Human Face Reconstruction and Recognition A Linear Algebra Approach

Kevin Ng and Suresh Subbiah
The University of Texas at Austin
Austin, Texas 78712 USA

May 12, 2000

Abstract

We discuss a linear algebraic approach to the two related problems of human face recognition and reconstruction. The central issue is to identify a subspace of all images of reduced dimensionality into which projections of the original image can be made to facilitate efficient image storage and reconstruction as well image recognition. Two new methods called Meanfaces and Svdfaces are proposed to address these problems. The new methods proposed are computationally significantly cheaper than standard linear algebraic methods such as Eigenfaces and Fisherfaces. The performance of the new algorithms is shown to be comparable to Fisherfaces and Eigenfaces.

1 Introduction

Automated human face recogition by machines and the related problem of storing human face images efficiently are of vital interest in many areas such as image databases, image based security systems and robotic vision. In the past decade several algorithms have been proposed to tackle these problems [7], ranging from traditional geometry based methods to modern probabilistic [5] and modular subspaces [6] based methods. In this article we focus on a class of linear algebraic methods proposed in [1], and extended for wider range of applicability in [2, 4]. These methods attempt to identify an appropriate subspace of the space containing all images into which full images can be projected. The projected images are typically 3-4 orders of magnitude smaller than the original specification of the image in conventional picture formats. The projected images can used for later reconstruction of the original image or for recognition. The choice of subspace is the centrally important problem in this approach.

It has been noted in the literature [4] that the choice of subspace depends on the intended application of the projected image. In [1, 2] the eigensubspace corresponding to non-zero eigenvalues of the covariance matrix of all the images is chosen. It is shown that this eigensubspace is extremely efficient for image reconstruction as well as useful for image recognition. In [4], using traditional Fisher Discriminant Analysis [9] it is shown that a different subspace, now known as Fisherspace, is very accurate for image recognition. We feel that the cost of computation of these approaches, especially for large data sets, has not been sufficiently addressed. In this article we propose two computaionally cheaper approaches to identifying alternate subspaces for image reconstruction and recognition, called Meanfaces and Svdfaces.

In both these approaches the set of training images is first clustered into a fixed number of classes using an automatic algorithm like k-means or manually based on prior knowledge about the subjects in the training set. Then in Meanfaces the average of each cluster is calculated and subtracted from the global average of the training set. The space spanned by these Meanfaces is shown to maximize the sactter between the clusters in section 3. In Svdfaces, one or two dominant singular vectors of each image cluster is computed. The space spanned by these vectors has information about each cluster mean from the dominant singular vector of that cluster and variance of each cluster from its second most dominant singular vector. In section 4 both these approaches are shown to perform better than Eigenfaces for image recognition and better than Fisherfaces for image reconstruction.

2 Background

In this section we discuss briefly the techniques of using PCA to construct Eigenfaces and Fisher Discriminant Analysis to construct Fisherfaces. Both these techniques are well known and we review them here to facilitate a comparision with the new algorithms for Meanfaces and Svdfaces described in the next section.

The dataset we consdier contains n images of human faces each with a resolution of $l \times b$ in 8-bit grayscale. We denote by N, $N = l * b$, the number of bytes it takes to store an image in the original format, in uncompressed form. We split the dataset into a training set with n_{tr} images and a testing set with n_{te} images. We have

$$n = n_{tr} + n_{te}. \tag{1}$$

We assume that the training set is partitioned into c classes. The partition may be done manually, based on prior knowledge about the content of the images. For e.g., in the classic image recognition problem, the training set consists of c persons, with n_j images of each person, taken under varying conditions of light, pose, and facial expression. There has also been some work done on images with different backgrounds and automatically isolating and extracting the human image from the whole picture. [8] The partition may also be done according to various characteristics of the image such as sex, race, presence of facial hair, presence of glasess etc. A $c - means$ type of clustering algorithm may also be used to automatically generate the clusters.

We denote an image as a column vector $\mathbf{x}$, with $\mathbf{x_i}$ denoting the i^{th} image, with $1 \leq i \leq n$. The symbol $\mathcal{C}_j$ denotes the j^{th} class, with $1 \leq j \leq c$. We introduce the mean image of the entire dataset $\mathbf{m}$ as

$$\mathbf{m} \equiv \frac{1}{n} \sum_{i=1}^{n} \mathbf{x_i}. \tag{2}$$

Similarly we have $\mathbf{m_j}$ the mean image of each cluster defined as

$$\mathbf{m_j} \equiv \frac{1}{n_j} \sum_{\mathbf{x_i} \in \mathcal{C}_j} \mathbf{x_i}. \tag{3}$$

We define as mean subtracted image $\bar{\mathbf{x}}_\mathbf{i}$ as

$$\bar{\mathbf{x}}_\mathbf{i} \equiv \mathbf{x_i} - \mathbf{m} \tag{4}$$

We assume that each image $\mathbf{x_i}$ and $\bar{\mathbf{x}}_\mathbf{i}$ are normalized.

