

Morphable Reflectance Fields for Enhancing Face Recognition

Ritwik Kumar¹, Michael Jones², Tim K. Marks²

¹School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA

²Mitsubishi Electric Research Laboratories, Cambridge, MA, USA

rkkumar@seas.harvard.edu, {mjones, tmarks}@merl.com

Abstract

In this paper, we present a novel framework to address the confounding effects of illumination variation in face recognition. By augmenting the gallery set with realistically relit images, we enhance recognition performance in a classifier-independent way. We describe a novel method for single-image relighting, Morphable Reflectance Fields (MoRF), which does not require manual intervention and provides relighting superior to that of existing automatic methods. We test our framework through face recognition experiments using various state-of-the-art classifiers and popular benchmark datasets: CMU PIE, Multi-PIE, and MERL Dome. We demonstrate that our MoRF relighting and gallery augmentation framework achieves improvements in terms of both rank-1 recognition rates and ROC curves. We also compare our model with other automatic relighting methods to confirm its advantage. Finally, we show that the recognition rates achieved using our framework exceed those of state-of-the-art recognizers on the aforementioned databases.

1. Introduction

Over the past decade, various methods have been proposed to address the problem of face recognition. These methods provide varying advantages ranging from performance benefits such as fast classification, high accuracy, low false acceptance rate, and fully automated processing to usability benefits such as public domain availability and ease of implementation. Given this landscape of techniques, rather than introducing yet another classifier, we propose a novel relighting model and gallery augmentation framework that can be used to enhance recognition rates of existing classifiers. This strategy builds on the advantages of the original classifier to achieve improved performance.

We will demonstrate that our method substantially improves the performance of a wide variety of existing classifiers, yielding results that exceed state-of-the-art recognition performance on three face datasets: CMU PIE [13],

Multi-PIE [6], and MERL Dome [20].

Here we focus on the illumination problem for face recognition. As many previous papers have shown, determining whether two images of a face are of the same person is an especially difficult problem when the images are taken under very different illuminations. There are three main classes of approaches to deal with this problem. The first is based on building classifiers that use illumination-invariant image features. Recent examples of approaches in this class include Local Binary Patterns [1], Local Gabor Binary Patterns [22], and Local Ternary Patterns [15].

The second class of approaches explicitly attempt to normalize out the effects of illumination from the images. These methods either assume a reflectance model (e.g., Lambertian) and attempt to remove the effects of light or adopt an image processing approach in which various steps are empirically chosen to provide desired output. Popular methods in this category include Quotient Image [12], Generalized Quotient Image [18], Total Variation Model [4], and Tan and Triggs lighting normalization [15].

Finally, there is the category of approaches that attempt to generate synthetic relit images that generalize from the given gallery set so as to match all possible illumination variations in the probe set (throughout this paper, *gallery set* refers to the known, enrolled images of faces, while *probe set* refers to the unknown, test images of faces). The newly generated images are added to the gallery set in the hope that each probe image will find a close match among at least one of the images in the augmented gallery. The 3D Morphable Models method [3], which involves fitting a 3D shape model and a texture model to a given input image, can be viewed as an online optimization-based example of gallery augmentation. Here the relighting is achieved by assuming Phong's model for reflectance. Since the introduction of 3D Morphable Models, various derivatives have been proposed, including [19] and [21]. Though effective in generating good quality relighting, these methods suffer from the requirements of manual initialization and cumbersome optimization, which reduce their attractiveness for recognition applications with large subject sets. In contrast,

methods such as [14] and [11] do not require explicit fitting of 3D shape models to the input images and hence are computationally more efficient and robust in practice. Note that some of these techniques, such as [11] and [12], can be used for both relighting and lighting normalization.

The second and third categories of approaches described above are in a sense inverses of each other. One attempts to remove any lighting effects while the other attempts to synthesize all lighting conditions. In many face recognition applications, the illumination of the gallery images can be at least roughly controlled. We argue that given a somewhat frontally lit gallery image and more harshly lit probe images, synthesizing harsh lighting conditions from the gallery image is simpler than the inverse problem of normalizing the harsh lighting conditions of the probe images. Thus, the approach we take lies in the third category. We propose a novel relighting method, which we call Morphable Reflectance Fields (MoRF), that works in the 2D domain, is fully automatic, and requires as few as one input image. In fact, all of the results presented in this paper use only a single input gallery image per person. Unlike [3], our technique does not require manual initialization and thus is a more attractive choice for large-scale recognition. Additionally, it is built on the recently proposed non-Lambertian Tensor Splines Model [2], allowing it to more expressively relight images than other automatic techniques ([11], [14]).

2. Morphable Reflectance Fields

We begin by considering the reflectance field for a face [5]. The *reflectance field* is an 8-dimensional function of the 4D incident light field arriving at the surface of the face and the 4D radiant light field leaving the surface. Here we consider only distant point light sources and frontal pose, which reduces our interest to only a 4D slice of the standard reflectance field: two dimensions for the surface coordinates of the radiant light and two for the angle of the incident light. Whereas previous work has defined the surface coordinates to be a 2D parameterization of the 3D points on the surface, our model treats the surface of a face as a 2D plane. Thus, our surface coordinates correspond exactly to 2D image coordinates. For convenience, we define the output of the reflectance field as a brightness value rather than the standard radiance value. We use the term *reflectance field* in this paper to denote this 4D slice of the standard reflectance field. Furthermore, we use *reflectance function* to refer to the 2D spherical function obtained by restricting the reflectance field to a fixed location (pixel). Hence, in this paper, the reflectance field is a 2D field of reflectance functions, one for each pixel.

