

# Active illumination and appearance model for face alignment

Fatih KAHRAMAN<sup>1</sup>, Muhittin GÖKMEN<sup>2</sup>, Sune DARKNER<sup>3</sup>, Rasmus LARSEN<sup>3</sup>

<sup>1</sup> Institute of Informatics, İstanbul Technical University, İstanbul, 34469, TURKEY  
e-mail: kahraman@be.itu.edu.tr

<sup>2</sup> Department of Computer Engineering, İstanbul Technical University, İstanbul, 34469, TURKEY  
e-mail: gokmen@itu.edu.tr

<sup>3</sup> DTU Informatics, Technical University of Denmark, DK-2800 Kgs. Lyngby, DENMARK  
e-mail: {sda, rl}@imm.dtu.dk

## Abstract

*Illumination conditions have an explicit effect on the performance of face recognition systems. In particular, varying the illumination upon the face imposes such complex effects that the identification often fails to provide a stable performance level. In this paper, we propose an approach integrating face identity and illumination models in order to reach acceptable and stable face recognition rates. For this purpose, Active Appearance Model (AAM) and illumination model of faces are combined in order to obtain an illumination invariant face localization. The proposed method is an integrated Active Illumination and Appearance Model (AIA) which combines identity, illumination and shape components in a single model and allows us to control them, separately. One of the major advantage of the proposed AIA model is that efficient model fitting is achieved, whilst maintaining performance against illumination changes. In addition to model fitting, images illuminated from different directions can easily be synthesized by changing the parameters related to illumination modes. The method provides a practical approach, since only one image with frontal illumination of each person for training, is sufficient. There is no need to build complex models for illumination. As a result, this paper has presented a simple and efficient method for face modeling and face alignment in order to increase the performance of face localization by means of the proposed illumination invariant AIA method for face alignment, such as the Active Appearance Models, invariant to changes in illumination. From the experimental results, we showed that the proposed AIA model provides higher accuracy than classical Active Appearance Model for face alignment in a point-to-point error sense.*

**Key Words:** AIA, alignment, illumination, face modeling, sub-spaces, active appearance.

## 1. Introduction

Face recognition is an active research topic in image processing and computer vision. Increasing numbers of studies have been observed in face recognition systems under real life condition However, profound difficulties

still need to be studied, including i) changes in pose, ii) changes in illumination source direction and strength, iii) changes in scale, iv) real-time constraints, and v) search in large face databases. The proposed algorithms in the literature can be classified into three categories as sketch-based, feature-based and appearance-based methods. Most of these algorithms assume that the face is imaged on a constant background, or that the face is already segmented from the background. Further, many methods require a frontal view of the face lit using homogeneous illumination. In particular, illumination variation is a challenging and a crucial problem for face recognition because the same individual with the same facial expression may have dramatically different appearances under various illumination conditions [1].

In this paper, we focus on the problems induced by varying illumination. Our primary aim is to eliminate the negative effect of illumination on the face recognition system performance through illumination-invariant face modeling. Several recent studies have centered around this issue: symmetric shape from shading [2], for the illumination cones method [3] theoretically explained the property of face image variations due to light direction changes. In this algorithm, both self shadow and cast-shadow were considered and its experimental results outperformed most existing methods. The main drawbacks of the illumination cone model are the computational cost and the strict requirement of seven input images per person.

Other directions in photometric stereo in face recognition include introducing a more general illumination model. Ramamoorthi [4] has proposed a spherical harmonic representation for face images under various lighting conditions. Basri [5] employs lighting using a spherical harmonic basis, and shows a low-dimensional linear subspace is quite effective for recognition. The harmonic images can easily be computed analytically given surface normals and the albedos. Zhou [6] extended a photometric stereo approach to unknown light sources. Lee [7] empirically found a set of universal illumination directions, images under which can be directly used as a basis for the 9 dimensional illumination subspace.

Shashua [8] employ a very simple and practical image ratio method to map the face images into different lighting conditions. This method is suitable for modeling the variation in facial appearance caused by diffuse reflection and the proposed method is simply the ratio of albedo between a face image and linear combination of basis images, for each pixel. Wang et al. [9] developed reflectance estimation methods by using the idea of the ratio of the original image and its smooth version. In the same research direction Zhang [10] and An [11] proposed new methods to extract an illumination invariant representation of a face images from a raw facial images.

Even though the proposed photometric normalization based representations increase recognition performance, it is not suitable to say that these representations provide complete invariance against illumination. There are many recent works on illumination invariant face recognition. An extensive review of illumination invariant face recognition approaches is given by Zou [12]. The most successful methods for the particular problem of face recognition under varying illumination are image-based [13] [14] [15] [16] [17]. Image-based methods are increasingly used in illumination invariant face recognition due to their robustness with respect to illumination variations [18]. In general, appearance-based methods require training images of individuals taken under different illumination conditions. A method proposed by Sim and Kanade [13] overcomes this restriction by using a statistical shape-from-shading model, in which images are generated for each individual under different lighting conditions, creating a database of images in a recognizer based on principal component analysis (PCA) [19].

Face alignment is also a very important step to extract good facial features to obtain high performance

in face recognition, expression analysis and face animation applications. Several face alignment methods have been propose. For shape alignment Kass et al [20] introduced Active Contour Models which is based on energy minimization; Kirby and Sirovich [21] described statistical modeling of grey-level appearance; Active Shape Models (ASM) [22] and Active Appearance Models (AAM) [23] [24], proposed by Cootes et al, are two successful models for object localization.