2.1 Eigenfaces

In [1, 2] Sirovich and Kirby show that the optimal basis for representing an image vector is given by the eigenspace corresponding to the nonzero eigenvalues of the covariance matrix of all images. Let

$A \equiv [\bar{\mathbf{x}}_\mathbf{1}, \bar{\mathbf{x}}_\mathbf{2}, ..., \bar{\mathbf{x}}_{\mathbf{n_{tr}}}]$ be a $N \times n_{tr}$ matrix containing all the training images. The covariance matrix is defined as

$$C \equiv \frac{1}{n_{tr}} A A^T, \tag{5}$$

where A^T denotes the transpose of A. The matrix C is of dimension $N \times N$ and rank $n_{tr} - 1$. Thus C has $n_{tr} - 1$ nonzero eigenvalues with N dimensional eigenvectors denoted by $\mathbf{e_i}$. The spectrum of C can be shown to be nonnegative. We index the eigenvectors in descending order of the eigenvalue associated with them, thus $\mathbf{e_1}$ is the eigenvectors corresponding to the dominant eigenvalue of C. The set of orthonormal vectors $\{\mathbf{e_i}\}$ which span a space of dimension $n_{tr} - 1$ form the eigensubspace into which images are projected. We denote by W_e the $N \times n_{tr}$ dimensional matrix formed by all the eigenfaces.

Using all the eigenfaces in the basis, any training image can be reconstructed perfectly. More significantly, using only a few of the dominant eigenvectors a good approximation of the original image may be reconstructed. Images from the testing set may also be reconstructed to varying degrees of accuracy (depending on how closely the testing images are to the training set) using a few dominant eigenfaces. In section 4 we present results on the use of eigenfaces for training image reconstruction and testing image recognition.

2.2 Fisherfaces

Fisher's Linear Discriminant analysis has been used traditionally to classify objects. The objective here is to find a set of basis vectors $\mathbf{f_j}$, such that the ratio J is maximised,

$$J = \frac{\sum_{j=1}^{c} ||\mathbf{m}_j - \mathbf{m}||}{\sum_{j=1}^{c} \sum_{\mathbf{x_i} \in \mathcal{C}_j} ||\mathbf{x}_i - \mathbf{m}_j||}, \tag{6}$$

where $||\ ||$ indicates the L^2 norm in the projected subspace, formed by the $\mathbf{f_j}$ vectors. This corresponds to finding a subspace that maximizes the ratio of distance between sum of class means to the sum of variance within each class mean.

To express criterion (6) in matrix notation we define the between-class-scatter matrix, S_B, as

$$S_B = \sum_{j=1}^{c} n_j (\mathbf{m}_j - \mathbf{m})(\mathbf{m}_j - \mathbf{m})^T \tag{7}$$

and the within-class-scatter matrix, S_W, as

$$S_W = \sum_{j=1}^{c} \sum_{\mathbf{x_i} \in \mathcal{C}_j} n_j (\mathbf{x}_i - \mathbf{m}_j)(\mathbf{x}_i - \mathbf{m}_j)^T \tag{8}$$

Criterion (6) is equivalent to determining the matrix W_f, defined as

$$\begin{aligned} W_f &= \max_W \frac{|W^T S_B W|}{|W^T S_W W|} && (9) \\ &= [\mathbf{f}_1, \mathbf{f}_2, \ldots \mathbf{f}_{c-1}] && (10) \end{aligned}$$

where the maximum is taken over all possible $N \times c - 1$ real matrices, and $|X|$ denotes determinant of a matrix X. It can be shown that the vectors $\mathbf{f}_j$ are the eigenvectors corresponding to the j^{th} largest nonzero eigenvalue of the of the follwing generalized eigenvalue problem

$$S_B \mathbf{f}_j = \lambda_j S_W \mathbf{f}_j, \quad j = 1, 2, \ldots c - 1 \tag{11}$$

