Image-based Face Recognition: Issues and Methods [1]

WenYi Zhao
Sarnoff Corporation
201 Washington Road
Princeton, NJ 08543
Email: wzhao@sarnoff.com

Rama Chellappa
Center for Automation Research
University of Maryland
College Park, MD 20742-3275
Email: rama@cfar.umd.edu

Abstract

As one of the most successful applications of image analysis and understanding, face recognition has recently gained significant attention, especially during the past several years. There are at least two reasons for such a trend: the first is the wide range of commercial and law enforcement applications and the second is the availability of feasible technologies after 35 years of research. Moreover, recent significant advances in multimedia processing has also helped to advance the applications of face recognition technology. Among the diverse contents of multimedia, face objects are particularly important. For example, a database software capable of searching for face objects or a particular face object is very useful. Another example is a security system that is able to automatically track human objects, and report their IDs.

Though tracking and recognizing face objects is a routine task for humans, building such a system is still an active research. Among many proposed face recognition schemes, image based approaches are possibly the most promising ones. However, the 2D images/patterns of 3D face objects can dramatically change due to lighting and viewing variations. Hence, illumination and pose problems present significant obstacles for wide applications of this type of approaches. In this chapter, we first review existing methods extensively. And then we propose using a generic 3D model to enhance existing systems. More specifically, we use the 3D model to synthesize the so-called prototype image *from a given image acquired under different lighting and viewing conditions. The advantages of this approach are computational simplicity and system robustness which are essential for many real applications.*

1 Introduction

1.1 Background

As one of the most successful applications of image analysis and understanding, face recognition has recently gained significant attention, especially during the past several years. This is evidenced by the emergence of specific face recognition conferences such as AFGR [24] and AVBPA [23], and systematic empirical evaluation of face recognition techniques (FRT), including the FERET [22, 27] and XM2VTS protocols [18]. There are at least two reasons for such a trend: the first is the

[1] WenYi Zhao was at the Center for Automation Research while this work was performed.

wide range of commercial and law enforcement applications and the second is the availability of feasible technologies after 35 years of research.

The strong demand for user-friendly systems which can secure our assets and protect our privacy without losing our identity in a sea of numbers is obvious. At present, one needs a PIN to get cash from an ATM, a password for a computer, a dozen others to access the internet, and so on. Although extremely reliable methods of biometric personal identification exist, e.g., fingerprint analysis and retinal or iris scans, these methods have yet to gain acceptance by the general population. A personal identification system based on analysis of frontal or profile images of the face is non-intrusive and therefore user-friendly. Moreover, personal identity can often be ascertained without the client's assistance. In addition, the need for applying FRT has been boosted by recent advances in multimedia processing along with others such as IP (Internet Protocol) technologies.

In summary, there exist tremendous opportunities and great challenges for FRT. The challenge facing FRT is to perform well under severe conditions. For example, a personal verification system might need to process a low-quality face image which might be acquired using a cheap PC-camera and transferred over IP. Or the image capture happens in an uncontrolled environment with bad lighting etc. On the other hand, the opportunity lies in the fact that multimedia is almost ubiquitous and face objects are among the most important multimedia contents. For example, you may want to search for the video clips from home video archives where your baby shows a cute pose. Multimedia applications based on face objects include: content-based applications, human-machine interactive applications, and security related applications, etc. For example, a database software capable of searching for face objects or a particular face object is very useful. Another example is a smart video conference system that is able to automatically track objects and enhance their appearances.

1.2 Face recognition technology

A general statement of face recognition problem can be formulated as follows: Given still or video images of a scene, identify or verify one or more persons in the scene using a stored database of faces. Available collateral information such as race, age, gender, facial expression and speech may be used in narrowing the search (enhancing recognition). The solution of the problem involves segmentation of faces (face detection) from cluttered scenes, feature extraction from the face region, recognition or verification. In identification problems, the input to the system is an unknown face, and the system reports back the decided identity from a database of known individuals, whereas in verification problems, the system needs to confirm or reject the claimed identity of the input face.

Various applications of FRT range from static, controlled format photographs to uncontrolled video images, posing a wide range of different technical challenges and requiring an equally wide range of techniques from image processing, analysis, understanding and pattern recognition. One can broadly classify the challenges and techniques into two groups: static and dynamic/video matching. Even among these groups, significant differences exist, depending on the specific application. The differences are in terms of image quality, amount of background clutter (posing challenges to segmentation algorithms), the availability of a well defined matching criterion, and the nature, type and amount of input from a user. A rich repository of research literature exists after 35 years of research. Particularly the last five-year experiences the most active research activities and rapid advances. For a up-to-date critical survey of still- and video-based face recognition research, please see [35].

1.3 Chapter organization

In this chapter, we present efficient techniques for processing and recognizing face images. We assume an image-based baseline system since image-based approaches are possibly the most promising and practical ones. However, the 2D images/patterns of 3D face objects can change dramatically due to lighting and viewing variations. Hence, the illumination and pose problems present significant obstacles for wide applications of this type of approaches. To overcome these issues, we propose using a generic 3D model to enhance existing image-based systems. More specifically, we use a 3D model to synthesize the so-called *prototype image* from a given image acquired under different lighting and viewing conditions. This enhancement enables the existing systems to handle both illumination and pose problems specific to face recognition under the following assumption: just **one** image per face object is available.

In the following section, we first review and categorize existing methods proposed to address the pose problem and the illumination problem. We then propose using a generic 3D face model to enhance existing systems in Section 3. Instead of being a full 3D approach which directly uses accurate 3D information which is not easy to obtain in practice, this approach synthesizes a 2D *prototype image* from a given 2D image acquired under different lighting and viewing conditions with the aid of a 3D model. The prototype image is defined as the frontal view of an object under frontal lighting. For the purpose of completeness, a brief introduction to one particular baseline subspace LDA (Linear Discriminant Analysis) system [40, 39] has been included. In Section 4, we feed prototype images into the subspace LDA system to perform recognition. Finally we conclude our chapter in Section 5.

2 Existing Face Recognition Techniques

Automatic face recognition consists of subtasks in a sequential manner: face detection, face segmentation/normalization, and face recognition/verification. Many methods of face recognition have been proposed [35]. Basically they can be divided into holistic template matching based systems, geometrical local-feature-based schemes, and hybrid schemes. Even though schemes of all these types have been successfully applied to the task of face recognition, they do have certain advantages and disadvantages. Thus an appropriate scheme should be chosen based on the specific requirements of a given task. For example, the Elastic Bunch Graph Matching (EBGM) based system [21] has very good performance in general. However,

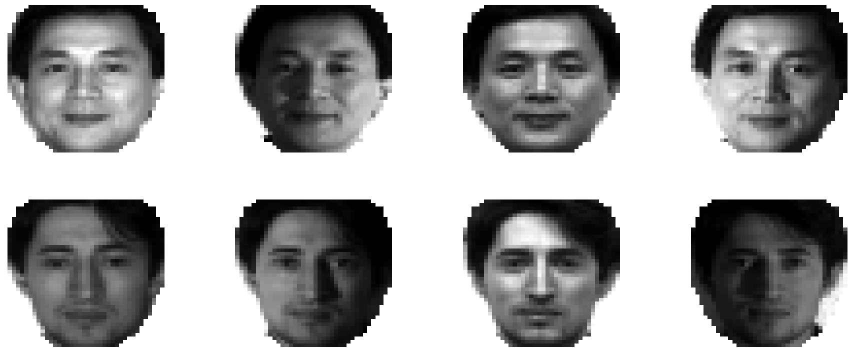


Figure 1: The Illumination problem.

it requires a large size image, e.g., 128×128. This severely restricts its possible application to video-based surveillance where the image size of the face area is very small. On the other hand, the subspace LDA system [40] works well with both large and small images, e.g., 96×84 or 24×21. It also has the best performance for verification tasks according to the most recent FERET test [27].

