A Practical Face Relighting Method for Directional Lighting Normalization

Kuang-Chih Lee[1] and Baback Moghaddam[2]

[1] University of Illinois at Urbana-Champaign,
Urbana, IL 61801, USA
[2] Mitsubishi Electric Research Laboratories
Cambridge MA 02139, USA

Abstract. We propose a simplified and practical computational technique for estimating directional lighting in uncalibrated images of faces in frontal pose. We show that this inverse problem can be solved using constrained least-squares and class-specific priors on shape and reflectance. For simplicity, the principal illuminant is modeled as a mixture of Lambertian and ambient components. By using a generic 3D face shape and an average 2D albedo we can efficiently compute the directional lighting with surprising accuracy (in real-time and with or without shadows). We then use our lighting direction estimate in a forward rendering step to "relight" arbitrarily-lit input faces to a canonical (diffuse) form as needed for illumination-invariant face verification. Experimental results with the Yale Face Database B as well as real access-control datasets illustrate the advantages over existing pre-processing techniques such as a linear ramp (facet) model commonly used for lighting normalization.

1 Introduction

In computer vision and specifically in face recognition, robust invariance to arbitrary illumination has presented a difficult challenge. Indeed, in large independent US government tests of the leading algorithms (*e.g.,* in FERET [10] and FRVT [9]) improper handling of variable (outdoor) lighting has been a key limiting factor in achieving recognition rates obtained in more controlled laboratory conditions.

In this paper, we address a difficult but routine problem in facial identification as applied to access control and forensic surveillance: we are given a photograph of a possibly unknown individual, appearing in fixed pose (*e.g.,* frontal) which was taken by an uncalibrated camera with unknown intrinsic parameters in an arbitrary scene with unknown and variable illumination. Without any 3D measurement (of subject or environment) and with only the single image provided, we are to match identities by comparing the facial image to that of another in a large database of (frontal) faces in *fixed* lighting (*e.g.,* diffuse or frontal). Equivalently, we must standardize all current and any future images in a growing database in order to simulate a common fixed illumination template suitable for robust pattern matching and illumination-invariant facial identification. Naturally, the canonical choice of illumination would consist of non-directional or diffuse (or at least frontal) lighting that would maximize visibility of all facial features.

Since our focus is on illumination-invariance, we acknowledge that all images have undergone *geometric* normalization prior to analysis: beginning with face detection (*e.g.,* [6]), feature detection (of eyes) followed by rigid transforms (scale, rotation and translation) to align all detected features. In addition, we assume that some form of *photometric* normalization may have already taken place in the form of a non-spatial *global* transform which is a function of intensity only (*e.g.,* gain, contrast, brightness). We also prepare to encounter any number of possible data sources: live video capture, archival photography, web imagery, family photo albums, passport and ID pictures, *etc.*

2 Background

Much of the research on illumination has focused on finding a compact low-dimensional subspace to capture lighting variations. Belhumeur & Kriegman [2] proved that under the Lambertian assumption, the image set of an object under all possible lighting conditions forms a polyhedral "illumination cone" in the image space. Georghiades *et al.* [5] demonstrated applications of this framework to face recognition under variable illumination. Ramamoorthi [11] presented a method to analytically determine the low-dimensional illumination subspace obtained with PCA. Basri & Jacobs [1] represent lighting using a spherical harmonic basis wherein a low-dimensional linear subspace is shown to be quite effective for recognition. Zhang & Samaras [16] have extended this framework with spherical harmonic *exemplars*. Lee *et al.* [7] have empirically found how to arrange physical lighting to best generate an equivalent basis for recognition.

A complementary approach is to generate a lighting invariant signature image. Although this technique cannot deal with large illumination changes, it does have the advantage that only one image per object is required in the gallery. Some of the earlier normalization techniques apply such an approach to face recognition, using histogram equalization or linear ramp subtraction to generate invariant templates [13]. Chen *et al.* [3] demonstrated that the image gradient direction is mostly illumination-insensitive and can be used in a probabilistic framework to determine the likelihood of two images coming from the same object. Zhao & Chellappa [17] took advantage of the near symmetry of faces to compute an illumination invariant prototype image for each individual without recovering albedos. Shashua & Riklin-Raviv [14] assumed that different faces have a common shape but different texture and computed an albedo ratio as an illumination-invariant signature.

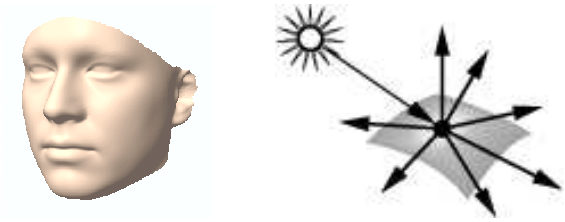
In computer graphics, object relighting has received much attention in recent years. An interesting application by Nishino & Nayar [8] is the use of corneal imaging for embedding realistic virtual objects (faces) into a scene, resulting in synthetic faces that are properly lit ("relit") in accordance with the estimated environmental lighting. Another example is the radiance environment map technique by Wen *et al.* [15] which renders relatively high-quality relighting of faces using the spherical harmonics approach [12].