From the previous discussion on Eigenfaces we know that that the space by the columns of W_e contain all the information to reconstruct a training image. Therefore we consider the projections of matrices S_B and S_W in the eigensubspace of matrix A. Since S_W is of rank $n_{tr} - c$ and we wish to obtain a nonsingular projection of S_W we choose only the first $n_{tr} - c$ columns of W_e to construct projections of S_W and S_B, defined as

$$\begin{aligned} \tilde{S}_B &= W_e^T S_B W_e && (12) \\ \tilde{S}_W &= W_e^T S_W W_e, && (13) \end{aligned}$$

where W_e is a $N \times n_{tr} - c$ matrix here. The Fisherfaces $\mathbf{f_j}$ may then be obtained as the usual eigenvectors corresponding to nonzero eigenvalues of the matrix $\tilde{S}_B \tilde{S}_W^{-1}$. Since the rank of $\tilde{S}_B$ is $c - 1$ there are atmost $c - 1$ Fisherfaces.

Fisherfaces are very effective in recognising test images and assiging them to the right cluster. However they perform poorly in reconstructing projected images. Results of our study on this algorithm is presented in section 4.

3 New approach

In this section we propose and analyze two simple methods to obtain appropriate subspaces for projecting the initial dataset. These algorithms are computationally much cheaper. Their performance will be discussed in section 5.

3.1 Meanfaces

The projection subspace for the Meanfaces approach is given by the span of $\bar{\mathbf{m}}_j$ defined as

$$\bar{\mathbf{m}}_j = \mathbf{m_j} - \mathbf{m} \quad \mathbf{j} = \mathbf{1}, \mathbf{2}, ..., \mathbf{c} \tag{14}$$

The rank of this space is $c - 1$ since we have the condition

$$\sum_{j=1}^{c} \bar{\mathbf{m}}_j = \mathbf{0} \tag{15}$$

We form an orthonormal set from the set of c $\bar{\mathbf{m}}_j$ vectors by using a Gram-Schmidt procedure. This results in a set of $c - 1$ vectors which we denote as $\mathbf{gs_j}$. The matrix S_B defined in the previous section has rank $c - 1$. From the arguments presented in the eigenfaces section we know that there will be $c - 1$ nonzero eigenvectors of $S_B S_B^T$, and a projection onto this eigenspace will maximize the distance between the cluster means as compared to any other $c - 1$ dimensional subspace. Now span of the set of eigenvectors of $S_B S_B^T$ is the same as the span of the $c - 1$ set of $\mathbf{gs_j}$ vectors, since both are linearly independent sets of size $c-1$ and the $\mathbf{gs_j}$ vectors can be expressed as a linear combination of the nonzero eigenvectors of $S_B S_B^T$. Since $c - 1 \ll n_{tr}$, we do not need to form the eigenvectors of $S_B S_B^T$ and choose a dominat few. We need all the nonzero eigenvectors of $S_B S_B^T$ and that eigensubspace is the same as the span of all the $\mathbf{gs_j}$ vectors. It is easy to see that the $\mathbf{gs_j}$much easier to compute than the eigenvectors of $S_B S_B^T$.

Is maximizing the distance between the cluster means a valid criterion for face recognition and/or reconstruction? For face recognition, since this method does not consider the variance within a cluster, is possible that the clusters get smeared out. We expect this method to be sufficient when the datset contains c very diffrent clusters, but the variance within a cluster is not high. This means that the images within a cluster should be similar. Also by using a k-nearest neighbour classifier, instead of a most nearest neighbour classifier we expect to improve the recognition accuracy. For reconstruction, the number of basis vectors that can be formed by this method is limited to $c - 1$. So unless the number of clusters is comparable to the n_{tr} not possible to obtain a good reconstruction, since even the number of Eigenfaces necessary to obtain a good approximation is comparable to n_{tr} [1, 2].

In the next section we show the results of applying Meanfaces to dataset. As can be expected from this analysis, for reconstruction, Meanfaces performs better than Fisherfaces but is not as optimal as Eigenfaces. For recognition, Meanfaces is better than eigenfaces, but not as good as Fisherfaces.

3.2 Svdfaces

To overcome the difficulty with a fixed number of basis vectors in the Meanfaces approach, we propose the use of Svdfaces. For each cluster $\mathcal{C}_j$ we construct a $N \times n_j$ matrix C_j formed by placing the vectors $\mathbf{x}_i$ of that cluster along each column. Then a fixed number of dominant singular vectors of this class is constructed. The constructed singular vectors are then orthogonalised by a standard Gram-Schmidt procedure. This is cheaper than obtaining eigenfaces, because only the dominant few singular vectors are needed and the singular decomposition is cluster by cluster. So we expect this method to scale well with a large number of clusters.