ASM uses local appearance models in order to find the candidate shape and a global model to constrain the searched shape. AAM combines constraints on both shape variation and texture variation in its characterization of facial appearance. In searching for a solution, it assumes linear relationships between appearance variation and texture variation and between texture variation and position variation. These two linear regression models are learned from training data. In the context of this paper, texture means the intensity patch contained in the face shape after warping to the mean face shape

Blanz [25] extended the standard AAM to a 3D morphable model (3DMM). The vertices in a 3DMM shape are much more dense than an AAM shape. The 3DMM achieved promising results for illumination invariant face recognition. However, fitting a dense model requires much higher computational effort, which is not suitable for real-time face recognition systems.

Standard AAM is sensitive to illumination, particularly if the lighting during testing is significantly different from the lighting during training [26]. To overcome the problem, we propose a general framework for face modeling under varying illumination conditions. The overview of the proposed method is depicted in Figure 1. We present a robust Active Appearance Model, called AIA, in which face illumination and a face identity sub-spaces are used to model the appearance of the image. Our proposed model combines constraints not only on both shape and texture variation but also on illumination variation in its characterization of facial appearance.

The rest of the paper is structured as follows: Section 2 introduces the identity and illumination modeling methodology to build AIA Model. The experimental results and the conclusion are presented in Sections 3 and 4, respectively.

## 2. Modeling of facial variations

One perspective on the problems induced by varying illumination is that commonly used linear normalization is no longer sufficient to counter-balance the illumination which can be perceived as an undesired noise contamination of the signal of interest: the face geometry and appearance, and ultimately the identity. In general, the illumination problem is quite difficult in face recognition literature. In the case of face recognition many approaches for this problem have been proposed [18].

In principal, the face localization and recognition problem is preferred to be robust to illumination, rotation variations, etc. In this study, we specifically address how to reduce the dependency on variation in illumination on localization performance. We treat the problem as an advanced normalization process, being able to estimate the contribution of a light source to a face image. The final aim of this work is to embed the illumination model into an Active Appearance Model and thus able to estimate and compensate for the actual light conditions. Since the subspace of illumination, identity and shape variations are merged upon AAM in order to obtain illumination-invariant localization of faces, the following subsection is devoted to present AAM in brief. The aim of later subsections is to present the individual variation sub-spaces and the method of combining each of the subspaces the Active Illumination and Appearance Model framework.

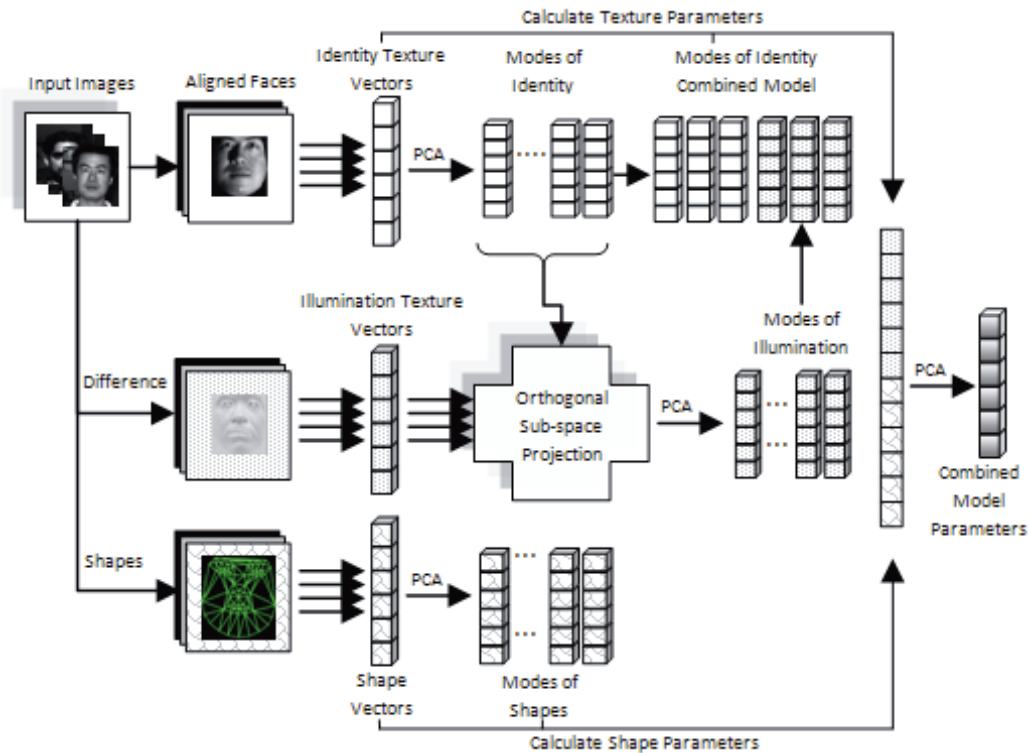


Figure 1. Overview of proposed method.

## 2.1. Active appearance model

Active Appearance Models are generative models capable of synthesizing images of a given object class. By estimating a compact and specific basis from a training set, model parameters can be adjusted to fit unseen images and hence perform image interpretation. The modeled object properties are usually shape and pixel intensities (which we denote as texture). AAM aims to find the optimal model parameters to represent the target image that belongs to the same object class by using an iterative scheme.

Training objects are defined by marking up each image with points of correspondence. Relying upon the landmarks, a triangulated mesh is produced for the reference position and orientation of the object. Before modeling variations, all shape vectors are normalized to a common reference shape frame by using Procrustes Analysis [27]. After obtaining the reference shape vector, all of the training images are warped to the reference shape by using a piecewise affine warping [28] which is defined between corresponding triangles to obtain normalized texture vectors.

