

Using Eye Reflections for Face Recognition Under Varying Illumination

Ko Nishino[†]

Peter N. Belhumeur[‡]

Shree K. Nayar[‡]

[†]Dept. of Computer Science
Drexel University
Philadelphia, PA 19104
kon@cs.drexel.edu

[‡]Dept. of Computer Science
Columbia University
New York, NY 10027
{belhumeur,nayar}@cs.columbia.edu

Abstract

Face recognition under varying illumination remains a challenging problem. Much progress has been made toward a solution through methods that require multiple gallery images of each subject under varying illumination. Yet for many applications, this requirement is too severe. In this paper, we propose a novel method that requires only a single gallery image per subject taken under unknown lighting. The method builds upon two contributions. We first estimate the lighting from its reflection in the eyes. This allows us to explicitly recover the illumination in the single gallery images as well as the probe image. Next, we exploit the local linearity of face appearance variation across different people. We represent the gallery images as locally linear montages of images of many different faces taken under the same lighting (bootstrap images). Then, we transfer the estimated combination of bootstrap images to synthesize each subject's face under the probe lighting to accomplish recognition. Finally, we show through tests on the CMU PIE database that we can achieve better recognition results using our lighting estimation method and locally linear montages than the current state-of-the-art.

1. Introduction

The effect of illumination variation on the appearance of faces is known to be one of the major obstacles for face recognition [17]. This is due to the fact that lighting variation often results in larger intra-personal face appearance variation than inter-personal variation [12].

Many previous approaches have shown that good recognition rates can be achieved even under extreme lighting [20, 2, 6, 10, 16, 25, 24]. However, most of these methods require multiple gallery (or training) images per subject. Each face has to be sampled under many different light directions such that the space of images under all possible lighting can be well reconstructed. For many applications this requirement is not practical. Rather, one would desire to accomplish recognition using only a single image per subject. This would not only allow recognition on the existing single image databases but also simplify enrollment procedures for future databases.

If the single gallery images are all the information we

have, we would have to somehow extract an illumination invariant feature from each gallery image as well as the probe (or test) image to accomplish recognition. However, it has been shown that such discriminative illumination invariant features do not exist [5]. One can also take an illumination normalization approach by deriving intrinsic images of each image [7, 26]. However, recovering intrinsic images from a single image is an inherently ill-posed problem and requires restrictive assumptions.

Alternatively, in addition to the single gallery images, one can collect relevant information from other peoples' faces. For instance, a set of 3-D laser scans of many different faces can be used to compute a statistical generative model of face geometry and photometry [4, 27, 23]. Then the generative face model can be fit to the images to estimate identity-dependent parameter values for recognition. However, such an approach is computationally expensive and difficult to scale since it requires 3-D information of many people.

Furthermore, while [4] reports excellent results on face recognition under varying illumination using single gallery images, it is only tested on images taken with ambient lighting in addition to a single point source¹. It has been shown in [10] that face recognition under single point light source is significantly more difficult compared to having multiple light sources². Yet face recognition under a single point source is also important, since we frequently encounter such cases, for instance, in images taken outdoors on a sunny day or in images taken indoors with the subject close to a dominant compact source.

Although it may be impractical to capture many images of each person who we would like to recognize, it is easy to prepare an image set in which faces of many different people are captured under many different lighting conditions. In fact, publicly available face databases, which we normally use to test face recognition algorithms, can readily be used to provide general information of face appearance variation

¹Note that the CMU PIE database [22] contains two sets of images taken under varying illumination ("illum" and "lights"). One set of images ("illum") are taken with single flashes while the other set ("lights") also has the room light turned on. We run our experiments on the "illum" set.

²In [11] it is shown that for the same data set we use, the error rate almost halves every time another point source is lit in the image and reaches almost zero with 11 point light sources. Similar results are reported on the Yale Face database B [6] in [10].

under varying illumination. Let us refer to these images as the *bootstrap images* [21]. Note that the people in the bootstrap set are not the same individuals as those in the gallery. In fact, we use an entirely different database for the bootstrap set. Thus, we have to prepare such a data set only once for any set of gallery images.

In this paper, we focus on face recognition under varying illumination using a single gallery image per subject taken under unknown lighting and a bootstrap set. Using these image sets, the task is to identify the person in the *probe image* taken under unknown illumination. Shashua et al. [21] tackle this problem by deriving an illumination invariant image termed the quotient image. However, their derivation relies on restrictive assumptions such as faces exhibit pure Lambertian reflection and different faces share exactly the same surface normal distribution. Zhou et al. [28] formulate this problem as a bilinear estimation of lighting and identity in each gallery and probe image. However, bilinear estimation is prone to an inherent ambiguity and thus difficult to solve. To our knowledge, [28] reports the best recognition results in this face recognition scenario. However, we will show in this paper that the recognition rates can be significantly improved.

We propose a novel method that builds upon two contributions to accomplish face recognition in this setup. The first contribution is the use of lighting information reflected in the eyes for face recognition. Recently, it was shown in [14] that lighting information of the scene in an image can be estimated from the eye captured in the image. We show that – for images in which the eyes are open – we are able to reliably estimate the lighting for each of the given single gallery images and the probe image. The explicit knowledge of lighting allows us to develop the recognition algorithm described below. It can also benefit many other face recognition methods that require the lighting to be estimated [4, 27, 23].

The second contribution is a simple yet effective recognition algorithm which exploits the local linearity of face appearance. It has been shown in the past that one person's face can be well represented as a linear combination of other peoples' faces [25, 24]. We take this idea one step further by exploiting the spatially localized linearity in the appearance variation of different faces taken under the same lighting. We show that by simply subdividing the images with a regular grid and by representing each individual subimage with the bootstrap images, we are able to better synthesize the appearance of one person's face under a novel (probe) lighting from a single exemplar image. The work of [16] exploits the spatially localized linearity of face appearance variation using modular eigenspaces in order to accomplish face recognition based on face parts such as the mouth and the nose. We, on the other hand, exploit the local linearity in order to synthesize the appearance of the entire face with non-linear photometric effects such as highlights and cast-shadows.