The dominant singular vector of each class contains information about the class mean $\mathbf{m_j}$, the next dominant vector contains information about the variance within each cluster and so on to higher moments. Buy choosing the first two singular vectors from each class we will a get $2c$ dimensional subspace for projection. We expect this give better results for image reconstruction. Results of this algorithm are presented in the next section.

4 Results

In this section we discuss the dataset we have used, the experiments e have tried and the results obtained.

We have used images from the freely available Yale Face Database for all our experiments. The images are available for download at *http://cvc.yale.edu/projects/yalefaces/yalefaces.html*. The original resolution of the images is 242×320 in 8-bit grayscale. There are 15 subjects, with 11 images of each person under varying conditions of light, pose and presence/absence of glasses. For our experiment we chose a fixed 10 of the 11 images per person and reduced the resolution of the image by a factor of two to obtain the number of bytes per image, $N = 19360$. We used a total of $n = 150$ images in our experiments. The variation among the images we used per subject is illustrated in figure 1.

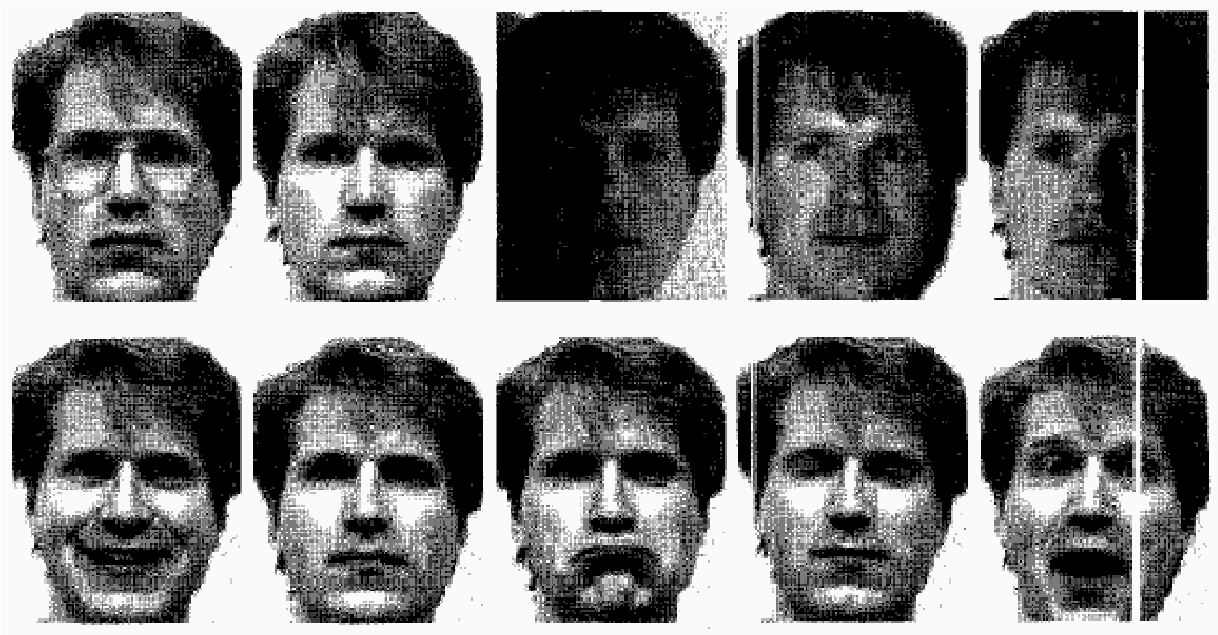

Figure 1: Subject 1 in 10 diffrerent poses and lighting condidtions. In the first image the subject is wearing glasses

We split the dataset into a training set and a testing set. In each experiment we choose tr, the number of images per person in the training set. We varied tr in the range $1 \leq tr \leq 10$, and studied the accuracy of the different methods as a function of the size of the training set. All images not part of the training set are placed in the testing set. We study four different algorithms, Eigenfaces, Fisherfaces, Meanfaces and Svdfaces, described in sections 2 and 3.

