Manifold Learning and its Applications:
Papers from the AAAI Fall Symposium (FS-09-04)

# Illumination Invariant Face Recognition on Nonlinear Manifolds

**Birkan Tunç**[†] **Muhittin Gökmen**[*]
[†] Istanbul Technical University, Informatics Institute
[*] Istanbul Technical University, Computer Engineering Department
İstanbul 34469, Turkey

**Abstract**

Face recognition under variable lighting conditions is recognized as one of the most problematic are of the recognition domain by various authors. Previous work suggested that image variations caused by parameters such as illumination, can be modeled by low dimensional subspaces. In this work, we propose a new scheme for recognition under a single variation. Using a generic manifold learning technique like LPP, we are able to construct coordinate systems for the underlying subspace with the help of an optimization step. We performed experiments with face recognition under changing illumination conditions.

## 1. Introduction

Appearance-based methods have been commonly used to achieve recognition in the literature (Zhao et al. 2003), (Turk and Pentland 1991). These methods simply try to find a low dimensional subspace in which face images can be separated easily depending on their identities. Original input images are $m \times n$ matrices that can be represented as a point in a $d = m \times n$ dimensional space. It is unreasonable to work in this d-dimensional space because of the high dimensionality. Methods like *eigenfaces* (Turk and Pentland 1991) try to find an orthogonal basis for the face space, and represent each face by coordinates in this space. Since only $k \ll d$ of the basis faces are used, the new dimensionality is much less than that of the original pixel space. In general, face images include large amount of information generated by various factors. Different images of same person may be altered extremely in different conditions depending on changes in these factors. Thus, appearance-based methods may suffer from uncontrolled environments.

To realize the robustness of the face recognition system, one must deal with the variations in the lighting conditions since illumination changes cause multimodality in the image space. Images of different people in the original data space are more closely located to each other than those of the same person under changing illumination conditions (Kim and Kittler 2005).

Illumination robust face recognition has been exhaustively studied by various authors (Shashua and Riklin-Raviv 2001), (Zhou and Chellappa 2003), (Georghiades, Belhumeur, and Kriegman 2001), (Zhang and Samaras 2006). Shashua and Riklin-Raviv proposed the quotient image as an lightning insensitive identity signature. The approach may fail when the probe image has an unpredictable shadow; however, it has the ability of recognizing probe images with illuminations different than that of gallery images. Technique only requires one gallery image per subject while using several images during the bootstrap phase. Zhou and Chellappa used extra constraints on albedo and surface normal to remove the shadow constraint. Georghiades et al. proposed an *illumination cone* model. They argued that set of images of an object in a fixed pose but under all lightning changes define a convex cone. They require a few images of a test identity to estimate its surface geometry and albedo map. After the estimation is completed, synthetic images with different illumination conditions can be rendered. All sets of Lambertian reflectance functions, which can be used to generate all kind of illumination conditions for Lambertian objects, were defined in (Basri and Jacobs 2003) and (Ramamoorthi 2002). They showed that by using only nine spherical harmonics, a wide variety of illumination can be approximated. In (Basri and Jacobs 2003), a methodology for recognition was also proposed. Zhang and Samaras exploited those spherical harmonics and represented excellent results for recognition.

Spherical harmonics approach may be stated as one of the most promising approaches to model the illumination variations (Zhang and Samaras 2006), (Lee, Ho, and Kriegman 2001). Yet, the main disadvantage of the approach is that it requires 3D scans of faces for the bootstrap phase. 3D imaging needs special equipments which can be very expensive. To get the coordinates in the harmonics space, they also need 2D images under various lightning conditions which is another problem with this approach. Most of the above techniques are specialized for illumination differences which may not be applied to other kind of parameter alterations like pose and expression. Usually, morphable models are employed to introduce viewpoint invariance.

Previous studies suggested that parameter changes like lighting constitute low dimensional subspaces (Chang, Hu, and Turk 2003), (Shashua, Levin, and Avidan 2002), (Seung and Lee 2000), (He et al. 2005). Manifold learning methods, that are superior to the linear dimensionality reduction techniques like PCA (Turk and Pentland 1991) when data lies on non-Euclidean geometries, can be utilized to work on these subspaces. He et al. developed an approximation to the Laplace-Beltrami operator defined on a manifold to construct linear maps which can be favored to get local coordinates for a nonlinear manifold.

