Face Alignment Under Variable Illumination

Yuchi Huang*† Stephen Lin* Stan Z. Li* Hanqing Lu† Heung-Yeung Shum*

* Microsoft Research Asia
† Institute of Automation, The Chinese Academy of Sciences

Abstract

This paper presents an approach to face alignment under variable illumination, an obstacle largely ignored in previous 2D alignment work. To account for illumination variation, our method employs two forms of relatively lighting-invariant information. Edge phase congruency is adopted to coarsely localize facial features, since prominent feature edges can be robustly located in the presence of shading and shadows. To accurately deal with features with less pronounced edges, final alignment is then computed from intrinsic gray-level information recovered using a proposed form of local intensity normalization. With this approach, our face alignment system works efficiently and effectively under a wide range of illumination conditions, as evidenced by extensive experimentation.

1. Introduction

The appearance of a face can change dramatically as the light condition changes, and sometimes variability due to illumination changes is greater than that owing to differences between individual faces [1]. To deal with this problem, several works have addressed the issue of lighting variation in face recognition, such as the Illumination Cone method proposed by Belhumeur and Kriegman [2] and the Quotient Image method proposed by Shashua et al. [3]. These methods assume that faces are already aligned, but there exists little work on how to perform this alignment automatically under variable illumination.

Lighting conditions can have a substantial effect on the robustness and accuracy of face alignment algorithms. Current alignment methods such as Active Shape Models (ASM)[4], Active Appearance Models (AAM)[5] and their extensions [6] attempt to model the appearance of important facial features, but feature search based on these models can become unstable when there exists significant shading and shadowing which can effectively mask subtle features and introduce misleading features as well. These illumination effects can confound the search process and lead the algorithm to local minima, even for relatively uniform lighting.

To reduce the misleading effects of shading, shadows and noise, we propose to use information on prominent edges in the initial stages of search. Generally a model point corresponds to an edge in its locality, and features having prominent edges, such as the eye boundaries, can in general be easily identified in the presence of shading and shadows. Since search based on prominent edge information can effectively locate some subset of the features even under variable illumination, it decreases the likelihood of poor convergence by robustly providing a rough face alignment. Although previous alignment methods require a good initialization to prevent poor shape convergence, we show that an edge-based approach can be robust to bad initializations.

Some edge features such as nose and lip boundaries, however, tend to exhibit gradual or low frequency edges that are sometimes not pronounced enough to be localized in an edge-based search, especially in the presence of shading and shadow variations. For such features, more detailed information is needed, so gray-level data is used. The gray-level information of features lying in deep shading or shadow, however, can become very subtle and difficult to distinguish from surrounding areas, which can lead to incorrect convergence. To deal with this problem, a *patch filtering* technique is proposed to perform a type of local intensity normalization to better recover intrinsic information from local regions. While dependence on only gray-level information throughout the entire search process can lead the alignment process astray, search using gray-level models gives good accuracy *for a good initialization*, given by the edge-based search.

Many methods for shape model fitting employ a hierarchical approach for efficiency and robustness of the search algorithm: at the coarsest level a rough alignment result is computed as a good initialization for a second level in which an accurate face contour is gradually located. Our method takes a similar approach using relatively illumination-invariant features. Our proposed technique initially uses features containing less information, specifically edge phase congruency, for coarse alignment because locating prominent edges is robust even when shading and shadows are present. After coarse alignment, our method utilizes features containing more detailed local information that is emphasized by patch filtering to obtain the final result. With

this proposed *bi-stage approach*, our face alignment system yields accurate, stable and efficient performance under a wide range of illumination conditions, as evidenced in extensive experimentation.

2 Feature Descriptors

In this section, we describe the two relatively illumination-invariant feature descriptors used in our alignment technique. The first one for coarse alignment is based on consistency in edge phases, and the second involves a filter for local intensity normalization.

2.1 Edge Filtering

Gradient-based edge detection methods are sensitive to edge magnitude and smoothness, which can be significantly affected by illumination conditions. In our algorithm, we instead employ phase congruency information in the frequency domain [7] to detect prominent edge features.

