FACE RECOGNITION UNDER GENERIC ILLUMINATION BASED ON HARMONIC RELIGHTING

LAIYUN QING

Graduate School of Chinese Academy Sciences, No.19, Yuquan Road,Beijing, 100039, China

lyqing@jdl.ac.cn

SHIGUANG SHAN, WEN GAO, BO DU

Institute of Computing Technology, CAS, No.6 Kexueyuan South Road, Beijing, 100080, China

{sgshan, wgao, bdu}@jdl.ac.cn

The performances of the current face recognition systems suffer heavily from the variations in lighting. To deal with this problem, this paper presents an illumination normalization approach by relighting face images to a canonical illumination based on the harmonic images model. Benefiting from the observations that human faces share similar shape, and the albedos of the face surfaces are quasi-constant, we first estimate the nine low-frequency components of the illumination from the input facial image. The facial image is then normalized to the canonical illumination by re-rendering it using the illumination ratio image technique. For the purpose of face recognition, two kinds of canonical illuminations, the uniform illumination and a frontal flash with the ambient lights, are considered, among which the former encodes merely the texture information, while the latter encodes both the texture and shading information. Our experiments on the CMU-PIE face database and the Yale B face database have shown that the proposed relighting normalization can significantly improve the performance of a face recognition system when the probes are collected under varying lighting conditions.

Keywords: face recognition; varying lighting; harmonic images; lighting estimation; illumination normalization.

1. Introduction

Face recognition has various potential applications in public security, law enforcement and commerce such as mug-shot database matching, identity authentication for credit card or driver license, access control, information security, and video surveillance. In addition, there are many emerging fields that can benefit from face recognition, such as human-computer interfaces and e-services, including e-home, online-shopping and online-banking. Related research activities have significantly increased over the past few years [5, 26].

However, face recognition remains a difficult, unsolved problem in general due to several bottlenecks, among which illumination change is one of the most challenging problems. It has been argued that the variations between the images of the same face due to illumination and viewing direction are almost always larger than the image variations due to the change in face identity [11]. These observations have been further verified by the evaluation of the state-of-the-art systems as well. The FERET tests revealed that the recognition performances of the best systems degraded significantly when the illuminations changed [12]. The recent FRVT 2002 had also shown that, there was a large decrease in performance on the outdoor probe categories even for the best commercial face recognition systems, though these systems were not sensitive to normal indoor lighting changes [13]. One of the characters of the outdoor images is the varying illuminations caused by the changes of the directional sunlight, which changes the appearance of the face significantly. This indicates that recognition of face images collected in outdoors needs to become a focus of research.

On lighting modeling, one recent important development is the harmonic images method, proposed by Basri [2] and Ramamoorthi [14] independently. By using spherical harmonics and signal-processing techniques, they analytically proved that the set of images of a convex Lambertian object obtained under arbitrary lighting conditions (including outdoor environments) could be approximated well by a 9-D subspace spanned by its nine harmonic images. Along the harmonic images strategy, this paper presents a model-based approach to identify faces robustly under generic illumination with the face relighting technology.

The rest of the paper is organized as follows. We review the previous work dealing with illumination variation in face recognition briefly in section 2. The details of the proposed face relighting method based on the harmonic images model are presented in section 3. Related issues on how to apply the relighting technology to face recognition are discussed in section 4. Finally, we show the experimental results of the proposed method in section 5, followed by some concluding remarks in the last section.

2. Previous Work

From the imaging principle, the appearance of a face depends on not only its intrinsic shape and texture, but also the extrinsic imaging conditions, such as illumination and the view direction. According to how to deal with the extrinsic imaging parameters, the methods dealing illumination problem in face recognition can be categorized into two fundamental categories: the model-based approaches and the statistics-based approaches [4].

The statistics-based approaches analyze the images directly using statistical methods and do not formally distinguish between the intrinsic and extrinsic parameters. This kind of approaches learn the distribution of the images collected under varying lighting conditions by statistical analysis of a large training set. Eigenface [20], Fisherface [3], and Bayesian method [10] are the typical methods belonging to this category. In this context, one of the most important observations is that the images of a convex Lambertian surface caused by varying illuminations lie in a low-dimensional subspace embedded in the image subspace. Especially, Hallinan reported that 5-D subspace would suffice to represent most of the image variations due to illumination changes including extreme cases [7].

The model-based approaches treat the extrinsic parameters as separate variables and model their functional role explicitly. These methods commonly build an explicit generative model of the variations of the face images, to recover the intrinsic features of the face: shape and/or albedo.

In the simplest case of a convex Lambertian object under distant illumination without attached and cast shadows, the appearances of the object can be completely described with a 3-D linear subspace defined with three input images taken with linearly independent lighting conditions [15]. However, for face recognition, usually only one single image is available. Then the Quotient Image method [16] was proposed, which assumed that the human faces were an ideal class of objects with same shape and rational span surface albedo. Because the assumption of this model would be too restrictive to be used for object recognition in more realistic settings, the rendered images will be unrealistic if shadows are ignored.

