Learning from Real Images to Model Lighting Variations for Face Images

Xiaoyue Jiang[1,2], Yuk On Kong[3], Jianguo Huang[1],
Rongchun Zhao[1], and Yanning Zhang[1]

[1] School of Computer Science, Northwestern Polytechnical University,
Xi'an, 710072, China
[2] School of Psychology, University of Birmingham, Birmingham B15 2TT, UK
[3] Department of Electronics and Informatics, Vrije Universiteit Brussel,
Brussels 1050, Belgium

Abstract. For robust face recognition, the problem of lighting variation is considered as one of the greatest challenges. Since the nine points of light (9PL) subspace is an appropriate low-dimensional approximation to the illumination cone, it yielded good face recognition results under a wide range of difficult lighting conditions. However building the 9PL subspace for a subject requires 9 gallery images under specific lighting conditions, which are not always possible in practice. Instead, we propose a statistical model for performing face recognition under variable illumination. Through this model, the nine basis images of a face can be recovered via maximum-a-posteriori (MAP) estimation with only one gallery image of that face. Furthermore, the training procedure requires only some real images and avoids tedious processing like SVD decomposition or the use of geometric (3D) or albedo information of a surface. With the recovered nine dimensional lighting subspace, recognition experiments were performed extensively on three publicly available databases which include images under single and multiple distant point light sources. Our approach yields better results than current ones. Even under extreme lighting conditions, the estimated subspace can still represent lighting variation well. The recovered subspace retains the main characteristics of 9PL subspace. Thus, the proposed algorithm can be applied to recognition under variable lighting conditions.

1   Introduction

Face recognition is difficult due to variations caused by pose, expression, occlusion and lighting (or illumination), which make the distribution of face object highly nonlinear. Lighting is regarded as one of the most critical factors for robust face recognition. Current attempt to handle lighting variation by either finding the invariant features or modeling the variation. The edge based algorithm [1] and the algorithm based on quotient image [2,3,4]belong to the first type. But these methods cannot extract sufficient features for accurate recognition.

Early work on modeling lighting variation [5,6] showed that a 3D linear subspace can represent the variation of a Lambertian object under a fixed

pose when there is no shadow. With the same Lambertian assumption, Belhumeur and Kriegman [7] showed that images illuminated by an arbitrary number of point light sources formed a convex polyhedral cone, i.e. the illumination cone. In theory, the dimensionality of the cone is finite. They also pointed out that the illumination cone can be approximated by a few properly chosen images. Good recognition results of the illumination cone in [8] demonstrated its representation for lighting variation. [9] indicated that lighting subspace of Lambertian object can be approximated by a linear subspace with dimension between three and seven. Recent research is mainly focused on the application of low-dimensional subspace to lighting variation modeling. With the assumption of Lambertian surface and non-concavity, Ramamoorith and Hanrahan [10] and Basri and Jacobs[11] independently introduced the spherical harmonic (SH) subspace to approximate the illumination cone. However, the harmonic images (basis images of SH subspace) are computed from the geometric and albedo information of the subject's surface. In order to use the SH subspace theory, a lot of algorithms applied the 3D model of faces to handling lighting variations [12,13,14,15,16]. However, recovering the 3D shape from images is still an open problem in computer vision.

Lee et al.[19] built up a subspace that is nearest to the SH subspace and has the largest intersection with the illumination cone, called the nine points of light (9PL) subspace. It has a universal configuration for different subjects, i.e. the subspace is spanned by images under the same lighting conditions for different subjects. In addition, the basis images of 9PL subspace can be duplicated in real environments, while those of the SH subspace cannot because its the basis images contain negative values. Therefore the 9PL subspace can overcome the inherent limitation of SH subspace. Since the human face is neither completely Lambertain nor entirely convex, SH subspace can hardly represent the specularities or cast shadows (not to mention inter-reflection). The basis images of 9PL subspace are taken from real environment, they already contain all the complicated reflections of the objects. Therefore the 9PL subspace can give a more detailed and accurate description of lighting variation.

In practice, the requirement of these nine real images cannot always be fulfilled. Usually there are fewer gallery images (e.g. one gallery image) per subject, which can be taken under arbitrary lighting conditions. In this paper, we propose a statistical model for recovering the 9 basis images of the 9PL subspace from only one gallery image. Zhang and Samaras [12] presented a statistical method for recovering the basis images of SH subspace instead. In their training procedure, geometric and albedo information is still required for synthesizing the harmonic images. In contrast, the proposed method requires only some real images that can be easily obtained in real environment. Since the recovered basis images of the 9PL subspace contain all the reflections caused by the shape of faces, such as cast shadows, specularities, and inter-reflections, better recognition results are obtained, even under extreme lighting conditions. Compared with other algorithms based on 3D model [12,15,16], the proposed algorithm is entirely a 2D algorithm, which has much lower computational complexity. The

proposed algorithm also has comparable recognition results. Note that we do not consider pose variation in this paper and assume that all subjects are in the frontal pose.

