Acquiring Linear Subspaces for Face Recognition

under Variable Lighting

Kuang-Chih Lee, Jeffrey Ho, David Kriegman

Abstract

Previous work has demonstrated that the image variation of many objects (human faces in particular) under variable lighting can be effectively modeled by low dimensional linear spaces, even when there are multiple light sources and shadowing. Basis images spanning this space are usually obtained in one of three ways: A large set of images of the object under different lighting conditions is acquired, and principal component analysis (PCA) is used to estimate a subspace. Alternatively, synthetic images are rendered from a 3D model (perhaps reconstructed from images) under point sources, and again PCA is used to estimate a subspace. Finally, images rendered from a 3D model under diffuse lighting based on spherical harmonics are directly used as basis images. In this paper, we show how to arrange physical lighting so that the acquired images of each object can be directly used as the basis vectors of a low-dimensional linear space, and that this subspace is close to those acquired by the other methods. More specifically, there exist configurations of k point light source directions, with k typically ranging from 5 to 9, such that by taking k images of an object under these single sources, the resulting subspace is an effective representation for recognition under a wide range of lighting conditions. Since the subspace is generated directly from real images, potentially complex and/or brittle intermediate steps such as 3D reconstruction can be completely avoided; nor is it necessary to acquire large numbers of training images or to physically construct complex diffuse (harmonic) light fields. We validate the use of subspaces constructed in this fashion within the context of face recognition.

K-C. Lee and J. Ho are with the Beckman Institute and Computer Science Department, University of Illinois at Urbana-Champaign, Urbana, IL 61801. D. Kriegman is with the Computer Science and Engineering Department at the University of California at San Diego, La Jolla, CA 92093-0114. E-mail: klee10@uiuc.edu, j-ho1@uiuc.edu, kriegman@cs.ucsd.edu.

I. INTRODUCTION

To build a robust and efficient face recognition system, the problem of lighting variation is one of the main technical challenges facing system designers. In the past few years, many appearance-based methods have been proposed to handle this problem, and new theoretical insights as well as good recognition results have been reported [1], [2], [3], [5], [7], [9]. The main insight gained from these results is that there are both empirical and analytical justifications for using low dimensional linear subspaces to model image variations of human faces under different lighting conditions. Early work showed that the variability of images of a Lambertian surface in fixed pose, but under variable lighting where no surface point is shadowed, is a three-dimensional linear subspace [9], [12], [17]. What has been perhaps more surprising is that even with cast and attached shadows, the set of images is still well approximated by a relatively low dimensional subspace, albeit with a bit higher dimension [5].

Under the Lambertian assumption and accounting for shadowing and multiple light sources, the set of images of an object under all possible lighting conditions forms a polyhedral cone, the illumination cone, in the image space [3]. Therefore, the illumination cone contains all the image variations of an object in fixed pose, and an accurate representation of the cone would be a powerful tool for recognizing objects across a wide range of illumination variations. Indeed, successful work on applying this theory to face recognition has been reported, e.g. [7]. For most objects, the exact illumination cone is very difficult to compute due to the large number of extreme rays that make up their cones, e.g., for a convex, Lambertian surface, there are $O(n^2)$ extreme rays, where n is the number of pixels. This complicates both quantitative and qualitative studies of the illumination cone.

However, several recent results have indicated that, although it provides a theoretical basis for discussions on illumination problems, the computation of the full illumination cone may be unnecessary. Using spherical harmonics and techniques from signal-processing, Basri and Jacobs have shown that for a convex Lambertian surface, its illumination cone can be accurately approximated by a 9-dimensional linear subspace that they called the harmonic plane [2], [14], [15]. The major contribution of their work is to treat Lambertian

reflection as a convolution process between two spherical harmonics representing the lighting condition and the Lambertian kernel. By observing that the Lambertian kernel contains only low-frequency components, they deduce that the first nine (low frequency) spherical harmonics capture more than 99% of the reflected energy. Using this nine-dimensional harmonic plane, a straightforward face recognition scheme can be developed, and results obtained in [2] are excellent. Recently Ramamoorthi [13] developed a novel method based on spherical harmonics to analytically compute low-dimensional (less than nine dimensional) linear approximations to illumination cones. His results give a theoretical explanation to many empirical results obtained earlier, e.g. [5].

For face recognition, one way to interpret Basri and Jacobs' result is that for each of the more than six billion human faces in the world, there exist nine "universal virtual" lighting conditions such that the nine "harmonic images" taken for each individual under these conditions are sufficient to approximate his/her illumination cone with the harmonic subspace H spanned by these images. These nine "harmonic lights" are not real lighting conditions because for some directions, the intensity is negative as specified by the spherical harmonic functions. Similarly, the nine "harmonic images" (basis images) are not real images because some of the pixel values (image irradiance) are negative. Therefore, these images must be the result of some computation from real images or rendered from a geometric model of a head under synthetic harmonic lighting. This requires knowledge of the object's surface normals and albedos before the harmonic subspace can be computed. On the other hand, simple linear algebra tells us that any set of nine linearly independent vectors (or images) in H is sufficient to recover the plane. This hints at the possibility of an easier way to obtain the linear subspace: that is, can we find a set of nine real images such that the linear subspace spanned by them coincides with the harmonic subspace? For all practical purpose, the answer to this question is 'no'. Any real image in H requires the lighting over the sphere of directions to be a smooth function specified by a linear combination of the first nine spherical harmonics, and it would be very difficult to physically construct such lighting conditions in a common laboratory or application environment. However, one can ask a different but related question: is there another 9-dimensional linear subspace R that can also provide a

good representation for face recognition? Can R be constructed in some canonical fashion, perhaps with nine physically and easily realized lighting conditions?

In this paper, we formulate the problem as follows. We will consider only single distant and isotropic light sources. Each such light source can be associated with a point on the unit sphere $s \in S^2$ indicating its direction. Let Ω denote a subset of the unit sphere S^2. The problem we wish to solve is the following: given Ω and a small integer d (typically nine or less); find a subset $\{s_1, \cdots, s_d\}$ of Ω such that the d corresponding lighting directions $\{l_{s_1}, \cdots, l_{s_d}\}$ and the associated d-dimensional subspace R generated by d images taken under these lighting conditions are a good approximation to the illumination cones of a collection of human faces. For computational reason, the set Ω is always assumed to have finite size, and in this paper Ω is either a uniformly sampled sphere or a uniformly sampled hemisphere.

Since we know that the harmonic subspace H is a good representation for face recognition under variable lighting, it seems reasonable to find a plane R close to H. To make this notion precise, we need a notion of distance between two planes. In our first algorithm, the distance between two planes (not necessarily of the same dimension) is defined to be the square-sum of the cosines of the principal angles between them. R is then defined as the plane that has the smallest distance to H. From the recognition standpoint, it is also preferable to require that the intersection between R and the illumination cone C be as large as possible. This condition is incorporated into our second algorithm. That is, we want to find a k-dimensional linear subspace R, with k ranging from 1 to 9, generated by elements in Ω such that the distance between R and H is minimized (in some way) while the (unit)-volume $R \cap C$ is maximized. In Section 3, we formulate both algorithms in terms of maximizing an objective function defined over Ω. Our end result is a set of k directions (points) in Ω, and R is spanned by the images taken under the lighting conditions specified by these k directions.

It turns out that the resulting k light source directions are qualitatively very similar for different individuals. By averaging the objective functions for different individuals and maximizing this new objective function, we obtain a sequence of configurations of light source directions, called the universal configurations, such

that on average, the linear spaces spanned by the corresponding images are a good approximation to the corresponding illumination cones. We demonstrate that by using these universal configurations, good face recognition results can be obtained. In some cases, as few as five training images per person are sufficient to produce reasonably accurate face recognition results, if a small error rate can be tolerated.

The main contribution of this paper is the demonstration, both theoretically and empirically, that it is possible to employ just a few real images (taken under single distant and isotropic light sources) to model the various illumination effects on human faces, provided that the light source locations are chosen carefully. From a practical standpoint, acquiring images under a single distant and isotropic light source is much easier and less costly than alternatives. That is, the linear subspace R is lot easier to obtain than either the harmonic subspace H or the illumination cone C. This is particularly appropriate for acquiring training images of individuals in a controlled environment such as a driver's license office, a bank, or a security office.

This paper is organized as follows. In Section 2, we briefly summarize the idea of [2] using a harmonic subspace H for face recognition. The relationship between H and the illumination cone [3] is explained. Our algorithms for computing R and the universal configuration are detailed in Section 3, and Section 4 presents experimental results. The final section contains a brief summary and conclusion of this paper. Preliminary results on this topic were presented in [10], [11].