For our face verification application however, there is really no need for high-quality graphics rendering or photo-realism. In fact, most 2D face recognition systems in existence today operate at low to moderate resolutions ($\approx$ 100 pixels across the face). Our relighting method can be categorized as an invariant template approach to illumination-invariance as discussed above (although we will also present an equivalent subspace formulation). As with [14] we assume a common underlying 3D shape for all individuals and utilize the albedo or diffuse reflectance (skin texture) as the main source of identity information. This, of course, is in keeping with the fact that most 2D face recognition systems do not measure 3D shape anyway.

Despite our simplifications, we will demonstrate that with low-resolution shape, approximate albedo and simple diffuse reflectance for relighting, it is possible to significantly improve the accuracy of face verification under moderate lighting variations encountered in real access-control operational scenarios. At the very least, it is our hope that others will find this algorithm to be a practical and superior alternative for lighting normalization.

3 Lighting Estimation

We use a Lambertian or "diffuse reflectance" (constant BRDF) illumination model for the face, as shown in Figure 1, despite the fact that sometimes there is some specular reflection (due to secretion of *sebum* oil by sebaceous glands in the skin). Nevertheless, this specular component is not always consistent and therefore of little use in a biometric analysis. Hence, our illumination model consists only of Lambertian and ambient components.



(a) (b)

Fig. 1. (a) 3D face shape with (b) a Lambertian reflectance model.

Specifically, let $I(x, y)$ be the intensity at pixel (x, y) corresponding to a point on the surface of a convex object (face or 3D face model) with Lambertian surface reflectance, which is illuminated by a mixture of ambient light and a (single) point light source at infinity $\mathbf{s} \in \mathbb{R}^3$ with intensity $|\mathbf{s}|$. We designate the unit surface normal $\mathbf{s}/|\mathbf{s}|$ as the direction *to* the light source (i.e., pointing out). This direction (*e.g.* in azimuth/elevation angles) is our main estimand of interest. The magnitude of the light source, on the other hand, is of little consequence in our analysis since it can be absorbed by the imaging system parameters modeling gain and exposure. We define $\rho(x, y)$ as the face albedo (or diffuse skin texture) and let $\mathbf{n}(x, y)$ be the unit surface normal of the point on the facial surface that projects onto the pixel $I(x, y)$ in the image (under orthography). Under the simple Lambertian (constant BRDF) model, a pixel's (monochrome) intensity is given by

$$I(x, y) = \alpha \ \{\rho(\mathrm{x}, \mathrm{y})[\max(\mathbf{n}(\mathrm{x}, \mathrm{y})^{\mathrm{T}}\mathbf{s}, 0) + \mathrm{c}]\} \ + \ \beta \tag{1}$$

where α and β represent intrinsic camera system parameters such as lens aperture and gain setting. In our analysis, (α, β) are essentially nuisance parameters which only affect the dynamic range (gain) and offset (exposure bias) of pixel intensity but *not* the lighting direction. Therefore, we can always set (α, β) to their default values of $(1, 0)$ with proper normalization. The parameter c represents the relative strength of the ambient lighting and we will show how it can be estimated in Section 4. The term $\max(\mathbf{n}^T\mathbf{s}, 0)$ above resets negative values of the (Lambertian) cosine factor to zero for surface points that are in shadow.

For simplicity's sake, we are assuming that a single (principal) light source alone is responsible for the majority of the observed directional lighting in the image (diffuse attenuation and/or shadowing) and that any other light sources present in the scene (diffuse or directional) are *non-dominant*, hence their overall contribution can be represented by a global ambient component with relative intensity c in Eq. (1). Nearly all 2D (view-based) face recognition systems are adversely affected by *directional* lighting, but to a much lesser extent by more subtle lighting effects [10]. Therefore, in most cases and for most algorithms the principal directional component is a more critical factor than any other lighting phenomena, especially when the other light sources are non-dominant. Therefore, accounting for this principal illumination factor by effectively "undoing" its effects can enhance verification performance.

Estimation of the principal lighting direction can be carried out with a least-squares formulation with the right simplifying assumptions, especially given the relatively simple illumination model in Eq. (1). More importantly, We can solve this problem rather efficiently (in closed form) with elementary matrix operations and dot-products. Specifically, let $\vec{I}$ be the column vector of pixel intensities obtained by stacking all the nonzero values of $I(x, y)$ and similarly define $\vec{\rho}$ to be the corresponding vectorized albedo map (diffuse texture). [3] We then form a 3-column shape matrix $\mathbf{N}$ by row-wise stacking of the corresponding surface normals. We then form the so-called *shape-albedo* matrix $\mathbf{A} \in \mathbb{R}^{\mathrm{p}\times 3}$ where each row $\mathbf{a}$ in $\mathbf{A}$ is the product of the albedo and the unit surface normal in the corresponding row of $\vec{\rho}$ and $\mathbf{N}$. Mathematically, this corresponds to the Hadamard (elementwise) matrix product $\circ$ as $\mathbf{A} = (\vec{\rho}\, \mathbf{1}_{1\times 3}) \circ \mathbf{N}$.