Our method for augmenting a gallery of face images has three parts: building a reflectance field model, fitting this model to the given gallery image(s), and using the fitted model to synthesize images under a variety of lighting con-

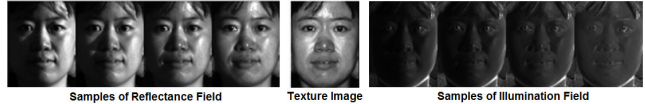


Figure 1. Our model factorizes a reflectance field as the product of a texture image and an illumination field. These images illustrate a reflectance and illumination field sampled in 4 lighting directions.

ditions. The first part of our technique involves building a reflectance field model, which captures a sufficient amount of the variation seen in actual facial reflectance fields so that it can be readily fit to any given input face. We call our model a Morphable Reflectance Field (MoRF) (see Section 2.1). Once we have a reflectance field model that is expressive enough to approximate any face image under any illumination, we fit this model to the input face image(s), as described in Section 2.2. After fitting, the input face is relit by changing the lighting direction input to the fitted reflectance field (see Section 4.1). The relit images are added to the gallery set to obtain an augmented gallery set.

Our method is fully automated. In practice, we currently require that for each person in the gallery, we are given at least one input image that is somewhat frontally lit. This requirement ensures that accurate pixelwise correspondences can be computed between the gallery image and a reference face using a 2D Morphable Model (2DMM) [7] (see Section 2.1.3). Additional gallery input images for each person are not required, but if provided they are assumed to be pixelwise aligned amongst themselves. In addition to estimating the reflectance field of the face, our method also recovers an estimate for the lighting condition in the given images. Our reflectance field model is different than the models used by existing techniques ([3, 19]) in that although it implicitly incorporates 3D shape information, it is fully defined by a 2D field of spherical functions. This allows the fitting of the model to be carried out without any 3D-to-2D projections.

2.1. Reflectance Field model

Our representation of the reflectance field breaks it down into two parts: 1) illumination-dependent appearance variation (which we refer to as the *illumination model* or *illumination field*) and 2) texture. Interestingly, the separation of a given reflectance field into an illumination model and a texture has not been well defined in the literature. Most often, the definition of texture and of the illumination model are dependent on the assumed bi-directional reflectance distribution function (BRDF). For instance, in the Lambertian model, the albedo—a constant scaling factor at each pixel—is commonly accepted as the texture, while the cosine term is considered to be the illumination function. The Lambertian reflectance function at a pixel is the product of the albedo and the illumination function. We use the term *illumination function* to refer to the function at each pixel that

Algorithm 1: Building the MoRF model $\{\hat{t}_j, \bar{T}, \hat{l}_i, \bar{\mathcal{L}}\}$

Input: *IllumTrainSet* of N_l subjects with k (≥ 9) directionally lit images each; *TexTrainSet* of N_t texture images; 2DMM with Reference Face F_r

Output: Mean texture image \bar{T} , eigen-texture images \hat{t}_j , mean illumination field $\bar{\mathcal{L}}$, eigen-illumination fields \hat{l}_i

- 1 **foreach** Subject $s \in$ *IllumTrainSet* **do**
- 2 Non-rigidly warp the k images to F_r (Sec. 2.1.3)
- 3 From the warped images, compute Tensor Splines representation $\Gamma_s = \{\gamma_{klm}\}$ at all pixels (Eq. 2) using the algorithm in [2]
- 4 Compute the illumination field
 $\mathcal{L}_s =$ vectorize(Γ_s /frontally lit image of s)
- 5 **foreach** Image $T_t \in$ *TexTrainSet* **do**
- 6 Non-rigidly warp T_t to F_r (Sec. 2.1.3)
- 7 $\mathcal{T}_t =$ vectorize (warped T_t)
- 8 $\bar{T} = (1/N_t) \sum_t \mathcal{T}_t$, $\bar{\mathcal{L}} = (1/N_l) \sum_s \mathcal{L}_s$
- 9 $\hat{l}_i =$ Eigenvectors($\sum_s (\mathcal{L}_s - \bar{\mathcal{L}})(\mathcal{L}_s - \bar{\mathcal{L}})^T / N_l$)
- 10 $\hat{t}_j =$ Eigenvectors($\sum_t (\mathcal{T}_t - \bar{T})(\mathcal{T}_t - \bar{T})^T / N_t$)

takes illumination direction as input and outputs a scalar brightness value. An *illumination field* is a field of such illumination functions (one for each pixel).