Using prior knowledge of the optimization space, AAMs can rapidly be fitted to unseen images with given a reasonable initialization. AAM uses principal component analysis (PCA) to model the variations of the shapes and textures of the images. Usage of PCA representation allows AAM to model and represent a certain image with very small number of parameters.

AAM works according to the following principle: An image is marked with  $n$  landmark points. The content of the marked object is analyzed based on a Principal Component Analysis (PCA) of both texture and

shape. The shape is defined by a triangular mesh and the vertex locations of the mesh. Mathematically the shape model is represented as follows:

$$\mathbf{x}_0 = ((x_1, x_2, x_3, \dots, x_K), (y_1, y_2, y_3, \dots, y_K)) \in \mathbf{R}^{2n}. \quad (1)$$

Texture is the intensities on these landmarks (color pixel values normalized to shape) and is represented with the vector,  $\mathbf{g}$ . Shape is reduced to a more compact form through PCA such that,

$$\mathbf{x} = \bar{\mathbf{x}} + \Phi_s \mathbf{b}_s. \quad (2)$$

In this form,  $x$  is the synthesized shape in the normalized frame,  $\Phi_s$  is a matrix that contains the  $t$  eigenvectors corresponding to the largest eigenvalues and  $\mathbf{b}_s$  is a  $t$ -dimensional vector of shape coefficients. By varying the parameters in  $\mathbf{b}_s$ , the synthesized shape can be varied. All training images warped to the reference shape and sampled into a vector to obtain the texture vectors. Prior to the PCA modeling of the texture we need to normalize all texture vectors. The aim of this normalization is to obtain zero mean and unit variance texture vectors. Texture model can now be obtained by applying PCA to the normalized textures,

$$\mathbf{g} = \bar{\mathbf{g}} + \Phi_g \mathbf{b}_g, \quad (3)$$

where  $\mathbf{g}$  is the synthesized texture,  $\bar{\mathbf{g}}$  is the mean texture and  $\mathbf{b}_g$  is a  $k$ -dimension vector of texture parameters. In the linear model of texture,  $\Phi_g$  is a set of orthogonal modes of variation.

To remove the correlation between shape and texture model parameters, a third PCA is applied on the combined model parameters, giving a further model,

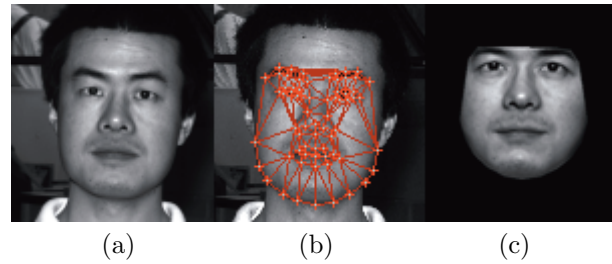
$$\mathbf{b} = \mathbf{Q} \mathbf{c} \quad (4)$$

where  $\mathbf{Q}$  is the eigenvectors and  $\mathbf{c}$  is a vector of appearance parameters controlling both the shape and the texture of the model. Note we do not use a mean vector in this model since the shape, texture and appearance model parameters have to zero mean. Due to the linear nature of the model, the shape and texture vectors can be expressed in terms of the appearance parameters  $\mathbf{c}$  as

$$\mathbf{x} = \bar{\mathbf{x}} + \Phi_s \mathbf{W}_s \mathbf{Q}_s \mathbf{c} \quad (5)$$

$$\mathbf{g} = \bar{\mathbf{g}} + \Phi_g \mathbf{Q}_g \mathbf{c} \quad (6)$$

where  $\mathbf{b} = [\mathbf{W}_s \mathbf{b}_s \quad \mathbf{b}_g]^T$  and  $\mathbf{Q} = [\mathbf{Q}_s \quad \mathbf{Q}_g]^T$ . In this form,  $\mathbf{W}_s$  is a diagonal matrix of weights for each shape parameter, allowing for the difference in units between the shape and the grey models. Generally  $\mathbf{W}_s$  is the square root of the ratio of the total intensity variation to the total shape variation. An example image can be synthesized for a given  $\mathbf{c}$  appearance vector by generating the shape normalized image from the vector  $\mathbf{g}$  and warping it using the control points described by  $\mathbf{x}$  vector. Appearance parameters vector,  $\mathbf{c}$ , is controls both the shape and the grey-levels of the model.  $\mathbf{Q}_s$  and  $\mathbf{Q}_g$  are the eigenvectors of the shape and texture models respectively. An image can be represented by a vector  $\mathbf{p}$  which is given as  $\mathbf{p} = [\mathbf{x} \quad \mathbf{g} \quad \mathbf{c}]^T$ . It's possible to synthesize a new image by changing parameter  $\mathbf{p}$ .



**Figure 2.** Example training image. (a) Input training image. (b) Annotated training image using 73 points. (c) Shape-compensated training image warped to a reference shape.

## 2.2. Building active illumination and appearance model

An identity and illumination model can be established after eliminating variation stemming from pose and shape of face dataset. Such an elimination method can be achieved by

- annotating prominent facial features and
- filtering out effects stemming from pose (translation, rotation and scaling) and shape by a piece-wise affine warp onto a given reference shape.

