

UMD Experiments with FRGC Data

Gaurav Aggarwal* Soma Biswas Rama Chellappa
Center for Automation Research
University of Maryland
College Park, MD 20742

Abstract

Although significant work has been done in the field of face recognition, the performance of state-of-the-art face recognition algorithms is not good enough to be effective in operational systems. Though most algorithms work well for controlled images, they are quite susceptible to changes in illumination and pose. Face Recognition Grand Challenge (FRGC) is an effort to examine such issues to suitably guide future research in the area. This paper describes the efforts made at UMD in this direction. We present our results on several experiments suggested in FRGC. We believe that though pattern classification techniques play an extremely significant role in automatic face recognition under controlled conditions, physical modeling is required to generalize across varying situations. Accordingly, we describe a generative approach to recognize faces across varying illumination. Unlike most current methods, our method does not ignore shadows. Instead we use them to our benefit by modeling attached shadows in our formulation.

1. Introduction

Humans use face as an important cue for identifying people. This makes automatic face recognition very important from the point of view of a wide range of commercial and law enforcement applications. Recent improvements in the accuracy of the face recognition algorithms in controlled scenarios has shifted the focus to more challenging tasks of achieving similar performance across illumination and pose variations. A detailed survey of various face recognition algorithms is presented in [14].

The goal of Face Recognition Grand Challenge (FRGC) is to improve the performance of still and 3D face recognition algorithms by an order of magnitude. Among the six experiments specified in FRGC [6], Experiment 4 involving faces imaged under varying illumination conditions is the most challenging. Accordingly, this paper focuses mainly on the issues involved with this experiment while briefly discussing the results obtained for other experiments.

Several researchers have attempted to obtain invariance to illumination by using image processing techniques like histogram equalization [10]. Some subspace based methods try to counter illumination variations by discarding the first few principal components [3]. These techniques do improve the accuracies of the respective algorithms but are usually ineffective in the case of a non-trivial change in illumination conditions.

The inability of such heuristics to handle illumination variation has led to the rise of generative (or analysis-by-synthesis) approaches for face recognition [4][16][11][5][9]. Broadly speaking, these techniques try to model the physical process of image formation by taking into consideration quantities like surface albedo, surface normals and illumination source direction. Though the recovery of shape and surface properties (reflectivity or albedo) from image(s) has been studied for a long time, its application to the problem of face recognition is fairly recent. An example in this category is the application of shape from shading (SFS) algorithms. SFS research typically assumes a constant albedo across an object which is usually not true and thus limits the use of the approach. Since then, there have been several advances which have led to the application of SFS for face recognition and rendering. Zhao *et al.* [13] present an SFS approach to recover both shape and albedo for a symmetric object from a single image. [11] uses singular value decomposition (SVD) to learn generative models of objects from a set of images taken under different, and unknown illuminations. Shashua *et al.* [9] perform recognition across varying illumination under an ideal-class assumption. All objects belonging to the ideal class are assumed to have the same shape. [5] uses illumination cone models for illumination-invariant face recognition. They require a small number of training images of each face under different illuminations to recover the shape and albedo of the face. Basri *et al.* [1] propose methods for recovering surface normals in a scene using images taken under general illumination conditions. Their work is based on [2], [7] which prove that the set of all Lambertian reflectance maps obtained with arbitrary distant illumination sources approximately lie in a 9D linear subspace. In [4], Blanz *et al.* perform face recognition across pose and illu-

*Partially supported by an NSF-ITR Grant 03-25119

mination by fitting a 3D morphable model to images. They use a set of textured 3D scans of heads for learning the model. [12] uses harmonic image exemplars to perform face recognition under varying lighting. Zhou *et al.* [16] generalize the traditional photometric approach to handle all the appearances of all the objects in a class. They impose a rank constraint on shapes and albedos in a class to separate the two from illumination effects using the factorization approach. Despite the advances made, most of the cited approaches have not been applied for the face recognition problem using a large dataset like the one provided in FRGC. This might be because many techniques require multiple, differently illuminated images of each face which are usually not present in most face datasets.