This paper is organized as follows. In Section 2, we briefly summarize the methods of low-dimensional linear approximation of the illumination cone, including the SH subspace and the 9PL subspace. The training of our statistical model and the application of the model for recovering basis images from only one gallery image are described in Sections 3 and 4 respectively. Section 5 is dedicated to the experimental results. The conclusion is given in Section 6.

2   Approximation of the Illumination Cone

Belhumeur and Kriegman [7] proved that the set of n-pixel images of a convex object that had a Lambertian surface illuminated by an arbitrary number of point light sources at infinity formed a convex polyhedral cone, called the illumination cone $\mathcal{C}$ in $\mathcal{R}^n$. Each point in the cone is an image of the object under a particular lighting condition, and the entire cone is the set of images of the object under all possible lighting conditions. Any images in the illumination cone $\mathcal{C}$ (including the boundary) can be determined as a convex combination of extreme rays (images) given by

$$I_{ij} = max(\tilde{B}\tilde{s}_{ij}, 0) \tag{1}$$

where $\tilde{s}_{ij} = \tilde{b}_i \times \tilde{b}_j$ and $\tilde{B} \in \Re^{n \times 3}$. Every row $\tilde{b}_i$ of $\tilde{B}$ is a three element row vector determined by the product of the albedo with the inward pointing unit normal vector of a point on the surface. There are at most $q(q-1)$ extreme rays for $q \leq n$ distinct surface normal vectors. Therefore the cone can be constructed with finite extreme rays and the dimensionality of the lighting subspace is finite. However, building the full illumination cone is tedious, and the low dimensional approximation of the illumination cone is applied in practice.

From the view of signal processing, the reflection equation can be considered as the rotational convolution of incident lighting with the albedo of the surface [10]. The spherical harmonic functions $Y_{lm}(\theta, \phi)$ are a set of orthogonal basis functions defined in the unit sphere, given as follows,

$$Y_{lm}(\theta, \phi) = N_{lm}P_l^m(\cos\theta)\exp^{im\phi} \tag{2}$$

where $N_{lm} = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}$, (θ, ϕ) is the spherical coordinate (θ is the elevation angle, which is the angle between the polar axis and the z-axis with range $0 \leq \theta \leq 180^o$, and ϕ is the azimuth angle with the range $-180^o \leq \phi \leq 180^o$). P_l^m is the associated Legendre function, and the two indices meet the conditions $l \geq 0$ and $l \geq m \geq -l$. Then functions in the sphere, such as the reflection equation, can be expanded by the spherical harmonic functions, which are basis functions on the sphere. Images can be represented as a linear combination of spherical harmonic functions. The first three order ($l \leq 3$) basis can account for 99% energy of the function. Therefore the first three order basis functions (altogether

9) can span a subspace for representing the variability of lighting. This subspace is called the spherical harmonic (SH) subspace .

Good recognition results reported in [11] indicates that the SH subspace $\mathcal{H}$ is a good approximation to the illumination cone $\mathcal{C}$. Given the geometric information of a face, its spherical harmonic functions can be calculated with Eq.(2). These spherical harmonic functions are synthesized images, also called harmonic images. Except the first harmonic image, all the others have negative values, which cannot be obtained in reality. To avoid the requirement of geometric information, Lee et al.[19] found a set of real images which can also serve as a low dimensional approximation to illumination cone based on linear algebra theory.

Since the SH subspace $\mathcal{H}$ is good for face recognition, it is reasonable to assume that a subspace $\mathcal{R}$ close to $\mathcal{H}$ would be likewise good for recognition. $\mathcal{R}$ should also intersect with the illumination cone $\mathcal{C}$ as much as possible. Hence a linear subspace $\mathcal{R}$ which is meant to provide a basis for good face recognition will also be a low dimensional linear approximation to the illumination cone $\mathcal{C}$. Thus subspace should satisfy the following two conditions [19]:

1. The distance between $\mathcal{R}$ and $\mathcal{H}$ should be minimized.
2. The unit volume ($vol(\mathcal{C} \cap \mathcal{R})$) of $\mathcal{C} \cap \mathcal{R}$ should be maximized (the unit volume is defined as the volume of the intersection of $\mathcal{C} \cap \mathcal{R}$ with the unit ball)

