An Active Illumination and Appearance (AIA) Model for Face Alignment

Anonymous CVPR submission

Paper ID ****

Abstract

Face recognition systems are typically required to work under highly varying illumination conditions. This leads to complex effects imposed on the acquired face image that pertains little to the actual identity. Consequently, illumination normalization is required to reach acceptable recognition rates in face recognition systems. In this paper, we propose an approach that integrates the face identity and illumination models under the widely used Active Appearance Model framework as an extension to the texture model in order to obtain illumination-invariant face localization.

1. Introduction

Face recognition is an active research topic in image processing and computer vision. Recent years have seen large efforts in searching for a face recognition system that is capable of working with images captured under ”real-life” conditions. However, profound difficulties remain unsolved including i) changes in pose, mimic, 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 either assume a constant background, or that the face images are 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 problem for face recognition. The same individual with the same facial expression may have dramatically different appearances under various lightning 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 [2] through illumination-invariant face modeling. Several recent studies are centered around this issue: symmetric shape from shading [3], illumination cones method [4] 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 [6]. Ramamoorthi [7] and Basri [8] proposed a spherical harmonic representation for face images under various lighting conditions. Basri et al [8] represent lighting using a spherical harmonic basis wherein a low-dimensional linear subspace is shown to be quite effective for recognition. The harmonic images can easily be computed analytically given surface normals and the albedos. Shashua [10] employ a very simple and practical image ratio method to map the face images into different lighting conditions.

There are many recent works on illumination invariant face recognition. The most successful methods for the particular problem of face recognition under varying illumination are image-based [9] [20] [21] [19] [18]. Image-based methods are increasingly used in illumination invariant face recognition due to their their robustness to illumination variations [17].

Generally, appearance-based methods require training images of individuals taken under different illumination conditions. A method proposed by Sim and Kanade [9] overcomes this restriction by using a statistical shape-from-shading model. Using this method they generate images of the each individuals under different lighting conditions to serve as database images in a recognizer based on PCA [11].

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 were proposed: for shape alignment Kass et al [12] introduced Active Contour Models which is based on energy minimization; Kirby and Sirovich [13] described statistical modeling of grey-level appearance; Active Shape Models (ASM) [14] and Active Appearance Models (AAM) [15], proposed by Cootes et al, are two successful models for object localization.

ASM [14] uses local appearance models to find the candidate shape and global model to constrain the searched shape. AAM [15] 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.

AAM is sensitive to illumination, particularly if the lighting during testing is significantly different from the lighting during training [16]. To overcome the problem, we propose a general framework for face modeling under varying illumination conditions. 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, Section 3 the proposed combined active appearance models. The experimental results and the conclusion are presented in Section 4 and 5, respectively.

2. Face and Facial Illumination Modeling

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 unwanted 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 image-understanding literature. In the case of face recognition, many approaches for this problem have been proposed. For more details about illumination models see [17].

We treat the problem as an advanced normalization process being able to estimate the contribution of a light source onto an arbitrary face image. This estimate can be employed partly in estimating the position of the light source and partly in re-lighting the face image and thus compensating for arbitrary illumination effects. Specifically, an illumination model is built from shape-compensated images of faces with known lighting conditions using a principal component analysis (PCA). The final aim of this work is to embed the illumination model into an Active Appearance Model and thus be able to estimate and compensate for the actual light conditions.

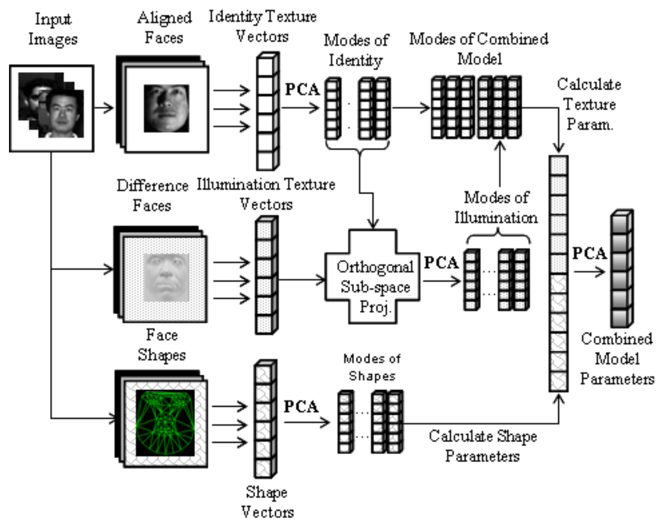

Figure 1. Overview of proposed method.