II. PRELIMINARIES

A. Illumination Cone

Let $\mathbf{x} \in \mathbb{R}^n$ denote an image with n pixels of a convex object with a Lambertian reflectance function illuminated by a single point source at infinity, represented by a vector $\mathbf{s} \in \mathbb{R}^3$ such that its magnitude $|\mathbf{s}|$ represents the intensity of the source and the unit normal $\mathbf{s}/|\mathbf{s}|$ represents the direction. Let $B \in \mathbb{R}^{n \times 3}$ be a matrix where each row $\mathbf{b}$ in B is the product of the albedo with unit normal for a point on the surface

projecting to a particular pixel in the image. Under the Lambertian assumption, $\mathbf{x}$ is given by

$$\mathbf{x} = \max(\mathrm{Bs}, 0) \tag{1}$$

where $\max(Bs, 0)$ sets to zero all negative components of the vector Bs. If the object is illuminated by k light sources at infinity, then the image is given by the superposition of the images that would have been produced by the individual light sources, i.e.,

$$\mathbf{x} = \sum_{i=1}^{k} \max(\mathrm{Bs_i}, 0). \tag{2}$$

Due to this superposition, the set of all possible images C of a convex Lambertian surface created by varying the direction and strength of an arbitrary number of point light sources at infinity is a convex cone. Furthermore, any image in the illumination cone C (including the boundary) can be determined as a convex combination of extreme rays (images) given by

$$\mathbf{x_{ij}} = \max(\mathrm{Bs_{ij}}, 0) \tag{3}$$

where $s_{ij} = \mathbf{b_i} \times \mathbf{b_j}$ are rows of B with $i \neq j$. It is clear that there are at most $m(m-1)$ extreme rays for $m \leq n$ distinct surface normals [7].

In computer vision, it has been a customary practice to treat the human face as a Lambertian surface. Although human faces are not convex, the degree of non-convexity is not serious enough to render the concept of the illumination cone inapplicable [7]. The only difference between the illumination cone of a human face and a convex object is that Equation 3 no longer accounts for all the extreme rays and there are extreme rays that are the result of cast shadows. Therefore, the formula for the upper bound on the number of extreme rays is generally more complicated than the quadratic expression $m(m-1)$ above. This poses a formidable difficulty for computing the exact illumination cone (i.e., specifying all the extreme rays). Instead, a subset

of the illumination cone can be computed by sampling lighting directions on the unit sphere, and Equation 1 is accompanied by ray tracing to account for the cast shadows.

B. *Lambertian Reflection and Spherical Harmonics*

In this section, we briefly summarize the recent work presented in [2], [14], [15], [19]. Consider a convex Lambertian object with uniform albedo illuminated by distant isotropic light sources, and p is a point on the surface of the object. Pick a local (x, y, z) coordinates system F_p centered at p such that the z-axis coincides with the surface normal at p, and let (θ, ϕ)[1]denote the spherical coordinates centered at p. Under the assumption of distant and isotropic light sources, the configuration of lights that illuminate the object can be expressed as a non-negative function $L(\theta, \phi)$. The reflected radiance at p is given by

$$
\begin{aligned}
r(p) & = \lambda \iint_S k(\theta) L(\theta, \phi) dA \\
& = \lambda \int_0^{2\pi} \int_0^{\pi} k(\theta) L(\theta, \phi) sin\theta d\theta d\phi
\end{aligned}
\tag{4}
$$

where λ is the albedo, and $k(\theta) = \max(\cos\theta, 0)$ is called the Lambertian kernel. A similar integral can be formed for any other point q on the surface to compute the reflected radiance $r(q)$. The only difference between the integrals at p and q is the lighting function L: at each point, L is expressed in a local coordinate system (or coordinate frame F_p) at that point. Therefore, considered as a function on the unit sphere, L_p and L_q differ by a rotation $g \in SO(3)$ that rotates the frame F_p to F_q. That is, $L_p(\theta, \phi) = L_q(g(\theta, \phi))$.

The spherical harmonics are a set of functions that form an orthonormal basis for the set of all square-integrable (L^2) functions defined on the unit sphere. They are the analogue on the sphere to the Fourier basis on the line or circle. The spherical harmonics, Y_{lm}, are indexed by two integers l and m obeying $l \geq 0$ and $-l \leq m \leq l$:

---

[1]To conform with the notation used in spherical harmonics literature, θ denotes the elevation angle and ϕ denotes the azimuth angle. In the next section, however, we will switch the roles of θ and ϕ.

$$Y_{lm}(\theta, \phi) = \begin{cases} N_{lm} P_l^{|m|}(cos\theta) cos(|m|\phi) & \text{if } m > 0; \\[2ex] N_{lm} P_l^{|m|}(cos\theta) & \text{if } m = 0; \\[2ex] N_{lm} P_l^{|m|}(cos\theta) sin(|m|\phi) & \text{if } m < 0; \end{cases} \tag{5}$$

where N_{lm} is a normalization factor guaranteeing that the integral of $Y_{lm} * Y_{l'm'} = \delta_{mm'}\delta_{ll'}$, and $P_l^{|m|}$ is the associated Legendre functions (its precise definition is not important here; however, see [20]). In particular, there are nine spherical harmonics with $l < 3$. One significant property of the spherical harmonics is that the polynomials with fixed l-degree form an irreducible representation of the symmetry group $SO(3)$, that is, a rotated harmonic is the linear superposition of spherical harmonics of same l-degree. For a 3D rotation $g \in SO(3)$:

$$Y_{lm}(g(\theta, \phi)) = \sum_{n=-1}^{l} g_{mn}^l Y_{ln}(\theta, \phi). \tag{6}$$

The coefficients g_{nm}^l are real numbers and are determined by g.

Expanding the Lambertian kernel $k(\theta)$ in terms of Y_{lm}, one has $k = \sum_{l=0}^{\infty} k_l Y_{l0}$. Because $k(\theta)$ has no ϕ-dependency, its expansion has no Y_{lm} components with $m \neq 0$. An analytic formula for k_l was given in [2], [15]. It can be shown that k_l vanishes for odd values of $l > 1$, and the even terms fall to zero rapidly; in addition, more that 99% of the L^2-energy of $k(\theta)$ is captured by its first three terms, those with $l < 3$. Because of these numerical properties of k_l, by Equation 4, any high-frequency ($l > 2$) component of the lighting function $L(\theta, \phi)$ will be severely attenuated. That is, the Lambertian kernel acts as a low-pass filter. Therefore, for a smooth lighting function L, the result of computing reflected radiance using Equation 4 can be accurately approximated by the same integral with L replaced by L', obtained by truncating the harmonic expansion of L at $l > 2$. Since rotations preserve the l-degree of the spherical harmonics (rf. Equation 6), the same truncated L' will work at every surface point.

C. Harmonic Images

From the above discussion, it follows that the set of all possible images of a convex Lambertian object under all lighting conditions can be well approximated by nine 'harmonic images', 'images' formed under lighting conditions specified by the first nine spherical harmonics. Except for the first spherical harmonic (which is a constant), all others have negative values and therefore, they do not correspond to real lighting conditions. Hence, the corresponding 'harmonic images' are not real images, and as pointed out by [2]: "they are abstractions." Knowing the object's geometry and albedos, these harmonic images can be synthesized using standard techniques, such as the ray-tracing.

For spherical harmonics, the spherical coordinates θ, ϕ are a little bit complicated to work with. Instead, it is usually convenient to write Y_{lm} as a function of x, y, z rather than angles. Each spherical harmonic $Y_{lm}(x, y, z)$ expressed in terms of (x, y, z) is a polynomial in (x, y, z) of degree l. The first nine spherical harmonics in Cartesian coordinates (with rounded constant coefficients) are

$$Y_{00} = 0.2821; \tag{7}$$

$$(Y_{11}; Y_{10}; Y_{1-1}) = 0.4886(x; z; y); \tag{8}$$

$$(Y_{21}; Y_{2-1}; Y_{2-2}) = 1.093(xz; yz; xy); \tag{9}$$

$$Y_{20} = 0.3154(3z^2 - 1); \tag{10}$$

$$Y_{22} = 0.5462(x^2 - y^2); \tag{11}$$

Figure 1 shows the rendered harmonic images for a face taken from the Yale Database. These synthetic images are rendered by sampling 1000 rays on a hemisphere, and the final images are the weighted sum of 1000 ray-traced images. Unlike [2] which only accounted for attached shadows, these harmonic images also include the effects of cast shadows arising from non-convex surfaces. Therefore, all nine harmonic images
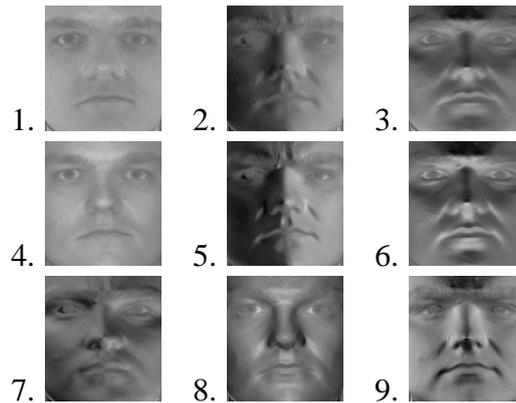
Fig. 1. The nine simulated harmonic images of a face from Yale Database. Light gray and dark gray respectively indicate the positive and negative pixel values. Since Y_{00} is a constant, the corresponding harmonic image simply scales the albedo values as shown in Picture 1. Pictures 4 is the harmonic image corresponding to $Y_{1-1} = z$, and has positive values for all pixels. Here, the image plane is defined as the xy-plane.
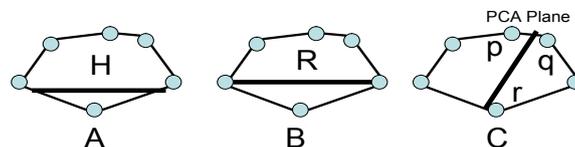

Fig. 2. An illustration of the cross section of an illumination cone C with the solid circles denoting the extreme rays. A) The intersection $C \cap H$ is shown as the dashed line. Notice that the intersection does not contain extreme rays, and H is parallel to the direction in which C is the thickest. B) A possible linear subspace passing through extreme rays that is good for face recognition. C) A PCA plane obtained by choosing a biased set of extreme rays p, q, and r as samples.

contain 3D information (i.e., the shadows) of the face. The values of the spherical harmonics at a particular point are computed easily using Equations 7–11.