We have chosen to use definitions of texture and of illumination model that are independent of any particular BRDF. We define the texture at a pixel to be the intensity value obtained from a frontally lit image. Given this definition of texture, the illumination function at a pixel is defined as the quotient function obtained by dividing the pixel’s reflectance function by the pixel’s texture. Thus,

$$R(x, y, \theta, \phi) = L(x, y, \theta, \phi) \cdot T(x, y) \quad (1)$$

where $R(x, y, \theta, \phi)$ is the reflectance field, $L(x, y, \theta, \phi)$ is the illumination field, $T(x, y)$ is the texture image, (x, y) is the pixel location, and (θ, ϕ) is the illumination direction.

We define texture to be the frontally lit image for pragmatic reasons: We need a large collection of texture images to build our model (see Section 2.1.2), and it is fairly easy to obtain frontally lit images. If such practical constraints can be overcome, alternate definitions of texture such as the mean of the reflectance function can be explored. Fig. 1 shows an example of one subject’s texture image along with a few samples of her reflectance field and illumination field.

2.1.1 Illumination model

Our illumination model estimates the illumination field using 3rd-order Tensor Splines [2]. In this framework, the

illumination function at pixel (x, y) is a spherical function that takes a lighting direction as input and outputs a scalar:

$$L(x, y, \theta, \phi) = \sum_{k+l+m=3} \gamma_{klm} (v_1)^k (v_2)^l (v_3)^m, \quad (2)$$

where γ_{klm} are real-valued tensor coefficients, the indices k, l, m are nonnegative integers, and the lighting direction in Cartesian coordinates is given by $v_1 = \sin(\theta)\cos(\phi)$, $v_2 = \sin(\theta)\sin(\phi)$, $v_3 = \cos(\theta)$. This illumination function has 10 tensor coefficients, $\Gamma = \{\gamma_{klm} \mid k + l + m = 3\}$ at each pixel, and in [2], it takes 9 or more images to recover these. Note that these coefficients at each pixel are scaled down by the frontally lit image in order to remove the texture. In practice, we found that normalizing the coefficient vector at each pixel to unit norm also provides a good approximation of the illumination function.

The Tensor Spline representation enables our illumination model to capture specularities and cast and attached shadows beyond the capabilities of the Lambertian model. Since we do not want to fit a 3D shape model, the Tensor Splines framework provides a seamless way to account for photo-effects caused by global shape, such as cast shadows.

In our work, unlike [2], we do not assume that we have 9 or more input images per person in the gallery. In fact, our method only requires a single image as input (though more can be used). In order to estimate the illumination field for the input image, we constrain the search space to be a linear combination of the illumination fields computed for a set of training faces. The training faces’ illumination fields are learned once, off-line, using the algorithm described in [2] that requires 9 illumination images per subject. The training images used to compute these illumination fields, as well as the input face image, are first warped into pixelwise correspondence (see Section 2.1.3). For efficiency, we apply principal component analysis (PCA) to the training illumination fields and use linear combinations of the first several principal components (plus the mean illumination field) to represent the illumination field of any person in the gallery.

The 3rd-order Tensor Splines representation for illumination fields contains 10 coefficients per pixel. For the j th training illumination field, we string all 10 coefficients at all of the M pixels into a single vector, \mathcal{L}_j , of length $10M$. Letting $\bar{\mathcal{L}}$ represent the mean of these training illumination field vectors, PCA yields orthonormal bases \hat{l}_i for the illumination model, and any illumination field \mathcal{L} can be written

$$\mathcal{L} = \left(\sum \alpha_i \hat{l}_i + \bar{\mathcal{L}} \right). \quad (3)$$

2.1.2 Texture model

We have defined texture to be the frontally lit image of a face. To fit our MoRF model to any face image, we must

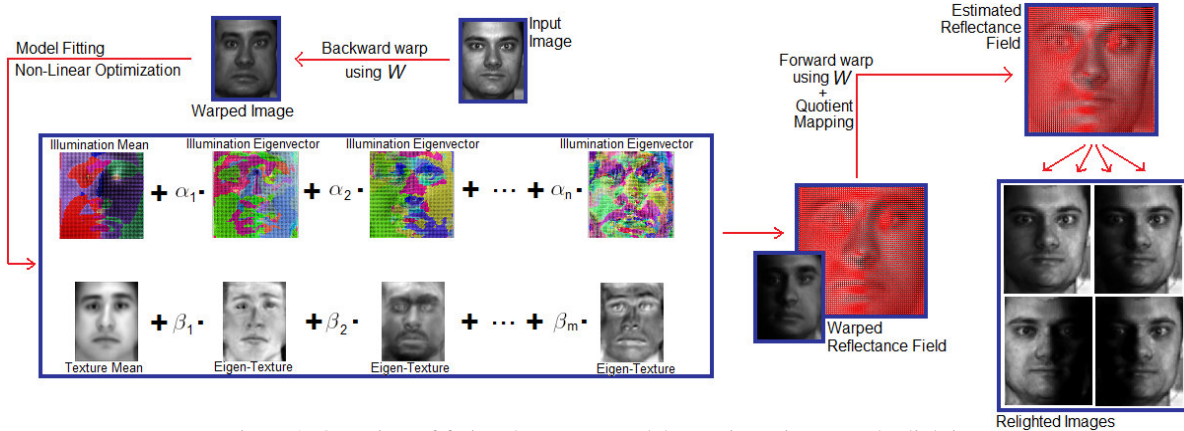


Figure 2. Overview of fitting the MoRF model to an input image and relighting

model how the textures of faces can vary across the population. To do this, we apply PCA to a set of training face texture images (all captured under frontal illumination), which are first put into pixelwise correspondence as described in Section 2.1.3. Any face texture image \mathcal{T} can then be expressed as a linear combination of basis textures (obtained using PCA), \hat{t}_j , plus the mean training texture, $\bar{\mathcal{T}}$:

$$\mathcal{T} = \left(\sum \beta_j \hat{t}_j + \bar{\mathcal{T}} \right). \quad (4)$$