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 (here denoted texture). Training objects are defined by marking up each example image with points of correspondence. Using prior knowledge of the optimization space, AAMs can be rapidly fitted to unseen images, given a reasonable initialization. Variability is modeled by means of principal component analysis (PCA). Prior to PCA modeling shapes are Procrustes aligned and textures are warped into a shape-free reference frame and sampled. Drawing samples from the respective PCA models of shape and texture can generate synthetic examples by warping the texture samples into the shape samples. Such synthetic examples can now be matched to an unseen image using a least-squares criterion in an iterative updating scheme.

2.2. Identity and Illumination Model

An identity and illumination model can be established after eliminating variation stemming from pose and shape of face dataset. This elimination is in the current work carried out by i) annotating prominent facial features, ii) filtering out effects stemming from pose (translation, rotation and scaling) and shape by a piece-wise affine warp onto a given reference shape. The remaining variation can now be modeled by a principal component analysis of these shape-compensated images by employing the Eckhart-Young the-

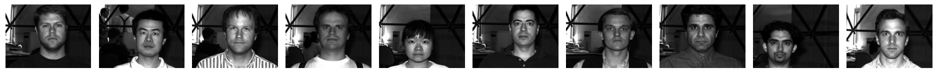

Figure 2. Identity face images of the Yale B dataset for ten individuals.

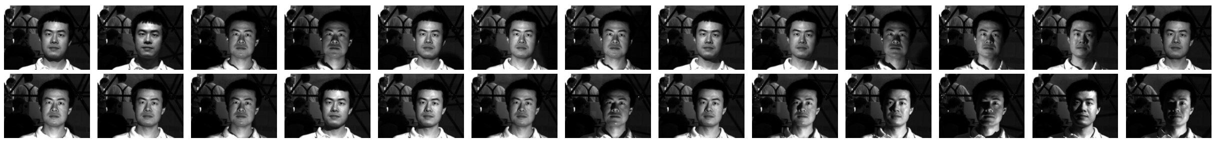

Figure 3. Face images of same person under 24 different lighting conditions (Yale B).

(a) (b)

Figure 4. 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; $b_i = -3\sqrt{\lambda}_j$, $b_i = 0$, $b_i = +3\sqrt{\lambda}_j$

orem (since the number of variables far exceeds the number of observations). The resulting principal scores thus give a compact parametrization of inter and intra-variability due to varying illumination.

Fig. 4 shows the first three modes of variation for (a) identity and (b) illumination parts of the model.

2.3. Building AIA Model

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 following texture model

$$\mathbf{t} = \mu + \mathbf{\Phi}_{\mathit{identity}} \boldsymbol{b}_{\mathit{identity}} + \mathbf{\Phi}_{\mathit{illum}} \boldsymbol{b}_{\mathit{illum}} \tag{1}$$

Here, μ denotes the average facial texture of a frontally illuminated face, Φ_{identity} spans the space of texture variations form the average, frontally illuminated face due to different identities, and Φ_{illum} spans the space due to illumination variations from the average, frontally illuminated face.

In order to estimate these parameters we used a training set from Yale B facial database which is comprised of two subsets, Set 1 and Set 2. We estimate μ and Φ_{identity} from Set 1 that contains frontally illuminated faces of different identity (Fig. 2). We can assume that $\boldsymbol{b}_{\mathit{illum}} = \mathbf{0}$ for Set 1. Therefore, μ can be estimated as the mean texture and Φ_{identity} as the set of p_{identity} eigenvectors of the texture covariance matrix corresponding to the largest eigenvalues.

Set 2 contains the illumination variations for a single individual (Fig. 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 Φ^*_{illum} as the set of p_{illum} 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 Φ_{identity} and Φ^*_{illum} span orthogonal subspaces. This is achieved by projecting Φ^*_{illum} into the the orthogonal subspace of Φ_{identity}, i.e.

$$\Phi_{\mathit{illum}} = [\boldsymbol{I} - \Phi_{\mathit{identity}} \Phi^T_{\mathit{identity}}] \Phi^*_{\mathit{illum}} \tag{2}$$

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.

$$\boldsymbol{d}_{\mathit{illum}} = [\boldsymbol{I} - \Phi_{\mathit{identity}} \Phi^T_{\mathit{identity}}] \boldsymbol{d}^*_{\mathit{illum}} \tag{3}$$

However, this is entirely equivalent to the procedure outlined above. The combined illumination and identity model in Eqn. (1) can be rewritten,

$$\boldsymbol{t} = \mu + \Phi_{\mathit{comb}} b_{\mathit{comb}} \tag{4}$$

where $\Phi_{comb} = [\Phi_{identity} \Phi_{illum}]$. The shape variation is estimated for the Yale B data set 1 in the usual way [15] and a combined shape and texture model is constructed.

