MASTER PAPER

APPLICATIONS OF CLASSIFICATION WITH TANGENT DISTANCE TO FACE RECOGNITION UNDER VARYING ILLUMINATION AND POSE CONDITIONS

Submitted by

Jen-Mei Chang

Department of Mathematics

In partial fulfillment of the requirements

for the degree of Master of Science in Mathematics

Colorado State University

Fort Collins, Colorado

Summer 2004

COLORADO STATE UNIVERSITY

July 12, 2004

WE HEREBY RECOMMEND THAT THE MASTER PAPER PREPARED UNDER OUR SUPERVISION BY JEN-MEI CHANG ENTITLED "APPLICATIONS OF CLASSIFICATION WITH TANGENT DISTANCE TO FACE RECOGNITION UNDER VARYING ILLUMINATION AND POSE CONDITIONS " BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN MATHEMATICS.

Committee on Graduate Work

Dr. Bruce Draper

Dr. Jeanne Duflot

Adviser: Dr. Michael Kirby

Department Head: Dr. Simon Tavener

ABSTRACT OF MASTER PAPER

APPLICATIONS OF CLASSIFICATION WITH TANGENT DISTANCE TO FACE RECOGNITION UNDER VARYING ILLUMINATION AND POSE CONDITIONS

We present a feature-invariant classification model that recognizes images under various analytic and nonanalytic transformations, in particular, in the category of face recognition where human faces to be recognized are seen under varying lighting conditions and viewpoints. Our method exploits the idea of tangent approximation to differentiable manifolds and makes use of the tangent distance to build a classifier that is invariant to changes in $2D$ images caused by the lighting conditions, pose, location of the camera, etc. It is important to note that there are two important ideas used in this paper that simplified the face recognition tasks significantly. First, this tangent space model does not require *a-prior* knowledge about the albedo functions and surface normals of the objects to be classified. That is, we work completely with $2D$ images of human faces and focus on the task of recognition only. Secondly, we do not require an analytic expression for the lighting and pose variations to create the image manifolds. We train our classifier on as many images as there are available and still achieve a reasonable recognition rate. Moreover, we employ local SVD to obtain the best tangent vectors for the tangent space and observe the effects of

recognition rate when building the classifier with different number of basis vectors obtained from SVD.

Jen-Mei Chang
Department of Mathematics
Colorado State University
Fort Collins, Colorado 80523
Summer 2004

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

One of the reasons why face recognition has received so much attention recently is because of the growing need for public surveillance in places like airports and governmental agencies. The ability to identify criminals in real-time has the potential to prevent disastrous events. Nowadays, aided by the power of modern computing machines, law enforcement agencies, casinos and airports, etc, can take a snapshot of a potential criminal and search through the vast gallery of images and match the new profile with any existing one in a matter of seconds under the assumption that the person has previously been seen before, i.e., is a member of the gallery. However, it is not always possible that a snapshot can be taken under ideal lighting conditions nor preferred poses. Thus, any recognition system that correctly classifies the identity of a novel subject while allowing changes in viewing conditions will be highly valuable. Another advantage of an identification system based on analysis of frontal images of the face is that it does not require participant's cooperation and knowledge contrary to fingerprint and iris analysis. See [12] for details in the face recognition literature.

Research on automatic machine recognition of faces started in the 1970's. The face recognition problem has been characterized as recognizing $3D$ objects, such as human faces, from $2D$ images. Most of the older methods are feature-oriented. That is, recognition schemes are based upon measurement of the distance of certain attributes of the human faces (e.g. distance from eyes to mouth) and therefore very insensitive to illumination and pose variations. It has been shown empirically in [1, 3, 6] that changes in images due to variations in lighting and pose can be greater than changes in subject's identity. Recent recognition models in dealing with lighting and

Figure 1: The first image on the left is a sample image for a particular face class. The middle face is from the the same face class as the first face with a different illumination condition. The third face is what Euclidean distance recognizes the middle face as. (Images obtained from Yale Face Database B)

pose variations can generally be categorized into two major classes, *generative* [2, 10, 8] and *invariance based* [11, 10, 7]. In this paper, we present an invariance based model that is insensitive to nuances in images induced by a group of nonanalytic transformations.

All classification schemes use one form of metric or another. Classification using the standard Euclidean distance has been shown to be highly unreliable especially in the recognition of human face images seen under varying viewing conditions (illumination and pose). Figure 1 gives an example of an undesirable classification based on Euclidean distance. The left and middle images belong to the same face class but a naive approach using the nearest neighbor algorithm (correlation) with standard Euclidean distance recognizes the second image as the third, which belongs to a different face class. We wish to incorporate a novel metric in our recognition model that will overlook the nuances created from varying viewing conditions. Figure 2 illustrates this idea. If we let E be the left image, Q the middle image, and P the right image seen in figure 1, it is clear that $||E-Q||_2^2 > ||P-Q||_2^2$, which is the undesirable result of the recognition discussed above. An ideal

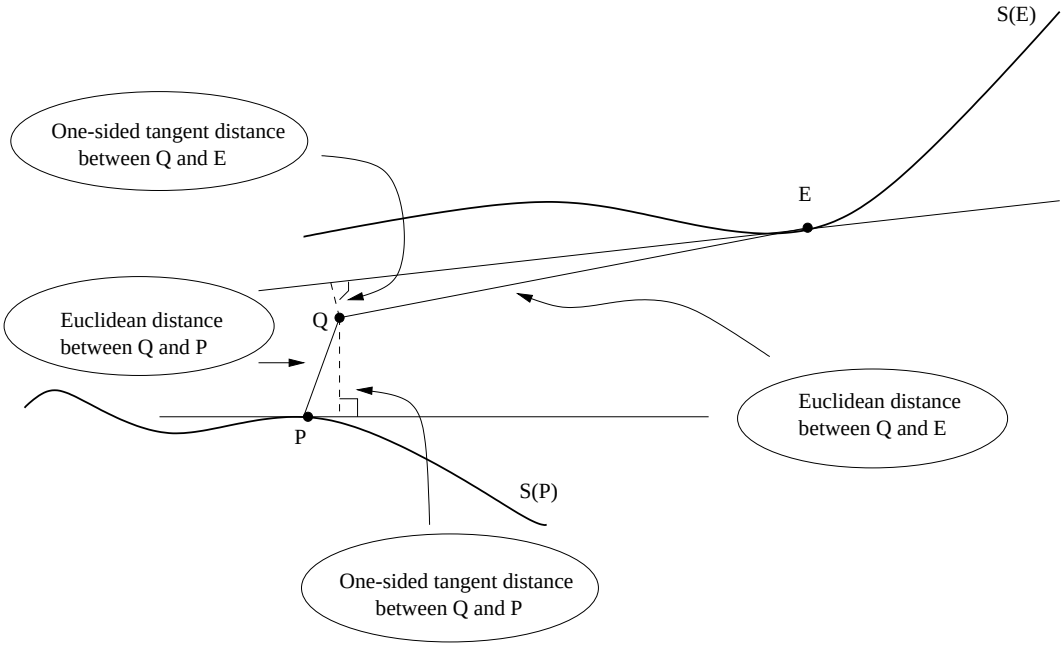

Figure 2: An illustrative example showing a probe pattern, Q, can be farther away from the prototype (E) of the same face class while closer to the prototype (P) of a different face class if using the Euclidean distance but vice versa if using the tangent distance. $S(E)$ is the set of training images for a particular face class characterized by E and $S(P)$ is the set of training images for a different face class characterized by P.

metric d should satisfy the relationship $d(E, Q) < d(P, Q)$, since both E and Q belong to the same face class. From Figure 2 it is apparent that the tangent distance has advantages over the Euclidean distance when data resides on manifolds. While most of the algorithms assume *a-prior* knowledge about the training set, motivated by [4] we use SVD to extract the geometry and statistics of the training set. Therefore, our recognition model is not limited to recognition of images under varying illumination and pose conditions only. It has been shown in [5] that recognition of handwritten digits using local SVD to determine tangent spaces provides a number of advantages. In other words, the tangent space model is able to allow a wide range of analytic and nonanalytic transformations, including rotation, scaling, horizontal translation, vertical translation, illumination, pose conditions, etc.

