Computer Vision and Image Understanding 115 (2011) 194–210

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Automatic face interpretation using fast 3*D* illumination-based *AAM* models

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

Instituto Nacional de Astrofísica, Óptica y Electrónica, Coordinación de Ciencias Computacionales, Luis Enrique Erro #1, 72840 Sta Ma. Tonantzintla. Pue., Mexico

ARTICLE INFO

Article history:
Received 10 July 2009
Accepted 8 November 2010
Available online 13 November 2010

Keywords:
Active appearance models
Face interpretation
3*D* face alignment
3*D* model fitting
Face modeling
Illumination modeling
Unconstrained face analysis

ABSTRACT

We present an innovative and fast approach for face interpretation invariant to lighting and pose. Our approach performs interpretation by fitting a parametric 3*D* face model to an input image using an optimization algorithm. The parameters obtained after the fitting process describe the appearance of the face. The fitting process is automatic and only requires a 2*D* position and a scale factor as initialization. The proposed model is a natural 3*D* extension of active appearance models and is based on modeling, separately and simultaneously, 3*D* pose, 3*D* shape, albedo, and lighting. Our model is capable of synthesizing faces with arbitrary 3*D* shape, 3*D* pose, albedo and lighting. In order to fit the model to an input image, we propose a fast optimization algorithm able to fit face images with non-uniform lighting and arbitrary pose. Our algorithm, based on a gradient descent approach, executes a fast update to the Jacobian by using the lighting parameters estimated in each iteration of the fitting process. We show that our method is able to accurately estimate the parameters of 3*D* shape and albedo, which are strongly related to identity. Experimental results, suggest that our model can be extended to face recognition under non-uniform lighting and variable pose.

1. Introduction

Automatic, fast and full interpretation of face images under variable conditions of lighting and pose is one of the more exciting and unsolved problems in computer vision. Interpretation is the inference of knowledge from an image. This knowledge covers relevant information, such as 3*D* 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 retrieving 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 Fig. 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 3*D* shape, 3*D* 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*, 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. Separating these two attributes is not an easy task into the context of sparse models, like *AAMs*. *AAMs* use a sparse set of vertices defining the shape. Texture is interpolated over that shape. In fact, a detailed dense set of surface normals, which is not present 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 used in the form of 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 range. In fact, lighting affects appearance more than identity and pose, and presents many degrees of freedom.

During a fitting process, an initial model is gradually modified in each iteration up to looking alike to the input image. Therefore,

* Corresponding author.
E-mail addresses: saraggi@ccc.inaoep.mx (S.E. Ayala-Raggi), robles@ccc.inaoep.mx (L. Altamirano-Robles), jcruze@ccc.inaoep.mx (J. Cruz-Enriquez).

1077-3142/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2010.11.005

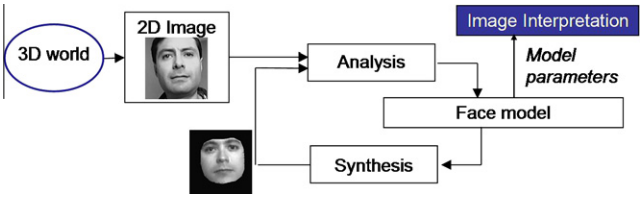

Fig. 1. Schematized flow of the analysis by synthesis approach.

if the lighting of the input image is too different from that 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 [11,23].

In this paper, we present an innovative 3*D* extension of *AAMs* based on an illumination model. By using interpolation, we incorporate a dense set of surface normals to our sparse 3*D 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: 3*D* shape, albedo, 3*D* pose and lighting.

Our approach estimates 3*D* shape, 3*D* 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 offers 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.

Applications to this kind of interpretation algorithms are many, such as face interpretation or recognition with robustness to lighting and pose, face and head tracking in video sequences, human–machine interfaces, algorithms aimed to obtain high levels of data compression in video conferencing, etc.

This paper proposes a novel way to cope with an important source of appearance variation which affects significatively face images: *lighting*. We anticipate that our 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, and expression, where lighting interferes seriously in the analysis process.

2. Related work

Face interpretation is the action of recovering relevant information from a single face image. In contrast to face recognition (a particular case of face interpretation where only identity must be determined), face interpretation can viewed as a more general concept including techniques capable of recovering important aspects such as 3*D* shape, texture, pose, albedo, illumination, expression, age, and gender. The problem of interpreting face images has been addressed in many ways. One of them is the paradigm known as *analysis by synthesis*, where interpretation is performed by synthesizing a face image as similar as possible to the face image to interpret. Recently, the interest in this kind of approaches has increased due to its natural capability of adaptation to novel images. Interpretation is performed by adapting a deformable model to a face image. This process of adaptation is achieved by adjusting iteratively a set of synthesis parameters. The parameters are used for interpretation when this fitting process concludes.

2.1. Approaches based on 3DMM: morphable models

In [7,8], Blanz et al. obtained detailed facial reconstructions by fitting a 3*D* dense model to a face image. Their model, known as 3*D* morphable model, or 3DMM, combines a dense 3*D* shape model and a texture model in order to estimate 3*D* shape and texture parameters related to the input image. In order to construct the model, detailed dense 3*D* scans of different persons are needed. These dense scans are used as a point to point correspondence for computing a statistic and parametric morphable model. 3DMMs are prone to fall in local minima when no proper initialization is given. To alleviate this condition, the convergence properties of 3DMM models have been improved by introducing multi-feature fitting, as in [28], which consists in considering not only pixel intensities, but also high level extracted features such as edges. Thus, it is easier to recover the correspondence between the input image and the model. In a similar way, the accuracy of the fit in 3DMM models has been improved by using *inverse compositional image alignment* (*ICIA*) in [26]. Despite the robustness and accuracy achieved with these new improvements, 3DMMs tend to be slow because of their thousands of vertexes needed for the shape description. On the other hand, 3DMMs are not capable to model every type of lighting. Original 3DMM approach is unable to model diffuse or multi-illuminant lighting.

2.2. Approaches based on 3DMM and lighting models

In an attempt for improving the ability of 3DMMs to handle every kind of lighting, Zhang and Samaras in [38] proposed a method for face recognition (a particular case of face interpretation). The algorithm proposed in [38] uses a 3DMM model to recover 3*D* shape from a single training image. Using the recovered 3*D* shape, the second step in the method is to synthesize a set of lighting basis images of the subject for many possible poses. Recognition is performed by returning the face from the training set for which there exists a weighted combination of basis images that is the closest to the test face image, which is done by solving a minimum squares problem as in [5]. Thus, 3DMM is used for synthesizing a huge set of basis images during the training stage. However, in the test stage, instead of using analysis by synthesis as it is done in the original 3DMM approach, recognition is performed by comparing with each one of the identities stored in this huge gallery. A drawback of this approach is a huge search space which is proportional to the number of considered poses per person and the number of possible identities to interpret. A simplification of the aforementioned method, was proposed by Yue et al. in [37]. Instead of computing basis images with different poses during the training stage, Yue et al., propose to use only basis images with frontal pose. A normalization of pose is done by warping the test face image to a frontal pose with the help of a few manually selected facial features. Thus, recognition is applied again as in [5], and the method is able to create a novel view of the recognized subject but with the same pose that the original test image, by applying a linear transformation of the spherical harmonics basis

images, which synthesizes rotated basis images. Although more efficient than the approach presented in [38], the approach in [37] is based on using the 3DMM model for building sets of frontal basis images per identity, and therefore its search space is again proportional in size to the number of identities to recognize, and unlike original 3DMM, it lacks for ability of adaptation to faces not contained into the training set. A complete 3DMM approach including lighting modeling and following the philosophy of analysis by synthesis, certainly should derive into a generic person-independent face interpretation system, due to the ability of adaptation. Nevertheless, an ideal 3DMM model, that is, at the same time, automatic, fast, person-independent, and capable of modeling every type of lighting (diffuse and multi-illuminant lighting) has not been proposed yet.