To solve for the unknown light source we use a matrix equation for least-squares minimization of the approximation error in Eq. (1) in the new vectorized form

$$\arg\min_{\mathbf{s}} \ \| \vec{I} - \alpha c\vec{\rho} - \mathbf{A}\mathbf{s} \|^2 \tag{2}$$

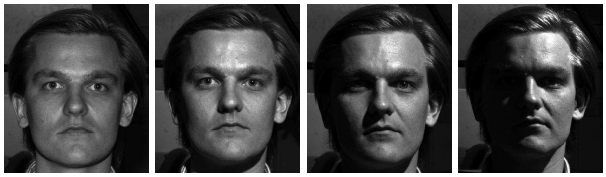
which yields the solution

$$\mathbf{s}^* = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T(\vec{I} - \alpha c\vec{\rho} - \beta) \tag{3}$$

Note that we are only interested in the unit light source vector $\mathbf{s}^*/|\mathbf{s}^*|$ for its direction and not the magnitude (which depends on the specific camera gain/exposure). Moreover, this estimation problem is well-behaved since it is heavily over-constrained: the number of nonzero elements in $\vec{I}$ ("observations") is on the order of $O(10^3)$ as compared to the 3 unknowns in $\mathbf{s}^*$ (in fact, since we only use the *direction* there are only 2 angular estimands: azimuth & elevation). Estimatation of the principal lighting direction is therefore quite stable with respect to noise and small variations in the input $\vec{I}$. Note that the derived matrix $\mathbf{A}$ comes from a *generic* shape and albedo and hence represents the entire frontal face object class. Assuming that it is adequately representative, there is no need to measure the exact shape (or even exact albedo) of an individual as long as all shapes (and albedos) are roughly equal to first order (*i.e.,* as far as lighting direction is concerned).

Furthermore, The pseudo-inverse $(\mathbf{A}^T\mathbf{A})^{-1}$ in Eq. (3) is directly proportional to the error covariance of the least-squares estimate $\mathbf{s}^*$ under Gaussian noise. If we further define the $\mathrm{p} \times 3$ matrix $\mathbf{P} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}$ we see that the only *on-line* computation in Eq. (3) is the projection of the input vector $\vec{I}$ on the 3 columns of $\mathbf{P}$ which are linearly independent. In fact, they are basic functions for the *illumination subspace* of our generic face (frontal face class). Moreover, we can always find an equivalent orthogonal basis for this subspace using a QR-factorization: $\mathbf{P} = \mathbf{Q}\mathbf{R}$, where the unitary matrix $\mathbf{Q}$ has 3 orthonormal columns spanning the same subspace as $\mathbf{P}$ and the 3×3 upper triangular matrix $\mathbf{R}$ now

[3] Without ambient light, zero-valued pixels are most likely in shadow and thus informative only if we use ray-casting to pinpoint the source. In practice, ambient light is always present and we use a nonzero threshold or a pre-set mask for pixel selection.



Subset A Subset B Subset C Subset D

Fig. 2. Images of an individual in the Yale Database B under the 4 directional lighting subsets. See [5] for more examples.

defines the quality of the estimates since $\mathbf{R}^{-1}$ is a Cholesky factor (matrix square root) of the error covariance. The QR factorization aids the interpretation and analysis of the estimation in terms of pixels and bases since the input image is directly projected onto the orthonormal basis $\mathbf{Q}$ to estimate the lighting direction (the QR decomposition also saves computation in larger problems). In Section 5 we will show an example of this factorization.

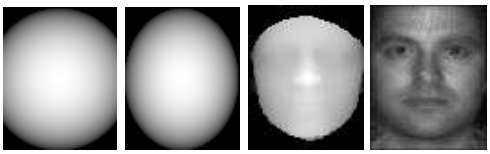
Since $\mathbf{P}$ and $\mathbf{Q}$ are independent of the input data they can be pre-computed once off-line and then stored. Also, the computational cost of using Eq. (3) is quite minimal (requiring only 3 image-sized dot-products) and since the subsequent relighting (see Section 4) is even less expensive (requiring a single dot-product), the lighting normalization process is very practical for real-time implementation (perhaps as part of a much larger face processing system).