Note that $\mathcal{C} \cap \mathcal{R}$ is always a subcone of $\mathcal{C}$; therefore maximizing its unit volume is equivalent to maximize the solid angle subtended by the subcone $\mathcal{C} \cap \mathcal{R}$. If $\{\tilde{I}_1, \tilde{I}_2, \cdots, \tilde{I}_k\}$ are the basis images of $\mathcal{R}$. The cone $\mathcal{R}_c \subset \mathcal{R}$ is defined by $\tilde{I}_k$,

$$\mathcal{R}_c = \{I | I \in \mathcal{R}, I = \sum_{k=1}^{M} \alpha_k \tilde{I}_k, \alpha_k \geq 0\} \tag{3}$$

is always a subset of $\mathcal{C} \cap \mathcal{R}$. In practice the subcone $\mathcal{C} \cap \mathcal{R}$ is taken as $\mathcal{R}_c$ and the subtended angle of $\mathcal{R}_c$ is maximized. $\mathcal{R}$ is computed as a sequence of nested linear subspace $\mathcal{R}_0 \subseteq \mathcal{R}_1 \subseteq \cdots \subseteq \mathcal{R}_i \subseteq \cdots \subseteq \mathcal{R}_9 = \mathcal{R}$, with $\mathcal{R}_k (k > 0)$ a linear subspace of dimension i and $\mathcal{R}_0 = \emptyset$. First, EC denotes the set of (normalized) extreme rays in the illumination cone $\mathcal{C}$; and EC_k denotes the set obtained by deleting k extreme rays from EC, where $EC_0 = EC$. With $\mathcal{R}_{k-1}$ and EC_{k-1}, the sets EC_k and R_k can be defined iteratively as follows:

$$\tilde{I}_k = \arg \max_{I \in EC_{k-1}} \frac{dist(I, \mathcal{R}_{k-1})}{dist(I, \mathcal{H})} \tag{4}$$

where $\tilde{I}_k$ denotes the element in EC_{k-1}. $\mathcal{R}_k$ is defined as the space spanned by $\mathcal{R}_{k-1}$ and $\tilde{I}_k$. $EC_k = EC_{k-1} \backslash \tilde{I}_k$. The algorithm stops when $\mathcal{R}_9 \equiv \mathcal{R}$ is reached. The result of Eq.(4) is a set of nine extreme rays that span $\mathcal{R}$ and there are nine directions corresponding to these nine extreme rays. For different subjects, the nine lighting directions are qualitatively very similar. By averaging Eq.(4) of different subjects and maximizing this function as follows:

$$\tilde{I}_k = \arg \max_{I \in EC_{k-1}} \sum_{p=1}^{N} \frac{dist(I^p, \mathcal{R}_{k-1}^p)}{dist(I^p, \mathcal{H}^p)} \tag{5}$$

where I^p denotes the image of subject p taken under a single light source. H^p is the SH subspace of subject p. $\mathcal{R}_{k-1}^p$ denotes the linear subspace spanned by images $\{\tilde{I}_1^p, \cdots, \tilde{I}_k^p\}$ of subject p. The universal configuration of nine light source direction is obtained. They are $(0,0)$, $(68,-90)$, $(74,108)$, $(80,52)$, $(85,-42)$, $(85,-137)$, $(85,146)$, $(85,-4)$, $(51,67)$[14]. The directions are expressed in spherical coordinates as pairs of (ϕ, θ), Figure 1(a) illustrates the nine basis images of a person from the Yale Face Database B [8].

(a)   (b)
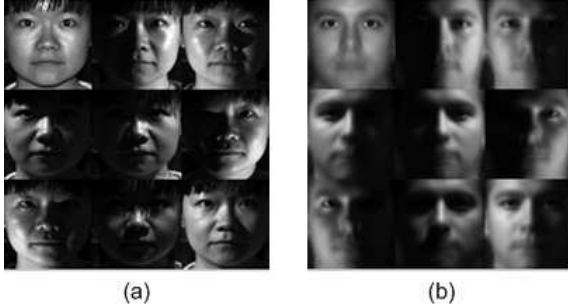
Fig. 1. the basis images of 9PL subspace. (a) images taken under certain lighting conditions can serve as the basis images of the object. (b) the mean images of the basis images estimated from the bootstrap data set.

3   Statistical Model of Basis Images

According to the universal configuration of lighting directions, we can apply nine images taken under controlled environment to spanning the 9PL linear subspace. However, even these nine images may not be available in some situations. Thus, we propose a statistical method for estimating the basis images from one gallery image. To build the statistical model, we must find the probability density function (pdf) of basis images and the pdf of the error term. Due to the limited amount of the training data, we use the bootstrap method to estimate the statistics of basis images. The recovering step is to estimate the corresponding basis images from one single image of a novel subject under arbitrary lighting conditions. For a given image, we first estimate its lighting coefficient. Then according to the maximum a posteriori (MAP) estimation, we obtain an estimation of the basis images. Finally, we apply the recovered subspace to face recognition. The probe image is identified as the face whose lighting subspace is closest in distance to the image.