D. Motivations

The main goal of this paper is to give a set of configurations of lighting directions such that the images taken under these lighting conditions can serve as a good linear basis for recognition. In the following paragraphs, we will explain, in terms of illumination cones and harmonic images, some of the heuristics that led us to believe the possibility of the existence of such configurations. The actual computational problems that produce the configurations will be described in the next section.

The good recognition results reported in [2] has indicated very clearly that the linear subspace H generated by the harmonic images is a good approximation to the illumination cone C [3]. Figure 2(A) gives a reasonable depiction of the relation between H and C. In particular, we can imagine geometrically that the

illumination cone is 'thick' in the directions parallel to H while it is 'thin' in directions perpendicular to H.

From its very definition, H can be considered as intrinsic to C, since both are completely determined by the object's shape and albedo as expressed through the B matrix. It is then natural to study the relation between H and C; in particular, how H intersects C and how the set $H \cap C$ is situated (or embedded) in C. If Figure 2(B) can be a guide, one interesting problem is to determine the set of extreme rays that are close to H as measured by the L^2 distance in the image space R^n, which have their linear span as close to H as possible, and whose intersection of their linear span and C is as large as possible. Any such set of extreme rays can replace the harmonic images and serve as a good basis for face recognition. Of course, for different persons there would be a different set of such extreme rays (images) because the associated illumination cones are different. That is, their locations in the image space are different. However, we can reasonably expect that the shape of the two illumination cones would be similar. With our common experience with human faces, it is reasonable to expect the following:

Heuristic: Let e_α^1, e_β^1 denote two images (extreme rays) of one face corresponding to the lighting directions l_α, l_β, and let e_α^2, e_β^2 correspond to two more images of another face under the same pair of lighting conditions. If the L^2-difference between e_α^1 and e_β^1 is small (large), then the L^2-difference between e_α^2 and e_β^2 should also be small (large).

That is, if two lighting conditions produce similar (dissimilar) images for one person, they will also produce similar (dissimilar) images for everyone else. This heuristic implies that if one illumination cone is "thick" in some directions, then for any other illumination cone, there will be "corresponding" directions in which it is "thick". At this point, it is natural to inquire on how such correspondences in directions can be realized. What perhaps has been neglected in the past is to regard an extreme ray both as an image *and* as a direction (the direction of the light source that generated it). With this understanding, it is straightforward to suspect that the lighting directions are responsible for this type of correspondences between extreme rays of different persons. That is, two extreme rays from different persons are considered to be in "correspondence" if they

are generated by the same lighting conditions (i.e., same directions).

Combining the above arguments, the following statement becomes plausible. Let $\{e_1, \cdots, e_k\}$ be a set of extreme rays for one face that is a good approximation of its illumination cone and harmonic subspace H and $\{l_1, \cdots, l_k\}$ are the corresponding lighting directions. For any other face, if $\{e'_1, \cdots, e'_k\}$ are extreme rays generated by $\{l_1, \cdots, l_k\}$, one should expect that $\{e'_1, \cdots, e'_k\}$ is a reasonable approximation of the illumination cone.

Of course, there are many ways to arrive at a k-dimensional linear subspace. The most common and straightforward way is to sample images in the cone and use the principal component analysis. However, principal component analysis depends heavily on the sample images used to define the correlation matrix, whose eigenvectors define the resulting PCA plane. A biased set of samples (e.g., a small number of samples) would produce a PCA plane that is not effective for face recognition, as illustrated in Figure 2(C).

III. Low Dimensional Linear Approximations of Illumination Cone

In this section, we detail our algorithm for computing R, a low dimensional linear approximation of an illumination cone. Given a model (human face), we assume that we have the detailed knowledge of its surface normals and albedos. Using the methods outlined in the previous section, we can easily render the model's harmonic images and construct the harmonic subspace H. Let C and EC denote the model's illumination cone and the set of normalized extreme (unit length) rays in the cone, respectively. For notational convenience, we will not make any distinction between an extreme ray (which is an image) and the direction of the corresponding light source; therefore, depending on the context, an element of EC can denote either an image or a light source direction.

For greater generality, we will slightly modify the formulation of our problem. Let Ω denote a finite subset of the unit sphere. In the following discussion, Ω will invariably denote a set of uniformly sampled points on the entire sphere or a set of uniformly sampled points on a hemisphere. Abusing the notation slightly, we will also call elements of Ω extreme rays. Following [7], the set Ω will be considered as a subset of EC, and all

computations pertained to the illumination cone will be carried out with the set Ω, instead of the complete set of extreme rays EC. This formulation allows a greater flexibility in applying our results to face recognition. For instance, if some prior lighting distribution is known (e.g., light sources are primarily frontal or lateral), Ω can represent a set of sampled points according to the distribution. With Ω now defined, the desired linear subspace R will be spanned by the extreme rays in Ω.

Let $GR(n, 9)$ denote the space of 9-dimensional linear subspaces of $\mathbb{R}^n$, the Grassmannian. Let $\mathbb{D}$ denote the subset of $GR(n, 9)$ consisting of 9-dimensional linear subspaces generated by the extreme rays in Ω. We define $\mathbb{D}$ to be our domain. Unlike $GR(n, 9)$, the space $\mathbb{D}$ is discrete and contains at most $\binom{e}{9}$ points, where e is the size of Ω. While the harmonic subspace H is generally in the set $GR(n, 9) \backslash \mathbb{D}$, the linear space R that we are after always resides in $\mathbb{D}$.

A. Computing Linear Subspace R

Since R is meant to provide a basis for a good face recognition method, we require R to satisfy the following condition:

Condition: The distance between R and H should be minimized.

Since we know that H is good for face recognition, it is reasonable to assume that any subspace close to H would likewise be good for recognition. What is needed now is an appropriate definition of the distance between two linear subspaces. One such notion of distance between two planes is the principal angles between them ([8], pp. 584-585). If A and B are matrices whose columns are orthonormal and span R and H, respectively, the cosines of the angles between the R and H are given by the singular values of $B^T A$. Let $\{\alpha_1, \cdots, \alpha_9\}$ be the singular values of $B^T A$, we can define the distance between R and H as

$$\mathrm{pDIST(H, R)} = 9 - \sum_1^9 \alpha_i^2 \tag{12}$$

Since $|\alpha_i| \leq 1$, we have $\text{pDIST}(H, R) = 0$ if and only if $R = H$, i.e., all $\alpha_i = 1$ or all angles between H and R are zero. Furthermore, $\text{pDIST}(H, L) \geq 0$ for any subspace L with dimension ≤ 9. The desired linear subspace R will be a global minimum of pDIST on $\mathbb{D}$. In the actual computation, Equation 12 is somewhat awkward to use, especially when the dimension of R is different from nine. Instead, we can try to maximize on $\mathbb{D}$ the square sum of the singular values:

$$\text{Sim}(H, R) = \sum_1^k \alpha_i^2 \tag{13}$$

with k the dimension of R. Clearly, this equation still makes sense when the dimension of R is no longer nine. In particular, $\text{Sim}(H, R) = k$ if and only if $R \subset H$.

A straightforward way to solve the problem is to evaluate Sim on the discrete set $\mathbb{D}$ and locate its maximum. Alas, although $\mathbb{D}$ is discrete, its size is prohibitively large. This prevents a direct solution to the problem and therefore, a local greedy algorithm to reach a reasonable approximation is needed. Instead, we compute R as a sequence of nested linear subspaces $R_0 \subseteq R_1 \subseteq \ldots \subseteq R_i \cdots \subseteq R_9 = R$ with $R_i, i > 0$ a linear subspace of dimension i and $R_0 \equiv \emptyset$ as follows. First, we let Ω_i denote the set obtained by deleting i extreme rays from Ω. It follows that $\Omega_0 = \Omega$. We will define R_i and Ω_i inductively. Assume that R_{i-1} and Ω_{i-1} have been computed. The sets Ω_i and R_i are defined as follows: let x_i denote the element in Ω_{i-1} such that

$$x_i = \arg \max_{x \in \Omega_{i-1}} \text{Sim}(x \oplus R_{i-1}, H). \tag{14}$$

R_i is defined as the space spanned by x_i, and R_{i-1}: $R_i \equiv x_i \oplus R_{i-1}$, and the set Ω_i is defined as $\Omega_{i-1} \backslash x_i$. The algorithm terminates after $R_9 \equiv R$ is computed.