3.1 Estimation Analysis

To evaluate our estimation technique we chose the Yale Face Database B which contains images of 10 individuals imaged under 45 different lighting conditions (our tests were performed on all 450 images). Following the protocol established in [5], the images were grouped into 4 subsets according to lighting direction angle with respect to the camera. The first two subsets cover the range $[0°, 25°]$, the third subset covers $[25°, 50°]$ and the fourth covers $[50°, 77°]$. Figure 2 shows sample images from the four different subsets in the Yale Face Database B.

These 450 images were manually cropped, geometrically aligned as best as possible and down-sampled to size 80×64 and then masked (but no photometric processing was performed). The choice of resolution was partly due to an operational specification (of a pre-existing face verification system) and also due to the challenge of working with a low resolution nearer to the limit at which surface normals can be reliably computed from a (potentially) noisy range map.

For our generic shape model we used the average 3D shape of 138 individuals in the DARPA HumanID database of 3D Cyberware scans [4]. The resulting average 3D face shape, seen in Figure 1(a), was first down-sampled and converted to an 80×64 range map (frontal depth) which was then smoothed using a Gaussian blur in order to reduce any quantization noise in its surface normals. Then, our generic face shape's surface normal map $\mathbf{n}(x, y)$ was computed and registered with the image plane (using the fixed eye positions only) thereby



Sphere Ellipse Face Albedo

Fig. 3. Range maps for 3 progressively more complex shapes: a sphere, an ellipse and a generic face (note its smoothness). The effective albedo image is shown on the right.

Table 1. Mean errors (in degrees) for the underlying shapes used in estimating lighting directions on subsets of the Yale database.

Light Source Estimation with Frontal Faces using Different Shapes								
Surface Geometry	Mean Errors (in azimuth and elevation)							
	subset **AB**		subset **C**		subset **D**		**ABCD** combined	
	AZ	EL	AZ	EL	AZ	EL	AZ	EL
Sphere	6.4	7.8	10.3	17.2	5.3	24.1	7.1	15.4
Ellipse	1.3	7.5	2.2	17.1	11.2	23.1	4.6	14.9
Face	2.3	4.9	3.0	6.7	5.5	9.2	4.5	5.5

aligning our standardized 2D input image $I(x, y)$ format with the surface normals of our 3D model. As a rough 2nd-order approximation to the generic face shape, we also computed range maps for a comparably-sized sphere and ellipse as shown in Figure 3.

For the average albedo we found it sufficient to use the facial *texture* obtained by averaging images of all 10 individuals in the Yale database with near-frontal lighting (subset A+B). By averaging (integrating) over all illuminations we (approximately) simulate *diffuse* or non-directional lighting. This average texture map was then processed (smoothed, de-noised, *etc.*) and normalized to the range $[0, 1]$ to serve as our generic face albedo. We should note that the albedo plays a secondary role (in estimating the illuminant) as compared to the geometry encoded in the surface normals. In fact, without going into the details, we mention that we were able to obtain almost as good estimates even with *constant* albedo. This is indicative of the overall stability (if not perfect accuracy) of this estimation problem. Figure 3 shows the generic face albedo we used for the Yale dataset.

In Table 1 we have summarized the mean estimation errors (in azimuth elevation degrees) for the 3 shapes in Figure 3 (sphere, ellipse, face). All 3 shapes were used to estimate lighting direction of all 450 images in the Yale Face Database B. Each error value reported in the table is the average difference between azimuth and elevation of the ground truth source and the computed lighting directions obtained from $\mathbf{s}^*$ in Eq. (3) with $c = 0$ as there is no ambient component in this dataset. From the totals column in the table (far right) the mean errors over *all* lighting directions are about 5 deg with the generic face and an average of about 10 deg (for both azimuth and elevation) with the sphere and ellipse. For

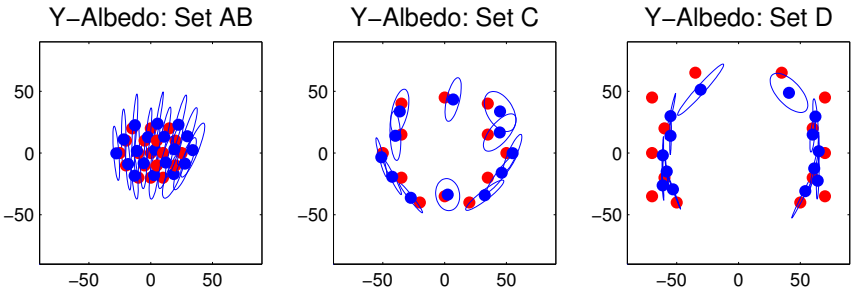


Fig. 4. Yale Database B lighting angle estimation plotted in degrees (elevation *vs.* azimuth). Red dots represent ground truth lighting angles. Blue dots are the mean estimates based on the 10 individuals. Blue dots are also the centers of the error covariances shown as blue ellipses (95% confidence interval). Results are grouped into 3 plots $\{\mathrm{AB}, \mathrm{C}, \mathrm{D}\}$ based on the 4 subsets shown in Figure 2.

near-frontal lighting (subset A+B) the difference between the two simple shapes and the face is relatively small in terms of the mean error.

