Face Recognition Under Varying Viewing Conditions with Subspace Distance

Jen-Mei Chang
Department of Mathematics and Statistics
California State University, Long Beach
1250 Bellflower Blvd.
Long Beach, California 90840-1001
jchang9@csulb.edu

Michael Kirby
Department of Mathematics
Colorado State University
Fort Collins, Colorado 80523-1874
kirby@math.colostate.edu

Abstract

We present a feature-invariant classification model that recognizes images under various analytic and nonanalytic transformations 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, which motivates the use of subspace distance to build a classifier that is invariant to changes in 2D images caused by the lighting conditions, pose, location of the camera, etc.

There are two important ideas used in this paper that simplified the face recognition tasks significantly. First, this subspace 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. 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. For this, we employ the local KL algorithm to obtain the best local basis vectors for the subspace and observe the effects of recognition rate when experimenting with different number of basis vectors.*

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 [1] 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 [2, 3, 4] 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 pose variations can generally be categorized into two major classes, *generative* [5, 6, 7] and *invariance- based* [8, 6, 9] models. In this paper, we present an invariance-based model that is insensitive to nuances in images induced by a group of nonanalytic transformations.

When the manifold is made up of images obtained via transformations (e.g., rotations, translations, scaling), the manifold has the differential topology of a Lie group. Thus it is possible to calculate an optimal linear approximation (subspace) that captures the relevant linear effects of deformation. This subspace, called the tangent space, typically offers a low-dimensional characterization on the images and contains nearly the same information as the original manifold for small transformations. Measuring the distance between images can then easily be done by forming their re-

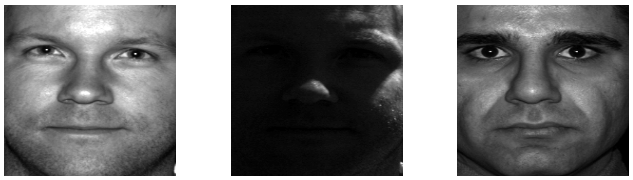

Figure 1. Left: a sample image for a particular face class in YDB. Middle: a sample image from the same face class but a different illumination condition. Right: image as what Euclidean distance recognizes the middle face as.

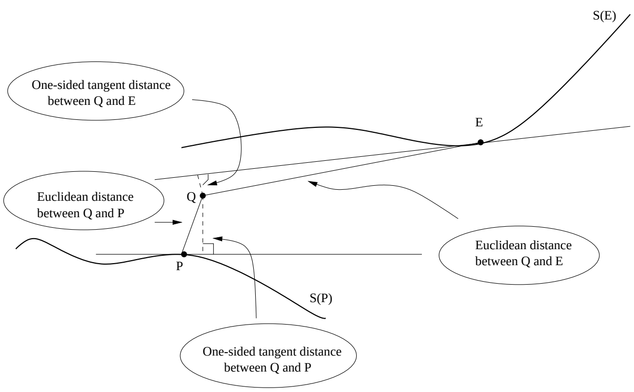

Figure 2. [8] $S(E)$ is the set of training images for a particular face class characterized by E and $S(P)$ is the set of training images for a different face class characterized by P. This is an illustrative example showing a probe pattern, Q, can be farther away from the prototype (E) of the same face class while closer to the prototype (P) of a different face class when the Euclidean distance is used. Vice versa when the tangent distance is used.

spective deformation manifold and tangent space followed by finding the gap distance between the two tangent spaces. This distance is called the two-sided tangent distance in [8]. This distance measure performs better than the standard Euclidean distance when images lie on a manifold.

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. 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 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 vectorize the left image by E, the middle image by Q, and the right image by P 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 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 require the knowledge of parameterized image transformations [10, 11], we use SVD to extract the geometry of the training set that is motivated by [12]. Therefore, our recognition model is not limited to recognition of images under varying illumination and pose conditions only. In other words, the subspace 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.

The training and testing sets are obtained from Yale Face Database B (YDB) [7]. In all empirical experiments that follow, 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. In the illumination experiments performed in Section 5, there are 10 different subjects with 64 illumination conditions for each subject and 9 distinct poses for each subject in the pose experiments.