Problem definition

To clarify the face recognition problem, we distinguish the pose and illumination problems so that in the illumination problem all the images are seen under a fixed frontal pose whereas in the pose problem all the images are seen under a fixed illumination condition. The training and testing sets are obtained from Yale Face Database B, see [8] for a complete description of the database.

Definition 1.1. *(Illumination problem) Given a gallery of $P \times N$ 2D digital images of P face classes under N illumination conditions, $\{x_i^j\}, 1 \leq i \leq N, 1 \leq j \leq P$ and a collection of probe images $\{Q_k\}$, $k \in \mathbb{Z}^+$. The illumination problem is to compute the distance of Q_k to representatives of each face class and assign Q_k the identity of the face class that has the shortest*

distance to Q_k. We denote the collection of $P \times N$ images the illumination database (X_I).

Definition 1.2. *(Pose problem) Given a gallery of $P \times M$ 2D digital images of P face classes under M pose conditions, $\{x_i^j\}, 1 \leq i \leq M, 1 \leq j \leq P$ and a collection of probe images $\{Q_k\}$, $k \in \mathbb{Z}^+$. The pose problem is to compute the distance of Q_k to representatives of each face class and assign Q_k the identity of the face class that has the shortest distance to Q_k. We denote the collection of $P \times M$ images the pose database (X_P).*

Note that in the illumination experiments performed in Section 5, $P = 10$ and $N = 64$, i.e., there are 10 different subjects with 64 illumination conditions for each subject. And in the pose experiments performed in Section 5, $P = 10$ and $M = 9$, i.e., there are 10 different subjects with 9 pose conditions for each subject.

The remainder of this paper is divided in the following way. In Section 2 we discuss some related researches that have been done on recognition models that handle varying illumination and pose conditions in the recent years. In Section 3 we briefly describe the mechanism of how tangent space and the associated tangent distance work in relation to our face recognition problem. In this section we also describe in detail how to build a classification model. In Section 4 we give a complete description of the databases we use to train and test the classifier. In Section 5 we present the testing results. In Section 6 we summarize our classification scheme and discuss its advantages and shortcomings. At the end of this paper, we include four MATLAB codes necessary to run all of the experiments done in this paper.

2 Related work

Many effective algorithms and theories have been developed in effort to solve the face recognition problem when subjects are seen under varying viewing conditions. Fraser and *et al* presented techniques for building a Bayes classifier that combine statistical information of the training data with tangent approximations to known analytic transformations [7]. Belhumeur *et al* developed the theory of illumination cone that gives a representation of the set of images seen under varying illumination conditions [8]. Basri and Jacobs interpret any $2D$ image in terms of reflectance functions of the surface normals and albedo, which are made up of spherical harmonics [2]. Shashua and Riklin-Raviv propose a way to extract illumination-invariant signature images that serve as basis images for the recognition process [10].

Fraser and *et al* propose in [7] a Bayes classifier that exploits *a-prior* knowledge of the known analytic transformations. The classifier treats the pose variations as one of the analytic transformation parameters whereas the illumination variations are treated as noise and handled by a stochastic model that fits the training data. In their application to the face recognition problem, the set of known transformations include rotation, scalings, vertical and horizontal translations. As a consequence, this classifier does not significantly improve the performance of face recognition when images are seen under varying illumination and pose conditions. It is worth mentioning that both [7] and this paper use tangent approximations for the image manifolds; however, we focus on the geometry of the training data while [7] emphasizes on the statistics of the training data.

One of the major results from [3] is that the set of images of an object of *any* shape with a more general reflectance function (contrary to Lambertian), seen under all possible illumination conditions forms a convex cone. Thus the illumination cone representation can be applied to a more general object classification problem. Also, it gives a way to render novel images and re-render images in the database. What makes this representation so useful is that it deals with two problems at once. Each face can be represented by a union of illumination cones where each cone is constructed for each distinct pose condition. This representation of faces gives a way to recognize images that are produced from a variety of viewing conditions (pose and illumination). Two of the shortcomings of this approach are the computational cost and the requirement of 7 images per person for training. Moreover, the recognition rate for images produced under more extreme viewing conditions is not as good compared to methods that deal with varying illumination conditions only.

The Lambertian reflectance method proposed in [2] performs recognition by finding the $3D$ model that best matches a $2D$ query image. It gives an analytic explanation to why illumination cone discussed in [8] can be approximated by a low dimensional ($9D$) linear subspace in the language of spherical harmonics. In general, a digitized $2D$ image is the net product of the reflectance function of the surface normal and the albedo function of the object and the lighting function (intensity and directions). Both lighting and reflectance can be described as functions on the surface of the sphere (human face is close to a half-sphere) and any piecewise continuous function on the surface of the sphere can be written as a linear combination of the spherical harmonics. Thus any $2D$ image can be described analytically in terms of spherical harmonics. This provides a generative model to

re-render images in the database. This model only accounts for attached shadow, which occurs when the inward-pointing surface normal has a negative dot product with the light source (or when object faces away from the light source), therefore recognition is not as good when testing on images that contain cast shadows. The recognition scheme requires computing a face model for each image in the database. That is, it needs to compute harmonic images and reflectance functions for each image in the database and the QR decomposition of the basis matrix consisting of harmonic images as column vectors. Then the matrix Q is used to form a projection matrix QQ^T followed by computing the residual of the projected image and the image. This is done for each image in the database and the query image is assigned the identity of the image that gives the smallest residual. Although the QR composition costs about half as much as SVD for thin rectangular matrices, the magnitude of the computation can get extensive for larger database.

The Quotient Image method [10] proposed by Shashua and Riklin-Raviv assumes Lambertian and ideal class of objects that have the same shape but differ in surface albedo function. One attractive feature of this recognition scheme is that the training set can be as small as 2 subjects with 3 images each (this is referred to as the *bootstrap set*), albeit the recognition result is better with a bigger bootstrap set. The Quotient Image of a face f is defined as the quotient (in the division sense) of a sampled image, f_s, of f and the product of the average of the bootstrap set and a set of appropriate scalars. The algorithm computes a quotient image for each distinct face class in the gallery and any probe face to be classified, then assign the identify of the probe face by correlation. This approach is both

generative and feature-invariant. It is generative since the definition of the Quotient Image gives rise to the image space of the face class by varying the values of the scalars mentioned above and it is illumination invariant by definition. Similar to the other two methods described above, a major drawback of this method is that it fails in case of shadows.