For a more revealing analysis of the face-based results above we must look a bit more closely at the individual estimates. Figure 4 plots the distribution of all lighting estimates for all 10 individuals in the database. By comparing the mean estimates (blue dots) with the ground truth (red dots) we see a high degree of correlation between the two. Furthermore, there is as a surprising degree of accuracy considering all the convenient half-true assumptions made in formulating the solution in Eq. (3). Clearly, the blue error covariance ellipses confirm variation with between individuals as expected (due to differing geometry/albedo) but these estimates are certainly "correct" in the trend, especially with respect to the azimuth. It is only at the most extreme angles (subset D) that we see the estimates "saturate" near a limiting angle of 50 deg (which is most likely due to a lack of resolution in depth and the spatial boundary of the mask).

In Figure 4 we also see, based on the orientation of the error covariances, that the *error* estimate (uncertainty) is much greater along elevation than it is along azimuth. This raises an interesting question: what is limiting the precision of elevation estimates? Since it cannot be a mathematical flaw in our derivation, it must be either the albedo or the shape that we are using to compute the matrix $\mathbf{A}$ in Eq. (3). To help explain this phenomenon, in Figure 5 we show two separate histograms for the azimuth and elevation of the surface normals in our generic face shape. Note that the variation in azimuth of the surface normals (± 50 deg) is greater than the variation in elevation (± 20 deg). This limited "bandwidth" in the elevation distribution reduces the effective precision of the estimates of elevation (*i.e.,* there are far fewer surface normals with high elevation angles). This is partly due to an intrinsic feature of the shape of the human face, which to a first-order approximation is an upright cylinder or ellipse (both of which span a limited range of elevation) and partly due to the limitation of the smoothed low-resolution face shape we are using (see subset D in particular in Figure 4).

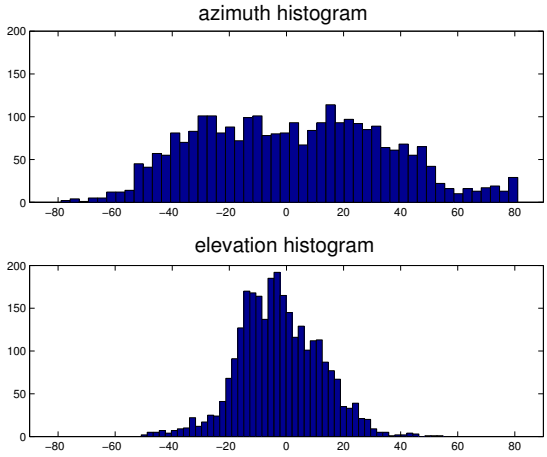


Fig. 5. The histograms of the azimuth and elevation angles of the surface normals of the average (generic) face shape.

4 Face Relighting

Given an estimate of the input image's directional lighting we can approximately "undo it" by estimating the individual's albedo (diffuse skin texture) and then relight this specific albedo (combined with a generic shape) under any desired illumination (*e.g.,* frontal or pure diffuse).

Whereas both generic shape and albedo were required in the inverse problem of estimating directional lighting, only generic shape is needed in the forward problem of relighting (as the input itself provides albedo information). Clearly, the basic assumption here is that all individuals have the same 3D geometry (that of our average shape). However, we find that moderate violations of this basic assumption are not highly critical to the verification performance since what is actually relighted to generate an invariant template is the facial texture of the individual herself and this texture carries most of the identity information for 2D face recognition. In fact, it is not possible to drastically alter the input image's albedo (skin texture) by using a (slightly) different 3D face shape. Therefore, despite the variations in geometry every individual's identity is mostly preserved as long as their face texture is retained.

Referring back to Eq. (1), once we have a lighting estimate $\mathbf{s}^*$ and our "plug-in" shape (surface normals of the average face) we can directly solve for albedo using

$$\rho^* = \frac{I - \beta}{\alpha(\mathbf{n}^T\mathbf{s}^* + c)}, \quad I \neq 0,\ \mathbf{n}^T\mathbf{s}^* \geq 0. \tag{4}$$

where from hereon we have suppressed the spatial indices (x, y) for all 2D-arrays (I, ρ and $\mathbf{n}$) for the sake of clarity. Notice that the estimated albedo ρ^* at a point (x, y) depends only on the corresponding pixel intensity $I(x, y)$ and the surface normal $\mathbf{n}(x, y)$. Thus, if a scene point is in shadow and there is no ambient

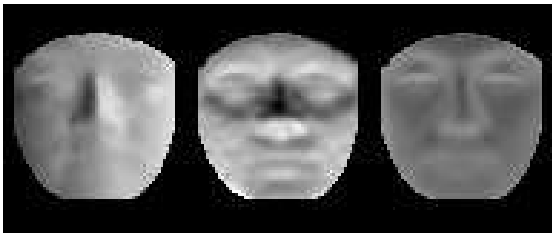


Fig. 6. The orthonormal basis $\mathbf{Q}$ in the QR-factorization of $\mathbf{P} = \mathbf{A}(\mathbf{A^T A})^{-1}$ in Eq. (3) for internal datasets I & II.

illumination ($c = 0$), I will be zero and $\mathbf{n}^T\mathbf{s}^*$ is negative. If so, the corresponding albedo cannot be estimated with Eq. (4) and a default (average) albedo must be substituted in for that pixel.