As with most greedy algorithms, the iterative process produces incrementally a k-dimensional linear subspace for each $k = 1, \cdots 9$. Each R_k can be regarded as a k-dimensional subspace that is reasonably close to H and hence, a reasonable linear subspace for face recognition under variable lighting. In Section 4, we will

Fig. 3. Five of the ten uncropped faces in the Yale database. The results for each individual are listed in Table III-B.

study a family of such nested linear subspaces $R_0 \subseteq R_1 \subseteq \ldots \subseteq R_i \cdots \subseteq R_9 = R$ for their effectiveness in face recognition.

B. Preliminary Experiments

In this subsection, we report our first results with the algorithm outlined above. In this experiment, Ω is a set of 1005 uniformly sampled points on S^2. For each sampled direction (point), we produce the corresponding extreme ray by rendering an image under a single directional source emanating from this direction (with intensity set to 1). This set of 1005 sampled extreme rays is used to define the domain for the maximization procedure specified by Equation 14. We have implemented our algorithm for computing the linear subspace R using the Yale Face Database B. For ten individuals, the Yale database contains a 3D model (surface normals and albedos) and 45 images under different lighting conditions of each person. Since the face is assumed to be Lambertian, the 3D model from the Yale Face Database allows quick rendering of a required image.

For five people in the database shown in Figure 3, the results of computing the 9-dimensional linear subspace R are shown in Table III-B. Since all lights are sampled from the unit sphere S^2, naturally we use the usual spherical coordinates to denote the light positions. The coordinates frame used in the computation is defined such that the center of the face is located at the origin, and the nose is facing toward the positive z-axis. The x and y axes are parallel to the horizontal and vertical axes of the image plane, respectively. The spherical coordinates are expressed as the pair (ϕ, θ) (in degrees), where ϕ is the elevation angle (angle between the polar axis and the z-axis) with range $0 \leq \phi \leq 180$, and θ is the azimuth angle with range $-180 \leq \theta \leq 180$. In all subsequent experiments, all results are reported with respect to this coordinates frame. It is worthwhile to note that the set of nine extreme rays chosen by the algorithm has a particular type of configuration. First, the first two directions chosen are frontal directions (with small values of ϕ).

The first ray chosen, by definition, is always the image that is closest to H and in most cases, it is the direct

frontal light given by $\phi = 0$. Second, after the frontal images are chosen, the next five directions are from

the sides (with $\phi \approx 90$). By examining the θ values of these directions, we see that these directions spread

in a quasi-uniformly manner around the lateral rim. Third, the eighth direction is always from behind (with

$\phi > 90$). This accounts for all the light coming from the hemisphere that is behind the face. And finally, the

last chosen direction seems to be random. It is important to note that it is by no mean clear a priori that our

algorithm based on minimizing the distance to H will favor such type of configurations. Furthermore and

most importantly, the resulting configurations across all individuals are very similar.

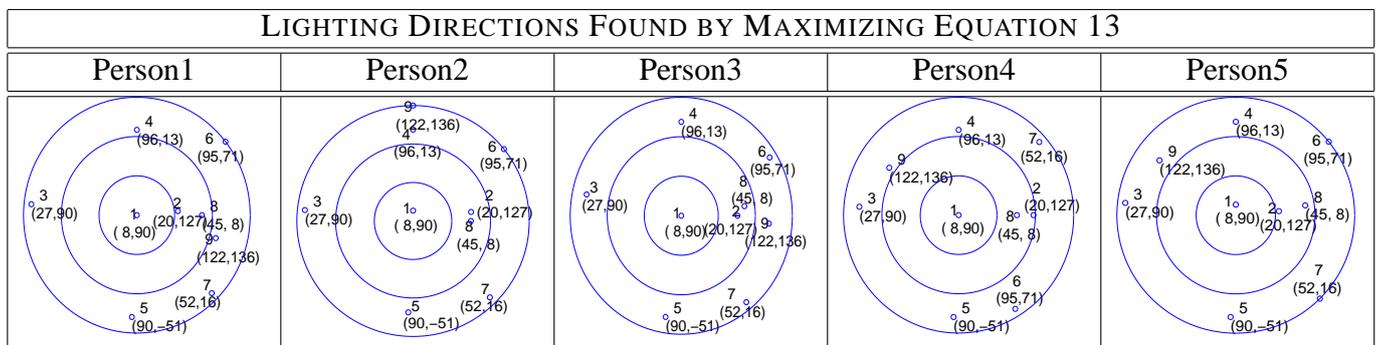| LIGHTING DIRECTIONS FOUND BY MAXIMIZING EQUATION 13				
Person1	Person2	Person3	Person4	Person5

Fig. 4. The nine lighting directions found by maximizing Equation 13 for five of the ten faces in Yale Face Database shown in Figure 3. The directions are represented in spherical coordinates (ϕ, θ) centered at the face. The first coordinate is the elevation angle with range $0 \le \phi \le 180$ and the second coordinate is the azimuth angle with range $-180 \le \theta \le 180$.

C. An Explicit Calculation

By maximizing (13) using a greedy algorithm, we have obtained a configuration of nine lighting directions

for each of the ten individuals in the Yale Face database. There are two prominent and distinctive patterns

emerged from the computations. First, the configurations are very similar (and in many cases, identical)

for different individuals. Second, the configurations are composed almost entirely of direct frontal lighting

directions and several lateral lighting directions. Because the size of the domain $\mathbb{D}$ is too large for a direct

maximization of (13), we have employed a straightforward greedy algorithm to reach the maxima; therefore,

there is a lingering doubt on whether the two prominent patterns we have observed is an artifact of the greedy

algorithm or that they are indeed the intrinsic properties of our solutions. In this subsection, we will maximize

Equation 13 directly without using a greedy algorithm. At the end, we will observe that the two prominent patterns still persist after direct computation, and the results will suggest strongly that the patterns are indeed intrinsic to our solutions.

To accomplish this, we have to drastically reduce the size of Ω from more than a thousand sample points down to only two dozens. We will compute a 5-dimensional linear subspace R generated by rays taken from two collections of sampled points on S^2, $\mathbb{U}$ and $\mathbb{U}'$. The first collection $\mathbb{U}$ contains 35 points. We place 4, 8, 10, 8 and 4 light source direction uniformly on the circles given by $\phi = 45°, 65°, 90°, 105°$ and $115°$, respectively. In addition, we place a light source at $\phi = 0°$, which is the direct frontal direction with respect to the face. We also define the smaller collection $\mathbb{U}'$ of 21 points, by placing 1 4, 6, and 10 points uniformly on the circles defined by $\phi = 0°, 45°, 90°$, and $125°$, respectively. Note that both collections contain lighting directions from behind the face (those with $\phi > 90°$) as well as frontal directions.

Our experiment is straightforward: for each collection, we enumerate all possible 5-dimensional linear planes generated by rays in the collection. For each ray in the collection, we render its corresponding image by using the normals and albedo values provided by the Yale Face Database B. We simplify the problem further, by assuming that R contains the frontal direction $\phi = 0°$. Therefore, there are $46,376$ and $4,845$ different planes for $\mathbb{U}$ and $\mathbb{U}'$, respectively. The final result is the plane that gives the largest $\mathrm{Sim}(R, H)$ value.

For $\mathbb{U}'$, we have rendered images of all ten persons in the Yale Face Database B. Out of $4,845$ different configurations, our algorithm consistently picks one particular configuration for all ten persons in the Yale Database. This configuration of five directions is (in spherical coordinates)

$$(\phi, \theta) : \{(0, 0), (90, -60), (90, -120), (90, 120), (90, 60)\}. \tag{15}$$

This configuration is symmetric with respect to the symmetric axis of the human face. It contains frontal, sides and top/down directions. For the much larger collection $\mathbb{U}$, we have also run the explicit algorithm on

Fig. 5. a. 21 images from the collection $\mathbb{IU}'$. b. The five images for the resulting configuration of five lighting directions $\{(0,0),(90,-60),(90,-120),(0,120),(0,60)\}$ obtained by maximizing Equation 13 explicitly.

the Yale Database, and the result is again quite consistent. Except in two of ten cases, our algorithm picks the same configuration for all individuals. See Figure 5.

In [9], Hallinan has shown empirically that there exists a reasonably good 5-dimensional approximation of the illumination cone C and a detailed characterization of these five basis images were given. These results are in good agreements with Hallinan's result [9]. However, our results follow directly from the computational problem defined in Section 3 while Hallinan's result is obtained empirically by acquiring a large number of images and applying PCA.

D. Maximizing The Intersection Volume

We have shown that by defining R as the linear subspace maximizing (13), for all individuals in the Yale Face Database, R can be formed by arranging single directional light sources in a special way. Although the algorithm employed a well-known concept of principal angles from linear algebra, there are two serious drawbacks. First, because of the singular value decomposition and Gramm-Schmidt process, the algorithm has a long computation time. For each individual, it takes approximately fifteen minutes to compute the nine

lighting directions on a 1GHz PC. Second, conceptually, the algorithm does not really tell us much about the geometric relation between R and the illumination cone. For instance, why there are lighting directions quasi-uniformly distributed along the lateral rim? Clearly, the answer to this question may depend on the geometric relation between R and the illumination cone. Since R is close to H, the geometric relation between R and the illumination cone can, in principle, be inferred from the relation between H and the illumination cone. However, this is clearly not very satisfactory and what we want is a formulation of the problem that takes into account the geometric relations between R and the entire illumination cone.