The latter two methods do not offer a solution to variations in pose while the first method does offer a solution but fails to provide an analytic expression for the pose transformation in the face recognition problem. A common shortcoming of all the approaches discussed above is the poor ability to deal with extreme illumination condition. We will see in section 5 that the tangent space model can recognize most of the images with extreme shadowing provided that the training set is big enough.

3 Tangent Space

3.1 Why tangent space and tangent distance and how it works

As discussed in Section 1, we wish to correctly identify two images of the same face class when one image is a transformed version of the other. If we assume that all the possible images of a person available in the training set forms a differentiable manifold in $\mathbb{R}^n$ where n is the length of the image vector and with Figure 2 in mind, we are essentially looking for a way to characterize the local behavior of the manifold at a prototypical point so that it will best match the incoming pattern. Further assume that the potential transformation function involved in creating the manifold is differentiable with respect to the transformation parameter α of length k where k is determined by the subspace approximation. For example, if we use a $7D$ subspace approximation of the manifold, then α will be a length 7 vector. This image manifold, $S_\alpha(E)$, is completely characterized by the prototype E and α since any point that lies on the manifold can be obtained by transforming E by a value of α. Note that we do not need to have any knowledge about the transformation parameter that the manifold inherits. Further assume $S_0(E) = E$. We can approximate this image manifold, $S_\alpha(E)$, by its Taylor expansion at E ($\alpha = 0$) in the following way, according to [11]:

$$S_\alpha(E) = S_0(E) + \left.\frac{\partial\, S_\alpha(E)}{\partial\, \alpha}\right|_{\alpha=0} \alpha + O(\alpha^2) \approx E + \left.\frac{\partial\, S_\alpha(E)}{\partial\, \alpha}\right|_{\alpha=0} \alpha \qquad (1)$$

The tangent space of the manifold at E is spanned by the columns of the Jacobian matrix $\frac{\partial\, S_\alpha(E)}{\partial\, \alpha}$, which we will denote by V_E (tangent vectors). The tangent vectors can be written as:

$$V_E = \left.\frac{\partial\, S_\alpha(E)}{\partial\, \alpha}\right|_{\alpha=0} = \lim_{\epsilon\to 0} \frac{S_\epsilon(E) - S_0(E)}{\epsilon} = \lim_{\epsilon\to 0} \frac{S_\epsilon(E) - E}{\epsilon} \qquad (2)$$

To clarify the dimensions of the tangent vectors V_E for the image manifold characterized by the prototype E and its associated scalars α, we consider the following example,

Example 3.1. *Let X be a set of 2D digital images where each image $x \in X$ is of pixel size 2×3. Then after concatenation by column, each x is of length 6. Assume $S_\alpha(x_0)$ is the image manifold obtained by transforming x_0 by various amounts of α where x_0 is a random element of X. If the tangent space approximation of $S_\alpha(x_0)$ at x_0 is 5 dimensional, then the tangent vectors V will be of dimension 6×5 and its associated scalars α will be of dimension 5×1 since for any point z on the tangent space,*

$$z = x_0 + V\alpha, \textit{ for some } \alpha$$

Therefore,

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \end{bmatrix} = \begin{bmatrix} V_1^{(1)} & V_1^{(2)} & \dots & V_1^{(5)} \\ V_2^{(1)} & V_2^{(2)} & \dots & V_2^{(5)} \\ \vdots & \vdots & \ddots & \vdots \\ V_6^{(1)} & V_6^{(2)} & \dots & V_6^{(5)} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{bmatrix}$$

Equation 2 gives a way to numerically compute the tangent vectors by taking the difference of a transformed image of E and E scaled by the change of the transformation.

It will be an emphasis of this paper how we construct the tangent vectors and determine the dimension of the subspace approximation. Once we have the tool to approximate a manifold linearly, we need a metric to measure the distance of a probe pattern to the tangent space of a manifold. This distance is called the one-sided tangent distance since it measures the distance from a *point* to a *tangent space*. See [11] for the discussion on the

two-sided tangent distance. To find the shortest distance of a probe pattern Q to the tangent space T_E of a manifold $S_\alpha(E)$ at E, we first notice that any point x in T_E can be written as

$$x = E + V_E\alpha, \text{ for some } \alpha \tag{3}$$

and since the one-sided tangent distance between E and Q is given by

$$D(E, Q) = \min_{x \in T_E} ||x - Q||_2^2 \tag{4}$$

we have

$$D(E, Q) = \min_{\alpha} ||E + V_E\alpha - Q||_2^2 \tag{5}$$

Computing a solution of Equation 5 amounts to a least squares problem. The necessary condition of the optimalization problem is that the partial derivative of $D(E, Q)$ with respect to α is equal to zero. The solution obtained this way will be a global minimum and unique.

Proposition 3.1. *Let E and Q be two points in $\mathbb{R}^n$ and $D(E, Q)$ is the one-sided tangent distance between E and Q given in Equation 5. If V_E is the tangent vectors of the image manifold at E, then the solution, α, to Equation 5 is such that*

$$V_E^T(E + V_E\alpha - Q) = 0 \tag{6}$$

Proof.

$$
\begin{aligned}
0 &= \frac{\partial D(E,Q)}{\partial \alpha} \\
&= \frac{\partial}{\partial \alpha}(E + V_E\alpha - Q)^T(E + V_E\alpha - Q) \\
&= \frac{\partial}{\partial \alpha}(E^T + \alpha^T V_E^T - Q^T)(E + V_E\alpha - Q) \\
&= V_E^T(E + V_E\alpha - Q) + (E + V_E\alpha - Q)^T V_E \\
&= < E + V_E\alpha - Q, V_E > + < V_E, E + V_E\alpha - Q > \\
&= 2 < E + V_E\alpha - Q, V_E > \\
&= 2V_E^T(E + V_E\alpha - Q) = 0
\end{aligned}
$$

since E, Q, and α are real vectors and V_E is a real matrix. This implies

$$V_E^T(E + V_E\alpha - Q) = 0$$

□

Now, the solution to Equation 6 is the α such that

$$V_E^T Q - V_E^T E = V_E^T V_E \alpha \tag{7}$$

This is precisely what we need to calculate the one-sided tangent distance of E and Q.

Note that the notion of the tangent distance can be extended to describe the distance between two subspaces. In the remainder of this paper, we will primarily work with the distance between subspaces and occasionally abuse the notion of the tangent distance in substitution of the subspace distances.

3.2 Apply tangent distance to the face recognition problem

We adopt a slightly different approach than [11] and [5] in obtaining the subspace approximation. Since we do not have an analytic expression for the illumination and pose transformations, we can not generate the image manifolds for training by applying the transformation function to a prototypical pattern. Instead, we assume that the set of images from the training set forms a differentiable surface. We now describe our classification process.