Illumination Cones [6] used seven distinct images of the same face to estimate its 3D shape and the albedo map by using a variant of the photometric stereo, considering both the attached and cast shadows. After the 3D reconstruction, the illumination cone of the object was constructed. The illumination cone algorithm was claimed to achieve the highest recognition rate under different illumination conditions. However, seven images for each face and the relatively complicated process may have prevented its wide application.

SFS (Shape From Shading) has also been investigated in face recognition under varying illuminations. However, SFS from one image with unknown illumination is an ill-posed problem. Therefore, researchers had exploited the classed-based methods, which made use of the prior knowledge of the “human face class” [1, 4, 18, 25], explicitly or implicitly. Statistical SFS [18] method learned the statistics of the basis images $b(x)$ and the residual image $e(x,s)$ for each pixel of the facial images from the bootstrap set. Symmetric SFS [25] exploited the symmetry of faces explicitly and assumed that all faces shared similar shape. More recently, the 3D Morphable model was presented in [4], which imposed the face specific constrains by statistically modeling the 3D face shape and texture through PCA model. Prior to the 3D morphable model, Atick had used PCA to estimate the parameters of the Eigen-head approximation of a real 3D head, based on the constant albedo assumption [1].

Recently, the 9-D linear subspace [2, 14] method was proposed. By using spherical harmonics and signal-processing techniques, Basri [2] and Ramamoorthi [14] showed that the set of images of a convex Lambertian object obtained under varying lighting conditions could be approximated by a 9-D subspace spanned by nine basis images of the object, called harmonic images, each of which corresponds to an image of the object illuminated under harmonic lights whose distributions are specified in terms of spherical harmonics. This discovery has greatly facilitated the modeling of generic illumination and provided the possibility to solve face recognition problem under varying lighting conditions, especially the outdoor environments. The 9-D subspace defined with harmonic images [2] and Harmonic Exemplars [24] provided the some solutions to recognize facial images under generic lighting conditions.

Based on the harmonic images model [2, 14], this paper presents a model-based approach to identify faces robustly under generic illumination with the face relighting technology. Given a facial image, we first estimate the nine low-order spherical harmonic coefficients of the illumination under which the image is captured. Then the face is re-rendered to a predefined canonical illumination to obtain the canonical form of the input face image. Finally, face recognition is achieved by matching the canonical form of the probe with those of the gallery images. Differing from REM Relighting [23], we apply the relighting method to face recognition rather than to computer graphics.

3. Face Relighting with Spherical Harmonic Images

Our work is based on the harmonic model proposed in [2] and [14]. We review the harmonic images briefly in 3.1. To analyze with the harmonic images, we need to know the albedo map and the normal map of the face. The information is obtained with the "human face class" constraints, which is introduced in subsection 3.2. Then lighting estimation based on the harmonic model and face relighting are given in subsection 3.3 and 3.4, respectively.

3.1. *The harmonic images model*

Given a convex Lambertian object in a distance isotropic illumination, its irradiance E is

$$\begin{aligned} E(\alpha,\beta) &= \int_{\Omega'} L(\theta_i,\phi_i)\cos\theta_i' d\Omega', \\ &= \int_{\Omega'} L\left(R^{\alpha,\beta}\left(\theta_i',\phi_i'\right)\right)\cos\theta_i' d\Omega', \end{aligned} \tag{1}$$

where (α,β) is the normal, L is the incident illumination, the limit of integration, Ω', is the local upper hemisphere and $R^{\alpha,\beta}$ is a rotation operation that rotates the local incident angle (θ_i',ϕ_i') to the global coordinate (θ_i,ϕ_i). Note that Eq.(1) is a rotational convolution.

Since the reflection Eq.(1) can be viewed as a convolution, it is natural to analyze it in frequency-space domain. The appropriate signal processing tools for the sphere are spherical harmonics, which are the equivalent to the Fourier series in 2-D. The limits of the integration of the Lambertian reflection are large (the local upper hemisphere), i.e., the irradiance is broad in spatial (angular) domain. Therefore, according to the Heisenberg uncertainty principle, it must localize in low frequency components in frequency domain. Analyzing the reflection Eq.(1) in frequency-space with spherical harmonics, Basri et al [2] and Ramamoorthi [14] proved that most energy of the irradiance was constrained in the three low order frequency components and got its frequency formula as

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

where $A_l(A_0 = \pi, A_1 = 2\pi/3, A_2 = \pi/4)$ [2, 14] are the spherical harmonic coefficients of Lambertian reflectance, L_{lm} are the coefficients of the incident light, and Y_{lm} are the spherical harmonic functions. The polynomial form of the spherical harmonic functions is defined as

$$\begin{aligned}
(x,y,z) &= (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta), \\
Y_{00}(\theta,\phi) &= \frac{1}{2}\frac{1}{\sqrt{\pi}}, \\
(Y_{1-1};Y_{10};Y_{11})(\theta,\phi) &= \frac{1}{2}\sqrt{\frac{3}{2\pi}}(y;z;x), \\
(Y_{2-2};Y_{2-1};Y_{21})(\theta,\phi) &= \frac{1}{2}\sqrt{\frac{15}{2\pi}}(xy;yz;xz), \\
Y_{20}(\theta,\phi) &= \frac{1}{4}\sqrt{\frac{5}{\pi}}(3z^2-1), \\
Y_{22}(\theta,\phi) &= \frac{1}{4}\sqrt{\frac{15}{2\pi}}(x^2-y^2).
\end{aligned} \tag{3}$$