To this end, we require R to satisfy the following two conditions (C denotes the illumination cone).

Condition 1: The distance between R and H should be minimized.

Condition 2: The unit volume $vol(C \cap R)$ of $C \cap R$ should be maximized (The unit volume is defined as the volume of the intersection of $C \cap R$ with the unit ball).

Before turning these two seemingly innocuous conditions into a workable computational problem, we need to spell out precisely how the volume of $C \cap R$ can be defined and computed. First, note that $C \cap R$ is always a subcone of C; therefore, maximizing its unit volume is the same as maximizing the (solid) angle subtended by it. If $\{x_1, \cdots, x_k\}$ with $x_i \in \Omega$ is a basis of R, the cone $R_C \subset R$ generated by x_i:

$$R_C = \{x | x \in R, x = \sum_{i=1}^{k} \alpha_i x_i, \alpha_i \geq 0\} \tag{16}$$

is *always* a subset of $C \cap R$. If $C \cap R = R_C$, $vol(C \cap R)$ can be computed easily, e.g., taking the determinant of $\{x_1, \cdots, x_k\}$. The following proposition will show that, in practice, R_C is indeed a worthy substitute for $C \cap R$.

Proposition III.1: If $C \cap R \neq R_C$, there exists a linear relation among the elements of Ω containing some of the basis elements of R, $\{x_1, \cdots, x_k\}$.

Proof: If $C \cap R \neq R_C$, there must exist $x \in R$ such that $x = \sum_{i=1}^{k} \alpha_i x_i$ with some $\alpha_i < 0$ and $x_i \in R$. On the other hand, because $x \in C \cap R$, x can be written as $x = \sum_{i=1}^{l} \beta_i x_i'$ with all $\beta_i \geq 0$ and $x_i' \in \Omega$. This

gives the following:

$$\alpha_1 x_1 + \cdots + \alpha_k x_k = \beta_1 x'_1 + \cdots \beta_l x'_l$$

Since β_i are all non-negative and there exists at least one negative α_i, moving the left hand side to the right, we have a non-trivial linear relations among the elements of Ω containing at least one element of $\{x_1, \cdots, x_k\}$. ∎

Since $|\Omega| \approx 1000$ and the image space R^n typically has dimension greater than $30,000$ (using 168-by-192 images), we expect that *most* of the extreme rays in Ω are linearly independent. In fact, using an Ω with 500 elements, we have observed that the linear subspace spanned by its extreme rays has 497 dimensions. That is, there are only three linear relations among the elements of Ω; therefore, if k is sufficiently small (which is our case), we can avoid linear relations involving our basis elements. What we have discussed so far shows that, in practice, we can take $C \cap R$ to be R_C, and we will maximize the (solid) angle subtended by the cone R_C.

To compute R, we can formulate the following computational problem. First, we let Ω_i denote the set obtained by deleting i extreme rays from Ω. R_i and Ω_i will be defined inductively in terms of R_{i-1} and Ω_{i-1} as follows:

Let x_i denote the element in Ω_{i-1} such that

$$x_i = \arg \max_{x \in \Omega_{i-1}} \frac{\mathrm{dist}(x, R_{i-1})}{\mathrm{dist}(x, H)}. \tag{17}$$

where as before R_i is defined as the space spanned by x_i and R_{i-1}, and the set Ω_i is defined as $\Omega_{i-1} \backslash x_i$. When computing R_1, we define $\mathrm{dist}(x, R_0) = \mathrm{dist}(x, \emptyset)$ to be 1. Therefore, the first element x_1 is the extreme ray in Ω that is closest to the harmonic subspace H. The algorithm terminates after $R_9 \equiv R$ is computed. In the equation above, the distance function dist between a point x and H or R_i is defined as the L^2 distance between a point and a linear subspace. Notice that the distance function dist is different from the distance function pDIST or Sim used earlier. At each stage, maximizing the numerator $\mathrm{dist}(x, R_{i-1})$ gives us a linear

subspace $R_i \equiv x \oplus R_{i-1}$ that has a large solid angle. This is balanced by the denominator which ensures that R does not deviate too far away from H. In principle, $x \notin H$ and $\mathrm{dist}(x, H)$ is always non-zero for all $x \in \Omega_i$.

E. Discussion and Results

To satisfy Condition 1, it is tempting to find the nine extreme rays that are closest to H and define R as the linear space spanned by these rays. We have observed that these nine extreme rays are generally clustered around the direct frontal direction, and the resulting linear space R is a poor approximation of the illumination cone. The explanation, according to Condition Two, is because the resulting intersection $R \cap C$ has small volume. Geometrically, using nearby (with respect to H) rays is no guarantee that R will be a good approximation of H. Indeed, one can easily create a counter-example in three dimensions showing the peril of choosing nearby rays for this purpose, and the situation becomes trickier when R has a large co-dimension (which is our case).

On the other hand, the collection of images that are produced by extreme lighting conditions (lighting from the sides, up/down or behind) generally produce large intersection volume $vol(C \cap R)$, see Figure 6. Notice that these four images are mutually orthogonal in the sense that their mutual L^2-inner product is 0. This is because the sets of pixels illuminated in each images are mutually disjoint. Therefore, they will produce the maximal possible value of $\sqrt{2}$ for $\mathrm{dist}(x, R_{i-1})$, and the solid angle subtended by them will be maximal. However, the resulting intersection $R \cap C$ is only on the *boundary* of C and does not contain the interior of C. To correct these pathological cases, we need Condition 1 to "pull the plane inside." Heuristically, the first condition favors lighting directions that are nearly frontal while the second condition favors lateral lighting conditions. In this sense, the two conditions actually complement each other.

The results of running the second algorithm on a set of 1005 uniformly sampled points on S^2 is reported in Table 7. As neither a singular value decomposition nor a Gramm-Schmidt process are computed, the algorithm runs two to three times faster than the previous algorithm. The general characteristic of the config-

Fig. 6. If R is the plane generated by these four images, the intersection volume $vol(C \cap R)$ will be large according to our definition. Note that these four images are mutually orthogonal in the sense that their mutual L^2-inner product is 0. This is quite obvious since their non-shadow parts never intersect.

urations obtained this time is similar to the configurations we have obtained earlier: The first three directions are concentrated in the frontal area and the next four directions are spread quasi-uniformly to the lateral area. According to the discussion in the previous paragraphs, the quasi-uniform spread in the lateral area occurred in both sets of results can be attributed geometrically to the fact that the intersection $R \cap C$ tends to have a large volume.
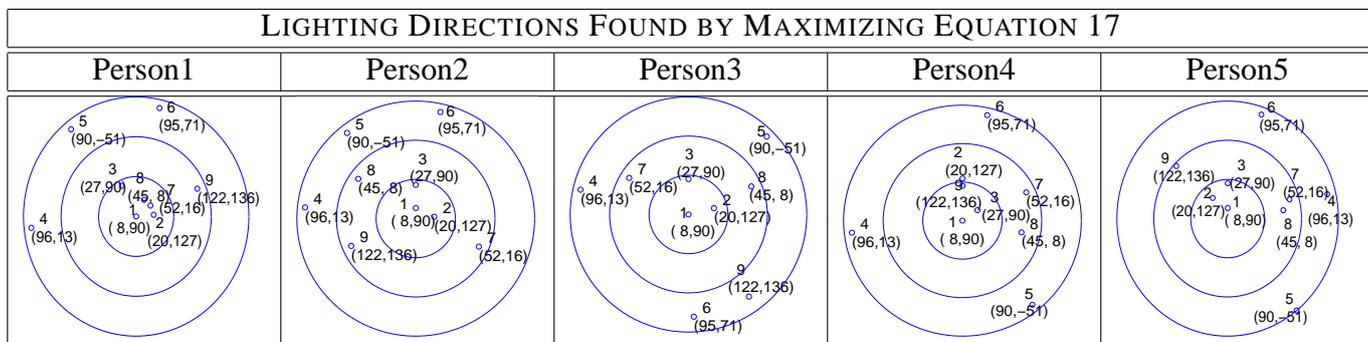
Fig. 7. The nine lighting directions found by maximizing Equation 17 for five of the ten faces in Yale Face Database shown in Figure 3

F. Computing a Universal Configuration

The results in the previous section demonstrate that, for each individual, there exists a configuration of nine lighting directions such that the linear subspace spanned by these images is a good linear approximation of the illumination cone. The configurations are qualitatively similar for different individuals with small variations in each lighting direction. It is then logical to seek a fixed configuration of nine lighting directions for all individuals such that for each individual, on average, the linear space spanned by the corresponding extreme rays is a good linear approximation to the illumination cone.