In this paper, we propose a novel approach based on the nonlinear manifold embedding to define a linear subspace for illumination variations. This embedding based framework utilizes an optimization scheme to calculate the bases of the subspace. Since the optimization problem does not rely on any physical properties, the framework can also be used for different types of factors such as pose and expression. The main advantage of the approach is that it only requires 2D images for bootstrap. We do not even need implicit 3D rendering.

In the bootstrap step, reduced coordinates for the face images under different lighting conditions are calculated by using LPP (He et al. 2005). After finding coordinates corresponding to illumination conditions, subspace bases are calculated by an optimization step. For each individual, a different subspace is defined. Details for this procedure are given in the Section 2.

A Statistical model for subspace bases is learned in the bootstrap step and this model is used during training for basis recovery. For each training individual which may be or not in the bootstrap gallery, basis images are recovered by computing the maximum a posteriori (MAP) estimate. To calculate the illumination coordinates of a novel face, the mapping coming from LPP is applied. Basis recovery is explained in Section 3.

Testing is applied by calculating the nearest subspace to the novel test image. Each test image is tried to be synthesized in subspaces belonging to different individuals and the distance between the real image and the synthetic one is used as a decision measure. Testing step is represented in Section 4.

Some final thoughts and concluding remarks are accessible in Section 5.

## 2. Subspace Analysis for Illumination

In this section, details on how to construct the illumination subspace is proposed. We need 2D images of several people under various illumination conditions to get basis statistics. Some example images used in the bootstrap are given in Figure 1. First step is to calculate the reduced coordinates corresponding to illumination variations. After coordinates are calculated, we need to learn a basis set for each identity.

### 2.1 Subspace coordinates

Following the previous works, we use a nine dimensional subspace; although, a more exhaustive work can be done to choose the number of dimensions. LPP is a manifold learning algorithm which can be run in a supervised mode. Given a set of images under a number of variations, it determines a subspace in which those variations are modeled by a coordinate system. The output of LPP is a mapping $\mathbf{M}$ that is used to compute the reduced coordinates $\mathbf{c}$ of a novel image $\mathbf{x}$ by

Figure 1: Example images used during the bootstrap

$$\mathbf{c} = \mathbf{M}\mathbf{x}$$

When the supervised mode is applied, LPP provides a common mapping $\mathbf{M}$ for different identities. The columns of the matrix $\mathbf{M}$ are regarded as the orthonormal basis of the subspace. In this work, we do not treat $\mathbf{M}$ as a basis set but only a mapping for coordinates. Instead of having a common mapping $\mathbf{M}$, we assume unified coordinates for each illumination type and different basis sets for each identity. The method for computing bases is given in the next section.

Since the supervised mode is used, coordinates of different individuals for the same illumination conditions are very close. In another words, images of different identities under same lighting have approximate $\mathbf{c}$ values. Average coordinates for each illumination condition are taken as global coordinates; therefore, there is only one unique reduced coordinate for each illumination condition. Figure 2 illustrates reduced coordinates in 2-dimensional subspace.

### 2.2 Subspace bases

The second step is to calculate basis images for each identity. We used the following design: for each identity there is a separate 9-dimensional subspace; coordinate systems for these different subspaces are same. Since we have all required coordinates from LPP but no basis for those subspaces, we are challenged to solve an inverse problem: given the basis coefficients, determine the basis images which will result the minimum reconstruction error.