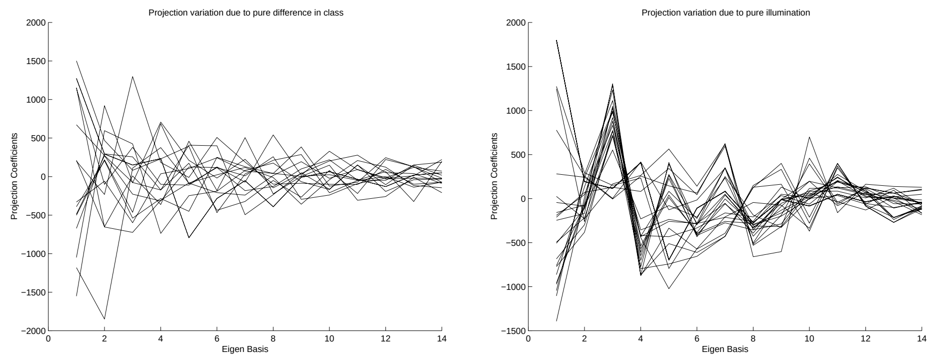
2.1 Two problems in face recognition

Despite the successes of many systems [19, 40, 21] based on the FERET test, many issues remain to be addressed. Among those issues the following two are prominent for most systems: 1) the illumination problem, 2) the pose problem.

The illumination problem is illustrated in Fig. 1 where the same face appears differently due to the change in lighting. More specifically, the changes induced by illumination could be larger than the differences between individuals, causing systems based on comparing images to misclassify the identity of the input image. This has been experimentally observed in [1] with a dataset of 25 individuals. We can also carry out some analysis. For example, the popular eigen-subspace projections used in many systems as features have been analyzed under illumination variation in [39]. The conclusions suggest that significant illumination changes cause dramatic changes in the projection coefficient vectors, and hence can seriously degrade the performance of subspace-based methods.

For comparison purposes, we plot the variations of projection coefficient vectors due to pure differences in class label (Fig 2(a)) along with the variations due to pure illumination change of the same class (Fig. 2(b)).

The pose problem is illustrated in Fig. 3 where the same face appears differently due to changes in viewing condition. Moreover, when illumination variation also appears in the face images, the task of face recognition becomes even more difficult (Fig. 3). In [39], an analysis and classification of various pose problems are performed using a reflectance model with varying albedo. Using such a model, the difficulty of the pose problem can be assessed and the efficacy of existing methods can be evaluated systematically. For example, the pose problem has been divided into three



(a) (b)

Figure 2: Change of projection vectors due to (a) class variation, and (b) illumination change.

Figure 3: The pose (and illumination) problem.

categories: 1) the simple case with small rotation angles, 2) the most commonly addressed case when there are a set of training image pairs (frontal and rotated images), and 3) the most difficult case when training image pairs are not available and illumination variations are present.

Difficulties due to illumination and pose variations have been documented in many evaluations of face recognition systems [22, 1, 39, 6]. An even more difficult case is the combined problem of pose and illumination variations. Unfortunately, this happens when face images are acquired in uncontrolled environments, for instance, in surveillance video clips. In the following, we examine the two problems in turn and review some existing approaches to these problems. More importantly, we point out the pros and cons of these methods so an appropriate approach can be applied to the specific task.

2.2 Solving the illumination problem

As a fundamental problem in image understanding literature, illumination problem is generally quite difficult and has been receiving consistent attentions. For face recognition, many good approaches have been proposed utilizing the domain knowledge, i.e., all faces belong to one face class. These approaches can be broadly divided into four types [39]: 1) heuristic methods including discarding the leading principal components, 2) image comparison methods where various image representations and distance measures are applied, 3) class-based methods where multiple images of one face under a fixed pose but different lighting conditions are available, and 4) model-based approaches where 3D models are employed.

2.2.1 Heuristic approaches

To handle the illumination problem, researchers have proposed various methods. Within the eigen-subspace domain, it has been suggested that by discarding the three most significant principal components, variations due to lighting can be reduced. And it has been experimentally verified in [6] that discarding the first few principal components seems to work reasonably well for images under variable lighting. However, in order to maintain system performance for normally lighted images, and improve performance for images acquired under varying illumination, we must assume that the first three principal components capture the variations only due to lighting. In [41], a heuristic method based on face symmetry is proposed to enhance system performance under different lighting.

2.2.2 Image comparison approaches

In [1], statistical approaches based on image comparison have been evaluated. The reviewed methods use different image representations and distance measures. The image representations used are: edge maps, derivatives of the gray level, images filtered with 2D Gabor-like functions, and a representation that combines a log function of the intensity with these representations. The different distance measures used are: point-wise distance, regional distance, affine-GL (gray level) distance, local affine-GL distance, and LOG point-wise distance. For more details about these methods and the evaluation database, please refer to [1]. One important conclusion drawn is that these representations are not sufficient by themselves to overcome the image variations. More recently, a new image comparison method proposed by Jacobs et al. [15] uses a measure robust to illumination change. Their method is based on the observation that the difference between two images of the same object is smaller than the difference between images of different objects. However this measure is not strictly illumination-invariant because the measure changes for a pair of images of the same object when the illumination changes.

2.2.3 Class-based approaches

With assumptions of Lambertian surfaces, no shadowing and three aligned images/faces acquired under different lighting conditions, a 3D linear illumination subspace for a person has been constructed in [5, 13, 29] for a fixed viewpoint. Thus under ideal assumptions, recognition based on the 3D linear illumination subspace is illumination-invariant. More recently, an illumination cone has been proposed as an effective method to handle illumination variations, including shadowing and multiple lighting sources [5, 11]. This method is an extension of the 3D linear subspace method [13, 29] and hence needs three aligned training images acquired under different lightings. One drawback to using this method is that we need more than three aligned images per person. More recently, a new method based on a quotient image has been introduced [26]. The advantage of this approach over existing similar approaches is that it only use a small set of sample images. This method assumes the same shape and different textures for faces of different individuals. An interesting energy function to be minimized is then formulated. Using this formulation, better results are rendered than using connectionist approaches.

2.2.4 Model-based approaches

In their paper [4], the authors suggest using Principal Component Analysis (PCA) as a tool for solving the parametric shape-from-shading problem, i.e., obtain the eigenhead approximation of a real 3D head after training on about 300 laser-scanned range data of real human heads. Though the ill-posed SFS problem is transformed into a parametric problem, they still assume constant albedo. This assumption does not hold for most real face images and we believe that this is one of the major reasons why most SFS algorithms fail on real face images.