The remainder of this paper is divided in the following way. In Section 2 we discuss some related researches that handle varying illumination and pose conditions in face recognition problems. In Section 3 we briefly describe how our subspace distance model is motivated by the ideas of tangent space and the associated tangent distance. Database description is given in Section 4 while empirical results are presented in Section 5. Finally, we conclude our findings in Section 6.

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 [9, 7, 5, 6].

Fraser and *et al.* propose in [9] 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 [9] and this paper use tangent approximations for the image manifolds.

One of the major results from [3] is that the set of images of an object of *any* shape with a general reflectance function, seen under all possible illumination conditions forms a convex cone. 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). A drawback of this approach is the computational cost. 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 [5] 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 [7] 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 [6] 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 subspace space model can recognize most of the images with extreme shadowing provided that the training set is big enough.

3. Tangent Space

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 dimension of the tangent space. 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 [8]:

$$\begin{aligned} S_\alpha(E) &= S_0(E) + \left.\frac{\partial S_\alpha(E)}{\partial \alpha}\right|_{\alpha=0} \alpha + O(\alpha^2) \\ &\approx E + \left.\frac{\partial S_\alpha(E)}{\partial \alpha}\right|_{\alpha=0} \alpha \end{aligned}$$

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:

$$V_E = \left.\frac{\partial S_\alpha(E)}{\partial \alpha}\right|_{\alpha=0} = \lim_{\varepsilon \to 0} \frac{S_\varepsilon(E) - S_0(E)}{\varepsilon}. \quad (1)$$

Equation 1 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. However, we adapt an alternative technique in determining the local tangent dimension and constructing the corresponding tangent vectors in this paper.

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* [8]. 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

$$x = E + V_E\alpha, \text{ for some } \alpha \quad (2)$$

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

$$D(E,Q) = \min_{x \in T_E} ||x - Q||_2^2 \quad (3)$$

we have

$$D(E,Q) = \min_{\alpha} ||E + V_E\alpha - Q||_2^2 \quad (4)$$

Computing a solution of Equation 4 amounts to a least squares problem. The necessary condition of the optimization 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.

Notice that this notion of the tangent distance can be extended to describe the distance between two subspaces. This is done simply replacing V_E by the basis vectors for the best local subspace approximation. It is the heart of this paper that we investigate how the classification rates change as we adjust the dimension of the subspace approximations.

As mentioned before, instead of opting an numerical approach, we adapt the *Local KL Algorithm* proposed in [12] for finding the local subspace dimension. Since we do not have an analytic expression for the illumination and pose transformations, we can not generate the image manifolds 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 for the illumination problem, where the database is denoted by X_I. Similar process is applied to the pose database, I_P, as well.

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 1, which requires either knowledge about the transformation parameters or having a fine sampled database, we adapt the method of local KL to obtain the best basis vectors as suggested in [12]. 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 $M_y(\varepsilon)$ for each ε-neighborhood of $I(y)$. Then observe the first singular values of each $M_y(\varepsilon)$, the second singular values of each $M_y(\varepsilon)$, 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. 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 [12] 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. We found that the tangent dimension is approximately 9 for the illumination database and 6 for the pose database. We do not claim that these numbers are representative as a whole since our data sets are not ideally sampled to best accommodate this approach. In obtaining the subspace distances, 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 k-dimensional subspaces where k is not necessarily equal to t.

4. Descriptions of the Databases

We test our face recognition method on 2 subsets of the YDB, which is specifically designed for testing recognition schemes incorporating illumination and pose varia-

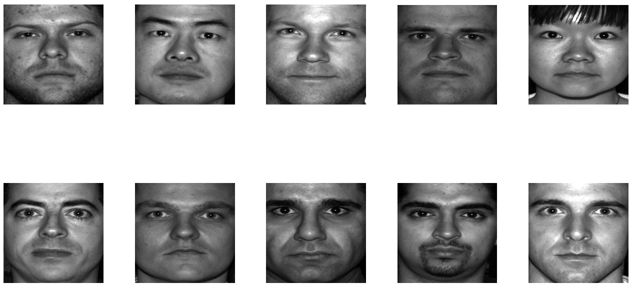

Figure 3. Example images of the 10 individuals in the YDB seen under frontal pose. Note that the images shown here are cropped to exclude as much ambient background and hair as possible.

tions. The original YDB 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 YDB. The pixel size of each image is 640×480. 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 151×151 and 241×181, respectively. See [7] for the original images in YDB.