Assuming the surface of human faces is convex Lambertian surface, let $\lambda(x,y)$ denotes the albedo of a point (x,y) in the image and $(\alpha(x,y),\beta(x,y))$ denotes its normal. The intensity of the point is

$$\begin{aligned}
I(x,y) &= \lambda(x,y)E(\alpha(x,y),\beta(x,y)) \\
&\approx \sum_{l=0}^{2}\sum_{m=l}^{l} L_{lm}\lambda(x,y)A_l Y_{lm}(\alpha(x,y),\beta(x,y)), \\
&= \sum_{l=0}^{2}\sum_{m=l}^{l} L_{lm} b_{lm}(x,y)
\end{aligned} \tag{4}$$

where

$$b_{lm}(x,y) = \lambda(x,y)A_l Y_{lm}(\alpha(x,y),\beta(x,y)), \tag{5}$$

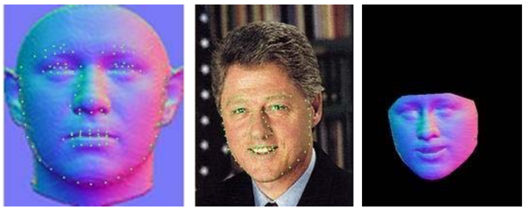
are the harmonic images of the face. The harmonic images can be viewed as the images of the face seen under harmonic lights. Harmonic light is a virtual light in which only one harmonic component is reserved.

3.2. *The harmonics images of a face*

To analyze with the harmonic images b_{lm}, one needs to know the albedo map $\lambda(x,y)$ and the normal map $(\alpha(x,y),\beta(x,y))$ of the given face. However, in many face recognition systems, only a single image is available. Therefore, in this paper, the albedo map and the normal map are obtained with the "human face class" constraints.

Human faces can be assumed rationally to have the similar shapes. This has always been used in many previous algorithms [17, 23, 25]. If the poses of the images are identical, the normal map of a given face can be obtained by aligning the facial image with the average facial normal map of that pose. Given a facial image, to create the correspondence between the average facial normal map and the image, we first create the correspondence between the feature points on the average normal map and those on the image. Then the rest pixels on the average normal map and the image are aligned by using image warping technique. In this paper, the feature points in the facial image are labeled automatically with an improved ASM method [22].

The average facial normal map and the warped normal map of a facial image are shown in Fig.1.



(a) (b) (c)

Fig. 1. The average facial normal map and the warped normal map. (a) The average facial normal map and its feature points. (b) The facial image and its feature points. (c) The warped normal map for the face in (b).

For albedo map, we assume that the albedo of the face surface is quasi-constant, i.e., the albedo map contains no low frequency components ($1 \le l \le 4$), except for DC component[a]. Though the albedo map of a face does not satisfy this constraint strictly, we find that we can still obtain good results in practice. Wen [23] had justified numerically that the low-frequency components of the albedo map of a face are very small. This can also be verified by vision that most regions of the face are skin with almost the same albedo.

Once the albedo map and the normal map of the face has been estimated, the harmonic images can be computed using Eq.(5).

3.3. *Lighting estimation based on the harmonic images*

To re-render the face to a canonical illumination, we need to recover the original illumination from the input image first.

Given an image of a face and its harmonic images, the only unknown variables in Eq.(4) are the nine illumination coefficients. So, lighting estimation is to estimate the nine low spherical harmonic coefficients of the illumination, L_{lm} ($0 \le l \le 2, -l \le m \le l$).

Given an input image $\mathbf{I}$ (a row vector of n elements, n is the number of the pixels in the image) of an object, if the harmonic images of the object are known, then the coefficients of the illumination $\mathbf{L}$ can be gotten by solving the least squares problem

$$\hat{\mathbf{L}} = \arg\min_{\mathbf{L}} \left\| \mathbf{BL} - \mathbf{I} \right\|, \tag{6}$$

where $\mathbf{B}$ denotes the harmonic images, arranged as a $n \times 9$ matrix. Every column of $\mathbf{B}$ contains one harmonic image b_{lm} as in Eq.(5). Note that the estimated lighting coefficients, $\hat{L}_{lm}$, are scaled up to a factor of the albedo of the skin. i.e., $\hat{L}_{lm} = \lambda_{00} L_{lm}$, where λ_{00} is the DC coefficient of the albedo map and L_{lm} is the coefficients of the real illumination.

3.4. *Face relighting based on the ratio image technique*

The ratio image technique [9, 17, 23] had been used to remove the material dependency of the diffuse object. Here we used the illumination ratio image to relight a face to a canonical illumination, given the original image and its illumination estimated.