Consider we are given $N$ input images of a single person for $N$ different lighting conditions under a fixed pose. Assuming a 1-dimensional subspace, the total reconstruction error is defined as

$$\mathcal{E} = \sum_{i=1}^{N} \|\mathbf{x}_i - c_{1,i}\varphi_1\| \tag{1}$$

where $\mathbf{x}_i$ is the input image under $i^{th}$ lighting condition, $c_{1,i}$ is the first coordinate term of the $i^{th}$ image, and $\varphi_1$ is the first basis image we trying to calculate. To reduce the condition

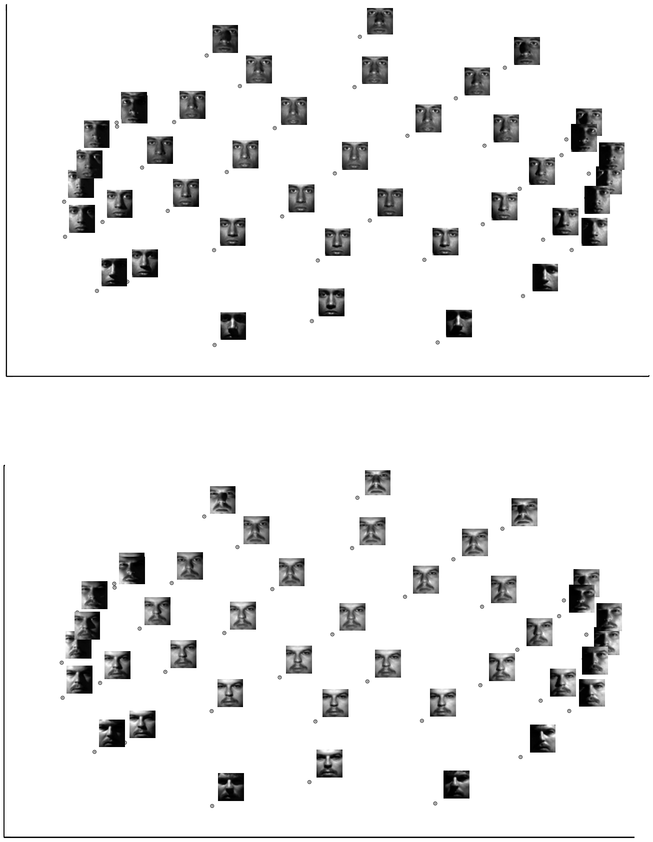

Figure 2: Global coordinates in 2-dimensional subspace for illumination variations. Coordinates of same lighting conditions through different identities are considered to be same

number related to the problem, we introduce a normalization constraint as $\varphi_1^T\varphi_1 = 1$. By re-arranging the terms and writing the minimization problem in the Lagrange Multiplier form, we get

$$\mathcal{L} = -2\mathbf{c}_1^T\mathbf{X}\varphi_1 + \varphi_1^T\varphi_1\mathbf{c}_1^T\mathbf{c}_1 + \lambda(\varphi_1^T\varphi_1 - 1) + \sum_{i=1}^{N}\mathbf{x}_i^T\mathbf{x}_i \tag{2}$$

here the matrix $\mathbf{X}$ includes image $\mathbf{x}_i$ as its $i^{th}$ row. $\mathbf{c}_1$ is the vector of the first coordinate terms, and $\lambda$ refers to the Lagrange multiplier. The last term $\sum_{i=1}^{N}\mathbf{x}_i^T\mathbf{x}_i$ can be omitted since it does not depend on the optimization variable $\varphi_1$. Minimization of this cost functional yields the following result.

$$\varphi_1 = \frac{\mathbf{X}^T\mathbf{c}_1}{\sqrt{\mathbf{c}_1^T\mathbf{X}\mathbf{X}^T\mathbf{c}_1}} \tag{3}$$

To calculate the second basis image, a similar formulation with an extra constraint $\varphi_1^T\varphi_2 = 0$ can be used. The minimization of the following functional

$$\mathcal{L} = \sum_{i=1}^{N}\|\mathbf{x}_i - c_{1,i}\varphi_1 - c_{2,i}\varphi_2\| + \lambda_1(\varphi_2^T\varphi_2 - 1) + \lambda_2(\varphi_2^T\varphi_1)$$

yields

$$\varphi_2 = \frac{\mathbf{Y}^T\mathbf{c}_2 - \alpha\varphi_1}{\sqrt{\mathbf{c}_2^T\mathbf{Y}\mathbf{Y}^T\mathbf{c}_2 - \alpha^2}}$$