2.3. Approaches based on active appearance models

As a counterpart of 3DMM, there is a faster technique for interpreting faces whose computational cost is significatively lower. This approach is known as active appearance models, or *AAMs*. *AAM*s were proposed by Cootes et al. in [10,11], and they are sparse generative models for fast and automatic 2*D* face alignment, often required in real time applications. In a similar way as is done in 3DMM models, in the *AAM* approach, pose and texture variation of the training samples is statistically encoded into a parametric model. The 3*D* pose of the face is encoded in the 2*D* shape parameters, resulting in shape parameters not independent of the face pose. This reduces the recognition capabilities of *AAM*s. Because of its 2*D* nature, original *AAM* allows only small out of the plane rotation, and it does not model neither directed light nor diffuse light. This scheme allows only an effective fit if the lighting during the test stage is equal or similar to the lighting used during the training phase. Original *AAM* models are based on considering a linear relationship between the residual error image and model parameters update, using a constant Jacobian during the whole fitting process. This additive approach of fitting is only a useful approximation which allows a fast convergence process. However, the inherent non-linearities of the image formation process produce an inaccurate fit. For solving the inaccuracies of the original additive method proposed in [11], Baker and Matthews et al. in [4,23], proposed the aforementioned *ICIA* for the 2*D AAM* fitting. The aim of *ICIA* is to improve the accuracy without sacrificing the speed. It is based on keeping a constant Jacobian, but the optimization algorithm updates the entire warp instead of updating the parameters additively. *ICIA* is based on composing the current warp with a computed incremental warp produced by parameter increments, instead of additively incrementing the parameters each iteration, as it is done in the original optimization scheme proposed in [10]. Nevertheless, *ICIA* is a person-specific fitting algorithm, and when it is used for multi-person environments, it gives the same performance as the original additive fitting technique. Due to its 2*D* nature, *ICIA* [23] cannot handle out of the image plane rotation and directed light [27]. Illumination has been considered in *AAM* by some researchers but only for the 2*D* case. In [18] and [22], Huang et al. and Legallou et al. respectively, propose methods for 2*D* face alignment under different illumination conditions. These approaches consist in preprocessing the image to eliminate the effect of lighting before applying the original *AAM* fitting scheme, however, their approaches are not able to recover 3*D* shape and lighting after the fitting process. Kahraman et al. in [19], propose an approach which integrates the original 2*D AAM* model (shape and texture) with a statistical illumination model created from applying *PCA* to a big set of images with different illuminations. Their model, called *AIA* (Active Illumination Appearance model) is 2*D* and consists of two linear subspaces: one for illumination, and another for identity. Although the model proposed in [19] includes a lighting modeling, it does not include the recovery of 3*D* information.

2.4. Approaches based on 3D active appearance models

Original *AAM*s encode the appearance caused by 3*D* pose and 3*D* shape as variation of 2*D* shape. To model the variation of 3*D* pose within the 2*D* approach, it would be necessary to collect data in every possible pose, and that is infeasible in practice. Besides, this mixture of shape and pose leads to the possibility of generating face instances with impossible features. Xiao et al. in [36], propose a 2*D* + 3*D AAM* which exploits the 2*D* shape and 3*D* shape models simultaneously. The shape instance generated by 2*D AAM* is modified to be consistent with a possible 3*D* shape. This constraint is formulated as part of the cost function. To combine this constraint into the original cost function, a balancing weighting is added. The value of this weighting constant is determined manually. Therefore, it is not a direct way for estimating 3*D* pose and shape.

Dornaika et al. in [12], proposed to fit a 3*D AAM* model to each frame of a video sequence for face tracking. They use a generic 3*D* human shape frame called *Candide* developed at Linköping University. The 3*D* model can be modified in 3*D* shape and texture by varying a set of parameters which define inter-person 3*D* shape, intra-person 3*D* shape associated with expression, and finally, texture. The fitting algorithm is similar to that proposed in [11], but it uses, as initialization, the synthetic face image obtained in the last frame. In [31], Sattar et al. proposed a fast method for estimating 3*D* pose based on fitting a 2.5*D AAM* model to a face image by using Simplex. 3*D* structure of training faces is obtained by using frontal and profile views of each face, and labeling both images with associated landmarks. In that work, only pose is estimated by varying six 3*D* pose parameters into an optimization procedure based on Simplex. Fitting additional parameters like pose and texture by using Simplex, certainly would increase the computational cost and the time needed to reach the convergence.

2.5. Approaches which consider lighting

The 3*D* approaches mentioned in the last section work fine in uniform lighting conditions. However, in non-controlled environments like outdoors, or even in most real situations, lighting is not uniform, and the need for modeling it becomes important. An automatic approach for face interpretation should include a lighting model in order to be able to accurately estimate pose, shape and albedo. Lighting has an infinite number of degrees of freedom. Then, in a combined lighting-texture scheme, the variation can not be reduced statistically by using *PCA*, as is done only with texture in [33] and in the classical *AAM* approach [11]. Furthermore, because lighting is independent from texture (albedo), both properties must be separated. The shading of a point over the face surface, depends mainly on the normal of this point and on its reflectance properties (albedo). Therefore, to accurately model the illumination, we require to accurately model the normals. In fact, it is difficult for a sparse 3*D* shape model like *AAM*, to separate shading from texture, because we only known the location of sparse vertices, as is mentioned in [28]. In our work, we show that, in fact, it is possible to model lighting in a sparse model, by interpolating the lacking surface normals within each triangle defined by sparse model vertices.

Regarding illumination, there are important advances in lighting modeling which have been applied to face recognition. However, these approaches, like those presented in [5,6,15,24,34] do not face the problem of face alignment, and 3*D* shape and albedo recovery simultaneously. Lighting over faces has been modeled by considering simplifications in the reflectance properties of face surfaces. Under the Lambertian assumption, the set of images of an

object under all possible lighting conditions forms a polyhedral cone, the illumination cone, in the image space, as it was shown by Belhumeur et al. in [6]. In a follow-up paper [15], Georghiades et al. reported that the illumination cones of human faces can be well approximated by low-dimensional linear subspaces. More recently, using spherical harmonics, Basri and Jacobs in [5], and simultaneously Ramamoorthi et al. in [24], have shown that for a convex Lambertian surface, its illumination cone can be accurately approximated by a 9-dimensional linear subspace. This model is known as 9D subspace model and is useful for its ability to model directed and non-directed light (multiple lights and diffuse light). According to this approach, any reflectance over a face can be approximated in 97.96% of accuracy using a linear combination of nine spherical harmonic reflectances, obtained from the surface normals and the albedos of the face surface.

2.6. Conclusions

Face interpretation has been faced through two paradigms: 3*DMM*s and *AAM*s. 3*DMM*s [7,8,25–29] 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 3*D* information simultaneously.

AAM models have been used for fast 2*D* face alignment under variable conditions of lighting [18,22,19], but not for estimation of 3*D* pose, 3*D* shape, albedo and illumination under non-uniform lighting conditions, which is still a challenging problem. In contrast, some authors [9,12,31,36] 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. In this paper, we propose 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 [11]. 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. Our fitting method called 3*D* illumination-based active appearance models have been published in [1–3] where they have been used for face alignment and pose estimation. In this paper, we present an extended and detailed explanation of the method, measuring its capability to recover 3*D* shape and albedo, and showing its capability to create unseen novel views of interpreted face images. Our experimental results, performed with real face images, show that the method can be extended to face recognition with invariance to lighting and pose.

3. Modeling lighting

Human face can be considered approximately as a convex surface with *Lambertian* reflectance [5,24]. In [5], 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 ith pixel of the object, in this case the face surface. In [5], 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 Eq. (1) we see that this precision is achieved with only nine harmonic images, and Eq. (1) can be expressed in matrix notation as

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

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

3.1. Forcing the lighting model to be positive

By using Eq. (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 [5], 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, Eq. (2) can be modified as

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

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 nine 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 [21], Lee et al., found an important result about how to approximate the illumination cone of lighting (see [14]) 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 9×9 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.

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

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

4.1. 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 3*D* face surfaces of different individuals, and their corresponding 2*D* albedo maps. This set of surfaces and albedo maps will be used to train models of 3*D* shape and 2*D* albedo, respectively.

4.1.1. 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* [17,32,35,13]. 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* [20]. 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.

4.2. Constructing the models of shape and albedo

In order to obtain a parametric 3*D* 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 3*D* 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 3*D* shape models first, by using an iterative algorithm based on Procrustes analysis (see Fig. 2).

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

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

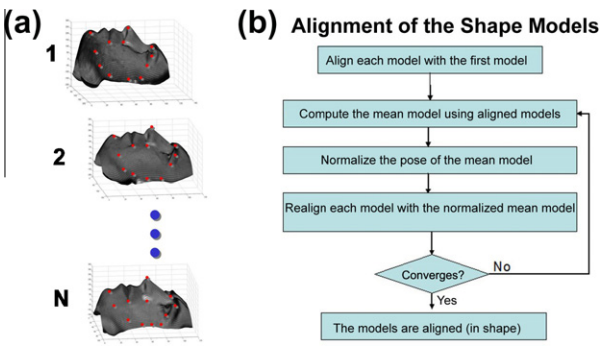

Fig. 2. The shape models (each one defined as the set of landmarks over a particular face surface) (a) must be aligned by using Procrustes Analysis ([30]) (b) before performing the statistical study of shape variation.

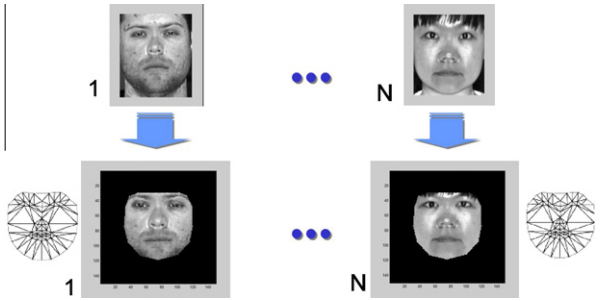

Fig. 3. Normalizing in shape the albedo images by warping the original albedo images into the 2*D* projection of the mean shape. Top: Original albedo images. Bottom: shape-normalized albedo images.

where $\bar{\mathbf{s}}$ 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 2*D* 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 Fig. 3.

A triangulation is designed to warp original images into the mean shape frame. Finally, any shape-normalized albedo image can be generated with

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

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

4.3. Synthesizing faces with novel appearances

By using Eq. (5), it is possible to synthesize an arbitrary albedo image λ and then warp it to the 2*D* projection of an arbitrary frontal shape generated with Eq. (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 $\mathbf{s}$. 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 nine basis reflectance images as is described in Section 3 by using Eq. (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.

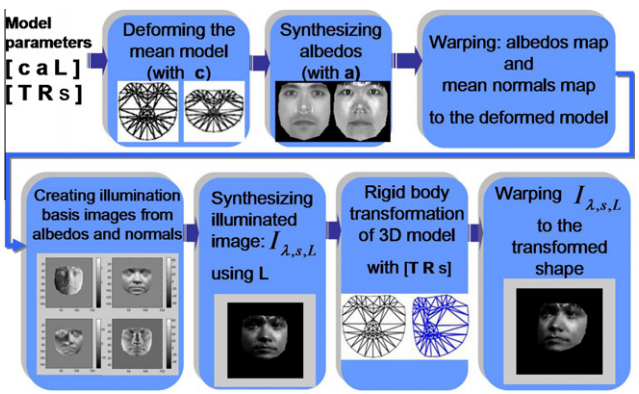

Fig. 4. Face synthesis process.

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 3D object onto a plane by a set of parallel rays orthogonal to the image plane.

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

5. Face alignment using the 3D-IAAM model

The original 2D AAM approach for face alignment presented in [11], 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 [11]. 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.

In this paper, 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.

5.1. 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 4.3.

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.

5.2. 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 $\mathbf{I}_{shape\ aligned}$.

5.3. 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 Eq. (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.

5.3.1. Albedo normalization by using a current estimation of parameters of albedo and illumination

In Section 3 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 Eq. (6) can be rewritten as

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

where λ 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 Eq. (9) and using $\mathbf{I}_{shape\ aligned}$ instead of $\mathbf{I}_{illuminated\ face}$,

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

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 Eq. (5),

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

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

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

where $\widehat{\mathbf{L}}$ is a vector containing the current estimated illumination parameters. We can rewrite Eq. (12) as

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

The residuals vector can be calculated as

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

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

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

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

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

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

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). $\mathbf{c}_0^T = (0,0,0,0,\ldots)$ 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, $\mathbf{a}_0^T = (0,0,0,0,\ldots)$ is the initial albedo parameters vector.

5.4. 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 ijth 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

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

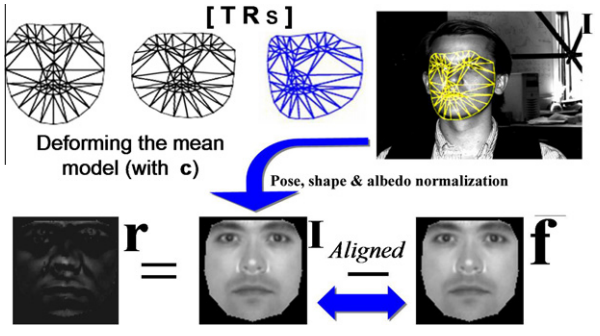

Fig. 5. Estimation of residuals during a step of the fitting process. The mean shape is deformed by using the current parameters $\mathbf{c}$ and $\boldsymbol{\tau}$ (top). Then, the region within the 2*D* projection of this new structure is warped from the test image to the reference mean shape frame (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.

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

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

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 [11], 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.

5.5. Iterative fitting algorithm

In [11], 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 [11] and in our 3*D* approach [1] when the lighting of the input face is considerably different from the lighting used during the training stage. In our 3*D* 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 3*D* 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, in this paper 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 Eq. (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 3*D* 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:

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

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:

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

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 $\mathbf{L}$:

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

where each column can be calculated as:

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

with $i = 1,2,\ldots,33$. 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 λ_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 $\mathbf{J}(\widehat{\mathbf{L}})$,

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

The iterative fitting algorithm is outlined in Fig. 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 2*D AAM* algorithm [11]. This first residual in combination with the Jacobian (which is a precalculated constant the first time) produces (such as it happens in [11]) 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.

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.

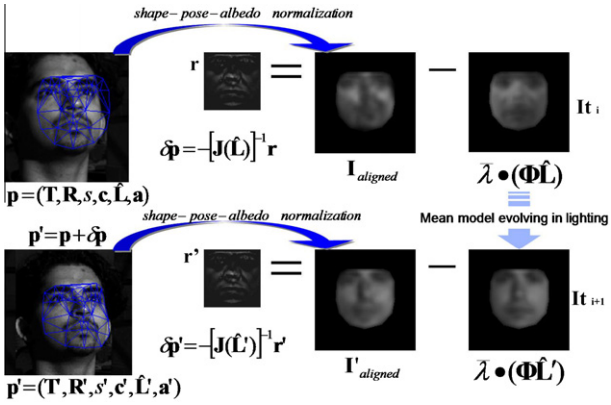

Fig. 7. Two consecutive iterations of the fitting process. During the iteration It_i a region in the test image is captured and normalized according to the current parameters $\mathbf{p}$ producing the image $I_{aligned}$. A residual $\mathbf{r}$ is calculated by comparing $I_{aligned}$ with a reference 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. 7 illustrates two consecutive iterations of the fitting process.

On the other hand, Fig. 8 shows the evolution of the model during the fitting process. Fig. 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.

6. Experimental results

In this section, a set of experiments designed to characterize and demonstrate the capability of the 3*D-IAAM* for performing face interpretation is reported. In particular, the experiments have the objective of demonstrating that the fitting algorithm is able to reconstruct automatically a face image from a single test image which can exhibit a wide range in pose, identity and illumination. We measure the ability of our model to recover 3*D* shape and albedo from face images under different lighting conditions. We show the performance of the fitting algorithm on both situations: using a constant Jacobian, such as is done in the original 2*D AAMs*, and using our proposed method for adapting the Jacobian to the particular present lighting. The results demonstrate a superior performance of the algorithm when an adaptive Jacobian is used.

6.1. Construction of a bootstrap set of face surfaces

6.1.1. Individuals used

We used the 10 identities contained in the Yale B Database [15]. Each subject in the database is photographed in six different poses. For each pose many different illuminations are available.

6.1.2. Constructing a set of face surfaces

The construction of a parametric 3*D* shape model requires of a set of 3*D* face surfaces. Each surface of this set must be manually labeled with a set of landmarks placed over features which are common to all human faces. The shape model will be created by applying a statistical analysis to the set of landmarks belonging to all the surfaces.

Therefore, an important step before the construction of the 3*D* shape model is the creation of a bootstrap set of face surfaces.

We have recovered the face surfaces by mean of the method of photometric stereo previously mentioned.

In practice, we have found that using 11 images per identity is a sufficient quantity for recovering surfaces which are visually similar to the real ones. To avoid surface reconstruction errors caused by cast shadows, we have carefully selected the direction of light sources to be less or equal than 20° from the camera axis.

Each one of these 11 images correspond to a different light source. Fig. 9 shows the recovered surfaces for individuals 1–10.

6.2. Creating a 3D shape model and a template of surface normals

In this approach, we model the 3D shape variation among human faces by applying *PCA* to a set of different shapes each one defined as a set of 3D landmarks over the face. We manually placed 50 landmarks over each one of the face surfaces. On the other hand, all 2D surface normals maps belonging to each individual, were reshaped over the 2D projection of the mean shape in order to obtain a mean map of surface normals. This mean map was used as a template for face synthesis during the construction of the set of basis reflectances matrices $\mathbf{\Phi}$ (see Section 5.5).

6.3. Setup for experiments

The test set for this experiments was composed by 60 real images (with a size of 320 × 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° from the camera axis. We choose six different illuminations for using with each one of the identities. Each illumination is generated by a single point light source, and its direction is specified with an azimuth angle and an elevation angle with respect to the camera axis, see Table 1.

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

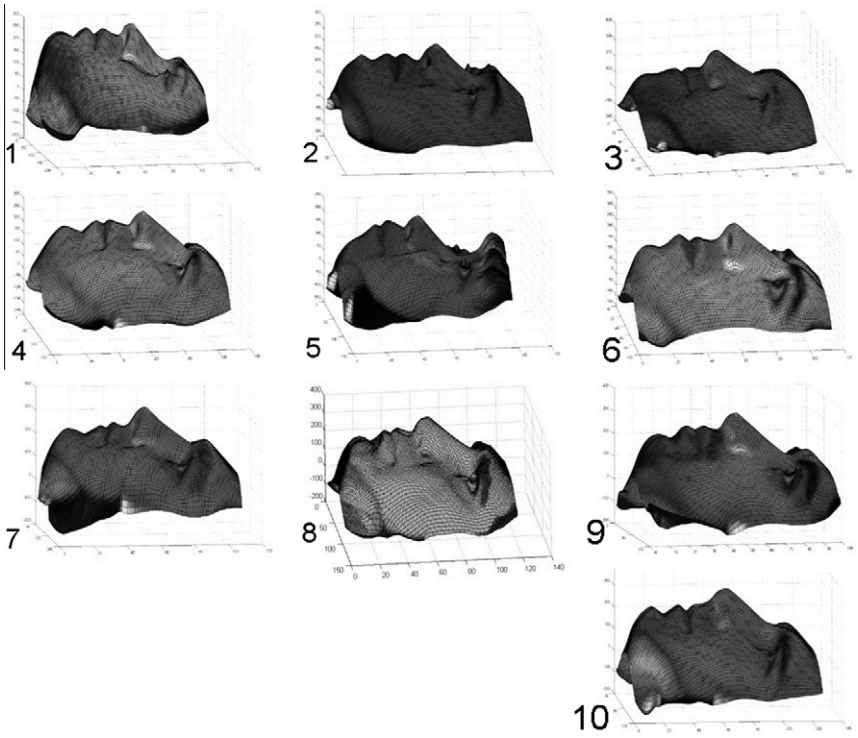

Fig. 9. Surfaces 1–10.

Table 1
Illuminations used for experiments.

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

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, 3*D* shape, illumination and albedo were always initialized in their mean state for all the alignments, i.e., rotation: $\mathbf{R}_0^T = [0,0,0]$, 3*D* 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 3*D* 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 ±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 four computations of the adaptive Jacobian. That is, the algorithm has been allowed to recompute the Jacobian only during the first four consecutive iterations.
2. *Group 2:* 60 alignments using the fitting algorithm programmed with two computations of the adaptive Jacobian. That is, the algorithm has been allowed to recompute the Jacobian only during the first two consecutive iterations.
3. *Group 3:* 60 alignments using the fitting algorithm programmed with a constant Jacobian.

Figs. 10 and 11 show 10 alignments for each one of the identities using pose number 6 and illumination *L*1, and 10 alignments using the same pose and illumination *L*6. In both figures, the first and fourth columns show the original face images, whereas the second and fifth columns show the synthetic faces produced by the fitting process. All the alignments showed in these figures are from Group 1 (four computations of the adaptive Jacobian).

Fig. 12 shows the alignments belonging to Group 1 (four computations of the Jacobian) for each one of the 6 illuminations for identity 7.

6.4. Experiment 1: measuring the convergence of the fitting algorithm

As a measure of similarity between the original face images and the synthetic images produced by our algorithm during the fit, we have used the *RMS* (Root Mean Square) error. In order to show that our algorithm converged in all the test alignments, we obtained a mean *RMS* error for each one of the first five iterations of the algorithm. Fig. 14a shows a decreasing *RMS* error depicted by the central blue curve line, whereas perpendicular straight line segments represent the standard deviation associated to each mean *RMS* error. We observe a mean *RMS* error of 11 gray levels at the first

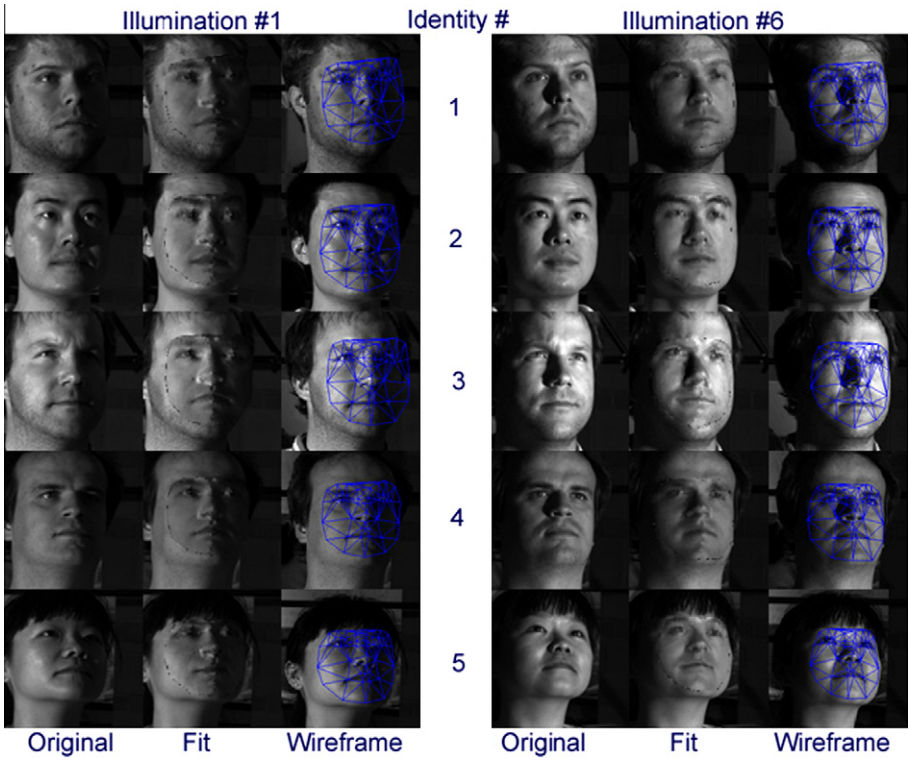

Fig. 10. Face alignments over face images of identities 1–5 with the illuminations L1 and L6.

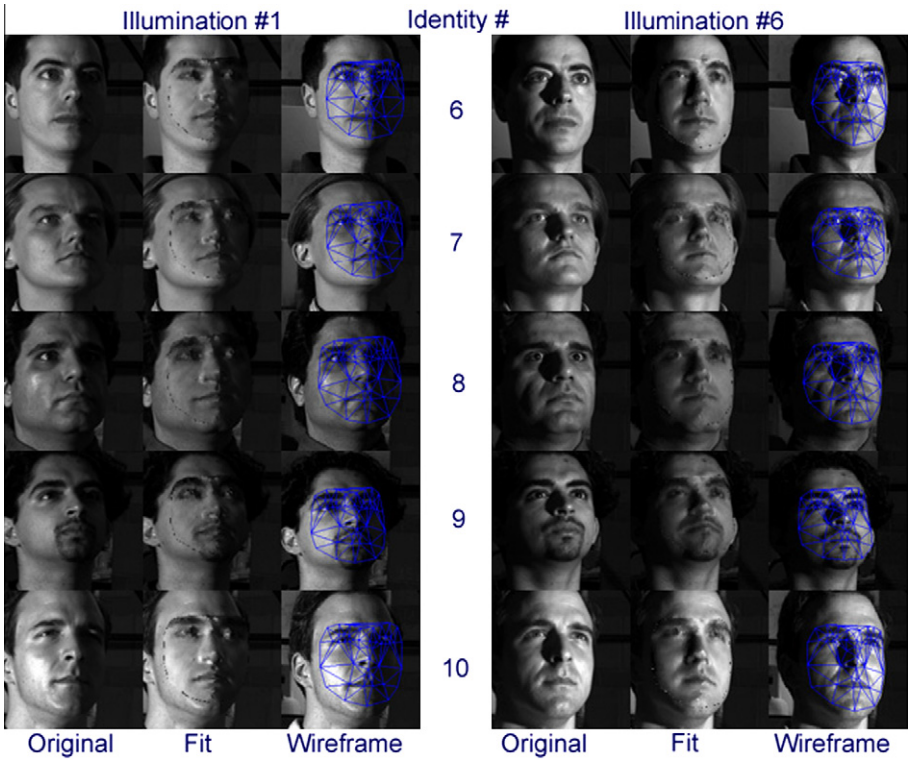

Fig. 11. Face alignments over face images of identities 6–10 with the illuminations L1 and L6.

iteration and a mean *RMS* error of 4 gray levels at the fifth iteration. The same thing occurs with the standard deviation from ±3 in the first iteration to ±1 in the fifth one. Fig. 13 shows the evolution of the model during 12 iterations of the fitting process for identity 7 with illumination 6.

6.5. Experiment 2: recovery of 3D shape and albedo and a measure of quality through identification

In order to measure the quality of the recovered 3*D* shape and albedo, we have considered that this quality is encoded into the

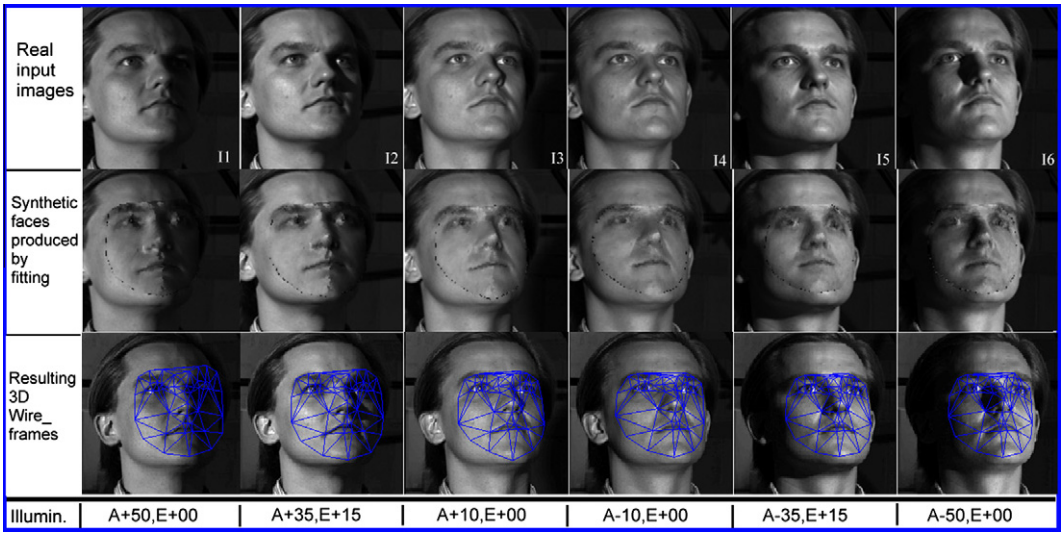

Fig. 12. Alignments for identity 7 with each one of the six different illuminations.

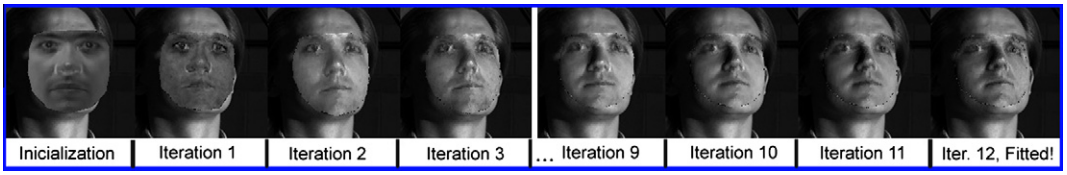

Fig. 13. Evolution of the fitting for identity 7 and lighting 6 in 12 iterations of the algorithm.

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 4.2).

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 [28]). This operation was performed separately for vectors of albedo and for vectors of shape.

If we denote as $\measuredangle(\hat{\mathbf{a}}, \mathbf{a}_i)$ 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:

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

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

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

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, Ω_i^s and Ω_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 $i = 1,2,\ldots,10$) represents the identity likelihood of the estimated albedo for each one of the 10 identities stored in the gallery. Similarly, IL_i^s (with $i = 1,2,\ldots,10$) represents the identity likelihood of the estimated shape for each one of the 10 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: 3*D* shape information of the original face is lost when the 2*D* image is formed. In fact, our fitting approach has to infer a probable shape. On the other hand, albedo which can be considered as 2*D* 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

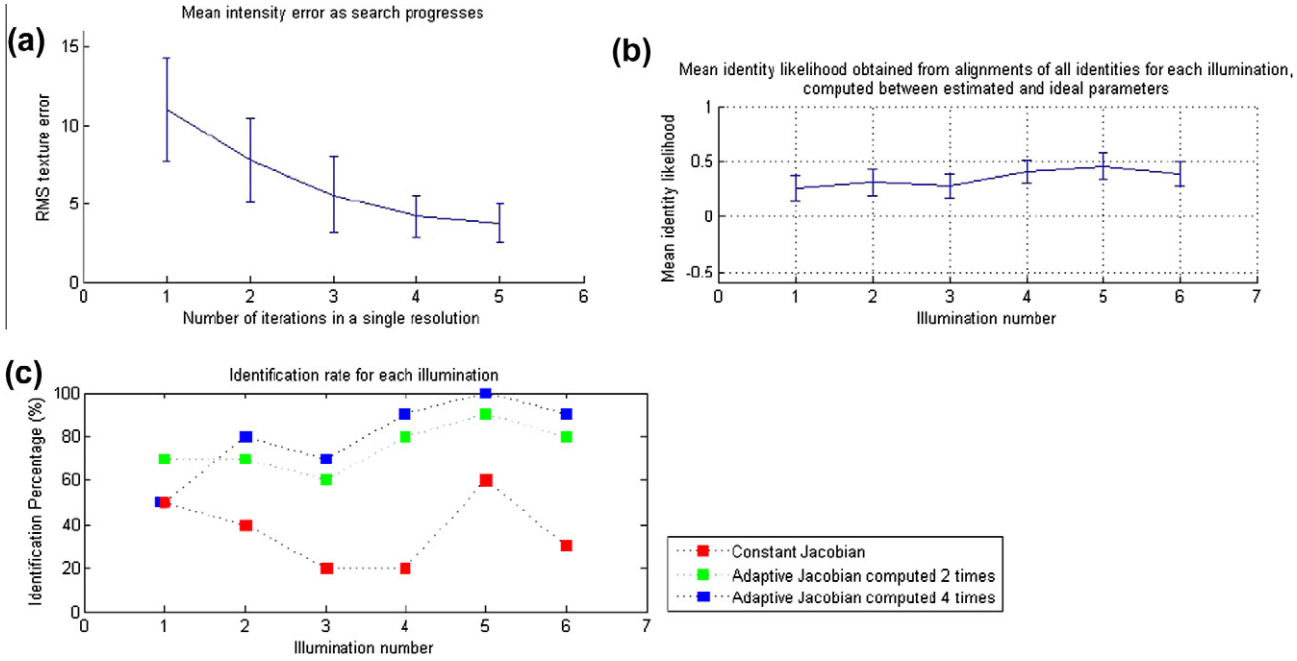

Fig. 14. (a) Evolution of RMS error in intensity difference. (b) Average (over the 10 identities) of the identity likelihood measured between estimated and ideal parameters. (c) Identification rates for each one of the six illuminations.

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) 10 alignments for test images of individuals $i = 1, 2, 3, \ldots, 10$ 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 Fig. 14b. 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). Fig. 14c 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.

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.

We measure the quality of the recovered parameters by means of the identification capability of our fitting algorithm. Therefore, it is convenient to chart the IL computed between the estimated parameters and each one of the 10 sets of real parameters for each identity (stored in the gallery). Each graph in Fig. 15 shows the mean IL measured with respect to each one of the 10 identities from the gallery. The mean IL is the average of the ILs for the 6 illuminations of a single individual. These graphs show the identification capability of our alignment algorithm. Most of the graphs show a peak of the IL value just in the identity number which corresponds to the actual identity of the aligned image. Hence, we can have an idea about the capability of classification of our algorithm. The graphs in Fig. 15 were made with the alignments from Group 1 (4 computations of the adaptive Jacobian).

The fitting algorithm programmed in *MATLAB* with four computations of the adaptive Jacobian takes an average time of 45 s in performing a fit, running in a Pentium IV computer with 2.4 GHz of speed and 3GB in RAM memory. The average number of iterations is 14.

6.6. Experiment 3: creating novel views from fitted faces

Using our approach, we are able to synthesize non-seen novel views of a recovered face from a face alignment. Figs. 16 and 17 show novel synthetic views (right) of the fitted faces (left) created by using the estimated parameters of albedo and shape recovered with the our fitting algorithm from the alignments of Group 1. We only have to modify the parameters of pose and lighting in order to create novel appearances of the same recovered identity. Note that the novel views on the right column have been intentionally illuminated by a lighting which is different from that estimated by the fit. This new lighting illuminates with more intensity the left side of the face.

6.7. Discussion

Since it is difficult to make a mathematical demonstration about the convergence of the fitting algorithm, we have shown through 60 different alignments that the algorithm in fact reduces the energy of

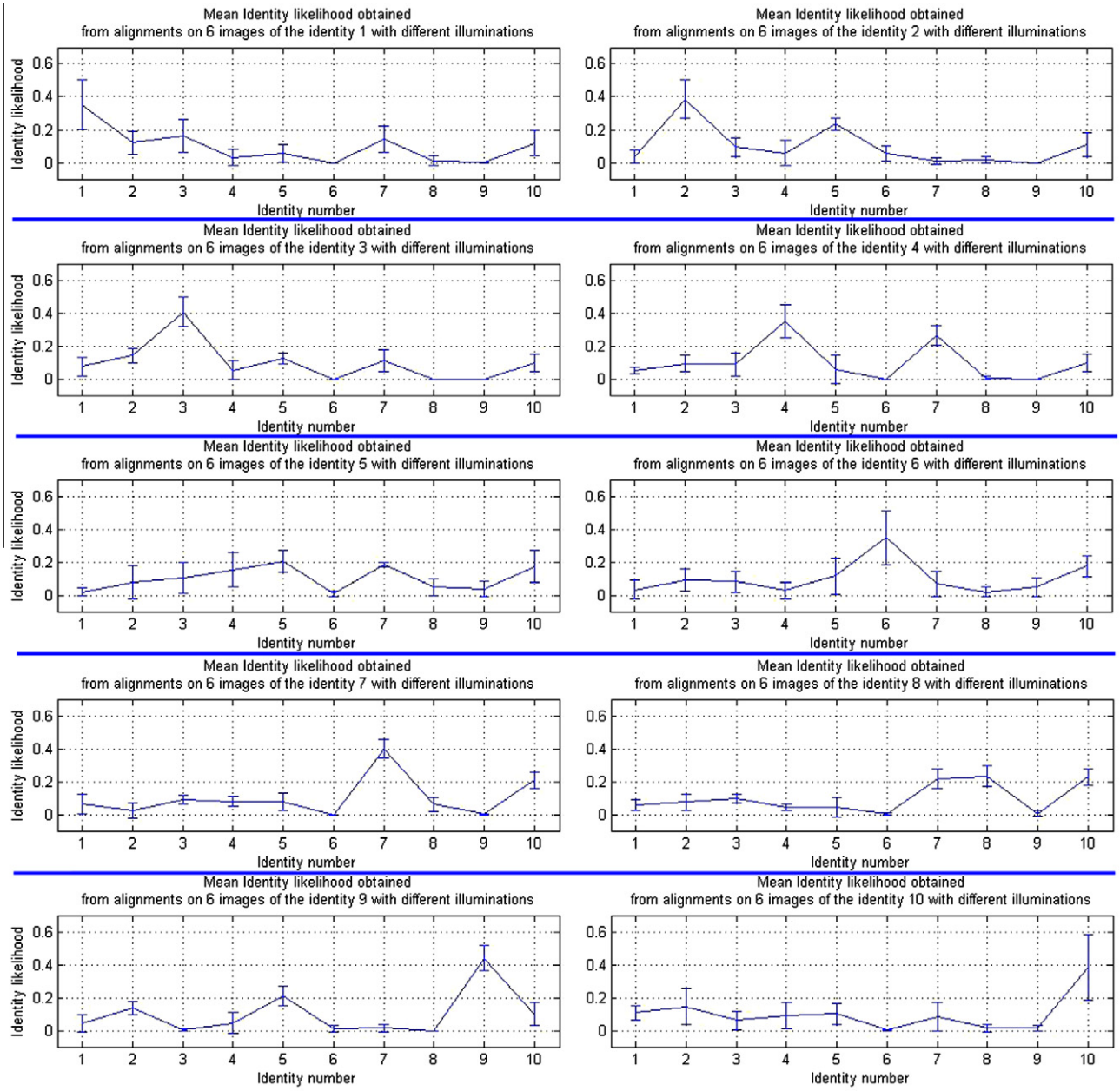

Fig. 15. Average (over the six illuminations) of the identity likelihood measured between estimated parameters and ideal parameters of each one of the 10 identities. The fitting process was performed over six images with different illumination of each identity (1–10).

the residual in each iteration as the optimization progresses. However, convergence cannot be reached when the initial translation is about 10 or more pixels far way from the point in the middle of the eyes, and the scale is greater or smaller about 20% than the actual scale value. In respect to the quality of estimated pose, we obtained an average error of about 3° between the pose of the reconstructions and the actual known pose of all the test images. We did not perform a more complete test of pose estimation because the lack of accurate and enough ground-truth pose data.

Regarding the quality of reconstruction, we can note from Fig. ?? that the use of an adaptive Jacobian improves the quality of reconstruction of the fitting algorithm. The more times the adaptive Jacobian is computed, the more precise will be the reconstruction.

On the other hand, Fig. 15 indicates a way to measure for each identity the probability of being the identity which appears in the picture. This probability or likelihood is a mean probability computed from six alignments of the same face but illuminated by different lightings. From the charts we observe that in some cases there are pairs of identities with similar values of likelihood like identities 4 with 7, identities 2 with 5, and identities 7 with 8. The explanation is related to the similarities between faces encoded within the model parameters.