Without loss of generality, we will describe the general classification scheme for illumination only. For each distinct face class y in the gallery, there corresponds to a subset U_y of X_I that we use for training and we assume U_y forms a differentiable surface. For each face class y, a tangent space is created at a random prototypical pattern $I(y)$. Note that the choice of $I(y)$ does not effect the results of the experiments done in Section 5. We then compute the basis vectors for the tangent space for each y. Instead of using Equation 2, which requires *a-prior* knowledge about the transformation parameter, we adopt the method of local SVD to obtain the best basis vectors as suggested in [4]. If the size of U_y is n_y for each y, then we create a difference matrix M_y such that its columns are made up of $w_y^i - I(y)$, where $w_y^i \in U_y$, $w_y^i \neq I(y)$ and $1 \leq i \leq n_y$. The $k-$dimensional basis vectors for the local subspaces are the k left singular vectors corresponding to the first k singular values from the singular value decomposition of M_y.

In general, if more training images are available, one would create a difference matrix $M_y(\epsilon)$ for each ϵ-neighborhood of $I(y)$. Then observe the first singular values of each $M_y(\epsilon)$, the second singular values of each $M_y(\epsilon)$, etc. The left singular vectors corresponding to the singular values that scale

linearly are then the tangent vectors. When the training images are sparse, it might not be possible to accurately identify the tangent vectors in such a way. However, the left singular vectors in the SVD of the difference matrices do form a set of best local basis vectors for the subspace approximation of the image manifold.

Now, for any probe pattern x that we wish to classify, we compute the one-sided tangent distance of x to each y by solving Equation 7 for each respective α_y of length k. If there exists a y_0 such that $D(x, y_0) < D(x, y)$ for all $y \neq y_0$, then assign x the identity of y_0.

It is shown in [4] that the dimension of the tangent approximation is given by the number of singular values of the difference matrices described above that scale linearly up to a ϵ-neighborhood of the prototypical pattern. Thus the tangent vectors are the left singular vectors that correspond to the singular values that scale linearly. A slight difference in our approach is that we do not limit ourselves in the dimension (t) of the tangent approximation determined by this scaling argument. Instead, we approximate the image manifold by a k-dimensional subspace where k is not necessarily equal to t. As a consequence, the distance between the probe patter x and each face class y is no longer a tangent distance when $k > t$.

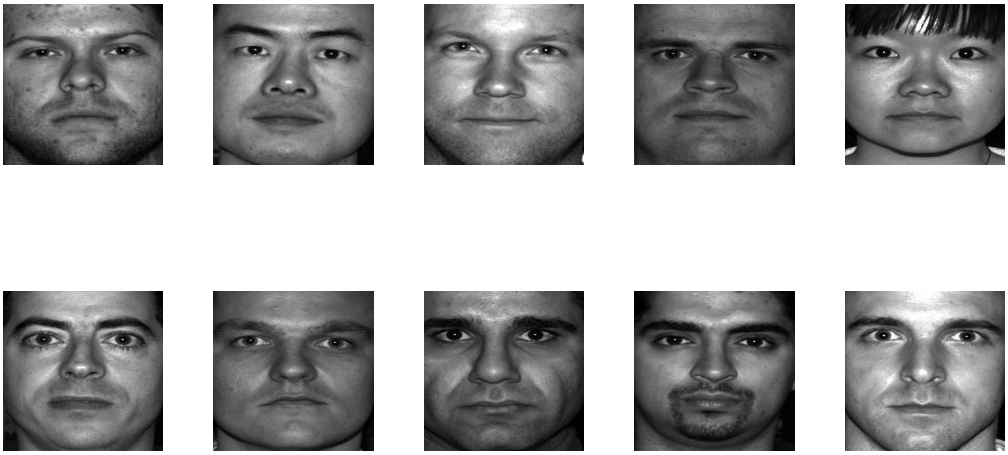

Figure 3: Example images of the 10 individuals in the Yale Face Database B seen under frontal pose. Note that the images shown here are cropped to exclude as much ambient background and hair as possible.

4 Descriptions of the Databases

We test our face recognition method on 2 subsets of the Yale Face Database B, which is specifically designed for testing recognition schemes incorporating illumination and pose variations. The original Yale Face Database contains 5760 single light source images of 10 individuals each seen under 576 viewing conditions (9 poses $\times$ 64 illumination conditions). See Figure 3 for example images of the 10 individuals in the Yale Face Database B. The pixel size of each image is 640(w) $\times$ 480(h). In order to ensure a better recognition rate, we cropped each image to exclude as much ambient background and hair as possible. The resulting pixel size of the images in the pose and illumination databases is 151 $\times$ 151 and 241 $\times$ 181, respectively. See [8] for the original images in the Yale Face Database B.

The first subset that we use for training and testing contains a total of 90 images of 10 individuals each seen under a fixed point light source with 9 poses. We will denote this subset the pose database (X_P). See Figure 4 for

Figure 4: Example images of a single individual in the Yale Face Database B seen under all 9 pose variations. (cropped)

example images of all possible pose variations. The second subset that we use for training and testing contains a total of 640 images of 10 individuals each seen under frontal pose with 64 lighting conditions. We will denote this subset the illumination database (X_I). See Figure 5 for example images of all possible illumination conditions.

It is worth mentioning that in each of the subsets we hold one parameter constant while vary the other. Namely, in the pose database, we fix the lighting condition while vary the pose conditions and vice versa. Moreover, unlike [8], we do not categorize our illumination database based on the angle the light source direction makes with the camera axis.

Figure 5: Example images of a single individual in the Yale Face Database B seen under all 64 illumination conditions. (cropped)

5 Experiments

To further reduce and compress the dimension of the databases $X_{P_{22801\times 90}}$ and $X_{I_{43621\times 640}}$ after preprocessing and concatenation, we factor X_P and X_I into their reduced SVD, i.e.,

$$\begin{aligned} X_{P_{22801\times 90}} &= U_{P_{22801\times 90}} S_{P_{90\times 90}} V^T_{P_{90\times 90}} \\ X_{I_{43621\times 640}} &= U_{I_{43621\times 640}} S_{I_{640\times 640}} V^T_{I_{640\times 640}} \end{aligned}$$

where U_P, U_I, V_P and V_I are orthonormal. Since $U^T U = I$ and UU^T is the orthogonal projection matrix onto the range of X, it is sufficient to perform the task of recognition with SV^T. For the remainder of the paper, we work with the orthogonal projection of the pose and illumination databases.

Now, each of the pose and illumination databases is of size 90×90 and 640×640, respectively. The training and the testing sets for the illumination database is constructed as the following. The first 64 columns of the matrix X_I give all 64 illumination conditions of the first person in the Yale Face Database B. The last 64 columns of X_I give all 64 illumination conditions of the tenth person in the database, etc. The images are not ordered in any specific way. Given X_I, we eliminate one image per person at a time and use the remaining 63 labeled images to build the subspace representation of each person. We then identify the class membership of the eliminated image with this classifier. Repeat this process for all 64 images of each person, therefore creating a testing set of 640 images.