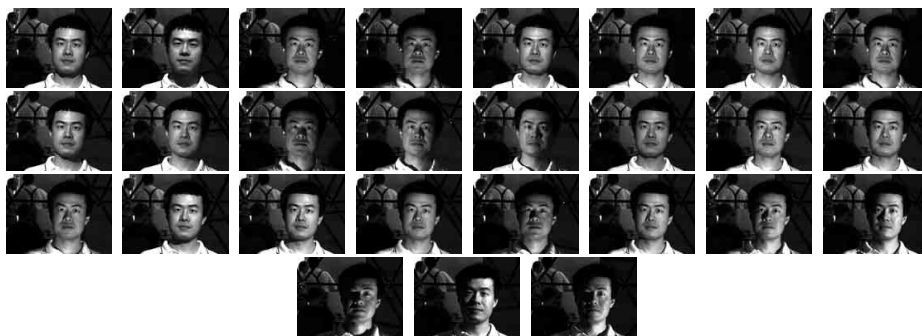
After carrying out the above steps, the undesired variations are removed from the face images. Therefore, the shape-compensated face images are obtained by which we intend to model the illumination and identity variations. Figure 2 shows the method of undesired affine variation removal process. Figure 2(b) shows an annotated training image and Figure 2(c) depicts the shape compensated face relying upon the method of piece-wise affine warp onto a given reference shape. Shape-compensated face images still have main variations which are illumination and identity. These variations are modeled by statistical analysis methods. In particular, Principal Component Analysis is used for the modeling. The resulting principal scores thus give a compact representation of both illumination and identity variations.

The appearance of a frontal facial image from a fixed view point is determined primarily by identity and illumination—two independent factors. It is conceivable that the 3D structure of the face may result in correlation between illumination pattern and the identity. However, since faces have roughly the similar geometry, any face image should have shadows at similar locations with similar 3D shapes caused by an unique point light source. Based on the principle of having similar geometry on faces, we choose to neglect this interaction between illumination and identity, thus allowing for a simple additive model. This simpler model, excluding the identity and illumination interaction, is assumed to have superior predictive power.

We assume an additive model for variations in facial texture due to different identities and different illuminations. Furthermore, we assume independence between identity and illuminations patterns leading to the texture model

$$g = \mu + \Phi_{id} b_{id} + \Phi_{il} b_{il}. \tag{7}$$

Here,  $\mu$  denotes the average facial texture of a frontally illuminated face,  $\Phi_{id}$  spans the space of texture variation, frontally illuminated face due to different identities, and  $\Phi_{il}$  spans the space due to illumination



**Figure 3.** Face images of same person, under 27 different lighting conditions (Yale B).

variation, frontally illuminated face. The first five PCA basis eigen-images of identity and illumination subspaces are shown in Figure 5. In order to estimate these parameters we used a training set from the Yale B facial database, which is comprised of two subsets, *Set 1* and *Set 2*. We estimate  $\mu$  and  $\Phi_{id}$  from *Set 1* that contains frontally illuminated faces of different identity (Figure 4). We can assume that  $\mathbf{b}_{il} = \mathbf{0}$  for *Set 1*. Therefore,  $\mu$  can be estimated as the mean texture and  $\Phi_{id}$  as the set of  $\mathbf{p}_{id}$  eigenvectors of the texture covariance matrix corresponding to the largest eigenvalues.

*Set 2* contains the illumination variations for a single individual (Figure 3). By subtracting the frontally illuminated face texture from all others we have, according to our model, removed all identity variations from this data set. Therefore, we can estimate a set of base vectors spanning this variation  $\Phi_{il}^*$  as the set of  $\mathbf{p}_{il}$  eigenvectors of the texture covariance matrix of these difference images corresponding to the largest eigenvalues. However, because our model requires independence between the identity and illumination texture space we must ensure that the estimates of  $\Phi_{id}$  and  $\Phi_{il}^*$  span orthogonal subspaces. This is achieved by projecting  $\Phi_{il}^*$  into the orthogonal subspace of  $\Phi_{id}$ , i.e.

$$\Phi_{il} = [I - \Phi_{id} \Phi_{id}^T] \Phi_{il}^* \quad (8)$$

For technical reasons we choose to project the observed illumination differences to the identity-orthogonal subspace prior to estimation of the illumination covariance, i.e.

$$\mathbf{d}_{il} = [I - \Phi_{id} \Phi_{id}^T] \mathbf{d}_{il}^* \quad (9)$$

However, this is entirely equivalent to the procedure outlined above. The combined illumination and identity model in Equation (7) can be rewritten,

$$\mathbf{g} = \mu + \Phi_{comb} \mathbf{b}_{comb} \quad (10)$$

where  $\Phi_{comb} = [\Phi_{id} \Phi_{il}]$ . The shape variation is estimated for the Yale B data set 1 in the usual way [23] and a combined shape and texture model is constructed. Figure 6 shows the first three modes of variation for (a) identity and (b) illumination parts of the model.



Figure 4. Identity face images of the Yale B dataset for ten individuals.

### 2.3. Optimization of AIA model

To fit the active appearance model automatically to the target image, we need to solve an optimization problem. AAM and AIA use same method for model searching. The search is treated as an optimization problem in which the difference between the synthesized image by the model and an input image is to be minimized. The method works by learning from an annotated set of training example. For each example in the training set, a number of known model displacements are applied and the corresponding difference vector recorded. Then multivariate regression is applied to model the relationship between image residual and the model displacement. The learnt regression model is then used to predict a movement of the target image model likely to give a better match. By adjusting the AIA parameters ( $\mathbf{b}_{comb}$ ) the model texture,  $\mathbf{g}_m$ , can be deformed to fit the image,  $\mathbf{g}_i$ . In this case the quadratic error norm is applied as optimization criterion which is proposed by Cootes [23]:

$$Error = \sum_{i=1}^n (g_i - g_m)^2 = \sum_{i=1}^n (\delta g_i)^2 = \|\delta g\|^2 = \|r(p)\|^2 \tag{11}$$