Though most of the approaches which use Lambertian model ignore the presence of shadows, there have been a few attempts to locate and reject shadows. Yuille *et al.* [11] treats shadows as outliers and remove them using robust statistics. They use a binary indicator variable to indicate shadows which is updated in an iterative manner using the current estimates of illumination source and shape-albedo vector. [16] uses a very similar approach to ignore shadows. They use pixel-wise reconstruction error to locate and thereby ignore shadows from the analysis. Note that these approaches do not incorporate shadows in the minimization framework, rather they detect and remove them to avoid the effects their presence can have on the error surface. In contrast, we model attached shadows in our method which reduces the ambiguity of the error surface. The hard non-linearity present in the Lambert’s law can account for the formation of attached shadows. Therefore, we do not ignore the nonlinearity as done in most previous generative approaches (probably to keep the formulation linear).

The paper is organized as follows: In section 2, we present results for Experiments 1-3 using Multiple-Exemplar Discriminant Analysis (MEDA) [15]. Section 3 evaluates the disadvantages of ignoring attached shadows as opposed to incorporating them in the formulation. The recognition algorithm is described in Section 3.3.

2. Multiple-Exemplar Discriminant Analysis

Experiments 1 and 2 in FRGC involve face images taken under controlled conditions. As there is no appreciable change in illumination/pose, one can deploy a suitable pattern classification technique to perform recognition. We use multiple-exemplar discriminant analysis proposed in [15] for these experiments. This is discussed briefly in this section.

Typical classification techniques like Linear Discriminant Analysis (LDA) represent each class by a single exemplar which is the sample mean of the class. In fact, LDA

makes an underlying assumption that each class follows a normal distribution with different mean but same covariance. This might not be a good assumption even for face images taken under controlled scenarios due to the inherent variations in faces and subtle changes in facial expressions. On the other hand MEDA represents each class by several exemplars. Moreover, instead of maximizing the between-class distance and minimizing the within-class distance as done in LDA, MEDA tries to maximize the between-class exemplar distance and minimize the distances between exemplars belonging to the same class. Mathematically, it can be shown as follows:

$$\Sigma_W = \sum_{i=1}^C \frac{1}{N_i^2} \sum_{j=1}^{N_i} \sum_{k=1}^{N_i} (x_j^i - x_k^i)(x_j^i - x_k^i)^T \quad (1)$$

where Σ_W is the within-class exemplar scatter, C is the number of classes, N_i is the number of exemplars in class i and x_j^i is the j th exemplar belonging to the i th class. Similarly, the between-class exemplar scatter can be written as:

$$\Sigma_B = \sum_{i=1}^C \sum_{j=1; j \neq i}^C \frac{1}{N_i N_j} \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} (x_k^i - x_l^j)(x_k^i - x_l^j)^T \quad (2)$$

Similar to LDA, the projection matrix W is computed by maximizing the criterion function

$$J_W = \frac{\det\{W^T \Sigma_B W\}}{\det\{W^T \Sigma_W W\}} \quad (3)$$

Given a test face y , its identity can be determined as

$$\arg \min_{i=1,2,\dots,N_i} \{|W^T(y - x^i)|^2\} \quad (4)$$

Please note that though the training requires multiple images per person, only one image per subject is needed in the gallery. Following sub-sections briefly describe results obtained using the FRGC data.

2.1 Experiment 1

This experiment involves recognizing a person from a single controlled test image given a gallery with one controlled still image per subject. In Version 1 of the FRGC data, the gallery consists of 152 subjects while there are 608 test images. There are 183 training images available with one or more images per person. To increase the number of images per person in the training data (as required in MEDA), we include mirror reflections of the training images. The correlation measure of the projection coefficients is taken as the similarity measure for recognition. Figure 1 shows the Cumulative Match Score (cms) plot obtained using this approach.