Let S_0 be the predefined canonical lighting condition. The illumination ratio image R for the lighting condition $\hat{S}$ is defined as:

[a] 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.

$$R(x,y)=\frac{I(x,y)}{I_0(x,y)}=\frac{\hat{E}(\alpha(x,y),\beta(x,y))}{E_0(\alpha(x,y),\beta(x,y))}, \tag{7}$$

where I and I_0 are the images taken under the lighting conditions of $\hat{S}$ and S_0, $\hat{E}$ and E_0 are the corresponding incident irradiances, (x,y) ranges over the whole image.

Given an input face image I and the estimated illumination $\hat{S}$, with the illumination ratio image R defined in Eq.(7), its canonical image under the predefined lighting condition S_0 can be derived by:

$$I_0(x,y)=\frac{I(x,y)}{R(x,y)}, \tag{8}$$

where (x,y) ranges over the whole image.

4. Robust Face Recognition Based on Facial Image Relighting

As our goal is face recognition, we discuss some issues about applying the proposed relighting algorithm to face recognition in this section. First, we consider two canonical illuminations, under which different information is encoded in the re-rendered canonical images for face recognition. Then the post-process to weaken the artifacts in the re-rendered canonical image is described in subsection 4.2.

4.1 *The Canonical Illuminations for robust face recognition*

The intuitive idea is that the intrinsic features of the face, albedo and/or shape, are suitable for face recognition under varying illuminations. Because the shape information is obtained by aligning the facial image with the average face model, the shape itself is not accurate enough for face recognition. So we discuss the other two possible cases: just the albedo and both the albedo and the shape.

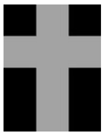
According to Eq.(4), the effect of the shape on the intensity of the image $I(x,y)$ is expressed by the irradiance $E(x,y)$. And according the Eq.(2), the irradiance $E(x,y)$ depends on the shape with the spherical harmonic function $Y_{lm}(\alpha(x,y),\beta(x,y))$. We can derive from the polynomial form of the spherical harmonic functions in Eq.(3) that 1) Y_{00} is a constant irrelevant to the shape; 2) The values of the Y_{lm} depend on the normal variable $(\alpha(x,y),\beta(x,y))$ when $l=1,2$. The effect of each spherical harmonic function Y_{lm} is scaled with the weight factor A_l. As $A_1 > A_2$, the dependence of the irradiance $E(x,y)$ on the shape lies mostly in the first order frequency components.

Therefore, if the canonical illumination contains the DC component only, the irradiance is independent on the face normal. Therefore, the image under the canonical illumination encodes just the albedo information. The illumination environment that contains only DC component is the uniform illumination, in which the intensities of light are equal in all directions. The ratio of the three low order energies of the uniform illumination is $1.0:0.0:0.0$. The irradiance of the uniform illumination is visualized in Tab. 1(a) with the corresponding harmonic coefficients listed after it.

However, generally, the natural illumination in the real world consists one or more primary light sources as well as the ambient lights. A typical case is one point light source with the ambient lights. The images captured under such illumination encode both the information of the albedo and the information of the shape. There are some images captured under such environment in the CMU-PIE face database [19]. We estimate the illumination from one of the images taken with a frontal flash with the room ambient lights on. The estimated illumination is given in Tab. 1(b) with the nine harmonic coefficients shown. The ratio of the three low order energies of the uniform illumination is $0.8:0.18:0.02$. Compared with that of the uniform illumination, more energy is contained in the first order frequency components. Therefore, more information about the face shape is encoded in the canonical images.

After defining the canonical illumination, the canonical image of any given original image I can be computed directly with the Eq. (8).

Tab. 1. The typical canonical illuminations

	Irradiance map	The nine spherical coefficients of the illumination									
			L_{00}	L_{11}	L_{10}	L_{1-1}	L_{21}	L_{2-1}	L_{2-2}	L_{20}	L_{22}
(a)		G	100	0	0	0	0	0	0	0	0
		R	100	0	0	0	0	0	0	0	0
		B	100	0	0	0	0	0	0	0	0
(b)		R	103	0.4	59	38	0.4	33	12	16	36
		G	89	-0.1	53	34	-0.1	32	7	26	37
		B	100	-2.0	62	39	-1.0	37	6	52	15

(a) The uniform illumination.
(b) A frontal flash with the room lights.

4.2. Post-process the limited dynamic range problem

Because digital image has limited dynamic range, ratio based re-lighting would have artifact effect where the intensities of some skin pixels are too low or nearly saturated. These pixels are called outliers. If we do not deal with these pixels, they may introduce confusion in recognition. Some other re-lighting based methods [21] avoid this problem by re-lighting the images in gallery to the lighting condition of the probe, because the lighting condition of the gallery is usually satisfying. However, this strategy needs to re-light all the images in the gallery for each probe. This is a high computational consumption method.

A constrained texture synthesis [23] can be used to alleviate the problem. However, the synthesized textures for the outliers add no more valuable information for recognition, so we use a much simpler method for face recognition, i.e., just discarding these pixels. This is like portion face recognition, except that the portion is automatically decided. We use the face part of the irradiance environment map E_{org} to detect the outliers. If the intensity of a pixel in irradiance map is too low, the corresponding pixel in the image must be in dark attached shadows and it is declared as an outlier. As the high lighting pixels are usually few, it will not affect recognition very much.