Similarly, we eliminate one image per person at a time and use the remaining 8 labeled images to build the subspace representation of each person in the pose database. We then identify the class membership of the eliminated image with this classifier. Repeat this process for all 9 images

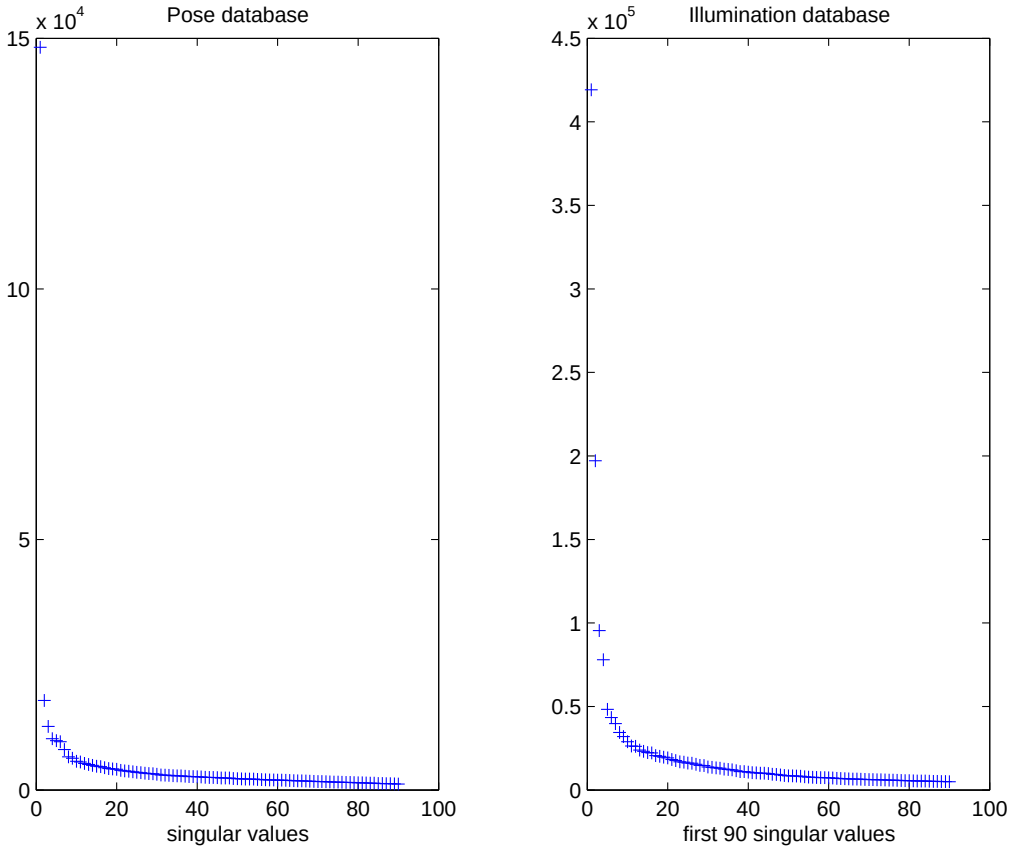

Figure 6: Left: singular values of the pose database. Right: singular values of the illumination database.

of each person, therefore creating a testing set of 90 images. In addition, in the extended pose database created later this section, we use 16 images per person to train.

5.1 Pose

Three kinds of experiments were conducted on the pose database. The first test, shown in Figure 7, was to confirm a major result in [5] that the recognition rate increases as the number of basis vectors used in the local subspace approximation increases. One can see that the recognition rate is greater than or equal to 90% when using 4 or more basis vectors. This

is consistent with the variance of the pose database, shown in Figure 6. The first three singular vectors contribute the most information about the geometry of X_P and the decline of the singular values seems to settle down after the 4th one.

The second test, shown in Figure 8 and produced with 7 basis vectors in the local subspace approximation, shows the 7 images out of a testing set of 90 images that our model fails to recognize. The magic number (number of basis vectors) 7 is chosen for the experiment since it achieves the lowest misclassification rate. The best recognition rate of 92.22% (83 out of 90) is obtained when using 7 basis vectors.

The third test, shown in Figure 9, illustrates how well the classifier learns versus the size of the training set. The recognition is performed on the entire pose database. The order of the probe images being classified follows from the order of the vectors in the data matrix, starting from the 90th one. So some of the images are duplicated in training and testing. One can see that even with only 6 training images, the classification rate is still over 90% (with 4 basis vectors).

To improve the recognition rate even further, one can expand the pose database by including the mirror image of each face, thus creating a symmetry - extended database of 180 images and introducing novel pose conditions to the original database. The method of expanding sparse database by introducing mirror images is proposed in [9]. To create a mirror image of a particular face, we first notice that each face is represented by a $2D$ digital image. A matrix representation of such an image contains the gray value of the face and its dimension is given by the pixel size of the image. We then identify the line of symmetry in the vertical direction of the image matrix

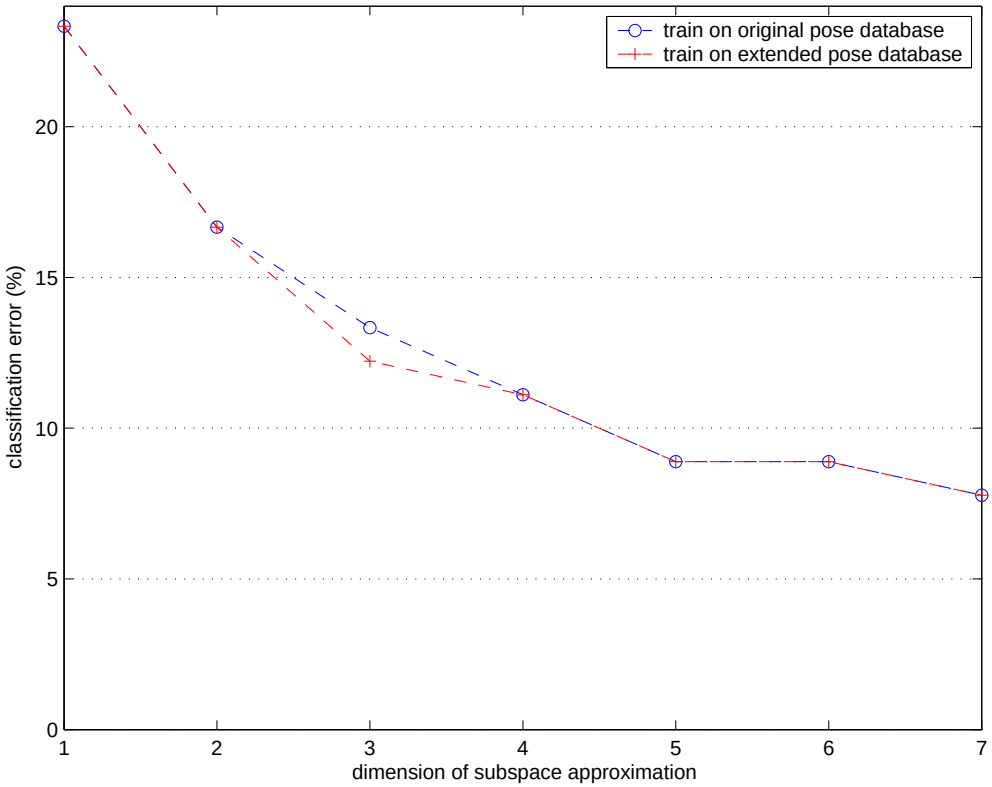

Figure 7: Misclassification rate versus the number of basis vectors used in the local subspace approximation when the classifier is trained on the original pose database and the symmetry-extended pose database.