In particular, for each single gallery image, we first esti-

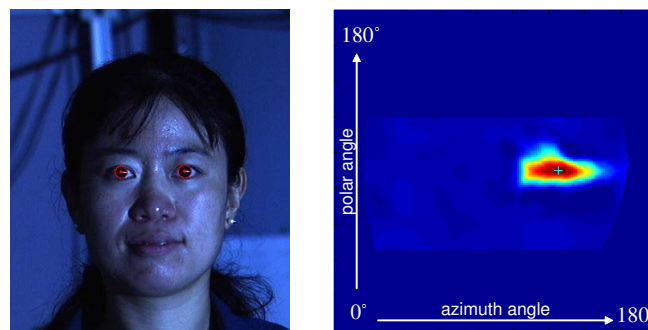


Figure 1: Left: One of the face images we used in our experiments. The image is cropped (the original size is 640×480). The estimated eye limbus are overlaid in red. Right: The environment map computed from the left eye. The intensity values are color coded with red indicating high and blue low. The estimated point source direction is depicted by a cyan cross.

mate the lighting from the eyes. Using the estimated gallery lighting, we synthesize images of each person in the bootstrap set taken under the same lighting by a simple linear combination. Then, we represent the gallery image as a locally linear montage of these bootstrap images taken under the same lighting. We do this by subdividing the images with a regular grid and computing linear coefficients for each subimage individually. In the recognition stage, given the probe image, we estimate the lighting from the eye and also prepare bootstrap images taken under the probe lighting. Then, by reusing the linear coefficients computed for each subimage of the gallery image, we synthesize an image of each potential identity taken under the probe lighting. These reconstructed images are compared to the probe image for recognition. We refer to this method as the *locally linear montage* method.

By exploiting these two factors, namely the lighting information recovered from the eyes and the local linearity of face appearance, we can significantly increase the accuracy of face recognition under varying illumination using single gallery images. We report recognition results on the CMU PIE database [22] using the Yale Face database B [6] as the bootstrap set. Compared to the result of the state-of-the-art algorithm reported on the same data [28], we reduce the error rate by more than 80%.

2. Estimating Lighting from the Eye

Recently, it was shown in [14] that lighting information of the scene in an image can be estimated from an eye captured in the image. It was shown that even from low resolution images, such as standard VGA images, one can recover an environment map of the scene which covers almost the entire frontal hemisphere. The computed environment map can be used as the illumination distribution of the scene.

In general, face recognition requires detecting the eyes and the mouth locations in the image as a pre-processing step

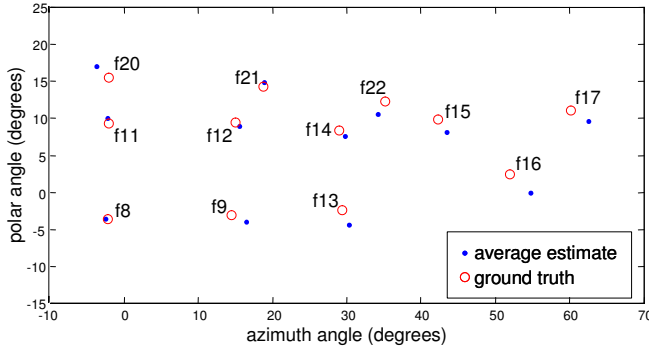


Figure 2: Average estimates over 68 people of the 12 different point source directions computed from the eyes. The symbol f# denotes the light source (flash) number.

to normalize the sizes of the faces. In our experiments, we assume that this pre-processing is already done with a face detection algorithm and we are provided with the rough locations of the eyes. Using the provided eye location as the initial estimate of the center of the eye, we use the algorithm described in [15] to fit an ellipse to the limbus (the contour of the cornea) in the image. The initial estimate of the ellipse parameters is set to a circle of pre-determined radius (8 pixels) for all images.

Figure 1 shows one of the face images used in our experiments and its estimated environment map. In the computed environment map, we thresholded the intensity values and computed the centroid of the remaining pixels which was used as the point source direction (see Figure 1). In the images of the CMU PIE database, since flashes were used as the point sources, the reflection in the cornea was often a short line rather than a point.

We evaluated the accuracy of the point source direction estimates on the $68 \times 12 = 816$ images we used in the CMU PIE database. We were able to estimate the light source direction from at least one of the eyes in 96.4% (786 of 816) of the images. In the remaining 3.6% images, the eyes were closed or nearly closed. As we discuss later, for such images, we can fall back to other methods to estimate the lighting and use it in the algorithm we describe in Section 4. Figure 2 shows the average estimates of the 12 different point source directions. Figure 3 shows the histogram of errors in the estimated polar and azimuth angles of the point source directions. The RMS error in polar and azimuth angles were 2.4° and 3.5° , respectively. The errors in polar angles were larger than those in the azimuth angles. This is mainly due to the fact that the eyelids often partially occlude the limbus and hence the estimate of the corneal orientation becomes less accurate in the polar direction.

These results show that, in images with open eyes, we can estimate the lighting with very good accuracy. While the lighting information estimated from the eyes may prove useful in many different face recognition algorithms [4, 27, 23], in this paper, we introduce a new recognition algorithm

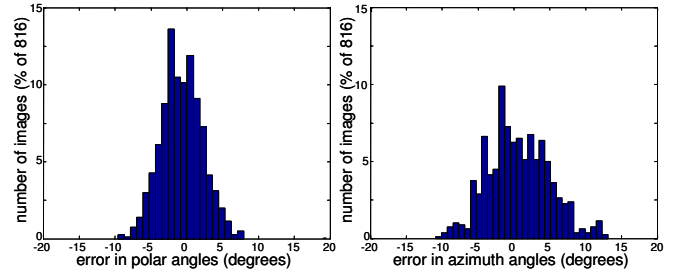


Figure 3: Histograms of polar and azimuth angle errors of the estimated point source directions. The RMS errors in polar and azimuth angles were 2.4° and 3.5° , respectively.

designed to explicitly exploit the known lighting.