4.1 Reconstruction

To determine the accuracy of reconstruction we fixed $tr = 10$ and had no images at all in the testing set. Using each of the four methods described previously we determine the projection subspace and project all the images into this subspace. Then we reconstruct the image using the projected vectors and the basis vectors of the subspace. We caclulate the reconstruction error $E(k, \mathbf{x})$ for each mean subtarcted image $\mathbf{x}$ as a function of the number of leading basis vectors, k, used in the reconstruction. The reconstruction error E for Eigenfaces, Meanfaces and Fisherfaces is defined as

$$E(k, \mathbf{x}_i) = || \sum_{j=1}^{k} a_{i,j} \mathbf{y}_j + \mathbf{m} - \mathbf{x}_i ||, \tag{16}$$

where $||\ ||$ denotes the L^2 norm in the N dimensional space, $\mathbf{y}_j$ is the j^{th} basis vector in the chosen algorithm and $a_{i,j}$ is the projection if the i^{th} mean–subtaracted image $\bar{\mathbf{x}}_i$, with the j^{th} basis vector, i.e.

$$a_{i,j} = \langle \mathbf{y}_j | \bar{\mathbf{x}}_i \rangle \tag{17}$$

with $< | >$ denoting the usual dot-product in N dimensional space. Since the projection subspace is constructed after subtraction of the global mean $\mathbf{m}$ in all these three methods, we have add $\mathbf{m}$ back before calculating the error with the original image $\mathbf{x}_i$. Note that for Fisherfaces the S_B matrix is global mean subtracted. For Svdfaces, since we do not subtract the global mean, the reconstruction error is defined as

$$E(k, \mathbf{x}_i) = || \sum_{j=1}^{k} a_{i,j} \mathbf{y}_j + - \mathbf{x}_i ||, \tag{18}$$

with $a_{i,j} = \langle \mathbf{y}_j | \mathbf{x}_i \rangle$ and the other symbols have their usual meaning. The average over all 150 images of the reconstruction error is shown in figure 2.

Note that Eigenfaces has a maximum dimensionality of $n_{tr} - 1 = 149$, while Fisherfaces and Meanfaces have a maximum dimensionality of $c - 1 = 14$ and SvdFaces have maximum dimensionalities of $c = 15$ and $2c = 30$ for when including one and two singular vectors per class respectively.

Dear Professor, Sailesh : We have not had time to make changes to the graph shown above. Right now I don't have reconstruction error for Eigenfaces for this dataset and also the reconstruction error for Meanfaces and Fisherfaces is not mean subtarcted. We will do those calculations over the weekend and send you an updated version of the report next week. Since the relation between curves may change by these new calculations, we make only general comments about which we are sure right now.

From figure 2 we can see that Fisherfaces do not offer significant improvement in reconstruction by increasing the number of basis vectors. This is because the Fisherfaces contain minimum information about the variation within a class. Also as expected Svdfaces with two singular vectors does better at reconstruction when compared to Svdfaces with only dominant singular vectors simply because the dimensionality of the space is doubled.

4.2 Recognition

In this experiment we study the percentage of testing images accurately recognised by each algorithm. Each testing image $\mathbf{x}$ is projected onto the subspace chosen by the algorithm. Then the L^2 norm of the projected test image with all the projected training images is calculated. The five training images which are closest (in L^2 norm) to the testing image are isolated. The testing image is detrmined to correspond to the person who has a majority in the list of 5 images. In case the list contains images of 5 different persons, the tie was broken arbitrarily. In this experiment tr, the number of training images per person, was varied as $1 \leq tr \leq 9$. For a fixed value of tr all the images not in the training set are in the testing set. We find that the results are strongly dependent on which tr images are placed in the training set and which $10-tr$ images are palced in the testing set. Note that in our experiment there are 10 images of each subject. To reduce the imapact of this dependency we considered 10 different ways of partioning the 10 images per person into two sets, one with tr training images and the other with $10-tr$ testing images per person, and averaging the recognition accuracy rate over all the 10 partitions and all the testing faces in each partition.

Figure 3 shows the result of such an experiment for Meanfaces and Svdfaces As expected, the recognition percentage increases as the number of training images are increased. At $tr = 9$ the recognition efficiency is 80%. The curve drops rather sharply around $tr = 4$ suggesting that a reasonable level of accuaracy (for these methods) may be acheived by keeping just 4 training images per person. It must be emphasised that the presence/absence of information about the overall mean image $\mathbf{m}$ in the projection baisis is irrelevant for face recognition. That is why Meanfaces and Svd1 give approximately the same recognition rate. It is well known that the dominant singular vector of a class j is similar to the class mean $\mathbf{m}_j$. In Meanfaces we subtract $\mathbf{m}$ from $\mathbf{m}_j$, but in Svd1 we do not. It is surprising that Svd2 performs slightly poorly when compared to Svd1. This may be because now the cluster means are not really maximised, while the within-cluster variance is not reduced at all. Note that algorithm Svd2 is not applicable below two training images per cluster.