2.3 Solving the pose problem

Researchers have proposed various methods to handle the rotation problem. Basically they can be divided into three classes [38]: 1) multiple images based methods when multiple images per person are available, 2) hybrid methods when multiple training images are available during training but only one database image per person is available during recognition, and 3) single image/shape based methods when no training is carried out. We have [8, 2, 31, 10] in the first type, and [32, 28, 17, 7] in the second type. Up to now, the second type of approach is the most popular one. The third approach does not seem to have received much attention.

2.3.1 Multi-image based approaches

Among the first class of approaches, one of the earliest is by Beymer [8], in which a template-based correlation matching scheme is proposed. In this work, pose estimation and face recognition are coupled in an iterative loop. For each hypothesized pose, the input image is aligned to database images corresponding to a selected pose.

The alignment is first carried out via a 2D affine transformation based on three key feature points (eyes and nose), and then optical flow is used to refine the alignment of each template. After this step, the correlation scores of all pairs of matching templates are used to perform recognition. The main restrictions of this method are 1) many images of different views per person are needed in the database, and 2) no lighting variations (pure texture mapping) or facial expressions are allowed, and finally 3) the computational cost is high since it is an iterative searching approach. More recently, an illumination-based image synthesis method [10] has been proposed as a potential method for robust face recognition handling both pose and illumination problems. This method is based on the well-known approach of *an illumination cone* [5] and can handle illumination variation quite well. To handle variations due to rotation, it needs to completely resolve the GBR (generalized-bas-relief) ambiguity when reconstructing the 3D shape.

2.3.2 Hybrid approaches

Numerous algorithms of the second type have been proposed and are by far the most popular ones. Possible reasons for this are: 1) it is probably the most successful and practical method up to now, 2) it utilizes prior class information. We review three representative methods here: the first one is the *linear class* based method [32], the second one is the graph matching based method [33], and the third is the view-based eigenface approach [25]. The image synthesis method in [32] is based on the assumption of linear 3D object classes and the extension of linearity to images which are 2D projections of the 3D objects. It extends the linear shape model from a representation based on feature points to full images of objects. To implement their method, a correspondence between images of the input object and a reference object is established using optical flow. Also correspondences between the reference image and other example images having the same pose are computed. Finally, the correspondence field for the input image is linearly decomposed into the correspondence fields for the examples. Compared to the parallel deformation scheme in [7], this method reduces the need to compute the correspondence between images of different poses. This method is extended in [28] to carry an additive error term for better synthesis. In [33], a robust face recognition scheme based on Elastic Bunch Graphic Matching (EBGM) is proposed. The authors basically assume a planar surface patch in each key feature point (landmark), and learn the transformation of 'jets' under face rotation. They demonstrate substantial improvements in face recognition under rotation. Also, their method is fully automatic, including face localization, landmark detection, and finally a flexible graph matching scheme. The drawback of this method is the requirement of accurate landmark localization which is not an easy task especially when illumination variations are present. The popular eigenface approach to face recognition has been extended to view-based eigenface method in order to achieve pose-invariant recognition [25]. This method explicitly codes the pose information by constructing an individual eigenface for each pose.

Despite their popularity, these methods have some common drawbacks: 1) they need many example images to cover all possible views, and 2) the illumination problem is separated from the pose problem.

2.3.3 Single image/shape based approaches

Finally, there is the third class of approaches which include low-level feature based methods, invariant feature based methods, and the 3D model based method. In [16], a Gabor wavelet based feature extraction method is proposed for face recognition and is robust to small-angle rotation. There are many papers on invariant features in the computer vision literature. To our knowledge, serious application of this technology to face recognition has not yet been explored. However, it is worthwhile to point out that some recent work on invariant methods based on images [3] may shed some light in this direction. For synthesizing face images under different appearances/lightings/expressions, 3D facial models have been explored in the computer graphics, computer vision and model-based coding communities. In these methods, face shape is usually represented either by a polygonal model or a mesh model which simulates tissue. However due to its complexity and computational cost, any serious attempt to apply this technology to face recognition has not yet been made except in [12].

3 3D Model Enhanced Face Recognition

Based on the assumption of one image per class available, solving the coupled illumination and pose problem is not a easy task. Most previous approaches either base on different assumptions or treat the two problems separately, hence it is necessary that we search for methods which can solve both problems simultaneously and efficiently. For example, the 3D model-based synthesis approaches used in computer graphics and coding communities are usually too complicated and expensive. Instead, we propose using a 3D model to enhance existing 2D approaches. To overcome the constant albedo issue [4] in modeling face objects, a varying albedo reflectance model is proposed [39]. Using this technique, we can convert any input images into prototype images, which are later fed into existing systems. In the recognition experiments carried out latter, we choose a particular baseline system based on its simplicity and efficiency : the subspace LDA system.

3.1 The subspace LDA system

The subspace LDA system in shown Fig. 4 [40]. It was proposed with the motivation of trying to solve the *generalization/overfitting* problem when performing face recognition on a large face dataset but with *very few* training face images available per class. Like existing methods, this method consists of two steps: first the face image is projected into a *face subspace* via Principal Component Analysis (PCA) where the subspace dimension is carefully chosen, and then the PCA projection vectors are projected into the LDA to construct a linear classifier in the subspace. Unlike other methods, the authors argue that the dimension of the face subspace is fixed (for a given training set) regardless of the image size as long as the image size surpasses the subspace dimensionality. The property of relative invariance of the subspace dimension enables the system to work with smaller face images without sacrificing performance. This claim is supported by experiments using normalized face images of different sizes to obtain different face subspaces [39]. The choice of such a fixed subspace dimension is mainly based on the characteristics of the eigenvectors instead of the eigenvalues. Such a choice of the subspace dimension enables the system to generate class-separable features via LDA from the *full* subspace representation. Hence the *generalization/overfitting* problem can be addressed to some extent. In addition, a weighted distance metric guided by the LDA eigenvalues was employed to improve the performance of the subspace LDA method. The improved performance of generalized recognition was demonstrated on FERET datasets [40] and the MPEG-7 content set in a proposal on robust face descriptor to MPEG-7 [37]. In [40], experiments were conducted to compare algorithm performance. The authors used a subset of FERET development dataset for training, FERET development dataset and other datasets for testing. The results show that subspace LDA (subLDA) is the best compared to PCA (with different eigenvectors: 15, 300, and 1000) and pure LDA algorithms. The detailed ranking of the algorithms is as follows: $subLDA300(85.2\%) > subLDA1000(80.8\%) > LDA(67.0\%) > PCA1000(58.3\%) > PCA300(57.4\%) > subLDA15(50.4\%) > PCA15(47.8\%)$. The numbers in the bracket are correct top-match scores.

The authors also reported a sensitivity test of the subspace LDA system. They took one original face image, and then electronically modified the image by creating occlusions, applying Gaussian blur, randomizing the pixel location, and adding an artificial background. Figure 5 shows electronically modified face images which were correctly identified.