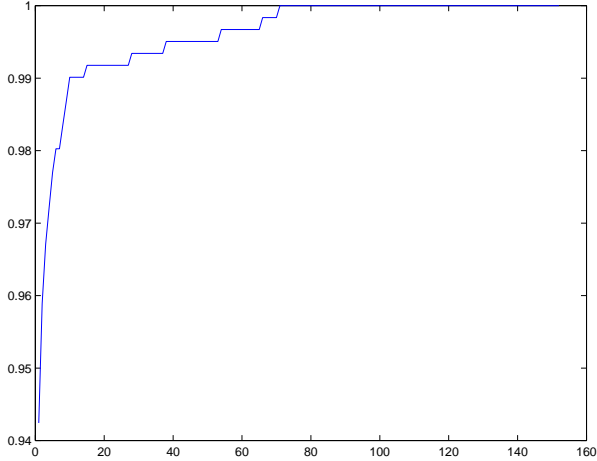


Figure 1: CMS curve obtained for Experiment 1 using MEDA for FRGC Ver 1 data

We used the same approach on FRGC Version 2 data which involves computing a similarity matrix of size 16028×16028 . Figures 2 and 3 show the Detection Error Trade-off (DET) curve and Receiver Operating Characteristic (ROC) curve, respectively, obtained using our approach.

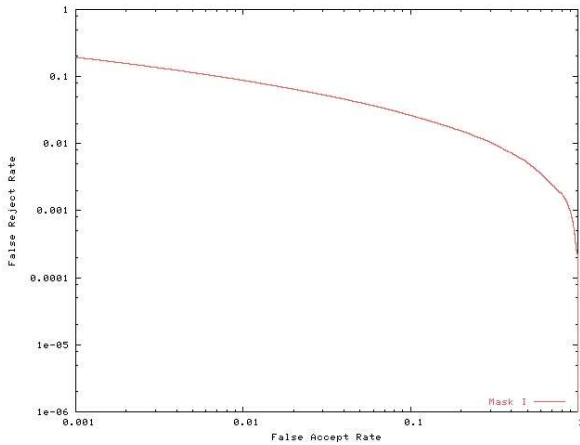


Figure 2: DET curve obtained for Experiment 1 using MEDA for FRGC Ver 2 data

2.2 Experiment 2

This experiment is similar to Experiment 1 other than the fact that we have four images per person both in gallery and probe sets to perform the recognition. The gallery consists of 152 subjects (152×4 images) while there are 608 test cases (608×4 images). As in Experiment 1, we use MEDA for recognition. The multiplicity of the image samples is ex-

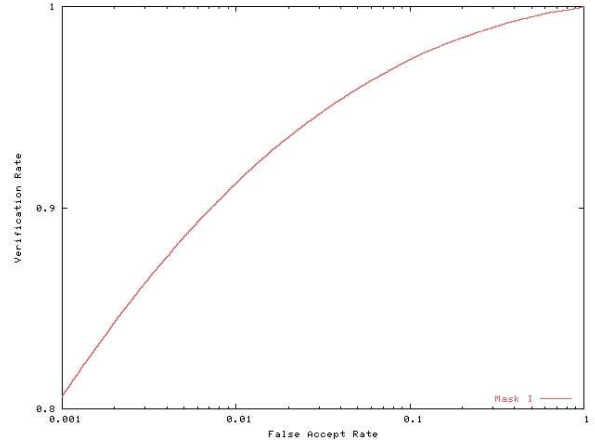


Figure 3: ROC curve obtained for Experiment 1 using MEDA for FRGC Ver 2 data

ploited with a simple polling scheme. As shown in Figure 4, the algorithm performs almost perfectly in this experiment.