By eliminating the PCA components with the smallest eigenvalues, the number of terms in the texture and illumination models can be chosen according to computational and quality requirements. We call the set $\{\hat{t}_j, \bar{\mathcal{T}}, \hat{l}_i, \bar{\mathcal{L}}\}$ our Morphable Reflectance Field (MoRF) model. The process for model building is summarized in Algorithm 1.

2.1.3 Pixelwise Correspondence

It is important for all of the illumination fields and textures of the MoRF model to be in pixelwise correspondence, so that linear combinations of illumination fields and textures yield other valid illumination fields and textures. To achieve this, we use a 2D morphable model (2DMM) [7]. A 2DMM consists of a reference face, a set of warp fields (obtained using the bootstrapping method proposed in [16]) that represent the ways 2D shapes of faces can vary, and a set of texture images (the same set of training images used for the texture model in Section 2.1.2) that represent the ways the shape-free appearances of faces can vary. A 2DMM is fit to an input image using a nonlinear optimization that finds the best linear combination of textures, such that when warped by the optimal linear combination of shapes, the input image is reconstructed with minimal L_2 error. The warp field, $W = (W_x, W_y)$, resulting from the optimal linear combination of shapes gives the pixelwise correspondences between the reference face of the 2DMM and the input image. $W_x(x, y)$ takes a pixel in the reference face image

Algorithm 2: Fitting the MoRF model to an input face

Input: K input images, I_k , where image I_{k^*} is lit from roughly frontal direction $(\theta_{k^*}, \phi_{k^*})$; MoRF Model $\{\hat{t}_j, \bar{\mathcal{T}}, \hat{l}_i, \bar{\mathcal{L}}\}$; Reference Face F_r

Output: Reflectance field R for input subject

- 1 Non-rigidly backward-warp I_k to F_r (Sec. 2.1.3)
 - 2 Optimize Eq. 5 to obtain parameters $\alpha_i, \beta_j, \theta_k, \phi_k$
 - 3 $\hat{R}(x, y) = (\sum_i \alpha_i \hat{l}_i + \bar{\mathcal{L}})(x, y) \cdot (\sum_j \beta_j \hat{t}_j + \bar{\mathcal{T}})(x, y)$ as in Eq. 1.
 - 4 Forward-warp \hat{R} to obtain R
 - 5 $I_f = \text{sample of } R \text{ in } (\theta_{k^*}, \phi_{k^*}) \text{ direction}$
 - 6 $R = R / I_f \times I_{k^*}$
-

and outputs the corresponding x position in the input image. $W_y(x, y)$ outputs the corresponding y position.

Before fitting the 2DMM to an input face image, the image is roughly cropped and rectified by running a face detector (we use a version of the Viola-Jones detector [17]) and feature detectors to find at least two facial feature points. These are used to compute a 2D similarity transform that aligns the face to a canonical scale, rotation and position. All of the training images for illumination fields and textures are cropped and rectified in this manner, resulting in an initial rough alignment.

The 2DMM is used to find pixelwise correspondences between the 2DMM reference face and each of the training images used to build the illumination fields. Thus, the illumination field eigenvectors, \hat{l}_i , are in correspondence because all of the training images were put in correspondence. Similarly, all of the training textures images are also warped into pixelwise correspondence with the 2DMM reference face to yield texture eigenvectors, \hat{t}_j , all in correspondence.

2.2. Model fitting

Given K images I_k of a face ($K \geq 1$) under unknown point-light sources, the problem now is to fit the MoRF

model (from Section 2.1) to the input image(s). The unknowns include the non-rigid deformation W to align the model to the input face, the lighting directions in each of the K images, the texture model coefficients, and the illumination model coefficients. We recover these unknown parameters by minimizing the following objective function:

$$E_1(W_x, W_y, \alpha_i, \beta_j, \theta_k, \phi_k) = \sum_k \sum_{(x,y)} \left\| I_k(W_x(x,y), W_y(x,y)) - D\left(\left(\sum_{i=1}^n \alpha_i \hat{l}_i + \bar{\mathcal{L}}\right)(x,y), \left(\sum_{j=1}^m \beta_j \hat{t}_j + \bar{\mathcal{T}}\right)(x,y), \theta_k, \phi_k\right) \right\|^2, \quad (5)$$

where W_x and W_y are the x and y components of the non-rigid deformation that gives correspondences from the reference face to the input images, (θ_k, ϕ_k) is the illumination direction of the k th input image, α_i are the illumination coefficients, and β_j are the texture coefficients. The function D takes the 10 coefficients of the estimated illumination function at (x, y) , the estimated scalar texture value at (x, y) , and the lighting direction (θ_k, ϕ_k) and computes the brightness at each pixel using the tensor splines basis:

$$D(\mathcal{L}(x, y), \mathcal{T}(x, y), \theta, \phi) = \mathcal{T}(x, y) \cdot [\mathcal{L}(x, y)^T S(\theta, \phi)] \quad (6)$$

where $S(\theta, \phi)$ is the vector of Tensor Spline basis functions, defined as

$$S(\theta, \phi) = [v_1^3 \ v_2^3 \ v_3^3 \ v_1^2 v_2 \ v_1^2 v_3 \ v_1 v_2^2 \ v_1 v_3^2 \ v_1 v_2 v_3 \ v_2^2 v_3 \ v_2 v_3^2]^T \quad (7)$$

where (v_1, v_2, v_3) are the Cartesian coordinates of the lighting direction (θ, ϕ) as defined in Section 2.1.1. The fitting procedure described above is graphically depicted in Fig. 2.

In addition to the objective function defined above, we constrain the search space for the illumination model further by adding the following Tikhonov regularizer to Eq. 5:

$$E_2(\alpha_i) = \lambda \cdot \sum_i^n \alpha_i^2, \quad (8)$$

where λ is the regularization parameter. This constraint prevents the estimated illumination field from straying too far from the training illumination fields and results in artifact-free relighted images.

We break down the process of recovering the unknowns into four steps. In the first step, the input images are aligned with the MoRF model. We apply the automatic face detector and the 2DMM described in Section 2.1.3 to the input image with somewhat frontal illumination to compute the non-rigid deformation parameters. These deformation parameters, W_x and W_y , are used to backward-warp the input image(s) into pixelwise correspondence with the eigen-illumination fields and eigen-textures of the MoRF model.

Database	Subjects	Gallery Set Images per Person	Probe Set Images per Person	Total
Extended Yale B [10]	38	-	-	-
CMU PIE [13]	67	1	20	1407
MERL Dome [20]	242	1	15	3872
Multi-PIE (Sess. 1) [6]	249	1	18	4731

Table 1. Various databases used in our experiments and the recognition setup. For all the datasets, all images used are in frontal pose with neutral expression. For each test dataset, a frontally lit image of each subject is used as the gallery image, while all other point-source-lit images of the subject are used as probe images.

Next, to compute the remaining unknowns we minimize

$$E(\alpha_i, \beta_j, \theta_k, \phi_k) = E_1(W_x, W_y, \alpha_i, \beta_j, \theta_k, \phi_k) + E_2(\alpha_i) \quad (9)$$

using a gradient-based nonlinear optimization technique. (In practice, MATLAB’s *fminunc* function with 200 iterations provides a good estimate of the unknown parameters.)

After recovering the unknowns, we have an estimate of the reflectance field of the input face, but it is still backward-warped (in pixelwise alignment with the reference face). As the third step in our model-fitting process, we forward-warp the estimated reflectance field using the deformation parameters computed earlier. Since the entire process involves two registration steps (warps), however, the resulting reflectance field provides images that appear grainy.

To remove these interpolation artifacts, we have incorporated a final step in the fitting process that we call quotient mapping. First, we generate a synthesized image from the estimated reflectance field using the same lighting direction as the somewhat-frontally-lit input image. (The lighting direction for this image was computed as part of the optimization procedure described above.) Next, we compute the quotient map by dividing the near-frontally lit image by its synthesized estimate. This quotient map is then used to scale the entire estimated reflectance field, which suppresses the artifacts introduced by interpolation during the warps of the reflectance field. The process for model fitting is summarized in Algorithm 2.

3. Experimental Setup

To test the proposed model, we used the Extended Yale B [10] dataset to build the MoRF model, since it contains a large variety of good quality point-source-lit images. We used images from the CMU PIE [13], Multi-PIE [6] and MERL Dome [20] face databases as test sets. Though our method can be used with multiple input gallery images, in all of the experiments in this paper we tackle the most difficult case: a single input gallery image. A brief description of the subsets of databases used is presented in Table 1. We have used only images with frontal pose, neutral expression, and illumination variation in our experiments. All of the images used were point-source lit.