3. Experimental Results

We tested the proposed method on the Yale Face Database B [5]. For the experiments we only used one face image for each lighting condition. The size of the face images are 640×480 pixels. The number of images under different lighting conditions for each individual is 29 in our experiments.

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 ± 60 degrees in the azimuth angle and ± 45 degrees in the elevation angle. To compare AAM and AIA methods, we choose the frontal face image under the standard lighting of each person as training images, other 29 images lighted under different for testing.

To examine if any outliers are included in the texture model, all faces are projected onto the first and second texture mode. All outliers are removed from the data set. Corrupted images in the database are also removed. All images that belong to the same individual are selected as an unseen test set and the remaining images are used as a training set to build the illumination model. 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 percent of the observed variation.

It is possible to synthesize a new face in different identities by changing the parameters of the identity model as shown in Fig. 4.(a). 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 as shown in Fig. 5.

Fig. 7 shows that illumination model constructed from one person's face images which are taken under different lighting conditions, can be used to model face lighting conditions of the other face images with different identities. This characteristic gives us a chance to estimate the lighting conditions and identity of a person by using the AIA model representation.

Using a ground truth given by a finite set of landmarks for each example performance can easily be assessed. In a leave-one-out setting this could be the same landmarks used for building the models. This calls out for a distance measure, $D(x_{gt}, x)$, that gives a scalar interpretation of the fit between the two shapes, the ground truth, x_{gt}, and the optimized shape, x. To assess 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. Mean pt.pt. error is expressed in Eqn. (5). This distance measure is here forth abbreviated to the point to curve error (pt.pt.).

Figure 5. Face re-lighting using combined model. Each row contains synthesized images that belong to the same identity. Images at each row are synthesized by changing illumination parameters in the AIA Model for the same identity.

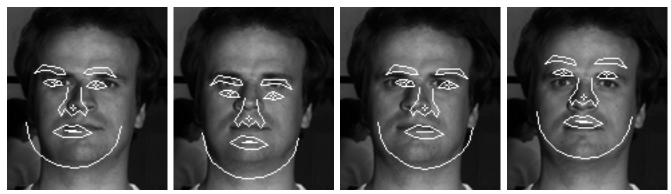

Figure 6. Four different initializations to test the sensitivity of localization performance of AIA and AAM to poor initialization.

$$D_{pt.pt.} = \sum \sqrt{(x_i - x_{gt,i})^2 + (y_i - y_{gt,i})^2} \qquad (5)$$

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, $r(t) = (r_x(t), r_y(t)), t\epsilon[0;1]$, of the landmarks from the ground truth, x_{gt}. Mean point to associated border error is given in Eqn. (6). This distance measure is here forth abbreviated to the point to curve error (pt.crv.).

$$D_{pt.crv.} = \frac{1}{n}\sum_{i=1}^{n} \min_t \sqrt{(x_i - r_y(t))^2 + (y_i - r_x(t))^2} \qquad (6)$$

The optimization scheme of AAM is inherently sensitive to initialization. AAM converges to the correct solution if good initialization is given, but it otherwise prone to the local minima. To calculate the accuracy of the segmentation, we applied same initializations to AAM and AIA. As initialization the ground truth pose is systematically displaced, $\pm$ 20 pixels in x and y coordinates (See Fig. 6), is

Table 1. Face segmentation results for test images (640×480).

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

performed. The comparative results are given in Table 1. It can be easily seen that from the results, 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. Fig. 8 and Fig. 9 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 Fig. 8 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 model approximation for the input images in Fig. 8. It is seen from the last columns of Fig. 8 and Fig. 9 that the synthesized faces are very close to input faces.

4. Discussion and Conclusion

This paper, proposes an approach that combines the face identity and face illumination models and embed 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.

The appearance of a frontal facial image for a fixed camera is determined primarily by identity and illumination - two independent factors. It is conceivable that the 3D structure of the face may result correlation between illumination pattern and the identity. However, since faces have roughly the same geometry we choose to neglect this interaction, thus allowing for a simple additive model. This simpler model excluding the identity and illumination interaction is assumed to have superior predictive power.

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 AAM 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.

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 conclusion, 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.