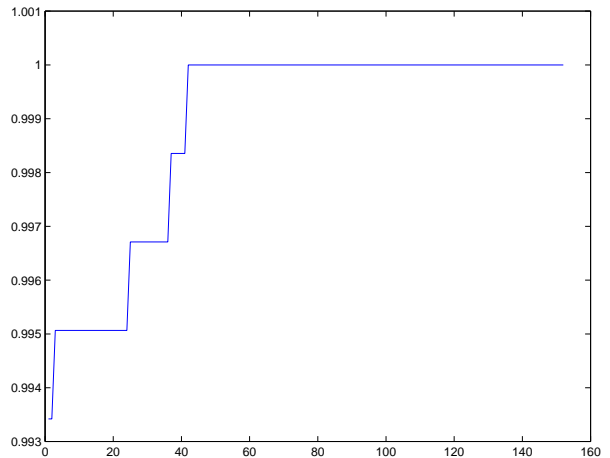


Figure 4: CMS plot obtained for Experiment 2 using MEDA

Figures 5 and 6 show the DET and ROC curves obtained, respectively, for Version 2 data. The similarity matrix in this case was 4007×4007 in size.

2.3 Experiment 3

This experiment involves 3D versus 3D face recognition. The provided data consists of both texture and shape information. We use MEDA to classify faces on the basis of texture and depth images separately. Though this is probably not the best way to exploit shape information, it works fine here due to the absence of any appreciable pose variation.

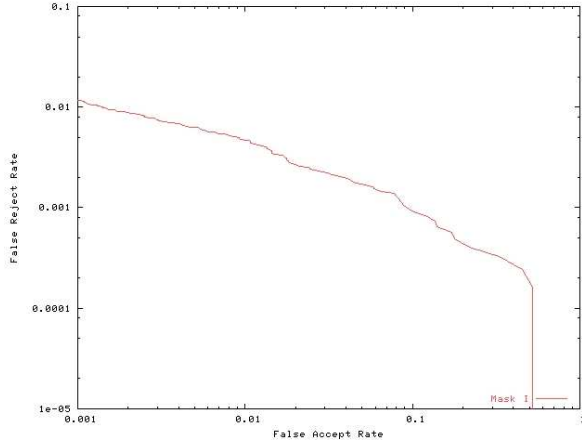


Figure 5: DET curve obtained for Experiment 2 using MEDA for FRGC Ver 2 data

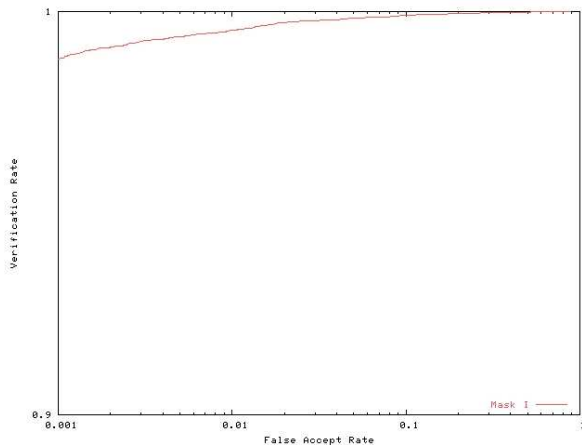


Figure 6: ROC curve obtained for Experiment 2 using MEDA for FRGC Ver 2 data

The two distance matrices obtained by applying MEDA to texture and depth images are fused simply as follows:

$$D_{fused} = D_{texture} \cdot * D_{shape} \quad (5)$$

where, $\cdot *$ is element-wise multiplication operator. The gallery consists of 152 subjects while there are 608 test images (texture and shape). Figure 7 shows the performance using this approach.