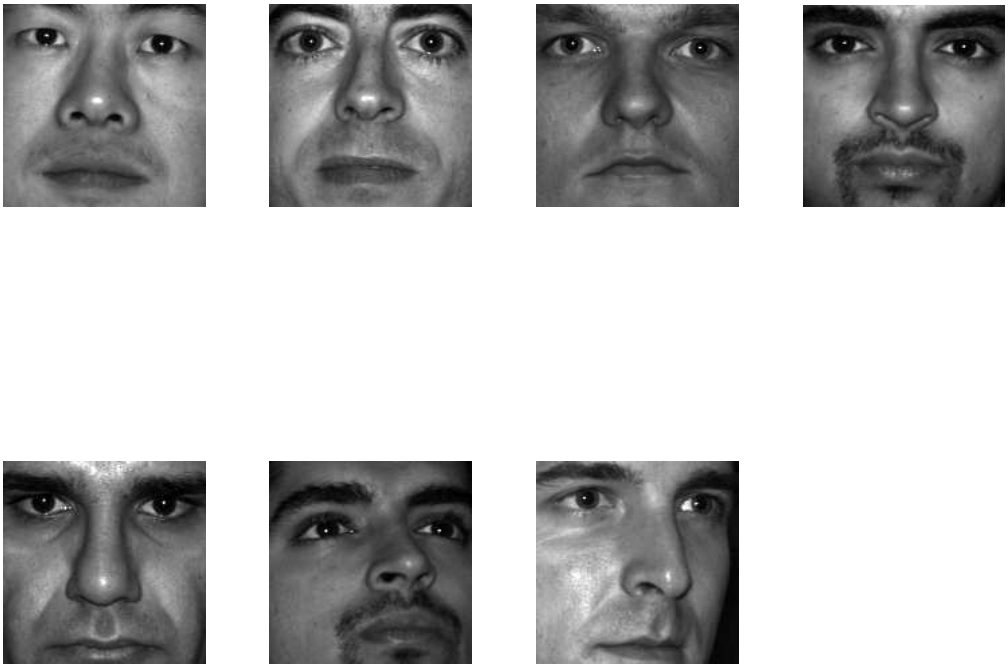

Figure 8: Faces in the original pose database that are misclassified by the classifier when it is trained on the original pose database and the extended pose database with 7 basis vectors in the local subspace approximation.

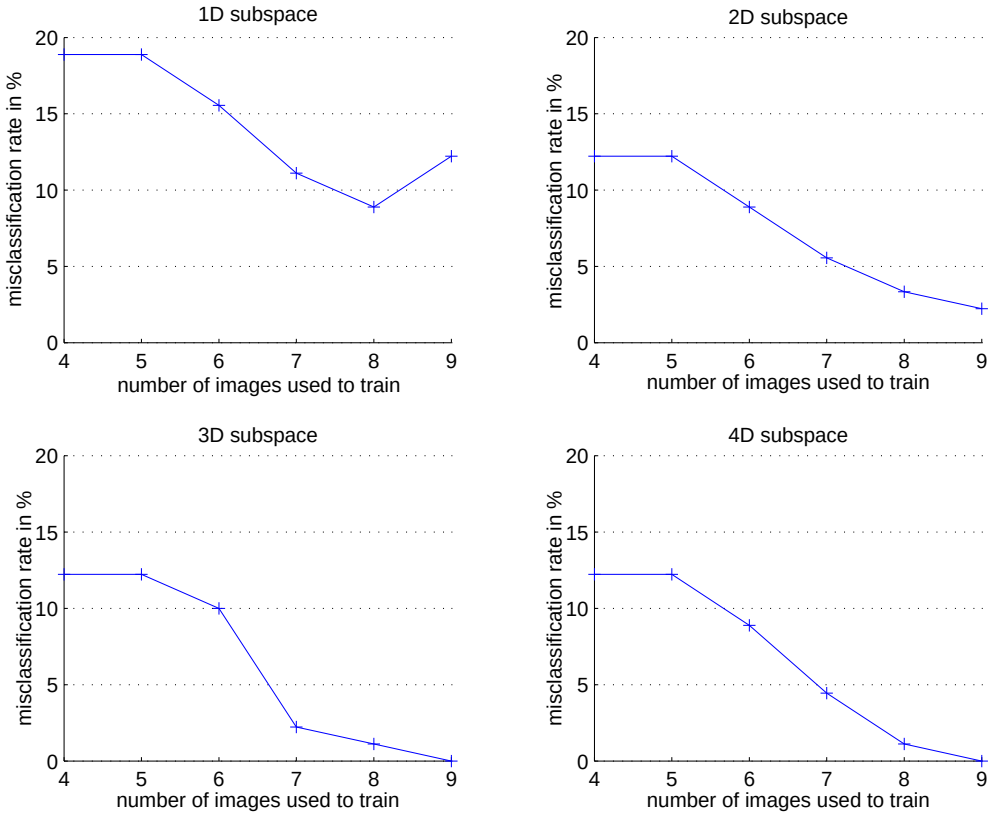

Figure 9: Misclassification rate in % versus the size of the training set when recognition is performed on the entire pose database X_P. Note that when the classifier is trained on all the available poses, it was able to recognize *all* 90 images in the database.

Figure 10: Left: an example image of an individual. Right: left image reflected about the vertical midline.

and flip the column vectors around the line of symmetry. The resulting matrix gives rise to the mirror image of this particular face. See Figure 10 for an illustrative example.

The same experiments are performed on this new database and results can be found in Figure 7 and Figure 8. Note that the images that are misclassified when the classifier is trained on the extended pose database are exactly the same as those misclassified when the classifier is trained on the original pose database. Recognition rate does not improve as we increase the size of the training set. Table 1 shows a sample of recognition rates when the classifier is trained on the original and extended pose databases. Note that the best recognition rate of 92.22% (83 out of 90) for the pose database is obtained when using 7 basis vectors and the classifier is trained on the extended pose database.

5.2 Illumination

First, we would like to discuss a natural way to extend the illumination database if needed. According to [3, 8], the set of images under all possible

dimension of the local subspace approx.	2	3	6	7
recog. rate in % when trained on X_P	81.11	86.67	91.11	92.22
recog. rate in % when trained on ext. X_p	81.11	87.78	91.11	92.22

Table 1: A sample of the recognition rates with various number of basis vectors used in the local subspace approximation when the classifier is trained on X_P and the extended X_P. Recognition performed on the pose database.

illumination conditions with a particular pose forms an convex polyhedral cone, $\mathcal{C}$. If we let s_1 and s_2 be two images of a person seen under two distinct lighting conditions, then $\mathcal{C}$ being convex implies

$$\alpha s_1 + (1-\alpha)s_2 = s \in \mathcal{C}, \text{ where } \alpha \in [0,1]$$

The lighting condition created in such a way are realized as true lighting condition, thus images are realistic. See Figure 11 for example images with various values of α. One can now expand the training set to any desirable size in such a way. Note that it is sufficient to create novel images with the orthogonal projection of X_I, since if V_i^T and V_j^T are the ith and jth column of V^T where i and j are two distinct lighting conditions of the same face class, then the image s created by a convex combination of V_i^T and V_j^T is given by

$$s = US(\alpha V_i^T + (1-\alpha)V_j^T). \tag{8}$$

Now, multiply U^T through Equation 8, we obtain

$$\begin{aligned} U^T s &= S(\alpha V_i^T + (1-\alpha)V_j^T) \\ &= \alpha S V_i^T + (1-\alpha) S V_j^T, \end{aligned}$$

which is the same convex combination of V_i^T and V_j^T in the projected space. Thus, it is sufficient to work with the projected space when creating novel images.