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 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 [7], we do not categorize our illumination database based on the angle the light source direction makes with the camera axis.

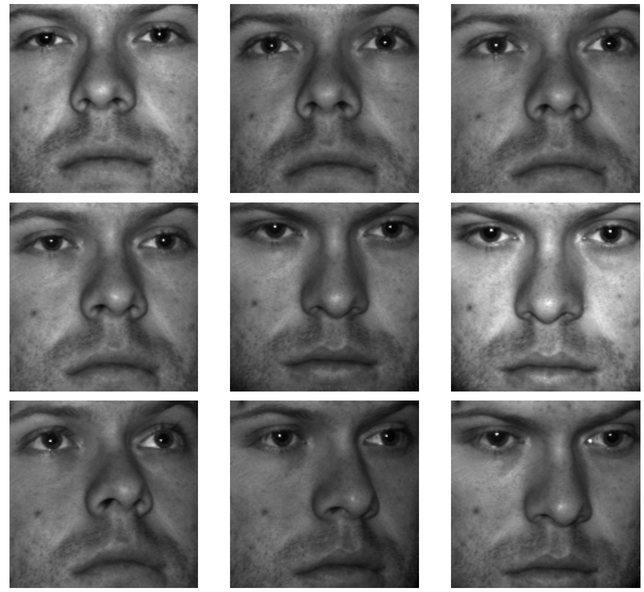

Figure 4. Example images of a single individual in the YDB seen under all 9 pose variations.

Figure 5. Example images of a single individual in YDB seen under all 64 illumination conditions.

5. Experiments

To further reduce 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 $U^TU = I$ 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, the pose and illumination databases are of size 90×90 and 640×640, respectively. Without loss of generality, we describe the process of constructing the training and the testing sets for the illumination database only. The training and the testing sets for the pose database is constructed in exactly the same way.

The first 64 columns of the matrix X_I give all 64 illumination conditions of the first person in YDB. The last 64 columns of X_I give all 64 illumination conditions of the

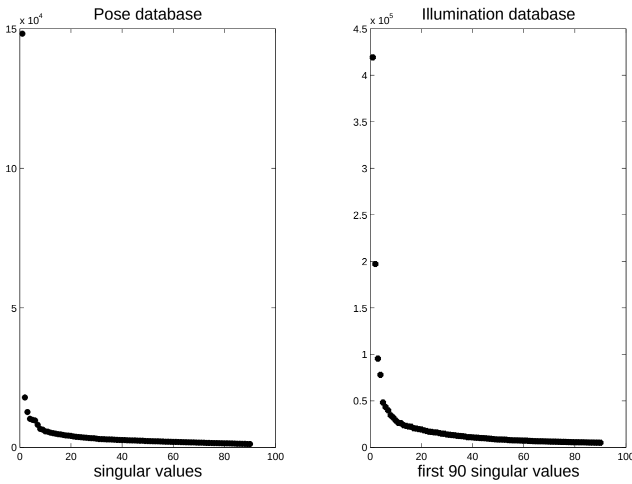

Figure 6. Left: singular values of the pose database. Right: singular values of the illumination database.

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 tangent space for 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.