3.2 A varying-albedo illumination model for face

In dealing with 2D-3D transformations, a physical illumination model is needed. There are many illumination models available, which can be broadly categorized into diffuse reflectance models and specular models [20]. Among these models, the Lambertian model is the most simple and popular one for diffuse reflectance and has been used extensively, especially in shape-from-shading (SFS) literature. With the assumption of Lambertian surface reflection and a single, distant light source, we have the following standard equations

$$I = \rho \cos \gamma,$$

or

$$I = \rho \frac{1 + pP_s + qQ_s}{\sqrt{1 + p^2 + q^2}\sqrt{1 + P_s^2 + Q_s^2}} \tag{1}$$

where γ is the angle between the outward normal to the surface $\vec{n} = (p, q, 1)$ and the negative illumination vector $-\vec{L} = (P_s, Q_s, 1)$ which represents the direction opposite to the distant light source, and ρ is the albedo. The surface orientation

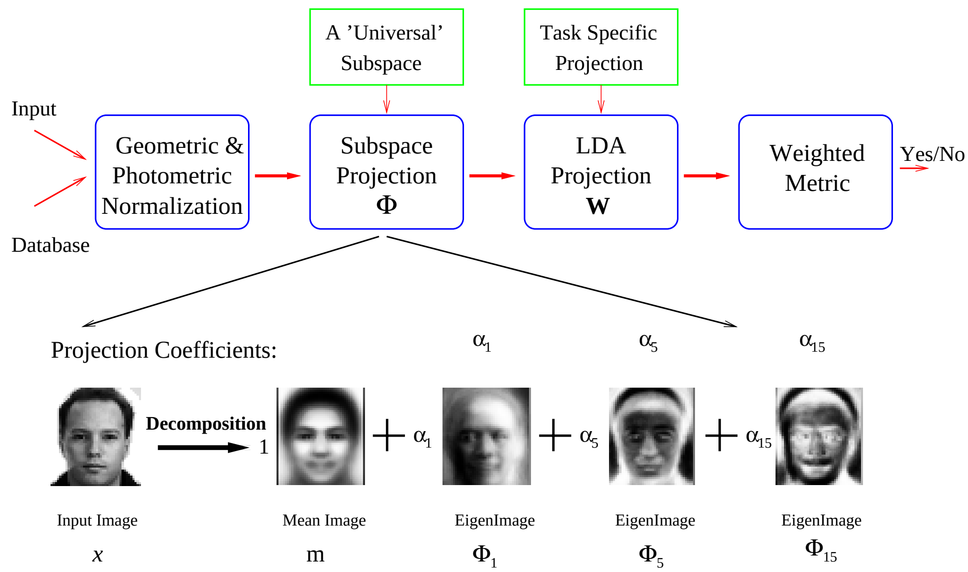


Figure 4: The subspace LDA face recognition system

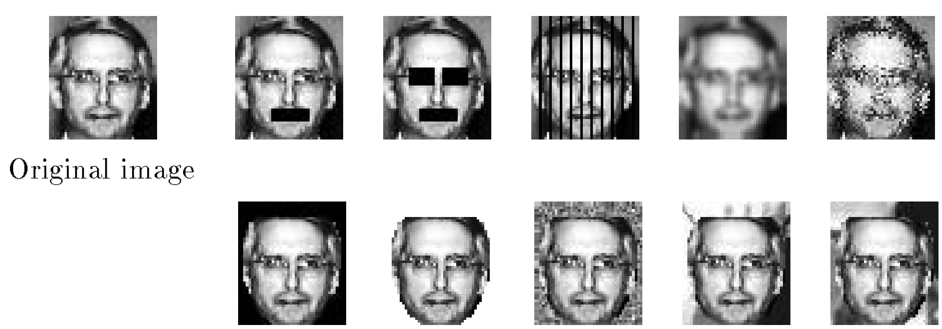


Figure 5: Electronically modified images which have been correctly identified.

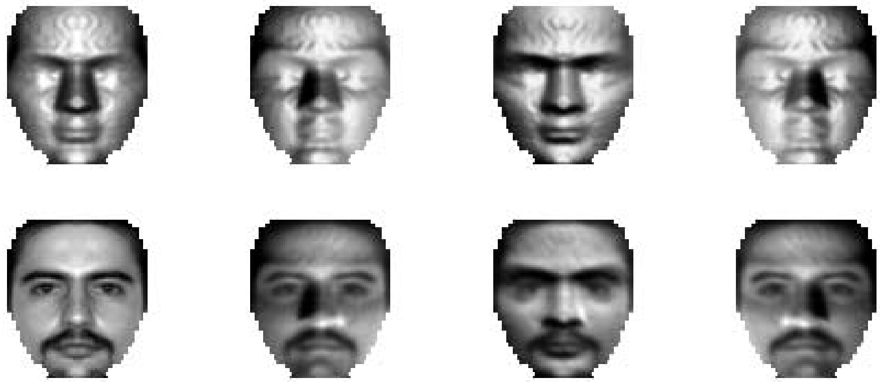


Figure 6: Image synthesis comparison under various lighting conditions. First row: constant-albedo Lambertian model; second row: varying-albedo Lambertian model.

can also been represented using two angles, slant and tilt. Similarly the light source can be represented by illuminant direction slant and tilt. The illuminant direction *slant* α is the angle between the negative $\vec{L}$ and the positive z-axis: $\alpha \in [0^0, 180^0]$; and the illuminant direction *tilt* τ is the angle between the negative $\vec{L}$ and the x-z plane: $\tau \in [-180^0, 180^0]$. To relate these angle terms to P_s and Q_s, we have $P_s = \tan\alpha\cos\tau$, $Q_s = \tan\alpha\sin\tau$.

Since we allow for arbitrary albedo, both (p, q) and ρ are functions of locations (x, y). However, we impose symmetry constraint for front-view face objects as follows (with an easily-understood coordinate system):

$$\begin{aligned} p[x,y] &= -p[-x,y] \\ q[x,y] &= q[-x,y], \end{aligned} \tag{2}$$

and

$$\rho[x,y] = \rho[-x,y]. \tag{3}$$

To show that varying-albedo Lambertian model is a good model, we compare the image synthesis results obtained using constant albedo and varying albedo assumptions. In Fig. 6, image synthesis results are compared one-by-one, i.e., a pair of images (in the same column) are synthesized exactly the same way except that one is using a constant-albedo model and the other is using a varying-albedo model. To obtain a realistic albedo we use a real face image and a generic 3D face model. To align this 3D model to the input image, we normalize both of them to the same size with two eye pairs are kept in the same fixed positions. Because the input image and model are not from the same object, we can see that some parts of the synthesized images are not perfect, for example, around nose region.

3.3 The Self-ratio image r_I

The concept of *self-ratio image* was initially introduced in [39] to address the additional parameter (albedo) issue. The idea of using two aligned images to construct a ratio has been explored by many researchers [15, 34]. But it was extended to a single image in [39]. Based on this concept, a new shape-from-shading (SFS) scheme has been developed. We can also it to help us to obtain the prototype images from given images.

