FACE RECOGNITION WITH HARMONIC DE-LIGHTING

Laiyun Qing[1,2], *Shiguang Shan*[2], *Wen Gao*[1,2]

[1]Graduate School, CAS, Beijing, China, 100080
[2]ICT-ISVISION Joint R&D Laboratory for Face Recognition, CAS, Beijing, China, 100080
Emails: {lyqing, sgshan, wgao}jdl.ac.cn

ABSTRACT

Evaluations of the state-of-the-art of both academic face recognition algorithms and commercial systems have shown that performance of most current technologies degrades due to the variations of illumination. We propose a novel technique for face recognition under generic illumination in this paper, namely, calibrating an input face image to an image under canonical illumination, named as face *de-lighting,* to reduce the negative effect of non-ideal illumination in the original image. The canonical illumination is defined as an illumination environment in which light is constant in every direction. Inspired by the low dimension effect of light on Lambertian surface and the compact representation of the canonical illumination in spherical frequency space (only the DC component needed), face de-lighting is achieved with spherical harmonics. Experiments show the effectiveness of the proposed method in both lighting estimation and face recognition.

1. INTRODUCTION

Much progress in face recognition has been made in the past few years. However, face recognition remains a difficult, unsolved problem in general. The performance of almost all current face recognition systems is heavily subject to the variations in the imaging conditions [1]. Illumination variation is one of the obstacles.

There have been many works dealing with illumination variation in face recognition. Low-dimensional lighting of face is the main cue. Eigenfaces [2] and Fisherfaces [3] [4] apply statistical learning to get the empirical low dimensional space of the face. These methods have demonstrated their easy implementation and accuracy. However, they perform well enough under the similar imaging conditions to those of the training images only.

Recently, an approach called generative method has been proposed in which a set of images of an object under varying illumination conditions is generated from a small number of training images of the object, assuming the Lambertian model. These methods aim at recovering the intrinsic feature of the face: shape and/or albedo. Based on whether the "face class" information is used or not, these approaches are categorized into two kinds: generic method and classed based method.

The former methods use no information about "face class". Therefore they can be applied in other objects than face in theory. Intrinsic images [5] and Illumination Cone [6] are instances of the generic methods except that Illumination Cone uses prior knowledge about the common shape of faces to resolve the three parameters of the GBR ambiguity. The disadvantage is that the generic methods usually need several images to work, which is not satisfied by many face recognition systems.

The class based methods use the prior knowledge of the common face model [7, 8, 9, 10] and usually need only a single face image. The prior knowledge is obtained with a boosting step or some other constrains. Sim's Statistical SFS [7] learn the statistics of *n(x)* and *e(x,s)* for each pixel of face images from the bootstrap set. Quotient image [8] assumes all faces have same shape and the shape information is learned from a training set. Zhao's SSFS exploits the symmetry of face explicitly and assume all faces share similar common shape [9]. Based on constant albedo assumption, Atick [10] uses PCA to solve the parameters of the Eigen-head approximation of a real 3D head. A 3D morphable model imposes class constrains is reported in [11].

Many of these methods assume simple light model. However, the natural illumination in the real world is highly complex, consisting of reflected light from every direction as well as distributed and localized primary light sources [12]. Basri [13] and Ramamoorthi [14] have got the analytic nine dimensional lighting space of a convex Lambertian surface expressed in terms of spherical harmonic. With the discovery that the effect of illumination on diffuse object is low dimension with analytic analysis, it won't be more difficult to deal with generic illumination than to deal with simple light source model. Recent face recognition algorithms such as Linear Subspace [13] and Harmonic Examplars [15] can recognize face images under arbitrary lighting conditions.

In this paper, we propose an algorithm for recognizing face under arbitrary lighting by face de-lighting based on the model proposed by Basri [13] and Ramamoorthi [14]. The problem of face de-lighting is stated as follows: given a face image under poor unknown illumination, calibrate it to a new image under the pre-defined canonical illumination. Then face recognition is achieved by matching the canonical form of the gallery images and the probes. Canonical illumination is defined as an illumination environment in which light is constant in every direction and it can be represented compactly with the DC component only in frequency domain. Differing form Wen's REM Relighting [16], we apply the method in face recognition rather than in computer graphics. In our work the feature points are automatically localized instead of manually labeled and the canonical illumination is virtual illumination that has only DC component.