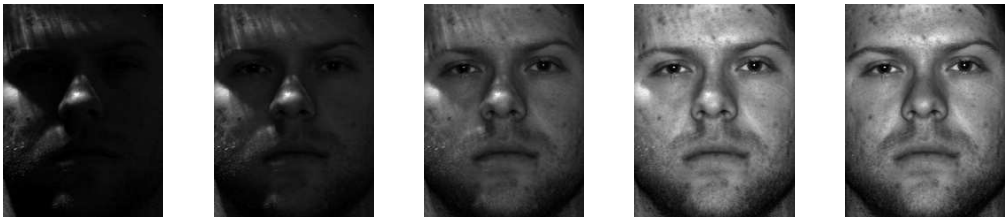

Figure 11: An illustrative example showing the images created by convex combinations of two distinct illumination conditions of a particular face object. From left to right: $\alpha = 0$ (first illumination condition), $\alpha = 0.1429$, $\alpha = 0.3810$, $\alpha = 0.6190$ and $\alpha = 1$ (second illumination condition).

dimension of the local subspace approx.	2	6	9	12
recognition rate in %	90.63	98.75	99.22	99.38

Table 2: A sample of the recognition rates with various number of basis vectors used in the local subspace approximation. Recognition performed on the illumination database.

A set of similar tests are done on the illumination database. See Figures 12, 13 and 14 for an illustration. First note that the four images shown in Figure 13 are almost impossible for human eye to recognize. Secondly, the best recognition rate of 99.38% (636 out of 640) is obtained when using 12 basis vectors. From Figure 6, we observe that the recognition rate will improve drastically if we take more than 4 basis vectors. See Table 2 for a sample of recognition rates. Note that the overall recognition rate is the best on X_I, then the symmetry-extended X_P and the original X_P. This is due to the decrease of the size of the training set.

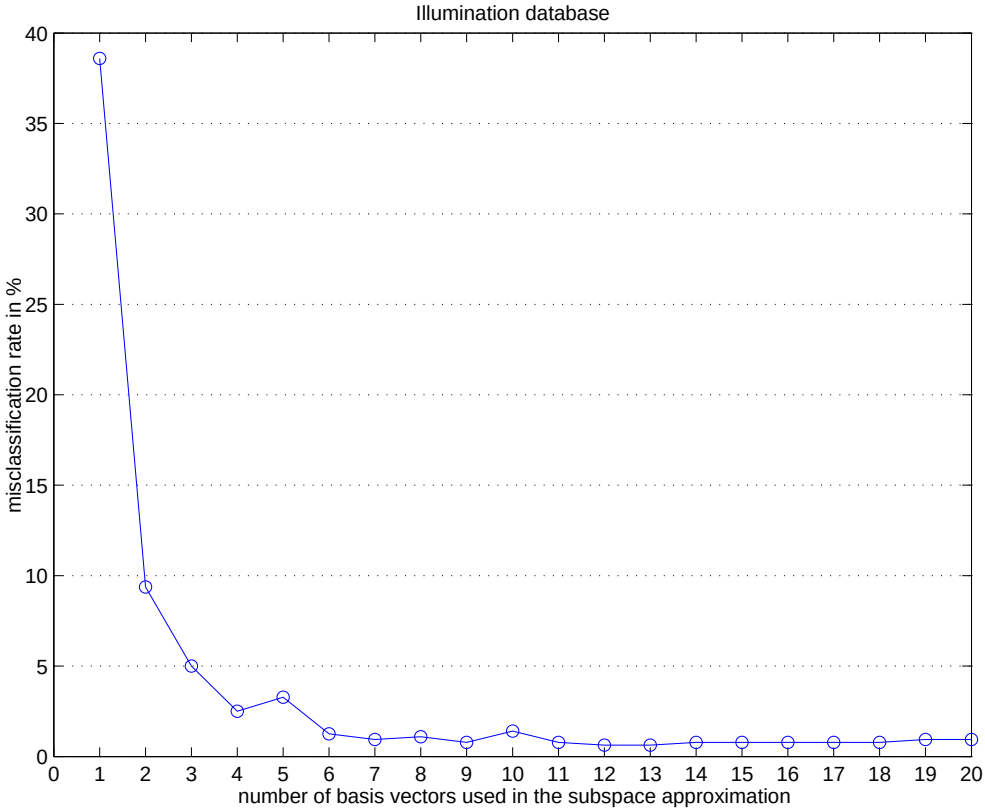

Figure 12: Misclassification rate versus the number of basis vectors used in the local subspace approximation for the illumination database.

Figure 13: Faces in the illumination database that are misclassified by the classifier with $12D$ local subspace approximation.

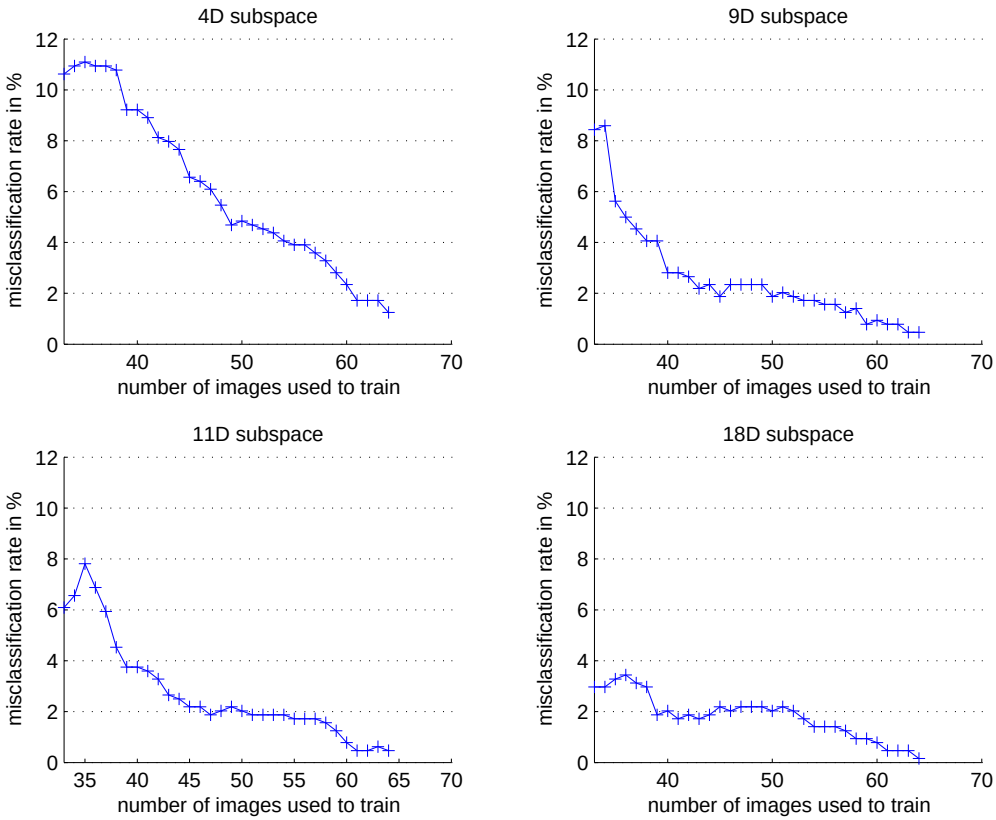

Figure 14: Misclassification rate in % versus the size of the training set when recognition is performed on the entire illumination database X_I.