where $\mathbf{Y} = \mathbf{X} - \mathbf{c}_1\varphi_1^T$ and $\alpha = \mathbf{c}_2^T\mathbf{Y}\varphi_1$. Following the same procedure, the $n^{th}$ basis becomes

$$\varphi_n = \frac{\mathbf{Y}^T\mathbf{c}_n - \sum_{i=1}^{n-1}\alpha_i\varphi_i}{\sqrt{\mathbf{c}_n^T\mathbf{Y}\mathbf{Y}^T\mathbf{c}_n - \sum_{i=1}^{n-1}\alpha_i^2}} \tag{4}$$

and similarly $\mathbf{Y} = \mathbf{X} - \sum_{i=1}^{n-1}\mathbf{c}_i\varphi_i^T$ while $\alpha_i = \mathbf{c}_n^T\mathbf{Y}\varphi_i$.

This formulation can be utilized to get basis images for each identity (a person in the bootstrap gallery). As mentioned before, for each identity, we get a different set of basis images to construct different subspaces. Since the coordinate vectors $\mathbf{c}_i$ are identical for each identity, these basis images reflect similar aspects of illumination variations through identities. Figure 3 represents examples of basis sets. After all bases are determined, statistics like mean vectors and covariance matrices are computed.

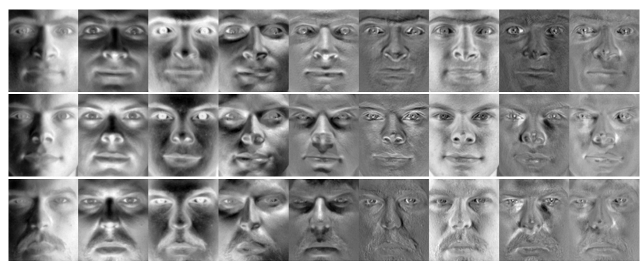

Figure 3: Example basis sets calculated by reconstruction error minimization

## 3. Recovery of Basis Images During Training

When we need to introduce a new person during training, the basis set for this individual is needed to be recovered. For basis recovery, we apply the same procedure defined in (Zhang and Samaras 2006) except for the part where basis coefficients are calculated. The governing equation for recovery is

$$\mathbf{x}(k) = \varphi(k)^T\mathbf{c} + e(k, \mathbf{c}) \tag{5}$$

which can be read as the pixel intensity $\mathbf{x}(k)$ at the pixel position $k$ can be generated by a linear combination of basis images plus an error term. One should be careful on realizing the meaning of the vectors $\varphi(k)$ and $\mathbf{c}$. Here $\varphi(k)$ is a 9-dimensional vector including intensities of nine basis images at the pixel position $k$. Similarly, $\mathbf{c}$ is another 9-dimensional vector corresponding the coordinates of the image $\mathbf{x}$ in the illumination subspace. The error term depends on the pixel position $k$ and the coordinates. Re-writing the equation in matrix form, we get $\mathbf{x} = \mathbf{B}\mathbf{c} + \mathbf{e}$ where $\mathbf{x}$ is d-dimensional image vector, $\mathbf{B}$ is the $(d \times 9)$-dimensional matrix including basis images as its columns, and $\mathbf{c}$ is again 9-dimensional coordinate vector. Every row of the matrix $\mathbf{B}$ corresponds to a vector $\varphi(k)$.

During the bootstrap, we compute mean vectors $\mu_\varphi(k)$ and covariance matrices $\mathbf{C}_\varphi(k)$ at each pixel position $k$ for bases calculated by the optimization scheme.

The error terms for images in the bootstrap gallery, can be calculated by

$$e(k, \mathbf{c}) = \mathbf{x}(k) - \varphi(k)^T \mathbf{c}$$

which is $\mathbf{c}$ dependent. For each $\mathbf{c}$ in the bootstrap gallery, statistics $\mu_e(k, \mathbf{c})$ and $\sigma_e^2(k, \mathbf{c})$ are computed through identities. There exists only a finite number of $\mathbf{c}$ vectors depending on the number of illumination conditions since we have a unified coordinate system.