Given nine basis images, we can reconstruct images under arbitrary lighting conditions as follows,

$$I = B\mathbf{s} + e(\mathbf{s}) \tag{6}$$

where $I \subset \Re^{d \times 1}$ is the image vector. $B \subset \Re^{d \times 9}$ is the matrix of nine basis images, every column of which is the vector of the basis image. $\mathbf{s} \subset \Re^{d \times 1}$ is the vector

of lighting coefficients which denotes the lighting conditions of the image. Error term $e(\mathbf{s}) \subset \Re^{d \times 1}$ is related to the pixels' position and lighting conditions.

For a novel image, we estimate its basis images through the maximum a posterior (MAP) estimation. That is

$$B_{MAP} = \arg \max_{B} P(B|I) \tag{7}$$

According to the Bayes rule

$$P(B|I) = \frac{P(I|B)P(B)}{P(I)} \tag{8}$$

where $P(I)$ is the evidence factor which guarantees that posterior probabilities would sum to one. Then Eq.(7) can become

$$B_{MAP} = \arg \max_{B} (P(I|B)P(B)) \tag{9}$$

In order to recover basis images from an image with Eq.(9), one should know the pdf of the basis images, i.e. $P(B)$, and the pdf of the likelihood, i.e. $P(I|B)$. Assuming the error term of Eq.(6) is normally distributed with mean $\mu_e(\mathbf{s})$ and variance $\sigma_e^2(\mathbf{s})$, we can deduce that the pdf of the likelihood $P(I|B)$ is also Gaussian with mean $B\mathbf{s} + \mu_e(\mathbf{s})$ and variance $\sigma_e^2(\mathbf{s})$ according to Eq.(6).

We assume that the pdf of the basis images B are Gaussians of means μ_B and covariances C_B as in [12,20]. The probability $P(B)$ can be estimated from the basis images in the training set. In our experiments, the basis images of 20 different subjects from the extented Yale face database B [8] are introduced to the bootstrap set. Note that, the basis images of every subject are real images which were taken under certain lighting conditions. The lighting conditions are determined by the universal configurations of the 9PL subspace. The sample mean μ_B and sample covariance matrix C_B are computed. Figure 1(b) shows the mean basis images, i.e. μ_B.

The error term $e(\mathbf{s}) = I - B\mathbf{s}$ models the divergence between the real image and the estimated image which is reconstructed by the low dimensional subspace. The error term is related to the lighting coefficients. Hence, we need to know the lighting coefficients of different lighting conditions. In the training set, there are 64 different images that taken under different lighting condition for every subject. Under a certain lighting condition, we calculate the lighting coefficients of every subject's image, i.e. $\mathbf{s}_k^p$ (the lighting coefficients of the pth subject's image under the lighting condition $\mathbf{s}_k$). For a training image, its lighting coefficients can be estimated by solving the linear equation $I = B\mathbf{s}$. The mean value of different subjects' lighting coefficients can be the estimated coefficients ($\bar{\mathbf{s}}_k$) for that lighting condition, i.e. $\bar{\mathbf{s}}_k = \sum_{p=1}^{N} \mathbf{s}_k^p / N$. Then, under a certain lighting condition, the error term the of the pth subject's image is

$$e_p(\bar{\mathbf{s}}_k) = I_k^p - B_p \bar{\mathbf{s}}_k \tag{10}$$

where I_k^p is the training image of the pth subject under lighting condition $\mathbf{s}_k$ and B_p is the basis images of the pth subject. Following the above assumption, we estimate the mean $\mu_e(\bar{\mathbf{s}}_k)$ and variance $\sigma_e^2(\bar{\mathbf{s}}_k)$ of the error term.

4   Estimating the Basis Images

As described in the previous section, the basis images of a novel image can be recovered by using the MAP estimation. Since the error term is related to lighting condition, we need to estimate the lighting condition, i.e. the lighting coefficients, of every image before calculating its basis images.

4.1   Estimating Lighting Coefficients

Lighting influences greatly the appearance of an image. Under similar illumination, images of different subjects will appear almost the same. The difference between the images of the same subject under different illuminations is always larger than that between the images of different subjects under the same illumination [21]. Therefore we can estimate the lighting coefficients of a novel image with an interpolation method. The kernel regression is a smooth interpolation method [22]. It is applied to estimating the lighting coefficients. For every training image, we have their corresponding lighting coefficients. For a novel image I_n, its lighting coefficient is given by

$$\mathbf{s} = \frac{\sum_{k=1}^{M} w_k \mathbf{s}_k^p}{\sum_{k=1}^{M} w_k} \tag{11}$$

$$w_k = \exp(-\frac{[D(I_n, I_k^p)]^2}{2(\sigma_{I_k^p})^2}) \tag{12}$$

where $D(I_n, I_k^p) = \|I_n - I_k^p\|_2$ is the L_2 norm of the image distance. $\sigma_{I_k^p}$ determines the weight of test image I_k^p in the interpolation. In the training set, every subject has 64 different images and there are altogether 20 different subjects. Thus, for a novel image, there are 20 images with similar illumination. In our experiment, we assign the farthest distance of these 20 images from the probe image to $\sigma_{I_k^p}$. $\mathbf{s}_k^p$ is the lighting coefficient of image I_k^p.