Here,  $n$  denotes the number of pixel samples of the shape-free texture vector. The method starts with the initial value for parameter vector  $p_0$ . By minimizing the difference between the target image and the image synthesized by the model, the method try to converge to optimal parameters  $p_{opt}$  for each iteration. We can re-write the  $r(p)$  using Taylor expansion as

$$r(p+\delta p) \cong r(p) + \frac{\partial r}{\partial p} \delta p. \tag{12}$$

Finally, to find optimal  $\delta p$ , to minimize  $\|r(p + \delta p)\|^2$ , we can define  $\delta p$  as

$$\frac{\partial r}{\partial p} \delta p = -r(p) \tag{13}$$

We wish to choose  $\delta p$  so as to minimize  $\|r(p + \delta p)\|^2$ . By equating equation 12 to zero we can obtain the solution

$$\delta p = -Rr(p), \text{ where } R = \left( \frac{\partial r^T}{\partial p} \frac{\partial r}{\partial p} \right)^{-1} \frac{\partial r^T}{\partial p} \tag{14}$$

In the AAM and AIA search procedure,  $\partial r / \partial p$  is defined as a fixed gradient matrix which is estimated at the optimal  $p$  for each training images and then all the parameters estimates are averaged to obtain one fixed





**Figure 5.** The first five PCA basis images of the identity and illumination subspaces. Upper row is the identity, and the lower row shows the illumination eigenfaces.

array. Since the fixed gradient matrix  $R$  is estimated once at model building time, then AIA use pre-computed  $R$  in all subsequent searches with the model. In practice the fixed gradient matrix  $R$  is estimated by a set of experiments on the training set. In the AIA and AAM search procedure, this estimation scheme is applied iteratively to both model and shape perturbation parameters.

### 3. Experimental results

We tested the proposed method on the Yale B Face Database [29]. The first experiment is implemented in order to show illumination modeling can be successfully achieved by means of the face images of only one individual which are taken from different lighting condition, whereas the poses are kept at frontal view. We choose just the frontal subset from the Yale B dataset, containing 300 images from 10 persons, each person has 30 frontal images under 30 different lighting conditions. To have a reasonable range of light source directions, we selected the light directions between  $\pm 60$  degrees in the azimuth angle and  $\pm 45$  degrees in the elevation angle.

To compare the performance of AAM and AIA methods, we choose the most ideally illuminated face image of each individual. The ideally illuminated face should be shadowless, and therefore it is required to use the frontally illuminated one as the identity information. The remaining 29 images under different lighting conditions are used for testing. In order to build up the identity subspace of the AIA, at least one single face being frontally illuminated is required for each individual. On the other hand, illumination part of the AIA is built using the remaining 29 images under different lighting conditions of each individual from the Yale B Dataset. To build illumination subspace of the AIA we used difference images which are aligned to the mean shape and then the frontally illuminated face image for each identity has been subtracted to remove identity information of each images to provide a data set primarily showing effects of varying illumination. In our experiments, we used same training and test images for AAM and AIA to provide a benchmark.

We manually labeled 300 images with the size of  $640 \times 480$  pixels. To establish the models 73 landmarks were placed on each face image; 14 points for the mouth, 12 points for the nose, 9 points for the left eye, 9 points for the right eye, 8 points for the left eyebrow, 8 points for the right eyebrow and 11 points for the chin. The warped images have approximately 33000 pixels inside the facial mask. Using normalized textures, we construct an 8-dimensional texture space to represent approximately 95% of the observed variation.

After building the identity and illumination subspaces of the AIA model, it is possible to synthesize a



**Figure 6.** Mode variation plots of the identity and illumination model. (a) Three largest identity modes from top to bottom:  $b_i = -3\sqrt{\lambda_j}$ ,  $b_i = 0$ ,  $b_i = +3\sqrt{\lambda_j}$ , (b) Three largest illumination modes from top to bottom:  $b_i = -3\sqrt{\lambda_j}$ ,  $b_i = 0$ ,  $b_i = +3\sqrt{\lambda_j}$ .

new face with different identities by only changing the identity parameters of the model as shown in Figure 6(a). Illumination models constructed from the face images belonging to one single individual which are taken under different lighting conditions, can be used to model the face lighting conditions of the other individuals' face images, as shown in Figure 6(b). This characteristic gives us a chance to estimate the lighting conditions and identify of a person by using the AIA model representation.

We can also synthesize faces for various illumination cases by tuning the parameters of the illumination model to obtain the re-lighted version of these faces. This is shown in Figure 7. The uppermost row of the Figure 7 presents an example of how the relighting aspect of AIA model works on mean face which is actually the average of the faces in Figure 4. The other two rows represent the same relighting process for different two identities. This paper demonstrates how an illumination subspace can be learned by applying lighting variations to just one individual and how this can be combined with a face sub space model learned from a database of frontally lit faces. This in fact constitutes an easily applicable way of adapting already learned face models to new lighting conditions.

The manually annotated landmarks are considered as the ground truth shape information. For each image the landmarks derived by the methods are compared by ground truth landmarks. A distance measure,  $D(x_{gt}, x)$ , gives a scalar interpretation of the fit between two shapes, the ground truth,  $x_{gt}$ , and the actual shape,  $x$ . To evaluate the performance using landmarks two distance measures are used. One of them is *point to point error*, defined as the Euclidean distance between each corresponding landmark. This distance measure is hence forth abbreviated to the point to point error (pt.pt.). Average pt.pt. error is expressed in the following equation:

$$D_{pt.pt.} = \sum \sqrt{(x_i - x_{gt,i})^2 + (y_i - y_{gt,i})^2}. \quad (15)$$



Figure 7. Face re-lighting using combined model.