If the energy of the DC component of the irradiance is larger compared with that of the first order, the detection of outliers is not necessary because the dark shadows are usually caused by the energy of the first order. For example, the ratio between the energy of the irradiance in DC component and the first order for a flash with the room lights (such as subset "lights" in CMU-PIE face database) and that of just a flash (such as "illum" subset in CMU-PIE face database) is 4.53 and 2.53, respectively. There are few artifacts in the re-rendered images of the images captured under a flash with the room lights. While for the images captured under just a flash, there are more artifacts. Therefore we need only do this post- procedure for images captured under the latter illumination.

Some examples of the detected outliers are illustrated in Fig. 2.

(a) (b) (c) (d)

Fig. 2. Outliers detection. (a) The input images. (b) The estimated face part of the irradiance environment maps. (c) The detected outliers marked with black pixels. (d) The re-rendered images under the uniform illumination.

5. Experimental Results

We select two publicly available face databases, the CMU-PIE face database [19] and the Yale B face database [6], for our experiments. Both of them contain face images with well-controlled lighting conditions.

The images in the CMU-PIE face database [19] vary in pose, illumination, and expression. We select the frontal images with lighting variations for the experiments. There are 21 flashes and the room lights to illuminate the face. The images are captured under the room lights on and off, resulting in 43 different conditions. There are total 68 persons included in the database. Some examples of the facial images under different illuminations are shown in Fig. 3 (the first row and the second row). For more details about the CMU-PIE face database, please refer to [19].

The Yale face database [6] contains images of 10 people under 9 poses and 64 illuminations per pose. We use 45×10 face images for 10 subjects in the frontal pose with each subject having 45 face images taken under different directional light sources. Fig. 4 (the first row and the second row) shows some of the images used in our experiments.

The results of the three main phases of the method: illumination estimation, relighting and face recognition are given in following subsections. Because most of the lighting conditions in the Yale B face database are included in the CMU-PIE face database, only the results of illumination estimation and face recognition on the Yale B face database are given.

5.1. Experimental results of illumination estimation

As the effect of the illumination on the appearance of the face is achieved though the irradiance. The results of the illumination estimation in this subsection are showed in the irradiance maps.

Fig. 3 and Fig.4 give some examples of the irradiance maps computed with the estimated nine spherical harmonic coefficients of the illuminations estimated from the images in the CMU-PIE face database and the Yale B face database respectively. Though the faces in the first row and the second row in the two figures are different in shape and albedo, the corresponding estimated irradiance maps are fairly similar.

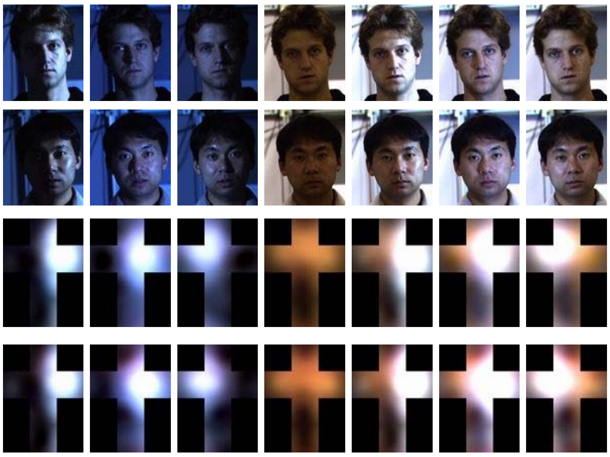


Fig. 3. Some examples of the estimated irradiance maps from the CMU-PIE face database. The first two rows are the images of two persons and the last two rows are the estimated irradiance maps. Each column of the first two rows is captured under the same illumination. Evidently, though the faces in the first two rows are different, the estimated irradiance maps are fairly similar.

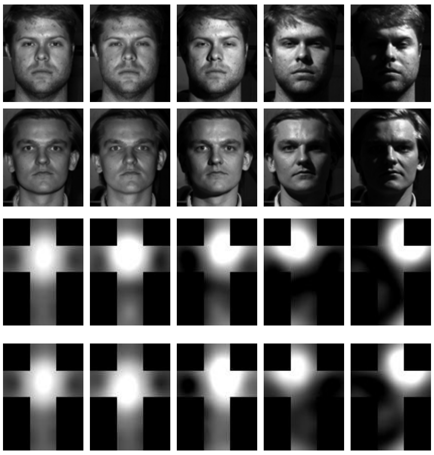
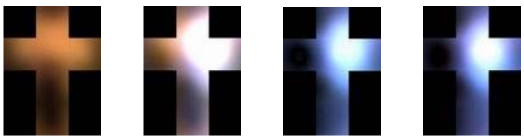


Fig. 4. Some examples of the estimated irradiance maps from the Yale B face database. The first two rows are the images of two persons and the last two rows are the estimated irradiance maps. Every column of the first two rows is captured under the same illumination. Evidently, though the faces in the first two rows are different, the estimated irradiance maps are fairly similar.

The CMU-PIE database includes the images taken under just the flash, under just the room lights and with both of them. Fig. 5 contains the estimated irradiance maps under just the room lights and the illuminations with both the room lights and one of the flashes. We also show the difference between those two irradiance maps and compare it with the irradiance map of the same flash with the room lights off. The two irradiance maps are fairly similar.