The remainder of this paper is organized as follows. We introduce the proposed face de-lighting method with spherical harmonics in section 2; Section 3 shows the experimental results about the de-lighting and face recognition; Section 4 concludes our paper and discusses the future work.

2. FACE DE-LIGHTING WITH SPHERICAL HARMONICS

The reflection equation is a rotational convolution and it's natural to analysis it in space-frequency domain. It has shown in [13] and [14] that the BRDF of Lambertian surface is a low-pass filter. The lower nine spherical harmonics is approximate enough for its irradiance environment map, that is,

$$\begin{aligned} E(\theta,\phi) &= \sum_{l=0}^{\infty}\sum_{m=-l}^{l} E_{lm}Y_{lm}(\theta,\phi), \\ &\approx \sum_{l=0}^{2}\sum_{m=-l}^{l} E_{lm}Y_{lm}(\theta,\phi), \\ &= \sum_{l=0}^{2}\sum_{m=-l}^{l} \Lambda_l A_l L_{lm}Y_{lm}(\theta,\phi). \end{aligned} \tag{1}$$

where Λ_l are the normalization factors and A_l are spherical harmonics coefficients of the diffuse reflection function. The analytic form of A_l and more details of deriving the space-frequency reflection equation are given in [13, 14].

Since the irradiance E is dominated by low frequency components of lighting, we need estimate only the lower nine spherical harmonic coefficients of lighting $L_{lm}(0 \le l \le 2)$.

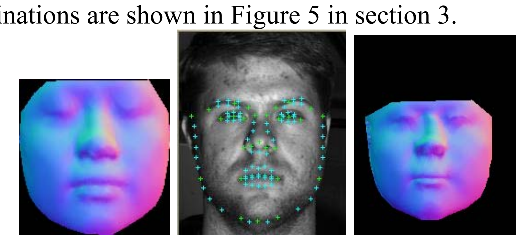
We define the canonical illumination as an illumination environment in which light is constant in every direction. It can be represented compactly using the DC spherical harmonics component only. After the original illumination is estimated, face de-lighting become simple in space-frequency domain with illumination ratio technology. The details will be introduced in section 2.2. In order to analyze with the spherical harmonics, the information of the surface normal or the geometric shape of the face is needed. This information is obtained by aligning the 2D image with a 3D generic mesh. The alignment is given in section 2.1.

2.1. Face alignment

Human faces can be assumed rationally have similar 3D shapes. This has always been used in many algorithms [8, 9, 16]. A generic 3D model of face is used in our implementation. Given a 2D image, to create the correspondence between the vertices of the mesh and the 2D image, we first create correspondence between the feature points on the mesh and the 2D image. Then the rest of the vertices on the mesh and the 2D image are aligned with image warping technique.

The feature points on the 2D image are marked automatically in the process of face detection. Our face detection method, named Face Center-of-Gravity Template [17], is based on some observations on the configure relationship between major face organs. The eyes are localized by growing a region window from the approximate center of the detected face and checking its characteristics. After eyes are located, we attempt to combine the ASM's (Active Shape Model) local texture models and AAM's (Active Appearance Model) global appearance models for sparse facial feature correspondence [18]. To integrate the local profile and global appearance constraints, the subspace reconstruction residual of the global texture is exploited to evaluate the fitting degree of the current model to the novel image.

The normal map of the 3D Mesh and the feature points on a face image is shown in Figure 1. Some results of our feature points locating under different illuminations are shown in Figure 5 in section 3.



(a) (b) (c)

Figure 1: The generic mesh and feature points: (a) the generic mesh; (b) the feature points of a face; (c) the warped mesh for the face in b).

2.2. Face de-lighting

In order to get a new image under canonical illumination, we first need to estimate the lighting condition of the original image.