To find such a configuration, we can modify our previous method slightly by computing the average of the

quotient in Equation 17 over all the available training models. With all the notations defined as above, we find the nested linear subspaces $R_0 \subseteq R_1 \subseteq \ldots \subseteq R_i \cdots \subseteq R_9 = R$ by computing each x_i such that

$$x_i = \arg \max_{x \in \Omega_{i-1}} \sum_{k=1}^{l} \frac{\text{dist}(\text{x}^\text{k}, \text{R}^\text{k}_{\text{i}-1})}{\text{dist}(\text{x}^\text{k}, \text{H}^\text{k})}. \tag{18}$$

Since we are computing Equation 17 for all the available face models (indexed by k) simultaneously, for each $x \in \Omega$, x^k denotes the image of model k taken under a single light source with direction x. Ω_i denotes the set obtained by deleting i elements from Ω. k indexes the available face models. H^k denotes the harmonic subspace of model k, and R^k_{i-1} represents the linear subspace spanned by the images $\{x_1^k, \cdots, x_i^k\}$ of model k under light source directions $\{x_1, \cdots, x_i\}$.

For the application to face recognition experiments in the next section, we will let Ω denote a set of 200 uniformly sampled points on the 'frontal hemisphere'. Since all the test images, both Yale Face Database and the CMU PIE Database, were taken with lighting directions located on the frontal hemisphere, Ω here is appropriately taken to contain points sampled only from the frontal hemisphere.

We call the resulting configuration of nine directions the universal configuration. These directions are $\{ (0,0), (68,-90), (74,108), (80,52), (85,-42), (85,-137), (85,146), (85,-4), (51,67) \}$. They along with the 200 samples on the hemisphere are plotted in Figure 8. In the next section, this set of nested linear subspaces $R_0 \subseteq R_1 \subseteq \ldots \subseteq R_i \cdots \subseteq R_9 = R$ will be applied in face recognition experiments.

IV. EXPERIMENTS AND RESULTS

In the previous section, we have computed a configuration of lighting positions based on the idea that the linear subspace formed by the images taken under these lighting conditions should have a large intersection volume with the illumination cone. The subspace is also required to be close to the harmonic subspace that is known to model the illumination cone well. The main result is a nested sequence of linear subspaces, $R_0 \subseteq R_1 \subseteq \ldots \subseteq R_i \cdots \subseteq R_9 = R$, with basis images consisting of images taken under lighting conditions specified by these lighting directions. In this section, we report on the results from a series of comprehensive
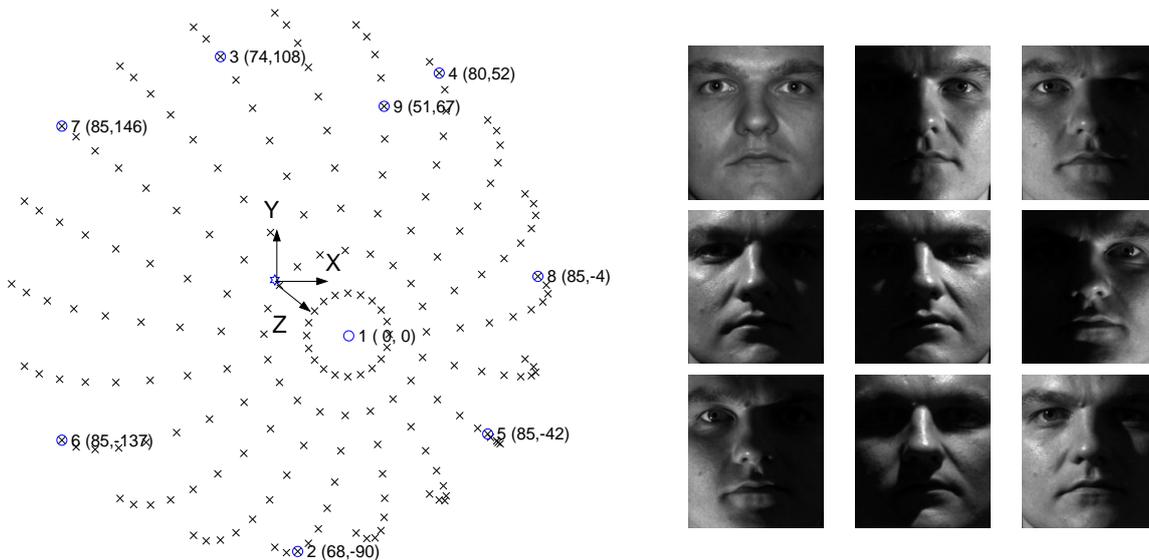
Fig. 8. **Left**: The universal configuration of nine light source directions with all 200 sample points plots on a hemisphere. **Right**: Nine images of a person illuminated by lights from the universal configuration.

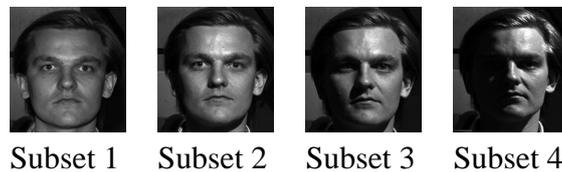

Subset 1   Subset 2   Subset 3   Subset 4

Fig. 9. Images of one of the ten individuals in the Yale database under the four subsets of lighting. See [7] for more examples.

experiments that validate our argument for favoring such configurations. First, we use the largest available subspace R_9 in the nested sequence above for face recognition as the choice of nine dimensions is largely motivated by the results in [2], [15]. Second, we use all the subspaces in the nested sequence for recognition. As shown below, these results demonstrate that subspaces with dimension greater than four all produce remarkably good recognition results. In the third subsection, we demonstrate experimentally that our lighting configuration is indeed special in the sense that it (almost) always provides a better face recognition performance compared with a randomly generated lighting configuration. In the last part of this section, we study the effectiveness of our subspaces in recognition experiments with diffuse illumination rather than just a single point source.

A. Recognition Experiments with Nine Points of Light

In this section, we apply the 9-dimensional subspace R_9 in a recognition experiment to see if the universal configuration of nine directions leads to effective face recognition compared to other published methods. For the experiments, we used images from the Yale Face Database B [7] that contains images of ten faces each under 45 different lighting conditions, and the test is performed on all of the 450 images. Following [7], the images are grouped into 4 subsets according to the lighting angle with respect to the camera axis. The first two subsets cover the angular range $0°$ to $25°$, the third subset covers $25°$ to $50°$, and the fourth subset covers $50°$ to $77°$.

Using the set of nine directions shown in Figure 8, we construct a linear subspace for each of the ten persons by taking the images of each person under these lighting conditions as the basis vectors of the linear subspace. In practice, the nine images should be real; however, due to the lack of real images acquired under these lighting directions, we offer two slightly different variations. In the first variation, which we call the Nine Points of Light (9PL) with simulated images, the required images are rendered using a geometric and albedo model from the Yale Database. In the second variation (9PL with real images), the required basis images are taken directly from the database: basis images are images whose lighting conditions are closest to the lighting conditions specified by our configuration. The recognition results of using our configuration of nine lighting directions together with recent illumination-insensitive recognition algorithms, such as Harmonic Images [2], Gradient Angle [4], and other methods reported previously in [7] are shown in Table I, ordered by decreasing overall error rate.

The correlation method, the Eigenface methods, the linear subspace method, and the cones methods were all trained using images from Subsets 1 and 2. The correlation method and Eigenface method are widely used face recognition techniques, and they serve as a baseline. Eigenfaces where the first three principal components are dropped is commonly used to remove the effects of lighting. The linear subspace method and illumination cones methods attempt to model the set of images of an object under differing lighting conditions. In the linear subspace method, the set is treated as a 3-D subspace while the Cones-attached

method is based on constructing an illumination cone that only accounts for attached shadows whereas Cones-cast accounts for cast shadows. Note that the recognition rate is perfect on this dataset for Cones cast. Using a 9-D subspace defined by rendered Harmonic images shows a 2.7% error rate for extreme lighting, and note that the 9PL method has a similar error rate (2.8%) when rendered images are used; hence, it is likely that the 9-D Harmonic planes and the 9-D subspaces produced using 9PL for each individual are indeed very close to each other. The gradient angle method introduced in [4] uses only a single training image, and compares images using a measure which is insensitive to lighting variation. Finally, when the subspace for the nine point algorithm is based on real images, there are again no errors.

All of the other methods reported in the table (except the Nearest Neighbor method) require considerable amounts of off-line processing on the training data, such as 3D reconstruction or eigen decomposition of the training data. For the Nine Points of Light method, no training is involved! The work is almost minimal as only nine images are needed.

At this point, it is reasonable to ponder whether it is important to represent an individual by the 9-D subspace spanned by the nine images or would it be sufficient simply to use the nine images as training images along with a straightforward classifier such as nearest neighbor (or perhaps something a bit more sophisticated such as Eigenfaces or a feed forward neural network). To answer this, we evaluated the recognition performance using the nearest neighbor classifier with the same nine normalized images as training data, and the results are also shown in Table I. Since most of the training samples are from Subset 4, nearest neighbor does reasonably well for Subset 4 with an error rate of 7.0%. However, unlike our method that measures the distance to the subspace, the nearest neighbor classifier does not generalize well to Subsets 1, 2, and 3.