References

[1] Belhumeur P. N. , Hespanha J. P., and Kriegman D. J. ”Eigenfaces vs. fisherfaces: Recognition using class specific linear projection”. IEEE Trans. Pattern Analysis and Machine Intelligence, 19(7):711-720, 1997. 1

[2] Lee J., Moghaddam B., Pfister H., Machiraju R., ”A Bilinear Illumination Model for Robust Face Recognition”, ICCV 2005, 1177-1184, 2005. 1

[3] Zhao W. and Chellappa R., ”SFS Based View Synthesis for Robust Face Recognition,” Proc. 4th Conf. on Automatic Face and Gesture Recognition, 2000. 1

[4] Georghiades A. S. and Belhumeur P. N., ”From Few to many: Illumination cone models for face recognition under variable lighting and pose”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 23, No. 6, pp. 643-660, 2001. 1

[5] Georghiades A. S., Belhumeur P. N., and Kriegman D. J., ”From Few to Many: Generative Models for Recognition Under Variable Pose and Illumination”, In AFGR, 2000. 4

[6] Wang H., Li Z. S., Wang Y., Zhang W., ”Illumination Modeling and Normalization for Face Recognition”, In Proc. AMFG 2003, 104-111, 2003. 1

[7] Ramamoorthi R., ”Analytic PCA Construction for Theoretical Analysis of Lighting Variability in Images of a Lambertian Object”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 24, No. 10, 2002. 1

Figure 7. Alignment results of the AIA and AAM for extremely illuminated test images. AIA alignment results are shown in the first two rows, standard AAM results are shown in the last two rows respectively.

(a) (b) (c) (d) (e) (f)

Figure 8. AIA model results while optimizing the appearance parameters for test images (unseen) and arbitrarily illuminated faces: a) input image, b) initial approximation (2nd iteration), c) 5th iteration, d) 9th iteration, e) 11th iteration, f) final approximation.

(a) (b) (c) (d) (e) (f)

Figure 9. AIA model results while optimizing the appearance parameters for extremely illuminated faces: a) input image, b) initial approximation (2nd iteration), c) 5th iteration, d) 9th iteration, e) 11th iteration, f) final approximation.

[8] Basri R. and Jacobs D., ”Photometric Stereo with General, Unknown Lighting”, IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Kauai, Vol. II: 374-381, 2001. 1

[9] Sim T., Kanade T., ”Combining models and exemplars for face recognition: An illuminating example”, In Proc. CVPR 2001 Workshop on Models versus Examplars in Computer Vision, 2001. 1

[10] Shashua A. and Riklin-Raviv T., ”The Quotient Image: Class-Based Re-Rendering and Recognition With Varying Illuminations”, IEEE Trans. Pattern Analysis and Machine Intelligence, pp. 129-139, 2001. 1

[11] Gross R., Baker S., Matthews I., and Kanade T., ”Face recognition across pose and illumination”, In Stan Z. Li and Anil K. Jain, editors, Handbook of Face Recognition. Springer-Verlag, 2004. 1

[12] Kass M., Witkin A., and Terzopoulos D., ”Snakes, Active contour models,” 1st Int. Conf. on Computer Vision, London, pp. 259-268, 1987. 1

[13] Kirby M. and Sirovich L., ”Application of the Karhunen-Loeve procedure for the characterization of human faces,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 1, pp. 103-108, 1990. 1

[14] Cootes T. F., Taylor C. J., Cooper D. H., and Graham J., ”Active Shape Models - their training and application,” Computer Vision and Image Understanding, vl.61, no.1, pp. 38-59, 1995. 1

[15] Cootes T. F., Edwards G. J., and Taylor C. J., ”Active appearance models,” IEEE Trans. Pattern Analysis and Machine Intelligence, 23(6):681-685, 2001. 1, 2, 4

[16] Ming Z., Stan Z. L., Chun C., ”Subspace analysis and optimization for active appearance model based face alignment,” The 6th Int. Conf. on Automatic Face and Gesture Recognition. Seoul, Korea, 2004. 2

[17] Zhao W. and Chellappa R., ”Face Processing:Advanced Modeling and Methods”, Academic Press, Elsevier, 2006. 1, 2

[18] 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 Computer Vision and Pattern Recognition, pp. 519-526, 2001. 1

[19] Chen H., Belhumeur P. and Jacops, ”In search of illumination invariants”, In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 1-8, 2000. 1

[20] Zhang L. and Samaras D., ”Face recognition under variable lighting using harmonics image examplars.”, In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 19-25, 2003. 1

[21] Basri, R. and Jacops D. ”Lambertian reflectance and linear sub-spaces”, IEEE Transaction on Pattern Analysis and Machine Intelligence, 25, pp.383-30, 2003. 1