Given an input image $\mathbf{I}$ (a column vector of n elements, n is the number of pixels in the image) of an object with constant albedo ρ, let $\mathbf{E_{lm}} = \Lambda_l A_l \mathbf{Y}_{lm}$ (a column vector of n elements, $0 \le l \le 2, -l \le m \le l$) denote the harmonic irradiance image and $\mathbf{E}$ is a $n \times 9$ matrix of $\mathbf{E_{lm}}$, then by solving the least squares problem

$$\min \| \mathbf{E}(\rho \mathbf{L}) - \mathbf{I} \|, \qquad (2)$$

we get the illumination coefficients vector scaled with the albedo $\rho\mathbf{L}$, which approximate the illumination.

In fact, this approximation will still be exact when the albedo contains no low frequency components (order $1 \le l \le 4$), except for the zero order, DC component[1]. Though the albedo of face does not strictly satisfy this constrain, we find that we obtain good results in practice. Wen [16] has justified numerically that the low-frequency components of texture map of faces are small except the DC component. This can also be verified by vision that most regions of the face are skin with almost the same albedo.

Once we have estimated the lighting condition of the original image, de-lighting it to the canonical illumination is just forward by rendering the face with the virtual illumination of L_{00}.

For any given point P at position (x, y) on the face, suppose its normal is (θ, ϕ), and $\rho(x, y)$ is its reflectance coefficient, then the intensities at P in the original image and the canonical image are respectively:

$$I_{org}(x, y) = \rho(x, y) \sum_{l=0}^{2} \sum_{m=-l}^{l} \Lambda_l A_l L_{lm} Y_{lm}(\theta, \phi), \qquad (3)$$

$$I_{can}(x, y) = \rho(x, y) \Lambda_0 A_0 L_{00}. \qquad (4)$$

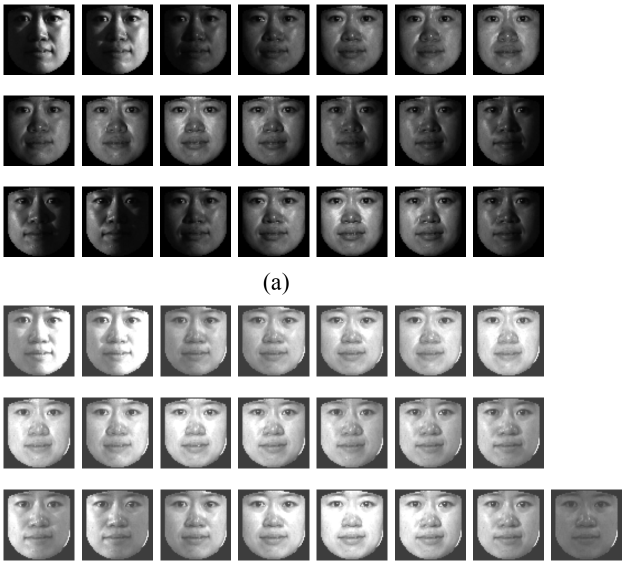
Given the original image, the canonical image is:

$$I_{can}(x, y) = \frac{\Lambda_0 A_0 (\rho L_{00})}{\sum_{l=0}^{2} \sum_{m=-l}^{l} \Lambda_l A_l (\rho L_{lm}) Y_{lm}(\theta, \phi)} \times I_{org}(x, y). \qquad (5)$$

Ramamoorthi [19] shows that only 5 or 6 coefficients can be stably recovered given only one image. As we need only the new image under canonical illumination, this does not affect us.

3. EXPERIMENTAL RESULTS

In the real world, illumination usually consists of an ambient light with perhaps one or more point sources. To obtain representative images of such cases, CMU-PIE database [20] includes face images both with the ambient lights on and off. To verify the results of lighting estimation and the effect of de-lighting on recognition under different illuminations, we select 22 frontal images captured with the ambient lights on and 21 images captured with the ambient lights off for each person in CMU-PIE database. There are 68 persons in CMU-PIE database. An example of the images of one person is illustrated in Figure 2 (we have masked the images for face recognition). For more details of the CMU-PIE database, please refer to [20].