(a) Room lights (b) With flash (c) Difference (d) Flash only

Fig. 5. An example of the subtracting of the estimated irradiance maps from the CMU-PIE face database. The difference irradiance map is quite similar to the estimated irradiance maps of the same flash with the room lights off.

As the locations of the feature points labeled by ASM are not correct in some cases, we need to know how the illumination estimation is sensitive to the alignment. We compute the sensitivity of the lighting estimation to alignment using the difference between the irradiance of the perfect alignment and that of the loose alignment. In perfect alignment, the feature points are adjusted manually. While in the loose alignment, only the location of the two eyes is used. The results of the automatic alignment labeled by ASM can be supposed to be between of the perfect alignment and the loose alignment. The sensitivity of the lighting estimation to alignment is measured as

$$S = \frac{\sqrt{\sum_{i=1}^{n} \left\| E_{perfect}(i) - E_{loose}(i) \right\|^2}}{\sum_{i=1}^{n} E_{perfect}(i)}, \tag{9}$$

where n is the number of the pixels in the irradiance maps, $E_{perfect}$ and E_{loose} are the estimated irradiance maps with the perfect alignment and the loose alignment, respectively.

The average sensitivities of different lighting conditions in the CMU-PIE face database are given in Fig. 6. The sensitivities are very small, which indicates that the lighting estimation is not sensitive to the alignment.

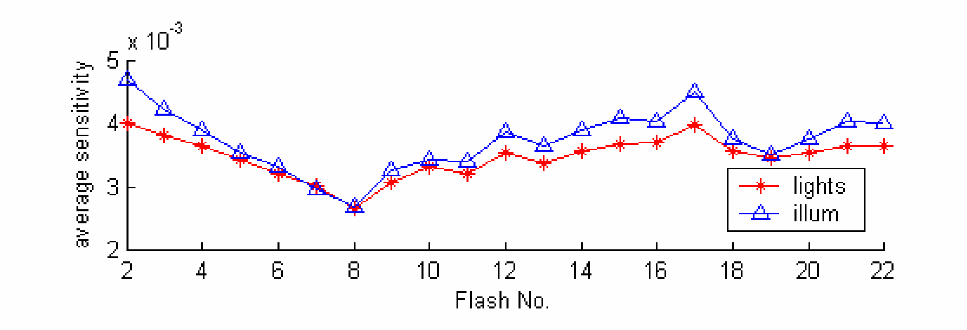


Fig. 6. The average sensitivity of the lighting estimation to alignment. “illum” are the illumination of a flash without the room lights in the CMU-PIE face database. “lights” are the illumination of a flash with the room lights. The sensitivities are small.

5.2. Experimental results of relighting

Some of the re-rendered images are given in Fig. 7. The re-rendered images in the last two rows look more similar than the original images in the first row.

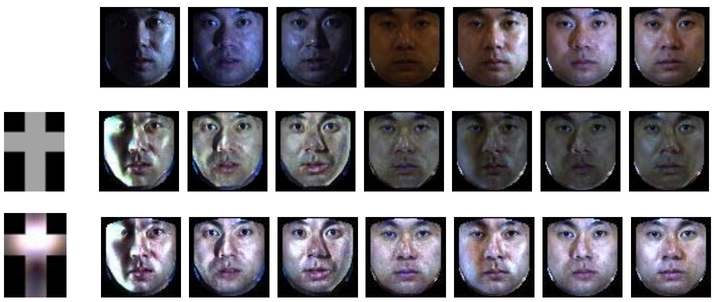


Fig. 7. The results of relighting. The first row illustrates the cropped facial images of the person in the second row of Fig. 3. The other two rows contain the corresponding re-rendered images under the uniform illumination and the room lights with a frontal flash, respectively.

Because of the limited dynamic range of the images, the relit images in the first three columns in Fig. 7 do not look realistic to some extent. If the dynamic range of the input image is more reasonable, i.e. the lighting of the input image is more satisfying; the relit images will look more realistic. This is verified by the images in the last four columns, which are illuminated by the room lights. More examples are given in Fig. 8, in which the inputs are images taken under good illumination.

Fig. 8. More results of relighting. The first row is the input facial images. The second row is the real facial images taken under different lighting conditions and the third is the corresponding re-rendered images (relighting the input image to the illumination of corresponding image in the second row). The corresponding images in the last two rows look similar.

5.3 Experimental results of face recognition

We verify the effect of relighting 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 with varying illuminations in the CMU-PIE face database are classified into two subsets: "illum" and "lights". There are some differences between the two subsets. In the "illum" subset, the room lights is turned off and the subject is not wearing glasses. While in the "lights" subset, the room lights is turned on. And the subject is wearing glasses if and only if they normally do. In both cases, there 21 flashes to illuminate the faces and the flash numbers are 02-21. The flash number “01” corresponds to no flash (only for the “lights” subset). See the reference [19] for more details of the meaning of the flash number. To test the variations in illuminations only, both the gallery and probe are from the same subsets. In the experiments, we select the flash “11” in each subset as gallery for the both subsets. As the illumination of the gallery released by the CMU-PIE face database (only with the room lights on and no flash) is the same as the flash “01” in subset “lights”, the results on the subset “lights” with the gallery flash “01” is also given in the paper.

