradiance Environment Maps, *Proc. ACM SIGGRAPH*, pp. 497-500

Romdhani, S., Pierrard, J.S., Vetter, T. (2005). 3D Morphable Face Model, a Unified Approach for Analysis and Synthesis of Images, In: *Face Processing: Advanced Modeling and Methods*. Elsevier.

Romdhani, S. (2005). Face image analysis using a multiple feature fitting strategy, In: *Ph.D. dissertation*, Univ. Basel, Basel, Switzerland.

Romdhani, S., Ho, J., Vetter, T., Kriegman, D.J. (2006). Face Recognition Using 3-D Models: Pose and Illumination. *Proceedings of the IEEE*, vol. 94, pp. 1977-1999

Ross, A. (2004). Procrustes analysis, In: *Technical Report*, Department of Computer Science and Engineering, University of South Carolina, SC 29208, www.cse.sc.edu/ songwang/CourseProj/proj2004/ross/ross.pdf.

Sattar, A., Aidarous, Y., Le Gallou, S., Séguier, R. (2007). Face Alignment by 2.5D Active Appearance Model Optimized by Simplex, *ICVS 2007*, Bielefeld University, Germany

Silver, W. (1980). Determining Shape and Reflectance Using Multiple Images, In: *Ph.D. dissertation*, Massachusetts Inst. of Technology, Cambridge.

Woodham, R. J. (1989). Photometric Method for Determining Surface Orientation from Multiple Images, In: *Shape From Shading*, B. K. Horn, Ed. Mit Press Series Of Artificial Intelligence Series. MIT Press, Cambridge, MA, pp. 513-531

Xiao, J., Baker, S., Matthews, I., Kanade, T. (2004). Real-Time Combined 2D+3D Active Appearance Model, *CVPR 2004*, vol. 2, pp. 535-542

Zhang, L., Samaras, D. (2006). Face Recognition from a Single Training Image under Arbitrary Unknown Lighting Using Spherical Harmonics, In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 351-363


**New Approaches to Characterization and Recognition of Faces**
Edited by Dr. Peter Corcoran

ISBN 978-953-307-515-0
Hard cover, 252 pages
**Publisher** InTech
**Published online** 01, August, 2011
**Published in print edition** August, 2011

As a baby, one of our earliest stimuli is that of human faces. We rapidly learn to identify, characterize and eventually distinguish those who are near and dear to us. We accept face recognition later as an everyday ability. We realize the complexity of the underlying problem only when we attempt to duplicate this skill in a computer vision system. This book is arranged around a number of clustered themes covering different aspects of face recognition. The first section presents an architecture for face recognition based on Hidden Markov Models; it is followed by an article on coding methods. The next section is devoted to 3D methods of face recognition and is followed by a section covering various aspects and techniques in video. Next short section is devoted to the characterization and detection of features in faces. Finally, you can find an article on the human perception of faces and how different neurological or psychological disorders can affect this.

**How to reference**
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Salvador E. Ayala-Raggi, Leopoldo Altamirano-Robles and Janeth Cruz-Enriquez (2011). Face Image Synthesis and Interpretation Using 3D Illumination-Based AAM Models, New Approaches to Characterization and Recognition of Faces, Dr. Peter Corcoran (Ed.), ISBN: 978-953-307-515-0, InTech, Available from: http://www.intechopen.com/books/new-approaches-to-characterization-and-recognition-of-faces/face-image-synthesis-and-interpretation-using-3d-illumination-based-aam-models

INTECH
open science | open minds

**InTech Europe**
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

**InTech China**
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen