

Towards an Illumination-Based 3D Active Appearance Model for Fast Face Alignment

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

Instituto Nacional de Astrofísica, Óptica y Electrónica,
Coordinación de Ciencias Computacionales
Luis Enrique Erro #1, 72840 Sta Ma. Tonantzintla. Pue., México
{saraggi, robles, jcruze}@ccc.inaoep.mx

Abstract. A novel 3D active appearance model invariant to illumination is presented. *3D-IAAM* (Tridimensional Illumination-based Active Appearance Model) is capable of representing human faces with any identity, pose and illumination condition and it was tested for face synthesis by creating faces with multiple identities, poses and illuminations. We also propose an illumination-invariant 3D face alignment algorithm based on our model which is suitable for fast estimation of 3D pose and structure of faces. In this work, we model the illumination to do the alignment, instead of eliminating its effects, with the possibility of obtaining additional information about the original lighting at the end of the fitting process.

Keywords: face alignment, active appearance models, pose detection.

1 Introduction

Developing fast 3D face alignment methods capable to work in real time despite the kind of lighting is a challenging problem. 3D Face alignment on 2D images has been solved with Morphable Models [1],[2],[3], however, although the quality of alignment is fine, the computational cost and processing time are high, and often it is necessary to use some anchor points manually located on the novel face to begin the fit. Active Appearance Models [4],[5], (*AAMs*) are generative models intended for fast 2D face alignment often required in real time applications. Identity can be estimated by fitting *AAMs*. Nevertheless, *AAMs* are 2D and sensitive to illumination, particularly when the lighting during the testing phase is different from illumination during the training phase. We propose an extension of the *AAMs* called *3D-IAAM* (Tridimensional Illumination-based Active Appearance Model) for fast 3D face alignment which estimates the 3D pose and structure of faces despite its illumination. Our model is built by 3D shape, albedo, pose and illumination parameters. To model illumination we use harmonic reflectances obtained from the surface normals and albedos maps of faces. We developed a face synthesizer able to create realistic faces with different poses, identities and lightings and an iterative face alignment algorithm based

on the fitting method proposed in [5], but using an adaptive to illumination gradient matrix instead of a constant gradient matrix.

2 Related Work

In [6], it is proposed a technique for obtaining 3D structure from motion using *AAMs* which are fitted with an algorithm originally proposed in [7]. This cited work does not include illumination. In [8] and [9] authors propose methods for 2D face alignment under different illumination conditions by preprocessing the image to eliminate the effect of lighting before applying traditional *AAM* fitting. Kahraman et al [10] propose an approach which integrates the classical *AAM* model (shape and texture) with an statistical illumination model. Their model, called *AIA* (Active Illumination Appearance model) is 2D and consists of two linear subspaces: one for illumination, and another for identity. In [11], it is proposed a 3D algorithm for face tracking in video sequences based on *AAM* models which uses a generic 3D human shape frame called *Candide* developed at Linköping University and does not include the problem of illumination. In [12], Sattar et al. propose a fast face alignment method based on a 2.5D *AAM* model optimized by Simplex. This technique does not consider illumination. On the other hand, to model lighting, Basri et al. [13] showed that any reflectance over a face can be approximated in 97.96% with a linear combination of 9 spherical harmonic reflectances, obtained from the surface normals and the albedos of the face surface. In summary, *AAM* models have been used for fast 2D face alignment under variable conditions of lighting but not for doing 3D pose estimation at the same time. Other authors have proposed extensions of *AAMs* to estimate 3D pose and structure but do not include illumination. We propose a fast and invariant to illumination 3D face alignment method based on *AAMs*, able to estimate 3D pose and structure.

3 Modeling Lighting

In [13], Basri et al show that any illumination over a face can be represented by a linear combination of n basis images

$$\mathbf{I} = \mathbf{B}\mathbf{H}^T\mathbf{L} \quad (1)$$

where \mathbf{L} is a vector containing n arbitrary parameters and \mathbf{B} is a matrix which columns are nine spherical harmonic images built by using an albedos map and a surface normals map. Columns in \mathbf{H} contain samples of the harmonic functions, whereas its rows contain the spherical harmonics transform of delta functions (punctual light sources). To obtain a good approximation we should use a large set of n punctual lights uniformly distributed around the sphere. However, in [14], Lee et al, showed that it is possible to achieve good results in face recognition using only $n = 9$ light punctual sources strategically distributed. This distribution can approximate any reflectance on a face. So, we can build a matrix H using nine deltas in order to project them into the spherical harmonics space.

4 Face Synthesis Using a 3D AAM Model Based on Illumination (3D-IAAM)

To build an statistical 3D shape model, we place 3D landmarks over the surface of faces with different identities. We define a *shape model* as the set of these landmarks. In order to obtain principal modes of shape variation (due to identity), we align the shape models and apply *PCA* to them, so we can generate an arbitrary model with

$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{Q}_s \mathbf{c} \quad (2)$$

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

$$\lambda = \bar{\lambda} + \mathbf{Q}_\lambda \mathbf{a} \quad (3)$$

where $\bar{\lambda}$ is the mean albedo image, \mathbf{Q}_λ is a matrix which contains principal albedo variation modes and \mathbf{a} is a vector of arbitrary parameters. Using the previous expression (eq. 3), it is possible to synthesize an arbitrary albedo and then warp it to the 2D projection of an arbitrary shape generated with expression 2. Now, we have a new face with an arbitrary albedo and shape. This face is not illuminated yet. In the same process of warping albedos to the new shape, it is also possible to warp the shape-normalized mean map of surface normals. Now we have an albedo image and a surface normals map shaped over the new shape. Using these maps (albedos and normals), we can create 9 basis reflectance images. Any illumination can be generated by a linear combination of these basis images using eq. 1. To give a pose to the model, we use the 3D landmarks of the new generated 3D shape. By applying a rigid body transformation (T, R, s) to these landmarks it is possible to give any pose and size to the created face. Finally we warp the frontal illuminated face to the 2D projection of the transformed 3D shape. See Fig. 1.

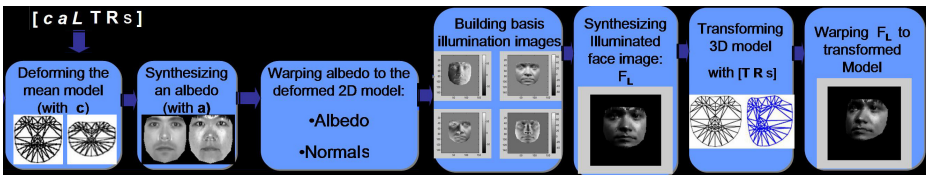


Fig. 1. Face synthesis using the 3D-IAAM direct model

5 Face Alignment Using 3D-IAAM Model

We can consider our face synthesizer as a transformation of the mean face which can result in an arbitrary face, $\mathbf{f} = T_{3D-IAAM}(\bar{\mathbf{f}})$, where $\bar{\mathbf{f}}$ is the mean face, and \mathbf{f} is the resulting synthetic face with arbitrary shape, albedo, illumination and pose. Following the same notation, we should use an inverse transformation for the alignment process:

$$\mathbf{r} = T_{3D-IAAM}^{-1}(\mathbf{I}) - \bar{\mathbf{f}} \quad (4)$$

We propose an extension of the iterative fitting algorithm in [5]. Our technique uses the $3D-IAAM$ inverse model in each iteration. During the fit, according to the parameters of the inverse model, the pixels inside a region in the image are sampled and transformed. So, the residuals image computed with (4) is a function of the model parameters \mathbf{p} . The first order Taylor expansion of (4) gives $\mathbf{r}(\mathbf{p} + \delta\mathbf{p}) = \mathbf{r}(\mathbf{p}) + \frac{\delta\mathbf{r}}{\delta\mathbf{p}}\delta\mathbf{p}$, here, $\mathbf{p}^T = (\mathbf{T}^T|\mathbf{R}^T|\mathbf{s}^T|\mathbf{c}^T|\mathbf{a}^T|\mathbf{L}^T)$, and the ij -th element of the matrix $\frac{\delta\mathbf{r}}{\delta\mathbf{p}}$ is $\frac{\delta r_i}{\delta p_j}$. We wish to choose $\delta\mathbf{p}$ such that it minimize $|\mathbf{r}(\mathbf{p} + \delta\mathbf{p})|^2$. Equating $\mathbf{r}(\mathbf{p} + \delta\mathbf{p})$ to zero leads to the solution

$$\delta\mathbf{p} = -\mathbf{R}\mathbf{r}(\mathbf{p}) \quad \text{where} \quad \mathbf{R} = \left(\frac{\delta\mathbf{r}^T}{\delta\mathbf{p}} \frac{\delta\mathbf{r}}{\delta\mathbf{p}}\right)^{-1} \frac{\delta\mathbf{r}^T}{\delta\mathbf{p}} \quad (5)$$

$\frac{\delta\mathbf{r}}{\delta\mathbf{p}}$ is actually a gradient matrix changing in each iteration. Recalculating it at every step is expensive. So, we assume it to be constant since it is been computed in a normalized reference frame. To compute a constant gradient matrix, we have to displace each parameter from its mean value and create synthetic images. Then, we calculate a weighted average of the residuals images for each displaced parameter as is shown in [5].

5.1 Inverse 3D-IAAM Model

In [13], it is shown that the albedos map really is a constant matrix which multiplies component by component to the harmonic reflectances matrix

$$\mathbf{I}_{illuminated\ face} = \mathbf{B}\mathbf{H}^T\mathbf{L} = (\lambda * \mathbf{R})\mathbf{H}^T\mathbf{L} \quad (6)$$

we use “ $*$ ” and “ $/$ ” for denoting component to component product and component to component division of matrixes (Hadamard product and division) respectively. This notation is commonly used in *MATLAB*. The normalization process applied to a region of the image to be compared with the mean model is:

1. Using the rigid body transformation parameters $(\mathbf{T}, \mathbf{R}, \mathbf{s})$ and the shape parameters \mathbf{c} , we sample a region in the image, and warp this region to the mean shape frame. We call this image as $\mathbf{I}_{shape\ aligned}$. Thus, we can estimate the albedos map which correspond to this image by using eq. 6

$$\hat{\lambda} = (\mathbf{I}_{shape\ aligned}) ./ \mathbf{R}\mathbf{H}^T\mathbf{L} \quad (7)$$

2. Using this estimated albedos map, we can derive an approximated mean albedos map $\bar{\lambda} \approx \hat{\lambda} - \mathbf{Q}_\lambda \mathbf{a}$
3. Finally, the image has been normalized in shape, illumination, albedo and pose according to the model parameters, $\mathbf{I}_{aligned} = (\bar{\lambda}) * \mathbf{RH}^T \mathbf{L}_{mean}$ where \mathbf{L}_{mean} is a predefined vector containing mean illumination parameters, and the expression to minimize will be $\mathbf{r} = \mathbf{I}_{aligned} - \bar{\mathbf{f}}$

The inverse model used during alignment is shown in Fig. 2.

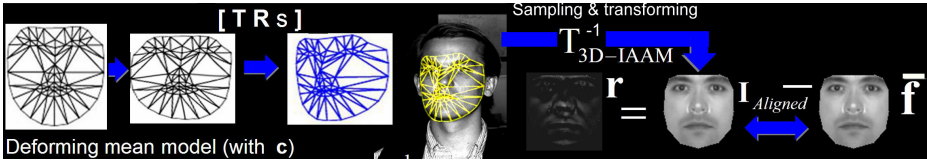


Fig. 2. Inverse 3D-IAAM model used for face alignment

5.2 Iterative Fitting Algorithm

In [5], the fitting process consists of computing a constant gradient matrix (Jacobian matrix) which is used during the fitting process. However when the lighting of the face to fit is very different from the lighting used in the training phase, the alignment fails. Therefore, we propose to use a gradient matrix adapted to the rough illumination estimated in the first iteration. This matrix can be quickly computed during the alignment process using the first estimated illumination parameters obtained from a precomputed constant gradient matrix. We can compute this gradient matrix by displacing each parameter (we use 33) in a suitable increment $p_i + \Delta p_i$. We have to obtain 9 basis images without albedo information: $\mathbf{B}_{p_i + \Delta p_i}$, here $i = 1, \dots, 33$. Increment Δp_i should be both positive and negative, thus, we will have 66 matrixes “B”. In a similar way, by displacing each parameter with a suitable increment $p_i - \Delta p_i$ (positive and negative) we obtain 33 albedo images for positive increments and 33 albedo images for negative increments. These albedo images do not have information about lighting: $\lambda_{p_1 + \Delta p_1}, \dots, \lambda_{p_{33} + \Delta p_{33}}$ and $\lambda_{p_1 - \Delta p_1}, \dots, \lambda_{p_{33} - \Delta p_{33}}$. These 66 matrixes and 66 albedo images are created during the training phase (*off-line*). During the alignment phase we can create a gradient matrix *on-line* according to a set of parameters of illumination \mathbf{L} , $\frac{\delta \mathbf{r}}{\delta \mathbf{p}} = [\frac{\delta \mathbf{r}_1}{\delta p_1} \dots \frac{\delta \mathbf{r}_{33}}{\delta p_{33}}]$ where $\frac{\delta \mathbf{r}_i}{\delta p_i} = [\frac{\delta \mathbf{r}_i^{(+)}}{\delta p_i} + \frac{\delta \mathbf{r}_i^{(-)}}{\delta p_i}] \times \frac{1}{2}$ and

$$\frac{\delta \mathbf{r}_i^{(+)}}{\delta p_i} = \frac{\mathbf{B}_{p_i + \Delta p_i} \mathbf{L} * \lambda_{p_i + \Delta p_i} - \mathbf{B}_{p_i} \mathbf{L} * \lambda_{p_i}}{\Delta p_i} \tag{8}$$

$$\frac{\delta \mathbf{r}_i^{(-)}}{\delta p_i} = \frac{\mathbf{B}_{p_i - \Delta p_i} \mathbf{L} * \lambda_{p_i - \Delta p_i} - \mathbf{B}_{p_i} \mathbf{L} * \lambda_{p_i}}{-\Delta p_i} \tag{9}$$

We obtain a rough estimate of the illumination parameters \mathbf{L} in the first iteration by using a precalculated constant gradient matrix. Then, we restart the iterations using this new created matrix (built from \mathbf{L}) during all the fitting process.

6 Experimental Results

We used Yale database which contains ten different identities with different poses and illuminations. To train a 3D shape model, we placed 3D landmarks over ten 3D face surfaces. We used 11 different images illuminated by different light sources each one. Solving a minimum square problem we estimated surface normals and albedos maps. From surface normals maps we reconstructed the surface of each face. We manually placed 50 landmarks over each surface. Our model uses 33 parameters: 6 for pose, 9 for 3D shape, 9 for illumination, and 9 for albedo. We validated our model qualitatively by synthesizing each one of the training faces in different poses and different lightings. Fig. 3 shows nine synthetic faces each one illuminated by a basis light source.



Fig. 3. Identity number 1 (from Yale database) illuminated by each one of the nine basis light sources

To evaluate face alignment we created two sets of synthetic faces: 13 faces illuminated on the left side, and 13 faces illuminated on the right side. All these with the same identity and different profile angles from -0.3 radians to 0.3 radians of rotation in y -axis. For measuring visual quality of the alignment we used the mean square error (MSE) as a distance measurement. Fig. 4(a) shows three from the thirteen alignments for faces with left illumination and Fig. 4(b) shows three from the thirteen alignments for faces with right illumination. Original synthetic faces are shown in the first row of Figs. 4(a) and 4(b), whereas aligned faces are shown in the second row. Figs. 4(c) and 4(d) show the estimated pose in the y -axis (profile angle) for left illuminated faces and right illuminated faces respectively. Fig. 5 shows results for a similar evaluation now using a fitting algorithm based on a light adaptive gradient matrix. These last results show an improvement in both MSE and profile pose estimation. To show that pose estimation is invariant to illumination and identity, we created 5 poses, each one with a different identity and illumination. Identities were selected randomly: 1, 3, 7, 8, and 10 from the 10 identities contained in the Yale database. Illuminations were randomly selected too from the basis illumination set: 1,3,4,7 and 8, see Fig. 3. Results are shown in Fig. 6.

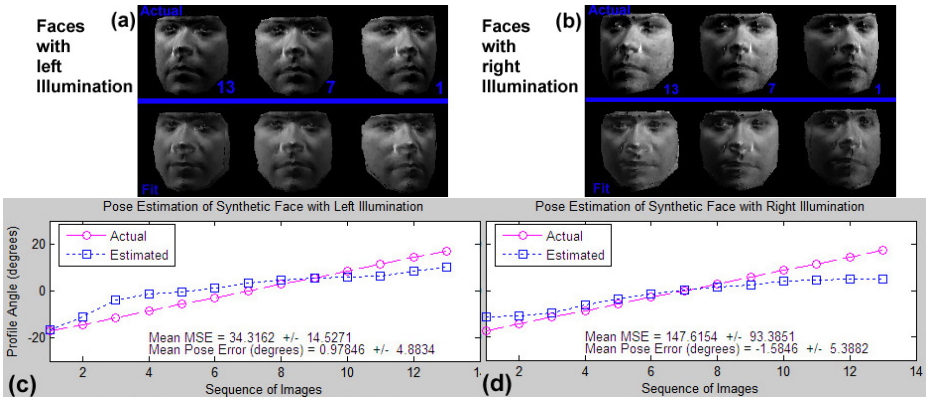


Fig. 4. (a) and (b): Face alignment using a constant gradient matrix. (c) and (d): Estimated poses using the fitting algorithm based on a constant gradient matrix.

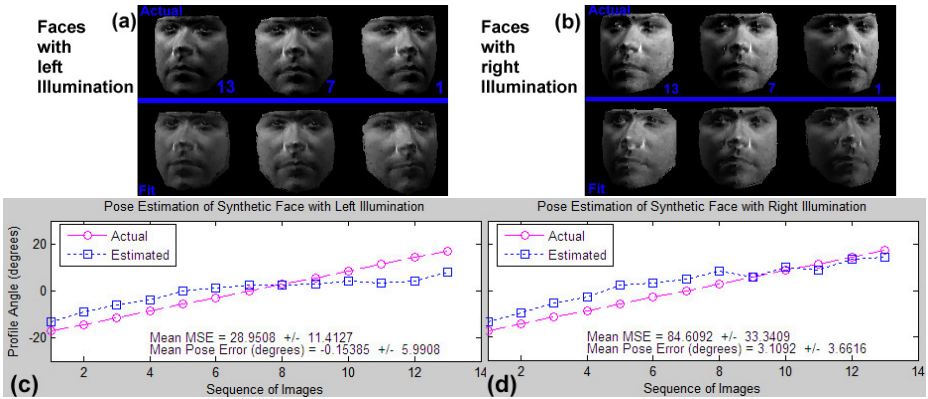


Fig. 5. (a) and (b): Face alignment using a light adaptive gradient matrix. (c) and (d): Estimated poses using a fitting algorithm based on a light adaptive gradient matrix.

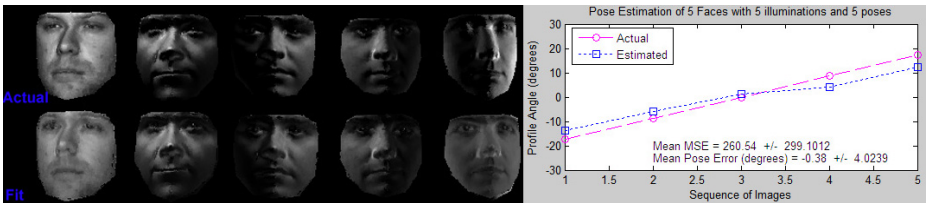


Fig. 6. Left: 5 aligned faces with different illuminations and poses using a light adaptive gradient matrix. Right: estimated poses.

7 Discussion

We estimated 3D poses of faces from 2D images by fitting a parametric 3D active appearance model based on illumination. We have to highlight three contributions here. The first one, is a direct 3D – IAAM model based on lighting, which is able to synthesize faces with multiple 3D spatial poses, 3D shapes, albedos and any possible illumination. The second one is an inverse 3D – IAAM model (used in our face alignment algorithm), which performs a whole normalization (pose, shape, albedo and lighting) of a region in the image to align. Finally, the third one is a fitting algorithm which improves the quality of alignment by creating a gradient matrix which is adaptive to lighting. Our model could be used for face recognition under arbitrary pose and illumination and it would be able of synthesizing new poses and new illuminations of an aligned face.

References

1. Blanz, V., Vetter, T.: A Morphable Model for the Synthesis of 3D Faces. In: *Siggraph 1999*, pp. 187–194 (1999)
2. Blanz, V., Vetter, T.: Face Recognition Based on Fitting a 3D Morphable Model. *IEEE Transaction Pattern Analysis and Machine Intelligence* 25, 1063–1074 (2003)
3. Romdhani, S., Ho, J., Vetter, T., Kriegman, D.J.: Face Recognition Using 3-D Models: Pose and Illumination. *Proceedings of the IEEE 94, 1977–1999* (2006)
4. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active Appearance Models. In: *Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407*, pp. 484–498. Springer, Heidelberg (1998)
5. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active Appearance Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 681–685 (2001)
6. Xiao, J., Baker, S., Matthews, I., Kanade, T.: Real-Time Combined 2D+3D Active Appearance Models. In: *CVPR 2004, vol. 2*, pp. 535–542 (2004)
7. Matthews, I., Baker, S.: Active Appearance Models Revisited. *International Journal on Computer Vision* 60, 135–164 (2004)
8. Huang, Y., Lin, S., Li, S.Z., Lu, H., Shum, H.Y.: Face Alignment Under Variable Illumination. In: *Proceedings of the FGR 2004*, pp. 85–90 (2004)
9. Le Gallou, S., Breton, G., García, C., Séguier, R.: Distance Maps: A Robust Illumination Preprocessing for Active Appearance Models. In: *VISAPP 2006, vol. 2*, pp. 35–40 (2006)
10. Kahraman, F., Gökmen, M., Darkner, S., Larsen, R.: An Active Illumination and Appearance (AIA) Model for Face Alignment. In: *CVPR 2007* (2007)
11. Dornaika, F., Ahlberg, J.: Fast And Reliable Active Appearance Model Search For 3d Face Tracking. In: *Proceedings of Mirage 2003*, pp. 10–11. INRIA Rocquencourt, France (2003)
12. Sattar, A., Aidarous, Y., Le Gallou, S., Séguier, R.: Face Alignment by 2.5D Active Appearance Model Optimized by Simplex. In: *ICVS 2007*, Bielefeld University, Germany (2007)
13. Basri, R., Jacobs, D.W.: Lambertian Reflectance and Linear Subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25, 218–233 (2003)
14. Lee, K.C., Ho, J., Kriegman, D.J.: Nine Points of Light: Acquiring Subspaces for Face Recognition under Variable Lighting. In: *CVPR 2001*, pp. 519–526 (2001)