4.2   Estimating the Error Term

The error term denotes the difference between the reconstructed image and the real image. This divergence is caused by the fact that the 9PL subspace is the low-dimensional approximation to the lighting subspace, and it only accounts for the low frequency parts of the lighting variance. The statistics of the error under a new lighting condition can be estimated from those of the error under known illumination, i.e. $\mu_e(\bar{\mathbf{s}}_k)$, $\sigma_e^2(\bar{\mathbf{s}}_k)$, also via the kernel regression method [20].

$$\mu_e(\mathbf{s}) = \frac{\sum_{k=1}^{M} w_k \mu_e(\bar{\mathbf{s}}_k)}{\sum_{k=1}^{M} w_k} \tag{13}$$

$$\sigma_e^2(\mathbf{s}) = \frac{\sum_{k=1}^{M} w_k \sigma_e^2(\bar{\mathbf{s}}_k)}{\sum_{k=1}^{M} w_k} \tag{14}$$

$$w_k = \exp(-\frac{[D(\mathbf{s}, \bar{\mathbf{s}}_k)]^2}{2[\sigma_{\bar{\mathbf{s}}_k}]^2}) \tag{15}$$

where $D(\mathbf{s}, \bar{\mathbf{s}}_k) = \|\mathbf{s} - \bar{\mathbf{s}}_2\|_2$ is the L_2 norm of the lighting coefficient distance. Like $\sigma_{I_k^p}$, $\sigma_{\bar{\mathbf{s}}_k}$ determines the weight of the error term related to the lighting coefficients $\bar{\mathbf{s}}_k$. Also, we assign the farthest lighting coefficient distance of these 20 images from the probe image to $\sigma_{\bar{\mathbf{s}}_k}$.

4.3 Recovering the Basis Images

Given the estimated lighting coefficients $\mathbf{s}$ and the corresponding error term $\mu_e(\mathbf{s})$, $\sigma_e^2(\mathbf{s})$, we can recover the basis images via the MAP estimation. If we apply the log probability, omit the constant term, and drop $\mathbf{s}$ for compactness, Eq.(9) can become

$$\arg\max_B \left(-\frac{1}{2}(\frac{I - B\mathbf{s} - \mu_e}{\sigma_e})^2 - \frac{1}{2}(B - \mu_B)C_B^{-1}(B - \mu_B)^T \right) \tag{16}$$

To solve Eq.(16), we estimate the derivatives,

$$-\frac{2}{\sigma_e^2}(I - B\mathbf{s} - \mu_e)\mathbf{s}^T + 2(B - \mu_B)C_B^{-1} = 0 \tag{17}$$

Then we rewrite Eq.(17) as a linear equation,

$$AB = b \tag{18}$$

where $A = \frac{\mathbf{s}\mathbf{s}^T}{\sigma_e^2} + C_B^{-1}$ and $b = \frac{I - \mu_e}{\sigma_e^2}\mathbf{s} + C_B^{-1}\mu_B$. The solution of the linear equation is $B = A^{-1}b$. Using the Woodbury's identity [25], we can obtain an explicit solution

$$B_{MAP} = A^{-1}b \tag{19}$$
$$= \left(C_B - \frac{C_B\mathbf{s}\mathbf{s}^T C_B}{\sigma_e^2 + \mathbf{s}^T C_B\mathbf{s}} \right) \left(\frac{I - \mu_e}{\sigma_e^2}\mathbf{s} + C_B^{-1}\mu_B \right)$$
$$= \left(\frac{I - \mu_B\mathbf{s} - \mu_e}{\sigma_e^2 + \mathbf{s}^T C_B\mathbf{s}} \right) C_B\mathbf{s} + \mu_B$$

From Eq.(19), the estimated basis image is composed of the term of characteristics, $\left(\frac{I - \mu_B\mathbf{s} - \mu_e}{\sigma_e^2 + \mathbf{s}^T C_B\mathbf{s}}C_B\mathbf{s} \right)$, and the term of mean, μ_B. In the term of characteristics, $(I - \mu_B\mathbf{s} - \mu_e)$ is the difference between the probe image and the image reconstructed by the mean basis images.