Edge phase information at each point is computed locally over multiple frequencies of logarithmic Gabor filters [8]. The amplitude and phase of the transform at a given Gabor wavelet scale is computed as

$$\begin{aligned} A_n(x) &= \sqrt{e_n(x)^2 + o_n(x)^2} \\ \phi_n(x) &= atan2(e_n(x), o_n(x)) \end{aligned} \quad (1)$$

$$[e_n(x), o_n(x)] = [S(x) * M_n^e, S(x) * M_n^o] \quad (2)$$

where M_n^e and M_n^o are the even-symmetric and odd-symmetric wavelets at scale n.

In [7], phase congruency in 1-D is computed as

$$PC(x) = \frac{\sum_n W(x)\lfloor A_n(x)\Delta\Phi_n(x) - T\rfloor}{\sum_n A_n(x) + \varepsilon}. \quad (3)$$

The value T of estimated energy due to noise is subtracted from the local energy to reduce the influence of noise, and a small constant ε is included in the denominator to avoid ill-conditioned calculation when all the Fourier amplitudes are small. $W(x)$ is a weighting function to devalue the phase congruency at locations where the spread of filter responses is narrow, because a point of phase congruency should be significant only if it occurs over a wide range of frequencies. A phase deviation function $\Delta\Phi_n(x)$ increases the sensitivity of the phase congruency measure.

For two-dimensional images, phase congruency is computed over several orientations in the frequency plane to detect all possible 2-D features:

$$PC_{2d}(x) = \frac{\sum_o\sum_n W_o(x)\lfloor A_{no}(x)\Delta\Phi_{no}(x) - T_o\rfloor}{\sum_o\sum_n A_{no}(x) + \varepsilon} \quad (4)$$

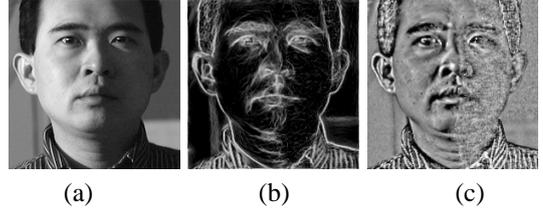where o denotes the index over orientations sampled uniformly over the frequency plane.

Figure 1: The log-intensity face image, its phase congruency image, and its patch-filtered image.

Because phase congruency identifies points in an image where the Gabor components are maximally in phase, it is relatively robust to noise and local variations caused by illumination. In Fig. 1(a,b), a log intensity face image and its normalized phase congruency image is displayed.

2.2 Patch Filtering

Since shading and shadows often diminish the appearance of features, they decrease the likelihood of correct convergence. To reduce the diminishing effects of shading and shadows, patch filtering is proposed for local intensity normalization, which makes the feature more distinguishable from its surrounding area, as demonstrated in Fig 1(c).

Our formulation of the patch filter begins with the Lambertian lighting model, which describes a gray-level image $I(x, y)$ as

$$I(x, y) = \rho(x, y)\boldsymbol{n}^T(x, y)\boldsymbol{s} \quad (5)$$

or more generally, the Lambertian model with shadows can be represented as

$$I = \min(\rho\boldsymbol{n}^T \sum_l \boldsymbol{s}_l, 0) = \min(\rho\boldsymbol{n}^T\mathbf{S}, 0) \quad (6)$$

where $\rho(x, y)$ is the reflectance (albedo) associated with point (x, y) in the image, $\boldsymbol{n}(x, y)$ denotes the surface normal of the object at (x, y), and $\boldsymbol{S}$ is the light source direction and intensity, which can be represented as a linear combination of multiple point light sources. This equation can be seen as a product of a reflectance component (ρ) and an illumination component ($\boldsymbol{n}^T\mathbf{S}$) as observed by Barrow and Tennenbaum [9].

From Retinex theory [10], the illumination image component can be approximated as the low frequency component of I, determined by convolution of the image with a low-pass Gaussian filter, which we denote as F_1. Dividing image intensities by this illumination component then yields an illumination-invariant descriptor:

$$R = \frac{I}{I * F_1}. \quad (7)$$

This descriptor normalizes a local patch with respect to illumination intensity, under the assumption that it is fairly
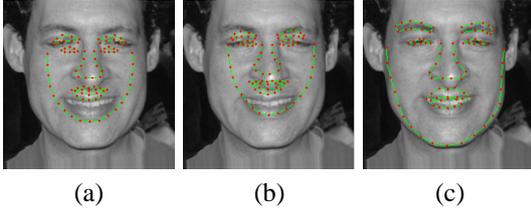
Figure 2: (a) Initialization #1: scale variation. (b) Results using gray-level feature. (c)Results using edge feature.
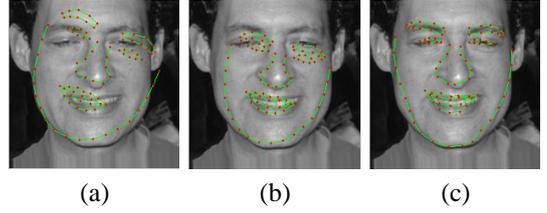