B. Recognition with Lower Dimensional Subspaces

As shown in [3], the actual dimension of an illumination cone is the number of distinct surface normals. Hence, for human faces, the actual dimension of the illumination cone is quite large; nevertheless, the previous results show that the illumination cone for a human face (under a fixed pose) admits a good approximation

TABLE I

THE ERROR RATES FOR VARIOUS RECOGNITION METHODS ON SUBSETS OF THE YALE FACE DATABASE B. SOME OF THE ENTRIES (INDICATED BY CITATION) WERE TAKEN FROM PUBLISHED PAPERS WHEREAS THE 9PL, HARMONIC IMAGES, AND NEAREST NEIGHBOR RESULTS ARE FROM OUR OWN IMPLEMENTATION.

COMPARISON OF RECOGNITION METHODS				
Method	Error Rate (%) vs. Illum.			
	Subset 1&2	Subset 3	Subset 4	Total
Correlation [7]	0.0	23.3	73.6	29.1
Eigenfaces [7]	0.0	25.8	75.7	30.4
Eigenfaces w/o 1st 3 [7]	0.0	19.2	66.4	25.8
Nearest Neighbor (9 images)	13.8	54.6	7.0	22.6
Linear subspace [7]	0.0	0.0	15.0	4.6
Cones-attached [7]	0.0	0.0	8.6	2.7
9PL (simulated images)	0.0	0.0	2.8	0.87
Harmonic Images	0.0	0.0	2.7	0.85
Gradient Angle [4]	0.0	0.0	1.4	0.44
Cones-cast [7]	0.0	0.0	0.0	0.0
9PL (real images)	0.0	0.0	0.0	0.0

by a 9-dimensional linear plane in the image space. The natural extension of this conclusion is to further reduce the dimension of the linear approximation and observe the resulting error rates.

We experimented with this type of dimensionality reduction by successively using each linear subspace in the nested sequence, $R_0 \subseteq R_1 \subseteq \ldots \subseteq R_i \cdots \subseteq R_9 = R$, for face recognition. For this experiment, we used an extended database with $1,710$ images of 38 individuals from the Yale Face Database B and C. As there are no recognition results reported in the literature for other methods using this extended database, we only report on our own method. The results are shown in Figure 10, and it is clear that the recognition rate is still reasonably good even for five dimensions.

As alluded to earlier, these results corroborate well with the much earlier results of [5], [9]. They have shown that using 5 ± 2 eigenimages is sufficient to provide a good representation of the images of a human face under variable lighting. The main distinctions between these earlier results and ours are 1) the linear approximations provided by the earlier work have always been characterized in terms of eigenimages. In contrast, our linear approximations are characterized by real images. 2) There is no report of recognition
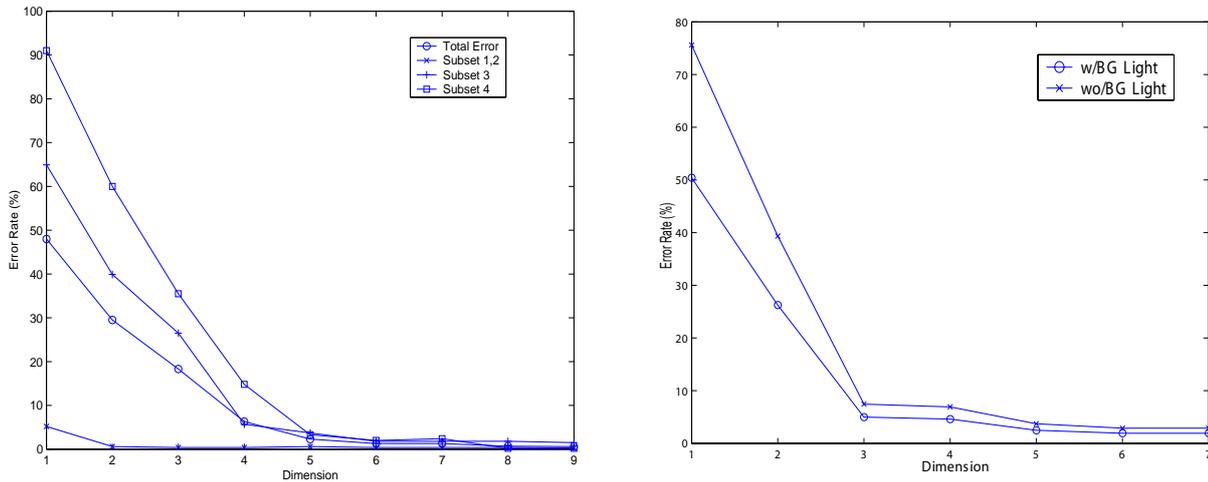
Fig. 10. The error rates for face recognition using successively smaller linear subspaces. The abscissa represents the dimension of the linear subspace while the ordinate gives the error rate. **Left**: In this experiment, the extended Yale Face Database, containing $1,710$ images of 38 individuals, was used. **Right**: In this experiment, the CMU PIE database, containing 1587 images of 69 individuals, was used. The database contains two sets of images. The first set of images were taken under single directional light sources without any background illumination. The recognition results for this set of images are marked with crosses. The other set of images were taken under same set of single directional light sources as the first set but with background illumination. The results for this set of images are marked with circles. Note that we have used a 7-dimensional subspace instead of the usual 9-dimensional subspace. The smallest error rates obtained by using 7-dimensional subspaces for recognitions with and without background illumination are 2.8% and 1.9%, respectively.

results in these earlier work while we have demonstrated that not only a good low-dimensional linear approximation of the illumination cone is possible but it also provides reasonably good face recognition results.

The experimental results reported so far have all used images in the Yale Face Database [7]. This database was designed primarily for studying illumination effects on face recognition. A more recent database designed for similar purposes is the PIE database from CMU (See [18]). We have tested our recognition algorithm on the PIE database, and the results are shown in Figure 10. For the illumination component of the PIE database, there are $1,587$ images of 69 individuals and 23 different illumination conditions. Due to the lack of surface normal and albedos estimates, we cannot render the images under the nine lighting directions specified in the previous section. Instead, we have decided to use some of the images provided by the database as the training images. For each lighting direction specified by our results, we choose the image in the PIE database with the closest lighting direction as the corresponding training image. Since the lighting directions used in acquiring the PIE database do not cover the sphere, two (out of nine) directions do not have an appropriately near-by direction in the PIE database. Therefore, we selected only seven images for each individual from the database as the basis of R. This leaves 16 test images for each individual.

C. Recognition Experiments with Randomly Generated Lighting Configurations

The experiments reported in the preceding subsections have demonstrated the effectiveness of the universal lighting configuration for face recognition. The universal lighting configuration shown in Fig. 8 is obtained by iteratively maximizing an objective function (rf. Equation 18). Although the implication of the optimization problem has been elucidated in the previous section, it is still natural to wonder whether there are other lighting configurations which are capable of generating the same results. For example, can a randomly generated lighting configuration have the same performance in face recognition as our universal configuration? We answer this question experimentally by randomly generating a large number of different lighting configurations and comparing the face recognition performance with the face recognition results reported earlier using the Universal Configuration.

In this experiment, we use the extended Yale Face Database ($1,710$ images for 38 people). In this Database, each individual has his/her images taken under 45 different lighting directions. We randomly generate configurations of five lighting directions among these 45 different lighting directions and the corresponding five images are taken to form the basis vector of a subspace. Therefore, for each randomly generated lighting configuration, there are five images for training and 40 images for testing. We randomly generate $16,000$ different configurations of five lighting positions and this number corresponds to roughly 1.5% of the total number of configurations $\binom{45}{5} = 1,221,759$.

Using our configuration of five lights (R_5), the recognition result is remarkable: with an error rate of 0.2%, only four images out of the total of $1,710$ images are incorrectly recognized. The mean error rate for randomly generated source directions is just under 10% (165 images) with the median of about 4%. Therefore, by randomly picking lighting conditions to form subspaces, we expect the error rate to be an order of magnitude larger than using our configuration. Most impressive is the fact that there are only three lighting configurations (out of the total $16,000$ tested configurations) that perform better than our configuration. Furthermore, these three configurations all share the same basic pattern with our configuration in the spatial distribution of their lighting directions. That is, a frontal lighting direction coupled with four (near) lateral directions.

D. Recognition with Ambient Lighting

So far, the empirical study of illumination effects on face recognition has focused on test images taken under single distant point light sources [7]. It has been conjectured earlier that with significant ambient lighting or multiple distant point sources, recognition should be simpler [3]. Using the illumination cone, it is straightforward to explain this conjecture. The images formed with ambient lighting or multiple distant point sources are located in the interior of the illumination cones. On the other hand, images formed by single direct distant sources belong to the boundary of the illumination cone. Therefore, the latter type of images are harder to recognize than the former type. In this subsection, we verify this conjecture experimentally using both the CMU PIE database and the Yale database.

The CMU PIE database contains two similar sets of images with exactly the same set of single distant lighting directions. The difference is that one set contains ambient illumination while the other does not. With ambient lighting, images generally do not contain hard (cast and attached) shadows caused by the geometry of the face; instead, soft shadows are present. We tested our method (with R of dimension seven) on the set of images with ambient lighting exactly as in the previous section, and the results are shown in Figure 10. Note the recognition results from testing images with ambient lighting are consistently better than the results without ambient lighting. In particular, using only one image (the image under the frontal lighting) as the training image, the error rate is about 50%. With 69 individuals in the database, a random pick will have an error rate of 98.5%.