6 Summary and Future work

We presented a feature-invariant classification model that offers a solution to recognition of images seen under varying illumination and pose conditions. For each face object y in the gallery, there corresponds to a tangent space (subspace) representation. An implemented version of the standard Euclidean distance is used to find the distance of a probe image x and y for each y in the gallery by calculating the distance from x to the tangent space (subspaces) of y. The basis vectors for the local subspace approximation are obtained via SVD. It follows that the identity of the probe image x is the identity of the face class that offers the shortest distance to x.

This model can be extended to deal with a combination of both illumination and pose variations. Although the recognition rate is enhanced by having abundant training images, we still achieve a reasonably high recognition rate when the size of the training set is as small as 6 and 32 for the pose and illumination database, respectively. Some of the advantages of the tangent space representation are that it is simple and cost-effective; the one-sided tangent distance is fairly easy to implement from the standard Euclidean distance and it is insensitive to relatively large nuances in images. A major shortcoming of this method is the requirement of a relatively bigger training set. However, it is quite easy to collect a reasonable size of the training set in reality. Our method improves drastically as we enlarge the size of the training set while remaining cost-effective.

References

[1] Yael Adini, Yael Moses, and Shimon Ullman. Face recognition: The problem of compensating for changes in illumination direction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):721–732, 1997.

[2] Ronen Basri and David W. Jacobs. Lambertian reflectance and linear subspace. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(2):218–233, 2003.

[3] P. N. Belhumeur and D. J. Kriegman. What is the set of images of an object under all possible lighting conditions? In *Proceedings of the 1996 Conference on Computer Vision and Pattern Recognition (CVPR '96)*, page 270. IEEE Computer Society, 1996.

[4] D. S. Broomhead, R. Jones, and G. P. King. Topological dimension and local coordinates from time series data. *J. Phys. A: Math. Gen*, 20:L563–L569, 1987.

[5] Jen-Mei Chang, Michael Kirby, Josh Ladd, Lucas Krakow, Ethan Murphy, and Steve Simske. Classification of images with tangent distance.

[6] David Dreisigmeyr and Michael Kirby. Classification on manifolds.

[7] A. M. Fraser, N. W. Hengartner, K. R. Vixie, and B. W. Wohlberg. Classification modulo invariance, with application to face recognition. *Journal of computational and graphical statistics*, 12(4):829–852, 2003.

[8] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(6):643–660, 2001.

[9] M. Kirby and L. Sirovich. Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(1):103–108, 1990.

[10] T. Riklin-Raviv and A. Shashua. The quotient image: class based re-rendering and recognition with varying illuminations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(2):129–139, 2001.

[11] P. Y. Simard, Y. A. Cun, J. S. Denker, and B. Victorri. Transformation invariance in pattern recognition - tangent distance and tangent propagation. *International Journal of Imaging System and Technology*, 11:181–194, 2001.

[12] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, 2003.

A Appendix

This appendix includes four MATLAB codes necessary in the experiments done in this paper.

Code #1

The projection.mat contains a single element P, which is the orthogonal projection of the original illumination database. The following code returns a misclassification rate for the illumination database. A similar code is also used for the pose database. The Pose.mat contains an element R, which is the orthogonal projection of the original pose database.

```
%clear
load projection
num_person = 10;
num_illu = 64;
num_train = 63;
num_tan = 12;
miss = zeros(num_illu,1);

order = [1 2 3 4 5 6 7 8 9 10];
%pattern is the one that's eliminated.
for elimination = 1:64
    for j = 1:num_person
       patterns(:,j,elimination) =
       P(:, elimination + num_illu*(j-1));
    end

%if the first image of each person is eliminated,
%then choose the 37th image as prototype for each person,
%otherwise the first image is the prototype.
    proto = [];
    if elimination == 1
        for i = 1:num_person
            proto(:,i) = P(:,(i-1)*num_illu+37);
        end
    else
        for i = 1:num_person
            proto(:,i) = P(:,(i-1)*num_illu+1);
        end
    end
```

```
    dataset = [];
    tan = [];
    for i = 1:num_person
        if elimination == 1
            dataset(:,:,i) =
            P(:,2+num_illu*(i-1):num_illu*i);
        else if elimination == num_illu
                dataset (:,:,i) =
                P(:,1+num_illu*(i-1):i*num_illu-1);
        else
                dataset(:,:,i) =
                    [P(:,(1+num_illu*(i-1):
                    (elimination-1)+num_illu*(i-1)))
                    P(:,(elimination+1)+num_illu*(i-1):
                    num_illu*i)];
        end
        end
            tan(:,:,i) =
            tanspace(proto(:,i),dataset(:,:,i),num_train,num_tan);
    end

    TD = [];
    for i = 1:num_person
        for j = 1:num_person
            TD(i,j,elimination) =
                tan_dis(proto(:,j),tan(:,:,j),
                patterns(:,i,elimination),0);
        end
    end
    for i = 1:num_person
        for j = 1:num_person
            if TD(i,j,elimination) == min(TD(i,:,elimination))
                index(i,elimination) = j;
            end
        end
        if index(i,elimination) ~= order(i)
            miss(elimination) = miss(elimination) + 1;
        end
    end
end
total_miss = sum(miss);
```

Code #2

This code calculates the k dimensional basis vectors given the manifold and the prototype where the local subspace approximation is created.

```
function [tan_vec] = tanspace(prototype,dataset,num_train, k)

for i = 1:num_train
    if dataset(:,i) ~= prototype
    tangvecs(:,i) = (dataset(:,i) - prototype);
    end
end

[tan_vec,S,V] = svds(tangvecs,k);
```

Code #3

We use only the one-sided tangent distance in this paper. Namely, we create only the tangent space at the image manifold with prototype E and compute the Euclidean distance of a probe image P with the tangent space of the manifold at the prototype E. Note that it is not needed to use the elastic constant k throughout the experiments since we do not need to create tangent vectors at P, therefore avoiding the possibility of tangent vectors of E and P being collinear.

```
function [output] = tan_dis(E,Le,P,k)

Lee = Le.'*Le;
A21 = -(1+k)*Le.';
A2 = ((1+k)^2)*Lee;
y = E-P;

b = A21*y;

[L,U,Perm] = lu(A2);

alpha = inv(U)*inv(L)*inv(Perm)*b;
Ep = E + Le*alpha;

output = norm(Ep-P)+k*(norm(Le*alpha,2));
```

Code #4

In order to improve the recognition rate in the pose database. We expand the training set by creating mirror image of each face. Let this new database be the extended database, which contains 18 images of each person. We construct the extended pose database by combining the projection of the original pose database and the projection of the mirror image database.

```
clear
load Pose
num_person = 10;
for i=1:size(faces,2)
    images(:,:,i) = reshape(faces(:,i),151,151);
    for j=1:151
        R_images(:,j,i) = images(:,151-j+1,i);
    end
    mirror_images(:,i) = reshape(R_images(:,:,i),151*151,1);
end
[U_mirror,S_mirror,V_mirror] = svd(mirror_images,0);

P_mirror = S_mirror*V_mirror.';
```