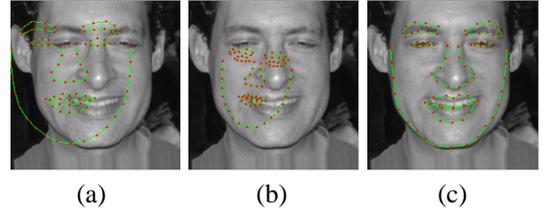Figure 3: (a) Initialization #2: rotation variation. (b) Results using gray-level feature. (c)Results using edge feature.

Figure 4: (a) Initialization #3: displacement variation. (b) Results using gray-level feature. (c) Results using edge feature.

even over the local patch. This division by the illumination component, however, can emphasize noise in the patch. To reduce this side effect, we filter out the high-frequency components in the numerator by convolving it with a low-pass filter F_2 with a larger passband than F_1:

$$R_1 = \frac{I * F_2}{I * F_1} \qquad (8)$$

where the division is pixel-wise.

Since the Retinex model assumes smooth variation of reflectance in a scene, sharp reflectance changes within the smoothing kernel of F_1 can distort the Retinex model of the illumination component. To exclude reflectance variations from the illumination component, Gross [9] employed an anisotropic filter instead of simple Gaussian smoothing. As a fast approximation to anisotropic filtering, our method uses a weighted Gaussian filter for F_2:

$$F_2 = \frac{1}{N} WG \qquad (9)$$

where G is the Gaussian kernel and N is a normalization factor such that

$$\frac{1}{N} \sum_\Omega WG = 1$$

where Ω is the kernel size. W is the weight function, modelled simply as a boolean function:

$$W(i,j) = \begin{cases} 1 & if \;\; I(i,j) \in M_1 \\ 0 & if \;\; I(i,j) \in M_2 \end{cases} \qquad (10)$$

where the convolution window is divided into two sub-regions M_1 and M_2 by a threshold $\tau = Mean(I_\Omega)$, and M_1 is the sub-region containing more pixels. In a smooth local region, the effect of this filter is similar to that a standard Gaussian filter. However, in an edge region, the filter kernel will convolve only with the primary local region M_1. The overall result of patch filtering is shown in Fig. 1(c). Although image noise is still amplified, the features become much more apparent than before patch filtering.

2.3 Feature Comparison

To distinguish the relative merits of the two relatively illumination-invariant features, their performance is mea-

sured with respect to initialization sensitivity and alignment accuracy.

To fit these features into the ASM search framework, we use as feature models the principal components of the phase congruency values or the filtered gray-level values in local windows centered on each feature point. The principal components analysis is computed from 200 images of size 200x200 under various non-extreme illumination conditions.

In an experiment performed on a different set of 200 images, three different poor initializations illustrated in Figs. 2-4 are used to test the sensitivity of these two features to initialization. Fig. 5 gives the statistical results of this experiment and Figs. 2-4 illustrate the difference in performance between the two methods. It is clear that the edge filtering method is more robust to poor initialization.
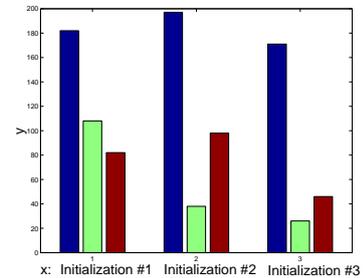