The method for the basis recovery requires only one image for each novel identity, and this fact simplifies the construction of an identity database in a real life application. Given a novel training image, one can calculate the basis images corresponding to its identity by a *maximum a posteriori* (MAP) estimate based on equation (5).

### 3.1 MAP estimate for basis recovery

Considering we already have coordinates for the novel image and statistics for the error term, the MAP estimate for image bases can be written as

$$\varphi_{MAP}(k) = argmax_{\varphi(k)} \quad P(\varphi(k)|\mathbf{x}(k))$$

Using Bayes' rule we get

$$\varphi_{MAP}(k) = argmax_{\varphi(k)} \quad P(\mathbf{x}(k)|\varphi(k)) \times P(\varphi(k))$$

here the denominator term $P(\mathbf{x}(k))$ is omitted since it is constant. We assume both the basis vector $\varphi(k)$ and the error term $e(k, \mathbf{c})$ are Gaussians $N(\mu_\varphi(k), \mathbf{C}_\varphi(k))$ and $N(\mu_e(k, \mathbf{c}), \sigma_e^2(k, \mathbf{c}))$. Therefore, the probability $P(\mathbf{x}(k)|\varphi(k))$ is another Gaussian $N(\varphi(k)^T \mathbf{c} + \mu_e, \sigma_e^2)$. Then, MAP estimate becomes

$$\begin{aligned} \varphi_{MAP}(k) &= argmax_{\varphi(k)} \quad N(\varphi(k)^T \mathbf{c} + \mu_e, \sigma_e^2) \\ &\times \quad N(\mu_\varphi(k), \mathbf{C}_\varphi(k)) \end{aligned} \tag{6}$$

Finally, the MAP estimate for $\varphi(k)$ is the solution to the following set of linear equations

$$\mathbf{A}\varphi(k) = \mathbf{b} \tag{7}$$

where $\mathbf{A} = \frac{1}{\sigma_e^2}\mathbf{c}\mathbf{c}^T + \mathbf{C}_\varphi^{-1}$ and $\mathbf{b} = \frac{(\mathbf{x} - \mu_e)}{\sigma_e^2}\mathbf{c} + \mathbf{C}_\varphi^{-1}\mu_\varphi$

The main attribute that recovered bases should have is that they should be invariant to illumination changes. In other words, using two images of the same identity under different lighting conditions must produce same or at least very similar set of basis. By this attribute, we try to prevent the decrease on the recognition rate depending on what kind of image is used during training. This fact can be examined on Figure 4. During testing, we realized that the basis sets are almost invariant to lighting except some extreme cases. When lighting causes severe shadowing effects, recovery process only generates images with an average identity which disturbs the performance of recognition.

For the MAP estimate, we assume that the coordinates of the novel image are already known. The next section explain how we get this coordinates along with the error statistics.

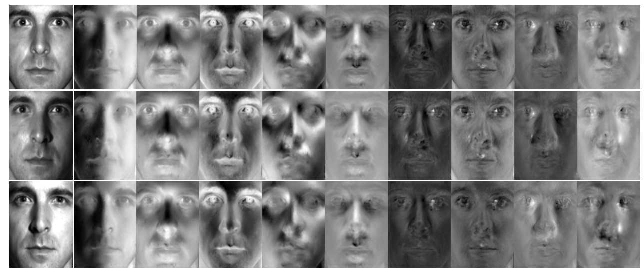

Figure 4: Recovered bases of same identity under variable lighting. The change in the basis depending on the lighting seems to be minimum

### 3.2 Coordinate estimation

Given a novel training image, the mapping $\mathbf{M}$ of LPP can be used to get coordinates. After $\mathbf{M}$ is applied, novel coordinates are determined by an weighted average in the subspace as follows:

$$\mathbf{c}_{new} = \frac{\sum_{i=1}^{kn} w_i \mathbf{c}_i}{\sum_{i=1}^{kn} w_i} \tag{8}$$

where the coordinate vectors $\mathbf{c}_i$ are from the bootstrap gallery. Weights $w_i$ are determined by $\|\mathbf{M}\mathbf{x}_{new} - \mathbf{c}_i\|$ where $\mathbf{x}_{new}$ is the training image. Having $\mathbf{M}$ in hand is a big advantage over previous works since it eliminates the need of calculating weights in image space as in (Zhang and Samaras 2006).