The estimated albedo is then relighted in order to generate our invariant (fixed-illumination) template I_o

$$I_o \;=\; \alpha_o \,\{\rho^*[\max(\mathbf{n}^T\mathbf{s}_o, 0) + c_o]\} + \beta_o \tag{5}$$

where $\mathbf{s}_o$ denotes the desired template's illumination (defaulted to on-axis frontal lighting) and c_o is the output ambient component. Similarly α_o and β_o designate the output (display/storage) image format parameters.

5 Verification Experiments

To evaluate the performance of our lighting normalization technique in an actual face verification test we used an internal face database belonging to our institution (had we chosen the same Yale Database B which has only 10 individuals, our performance results would have been statistically insignificant). Our data comes from an unconstrained and realistic access-control operational scenario inside a large industrial plant. The data was captured on two different days and in two separate locations in the facility with no specific lighting control. Approximately 500 individuals were automatically detected from a surveillance video using an enhanced face/feature-detector based on [6] and then geometrically aligned and cropped to our 80×64 image format and stored on disk (no photometric pre-processing was performed).

We put together and processed (by averaging and filtering) a suitably smooth and diffuse average face albedo (texture) for this dataset which is shown in Figure 9(b). Then we computed the corresponding shape-albedo matrix $\mathbf{A}$ in Eq. (3) using the same average 3D shape used with the Yale database in Section 3. Figure 6 shows the orthogonal illumination subspace of this database's generic face. Notice how both the horizontal and vertical lighting directions are decoupled into separate bases. The third basis image is the global intensity or amplitude (recall that we are interested in the two degrees of freedom in azimuth and elevation). The mottled appearance in the 2nd basis image is mostly due to the

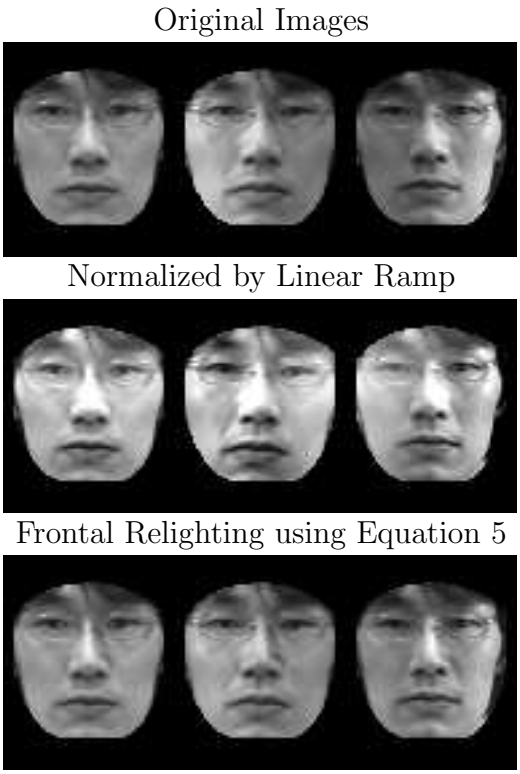


Fig. 7. Comparison of the linear ramp and our relighting method. Three images from different illuminations: frontal, left and right directional lighting (shown left to right).

quantization noise in the surface normals (we did not wish to smooth the surface normals too much and this obviously shows).

Unlike the Yale Database B, our internal dataset *does* have ambient lighting which we can model with the ambient parameter c in Eq. (3. By using a representative set of N training images we can (numerically) estimate the ambient component using the optimality criteria

$$c^* \;=\; \arg\min_{c} \sum_{i=1}^{N} \left| \rho_i(c) - \frac{1}{N} \sum_{i=1}^{N} \rho_i(c) \right|^2 \tag{6}$$

where $\rho_i(c)$ denotes the albedo of the i-th training image estimated with the relative ambient intensity c in Eq. (3).