(a)

(b)

Figure 2: The images of one person under different illuminations in CMU-PIE database: (a) images with ambient lighting off, flash no. ranging from 2-22, top to down, left to right; (b) images with ambient on, flash no. ranging from 2-22, top to down, left to right, the last image is captured with no flash (flash no is 1).

3.1. The experimental results of lighting estimation

The results of lighting estimation of the 43 illuminations are given in Figure 3. In order to eliminate the effect of different average abledos of different persons, the last eight illumination coefficients are divided by the DC coefficient.

The standard deviations of the coefficients are very small, which indicates that our light estimation is very stable. The physical meanings of the nine illumination coefficients are very explicit. The DC coefficient represents the average lighting energy. The lighting environments with ambient illumination on have larger energy than those with ambient illumination off. Therefore the mean coefficient in Figure 3(c) is smaller than the corresponding coefficient in Figure 3(a). L_{11} denotes the direction of the primary light in x-axis. For

[1] The spherical harmonic coefficients of a product of irradiance and albedo, as in the basis functions Y_{lm}, are determined by a Clebsch-Gordan expansion of the product of spherical harmonics. To ensure that orders 0,1 and 2 of the image correspond to irradiance coefficients scaled by the DC term of the albedo, assuming the only relevant irradiance coefficients are orders 0,1 and 2, we require orders 1-4 of the albedo vanish.

example, L_{11} of Flash No.2 and Flash No.17 have the similar absent value but different sign, which indicates that they are opposite in x-axis direction. L_{10} and L_{1-1} denotes the direction of the primary light in z-axis and y-axis respectively.

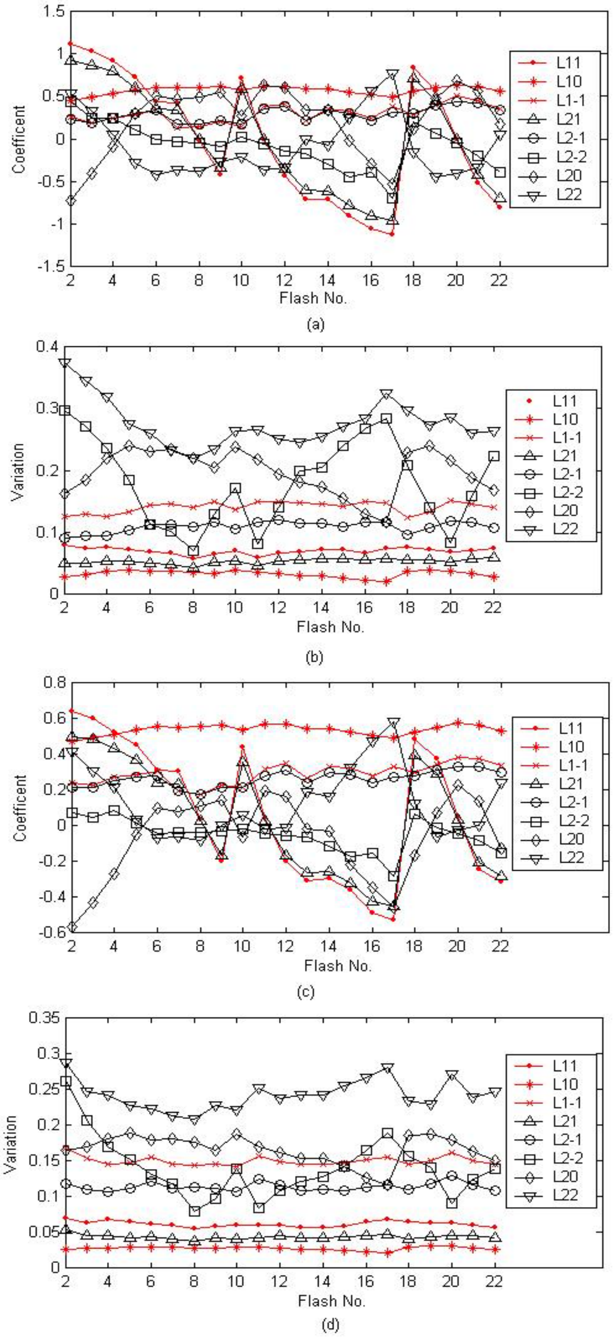


Figure 3: The results of lighting estimation of PIE Face Database. (a) and (b) are the mean and the standard deviation of the coefficients of lighting with ambient light off. (c) and (d) are those of lighting with ambient light on.