In equation (7), we also need $\mu_e(k, \mathbf{c}_{new})$ and $\sigma_e^2(k, \mathbf{c}_{new})$ which both depend on new coordinates. Same weights can be used to calculate those values as weighted averages of previous $\mu_e(k, \mathbf{c}_i)$ and $\sigma_e^2(k, \mathbf{c}_i)$ determined during bootstrap.

### 3.3 Recognition scheme

In the training, subspace bases are derived for each identity and stored in the database. Since there is no constraint for orthogonality of the basis images during recovery, recovered basis do not establish an orthogonal set. Therefore, we apply singular value decomposition (SVD) after the recovery.

When a test image is arrived, it is tried to synthesized in each subspace corresponding to different identities. Distance between synthesized image and the real one is used as a decision measure. The image synthesis is simply accomplished as

$$\mathbf{x}_s = \mathbf{Q}\mathbf{Q}^T\mathbf{x}_t$$

where $\mathbf{x}_t$ and $\mathbf{x}_s$ are testing and synthesized images, respectively. $\mathbf{Q}$ is the orthonormal basis set. The identity with the minimum distance $\|\mathbf{x}_s - \mathbf{x}_t\|$ is assigned as the recognition result.

## 4. Experimental Results

The Extended Yale Face Database B (Georghiades, Belhumeur, and Kriegman 2001), (Lee, Ho, and Kriegman 2005) is used as the bootstrap gallery. There are 28 people with 45 different illumination variations in this database. We used 41 types of illumination since there are several corruptions during image acquisition for the excluded 4 lighting conditions. Whole data set is processed with LPP to get unified coordinates for those 41 illumination type.

The basis images for 28 identities are computed using the optimization scheme defined in Section 2.The mean vectors $\mu_\varphi(k)$ and covariance matrices $\mathbf{C}_\varphi(k)$ are calculated along with statistics for the error term depending on each illumination type.

The training and the testing is applied using The Yale Face Database B which includes 10 people with again 45 lighting conditions. These 10 people are not included in the bootstrap gallery. This database is commonly used in the literature for testing purposes. There are 4 sets of images grouped according to lighting conditions. Four subsets cover lighting angles between $O^o - 12^o$, $13^o - 25^o$, $26^o - 50^o$, $51^o - 77^o$ respectively. The last subset includes severe conditions with extreme shadowing. Figure 5 shows some example images from each subset. We used all 45 illumination types during training and testing.

During training, we randomly select one image for each identity from subsets 1-2. Then the testing is applied with images from all subsets. We repeated this process 10 times, and averages were taken. Final error rates are given on Table 1 along with previous results from the literature. Our error rates are comparable with other methods considering their extensive bootstrap and/or training requirements. There are several important advantages that should be emphasized. Only our method and the harmonics approach can work with single images of training identities. Other methods require several images of a person which causes scalability problems in real life applications. The harmonics approach along with the illumination cone method need 3D images of the bootstrap gallery explicitly or implicitly. Our method only need 2D images under several lighting conditions. Coordinate values calculated by the manifold learning step are more reliable compared to kernel regression in the image space. Finally, both the harmonics and the cone approaches are constructed over some physical aspects of the lighting that makes them applicable only for illumination variations. We use a general subspace analysis that can be used for all kind of variations which are known to be lie on nonlinear manifolds.

Table 1: Face recognition error rates of previous methods compared with our method. Previous values are taken from (Zhang and Samaras 2006)

| Methods | Subset 1-2 | Subset 3 | Subset4 |
|---|---|---|---|
| Correlation | 0.0 | 23.3 | 73.6 |
| Eigenfaces | 0.0 | 25.8 | 75.7 |
| Linear Subspaces | 0.0 | 0.0 | 15.0 |
| Cones-attached | 0.0 | 0.0 | 8.6 |
| Cones-cast | 0.0 | 0.0 | 0.0 |
| 9PL | 0.0 | 0.0 | 2.8 |
| Spherical Harmonics | 0.0 | 0.3 | 3.1 |
| Our Method | 0.0 | 1.3 | 6.5 |