Let us substitute (3,2) into the equations for $I[x,y]$ and $I[-x,y]$, and then add them, giving

$$I[x,y] + I[-x,y] = 2\rho \frac{1 + qQ_s}{\sqrt{1+p^2+q^2}\sqrt{1+P_s^2+Q_s^2}}. \tag{4}$$

Similarly we have

$$I[x,y] - I[-x,y] = 2\rho \frac{pP_s}{\sqrt{1+p^2+q^2}\sqrt{1+P_s^2+Q_s^2}}. \tag{5}$$

To simplify the notation, let us define $I_+[x,y] = \frac{I[x,y]+I[-x,y]}{2}$ and $I_-[x,y] = \frac{I[x,y]-I[-x,y]}{2}$. Then the *self-ratio image* r_I can be defined as

$$r_I[x,y] = \frac{I_-[x,y]}{I_+[x,y]} = \frac{pP_s}{1+qQ_s}. \tag{6}$$

Solving for shape information using (6) combined with (1) is called symmetric SFS [39]. The main result of symmetric SFS is the following theorem [39]:

Theorem 1 *If the symmetric depth z is a $\mathcal{C}^2$ surface and the symmetric albedo field is piece-wise constant, then both the solution for shape (p,q) and albedo ρ are unique except in some special conditions. Significantly, the unique global solution consists of unique local solutions at each point simultaneously obtained using the intensity information at that point and the surrounding local region under the assumption of a $\mathcal{C}^2$ surface.*

When symmetric objects are rotated, we can not directly apply the symmetric SFS theorem/algorithm. However, we can generate a virtual front-view from the given image. And the virtual front-view image can be obtained using the following relation between the rotated (in the x-z plane about the y-axis) image $I^\theta[x',y']$ and the original image $I[x,y]$ [39]

$$\begin{aligned} I^\theta[x',y'] &= 1_{z,\theta} I[x,y](\cos\theta - p[x,y]\sin\theta) \\ &\quad \times \tfrac{\tan(\theta+\theta_0)P_s + \frac{q\cos(\theta_0)}{\cos(\theta+\theta_0)}Q_s + 1}{pP_s + qQ_s + 1}, \end{aligned} \tag{7}$$

where $\tan\theta_0 = p[x,y]$ and $1_{z,\theta}$ is the indicator function indicating possible occlusion determined by the shape and rotation angle. After the virtual front-view is obtained, the symmetric SFS algorithm can then be applied. Of course, this is useful only if we can solve the correspondence problem: $[x,y] \rightarrow [x',y']$. In the following section, we propose using a 3D model to obtain an approximate solution to this problem.

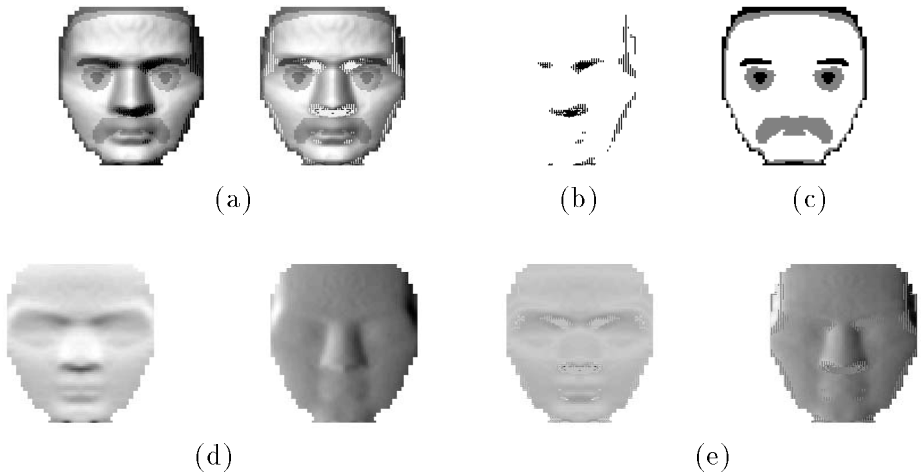


Figure 7: Face shape and image reconstruction results using symmetric SFS: the piece-wise constant albedo case. The plots in the left-half of the first row are input and reconstructed images (a). Plot (b) represents the shadow map in the input images which explains the holes in the reconstructed images. Plot (c) is the recovered albedo filed. The plots in the second row are the true partial derivatives (d) and recovered partial derivatives (e).

3.4 Using a generic 3D face shape

In theory, we can apply symmetric SFS to recover the complete 3D shape of a symmetric object. And we have proposed simple algorithms which work well on objects with complex/face shape and piece-wise albedo field. In Fig. 7 we plot the input and reconstructed face images, partial derivatives side-by-side for the piece-wise constant albedo case. However we still have difficulties in recovering the case when both the shape and albedo are arbitrary. Moreover, there are many practical issues that prevents us applying it for face recognition: 1) the unique solution might be very sensitive to possible violations of the assumptions such as $\mathcal{C}^2$ surface (possibly caused by digitizing the surface), 2) the solution might be sensitive to noise in measurement I (and hence r_I), and 3) the solution might be sensitive to possible violations of single light source assumption. To be more practical for face recognition, we propose using a simple 3D model to by-pass this 2D-to-3D process. This technique has been successfully applied in [36] to address pure illumination problem with pose fixed. This technique can also be extended for rotated face images [38].

Front-view Case Let us write the image equation for the prototype image I_p with $\alpha = 0$:

$$I_p[x, y] = \rho \frac{1}{\sqrt{1 + p^2 + q^2}}. \tag{8}$$

Comparing (4) and (8), we obtain

$$I_p[x,y] = \frac{K}{2(1+qQ_s)}(I[x,y]+I[-x,y]), \tag{9}$$

where K is a constant equal to $\sqrt{1+P_s^2+Q_s^2}$. This simple equation directly relates the prototype image I_p to $I[x,y]+I[x,-y]$ which is already available. It is worthwhile to point out that this *direct computation* of I_p from I offers the following advantages over the two-step procedure which first uses SFS to recover 3D information and then synthesizes new 2D images:

- There is no need to recover the varying albedo $\rho[x,y]$.
- There is no need to recover the *full* shape gradients (p,q).

The only parameter that needs to be recovered is the partial shape information q and we approximate this value with the partial derivative of a 3D face model. To guarantee good synthesis quality, we use the self-ratio image equation (6) as a consistency checking tool.

Rotated Case Combining (7) and (9), we have a direct relation between a rotated (in the x-z plane about the y-axis) image $I^\theta[x',y']$ and the prototype image I_p [38]

$$\begin{aligned} I^\theta[x',y'] &= 1_{z,\theta} I_p[x,y](\cos\theta - p[x,y]\sin\theta)\frac{1}{\sqrt{1+P_s^2+Q_s^2}} \\ &\times[\tan(\theta+\theta_0)P_s + \tfrac{q\cos(\theta_0)}{\cos(\theta+\theta_0)}Q_s + 1], \end{aligned} \tag{10}$$

To actually apply the above techniques for face recognition, we need the illumination direction of the light source and the face pose. In this way, we do not need to synthesize *all* views and illuminations in our database in order to recognize input images under various viewing and illumination conditions. Instead, we can synthesize the prototype view defined in the database from an input image acquired under different views and illumination directions.

3.5 Light source and pose Estimation

Frontal-view case Many source-from-shading algorithms are available, but we found that none of them work well for both the tilt and slant angles [39]. Instead, we propose a new model-based symmetric source-from-shading algorithm [38]. Basically we can formulate a minimization problem as

$$(\alpha^*,\tau^*) = \arg_{\alpha,\tau}\min(r_{I_{M_F}}(\alpha,\tau)) - r_I)^2. \tag{11}$$

where r_I is the self-ratio image, and $r_{I_{M_F}}$ is the self-ratio image generated from the 3D face model M_F given hypothesized α and τ. One advantage of using a 3D face model is that we can take into account both attached-shadow and cast-shadow effects, which are not utilized in the traditional statistics-based methods.