4.4 Recognition

The most direct way to perform recognition is to measure the distance between probe images and the subspace spanned by the recovered basis images. Every column of B is one basis image. However, the basis images are not orthonormal

vectors. Thus we perform the QR decomposition on B to obtain a set of orthonormal basis, i.e. the matrix Q. Then the projection of probe image I to the subspace spanned by B is $QQ^T I$, and the distance between the probe image I and the subspace spanned by B can be computed as $\|QQ^T I - I\|_2$. In the recognition procedure, the probe image is identified as the subspace with minimum distance from it.

5   Experiments

The statistical model is trained by images from the extended Yale Face Database B. With the trained statistical model, we can reconstruct the lighting subspace from only one gallery image. This estimation is insensitive to lighting variation. Thus, recognition can be achieved across illumination conditions.

5.1   Recovered Basis Images

To recover the basis images from a single image, the lighting coefficients of the image are estimated first. Then we estimate the error terms of the image Finally, the basis images of the image can be obtained with Eq.(19).
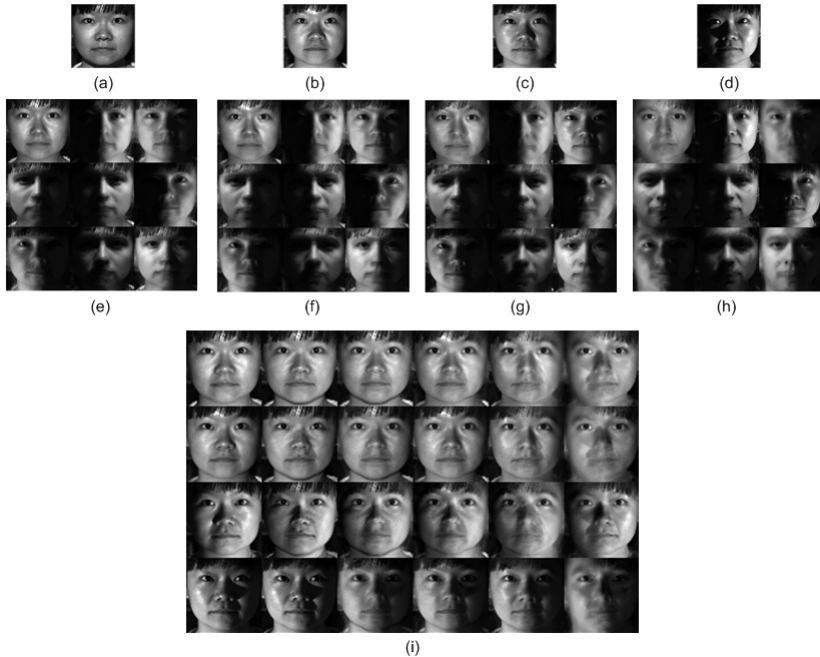
Fig. 2. Recovered basis images. (a)~(d) are images in subset 1~4 of Yale Face Database B respectively. (e)~(h) are recovered basis images from image (a)~(d) respectively. (i) are the reconstruction results: from left to right, the columns are the original images, the reconstruction results from the real basis images and the estimated basis images(e)~(h), respectively.

Although the images of the same object are under different lighting conditions, the recovered basis images should be similar. The probe images are from the Yale face database B. There are 10 subjects and 45 probe images per subject. According to the lighting conditions of the probe images, they can be grouped into 4 subsets as in [8]. The details can be found in Table 1. From subset1 to subset4, the lighting conditions become extreme. For every subject, we recover its basis images from only one of its probe images each time. Then we can obtain 45 sets of basis images for every subject. Fig.2(e)∼(h) are the basis images recovered from an image of each subset. $\bar{\sigma}_{basis}$(the mean standard deviation of the 45 sets of basis images of 10 subjects) is 7.76 intensity levels per pixel, while $\bar{\sigma}_{image}$(the mean standard deviation of the original 45 probe images of 10 subjects) is 44.12 intensity levels per pixel. From the results, we can see that the recovered basis images are insensitive to the variability of lighting. Thus we can recover the basis images of a subject from its images under arbitrary lighting conditions. Fig.2(i) are the reconstruction results from different basis images.The reconstructed images also contain shadows and inter-reflections because the recovered basis images contain detailed reflection information. As a result, good recognition results can be obtained.

Table 1. The subsets of Yale Face Database B

	subset1	subset2	subset3	subset4
illumination	0∼12	13∼25	26∼50	50∼77
Number of images	70	120	120	140

5.2   Recognition

Recognition is performed on the Yale Face Database B [8] first. We take the frontal images (pose 0) as the probe set, which is composed of 450 images (10 subjects, 45 images per subject). For every subject, one image is used for recovering its lighting subspace and the 44 remaining images are used for recognition. The comparison of our algorithm with the reported results is shown in Table 2.