Figure 5: Comparison of different search features. The blue bar is for the edge-based method, the green bar is for the gray-level method, and the red bar is for the original ASM method. The y-axis represents the number of images out of 200 on which shape points converge to relatively correct positions, as opposed to images on which shape points converge to totally wrong positions.
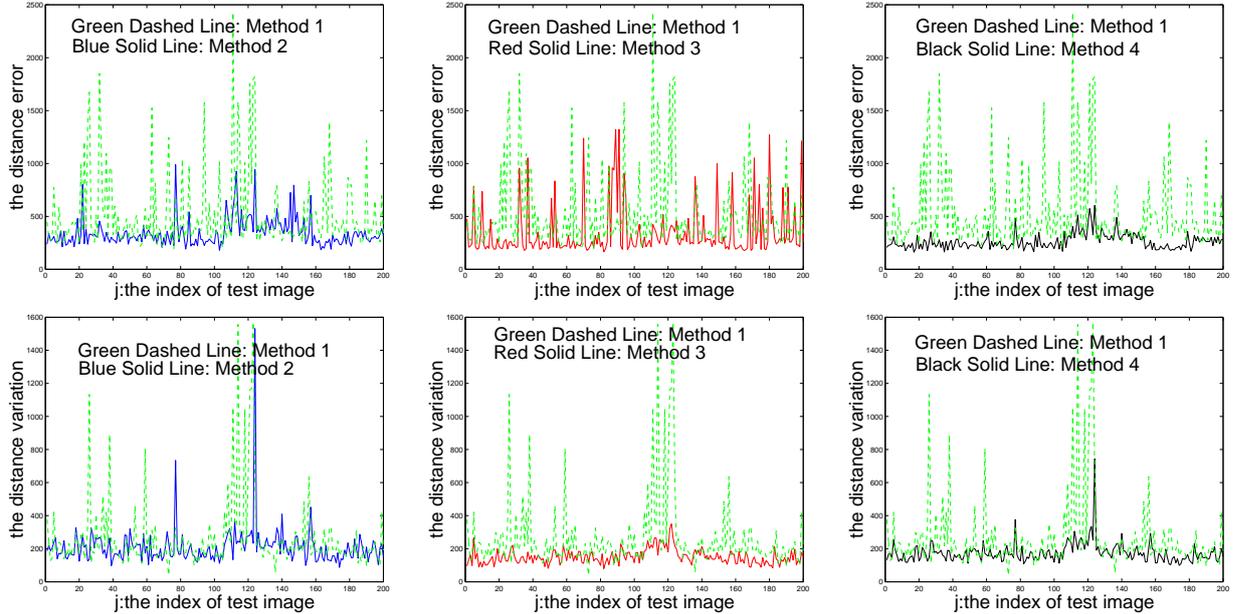
3

Figure 6: Statistical comparison with ASM. Top row compares accuracy, and bottom row compares stability. Method 1: ASM. Method 2: Edge-based alignment. Method 3: Gray-level method. Method 4: Bi-stage alignment.

From the experimental results presented in Sec. 4, it is also apparent that the patch filtering method provides higher accuracy for final alignment, since it improves upon the results given by edge filtering. The different merits of these two methods motivates us to employ them at different stages of the alignment process. The search process switches from edge filtering to patch filtering when edge filtering has converged, as determined for each point if after an iteration its change in position falls below a specified threshold. Termination of the patch filtering method is determined in the same manner.

3 Implementation

The multi-resolution implementation of our bi-stage alignment method is summarized in the following steps. As in many ASM implementations, the number of resolutions we use is $L = 4$ and the size of the search window is 5×1.

1. For each training and test image, a Gaussian image pyramid is built. The base image is denoted as level 0, and the roughest image is taken as level L. Similar to the original ASM method, a statistical shape model is built from the training images using PCA. For each level of the pyramid, the PCA models of the edge features and the patch features are each computed from the training images.

2. An initialization for each test image is determined. For images under non-extreme illumination, the initial shape can be given by a face detection algorithm. For images under extreme illumination, the initial shape is provided manually or could be provided by color-based detection methods.

3. In the search phase,

    (a) Set $l = L$.

    (b) If $l = L$, use the edge feature. Otherwise, use the patch feature.

    (c) Search the positions of all points until 90% of the points converge, and then project the shape into the PCA shape subspace.

    (d) If $l > 0$, then decrement l by one and return to (b).