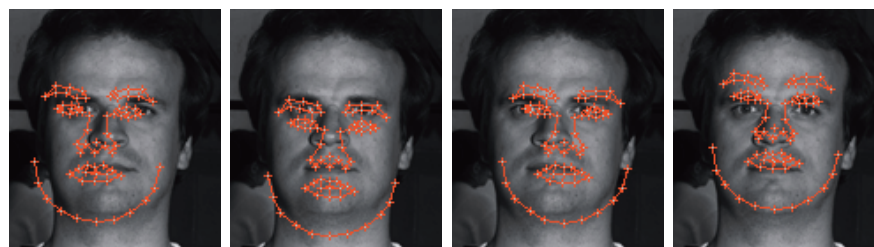


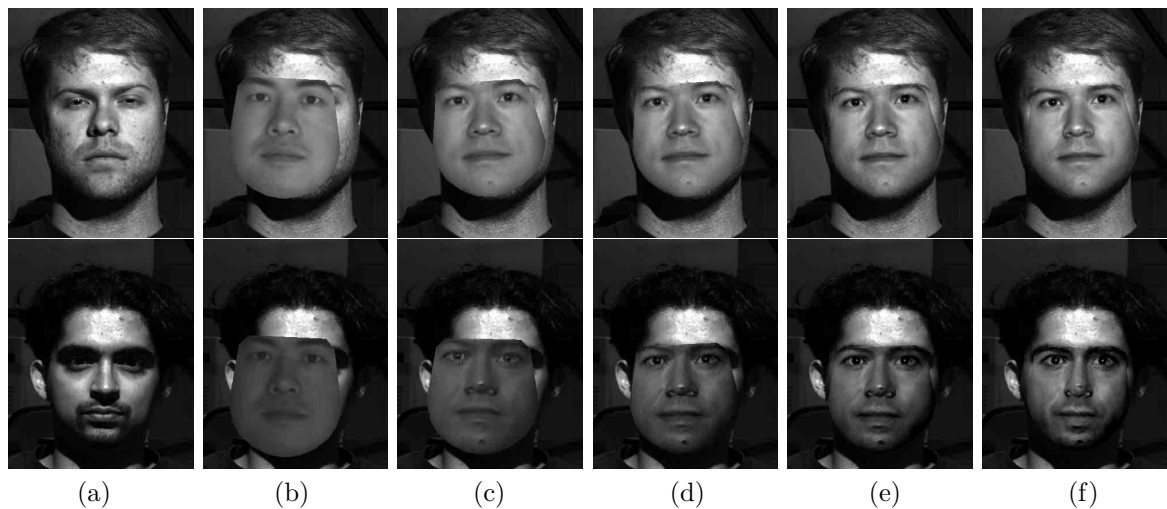
Figure 8. Four different initializations to test the sensitivity of localization performance of AIA and AAM to poor initialization.

The other distance measure is *point to curve error*, defined as the Euclidean distance between a landmark of the fitted shape,  $x$ , to the closest point on the border given as the linear spline,  $r(t) = (r_x(t), r_y(t))$ ,  $t \in [0, 1]$ , of the landmarks from the ground truth,  $x_{gt}$ . This distance measure is abbreviated to the point to curve error (pt.crv.). Average point to associated border error is given as

$$D_{pt.crv.} = \frac{1}{n} \sum_{i=1}^n \min_t \sqrt{(x_i - r_y(t))^2 + (y_i - r_x(t))^2}. \quad (16)$$

The optimization scheme of AAM is inherently sensitive to initialization. AAM converges to the correct solution if it is correctly initialized. To calculate the accuracy of the alignment, we applied the same initializations to AAM and AIA. As initialization the ground truth pose is systematically displaced,  $\pm 20$  pixels in  $x$  and  $y$  coordinates (See Figure 8), at each step the average distance of all initializations are calculated. The comparative results are given in Table 1. It can be easily seen that AIA considerably outperforms original AAM. To match a given image and the model, an optimal vector of parameters are searched by minimizing the difference between synthetic image and input image. Figure 9 and Figure 10 illustrate the optimization and search procedures for fitting the model to the input images.

Examples of the optimization/search results of the proposed method are shown in Figure 9 where the first column is the arbitrarily illuminated unseen images from test dataset and the remaining images are the optimization iterations and rendering of the fitting results for each iteration. The last column presents final



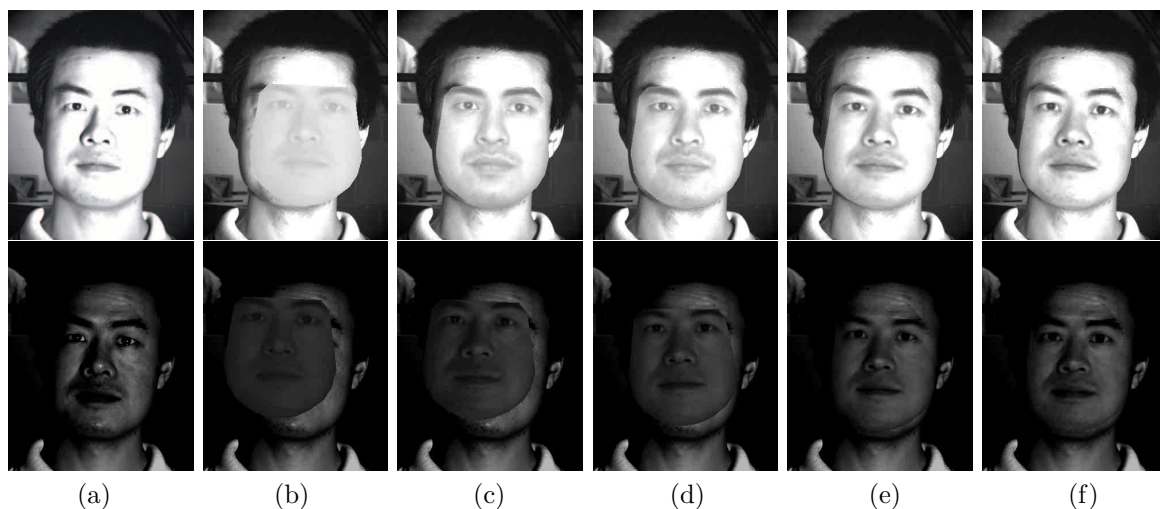
**Figure 9.** AIA model results while optimizing the appearance parameters for test images (unseen) and arbitrarily illuminated faces: (a) input image, (b)–(f) optimization iterations of the AIA.