Our algorithm reconstructed the 9PL subspace for every subject. The recovered basis images also contained complicated reflections on faces, such as cast shadows, specularities, and inter-reflection. Therefore the recovered 9PL subspace can give a more detailed and accurate description for images under different lighting conditions. As a result, we can get good recognition results on images with different lighting conditions. Also, the reported results of 'cone-cast', 'harmonic images-cast' and '9PL-real' showed that better results can be obtained when cast shadows were considered. Although [15,16] also use only one image to adjust lighting conditions, they need to recover the 3D model of the face first. The performance of our algorithm is comparable to that of these algorithms, which are based on high-resolution rendering [15,16] and better than that of those algorithms based on normal rendering [14]. Our algorithm is a completely 2D-based approach. Computationally, it is much less expensive compared with

Table 2. The Recognition Error Rate of Different Recognition Algorithms on Yale Face Database B

Algorithms	subset1&2	subset3	subset4
Correlation[8]	0.0	23.3	73.6
Eigenfaces[8]	0.0	25.8	75.7
Linear Subspace[8]	0.0	0.0	15.0
Cones-attached[8]	0.0	0.0	8.6
Cones-cast[8]	0.0	0.0	0.0
harmonic images-cast[8]	0.0	0.0	2.7
3D based SH model [12]	0.0	0.3	3.1
BIM(30 Bases)[15]	0.0	0.0	0.7
Wang *et al.*[16]	0.0	0.0	0.1
Chen *et al.*[17]	0.0	0.0	1.4
9PL-real[19]	0.0	0.0	0.0
our algorithm	0.0	0.0	0.72

those 3D based methods. The basis images of a subject can be directly computed with Eq.19 while the recognition results are comparable to those from the 3D-based methods.

5.3   Multiple Lighting Sources

An image taken under multiple lighting sources can be considered as images taken under a single lighting source being superimposed. Through interpolation, the lighting coefficients of images taken under single lighting are linearly combined to approximate those of the image taken under multiple-lighting. Here we also apply the statistical model trained on the extended Yale Database B to basis images estimation. Similarly the lighting coefficients of images are estimated through interpolation. Then the error term can be estimated according to the lighting coefficients. Finally, the basis images are recovered.

In the PIE face database [23], there are 23 images per subject taken under multiple lighting sources, and altogether 69 subjects. We recover 23 sets of the basis images from the 23 images of every subject respectively. With these estimated basis images, we perform recognition on the 1587 images (23 images per person) 23 times. We also estimate basis for images in the AR database [24]. We select randomly 4 images under different illumination per subject (image 1, 5, 6, 7) and recover the respective basis images from those images. Recognition is performed on 504 images (126 subjects and 4 images per subject) 4 times. Samples of the recovered basis images from images in the PIE and AR databases are shown in Fig.3. The average recognition rates, the mean standard deviation of the recovered basis images ($\bar{\sigma}_{basis}$) and the mean standard deviation of the gallery images ($\bar{\sigma}_{images}$) are presented in Table 3. Also [12] reported a recognition rate of 97.88% on part of PIE and [18] reported his recognition rate as 97% on PIE database. Our recognition results are better. The results show that the statistical model trained by images taken under a single lighting source can also be generalized to images taken under multiple lighting sources.

Table 3. Recognition Rate on Different Databases

Face Database	PIE	AR
$\bar{\sigma}_{basis}$	11.01	11.34
$\bar{\sigma}_{image}$	285	38.59
Recognition rate	98.21%	97.75%

(a)　　　　　(b)　　　　　(c)　　　　　(d)

(e)　　　　　(f)　　　　　(g)　　　　　(h)
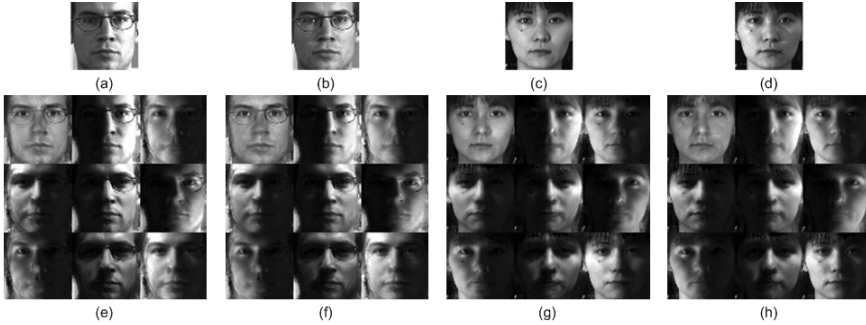
Fig. 3. Recovered basis images. (a) and (b) are images in PIE database, (e) and (f) are estimated basis images from image (a) and (b), respectively. (c) and (d)are images in AR database, (g) and (h) are estimated basis images from image (c) and (d), respectively.