We have also shown the additive of light. The difference between the estimated light with ambient light on and the sum of corresponding light and the ambient light are shown in Figure 4. The differences are very small (the average DC coefficient is 89.33213).

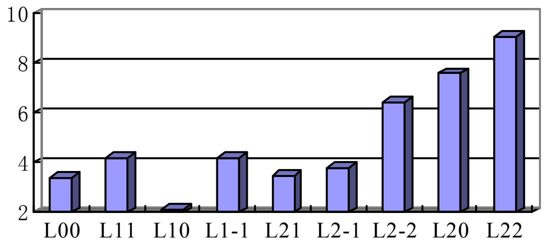


Figure 4: The difference between the flash with ambient light on and the sum flash and ambient light.

3.2. The experimental results of face de-lighting

The results of de-lighting on both synthesized face images and real face images in CMU-PIE database are given in this subsection.

The synthesized faces are rendered using Debevec's Facial Reflectance Field Demo [21], which can be downloaded freely at http://www.debevec.org/FaceDemo/. Figure 5 has shown some of the results. The top row are the four original images rendered under an outdoor illumination map and the other four images under an indoor environment map, each by rotating the environment map 90 degrees. The images in the bottom row are the respective de-lit images. Though the variations in lighting are large between the original images, the de-lit images are almost the same.

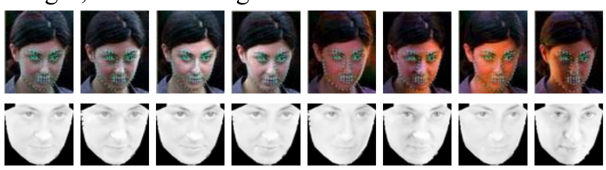
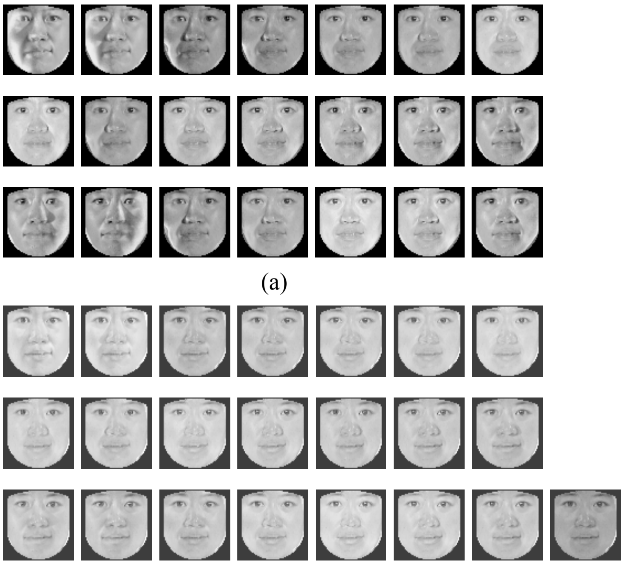


Figure 5: Face de-lighting for synthesized face images

We also test our method on real face images in CMU-PIE database. The de-lit images of images in Figure 2 are given in Figure 6. The results of some images in Figure 6(a) are not as good as those in Figure 5 and Figure 6(b), because the lighting condition is point light source rather than natural illumination. Fortunately, such cases are only existed with strictly controlled situations and they are rare in the real world. We will see that even in these extreme cases, the de-lit images achieve better results in both human vision and machine recognition.



(a) (b)

Figure 6: The de-lit images of images in CMU-PIE database.