3. Linear Reconstruction

It has been shown in the past that one person's face can be well-represented as a linear combination of other peoples' faces [25, 24]. Since we have already estimated the lighting for each gallery image and probe image from the eyes, we can represent each image as a linear combination of other peoples' images in the bootstrap set taken under the same lighting as follows.

Once we estimate the lighting for each gallery image or the probe image from the eyes, we are able to represent the estimated lighting as a linear combination of the lighting sampled in the bootstrap set. Let us denote the estimated point source direction in either the gallery image or the probe image with a 3×1 vector $\mathbf{s}$. Similarly, we will denote the point source directions sampled in the bootstrap set with $\mathbf{s}^j$ ($j = 1, \dots, K$). Then we can compute the *lighting coefficients* $\mathbf{c} = \{c^1, \dots, c^K\}$ which are the linear coefficients of $\mathbf{s}^j$ for $\mathbf{s}$. Since lighting is generally frontal in face images, it is natural to assume $c^j \geq 0$. Hence we solve a non-negative minimization problem to estimate the lighting coefficients:

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c}} \|\mathbf{s} - \sum_j c^j \mathbf{s}^j\|_{L_2}, \quad (1)$$

subject to $\hat{c}^j \geq 0$. If we have a complex lighting, for instance consisting of area light sources, we can estimate and represent it as a set of point sources by sampling the environment map computed from the eye. Then we can compute the corresponding lighting coefficients for each point source using Eq. (1).

Once we compute the lighting coefficients $\hat{\mathbf{c}}$, we can use them to synthesize face images of the people in the bootstrap set under the gallery/probe lighting $\mathbf{s}$. Let us denote the bootstrap images by $\mathbf{x}_b^{i,j}$ where j ($j = 1, \dots, K$) represents one point source direction sampled in the bootstrap set and i ($i = 1, \dots, M$) is the identity of the person in the bootstrap set. In our experiments, we use the Yale Face database B [6] as the bootstrap set in which we have $M = 10$ and $K = 64$. Then, we can synthesize bootstrap images under the given lighting $\mathbf{s}$ as

$$\mathbf{x}_b^{i,\hat{\mathbf{s}}} = \sum_j^K \hat{c}^j \mathbf{x}_b^{i,j}, \quad (2)$$

where $\hat{\mathbf{s}}$ is the estimate of $\mathbf{s}$ using $\hat{\mathbf{c}}$ in Eq. (1). Note that we have the light direction $\hat{\mathbf{s}}$ in the superscript so that it is easier to see that the left hand side is a bootstrap image of person i taken under the lighting $\hat{\mathbf{s}}$.

Then, we can estimate the linear coefficients $\hat{\mathbf{w}}$ to represent the gallery/probe image $\mathbf{x}$ with these bootstrap images by solving a simple linear estimation problem:

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \|\mathbf{x} - \sum_i^M w^i \mathbf{x}_b^{i,\hat{\mathbf{s}}}\|_{L_2}. \quad (3)$$

Here again, we assume that the linear coefficients are non-negative and solve Eq. (3) subject to $\hat{w}_i \geq 0 (i = 1, \dots, M)$. Also, we normalize each image such that $\|\mathbf{x}\| = 1$ to factor out the brightness variation in lighting among images and differences in the overall reflectivities of the different faces.

Let us refer to the linear coefficient set $\hat{\mathbf{w}}$ as the *identity coefficients* since it is unique for each individual. Zhou et al. [28] also estimate the lighting coefficients and identity coefficients to represent the gallery image and the probe image as a linear combination of bootstrap images. However, they simultaneously estimate both coefficients in a bilinear formulation assuming pure Lambertian reflection. It is, however, well-known that bilinear estimation suffers from ambiguities [9]. It has been shown that lighting estimation from shading in images suffer from bas-relief ambiguity [3]. Especially for face images, since Lambertian reflection acts as a low-pass filter in the frequency domain [1, 18], simultaneous estimation of both the identity and lighting in Eq. (1) and (3) is prone to local minima. In fact, as we will demonstrate later, even when the probe image and the gallery image are exactly the same, it is not guaranteed that the bilinear approach will estimate the same identity and lighting coefficients. Note that we do not suffer from such ambiguities, since the lighting is explicitly recovered from the eyes and used to estimate the lighting coefficients separately from the identity coefficients.

Since the identity coefficients are independent of the illumination, we can reuse the identity coefficients $\hat{\mathbf{w}}_g^k$ computed from the gallery image of person k , to synthesize his/her face image under the probe lighting $\mathbf{s}_p$:

$$\hat{\mathbf{x}}_p^k = \sum_i^M \hat{w}_g^{i,k} \mathbf{x}_b^{i,\hat{\mathbf{s}}_p}. \quad (4)$$

Here the probe lighting $\mathbf{s}_p$ is computed from the eye and its estimate $\hat{\mathbf{s}}_p$ is computed using Eq. (1). Once we reconstruct each person's face image under the probe lighting (*reconstructed gallery image*), we can compare them to the actual probe image to identify the person. We will describe this method and its related assumptions in detail in the following section.