To further verify this conjecture, we will again turn to simulated images. The basic idea of the next set of experiments is simple. We start with a single distant light source and try to simulate the effect of successively turning on more light sources. The main question that needs to be resolved by the experiments is whether adding more light sources indeed makes the recognition task easier.

More precisely, for each experiment we will define a sequence of twelve lighting conditions, $L = \{L^1, \cdots, L^{12}\}$. The first lighting condition, L^1, consists of only one single light source. For $1 \leq k \leq 11$, an additional source l_k will be added to the sequence successively.

$$L^{k+1} = L^k \cup \{l_k\} \tag{19}$$

Since the Yale Face Database is used in the experiment, the light sources l_k will all be taken from the light sources present in the database. For each individual and each lighting condition L^k, we simulate the image taken under L^k by taking a suitable linear combinations of images in the database. More precisely, if the lighting condition L^k consists of k different single distant direct light sources, $\{l_1, \cdots, l_k\}$, for each individual, we simulate the image taken under L^k by taking the average of the corresponding images (real, not rendered) in the Yale Face database:

$$I_{L^k} = \frac{I_{l_1} \cdots + I_{l_k}}{k}, \tag{20}$$

where I_{l_i} is the image in the database taken under the light source direction l_i. These simulated image will constitute the test images for the experiment. As before, we will use the nested sequence of linear subspaces $R_0 \subseteq R_1 \subseteq \ldots \subseteq R_i \cdots \subseteq R_9 = R$ in the recognition experiments.

In this experiment, 50 randomly generated sequences of L were selected, and Figure 11 shows one example of a sequence of seven images of an individual from the Yale Face Database under a randomly generated sequence of lighting conditions L. Please notice the softening of the shadows as the sequence progresses.

The average recognition results are shown in Figure 12. These results support the conjecture stated earlier. That is, adding lights that can remove shadows does decrease the error rate and makes recognition task easier. We can see that the error rate goes below 10% when enough lights are turned on, even though the method was trained with a single frontaly illuminated image. The results show that the ambient lighting conditions from those randomly generated testing sequences make the recognition problem particularly easy.
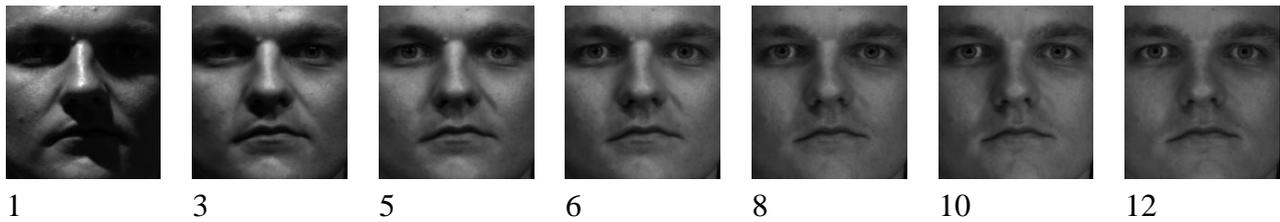
Fig. 11. An example of seven images from a twelve image sequence Ln creating by turning on succesively more distant light sources. The numeral indicates the number of distant sources when synthesizing each image. The gradual disappearance of shadows is noticeable, particularly in the areas around the eye sockets and the lower cheek.
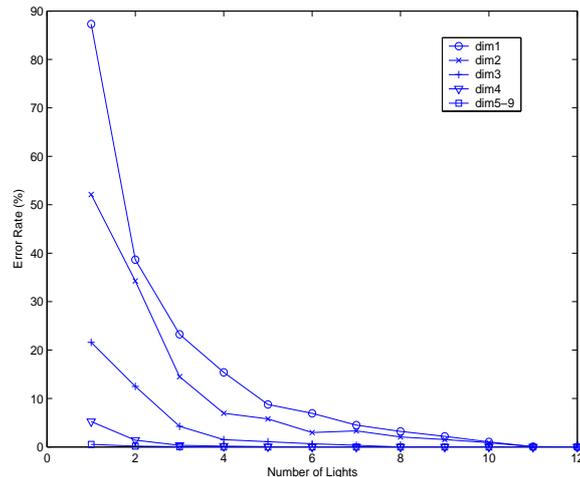

Fig. 12. The error rates for face recognition under the sequence of lighting conditions that resolves soft shadows at the end. The abscissa represents the number of single distance light sources used. The figure includes all the results using all nine linear subspaces. From dimension 5 to 9 of those linear subspaces, the error rates are almost zero.

V. Conclusions

We have shown that there exists configurations of single light source directions that are effective for face recognition. Depending on the difficulty of the lighting condition (e.g., strong or weak ambient lighting) and the degree of accuracy required, the number of single light source directions in the configuration can range from five to nine. The linear subspace spanned by the corresponding images is a good approximation to the illumination cone, and it provides good face recognition results under a wide range of difficult lighting conditions. We obtain the set by maximizing a function defined on the set of extreme rays of the illumination cone. Our result provides a recipe for building a simple but robust face recognition system. By taking several images of each individual with single light sources emanating from these directions, our results show that these nine images are already sufficient for the task of recognizing faces under different illumination condi-

tions. The usual complicated intermediate steps, such as the 3D reconstruction, can be completely avoided. Recently, Schechner et al. [16] pointed out that taking a set of images under multiplexed illumination rather than by a collection of single point light sources, the signal-to-noise ratio (SNR) will be reduced. Without noise, the resulting images acquired under multiplex illumination would span the exact same subspace as with just the nine single source images. But presumably, the lower SNR would lead to lower error rates.

One surprising conclusion of our work is that for modeling the *effect of illumination* on human faces, linear superpositions of a few directional sources are likely to be sufficient and effective. The basis images invariably contains one or two frontally-lit images and four to five laterally-lit images. Furthermore, as few as one or two frontally-lit images may already be sufficient as the training images if the lighting conditions are known to contain strong ambient components. This hints at an interesting venue for future research. Suppose some prior knowledge on the lighting distribution is known. An interesting problem is how to design a recognition algorithm that requires the minimal number of training images and under what conditions should these be acquired.

REFERENCES

[1] Y. Adini, Y. Moses, and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction. In *IEEE Trans. PAMI*, volume 19, pages 721–732, July 1997.

[2] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. In *Int. Conf. on Computer Vision*, volume 2, pages 383–390, 2001.

[3] P. Belhumeur and D. Kriegman. What is the set of images of an object under all possible lighting conditions. In *Int. J. Computer Vision*, volume 28, pages 245–260, 1998.

[4] H. Chen, P. Belhumeur, and D. Jacobs. In search of illumination invariants. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 1–8, 2000.

[5] R. Epstein, P. Hallinan, and A. Yuille. 5+/-2 eigenimages suffice: An empirical investigation of low-dimensional lighting models. In *PBMCV*, 1995.

[6] A. Georghiades, D. Kriegman., and P. Belhumeur. Illumination cones for recognition under variable lighting: Faces. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, 1998.

[7] A. Georghiades, D. Kriegman, and P. Belhumeur. From few to many: Generative models for recognition under variable pose and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40:643–660, 2001.

[8] G. Golub and C. van Loan. *Matrix Computation*. The John Hopkins Univ. Press, 1989.

[9] P. Hallinan. A low-dimensional representation of human faces for arbitrary lighting conditions. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 995–999, 1994.

[10] J. Ho, K. Lee, and D. Kriegman. On reducing the complexity of illumination cones for face recognition. In *IEEE CVPR Workshop on Identifying Objects Across Variations in Lighting: Psychophysics and Computation*, pages 56–63, 2001.

[11] K. Lee, J. Ho, and D. Kriegman. Nine points of lights: Acquiring subspaces for face recognition under variable lighting. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 519–526, 2001.

[12] S. Nayar and H. Murase. Dimensionality of illumination in appearance matching. *IEEE Conf. on Robotics and Automation*, 1996.

[13] R. Ramamoorthi. Analytic PCA construction for theoretical analysis of lighting variability in images of a lambertian object. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, October 2002.

[14] R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment. In *Proceedings of SIGGRAPH*, pages 497–500, 2001.

[15] R. Ramamoorthi and P. Hanrahan. A signal-processing framework for inverse rendering. In *Proceedings of SIGGRAPH*, pages 117–228, 2001.

[16] Y. Y. Schechner, S. K. Nayar, and P. N. Belhumeur. A theory of multiplexed illumination. In *Int. Conf. on Computer Vision*, 2003.

[17] A. Shashua. On photometric issues in 3D visual recognition form a single image. *Int. J. Computer Vision*, 21:99–122, 1997.

[18] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression (pie) database. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pages 53–58, 2002.

[19] P. Sloan, J. Kautz, and J. Snyder. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. In *Proceedings of SIGGRAPH*, pages 527–536, 2002.

[20] W. Strauss. *Partial Differential Equations*. John Wiley & Sons, Inc, 1992.

[21] M. Turk and A. Pentland. Eigenfaces for recognition. *J. of Cognitive Neuroscience*, 3(1):71–96, 1991.

[22] W. Y. Zhao and R. Chellappa. Symmetric shape-from-shading using self-ratio image. *Int. J. Computer Vision*, 45(1):55–75, 2001.