3.3. The experimental results of face recognition

We verify the effect of de-lighting on face recognition in this subsection. The simplest normalized correlation is exploited as the similarity between two images. Face recognition is achieved by finding a nearest neighbor based on the image similarity.

The images are divided into 4 subsets according to the greater of the longitudinal and latitudinal angles of the flash direction from the camera axis—Subset 1(f06~f09, f11, f12, f20), Subset 2(f05, f10, f13, f14, f19, f21), Subset 3(f04, f15, f18, f22), and Subset 4(f02, f03, f16, f17).

The experimental results are illustrated in Table 1. The images in set *a* are images with ambient lights off as in Figure 2(a) and Figure 6(a) and the images in set *b* are images with ambient lights on as in Figure 2(b) and Figure 6(b). The numeral in the blank is the number of the flash.

We can see that the recognition rates are higher after de-lighting in all the cases. The best case is the gallery is the b(11) after de-lighting, in which is recognition is perfect for all the four subsets. The experimental results are promising when the lighting models of the gallery and probe are the same (all are with ambient lights on or all are with ambient lights off). But for the case of recognition of images in set *a* with gallery images in set *b*, the results are not so good. There are probably several reasons. One is the automatically feature points extraction. The feature points labeling is not all correct therefore the information of the shape of the faces are not correct. The labeling results are good enough for lighting estimation but it's bad for delighting in the case of extreme lighting conditions. The other reason is the limited dynamic range in the digitized images. There are some artifacts in the de-lit images if the pixel intensity in the original image is too low or saturation. Wen [16] used constrained texture synthesis [22] to alleviate the low dynamic range, which can also be applied for our purpose. Because our purpose is face recognition, maybe some simpler methods can also work. There are some variations about dressing such as glasses between image set *a* and set *b*. Considering all these reasons, if we applied the more robust face recognition techniques such as LDA, the performance of the system is expected to be better.

Table 1: Recognition rate comparisons of before delighting and after delighting on CMU-PIE database

gallery	probe	de-light	Performance of Subset No. (%)			
			1	2	3	4
a(11)	a	No	96	58	24	9
		Yes	100	100	97	73
	b	No	63	37	24	11
		Yes	88	70	55	40
b(11)	a	No	71	35	20	7
		Yes	79	75	67	45
	b	No	100	97	85	45
		Yes	100	100	100	100
b(1)	a	No	24	26	26	16
		Yes	65	56	55	43
	b	No	74	83	94	80
		Yes	96	100	100	100

4. CONCLUSIONS

Though simple lighting models are assumed in many vision systems, the natural illumination in the real world is very complex. This complexity brings us some advantages rather than troubles in dealing with lighting variation. The quality of the images under natural illumination is generally better than that of images under strictly controlled point light source. With the discovery that the effect of illumination on diffuse object is low dimension with analytic analysis, it won't be more difficult to deal with natural illumination than to deal with simple light source models.

Based on these observations, we propose a technique for face de-lighting under generic illumination, namely, calibrating the input face image to an image under canonical illumination, to reduce the negative effect of poor illumination in the original image. Inspired by the low dimension effect of illumination map of Lambertian surface and the compact representation of the canonical illumination in spherical frequency space analysis, face de-lighting is achieved with spherical harmonics by only remaining the DC component. Experimental results show that the proposed method is effective on both lighting estimation and face recognition. Experiments on larger face databases are ongoing.

The results of face recognition are still not practical enough in the case of extreme illumination. Some more robust face de-lighting based on the lighting estimation is our next task.

5. ACKNOWLEDGEMENTS

This research is partially sponsored by Natural Science Foundation of China under contract No.60332010, National Hi-Tech Program of China (No. 2001AA114190 and 2002AA118010), and ISVISION Technologies Co., Ltd.

REFERENCES

[1] P.J. Phillips and H. Moon, et al, “The FERET Evaluation Methodology for Face-Recognition Algorithms”, *IEEE Trans. on PAMI*, Vol.22, No.10, pp1090-1104, 2000.