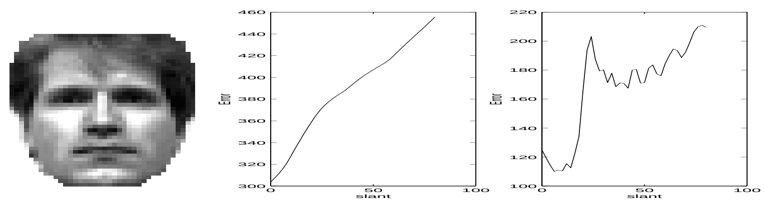


Figure 8: Comparison of model-based source-from-shading algorithms. The correct slant value was recovered using the algorithm (11) (right figure), while it was missed using algorithm (12) (middle figure).

Compared to other model-based methods[4], this method produces better results since it adopts a better model. Similarly, other methods can be formulated as a minimization problem

$$(\alpha^*, \tau^*) = \arg_{\alpha,\tau} \min(I_M(\alpha, \tau)) - I)^2. \tag{12}$$

where I is the input image, and I_M is the image generated from a 3D generic shape M based on Lambertian model (1) with constant albedo given hypothesized α and τ. For a simple comparison of these two model-based methods, we ran both these algorithms on real face images. In Fig. 8, we plot one face image along with the error-versus-slant curve for each method. As can be seen, the correct (subjective judgment) value of slant (8^0) has been recovered by the symmetric method (11). However, it is missed using (12). This new symmetric source-from-shading method has been successfully applied to more than 150 real face images as the pre-processing step prior to illumination-normalization for face recognition [36].

Rotated Case Most existing face pose estimation algorithms use some prior class knowledge, that is all face object are similar. Instead of using 2D example images, we propose using a simple 3D model to estimate the pose. Further, to incorporate the estimation of the light source we formulate the following problem

$$(\theta^*, \alpha^*, \tau^*) = \arg_{\theta,\alpha,\tau} \min[I^R_{M_F}(\theta, \alpha, \tau) - I^R]^2. \tag{13}$$

However, such formulations ignore the reality of a varying albedo. To better address this problem, the *self-ratio* image is used. However, in order to apply this method to rotated images of symmetric objects, we need additional processing. Using (7), we can formulate a new estimation problem:

$$(\theta^*, \alpha^*, \tau^*) = \arg_{\theta,\alpha,\tau} \min[r_{I_{M_F}}(\alpha, \tau) - r_{I^F}(\theta, \alpha, \tau)]^2, \tag{14}$$

where $r_{I^F(\theta,\alpha,\tau)}$ is the self-ratio image for the virtual frontal view generated from the original image I_R via image warping and texture mapping using (7).

4 Experiments

4.1 Shadow and implementation issues

One important issue we have not discussed in detail is the attached-shadow and cast-shadow problem. By definition, attached-shadow points are those where the image intensities are set to zero because $(1 + pP_s + qQ_s) \leq 0$. A cast-shadow is the shadow cast by the object itself. It has been shown in [39] that the shadow points can still be utilized in both source estimation and image rendering. For example in the case of source estimation, one advantage of using a 3D face model is that we can take into account both attached-shadow and cast-shadow effects, which are not utilized in the traditional statistics-based methods. However, these points contribute significantly and correctly to the computation of slant and tilt angles. Hence the model-based method can produce a more accurate estimate if the 3D face model is a good approximation to the real 3D face shape.

In addition to these shadow points, we need to single out the "bad" points, or outliers in statistical terms, for stable source estimation and prototype image rendering. This is because we need to compute the self-ratio image which may be sensitive to image noise. Let us denote the set of all "bad" points by $\mathcal{B}$; at these points the values cannot be used. From a robust statistics point of view, these "bad" points are outliers. Hence our policy for handling these outliers is to reject them and mark their locations. We then use values computed at good points to interpolate/extrapolate at the marked bad points. Many interpolation methods are available such as nearest-neighbor interpolation, polynomial interpolation, spline interpolation etc. Since we may have an irregular structure of good points, we use triangle-based methods. For example, to detect these bad points in the process of computing the prototype images, we employ the *consistency check*

$$\mathcal{B} = \{I[x, y] \mid \ |r_I - \frac{p_{M_F} P_s}{1 + q_{M_F} Q_s}| > \epsilon\}. \tag{15}$$

4.2 Solving the illumination problem

4.2.1 Rendering prototype images

The faces we used in our experiments are from the FERET, Yale and Weizmann databases. The Yale database contains 15 persons including four images obtained under different illuminations. The Weizmann database contains 24 persons also including four images obtained under different illuminations. We have applied our light-source estimation and direct prototype image rendering method to more than 150 face images from the Yale and Weizmann databases. Though the purpose of rendering prototype images is to improve the recognition performance, we would like to visualize the quality of the rendered images and compare them to the images obtained using a local SFS algorithm [30]. These results (Fig. 9) clearly indicate the superior quality of the prototype images rendered by our approach. More rendered prototype images using only the direct computation are plotted in Fig. 10.

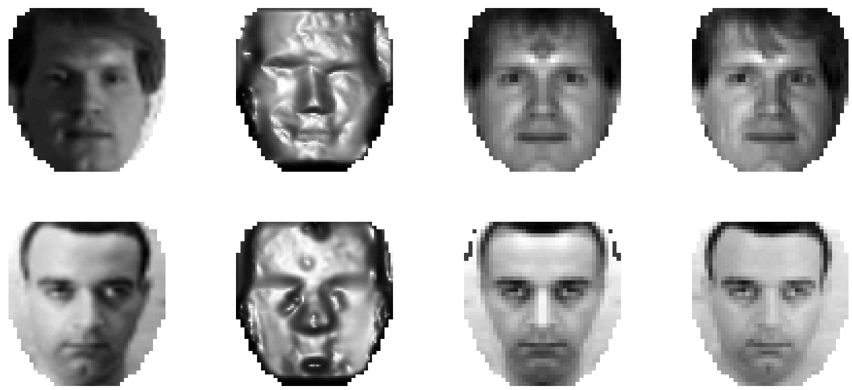


Figure 9: Image rendering comparison.

4.2.2 Enhancing face recognition

In this experiment we demonstrate that the generalized/predictive recognition rate of subspace LDA can be greatly enhanced. We conducted two independent experiments on the Yale and Weizmann databases. For the Yale database, we have a testing database composed of a gallery set containing 486 images from several face databases, including 15 (one image per class) from the Yale database, and a probe set containing 60 images also from the Yale database. For the Weizmann database, we have a testing database composed of a gallery set containing 495 images from several face databases, including 24 (one image per class) from the Weizmann database, and a probe set containing 96 images from the same database. Figure 11 shows the significant improvement in performance using the prototype images in both databases.

4.3 Solving the pose problem

Currently, light source and pose estimation (14) is not carried out. Instead very rough pose information is given manually. The light source is also assumed to be frontal though in fact it may be not. Basically we have only experimented on model-based pure image warping and plan to implement full SFS-based view synthesis in near future. However, as we have shown in [38], this is a good approach for eigen-subspace based method even when Lambertian model and frontal lighting are assumed.