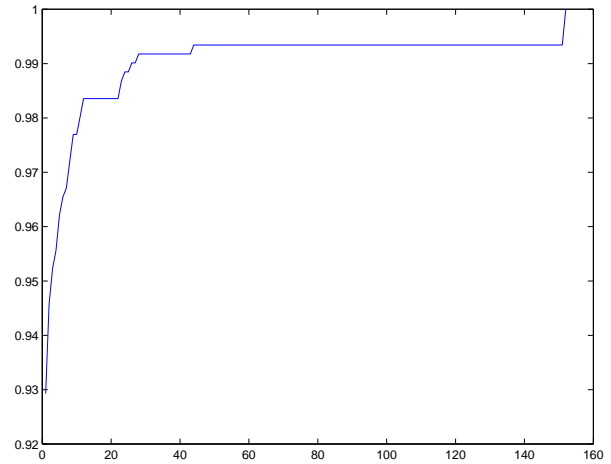


Figure 7: CMS plot obtained for Experiment 3 using MEDA

Though MEDA works well for controlled cases, it is not meant to handle variations in illumination etc. This motivates us to model the physical process of image formation. In the next section, we present a generative approach to handle arbitrary illumination variations like the one present in images for Experiment 4 in FRGC.

3. Face Recognition across Illumination Variations

We model a face as a Lambertian surface. Therefore, it is worthwhile to address the issue of the often ignored non-linearity in the Lambert's law. The diffuse component of the reflection of a surface is often modeled using the popular Lambert's Law. For example, Blanz and Vetter[4] use the following equation to model the diffuse component

$$L_{r,k} = R_k \cdot L_{r,dir} \cdot \langle n_k, l \rangle, \quad (6)$$

where R_k is the red component of the diffuse reflection coefficient, $L_{r,dir}$ is the red channel of the directed light, n_k is the surface normal at point k and l is the direction of the light source. Similarly, the generalized photometric stereo method [16] uses

$$h = \rho n^T s, \quad (7)$$

where ρ is the surface albedo, n is the surface normal, and s is the light source direction (multiplied by the intensity), as the rule for image formation. A close look at these equations reveals that the Lambert’s law is assumed to be linear in both these models. If used in its pure form, the non-linearity in the Lambert’s law would have made (7) to be

$$h = \rho \max(n^T s, 0) \quad (8)$$

Quite clearly, the linearity assumption is valid as long as the directed light source is in front of the surface for all its points. In general, objects like faces do not have all the surface points facing the illumination source which leads to the formation of shadows (commonly known as form/attached shadows). The cast and attached shadows are often ignored to keep the subspace of the observed images in a three [8] or with the addition of an ambient component [11], four dimensional linear subspace. Therefore, several generative approaches either ignore this non-linearity completely or ignore the shadow pixels. Here we present a simple illustration to highlight the role attached shadows can play.

3.1. Illustration 1

Suppose the goal is to estimate the illumination source from a single face image given the shape and albedo of the face. We explore three approaches for this task: the first approach ignores the non-linearity completely, the second one uses the linear rule but ignores the shadow pixels and the last one uses the Lambert’s Law in its pure form. The accuracy of the global minimum and its ambiguity on the error surface is taken as the criterion for the goodness of the method. The analytical expressions for the error function using the three options can be written as :

$$\text{Completely linear:} \quad \varepsilon(s) = \| h - \rho n^T s \|^2 \quad (9)$$

$$\text{Shadow pixels ignored:} \quad \varepsilon(s) = \| \tau \circ (h - \rho n^T s) \|^2 \quad (10)$$

$$\text{Non-linear rule:} \quad \varepsilon(s) = \| h - \max(\rho n^T s, 0) \|^2 \quad (11)$$

where, $\varepsilon(s)$ is the error with s as the illumination source direction, $h_{d \times 1}$ is the vectorized input image, ρ is the albedo vector, $n_{3 \times d}$ contains the surface normals, and $\tau_{d \times 1}$ is the shadow indicator vector which is 0 for the shadow pixels and 1 for the rest. Clearly, the linear method penalizes the correct illumination at the shadow pixels by having non-zero error values for those pixels. On the other hand, when shadows are ignored, the illuminations which produce wrong values for the shadow pixels do not get penalized there. As the set of all possible normals lies on the surface of a unit sphere, we use a sphere to display the computed error functions. Figure 8 shows the error surfaces for the three methods for a given face image. The lower the error is for a hypothesized illumination direction s , the darker