We conducted experiments with the given single-image

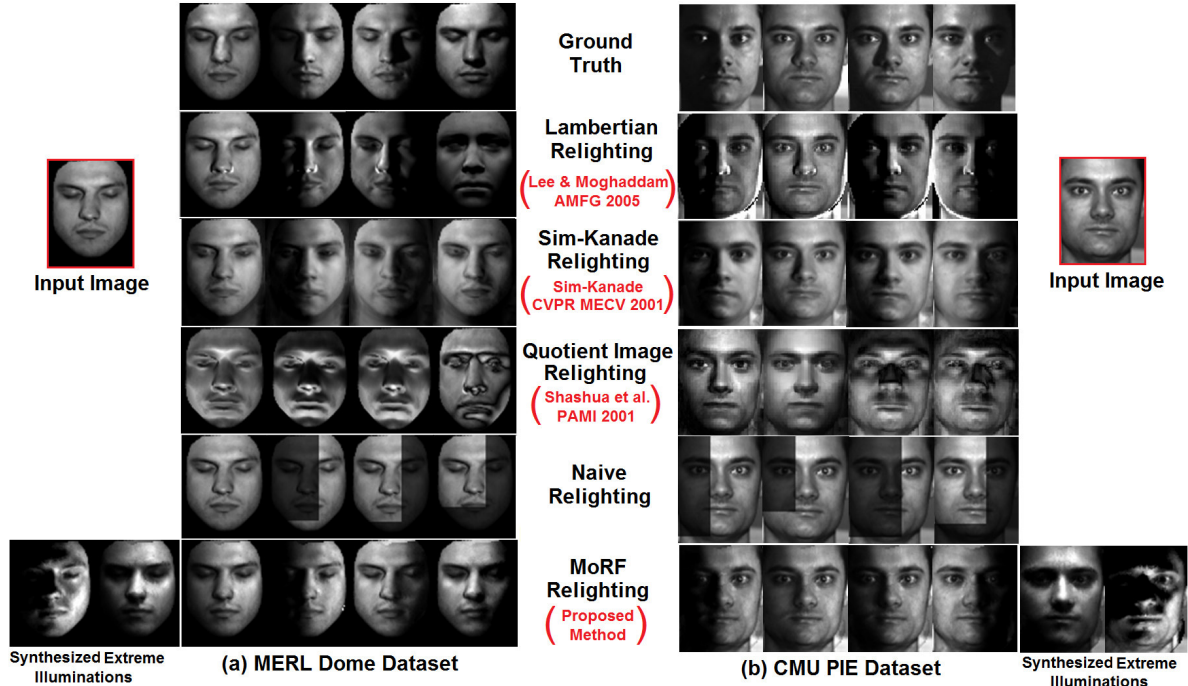


Figure 3. Comparison of images relighted using various methods (see Section 3.1) with ground truth images from the CMU PIE and MERL Dome databases. The single input images used are shown in the red boxes.

gallery, as well as with a gallery that we augmented by adding to the gallery set relit images that were synthesized from the single input image. In every gallery augmentation experiment described in Section 4.2, the gallery was augmented with the described number of relit images, whose illumination directions are sampled uniformly from the range -60° to $+60^\circ$ in both the azimuth and elevation angles.

To compare a probe image to an augmented gallery set for a single subject, the probe is compared pairwise to each gallery image in the augmented set and the maximum similarity score is returned as the overall similarity score.

Success of any scheme based on gallery set augmentation is contingent on its ability to limit the number of false acceptances as new images are added. In our case, this depends on the quality of relighting. If the relighting is not realistic, it is likely that the newly added images in the gallery set will provide more close matches to an incorrect subject’s probe rather than aid in recognizing the correct subject. The results, in Section 4, show that our MoRF relighting method is successful in keeping the false acceptance rate in check while boosting the rank-1 recognition rate.

3.1. Relighting Methods

We compare the performance of our **MoRF** relighting framework with three other relighting methods. The first is **Lambertian Relighting** [11], a method for single-image relighting that boasts a simple implementation with high performance. The method uses an average 3D head model and the Lambertian reflectance assumption to estimate the

lighting direction and albedo of an input face image which can then be used to render the face under any illumination direction. The second relighting method we implemented is **Sim-Kanade Relighting** [14], a learning-based method that models non-Lambertian lighting effects using Gaussian error functions. We used the Extended Yale B dataset to train this method. We call the third method we implemented **Naïve Relighting**. This simple relighting methods darkens (to 30% of the original grayscale values) a part of the image to give the rough appearance that part of the image is in shadow due to lighting. We simply divide the image into rectangular segments and generate new images with pixels in some rectangles darkened, to determine whether the more complicated relighting schemes proposed here (and elsewhere) are better than naïvely darkening segments of the images to simulate shadowing. We also experimented with relighting using the **Quotient Image** [12] but have not included recognition results from that method since its low quality leads to extremely low recognition rates. Note that we have compared results only with other relighting techniques that do not require any manual intervention.

3.2. Classifiers

In order to demonstrate that our MoRF relighting and gallery augmentation framework is helpful independent of the particular classifier used, we chose five different classifiers for our tests. The first is the L_1 distance-based **Nearest Neighbor** classifier, used as a baseline technique. The next is an implementation of the **Haar-like Feature (HF) Rec-**

Relighting method	Size of Augmented Gallery							
	256	128	64	32	16	8	4	1
Naive	92.5	92.2	91.8	93.1	92.8	90.4	90.9	90.1
Lambertian [11]	92.7	92.2	92.3	91.6	90.7	90.2	91.3	90.1
MoRF	98.4	98.4	98.1	97.4	97.9	96.6	94.8	90.1

Table 2. Effect of gallery augmentation size: Face Recognition rates for the CMU PIE dataset with the HF Recognizer.

ognizer [8]. This classifier is trained using AdaBoost to discriminate same-face image pairs from different-face image pairs. It was trained on a large set of face images collected at our lab that did not include any of the Yale B Extended, PIE, MERL Dome, or Multi-PIE face images. Recently, **Local Binary Patterns (LBP)** [1] and **Local Ternary Patterns (LTP)** [15] have been reported to be fairly successful in illumination-invariant face recognition. These methods use local histograms of features with the χ^2 distance measure. Finally, we also include results obtained using the **Volterrafaces** [9] technique, an example of subspace transformation-based approaches to face recognition.