We have observed through experimentation that this confusion caused by similar probabilities is greater for shape than for albedo. A logical explanation could be that in contrast to albedo, the information regarding to the 3*D* shape is lost when the image formation process is carried out. The fitting method proposed in this paper is capable of inferring 3*D* shape based on *a priori* knowledge obtained during the training stage. Hence, and since there can be two human faces similar in 2*D* appearance but different in depth, it is reasonable that the *depth ambiguity problem* (see [16] for details) might produce errors in the recovery of 3*D* shape.

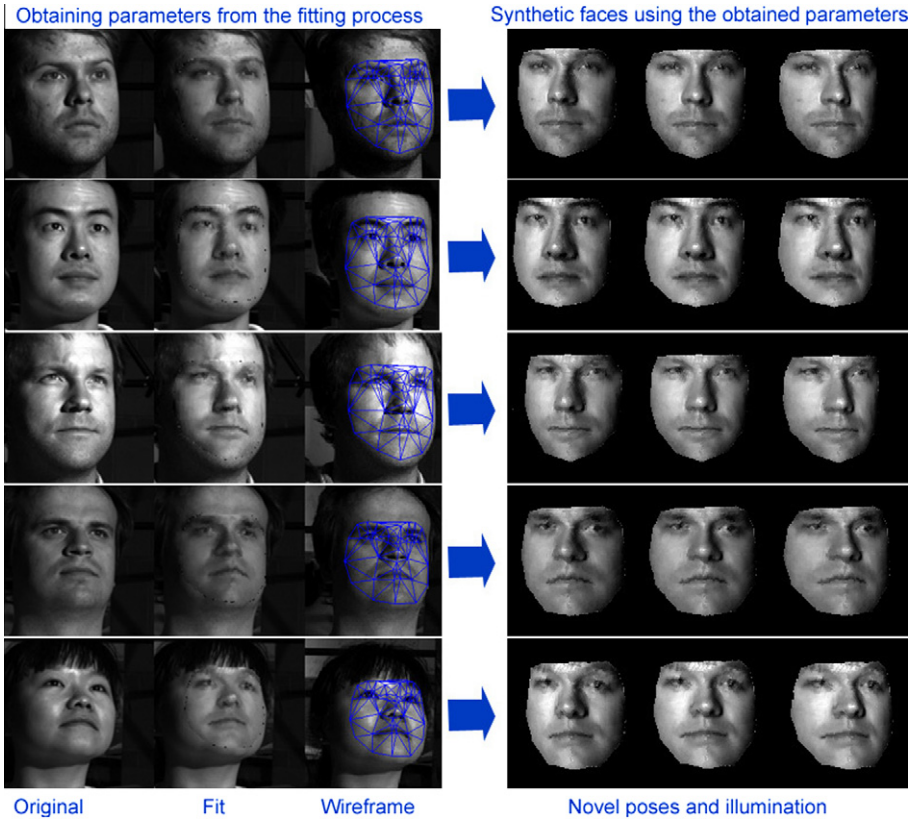

Fig. 16. Novel views of aligned faces by modifying pose and illumination parameters. The alignments were performed over identities 1–5.

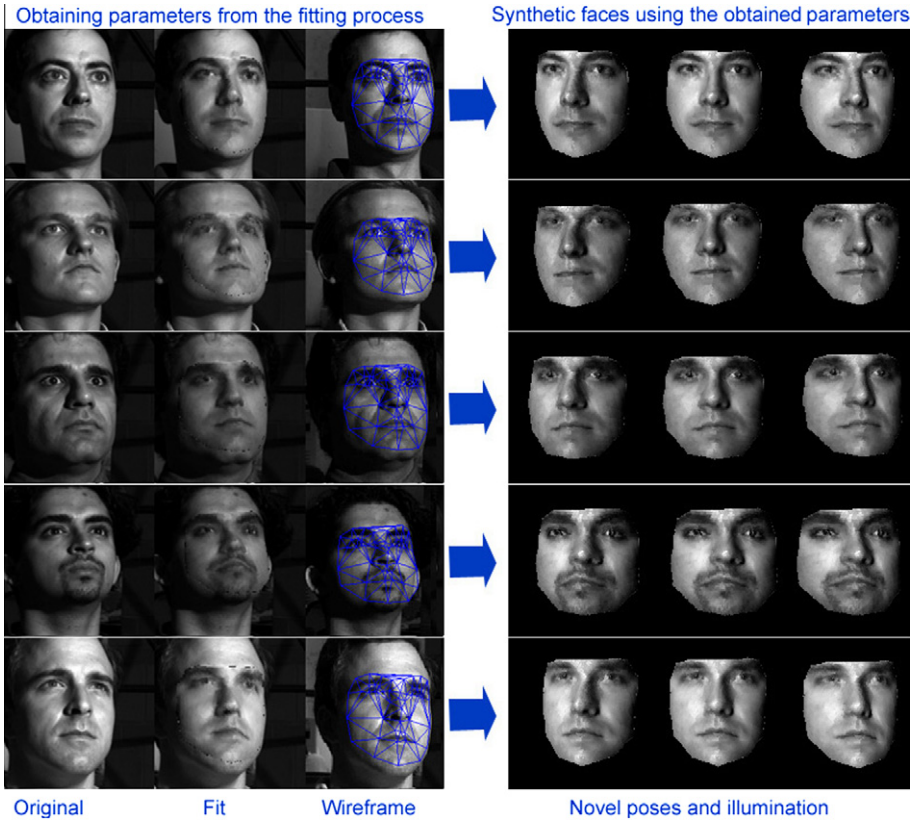

Fig. 17. Novel views of aligned faces by modifying pose and illumination parameters. The alignments were performed over identities 6–10.

Finally, our alignment method is capable of synthesizing novel views of a fitted face. That is an important issue because our model is parametric and recovered parameters provide us with relevant information about the face in the picture. So, creating novel views of that face but using arbitrarily illuminations and poses can be an important task for face interpretation purposes.

7. Conclusions and future work

This paper 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, 3*D*, 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 variation, our fitting method estimates simultaneously, in each iteration, the appropriate increments for parameters of 3*D* shape, 3*D* pose, albedo and lighting. At the end of the fitting process our proposed algorithm provides us with a compact set of parameters of 3*D* pose, 3*D* shape, albedo and lighting which describe the test image.

Our 3*D* face alignment algorithm is robust to non-uniform lighting and is able to fit to different identities with different albedo, shape, pose and illumination. The proposed model is parametric, and we claim that the parameters of 3*D* shape and albedo, recovered by our fitting algorithm, contain the necessary information for identification purposes. By using these estimated parameters, it is possible to identify the individual in the image.

Our 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 parameters of 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, 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.

We evaluated the performance of our fitting method by measuring the quality of the recovered parameters of 3*D* shape and albedo. In order to accomplish that process, we compared the estimated parameters obtained in the test stage to those stored in the gallery for each one of the identities. Then, we obtained identification rates for each one of the lightings used in the experiments. Our results, both quantitative and qualitative, show that the method is able to align a 3*D* 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 with novel poses and novel illuminations from the estimated parameters obtained after the fitting process.

The main contribution of this work is a method for face interpretation based on analysis by synthesis. Our approach is aimed to recover the 3*D* pose, 3*D* shape, albedo and lighting from a single real face image by using the 3*D*-*IAAM* model. The method includes a novel way to normalize the albedo and a novel way for updating the Jacobian in each iteration. Both tasks use the last estimated illumination parameters.

In our approach, the process of creating synthetic faces is used during the synthesis of the basis reflectance images created during the training stage. That set of 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 2*D*, but because our shape model is 3*D*, normals can be reoriented according to the new 3*D* position of each triangular facet. A more accurate representation of lighting should improve the recovery of 3*D* 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 future work. With a careful selection of the faces for the training set, our method can be extended to a generic person-independent automatic 3*D* 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 can be a suitable part of a complete biometric system for identity recognition.