This reconstruction approach using Eq. (4) allows us to assess the quality of subsequent recognition by looking at

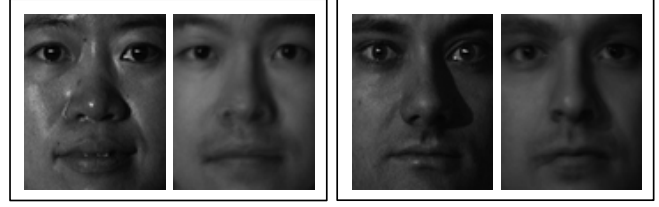


Figure 4: Two example results of the linear reconstruction method described in Section 3. In each example, the left image is the probe image and the right image is the reconstructed gallery image (reconstructed image of the person in the gallery image under the lighting in the probe image). Although the overall appearance is similar to the probe image, the reconstructed gallery image does not include the details of the face.

how well each individual's face appearance is synthesized under the probe lighting. Figure 4 shows two probe images and the same persons' reconstructed gallery images. Although the reconstructed gallery images do capture the overall appearance of each person's face, they are very smooth and the details of each individual's face are lacking. In order to accomplish recognition with high accuracy, we need to reconstruct the images with finer details – details that captures each individual's facial characteristics as well as correct illumination effects including cast-shadows and highlights.

4. Locally Linear Reconstruction

In order to see why the linear reconstruction method fails to reconstruct the appearance of each individual under the probe lighting well, let us first take a geometrical view of the method.

Each image in the gallery, probe, and bootstrap set can be considered as a unique point in $\mathbb{R}^n$. Then, computing identity coefficients $\hat{\mathbf{w}}$ using Eq. (3) corresponds to projecting the gallery image onto the hyperplane formed by the bootstrap images taken under the same lighting (gallery lighting) and computing its relative coordinates with respect to the bootstrap image points on this hyperplane. The relative coordinates of this gallery image projection are then used to reconstruct a point on the hyperplane formed by the bootstrap images taken under the probe lighting, see Eq. (4). The underlying claim made here is that the image points of faces taken under the same lighting can be well-approximated with a linear hyperplane, and the relative position of one image point with respect to other image points on this hyperplane is invariant to the lighting change.

Clearly, this assumption does not hold for images of faces. Because of non-linear appearance variations due to specular reflections and cast-shadows, we cannot expect that lighting variations will preserve the relative coordinate system of image points of different faces. However, we can make simple modifications to the way we compute the relative coordinates such that they are well preserved under illumination variations.

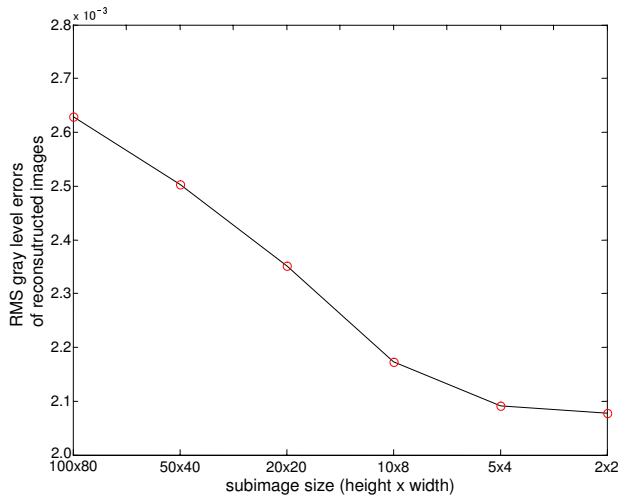


Figure 5: Example results of varying the amount of subdivision of images used to reconstruct the gallery images under the probe lighting. The original image size was 100×80 and we compared the reconstruction errors for 6 different levels of subdivision. The RMS error of gray level values of the reconstructed images decreases as the images are subdivided into smaller subimages.

4.1. Spatially Local Linearity

While images of one person's face taken under varying illumination have been shown to lie close to an extremely low-dimensional subspace [6, 8, 1, 18], face images of different people taken under the same lighting do not exhibit the same behavior [24, 25]. Therefore, approximating the gallery/probe image using the hyperplane formed by the set of bootstrap images would simply not yield a good reconstruction. Once again, consider Figure 4. With only a modest number of bootstrap images it is difficult to represent the fine details in face images – details that are crucial for recognition.

However, if we observe locally in the spatial domain of face images, we can expect the intensity variation to be of much lower-dimensional compared to the variation of the entire image. In other words, while the appearance variation of the entire face across different people is highly non-linear, local image regions exhibit spatially localized linearity because of similar geometric and photometric properties³ [13]. In [16], this spatially local linearity of face appearance variation across different people has been exploited for constructing low-dimensional subspaces of facial parts such as the nose and mouth. Here, we can exploit this local linearity to reduce the dimensionality of the subspace formed by all images points in which we compute the relative coordinates. In other words, we can make the linear hyperplane approximation of the image point of faces taken under the same lighting work better by reducing the dimensionality of the subspace in which those image points lie.

³In [21], the authors assume pure Lambertian reflection and that the surface normals are exactly the same among different people and only the albedo varies. Note that we are not making such restrictive assumptions.

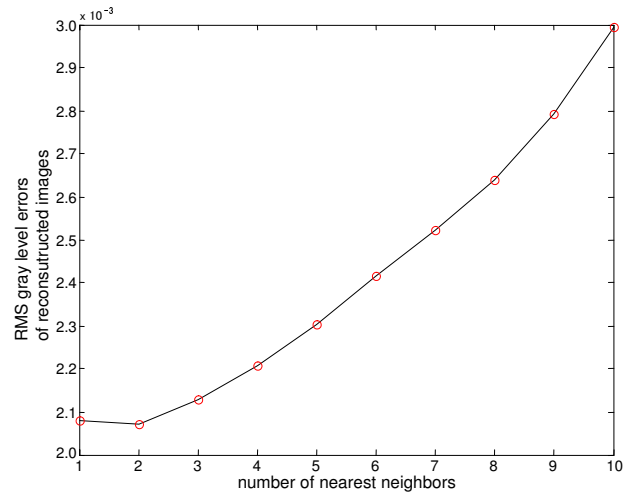


Figure 6: Example results of varying the number of nearest neighbors used to reconstruct the gallery images under the probe lighting. The RMS error of gray level values of the reconstructed images decreases when only neighboring bootstrap images are used (10 corresponds to using all images). The optimal number of nearest neighbors was 2 for this specific example.