We now demonstrate the range of our estimation and relighting capability using a small subset of our internal database which we call Dataset I. This subset has frontal views of 32 individuals under three (maximally) different lighting conditions: frontal lighting (for the gallery images) and left and right directional lighting (for the probe images) both of which are mixed in with some (unknown)

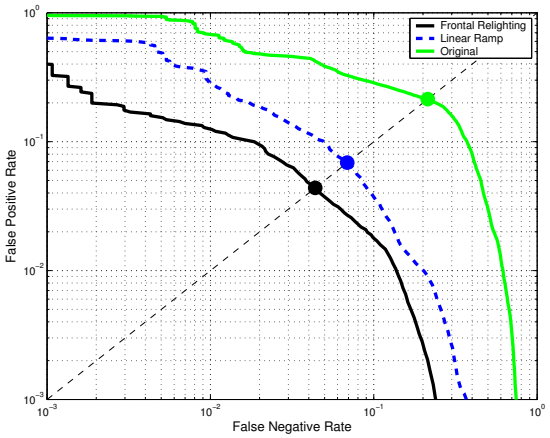


Fig. 8. The ROC curves of the face verification test based on L_2 distance in internal Dataset I. The diagonal (dashed) designates the EER line.

proportion of ambient illumination. For each lighting condition 10 images of each individual were taken for a total of 960 images.

Figure 7 shows a representative sample of our frontal relighting technique applied to the gallery and probe images (left, then center and right, respectively) of one of the 32 individuals in Database I. The top row shows the original (unprocessed) input images and the bottom row shows our relighting results. Clearly, the directional component of the illumination has been successfully removed and the frontally relit image has been generated in its place (which incorporates an ambient component estimated with Eq. (6). In contrast, the middle row shows the results obtained with *linear ramp* normalization (also known as a "facet model") which fits a plane to the image intensities (using least-squares) which is then subtracted from the input in order to remove the main illumination gradient. The residual image is then normalized back to the original intensity range. The saturation observed is a common feature of linear ramp normalization since this technique can not represent shape variations other than its own implicit Lambertian flat surface. The linear ramp model is also quite susceptible to the specular highlights which can be rather significant outliers in the linear fit. We also compared our results to histogram equalization (not shown) but the ROC curve was worse than the linear ramp.

Of course visual inspection of the apparent "good" quality of the relighted images is (for us) a secondary issue (unlike say, in computer graphics). We are primarily interested in improving the *accuracy* of face verification under arbitrary illumination. Therefore, we compared the verification performance of a given algorithm with our relighted images versus with the original (raw) images. We specifically wanted to illustrate a phenomenon which was universal, simple to understand, made the least algorithmic assumptions and could be easily du-

(a) (b)

Fig. 9. (a) Database II sample images of an individual in gallery (left) and probes (center and right). Also shown is (b) the mean "albedo" image used.

plicated by others. Therefore, we chose a "model-free" matching metric in the form of simple L_2 norm of the difference between the probe and gallery images or equivalently the Euclidean distance in pixel space. Figure 8 shows the resulting receiver operating characteristic curve (ROC) obtained using the inverse of the L_2 norm as a similarity metric for verification. We see that the commonly reported performance measure, the "equal error rate" (EER), drops from approximately 20% with raw imagery down to 4% with our frontal relighting technique (an improvement by a factor of 5). By comparison, the linear ramp normalization achieves approximately 7% EER. Moreover, the ROC curve for frontal relighting is superior to the linear ramp at every operating point specified by a given pair of false accept rate (FAR) and false reject rate (FRR). Although the improvement trend is quite clear and pronounced in this test, we acknowledge that these results are based on only 32 individuals may not generalize to larger datasets (since it is probably too optimistic).

In order to obtain a more statistically significant performance measure, we assembled the largest collection of images in our internal dataset, called Database II, to accurately calibrate and benchmark the performance advantage provided by using our lighting normalization. Database II consists of a gallery of 3,444 images of 496 individuals and a probe set of 6,663 images of 363 individuals not all of whom are in the gallery (*i.e.*, impostors). Sample images for one individual can be seen in Figure 9. Notice the bright highlights and specular reflection off the spectacles and the relatively overhead (high elevation) lighting direction in the center image (including shadows of the lens and the rims of the glasses on the cheeks). Obviously, the lighting in these images is much more complex than simple directional illumination with Lambertian reflectance.

Figure 10 shows the ROC curve for the verification test with Dataset II. These results were computed based on the statistics of almost 23 million probe-to-gallery matches yielding a relatively high degree of statistical significance. Despite the fact that these images do not contain a single directional lighting component and moreover exhibit some non-Lambertian phenomena (specularity, shadowing, inter-reflection, *etc.*) we see a similar improvement trend with the same rank ordering of performance under the two candidate lighting normalizations performed. Specifically, we see the raw data yield an EER of 10% (lower than the 20% in the previous figure since although the illumination is more com-