The data set we used here is drawn from FERET and Stirling databases [40]. To compare, the quality of this dataset is lower than the one reported in [32] (which is also used in several subsequent works on face recognition such as [26]) and the size of normalized images we are using is much smaller than those in [32] and other. There are 108 pairs of face images: front view and quarter-profile view, all normalized to the size of 48×42 w.r.t. the two eyes. The poses of these faces are not quite

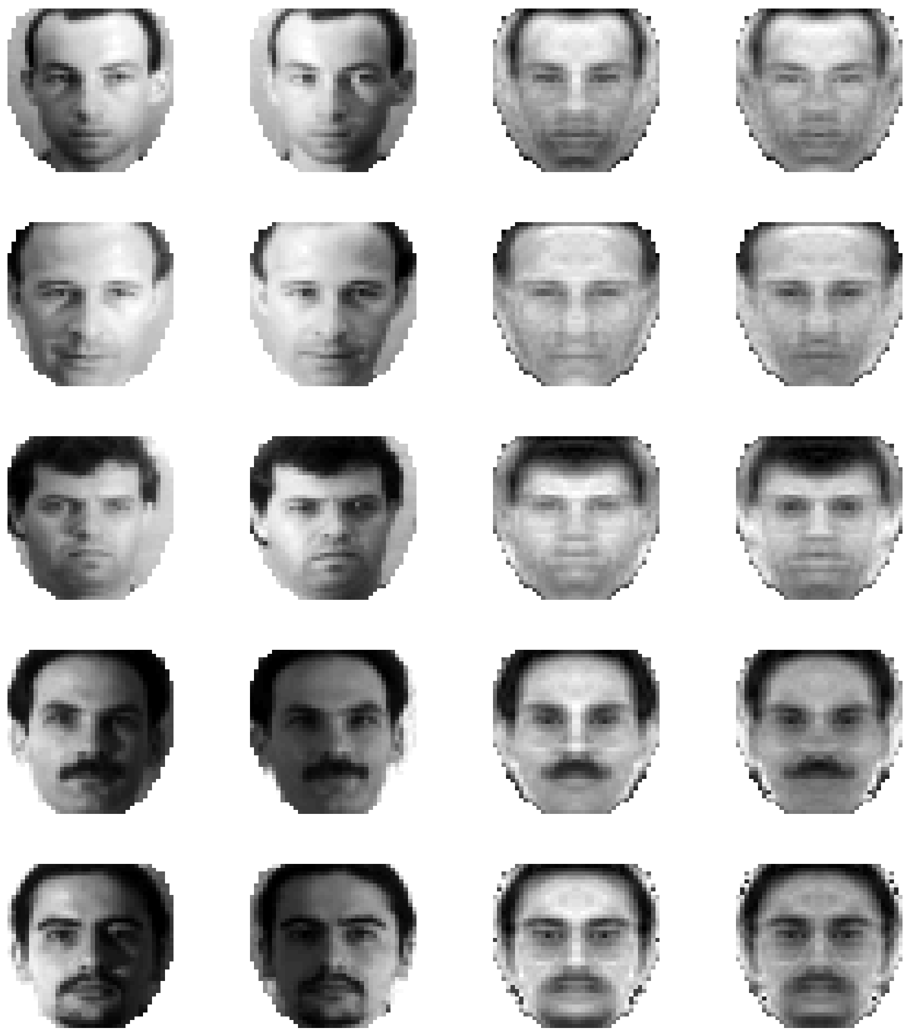
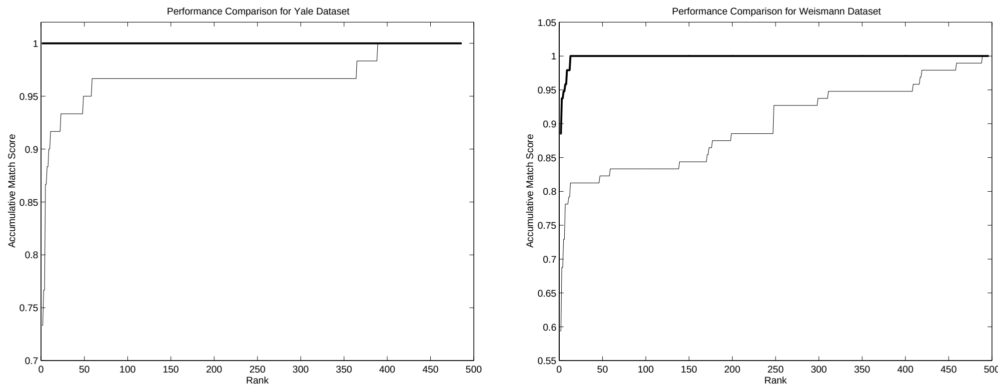


Figure 10: Rendering the prototype image. The images with various lighting are in the first two columns, while the corresponding prototype images are shown in the last two columns respectively.



(a) (b)

Figure 11: Enhancing the subspace LDA. The thin lines represent the cumulative scores when applying the existing subspace LDA to the original images, while the thick lines represent the scores when applying it to the prototype images. The curves in (a) are for the Yale face database, the curves in (b) are for the Weizmann database.

consistent, and we only apply a unique rotation angle picked manually to all images.

4.3.1 Visualization of image synthesis

As we mentioned earlier, the poses of these faces are not consistent and only one unique rotation angle is chosen for all the images. Hence some of the synthesis results are good (the first three columns in Fig. 12) if the actual rotation angle agrees with the preset value, and some are bad (the last three columns in Fig. 12).

4.3.2 Comparison of recognition results

To test and compare the efficacy of various methods for robust face recognition, we have tried two subspace LDA methods: I. Subspace LDA [40] on the original images, and II. Subspace LDA on the synthesized frontal images.

As mentioned before, the database we used have 108 pairs of images, of which only about 42 pairs are good images in terms of the correctness of rotation angle we manually picked (refer to Fig. 12). We use all the frontal views as the database, and all the rotated images as the testing images. We report the recognition performances of subspace LDA on the original images and on the virtual frontal images in Table 1. Some conclusions can be drawn here: using the virtual frontal views the performance of subspace LDA, which does not have the generalization problem and does not need retraining of the subspace and the LDA classifier, can be improved, and the extent of the improvement depends on the quality of the virtual views.

Figure 12: Some images used in the database. The first row are the rotated views, the second row are the synthesized frontal views, and the third view are the real frontal views. The first three columns are good image pairs, while the last three columns are bad pairs.

Method	I	II
Score (%)	39.8/61.9	46.3/66.7

Table 1: Performance comparison of subspace LDA on original images (I) and on virtual images (II). The scores on the right are for the 42 good images, and the scores on the left are for all 108 images.

5 Conclusion

We have proposed simple and efficient techniques for processing and recognizing face objects. The characteristics of these techniques are very suitable for many applications. We first identified two key issues in the face recognition literature: the illumination and pose problems. We then examined existing methods of handling these two problems extensively. To handle pose and illumination problems in a uniform framework, we proposed a reflectance model with varying albedo for 3D face and introduced a new concept, the self-ratio image. Finally, we proposed using a 3D model to synthesize the prototype image from a given image under any lighting and viewing conditions. This technique alone can be used to synthesize new images, i.e., enhancing appearance. Adding this technique into existing subspace LDA system, we basically propose an enhanced system. In the future, we plan to improve our method by deforming the 3D model to fit individuals better or using multiple 3D face models as in [4].

References

[1] Y. Adini, Y. Moses, and S. Ullman, "Face Recognition: The problem of Compensating for Changes in Illumination Direction," *IEEE Trans. on PAMI,* Vol. 19, pp. 721-732, 1997.