Since the faces are pre-aligned, we can simply subdivide the images into rectangular subimages with a regular grid and compute identity coefficients for each individual subimage. In other words, we will first represent the gallery image as a montage of bootstrap images taken under the same lighting. Then, we use the same combination with the bootstrap images taken under the probe lighting to reconstruct a montage of the same person under the probe lighting. Figure 5 shows example results of varying the degree to which the images are subdivided to reconstruct the gallery images under the probe lighting. In this example, the gallery lighting was f8 and the probe lighting was f15 (see Figure 2 for lighting directions). The RMS errors of gray level values were computed over 68 people (note that each image is normalized such that the total energy is 1). The original image size was 100×80 and we compared the reconstruction errors for 6 different levels of subdivision. As can be seen in the results, the smaller the subimages become the better we can reconstruct the image. In the example shown in Figure 5, the best result was achieved when the images were subdivided into 2×2 subimages. Note that, for different combinations of lighting, the optimal size of subimages might slightly vary. In our experiments, we use a fixed size (2×2) for all tests.

4.2. Reconstruction from Neighbors

Recently, in the context of manifold embedding, Saul et al. [19] showed that one can effectively compute a low-dimensional embedding of image points in high-dimensional space by assuming that nearby points remain nearby. In our case, we would expect that faces that are close by in the image space would have more similar shape and photometric properties than those faces that are far apart. As a result, we

expect that nearby faces to one face in the image space under one lighting would be better representative for that face under another lighting. To this end, we follow [19] and only use neighboring image points to represent an image point of interest in $\mathbb{R}^n$ (or $\mathbb{R}^{\hat{n}}$, where $\hat{n} = \#$ of pixels in each subimage). Instead of using all bootstrap images taken under the same lighting to compute the identity coefficients using Eq. (3), we will find q nearest neighbors to the gallery image and use only those bootstrap images to compute the identity coefficients⁴. Then we will use the same q people's face images taken under the probe lighting to reconstruct the gallery image under the probe lighting from the computed identity coefficients using Eq. (4).

Figure 6 shows example results of varying the number of nearest neighbors used to reconstruct probe images. In this example, the gallery lighting was f8 and the probe lighting was f15 (see Figure 2 for lighting directions). The images were subdivided into 2×2 subimages. The RMS errors of gray level values were computed over 68 probe images (note that each image is normalized such that the total energy is 1). Since the bootstrap set (Yale database B) has 10 people, 10 nearest neighbors corresponds to using all bootstrap images under the same lighting. As can be seen, using only 2 nearest neighbors yields the best reconstruction.

The optimal number of nearest neighbors depends on how densely the image points are sampled as well as the dimensionality of the subspace that all image points form. As a result, it is difficult to automatically determine the number of nearest neighbors to be used. This remains as future work. In our experiments, we use $q = 3$ which we found to yield best results.

4.3. Locally Linear Montage

Let us now summarize the face recognition method we propose. We will refer to the proposed algorithm as the *locally linear montage* method, where locally linear refers to both spatially within the image and also locally in the image space. Also, remember that we need to first explicitly estimate the lighting in order to use this method, which we do from the eyes in the images.

Given a set of gallery images $\mathbf{x}_g$ and bootstrap images $\mathbf{X}_b = \{x_b^{i,j} | i = 1, \dots, M, j = 1, \dots, K\}$, we would like to identify the person in the probe image $\mathbf{x}_p$. Note that the gallery images need not be taken under the same lighting.

1. For each gallery image $\mathbf{x}_g^k$
 - (a) Estimate the lighting s_g^k from the eye.
 - (b) Compute the lighting coefficients $\hat{\mathbf{c}}_g^k$ using Eq. (1).
 - (c) Synthesize bootstrap images taken under lighting $\hat{s}_g^k$ as $\mathbf{X}_b^{\hat{s}_g^k}$ using Eq. (2).

⁴We use the Euclidean distance for measuring distances between the images.

- (d) Subdivide both the gallery image and the bootstrap images into rectangular patches: $\mathbf{x}_g^{k,m}$, $\mathbf{X}_b^{\hat{s}_g^k, m}$ where m is the patch number.
- (e) Compute the identity coefficients $\hat{\mathbf{w}}^{k,m}$ using Eq. (3) for each rectangular patch m with q nearest neighbor bootstrap images.

2. Given the probe image $\mathbf{x}_p$
 - (a) Estimate the lighting s_p from the eye.
 - (b) Compute the lighting coefficients $\hat{\mathbf{c}}_p$ using Eq. (1).
 - (c) Synthesize bootstrap images taken under lighting $\hat{s}_p$ as $\mathbf{X}_b^{\hat{s}_p}$.
 - (d) Subdivide the bootstrap images taken under the probe lighting $\mathbf{X}_b^{\hat{s}_p}$ into rectangular patches $\mathbf{X}_b^{\hat{s}_p, m}$.
3. For each person k in the gallery image set, reconstruct the image under the probe lighting $\hat{\mathbf{x}}_p^k$ with $\hat{\mathbf{w}}^{k,m}$ and $\mathbf{X}_b^{\hat{s}_p, m}$ using Eq. (4) for each patch.
4. Compare every reconstructed image $\hat{\mathbf{x}}_p^k$ with the probe image $\mathbf{x}_p$. The identity k that gives the smallest error in L_2 is the estimate of identity.

Note that step 1 is an off-line procedure and the actual recognition process (step 2 to 4) only involves simple multiplications and additions which can potentially be implemented to run in real time.