The experimental results of face recognition on the CMU-PIE database are listed in Tab. 2. Two canonical illuminations discussed in section 4.1 are compared: the uniform illumination and a frontal flash with the room lights. As the locations of the feature points labeled by ASM are not correct in some cases, e.g., in the cases of extreme illuminations, we need to know how the face recognition is sensitive to the location of the feature points. The results of face recognition with loose alignment using the location of the two eyes are given also in Tab. 2, which can be compared with those of the perfect feature points adjusted manually.

Tab. 2. Error rate comparisons between different canonical illuminations and different alignments on the CMU-PIE database.

Gallery	Probe	Canonical illuminations	Error rate (%)	
			Perfect alignment	Loose alignment
“illum”(f11)	“illum”	No	47.1	47.1
		Uniform illumination	4.0	13.7
		A frontal flash with room lights	3.4	13.5
“lights”(f11)	“lights”	No	14.3	14.3
		Uniform illumination	0.0	0.0
		A frontal flash with room lights	0.0	0.0
“lights”(f01)	“lights”	No	18.5	18.5
		Uniform illumination	1.5	0.0
		A frontal flash with room lights	1.3	0.0

Following [6], we group the facial images in Yale B face database into four subsets. Each subset contains images illuminated from a specific range of directions. Please refer to [6] for more information about this grouping. The gallery is the images taken with a frontal point source.

The experimental results of face recognition on the Yale B face database with perfect alignment and loose alignment are given in Tab. 3 and Tab .4, respectively. The results are a little worse than that of the Illumination Cone [6], 9PL [8], and Harmonic Exemplars [24], which needed 7~9 images for gallery. Compared with the illumination normalization in [21], which used only one image for gallery as in our method, our results are better.

Tab. 3. Error rate comparisons between different canonical illuminations with perfect alignments on the Yale B face database.

Canonical illuminations	Error rate (%)			
	Subset 1	Subset 2	Subset 3	Subset 4
No	0.0	2.5	25.0	60.0
Uniform illumination	0.0	0.0	5.0	19.6
A frontal flash with room lights	0.0	0.0	0.0	10.0

Tab. 4. Error rate comparisons between different canonical illuminations with loose alignments on the Yale B face database.

Canonical illuminations	Error rate (%)			
	Subset 1	Subset 2	Subset 3	Subset 4
No	0	2.5	25.0	60.0
Uniform illumination	0	0	8.3	20.0
A frontal flash with room lights	0	0.8	8.3	18.6

Several conclusions can be derived from the experimental results of face recognition above.

1. The error rates are much lower after relighting in all of the cases, which indicates that the relighting strategy does improve the performance of the face recognition system.
2. The best canonical illumination is the illumination of a frontal flash and the room lights. The possible reason is that both the texture and the shape information of the face are encoded and used for face recognition. However, the differences of the performances between the two canonical illuminations are not very large, which implies that the most valuable information for recognition is the texture information.
3. The error rates of loose alignment do not increase significantly compared with those of the perfect alignment. This indicates that we can use the loose alignment for a real application when the perfect feature points are not available. The decrease of the canonical illumination of a frontal flash with room lights are more than that of the uniform illumination.
4. The error rate of the two canonical illumination with loose alignment are almost the same, which indicates that if the shape are not elaborate, the shape information encoded in the canonical form under the canonical illumination of a frontal flash with room lights are little value to face recognition.

6. Conclusion and Future work

Based on the recent development that the images of a diffuse object under varying illuminations lie in a nine-dimensional subspace analytically computable through the spherical harmonic images, we propose a technique for face relighting under generic illumination, i.e., calibrating the input face image to a predefined canonical illumination, in order to reduce the negative effect of the varying illumination in face recognition.

According to the harmonic images theory, the nine lower frequency components have covered most of the energy of the illumination for a Lambertian surface. Therefore, illumination estimation is to estimate the nine low spherical harmonic coefficients of the illumination. Given a single facial image, this is achieved by exploiting the prior knowledge of the human faces to get the harmonic images of the given face. After illumination estimation, the facial image is re-rendered to a predefined canonical illumination by altering the nine harmonic lighting coefficients. We consider two different kinds of canonical illuminations: the uniform illumination and a flash with the ambient lights, among which the former encodes merely the facial texture information, while the latter encodes both the facial texture and shape information. Experiments are conducted to compare the recognition performance using the original face and the re-rendered images. Our experiments on the CMU-PIE face database and the Yale B face database have shown that the proposed method improves impressively the performance of a face recognition system in the case of varying lighting conditions.

Though our experiments have shown that imprecise feature alignment would not degrade the final recognition results significantly, more accurate alignment would facilitate the 3D shape recovery and the subsequent recognition. Therefore, one of our future efforts will be the accurate alignment, especially under the non-ideal imaging environment.