the surface looks at the corresponding point on the sphere. The global minimum is far from the true value using the first approach but is correct up to a discretization error for the second and third approaches. In fact, the second and third methods will always produce the same global minimum (assuming τ is correct), but the global minimum will always be less ambiguous in the third case because several wrong hypothesized illumination directions do not get penalized enough in the second approach due to the exclusion of the shadow pixels (Figure 8) .

3.2 The case of Multiple Light Sources

The above analysis implicitly assume that there is only one distant light source illuminating the face. Though the assumption is valid for datasets like PIE, it does not hold for most realistic scenarios. We now explore the impact of using the *linear* Lambert’s law for images illuminated by multiple light sources. Using the *linear* Lambert’s law, an image illuminated by k different light sources can be represented as:

$$h = \sum_{i=1}^k \rho n^T s_i = \rho n^T \sum_{i=1}^k s_i = \rho n^T s^* \quad (12)$$

where, $s^* = \sum_{i=1}^k s_i$. This shows that under the linear assumption, multiple light sources can be replaced by a suitably placed single light source without having any effect on the image. This is a bit counter-intuitive as can be seen in a simple two source scenario:

$$h = \rho n^T s_1 + \rho n^T s_2 \quad (13)$$

Now if $s_1 = -s_2$

$$h = \rho n^T (s_1 - s_2) = 0 \quad (14)$$

Thus the linear assumption can make the effect of light sources interfere in a destructive manner and give bizarre outcomes. Note that the negativity comes because of the direction and not because of the intensity of the light source. Quite clearly, the harm done by the linearity assumption is proportional to the angle subtended by the light sources at the surface.

Though the above discussion concludes that the Lambert’s law in its pure form is more appropriate than the other variants, it is only of academic interest if inclusion of the non-linearity does not improve the recognition results.

3.3 Recognition Algorithm

We extend the approach proposed in [16]. We first present a quick overview of the method and then highlight the impact of the non-linearity on the approach. Using Lambert’s law

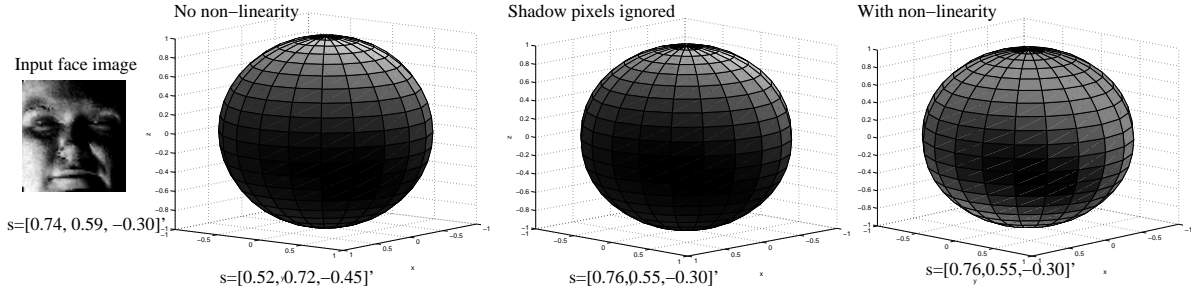


Figure 8: The error surfaces for the estimation of the light source direction given a face image of known shape and albedo. The three plots correspond to the three approaches described in Illustration 1. The lower the error is for a particular illumination direction, the darker the error sphere looks at the point corresponding to that direction. The true and estimated values of the illumination direction are listed along with the plots.