4 Results

To test the performance of our alignment system, we do substantial experimentation on two groups of images, under general illuminations without significant facial shadows and under extreme illuminations, which consist of a single point light source at a large angle from the viewing direction. Because the bi-stage alignment algorithm requires some additional time to compute the edge and patch features in the image, it is slightly slower than the original ASM search scheme, but it nevertheless takes only about 0.5 to 0.8 seconds to align a face in a 200×200 image on a P-4 1.4G computer with 256M memory.
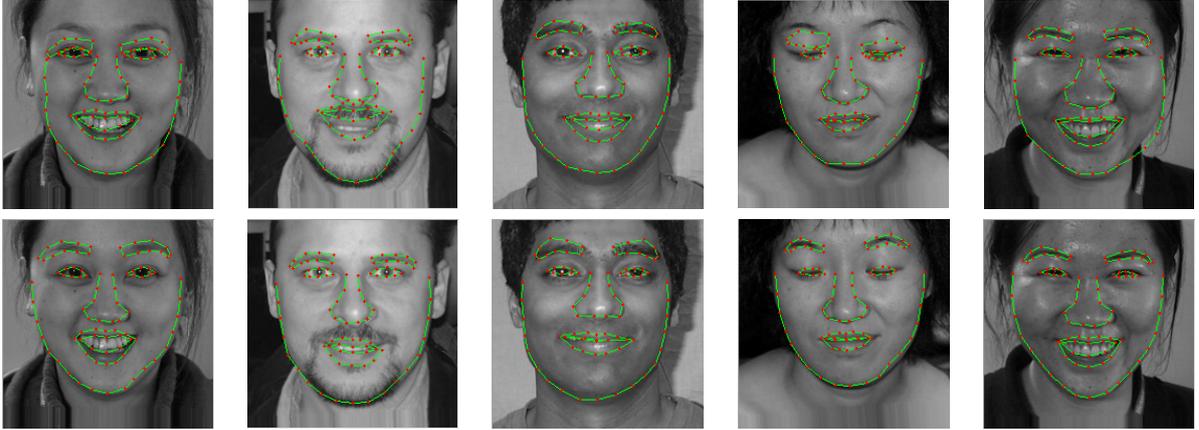
4

Figure 7: Comparison under good illumination. Top row: ASM. Bottom row: Bi-stage method.

Figure 8: Comparison under illumination that causes significant shading variation. Top row: ASM. Bottom row: Bi-stage method

4.1 Results under General Illumination

We manually labelled 400 images under general illumination, each of size 200×200. Of these images, 200 were used for training and the remaining 200 for testing. Even though the faces are fairly well illuminated, some of these images present problems to ASM. In this section, we compare the accuracy and stability of our algorithm to the original ASM method on the 200 test images.

To measure accuracy, the distance between the searched feature positions (x_{k1}, y_{k1}) and the manually annotated feature points (x_{k2}, y_{k2}) is taken as the estimated alignment error:

$$D = \sum_k \sqrt{(x_{k1} - x_{k2})^2 + (y_{k1} - y_{k2})^2} \qquad (11)$$

For each of the 200 test images, indexed by j, the top row of Fig. 6 plots the error of ASM (green) and the error of our algorithm (red).

As a rough measure of stability, we input the manually aligned shape to the alignment algorithms as the initialization, and then observe the variation between this initialization and the resulting shapes after search. The bottom row of Fig. 6 exemplifies the greater stability of our method in comparison to the original ASM method.

Although the edge-based method is more robust than the gray-level method to poor initialization as illustrated in Fig. 5, it has lower accuracy and stability than bi-stage alignment. The gray-level method has lower accuracy than the bi-stage method but it gives good stability when given a good initialization.

Additional comparisons with ASM under good illumination are given in Fig. 7 for examples with exaggerated expressions, facial hair, or unusually-shaped features. Fig. 8 displays results under illumination that causes much shading variation on the faces.
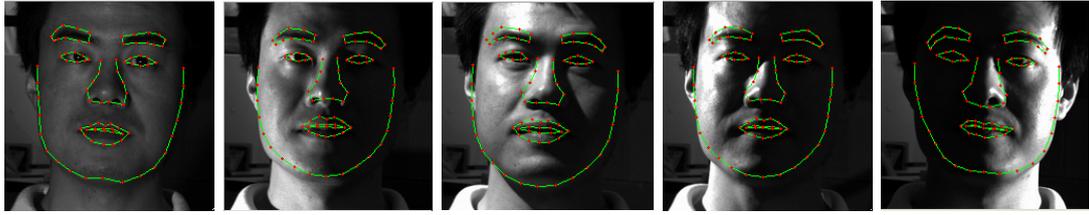
5