5. Experimental Results

In order to conduct a comprehensive evaluation and a fair comparison with the state-of-the-art method of Zhou et al. [28], we evaluated our algorithm on the same setup used in [28]. We used Yale Face database B [6] which has 10 people taken under 64 different point source directions as the bootstrap images. Following [28], we ran recognition tests on the 12 light source subset of CMU PIE database which has 68 people [22]. As in [28], we tested our algorithm on all possible combinations of gallery lighting and probe lighting.

We ran three recognition methods on the data set: the bilinear estimation method [28], the global linear reconstruction method described in Section 3, and the locally linear montage method given in Section 4. The table shown in Figure 7 shows the recognition rates for all combinations of gallery and probe lightings⁵.

⁵The images which we could not estimate the lighting from the eyes (3.4% of the images) were not used in the experiments shown in Figure 7 for fair and clear comparison. However, in practice, we can fall back to the bilinear estimation in [28] to estimate the lighting for such images and use it in our locally linear montage method. The results did not change with this hybrid method (92.8%) while the bilinear method ran on all images also did not change (67.4%).

		gallery lighting																avg.
		f8	f9	f11	f12	f13	f14	f15	f16	f17	f20	f21	f22					
probe lighting	f8	95.6/100.0/100.0	83.8/98.5/100.0	89.7/97.1/100.0	83.8/91.2/100.0	60.3/73.5/96.9	61.8/63.2/93.8	38.2/45.6/84.6	30.9/29.4/67.7	14.7/32.4/54.7	76.5/88.2/100.0	72.1/79.4/98.4	41.2/44.1/92.1	62.4/70.2/90.7				
	f9	86.8/98.5/100.0	97.1/100.0/100.0	85.3/92.6/100.0	82.4/98.5/100.0	82.4/92.6/100.0	79.4/80.9/100.0	60.3/66.2/98.5	38.2/55.9/98.5	35.3/44.1/84.9	72.1/77.9/100.0	82.4/89.7/100.0	60.3/63.2/96.9	71.8/80.0/98.2				
	f11	86.8/98.5/100.0	76.5/97.1/100.0	94.1/100.0/100.0	95.6/92.6/100.0	72.1/66.2/98.5	69.1/63.2/100.0	39.7/48.5/95.5	30.9/36.8/77.6	23.5/35.3/72.2	94.1/88.2/100.0	82.4/79.4/100.0	54.4/51.5/98.5	68.3/71.4/95.2				
	f12	72.1/89.7/100.0	88.2/100.0/100.0	85.3/94.1/100.0	94.1/100.0/100.0	79.4/85.3/100.0	80.9/89.7/100.0	64.7/73.5/100.0	48.5/51.5/89.6	39.7/42.6/88.9	86.8/95.6/98.4	94.1/100.0/98.5	72.1/73.5/100.0	75.5/83.0/97.9				
	f13	60.3/82.4/89.2	83.8/92.6/100.0	54.4/66.2/91.0	70.6/92.6/100.0	92.6/100.0/100.0	92.6/98.5/100.0	86.8/89.7/98.5	69.1/82.4/89.6	44.1/55.9/83.3	63.2/61.8/89.1	75.0/79.4/98.5	89.7/86.8/98.5	73.5/82.4/94.8				
	f14	58.8/73.5/90.8	82.4/88.2/98.5	58.8/70.6/95.5	88.2/95.6/98.5	89.7/97.1/98.5	95.6/100.0/100.0	89.7/100.0/98.5	66.2/77.9/92.5	63.2/57.4/90.7	63.2/69.1/96.9	83.8/89.7/98.5	92.6/100.0/98.5	77.7/84.9/96.4				
	f15	30.9/44.1/72.3	58.8/70.6/92.4	41.2/39.7/82.1	54.4/75.0/100.0	79.4/86.8/100.0	80.9/100.0/100.0	95.6/100.0/100.0	80.9/94.1/98.5	76.5/64.7/96.3	44.1/45.6/78.5	58.8/70.6/95.5	95.6/98.5/100.0	66.4/74.1/93.0				
	f16	26.0/36.8/86.2	39.7/51.5/90.9	23.5/29.4/80.6	39.7/45.6/91.0	64.7/79.4/97.0	72.1/73.5/100.0	89.7/95.6/100.0	95.6/100.0/100.0	89.7/73.5/100.0	25.0/26.5/78.5	42.6/38.2/92.4	73.5/77.9/100.0	56.4/60.7/93.1				
	f17	26.5/32.4/73.6	33.8/42.6/88.7	22.1/29.4/85.2	33.8/45.6/92.6	45.6/55.9/92.6	41.2/57.4/96.3	75.0/72.1/100.0	94.1/80.9/100.0	187.1/100.0/100.0	20.6/36.8/75.9	26.5/50.0/92.6	54.4/54.4/100.0	47.5/54.8/91.5				
	f20	72.1/83.8/98.4	70.6/73.5/95.2	94.1/86.8/100.0	79.4/92.6/96.9	60.3/54.4/93.8	67.6/70.6/95.3	45.6/50.0/83.1	32.4/26.5/58.5	19.1/38.2/57.4	95.6/100.0/100.0	83.8/97.1/100.0	63.2/50.0/95.4	65.3/68.6/89.5				
	f21	66.2/72.1/95.2	79.4/92.6/98.4	77.9/72.1/98.5	91.2/98.5/100.0	86.8/76.5/96.9	89.7/86.8/98.5	64.7/72.1/97.0	50.0/47.1/81.8	42.6/50.0/68.5	76.5/98.5/98.5	94.1/100.0/100.0	83.8/79.4/98.5	75.2/78.8/94.3				
	f22	52.9/52.9/82.5	55.9/73.5/90.6	47.1/42.6/89.2	70.6/70.6/93.8	88.2/88.2/96.9	88.2/97.1/98.5	83.8/100.0/97.0	70.6/76.5/86.4	61.8/48.5/88.9	52.9/54.4/92.3	75.0/76.5/100.0	97.1/100.0/100.0	70.3/73.4/93.0				
	avg.		60.8/72.1/90.7	70.8/81.7/96.2	64.5/68.4/93.5	73.7/83.2/97.7	75.1/79.7/97.6	76.6/81.7/98.5	69.5/76.1/96.1	58.9/63.2/86.7	50.6/53.6/82.2	64.2/70.2/92.3	72.5/79.2/97.9	73.2/73.3/98.2	67.5/73.5/94.0			

Figure 7: Recognition results of bilinear estimation [28] / global linear reconstruction (Section 3) / locally linear montage. The locally linear montage method achieves over 25% better recognition rate compared to the bilinear estimation method.