[2] M. Turk and A. Pentland, “Eigenfaces for Recognition” *Journal of cognitive neuroscience,* vol. 3, no. 1, pp71-86, 1991.1

[3] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman, “Eigenfaces vs Fisherfaces: recognition using class specific linear projection”. *IEEE Trans. on PAMI*, vol.20, No.7, 1997.7.

[4] R. Gross, I. Matthews and S. Baker, “Eigen Light-Fields and Face Recognition Across Pose”, *Proc. FG’02,* 2002.

[5] Y. Weiss, “Deriving intrinsic images from image sequences”. *Proc. ICCV’01,* 2001.

[6] A.S. Georghiades, P.N. Belhumeur and D.J. Kriegman, “From Few to Many: Illumination Cone Models for Face Recognition under Differing Pose And Lighting”, *IEEE Trans. on PAMI,* Vol.23, No.6, pp643-660, June 2001

[7] T. Sim, T. Kanade, “Combining Models and Exemplars for Face Recognition: An Illuminating Example,” *Proc. Workshop on Models versus Exemplars in Computer Vision, CVPR’01*, 2001.

[8] A. Shashua and T. Riklin-Raviv, “The Quotient Image: Class-Based Re-Rendering And Recognition With Varying Illuminations”, *IEEE Trans. on PAMI,* vol.23, no.2, pp.129-139, 2001.2

[9] W. Zhao and R. Chellappa, “Illumination-insensitive Face Recognition using Symmetric Shape-from-Shading”. *Proc. CVPR’*00, 2000.

[10] J. Atick, P. Griffin, and A. Redlich, “Statistical Approach to Shape from Shading: Reconstruction of 3D Face Surfaces from Single 2D Images”, *Journal of Neural Computation*, vol. 8, no. 6, pp. 1321—1340, 1996.

[11] V. Blanz and T. Vetter, “Face Recognition Based on Fitting a 3D Morphable Model”, *IEEE Trans. on PAMI*, vol. 25, no. 9, pp.1-12, 2003.

[12] R. O. Dror, T. K. Leung, E. H. Adelson, and A. S. Willsky, "Statistics of real-world illumination". *Proc. CVPR’01*, December 2001.

[13] R. Basri, D. Jacobs, “Lambertian Reflectance and Linear Subspaces,” *Proc. ICCV’01*, 2001, pp. 383-390.

[14] R. Ramamoorthi and P. Hanrahan, “An efficient representation for irradiance environment maps”, *Proc. SIGGRAPH’01*, August 2001, pp.497-500.

[15] L. Zhang, D. Samaras, "Face Recognition Under Variable Lighting using Harmonic Image Exemplars", *Proc. CVPR’03*, June 2003, vol. I, pp.19-25.

[16] Z. Wen, Z. Liu and T. S. Huang, "Face Relighting with Radiance Environment Maps", *Proc.CVPR’03*, 18-20 June 2003, vol II, pp.158-165.

[17] S.G. Shan, W. Gao, W. Wang, et al., “Enhanced Active Shape Models with Global Texture Constraints for Face Image Analysis”, *ISMIS’03*, Oct. 2003. pp. 593-597.

[18] W. Wang, S. Shan, W. Gao, et. al., “An Improved Active Shape Model For Face Alignment”, *Proc. ICMI’02*, pp523-528, 2002.

[19] R. Ramamoorthi, Analytic PCA Construction for Theoretical Analysis of Lighting Variability in Images of a Lambertian Object. *IEEE Trans. on PAMI*, vol. 24, no. 10, pp. 1-12, 2002.

[20] T. Sim, S. Baker, and M. Bsat, “The CMU Pose, Illumination, and Expression (PIE) Database”, *Proc. FG’02*, May 2002, pp. 53-58.

[21] P.E. Debevec et. al. “Acquiring the reflectance field of a human face”, *Proc. SIGGRAPH’00,* July 2000, pp. 145–156.

[22] A. Hertzmann et. al. “Image analogies”, *Proc. SIGGRAPH’01*, August 2001, pp. 327–340.