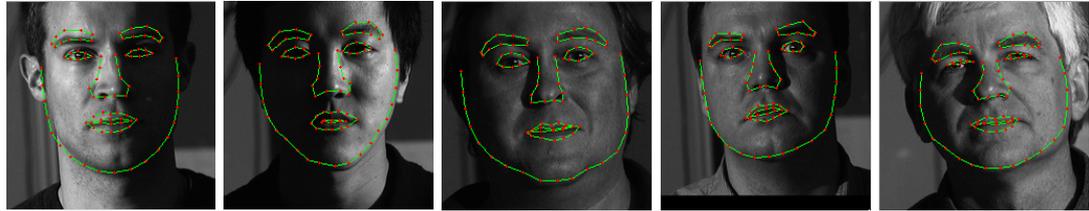Figure 9: Alignment results for one person under varying illumination.

Figure 10: Alignment results for different people under extreme illumination.

4.2 Results under Extreme Illumination

We selected images from the CMU PIE database [10] and YALE FACE DATABASE B [13] to test our method on extreme illuminations. Since ASM collapses entirely on extreme images, a statistical comparison between ASM and our algorithm is not meaningful. Experiments on these images show that our system can give reasonable results under various shadings and shadows, as exemplified in Fig. 9. Even for images that are heavily shadowed and require great care to align manually, our algorithm can often work effectively, as shown in Fig. 10. For some images with significant shadowing, although our algorithm may not accurately locate some of the feature points, it rarely collapses to a totally wrong result.

5 Conclusion

This paper addresses the problem of face alignment under variable illumination using two relatively illumination-invariant features at different levels of an alignment algorithm. This approach allows our system to handle not only illumination variations, but also poor initializations. Experiments have demonstrated the robustness and accuracy of this method, even for a fair number of images under extreme illumination conditions.

References

[1] Y. Moses, Y. Adini, and S. Ullman. Face Recognition: The Problem of Compensating for Changes in Illumination Direction, Proc. Euro. Conf. Comp. Vision, pp. 286-296, 1994.

[2] P. Belhumeur and D. Kriegman. What Is the Set of Images of an Object under All Possible Illumination Conditions? Int. J. Comp. Vision, vol. 28, no. 3, pp. 245-260, 1998.

[3] A. Shashua and T. Riklin-Raviv. The Quotient Image: Class-Based Rerendering and Recognition with Varying Illuminations, IEEE Trans. on PAMI, 23(2): 129-139, 2001.

[4] T.F. Cootes, C. J. Taylor, D. H. Cooper and J. Graham. Active shape models - their training and application, Comp. Vision and Image Understanding, 61(1):389, January 1995.

[5] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance model. In Proc. 5th Euro. Conf. on Comp. Vision, Freiburg, Germany, 1998.

[6] T.F.Cootes and P. Kittipanya-ngam. Comparing Variations on the Active Appearance Model Algorithm, Proc. BMVC2002, Vol.2, pp.837-846.

[7] Peter Kovesi. Image Features From Phase Congruency. Videre. Volume 1, Number 3, Summer 1999

[8] D.J. Field. Relations between the Statistics of Natural Images and the Response Properties of Cortical Cells. JOSA A, 4(12): 2379-2394, December 1987.

[9] H.G. Barrow and J. Tenenbaum. Recovering intrinsic scene characteristics from images. Computer Vision Systems, pp. 3-26. Academic Press, 1978.

[10] Daniel J. Jobson, Zia-ur Rahman, and Glenn A. Woodell, " Properties and Performance of a Center/Surround Retinex", IEEE Trans. on Image Proc., Vol. 6, No. 3, 1997, pp. 451-462

[11] Ralph Gross, Vladimir Brajovie, "An Image Preprocessing Algorithm for Illumination Invariant Face Recognitoin", Int. Conf. on Audio and Video Based Biometric Person Auth., pp. 10-18, 2003

[12] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Data-base", Proc. IEEE Int. Conf. on Auto. Face and Gesture Recog., May, 2002

[13] Athinodoros S. Georghiades and Peter N. Belhumeur, "From Few to many: Illumination cone models for face recognition under variable lighting and pose", IEEE Trans. PAMI, Vol. 23, No. 6, pp 643-660, 2001