In the diagonal of the results, one can see that the bilinear estimation method does not achieve 100% recognition rates while the other two methods do. These are the cases when exactly the same images are used for both the gallery and the probe. We ran the bilinear alternating minimization described in [28] with random initial estimates, since there are no legitimate reasons to use a constant initial value for either the identity or lighting coefficients. These results show that the ambiguity between lighting and identity in the bilinear estimation is difficult to resolve. The other two methods which explicitly estimates the lighting from the eyes do not suffer from such ambiguity.

As shown in Figure 7, once the lighting estimated from the eyes is used (global linear reconstruction), the overall recognition rate increases 6% compared to the bilinear method. Furthermore, by exploiting the local linearity of face appearance taken under the same lighting in the locally linear montage method, we achieve 94% recognition rate. In [28], the authors increase the number of people in the bootstrap images to 100 by using synthetic face images rendered from the Blanz and Vetter 3-D Face database [4]. This yields 93% overall recognition rate. This recognition rate is about the same as what we can achieve with only 10 people in the bootstrap set. We expect that the locally linear montage method can scale better.

Figure 8 shows several reconstructed gallery images under a specific probe lighting. By comparing the third row (global linear reconstructions) with the second row (bilinear reconstructions), one can see that the overall shading on the face is more accurate once the lighting information recovered from the eyes is used. This can be observed by looking at the sharpness of the shadow boundaries cast by the noses. In the locally linear montages (the fourth row), the facial features of each individual is recovered in detail, as well as the non-linear effect of highlights and cast-shadows.

The locally linear montages are blocky since each individual subimage is currently handled independently. In order to achieve even better reconstruction we would need to exploit the spatial correlation of face appearance variation among different people. For instance, instead of subdividing the images with a regular grid and handling each subimage independently, we would have to simultaneously estimate both the optimal spatial partitioning/segmentation and the identity coefficients such that the reconstruction errors

are minimized. However, such an optimization is extremely difficult to solve. We are currently investigating efficient algorithms which will yield sub-optimal but effective solutions to circumvent this problem.

6. Conclusion

In this paper, we first showed that the lighting information recovered from the eyes in face images can be of significant use for face recognition under varying illumination. Next, we showed that the explicit information of lighting allows us to exploit the local linearity of face appearances of different people taken under the same lighting. By reconstructing the gallery image as a locally linear montage of bootstrap images taken under the probe lighting using the combinations computed for the single gallery image, we are able to better synthesize each person's facial details and non-linear illumination effects. We showed that these two contributions result in significantly better recognition rates through a comprehensive test on face recognition with varying illumination from single gallery images.

Acknowledgments

This work was conducted at the Columbia Vision and Graphics Center of Columbia University while the first author was affiliated with Columbia University as a postdoctoral research scientist. This work was supported by NSF IIS 00-85864 and NSF IIS 03-08185 in part.

References

- [1] R. Basri and D.W. Jacobs. Lambertian Reflectance and Linear Subspaces. *IEEE TPAMI*, 25(2):218–233, 2003.
- [2] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE TPAMI*, 19(7):711–720, 1997.
- [3] P.N. Belhumeur, D.J. Kriegman, and A.L. Yuille. The Bas-Relief Ambiguity. *IJCV*, 35(1):33–44, 1999.
- [4] V. Blanz and T. Vetter. Face Recognition Based on Fitting a 3D Morphable Model. *IEEE TPAMI*, 25:1063–1074, 2003.
- [5] H.F. Chen, P.N. Belhumeur, and D.W. Jacobs. In Search of Illumination Invariants. In *IEEE CVPR*, volume 2, pages 254–261, 2000.
- [6] A.S. Georgiades, P.N. Belhumeur, and D.J. Kriegman. From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. *IEEE TPAMI*, 23(6):643–660, 2001.