(the linear version), the intensity of a pixel can be written in terms of its albedo, shape and an illumination source as

$$h_i = (\rho n^T)_i s = t_i^T s \quad (15)$$

Suppose the image has d pixels, then

$$h_{d \times 1} = [h_1, h_2, \dots, h_d]^T = T_{d \times 3} s_{3 \times 1} \quad (16)$$

where, $T = [t_1 t_2 \dots t_d]^T$ is the object specific shape-albedo matrix. If T can be represented as a linear combination of m basis $T_i^T s$, we get

$$\begin{aligned} T &= f_1 T_1 + f_2 T_2 + \dots + f_m T_m \\ &= [T_1 T_2 \dots T_m] (f \otimes I_3) \\ &= W (f \otimes I_3) \end{aligned} \quad (17)$$

where $W = [T_1 T_2 \dots T_m]$ is the class specific shape-albedo matrix and I_3 is the 3×3 identity matrix. In this formulation, vector $f = [f_1 f_2 \dots f_m]^T$ is treated as the illumination-free identity vector. Given n different objects under different (and unknown) illumination conditions, Zhou *et al.* [16] estimate W (up to an invertible matrix) by solving a rank $3m$ problem using the factorization approach. The ambiguity is resolved using symmetry and integrability constraints. The interested reader is referred to [16] for the complete derivation. The average recognition results reported in [16] increased from 67% to 93%, when the W matrix is estimated from Vetter's 3D data [4] instead of the approach mentioned above.

Our main focus here is to highlight the importance of the non-linearity in the Lambert's law and to generalize the approach to handle multiple illumination sources. Therefore, we generate the shape-albedo matrix W using Vetter's 3D data for all our experiments. As opposed to [16], we take into account the inherent hard non-linearity present in the Lambert's law. Given shape-albedo matrix W , the recovery

of the identity vector f and illumination s can be posed as an optimization problem as follows:

$$\min_{f, s} \varepsilon(f, s) \equiv \|h - h_{rec}\|^2 + (1^T f - 1)^2 \quad (18)$$

$$\text{where, } h_{rec} = \sum_{i=1}^m f_i \max(T_i s, 0) \quad (19)$$

The second term is included in the error function to take care of scale ambiguity between f and s . Please note that s is not a unit vector as it contains the intensity of the illumination source also. The minimization is performed using an iterative approach, fixing f for optimizing ε w.r.t. s and fixing s for optimization w.r.t. f . In each iteration, f can be estimated by solving a linear least-squares (LS) problem but a non-linear LS solution is required to estimate s . The non-linear optimization is performed using the *lsqnonlin* function in MATLAB which is based on the interior-reflective Newton method. For most faces, the function value did not change much after 4-5 iterations. Therefore, the iterative optimization was always stopped after 5 iterations. The whole process took about 5-7 seconds per image on a normal desktop.

3.4 Experiments

We perform recognition experiments across illumination using the frontal faces from the PIE dataset. The correlation coefficient of the identity vectors is taken as the measure of the similarity between face images. Table 1 shows the recognition results obtained using this approach. Recognition is performed across illumination with images from one illumination condition from the PIE dataset forming the gallery set while images from another illumination condition forming the probe set. Each gallery/probe set contains one frontal image per subject taken in the presence of a particular light source (there are 68 subjects in each

Gallery	f_{08}	f_{09}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}	f_{17}	f_{20}	f_{21}	f_{22}	Average	Average from [16]
Probe														
f_{08}	-	100	100	100	96	97	81	72	50	100	97	84	90	88
f_{09}	100	-	100	100	100	99	97	96	75	100	100	97	97	94
f_{11}	100	100	-	100	100	97	94	78	63	100	99	94	94	93
f_{12}	100	100	100	-	100	100	100	99	90	100	100	100	99	97
f_{13}	97	100	100	100	-	100	100	100	96	100	100	100	99	99
f_{14}	94	100	100	100	100	-	100	100	99	100	100	100	99	99
f_{15}	88	97	97	100	100	100	-	100	100	97	100	100	98	96
f_{16}	74	90	81	93	100	100	100	-	100	76	97	100	93	89
f_{17}	59	74	63	87	99	99	100	100	-	71	94	100	87	75
f_{20}	99	100	100	100	100	99	96	82	71	-	100	97	95	93
f_{21}	97	100	100	100	100	100	100	99	96	100	-	100	99	98
f_{22}	93	100	99	100	100	100	100	100	99	99	100	-	99	98
Average	92	97	95	98	100	99	97	94	87	95	99	98	96	-
Average from [16]	89	93	92	96	98	99	96	91	80	91	96	98	-	93