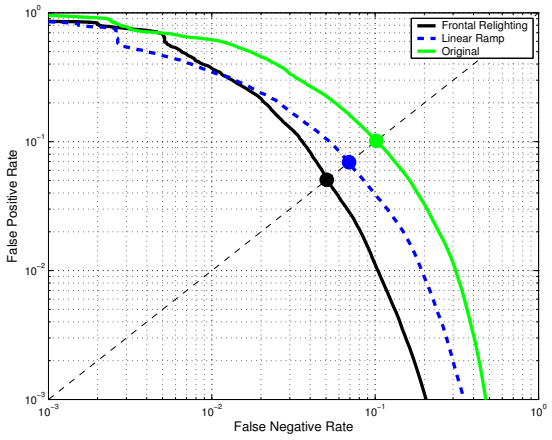


Fig. 10. The ROC of the face verification test based on L_2 distance in internal Dataset II. The diagonal (dashed) designates the EER line. The ROC for histogram equalization (not shown) is similar to that of the linear ramp but slightly worse.

plex it is still mostly near-frontal and less extreme than in Figure 7). As seen in the figure, our relighting technique achieves an EER of 5%, compared to 7% with the linear ramp normalization. Considering the size of this test and the subsequent accuracy of the performance estimates (not just the EER but the entire ROC curve) it is perhaps surprising that our simple illumination model which makes a few simplifying assumptions that are clearly violated here, still manages to provide a sizable performance advantage in complex lighting; especially with imagery in which *directional* lighting does not seem (at first glance) to be the main complicating factor adversely affecting the verification performance.

6 Conclusions

The contributions of this paper are essentially two-fold: first, a simple and practical method for estimating the dominant directional light source in a photometrically uncalibrated image of a face (whose exact shape and albedo are basically unknown) and secondly, a fast and efficient relighting technique for normalizing the image for illumination-invariant template matching and recognition. The necessary computations require less than 5 image-sized dot-products.

Furthermore, we have demonstrated the superiority of our technique in experiments with both public datasets such as Yale Face Database B and our own internal datasets of realistic access-control imagery which exhibits complex real-world illumination environments. This performance enhancement is directly due to a tighter clustering of an individual's images in image space, which will very likely help more sophisticated matching algorithms achieve illumination invariance.

Our results demonstrate that (relatively) robust estimation of lighting direction and subsequent normalization are not only possible with simplified calculations but are also quite feasible for online use. The total computational cost for estimation and relighting is only a few dot-products. Consequently, this methodology is a viable alternative for real-time applications while being superior to linear ramp and histogram equalization techniques currently in use.

References

1. R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. In *Int'l Conf. on Computer Vision*, volume 2, pages 383–390, 2001.
2. P. Belhumeur and D. Kriegman. What is the set of images of an object under all possible lighting conditions. In *Int'l J. Computer Vision*, volume 28, pages 245–260, 1998.
3. H. Chen, P. Belhumeur, and D. Jacobs. In search of illumination invariants. In *Proc. IEEE Conf. on Computer Vision & Pattern Recognition*, pages 1–8, 2000.
4. DARPA. The HumanID 3D Face Database, courtesy of Prof. Sudeep Sarkar, University of South Florida, Tampa Fl. USA.
5. A. Georghiades, D. Kriegman, and P. Belhumeur. From few to many: Generative models for recognition under variable pose and illumination. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 40:643–660, 2001.
6. M. J. Jones and P. Viola. Fast multi-view face detection. In *Proc. IEEE Conf. on Computer Vision & Pattern Recognition*, 2003.
7. K. C. Lee, J. Ho, and D. Kriegman. Nine points of light: Acquiring subspaces for face recognition under variable lighting. In *Proc. IEEE Conf. on Computer Vision & Pattern Recognition*, pages 519–526, 2001.
8. K. Nishino and S. K. Nayar. Eyes for relighting. In *Proceedings of SIGGRAPH*, 2004.
9. J. Phillips. Face Recogntion Vendor Test 2002 report. Technical report, National Institute of Standards and Technology, March 2003.
10. J. Phillips, H. Moon, S. Rizvi, and P. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.
11. R. Ramamoorthi. Analytic PCA construction for theoretical analysis of lighting variability in images of a lambertian object. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24, October 2002.
12. R. Rammamoorthi and P. Hanrahan. A signal processing framework for inverse rendering. In *Proceedings of SIGGRAPH*, 2001.
13. H. Rowley and T. Kanade. Neural network0based face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998.
14. A. Shashua and T. Riklin-Raviv. The quotient image: Class-based re-rendering and recognition with varying illuminations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(2):129–139, 2001.
15. Z. Wen, Z. Liu, and T. S. Huang. Face relighting with radiance environment maps. In *Proc. IEEE Conf. on Computer Vision & Pattern Recognition*, 2003.
16. L. Zhang and D. Samaras. Face recognition under variable lighting using harmonic image exemplars. In *Proceedings of Computer Vision and Pattern Recognition*, pages 19–25, 2003.
17. W. Y. Zhao and R. Chellappa. Symmetric shape-from-shading using self-ratio image. *Int'l J. Computer Vision*, 45(1):55–75, 2001.