6 Conclusion

The 9PL provides a subspace which is useful for recognition and is spanned by real images. Based on this framework, we built a statistical model for these basis images. With the MAP estimation, we can recover the basis images from one gallery image under arbitrary lighting conditions, which could be single lighting source or multiple lighting sources. The experimental results based on the recovered subspace are comparable to those from other algorithms that require lots of gallery images or the geometric information of the subjects. Even in extreme lighting conditions, the recovered subspace can still appropriately represent lighting variation. The recovered subspace retains the main characteristics of the 9PL subspace.

Based on our statistical model, we can build the lighting subspace of a subject from only one gallery image. It avoids the limitation of requiring tedious training or complex training data, such as many gallery images or the geometric information of the subject. After the model has been trained well, the computation for recovering the basis images is quite simple and without the need of 3D models. The proposed framework can also potentially be used to deal with pose and lighting variations together, with training images in different poses taken under different lighting for building the statistical model.

Acknowledgement

References

1. Guo, X., Leung, M.: Face recognition using line edge map. IEEE Trans. Pattern Recognition and Machine Intelligence 24(6), 764–799 (2002)
2. Shashua, A., Tammy, R.: The quotient images: class-based rendering and recognition with varying illuminations. IEEE Trans. Pattern Recognition and Machine Intelligence 23(2), 129–139 (2001)
3. Gross, R., Brajovic, V.: An image processing algorithm for illumination invariant face recognition. In: 4th International Conference on Audio and Video Based Biometric Person Authentication, pp. 10–18 (2003)
4. Wang, H., Li, S.Z., Wang, Y.: Generalized quotient image. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (2004)
5. Hallinan, P.: A low-dimensional representation of human faces for arbitrary lighting conditions. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 995–999 (1994)
6. Nayar, S., Murase, H.: Dimensionality of illumination in appearance matching. In: Proc. IEEE Conf. Robotics and Automation, pp. 1326–1332 (1996)
7. Belhumeur, P., Kriegman, D.J.: What is set of images of an object under all possible lighting conditions? In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 270–277 (1996)
8. Georghiads, A., Belhumeur, P., Kriegman, D.: From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Recognition and Machine Intelligence 23(6), 643–660 (2001)
9. Yuille, A., Snow, D., Epstein, R., Belhumeur, P.: Determing generative models of objects under varying illumination: shape and albedo from multiple images using SVD and integrability. International Journal of Computer Vision 35(3), 203–222 (1999)
10. Ramamoorthi, R., Hanrahan, P.: On the relationship between radiance and irradiance: determine the illumination from images of a convex Lambertian object. J. Optical. Soc. Am. A 18(10), 2448–2459 (2001)
11. Basri, R., Jacobs, D.: Lambertian reflectance and linear subspaces. IEEE Trans. Pattern Recognition and Machine Intelligence 25(2), 218–233 (2003)
12. Zhang, L., Samaras, D.: Face recognition under variable lighting using harmonic image exemplars. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (2003)
13. Wen, Z., Liu, Z., Huang, T.: Face relighting with radiance environment map. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (2003)
14. Zhang, L., Wang, S., Samaras, D.: Face synthesis and recognition from a single image under arbitrary unknown lighting using a spherical harmonic basis morphable model. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (2005)
15. Lee, J., Moghaddam, B., Pfister, H., Machiraju, R.: A bilinear illumination model for robust face recognition. In: Proc. IEEE International Conference on Computer Vision (2005)
16. Wang, Y., Liu, Z., Hua, G., et al.: Face re-lighting from a single image under harsh lighting conditions. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (2007)
17. Chen, H.F., Belhumeur, P.N., Jacobs, D.W.: In search of illumination invariants. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition (2000)
18. Zhou, S., Chellappa, R.: Illuminating light field: image-based face recognition across illuminations and poses. In: Proc. IEEE Intl. Conf. on Automatic Face and Gesture Recognition (May 2004)

19. Lee, K., Ho, J., Kriegman, D.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans. Pattern Recognition and Machine Intelligence 27(5), 684–698 (2005)
20. Sim, T., Kanade, T.: Combining models and exemplars for face recognition: an illumination example. In: Proc. Of Workshop on Models versus Exemplars in Computer Vision, CVPR 2001 (2001)
21. Adini, Y., Moses, Y., Ullman, S.: Face recognition: the problem of compensating for changes in illumination directions. IEEE Trans. Pattern Analysis and Machine Intelligence 19(7), 721–733 (1997)
22. Atkenson, C., Moore, A., Schaal, S.: Locally weighted learning. Artificial Intelligence Review (1996)
23. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression (pie) database. In: Proc. IEEE International Conference on Automatic Face and Gesture Recognition (May 2002)
24. Martinez, A.M., Benavente, R.: The AR face database. CVC Tech. Report No.24 (1998)
25. Scharf, L.: Statistical signal processing: detection, estimation and time series analysis, p. 54. Addison-Wesley, Reading (1991)