4. Results

In every experiment reported here, the system is given a single gallery image per subject, which is frontally lit, and which is used to synthesize any relit images.

4.1. Relighting

Though not the central focus of our work, here we qualitatively compare the quality of the relit images with the ground truth images. Fig. 3 presents images of two subjects, one from the MERL Dome database and one from the CMU PIE database, that were relit using our MoRF model. In both cases, the single image in the red box was used as the input. For comparison, we also include relighting results from three other methods: Lambertian Relighting [11], Sim-Kanade Relighting [14], and Quotient Image Relighting [12]. It is apparent from Fig. 3 that our relighting method provides more realistic images than those generated by these other methods. To challenge our MoRF model further, we also use it to synthesize relit images of each face under two extreme illuminations, for which there were no ground truth images available. To confirm that the reflectance fields are person-specific, we verified that the estimated coefficients are significantly different from one person to another.

4.2. Face Recognition

An important parameter in a gallery set augmentation scheme is the number of synthesized images to add to the gallery. This can affect performance in three ways. As gallery set size increases, the likelihood of a correct close match increases, the likelihood of a false acceptance increases, and the classification time increases. Assuming

that the classification time is not critical (since modern commercial classifiers work quite fast), here we focus on the impact on recognition rates and false acceptance rates.

The first set of results, presented in Table 2, shows how face recognition rates change as the number of images used to augment the gallery increases. Here we have used the HF recognizer with our MoRF relighting method (as well as two other relighting methods) on the CMU PIE dataset. Note in Table 2 that across different relighting schemes, the recognition rates improve as the number of synthesized images is increased. Trends similar to those shown in Table 2 were observed for other datasets and classifiers but are not included here due to space constraints.

For the rest of the experiments, the basic recognition setup is given in Table 1. For gallery augmentation, we fix the number of images to 64, since it provides a reasonable trade-off between the memory/time requirements and recognition rates. In Table 3, we compare various combinations of relighting methods and recognizers on the CMU PIE, MERL Dome, and Multi-PIE databases. Notice that across the databases, recognition methods, and relighting techniques, our MoRF relighting method provides the best results in terms of rank-1 recognition rates.

Rank-1 recognition rates are widely used in the literature, but they do not provide much insight into the impact of gallery augmentation on False Acceptance Rates (FAR). We explore this aspect of gallery augmentation using receiver operating characteristic (ROC) curves, which plot False Rejection Rates (FRR) against FAR as the recognition threshold is varied. In Fig. 4, we show ROC curves corresponding to the best three classifiers (HF, LBP and LTP) from Table 3. The plots have been truncated on both axes to save space. Fig. 4 demonstrates that gallery set augmentation using our MoRF relighting outperforms the existing methods across almost all recognizers and databases.

Finally, we compare the best results obtained using our MoRF framework to state-of-the-art classifier performance on the tested databases (Table 4). We included results from the Nearest Neighbor classifier (baseline), the HF classifier, and the Enhanced Local Texture Features (ELTF) classifier [15]. ELTF takes an illumination normalization approach, then classifies the images using Local Ternary Pattern features with a χ^2 distance measure. (Actually, [15] used a different distance measure, but we got better results using χ^2 .) Results for all of the above methods were obtained using the given single-image gallery. The last two rows in Table 4 present the results obtained using our MoRF relighting-based gallery augmentation with the LTP and Volterrafaces recognizers. (The Volterrafaces classifier was used only in the augmented-gallery experiment because it cannot be used with a single gallery image.) Table 4 shows that recognition rates achieved using our MoRF framework readily exceed rates achieved by state-of-the-art classifiers.

Relighting Method	CMU PIE				MERL Dome				Multi-PIE			
	NN (L_1)	HF [8]	LBP [1]	LTP [15]	NN (L_1)	HF [8]	LBP [1]	LTP [15]	NN (L_1)	HF [8]	LBP [1]	LTP [15]
None (single image)	59	90	93	93	43	71	66	67	18	70	63	63
Naïve	70	92	94	93	54	78	66	68	31	72	63	64
Sim-Kanade [14]	74	94	94	95	66	83	68	70	65	77	69	70
Lambertian [11]	81	92	96	97	78	81	85	86	43	71	73	74
MoRF	94	98	99	99	80	83	88	90	66	77	77	78

Table 3. Face recognition rates for the CMU PIE, MERL Dome and Multi-PIE datasets. Each column heading indicates the classifier used.