5.1. Pose

Three kinds of experiments were conducted on the pose database. The first test, shown in Figure 7, was to confirm 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 5 or more basis vectors. This is consistent with the variance of the pose database, shown in Figure 6. The first four 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 5th 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 learnt empirically. Indeed, the best recognition rate of 92.23% (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. 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, we expand our pose database by including the mirror image of each face, thus creating a symmetry - extended database of 180

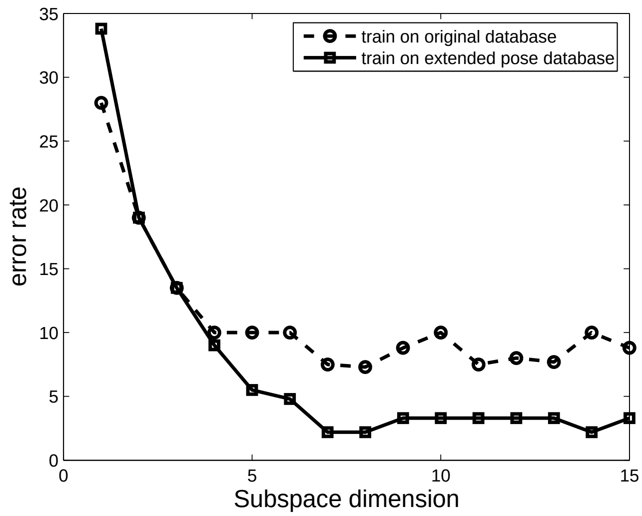

Figure 7. Misclassification rate versus the number of basis vectors used in the local subspace approximation when the classifier is trained on the original pose database and the symmetry-extended pose database.

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

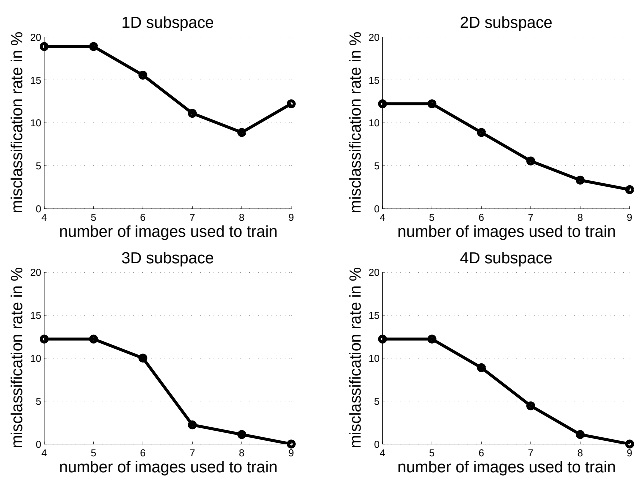

Figure 9. Misclassification rate in % versus the size of the training set when recognition is performed on the entire pose database X_P. Note that when the classifier is trained on all the available poses, it was able to recognize *all* 90 images in the database.

images and introducing novel pose conditions to the original database. The method of expanding sparse database by

local subspace dim	2	6	7	8
recog. rate on X_P	81.1	90	92.2	91.1
recog. rate on ext. X_p	81.1	95.6	97.8	97.8

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.

Figure 10. Faces in the original database that are misclassified by the classifier when it is trained on the symmetry-extended pose database with $7D$ local subspace approximation.

introducing mirror images is proposed in [13]. 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 and flip the column vectors around the line of symmetry. The resulting matrix gives rise to the mirror image of this particular face.

The same experiments are performed on this new database and results can be found in Figure 7 and Figure 10. 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 97.78% (88 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

A set of similar tests are done on the illumination database. See Figures 11, 12 and 13 for an illustration. First note that the four images shown in Figure 12 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, then the original X_P. This is because of the size of the training set decreases from 63 to 17 to 8 images per person.

local subspace dim	2	6	9	12
recognition rate	90.6	98.8	99.2	99.4

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.