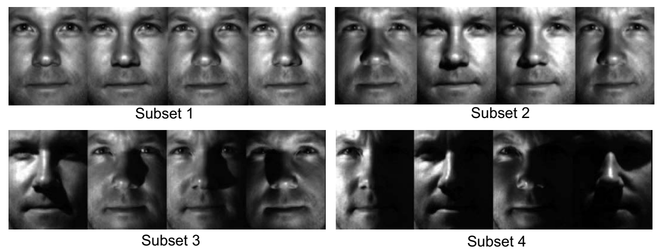


Figure 5: Some example images of four subsets used during testing. Illumination conditions in subset 4 cause severe shadowing

## 5. Conclusion

In this study, we propose a new scheme for face recognition under a single variation. The main advantage of the proposed approach is its ability of working with only one single image during training. The approach does not require 3D scans of faces to accomplish this task. Using a generic manifold learning technique like LPP, we are able to construct coordinate systems for the underlying subspace with the help of an optimization step.

We performed some tests with changing illumination conditions. Initial results are not better than the previous ones; however, considering their requirements like 3D scans and multiple training images, the results seem promising. The performance can be improved by introducing a more complex optimization scheme. We also examined that the size of bootstrap gallery greatly affects the error rates.

The proposed method can be used for any kind of variation which has an underlying non-linear manifold describing the behavior of the corresponding parameter.

## Acknowledgments

This work was partially supported by The Scientific and Technological Research Council of Turkey.

## References

Basri, R., and Jacobs, D. 2003. Lambertian reflectance and linear subspaces. *Pattern Analysis and Machine Intel-*

*ligence* 25(2):218–233.

Chang, Y.; Hu, C.; and Turk, M. 2003. Manifold of facial expression. In *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 28–35.

Georghiades, A.; Belhumeur, P.; and Kriegman, D. 2001. From few to many: illumination cone models for face recognition under variable lighting and pose. *Pattern Analysis and Machine Intelligence* 23(6):643–660.

He, X.; Yan, S.; Hu, Y.; Niyogi, P.; and Zhang, H.-J. 2005. Face recognition using laplacianfaces. *Pattern Analysis and Machine Intelligence* 27(3):328–340.

Kim, T.-K., and Kittler, J. 2005. Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. *Pattern Analysis and Machine Intelligence* 27(3):318–327.

Lee, K.-C.; Ho, J.; and Kriegman, D. 2001. Nine points of light: Acquiring subspaces for face recognition under variable lighting. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on* 1:519.

Lee, K.; Ho, J.; and Kriegman, D. 2005. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intelligence* 27(5):684–698.

Ramamoorthi, R. 2002. Analytic pca construction for theoretical analysis of lighting variability in images of a lambertian object. *Pattern Analysis and Machine Intelligence* 24(10):1322–1333.

Seung, S. H., and Lee, D. D. 2000. Cognition: The manifold ways of perception. *Science* 290(5500):2268–2269.

Shashua, A., and Riklin-Raviv, T. 2001. The quotient image: class-based re-rendering and recognition with varying illuminations. *Pattern Analysis and Machine Intelligence* 23(2):129–139.

Shashua, A.; Levin, A.; and Avidan, S. 2002. Manifold pursuit: A new approach to appearance based recognition. In *Procs of International Conference on Pattern Recognition (ICPR'02)*, volume 3.

Turk, M., and Pentland, A. 1991. Face recognition using eigenfaces. In *Procs. of the Computer Vision and Pattern Recognition (CVPR'91)*, 586–591.

Zhang, L., and Samaras, D. 2006. Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics. *Pattern Analysis and Machine Intelligence* 28(3):351–363.

Zhao, W.; Chellappa, R.; Rosenfeld, A.; and Phillips, P. J. 2003. Face recognition: A literature survey. *ACM Computing Surveys* 399–458.

Zhou, S. K., and Chellappa, R. 2003. Rank constrained recognition under unknown illuminations. *Analysis and Modeling of Faces and Gestures (AMFG'03)* 11–18.