Table 1: Recognition results on the PIE dataset. The averages from [16] are included for comparison. f_i denotes images taken with a particular flash ON as labeled in PIE. Each $(i, j)^{th}$ entry in the table shows the recognition rate obtained with the images from f_j as gallery while from f_i as probe.

gallery/probe set). Each entry in the table shows the recognition rate achieved for one such choice of gallery and probe set. The averages from [16] are shown for comparison. For fair comparison, we show results only across the illumination scenarios displayed in [16]. The recognition performance with the inclusion of the non-linearity in the Lambert’s law is almost always better or same. The overall average performance is up from 93% to 96%. The improvement is significant in cases involving difficult illumination conditions (with lots of shadows) like the flash f_{17} in the PIE dataset. This shows that though the estimation becomes slightly more difficult, the recognition rate improves with the inclusion of the non-linearity.

Though the algorithm works well on the PIE dataset as shown, it has certain limitations which need to be taken care of before it can be applied for more realistic images like the ones present in FRGC. First of all it makes single light source assumption which is not usually valid. Moreover, it expects the images to be accurately registered. (We have an extension of this algorithm that can handle arbitrary number of light sources which will be presented at International Conference on Computer Vision (ICCV), 2005) Once these issue are taken care of, we are confident that the method will be quite useful for practical applications.

Acknowledgements

We thank Dr. Shaohua Kevin Zhou and Narayanan Ramathanan for useful discussions and feedback on the work.

References

- [1] R. Basri and D. Jacobs. Photometric stereo with general, unknown lighting. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 374–381, 2001.
- [2] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25:218–233, 2003.
- [3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19:711–720, 1997.
- [4] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, September 2003.
- [5] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(6):643–660, June 2001.
- [6] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview

of the face recognition grand challenge. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

- [7] R. Ramamoorthi and P. Hanrahan. On the relationship between radiance and irradiance: determining the illumination from images of convex Lambertian object. *Journal of the Optical Society of America A*, pages 2448–2459, October 2001.
- [8] A. Shashua. On photometric issues in 3d visual recognition from a single 2d image. *International Journal of Computer Vision*, 21:99–122, 1997.
- [9] A. Shashua and T. R. 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, February 2001.
- [10] K. Sung and T. Poggio. Example-based learning for view-based human face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20:39–51, 1997.
- [11] A. L. Yuille, D. Snow, R. Epstein, and P. N. Belhumeur. Determining generative models of objects under varying illumination: Shape and albedo from multiple images using svd and integrability. *International Journal of Computer Vision*, 35(3):203–222, 1999.
- [12] L. Zhang and D. Samaras. Face recognition under variable lighting using harmonic image exemplars. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 19–25, 2003.
- [13] W. Zhao and R. Chellappa. Symmetric shape from shading using self-ratio image. *International Journal of Computer Vision*, 45(1):55–75, October 2001.
- [14] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.
- [15] S. Zhou and R. Chellappa. Multiple-exemplar discriminant analysis for face recognition. In *International Conference on Pattern Recognition(ICPR)*, Cambridge, UK, August 2004.
- [16] S. Zhou, R. Chellappa, and D. Jacobs. Characterization of human faces under illumination variations using rank, integrability, and symmetry constraints. In *European Conference on Computer Vision*, 2004.