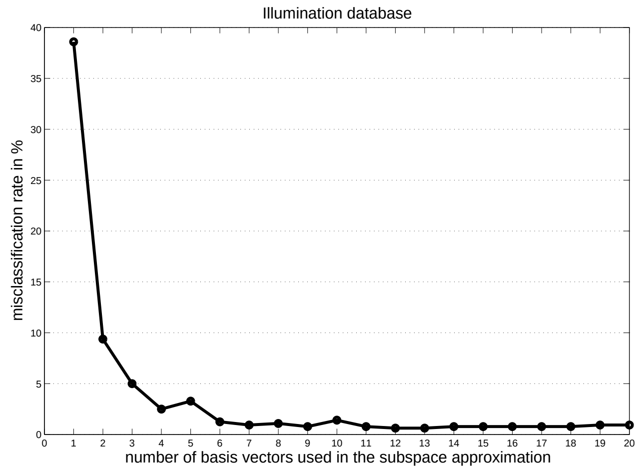

Figure 11. Misclassification rate versus the number of basis vectors used in the local subspace approximation for the illumination database.

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

6. Summary

As an attempt to attack the face recognition problem under varying illumination and pose conditions, the tangent space/distance model was used to motivate a subspace model. The notion of tangent/subspace distance is easy to comprehend as it is built upon and implemented from the widely used Euclidean-norm. The concept of local invariance being the deciding factor in pattern classification may seem a bit unsteady, but in fact, is overly important. For this purpose, we adapt the local KL algorithm to extract basis information for the best subspace approximation for each training subject in the gallery. A central message of the paper is the fact that recognition rates improve as we vary the subspace dimensions. Although the recognition rate on the extremely cast-shadowed images is highly satisfactory, the proposed method can certainly benefit from a better sampled training set. In the world we live in nowadays, this is no longer infeasible.

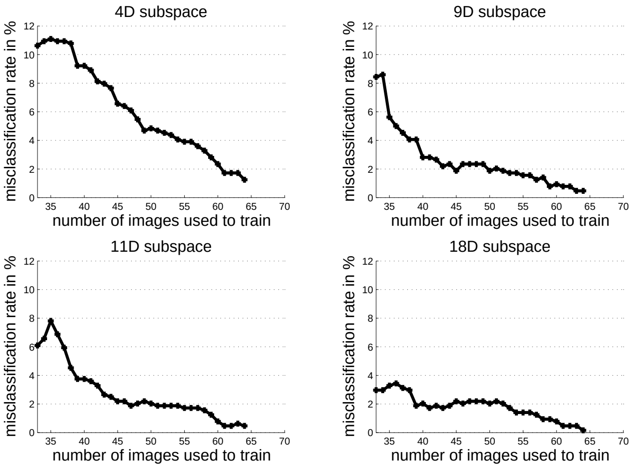

Figure 13. Misclassification rate in % versus the size of the training set when recognition is performed on the entire illumination database X_I.

References

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[2] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction," *IEEE Tran. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 721–732, 1997.

[3] P. Belhumeur and D. Kriegman, "What is the set of images of an object under all possible lighting conditions," *IJCV*, vol. 28, pp. 270–277, 1998.

[4] D. Dreisigmeyr and M. Kirby, "Classification on manifolds."

[5] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Analyis and Machine Intelligence*, vol. 25, no. 2, pp. 218–233, 2003.

[6] T. Riklin-Raviv and A. Shashua, "The quotient image: Class based re-rendering and recognition with varying illuminations," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 129–139, 2001.

[7] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[8] P. Simard, Y. Cun, J. Denker, and B. Victorri, "Transformation invariance in pattern recognition - tangent distance and tangent propagation," *IJIST*, vol. 11, pp. 181–194, 2001.

[9] A. Fraser, N. Hengartner, K. Vixie, and B. Wohlberg, "Classification modulo invariance, with application to face recognition," *J. of Computational and Graphical Statistics*, vol. 12, no. 4, pp. 829–852, 2003.

[10] A. Fitzgibbon and A. Zisserman, "Joint manifold distance: a new approach to appearance based clustering," in *Proc. Computer Vision and Pattern Recognition Conf.* IEEE Computer Society, 2003, pp. 26–36.

[11] N. Vasconcelos and A. Lippman, "A multiresolution manifold distance for invariant image similarity," *IEEE Trans. Multimedia*, vol. 7, no. 1, pp. 127–142, 2005.

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

[13] M. Kirby and L. Sirovich, "Application of the Karhunen-Loève procedure for the characterization of human faces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103–108, 1990.