Face Recognition Method	CMU PIE	MERL Dome	Multi-PIE
Nearest Neighbor (L_1)	58.8	42.8	17.8
HF Recognizer [8]	90.1	71.2	69.7
ELTF [15]	98.1	82.9	83.6
MoRF + LTP + χ^2 [1]	99.1	89.6	78.3
MoRF + Volterrafaces [9]	100	99.4	99.3

Table 4. Comparison with the state of the art

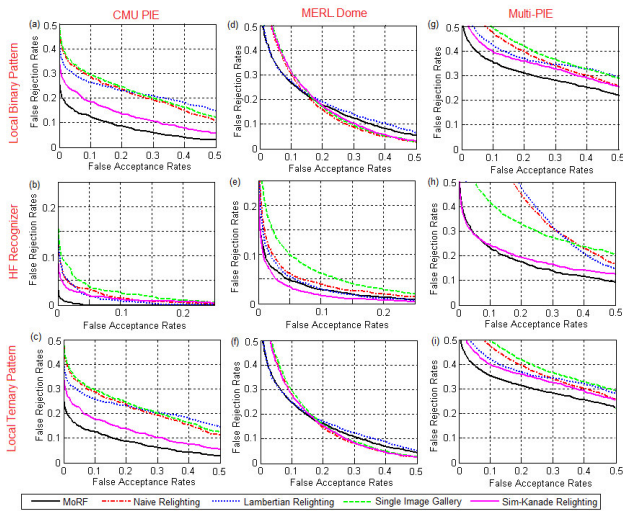


Figure 4. ROC curves for the CMU PIE, MERL dome and Multi-PIE datasets.

5. Conclusion

In this paper, we have demonstrated that realistic relighting can be effectively used to enhance recognition via gallery set augmentation. We have presented a novel automatic single-image (or multiple-image) relighting technique, MoRF, that uses a Tensor Splines-based reflectance model and PCA-based model fitting to achieve recognition and relighting performance superior to other automatic relighting methods. Our method is purely 2D-based and does not require cumbersome 3D-to-2D projections or manual initialization. We also show that the recognition performance obtained using our MoRF relighting and gallery augmentation framework exceeds that of various state-of-the-art classifiers across databases. In the future, we plan to extend our work to handle variations in pose, as well as input gallery images with more complex lighting.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE PAMI*, 28(12):2037–2041, 2006.

[2] A. Barmpoutis, R. Kumar, B. C. Vemuri, and A. Banerjee. Beyond the lambertian assumption: A generative model for apparent brdf fields of faces using anti-symmetric tensor splines. *CVPR*, 2008.

[3] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE PAMI*, 25(9):1063–1074, 2003.

[4] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. Huang. Total variation models for variable lighting face recognition. *IEEE PAMI*, 28:1519–1524, 2006.

[5] P. Debevec, T. Hawkins, C. Tchou, H. Duiker, and W. Sarokin. Acquiring the reflectance field of a human face. *ACM SIGGRAPH*, 2000.

[6] R. Gross, I. Matthews, J. Cohn, S. Baker, and T. Kanade. The CMU Multi-Pose, Illumination, and Expression (Multi-PIE) face database. *Tech Rep TR-07-08, CMU*, 2007.

[7] M. Jones and T. Poggio. Multidimensional morphable models: A framework for representing and matching object classes. *IJCV*, 29(2):107–131, 1998.

[8] M. Jones and P. Viola. Face recognition using boosted local features. *Tech Report TR2003-25, MERL*, 2003.

[9] R. Kumar, A. Banerjee, and B. C. Vemuri. Volterrafaces: Discriminant analysis using volterra kernels. *CVPR*, 2009.

[10] K.-C. Lee, J. Ho, and D. J. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE PAMI*, 27(5):684–698, 2005.

[11] K.-C. Lee and B. Moghaddam. A practical face relighting method for directional lighting normalization. *AMFG*, 2005.

[12] A. Shashua and T. Riklin-Raviv. The quotient image: Class-based re-rendering and recognition with varying illuminations. *IEEE PAMI*, 23(2):129–139, 2001.

[13] T. Sim, S. Baker, and M. Bsat. The CMU Pose, Illumination, and Expression database. *IEEE PAMI*, 25(12):1615–1618, 2003.

[14] T. Sim and T. Kanade. Combining models and exemplars for face recognition: An illuminating example. *CVPR Wkshp Models vs Exemplars in Computer Vision*, 2001.

[15] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *AMFG*, pages 168–182, 2007.

[16] T. Vetter, M. J. Jones, and T. Poggio. A bootstrapping algorithm for learning linear models of object classes. *CVPR*, pages 40–46, 1997.

[17] P. Viola and M. Jones. Robust real-time face detection. *IJCV*, 57:137–154, 2004.

[18] H. Wang, S. Li, and Y. Wang. Generalized quotient image. In *CVPR*, 2004.

[19] Y. Wang, L. Zhang, Z. Liu, G. Hua, Z. Wen, Z. Zhang, and D. Samaras. Face relighting from a single image under arbitrary unknown lighting conditions. *IEEE PAMI*, 31(11):1968–1984, 2009.

[20] T. Weyrich, W. Matusik, H. Pfister, B. Bickel, C. Donner, C. Tu, J. McAndless, J. Lee, A. Ngan, H. W. Jensen, and M. Gross. Analysis of human faces using a measurement-based skin reflectance model. *ACM SIGGRAPH*, pages 1013–1024, 2006.

[21] L. Zhang and D. Samaras. Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics. *IEEE PAMI*, 28(3):351–363, 2006.

[22] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang. Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition. *ICCV*, pages 786–791, 2005.