References

[1] S. Ayala-Raggi, L. Altamirano-Robles, J. Cruz-Enriquez, Towards an illumination-based 3*D* active appearance model for fast face alignment, in: CIARP, 2008, pp. 568–575.
[2] S. Ayala-Raggi, L. Altamirano-Robles, J. Cruz-Enriquez, Interpreting face images by fitting a fast illumination-based 3*D* active appearance model, in: Mirage, INRIA, France, 2009, pp. 368–379.
[3] S. Ayala-Raggi, L. Altamirano-Robles, J. Cruz-Enriquez, Recovering 3*D* shape and albedo from a face image under arbitrary lighting and pose by using a 3*D* illumination-based AAM model, in: ICIAR, Halifax, Canada, 2009, vol. 5627, pp. 584–593.
[4] S. Baker, I. Matthews, Equivalence and efficiency of image alignment algorithms, in: Proceedings IEEE Conf. on Computer Vision and Pattern Recognition, 2001.
[5] R. Basri, D.W. Jacobs, Lambertian reflectance and linear subspaces, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2) (2003) 218–233.
[6] P. Belhumeur, D. Kriegman, What is the set of images of an object under all possible illumination conditions, International Journal of Computer Vision 28 (3) (1998) 245–260.
[7] V. Blanz, T. Vetter, A morphable model for the synthesis of 3*D* faces, in: Siggraph, 1999, pp. 187–194.
[8] V. Blanz, T. Vetter, Face recognition based on fitting a 3*D* morphable model, IEEE Transaction Pattern Analysis and Machine Intelligence 25 (9) (2003) 1063–1074.
[9] C.W. Chen, C.C. Wang, 3*D* active appearance model for aligning faces in 2*D* images, in: IROS, 2008, pp. 3133–3139.
[10] T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, in: ECCV, LNCS, vol. 1407, Springer, 1998, pp. 484–498.
[11] T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (6) (2001) 681–685.

[12] F. Dornaika, J. Ahlberg, Fast And reliable active appearance model search for 3D face tracking, in: Proceedings of Mirage, INRIA, France, 2003, pp. 10–11.
[13] D.A. Forsyth, J. Ponce, Computer Vision: A Modern Approach, Prentice Hall, US, 2002.
[14] A.S. Georghiades, D.J. Kriegman, P.N. Belhumeur, Illumination cones for recognition under variable lighting: faces, in: IEEE CVPR 1998, 1998, p. 52.
[15] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: Illumination cone models for face recognition under variable lighting and pose, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (6) (2001) 643–660.
[16] G.G. Gordon, 3D Pose Estimation of the Face From Video, Face Recognition: From Theory to Applications, Springer-Verlag, 1998. pp. 433–455.
[17] B.K.P. Horn, R.J. Woodham, W.N. Silver, Determining Shape and Reflectance Using Multiple Images, Technical Report, A.I. Laboratory Memo 490, Cambridge, MA, MIT, August 1978.
[18] Y. Huang, S. Lin, S.Z. Li, H. Lu, H.Y. Shum, Face alignment under variable illumination, in: Proceedings of the IEEE FGR, 2004, pp. 85–90.
[19] F. Kahraman, M. Gökmen, S. Darkner, R. Larsen, An active illumination and appearance (AIA) model for face alignment, in: CVPR, 2007.
[20] P. Kovesi, Shapelets correlated with surface normals produce surfaces, in: ICCV, 2005, pp. 994–1001.
[21] K.C. Lee, J. Ho, D.J. Kriegman, Nine points of light: acquiring subspaces for face recognition under variable lighting, in: CVPR, 2001, 519–526.
[22] S. Le Gallou, G. Breton, C. García, R. Séguier, Distance maps, a robust illumination preprocessing for active appearance models, in: VISAPP, vol. 2, 2006, pp. 35–40.
[23] I. Matthews, S. Baker, Active appearance models revisited, International Journal on Computer Vision 60 (2) (2004) 135–164.
[24] R. Ramamoorthi, P. Hanrahan, An efficient representation for irradiance environment maps, in: Proc. ACM SIGGRAPH, 2001, pp. 497–500.
[25] S. Romdhani, V. Blanz, T. Vetter, Face identification by fitting a 3D morphable model using linear shape and texture error functions, in: ECCV, 2002, pp. 3–19.
[26] S. Romdhani, T. Vetter, Efficient, robust and accurate fitting of a 3D morphable model, in: ICCV, vol. 2, 2003, p. 59.
[27] S. Romdhani, J.S. Pierrard, T. Vetter, 3D morphable face model, a unified approach for analysis and synthesis of images, in: Face Processing: Advanced Modeling and Methods, Elsevier, 2005.
[28] S. Romdhani, Face Image Analysis Using a Multiple Feature Fitting Strategy, Ph.D. dissertation, Univ. Basel, Basel, Switzerland, January 2005.
[29] S. Romdhani, J. Ho, T. Vetter, D.J. Kriegman, Face recognition using 3-D models: pose and illumination, in: Proceedings of the IEEE, vol. 94, 2006, pp. 1977–1999.
[30] A. Ross, Procrustes analysis, in: Technical Report, Department of Computer Science and Engineering, University of South Carolina, SC 29208, 2004. <www.cse.sc.edu/songwang/CourseProj/proj2004/ross/ross.pdf>.
[31] A. Sattar, Y. Aidarous, S. Le Gallou, R. Séguier, Face alignment by 2.5D active appearance model optimized by simplex, in: ICVS, Bielefeld University, Germany, 2007.
[32] W. Silver, Determining Shape and Reflectance Using Multiple Images, Ph.D. dissertation, Massachusetts Inst. of Technology, Cambridge, 1980.
[33] M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3 (1991) 71–86.
[34] Y. Wang, Z. Liu, G. Hua, Z. Wen, Z. Zhang, D. Samaras, Face re-lighting from a single image under harsh lighting conditions, in: CVPR, June 2007.
[35] R.J. Woodham, Photometric method for determining surface orientation from multiple images, in: B.K. Horn (Ed.), Shape From Shading, Mit Press Series Of Artificial Intelligence Series, MIT Press, Cambridge, MA, 1989, pp. 513–531.
[36] J. Xiao, S. Baker, I. Matthews, T. Kanade, Real-time combined 2D + 3D active appearance models, in: CVPR, vol. 2, 2004, pp. 535–542.
[37] Z. Yue, W. Zhao, R. Chellappa, Pose-encoded spherical harmonics for face recognition and synthesis using a single image, EURASIP Journal on Advances in Signal Processing (2008) (Article ID 748483).
[38] L. Zhang, D. Samaras, Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (3) (2006) 351–363.

Salvador Ayala-Raggi received the BS degree in Electronics from the Autonomous University of Puebla, Puebla, Mexico, in 1993. He obtained the MS degree in electrical engineering from the National Autonomous University of Mexico in 1996. From 1996 to 2005 he was head of technology development at PCTV, a company of television broadcast in Mexico. He was a professor in the Faculty of Engineering at the National Autonomous University of Mexico from 2000 to 2005. He received the PhD degree in the Computer Science from the National Institute of Astrophysics Optics and Electronics in Mexico. From march 2010 to date, he joined to the Computer Vision Laboratory in the same institution, where he is a scientist leader of high technology projects aimed to public, government and industrial sectors in Mexico. His research interests include artificial intelligence, computer vision, and linear subspaces for face image interpretation.

Leopoldo Altamirano-Robles received the BS degree in computer science from the Autonomous University of Puebla, Puebla, Mexico, in 1985. He obtained the MS degree in electrical engineering from CINVESTAV at the National Polytechnical Institute, Mexico, 1991. He received the PhD degree from the Faculty of Informatics, Technical University of Munich, Germany, 1996. From 1997 to date, he is a titular researcher in the computer science department at the National Institute of Astrophysics Optics and Electronics, Mexico. He has been head of the Computer Vision Laboratory in the same institute since 2000. He is author of more than twenty papers in international journals and about forty papers in international conferences. His research interests include 3D image analysis and industrial applications of computer vision.

Janeth Cruz-Enriquez received the BS degree in computer systems engineering from the Tuxtla Gutierrez Technological Institute, Chiapas, Mexico, in 1998. She obtained the MS degree and the PhD degree in computer science from the National Institute of Astrophysics Optics and Electronics, Puebla, Mexico, in 2001 and 2006 respectively. From 2006 to date, she joined to the Computer Vision Laboratory in the same institution, where she is leader of high technology projects aimed to public, government and industrial sectors in Mexico. Her research interests include computer vision and biometric algorithms for recognition.