Acknowledgements

This research is partly sponsored by Natural Science Foundation of China (under contract No. 60332010), National Hi-Tech Program of China (No.2001AA114190 and No. 2002AA118010), "100 Talents Program" of CAS, Open Project (KP0706200377) of Beijing Municipal key Laboratory on Multimedia and Intelligent Software Technology, BJUT, and ISVISION Technologies Co., Ltd. Portions of the research in this paper use the CMU-PIE face database and the Yale B face database. The authors wish to thanks everyone involved in collecting these data. The authors would like to thanks W. Wang for ASM algorithm.

References

1. 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*, 8(6) (1996) 1321-1340.
2. R. Basria and D. Jacobs, Lambertian Reflectance and Linear Subspaces, *Proc. 8th IEEE Int. Conf. Computer Vision,* 2001, pp. 383-390.
3. P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, “Eigenfaces vs Fisherfaces: recognition using class specific linear projection”, *IEEE Trans. PAMI*, 20(7) 1997 711-720.
4. V. Blanz and T. Vetter, “Face Recognition Based on Fitting a 3D Morphable Model”, *IEEE Trans. PAMI*, 25(9) (2003), 1-12.
5. R. Chellappa, C. L. Wilson, and S. Sirohey, “Human and Machine Recognition of Faces: A Survey”, *Proc. the IEEE*, 83(5) (1995) 705-740.
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. PAMI*, 23(6) (2001) 643-660.
7. P. Hallinan, A low-dimensional representation of human faces for arbitrary lighting conditions, *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'94)*, 1994, pp. 995-999.
8. K. C. Lee, J. Ho and D. Kreigman, Nine Points of Light: Acquiring Subspaces for Face Recognition under Variable Lighting, *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'01)*, 2001, Vol. I, pp.519-526.
9. Z. Liu, Y. Shan, and Z. Zhang, “Expressive expression mapping with ratio images”, *Proc.Siggraph’01*, 2001, pp. 271–276.

10. B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian Face Recognition", *Pattern Recognition*, 33(11) (2000) 1771-1782.

11. Y. Moses, Y. Adini, and S. Ullman, "Face Recognition: the Problem of Compensating for Changes in Illumination Direction", *Proc. 3rd European Conf. Computer Vision (ECCV'94)*, 1994, Vol. I, pp. 286-296.

12. P. J. Phillips, H. Moon, S. A. Rizvi, *et. al.*, "The FERET Evaluation Methodology for Face-Recognition Algorithms", *IEEE Trans. PAMI*, 22(10) (2000) 1090-1104.

13. P. J. Phillips, P. Grother, R. J. Micheals, *et. al.*, "FRVT 2002: Evaluation Report", http://www.frvt.org/DLs/FRVT_2002_ Evaluation_Report.pdf, March 2003.

14. R. Ramamoorthi and P. Hanrahan, "On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object", *Journal of the Optical Society of America A*, 18(10) (2001) 2448-2459.

15. R. Ramamoorthi, "Analytic PCA Construction for Theoretical Analysis of Lighting Variability in Images of a Lambertian Object", *IEEE Trans. PAMI*, 24(10) (2002) 1-12.

16. A. Shashua, "On photometric issues in 3D visual recognition from a single 2D image", *International Journal of Computer Vision (IJCV)*, 21(1-2) (1997) 99–122.

17. A. Shashua and T. Riklin-Raviv, "The Quotient Image: Class-Based Re-Rendering and Recognition with Varying Illuminations", *IEEE Trans. PAMI*, 23(2) (2001) 129-139.

18. T. Sim and T. Kanade, "Combining Models and Exemplars for Face Recognition: An Illuminating Example", *Proc. CVPR 2001 Workshop on Models versus Exemplars in Computer Vision*, 2001.

19. T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database", *Proc. 5th IEEE International Conference on Automatic Face and Gesture Recognition (FGR'02)*, 2002, pp.53-58.

20. M. Turk and A. Pentland, "Eigenfaces for Recognition", *Journal of cognitive neuroscience*, 3(1) (1991) 71-86.

21. H. Wang, S. Li, Y. Wang, and W. Zhang, "Illumination Modeling and Normalization for Face Recognition", *Proc.IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures*, 2003, pp.104-111.

22. W. Wang, S. Shan, W. Gao, *et. al.*, "An Improved Active Shape Model for Face Alignment", *Proc. 4th IEEE Int. Conf. Multimodal Interfaces (ICMI'02*, 2002, pp. 523-528.

23. Z. Wen, Z. Liu and T.S. Huang, Face Relighting with Radiance Environment Maps, *Proc.IEEE Conf. Computer Vision and Pattern Recognition (CVPR'03)*, 2003, Vol. II, pp. 158-165.

24. L. Zhang and D. Samaras, "Face Recognition under Variable Lighting using Harmonic Image Exemplars", *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'01)*, 2003, Vol. I, pp. 19-25.

25. W. Zhao and R. Chellappa, "Illumination-insensitive Face Recognition using Symmetric Shape-from-Shading", *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR'00)*, 2000, Vol. 1, pp.1286-1293.

26. W. Zhao, R. Chellappa, A. Rosenfeld, and J. Phillips, "Face Recognition: A Literature Survey", Technical Report, CFAR-TR00-948, University of Maryland, 2000. (Revised 2002)