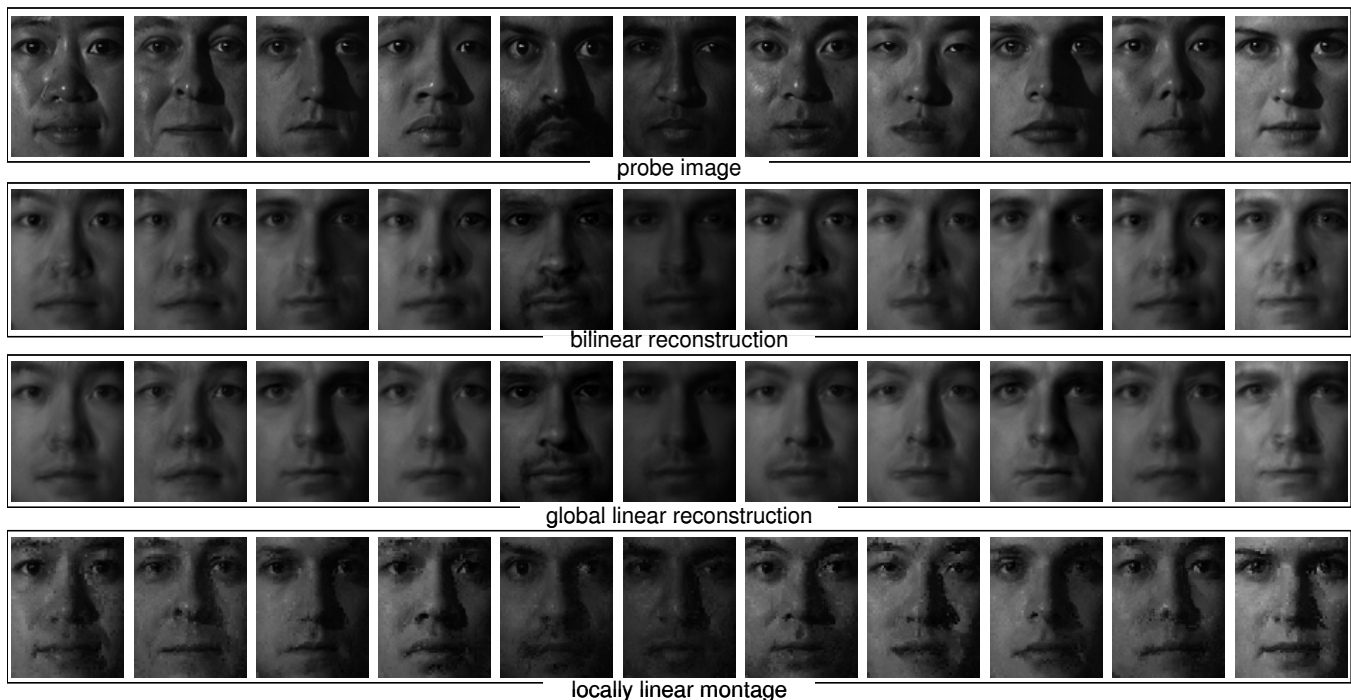


Figure 8: Reconstruction results of the three different methods. First row: probe images, second row: bilinear reconstruction [28], third row: global linear reconstruction (Section 3), and fourth row: locally linear montage. The locally linear montages capture both the details of each individual face and the illumination effects including shading, cast shadows and highlights.

- [7] R. Gross and V. Brajovic. An Image Processing Algorithm for Illumination Invariant Face Recognition. In *Int'l Conf. on Audio- and Video-Based Biometric Person Authentication*, pages 10–18, 2003.
- [8] P. Hallinan. A Low-Dimensional Representation of Human Faces for Arbitrary Lighting Conditions. In *IEEE CVPR*, pages 959–999, 1994.
- [9] J.J. Koenderink and A.J. van Doorn. The Generic Bilinear Calibration-Estimation Problem. *IJCV*, 23(3):217–234, 1997.
- [10] K.-C. Lee, J. Ho, and D. Kriegman. Nine Points of Light: Acquiring Subspaces for Face Recognition under Variable Lighting. In *IEEE CVPR*, pages 519–526, 2001.
- [11] K.-C. Lee, J. Ho, and D. Kriegman. Acquiring Linear Subspaces for Face Recognition under Variable Lighting. *IEEE TPAMI*, 27(5):684–698, May 2005.
- [12] Y. Moses, Y. Adini, and S. Ullman. Face Recognition: The Problem of Compensating for Changes in Illumination Direction. In *ECCV*, pages 286–296, 1994.
- [13] S.K. Nayar, P.N. Belhumeur, and T.E. Boult. Lighting Sensitive Display. *ACM TOG*, 23(4):963–979, 2004.
- [14] K. Nishino and S.K. Nayar. Eyes for Relighting. *ACM TOG*, 23(3):704–711, 2004.
- [15] K. Nishino and S.K. Nayar. The World in Eyes. In *IEEE CVPR*, volume I, pages 444–451, 2004.
- [16] A. Pentland, B. Moghaddam, and T. Starner. View-Based and Modular Eigenspaces for Face Recognition. In *IEEE CVPR*, pages 84–91, 1994.
- [17] P.J. Phillips. Human Identification Technical Challenges. In *ICIP*, volume 1, pages 49–52, 2002.
- [18] R. Ramamoorthi. A Signal-Processing Framework for Reflection. *ACM TOG*, 23(4):1004–1042, 2004.
- [19] L.K. Saul and S.T. Roweis. Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds. *JMLR*, 4:119–155, Dec. 2003.
- [20] G. Shakhnarovich and B. Moghaddam. *Handbook of Face Recognition*, chapter Face Recognition in Subspaces. Springer-Verlag, 2004.
- [21] A. Shashua and T. Riklin-Raviv. The Quotient Image: Class Based Re-rendering and Recognition With Varying Illumination. *IEEE TPAMI*, 23(2):129–139, 2001.
- [22] T. Sim, S. Baker, and M. Bsat. The CMU Pose Illumination and Expression Database. *IEEE TPAMI*, 25(12):1615–1618, 2003.
- [23] T. Sim and T. Kanade. Combining Models and Exemplars for Face Recognition: An Illumination Example. In *CVPR Workshop on Models versus Exemplars in Computer Vision*, 2001.
- [24] L. Sirovitch and M. Kirby. Low-Dimensional Procedure for the Characterization of Human Faces. *JOSA, A*, 2:519–524, 1987.
- [25] M. Turk and A. Pentland. Eigenfaces for Recognition. *Journal of Cog. Neuroscience*, 3(1):71–96, 1991.
- [26] H. Wang, S.Z. Li, and Y. Wang. Generalized Quotient Image. In *IEEE CVPR*, volume 2, pages 498–505, 2004.
- [27] L. Zhang and D. Samaras. Face Recognition Under Variable Lighting using Harmonic Image Exemplars. In *IEEE CVPR*, volume I, pages 19–25, 2003.
- [28] S. Zhou and R. Chellappa. Rank Constrained Recognition from Unknown Illuminations. In *IEEE AMFG*, 2003.