model approximation for the input images. It can be observed from the last columns that the synthesized faces are very close to input faces.

The AIA optimization is performed as a set of iterations, in each iteration, the model, predicts a set of changes in model parameters leading to a better model to image fit. Using equation (14) we can suggest a correction to make in the model parameters based on a measured residual  $r$ . This allows us to construct an iterative algorithm for solving our optimization problem. This procedure is repeated until no improvement is made to the error,  $\|r(p)\|^2$ . This iterative search procedure is repeated until termination condition is met. Convergence is declared when an error measure is less than a suitable threshold or maximum iteration number is occurred. The minimized error is the sum of the quadratic errors between real image and generated model on each pixel. Usually AIA model would converge in 5–10 iterations; in our implementation a maximum of 30 iterations was allowed in the optimization procedure. We set the convergence level threshold as 0.01 to declare convergence of the algorithm. The optimization scheme of AIA is inherently sensitive to initialization. Please notice that AIA and AAM converge to the correct solution if they are correctly initialized. The complexity of the proposed algorithm is  $O(\mathbf{b}_{comb} \times \mathbf{g}_m)$ . Computation complexity is also measured by computation time of the algorithm. All the computations are performed on Intel Centrino Duo machine (Toshiba Tecra A7, 2GHz with Windows XP) and the AIA search took about 2 seconds.

#### 4. Discussion and conclusion

This paper proposes an approach that combines the face identity and face illumination models, and embeds them into the widely used Active Appearance Model framework as an augmentation to the texture model in order to obtain illumination-invariant localization of faces. In classic AAM formulation there are only two variations, texture and shape. We add illumination variation into the AAM framework in order to build a new combined model containing both identity and illumination. Experimental results using the Yale B database demonstrated the feasibility of the proposed method, showing a significant increase in face localization accuracy.



**Figure 10.** AIA model results while optimizing the appearance parameters for extremely illuminated faces: (a) input image, (b)–(f) optimization iterations of the AIA.

**Table 1.** Face segmentation results for test images ( $640 \times 480$ ).

	Standard AAM	Proposed AIA
Mean pt.-pt. Error	$23.90 \pm 0.38$	$8.85 \pm 0.64$
Mean pt.-crv. Error	$14.70 \pm 0.24$	$5.60 \pm 0.46$
Median pt.-pt. Error	21.62	5.53
Median pt.-crv.	14.29	3.54

The comparative alignment results of the standard AAM and proposed AIA are given in Figure 11 for extremely illuminated test images.

The experiments show that our AIA model can synthesize extremely illuminated faces successfully. For recognition purpose there is no need to use full combined parameters; one needs only the identity part of the final converged combined parameter vector. So, after the AIA is converged, it is easy to re-construct face images using only identity part. In addition to this, we have also illumination parameters. The illumination vector can be used to analyze global lighting (location of light source etc.) and the AIA model can be used to re-light of arbitrarily illuminated input faces.

The method is simple to implement when the basic AAM elements is at hand, e.g. shape alignment, piece-wise affine warping, PCA etc, is also computationally efficient to apply. By being driven by traditional 2D face images in a controlled light setup the method does not require the complex machinery of 3D face model to estimate and synthesize the effects of varying illumination. However, this obviously comes at the cost of establishing a sophisticated controlled light setup for training the system. Luckily, such data sets are now readily available. Hence, we cannot stress our appreciation enough of the Yale B dataset employed in this work.

In this study, a novel method is proposed to automatically align a face image from an image captured under arbitrary lighting conditions. The method requires only one image with frontal illumination of each person for training, which means that it is very practical. There is no need to build complex models for illumination. We demonstrate how an illumination subspace can be learned by applying lighting variations to just one individual



**Figure 11.** Standard AAM alignment results are shown in the first two rows, AIA results are shown in the last two rows respectively for extremely illuminated test images.

and how this can be combined with a face (sub)space model learned from a database of frontally (or if you wish identically) lit faces. This in fact constitutes an easily applicable way of adapting already learned face models to new lighting conditions. As a result, this paper has presented a simple and efficient method for face modeling and face alignment with the primary application of rendering current state-of-the-art methods for face localization, such as the Active Appearance Models, invariant to changes in illumination.

## Acknowledgements

We are grateful to Mikkel B. Stegmann for his early contributions to our ideas. We would like to thank Cenk D. Mendi for his helpful comments on this manuscript. This work is partly supported by the State Planning Agency (DPT) of Turkey and the National Scientific and Research Council of Turkey, project no: EEEAG-104E121.