[2] S. Akamatsu, T. Sasaki, H. Fukamachi, N. Masui, and Y. Suenaga, "An Accurate and Robust Face Identification Scheme," In *Proc. International Conference on Pattern Recognition,* pp. 217-220, The Hague, The Netherlands, 1992.

[3] R. Alferez and Y.F Wang, "Geometric and Illumination Invariants for Object Recognition," *IEEE Trans. on PAMI*, Vol. 21, pp. 505-536, 1999.

[4] J. Atick, P. Griffin, and N. Redlich, "Statistical Approach to shape from shading: Reconstruction of three-dimensional face surfaces from sigle two-dimensional images," *Neural Computation*, Vol. 8, pp. 1321-1340, 1996.

[5] P.N. Belhumeur and D.J. Kriegman, "What is the Set of Images of an Object Under All Possible Lighting Conditions?" In *Proc. Conference on Computer Vision and Pattern Recognition*, San Juan, PR, pp. 52-58, 1997.

[6] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. on PAMI,* Vol. 19, No. 7, pp. 711-720, 1997.

[7] D.J. Beymer and T. Poggio, "Face Recognition from One Example View," in *Proceedings, International Conference on Computer Vision,* pp. 500-507, 1995.

[8] D.J. Beymer, "Face Recognition Under Varying Pose," Technical Report 1461, MIT Artificial Intelligence Laboratory, 1993.

[9] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces, A Survey," *Proc. IEEE*, Vol. 83, pp. 705-740, 1995.

[10] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, "Illumination-Based Image Synthesis: Creating Novel Images of Human Faces Under Differing Pose and Lighting," In *Proc. Workshop on Multi-View Modeling and Analysis of Visual Scenes,* pp. 47-54, 1999.

[11] A.S. Georghiades, D.J. Kriegman, and P.N. Belhumeur, "Illumination Cones for Recognition Under Variable Lighting: Faces," in *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 52-58, 1998.

[12] G. Gordon, "Face Recognition Based on Depth Maps and Surface Curvature," in *SPIE Proceedings, Vol. 1570: Geometric Methods in Computer Vision*, pp. 234–247, 1991.

[13] P. Hallinan, "A Low-Dimensional Representation of Human Faces for Arbitrary Lighting Conditions," in *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 995-999, 1994.

[14] B.K.P. Horn and M.J. Brooks, *Shape from Shading,* MIT Press: Cambridge, MA, 1989.

[15] D.W. Jacobs, P.N. Belhumeur, and R. Basri, "Comparing Images under Variable Illumination," in *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 610-617, 1998.

[16] B. S. Manjunath, R. Chellappa, and C. v. d. Malsburg, "A Feature Based Approach to Face Recognition," in *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 373–378, 1992.

[17] T. Maurer and C.v.d. Malsburg, "Single-View Based Recognition of Faces Rotated in Depth," in *Proc. International Workshop on Automatic Face and Gesture Recognition*, pp. 176-181, 1996.

[18] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The Extended M2VTS Database," in *Proc. International Conference on Audio- and Video-Based Person Authentication*, pp. 72-77, 1999.

[19] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation," *IEEE Trans. on PAMI,* Vol. 19, No. 7, pp. 696-710, 1997.

[20] Nayar, S.K., Ikeuchi, K., and Kanade, T. 1991. Surface Reflection: Physical and Geometrical Perspectives. *IEEE Trans. on PAMI*, Vol. 13, pp. 611-634.

[21] K. Okada, J. Steffans, T. Maurer, H. Hong, E. Elagin, H. Neven, and C.v.d. Malsburg, "The Bochum/USC Face Recognition System and How it Fared in the FERET Phase III Test," *Face Recognition: From Theory to Applications*, Eds. H. Wechsler, P.J. Phillips, V. Bruce, F.F. Soulie and T.S. Huang, Springer-Verlag, pp. 186-205, 1998.

[22] P.J. Phillips, H. Moon, P. Rauss, and S.A. Rizvi, "The FERET Evaluation Methodology for Face-Recognition Algorithms," in *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 137-143, 1997.

[23] Proceedings of the International Conferences on Audio- and Video-Based Person Authentication, 1997,1999.

[24] Proceedings of the International Conferences on Automatic Face and Gesture Recognition, 1995-1998.

[25] A. Pentland, B. Moghaddam, T. Starner, "View-based and Modular Eigenspaces for Face Recognition," in *Proc. Conference on Computer Vision and Pattern Recognition*, 1994.

[26] T. Riklin-Raviv and A. Shashua, "The Quotient Image: Class Based Re-rendering and Recognition With Varying Illuminations," in *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 566-571, 1999.

[27] S.A. Rizvi, P.J. Phillips, and H. Moon, "The FERET Verification Testing Protocol for Face Recognition Algorithms," in *Proc. International Conference on Automatic Face and Gesture Recognition*, 1998.

[28] E. Sali and S. Ullman, "Recognizing Novel 3-D Objects Under New Illumination and viewing Position Using A Small Number of Example Views or Even A Single View," In *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 153-161, 1998.

[29] A. Shashua, "Geometry and Photometry in 3D Visual Recognition," PhD Thesis, MIT, Cambridge, 1994.

[30] P.S. Tsai, and M. Shah, "A Fast Linear Shape from Shading," In *Proc. Conference on Computer Vision and Pattern Recognition*, Urbana/Champaign, IL, pp. 459-465, 1992

[31] S. Ullman and R. Basri, "Recognition by Linear Combinations of Models," In *IEEE Trans. on PAMI,* Vol. 13, pp. 992-1006, 1991.

[32] T. Vetter and T. Poggio, "Linear Object Classes and Image Synthesis From a single Example Image," *IEEE Trans. on PAMI*, Vol. 19, No. 7, pp. 733-742, 1997.

[33] L. Wiskott, J.-M. Fellous, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. on PAMI*, Vol. 19, pp. 775-779, 1997.

[34] L.B. Wolff and E. Angelopoulou, "3-D Stereo Using Photometric Ratios," In *Proc. European Conference on Computer Vision*, pp. 247-258, Springer, Berlin, 1994.

[35] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face Recognition: A Literature Survey," Technical Report CS-TR-4167, Center For Automation Research, Unversity of Maryland, 2000.

[36] W. Zhao and R. Chellappa "Illumination-Insensitive Face Recognition Using Symmetric Shape-from-Shading," In *Proc. Conference on Computer Vision and Pattern Recognition*, pp. 286-293, 2000.

[37] W. Zhao, D. Bhat, N. Nandhakumar, and R. Chellappa, "A Reliable Descriptor for Face Objects in Visual Content," *Journal of Signal Processing: Image Communications,* Vol. 16, pp. 123-136, 2000.

[38] W. Zhao and R. Chellappa, "SFS Based View Synthesis for Robust Face Recognition," *Proc. International Conference on Automatic Face and Gesture Recognition*, 2000.

[39] W. Zhao, *Robust Image Based 3D Face Recognition*, PhD Thesis, University of Maryland, 1999.

[40] W. Zhao, R. Chellappa, and P.J. Phillips, "Subspace Linear Discriminant Analysis for Face Recognition," Technical Report CAR-TR-914, Center for Automation Research, University of Maryland, 1999.

[41] W. Zhao, "Improving the robustness of face recognition," in *Proc. International Conference on Audio- and Video-based Person Authentication*, pp. 78-83, 1999.