## References

- [1] Belhumeur P. N., Hespanha J. P. and Kriegman D. J., "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection", IEEE Trans. on PAMI, Vol. 19, No.7, pp. 711-720, 1997.
- [2] Zhao W. and Chellappa R., "SFS Based View Synthesis for Robust Face Recognition," In Proc. of 4th Conf. on AFGR, pp. 285-292, 2000.
- [3] Georghiades A. S. and Belhumeur P. N., "From Few to many: Illumination cone models for face recognition under variable lighting and pose", IEEE Trans. on PAMI, Vol. 23, No. 6, pp. 643-660, 2001.
- [4] Ramamoorthi R., "Analytic PCA Construction for Theoretical Analysis of Lighting Variability in Images of a Lambertian Object", IEEE Trans. PAMI, Vol. 24, No. 10, 2002.
- [5] Basri R. and Jacobs D., "Photometric Stereo with General, Unknown Lighting", IEEE Conf. on CVPR, Kauai, Vol. 2, pp. 374-381, 2001.
- [6] Zhou S., Aggarwal G., Chellappa R. and Jacobs D., "Appearance characterization of linear lambertian objects, generalized photometric stereo, and illumination invariant face recognition", IEEE Trans. on PAMI, Vol. 29, 230-245, 2007.
- [7] Lee J., Moghaddam B., Pfister H. and Machiraju R., "A bilinear illumination model for robust face recognition" IEEE Conf. on ICCV, pp. 1177-1184, 2005.
- [8] Shashua A. and Riklin T., "The Quotient Image: Class-Based Re-Rendering and Recognition With Varying Illuminations", IEEE Trans. on PAMI, pp. 129-139, 2001.
- [9] Wang H., Li S. Z. and Wang Y., "Generalized quotient image", IEEE Proceeding Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 498-505, 2004.
- [10] Zhang Y., Tian J., He X. and Yang X., "MQI Based Face Recognition Under Uneven Illumination", Advances in Biometrics, Vol. 4642, pp. 290-298, 2007.
- [11] An G., Wu J. and Ruan Q., "Kernel TV-Based Quotient Image Employing Gabor Analysis and Its Application to Face Recognition", IEICE Transactions on Information and Systems, Vol. E91-D, No. 5, pp. 1573-1576, 2008.
- [12] Zou X., Kittler J. and Messer K., "Illumination Invariant Face Recognition: A Survey", First IEEE International Conference on Biometrics: Theory, Applications, and Systems, pp. 1-8, 2007.
- [13] Sim T. and Kanade T., "Combining models and exemplars for face recognition: An illuminating example", In Proc. CVPR 2001 Workshop on Models versus Exemplars in Computer Vision, 2001.
- [14] Zhang L. and Samaras D., "Face recognition under variable lighting using harmonics image exemplars", In Proc. IEEE Conf. on CVPR, pp. 19-25, 2003.

- [15] Basri R. and Jacobs D., "Lambertian reflectance and linear sub-spaces", IEEE Trans. on PAMI, 25, pp.383-30, 2003.
- [16] Chen H., Belhumeur P. and Jacobs D., "In search of illumination invariants", In Proc. IEEE Conf. on CVPR, pp. 1-8, 2000.
- [17] Lee K., Ho J. and Kriegman D., "Nine points of lights:acquiring sub-spaces for face recognition under variable lighting", In Proc. IEEE Conf. on CVPR, pp. 519-526, 2001.
- [18] Zhao W. and Chellappa R., "Face Processing:Advanced Modeling and Methods", Academic Press, Elsevier, 2006.
- [19] Gross R., Baker S., Matthews I., Kanade T., "Face recognition across pose and illumination", Handbook of Face Recognition, Springer-Verlag, 2004.
- [20] Kass M., Witkin A., Terzopoulos D., "Snakes, Active contour models," 1st Int. Conf. on Computer Vision, pp. 259-268, 1987.
- [21] Kirby M., Sirovich L., "Application of the Karhunen-Loeve procedure for the characterization of human faces," IEEE Trans. on PAMI, Vol. 12, No. 1, pp. 103-108, 1990.
- [22] Cootes T. F., Taylor C. J. , Cooper D. H. , Graham J., "Active Shape Models - their training and application," Computer Vision and Image Understanding, Vol. 61, No. 1, pp. 38-59, 1995.
- [23] Cootes T. F., Edwards G. J., Taylor C. J., "Active appearance models," IEEE Trans. on PAMI, Vol. 23, No.6, pp. 681-685, 2001.
- [24] Stegmann M. B., Ersboll B. K., Larsen R., FAME - A Flexible Appearance Modelling Environment, IEEE Transactions on Medical Imaging, vol. 22(10), pp. 1319-1331, 2003.
- [25] Blanz V. and Vetter T., "Face recognition based on fitting a 3D morphable model", IEEE Trans. on PAMI, Vol. 25, No.9, pp. 1063-1074, 2003.
- [26] Ming Z., Stan Z. L., Chun C., "Subspace analysis and optimization for active appearance model based face alignment," In Proc. of 6th Conf. on AFGR, pp.290-295, 2004.
- [27] Goodall C., "Procrustes methods in the statistical analysis of shape", Journal of Royal Statistical Society, Vol. 53, pp. 285-339, 1991.
- [28] Glasbey C. A. and Mardia K. V., "A review of image warping methods", Journal of Applied Statistics, Vol. 25, pp. 155-171, 1998.
- [29] Georghiades A. S., Belhumeur P. N., and Kriegman D. J., "From Few to Many: Generative Models for Recognition Under Variable Pose and Illumination", In Proc. of 4th Conf. on AFGR, 2000.