Due to machine and time constraints we were unable to run such extensive averages for the Eigenfaces and Fisherfaces algorithms. These methods are computationally much more expensive as discussed in section 5. In figure 4 we present results from the same experiment as above but with only 2 partitions considered for each value of tr. We still average over all images the testing set, but the number of partions of the 10 images of each person for a fixed value of tr is reduced from 10 in the previous experiment to 2. The dependence on the partition is clearly visible in the non-monotonic nature of the curves. By carrying out a more extensive average we expect to smooth out the non-monotonic feature. Figure 4 shows the results of this experiment when applied to Eigenfaces, Fisherfaces and Meanfaces. It is interesting to note that Meanfaces performs better than Eigenfaces for intermediate values of tr. Previously many authors [3] have used Eigenfaces for face recognition. Our results show that better results may be obtained by computationally much cheaper method. The dip in the Eigenfaces and Meanfaces curves at $tr = 9$ is due to having a testing image which is strongly shadowed on one side for all the subjects. Since the total number of testing images for all runs when $tr = 9$ is only 20, and 10 of them are stringly shadowed, the recognition rate drops close to 50% because most of the strongly shadowed images are not recognised correctly by these two algorithms. Fisherfaces meanwhile accomodates shadowed images significantly better, but the presence of shadowed images causes some degradation in recognition accuracy for Fisherfaces too.

5 Discussion and Evaluation

In this section we discuss the computation required to calculate the basis vectors for each algorithm.

Eigenfaces It is well known that computation of eigenvectors is a $O(n^3)$ problem, where n is the dimension of the matrix to be diagonalised. In this approach a n_{tr} dimensional matrix needs to be diagonalized, so the computation is of order $O(n_{tr}^3)$. Also the overall mean image has to be subtarcted from the training set, which is a order $O(N * n_{tr})$ operation. For approximate reconstruction all $n_{tr} - 1$ eigenvectors are not needed, but for the study reported in [1] with $n_{tr} = 143$, approximately 40 eigenvectors were needed, which is of the same order of magnitude. Currently, it is clearly known how the number of required eigenvectors scales with n_{tr}.

Fisherfaces This algorithm is the most expensive of approaches considered in this paper. In the first step $n_{tr} - c$ eigenvectors of a n_{tr} dimensional matrix has to be constructed. Then four matrix multiplications are required to project S_B and S_W into the eigenspace. This requires approximately order $O(N^3)$ operations. The $\tilde{S}_B$ has to be inverted which is approximately $O\left((N-c)^3\right)$. Finally a c dimensional matrix has to be diagonalized, which is relatively cheap since $c \ll N$. The total cost is of order $O(N^3 + n_{tr}^3)$.

Meanfaces This method is the cheapest and requires order $O(N*n_{tr})$ operations to calculate the mean images and the another $O(N*n_{tr})$ operations to orthogonalise the meanimages.

Svdfaces Calculation of the dominant singular vectors of a $N \times n_j$ matrix is relatively cheap, compared to a complete eigenspace decomposition. As before the orthogonalization takes order $O(N*n_{tr})$ operations.

These evaluations lead us to conclude that Meanfaces and Svdfaces are significantly cheaper to compute and therefore can be applied to large training sets.

6 Future Work

We see two directions in which the work on this report can be improved upon.

Since the new approaches proposed in this article have computation efficiency as their cheif strength, it is imperative that a large scale study be undertaken with significantly more number of training images. A large number of images (around 4000) are available at the Purdue face database, available at *http://rvl1.ecn.purdue.edu/ aleix/aleix_face_DB.html*

The application of wavelets to reduce the size of an image vector from N to something smaller is also promising. This is because wavelets operate on the principle of storing the difference between neighbouring pixels at various scales. At small scales pixel maps of face images have very similar values for large areas of the human face. Therefore a few levels of wavelet decomposition should result in a seignificant reduction in N. This feature will also help in analyzing large databases of face images.

7 Conclusion

We have proposed two new algorithms to tackle the problems of face recognition and reconstruction/efficient storage. Both these methods are computationally much cheaper than standard Eigenfaces and Fisherfaces methods. It is well-known that Eigenfaces performs better for face reconstruction than Fisherfaces, while Fisherfaces performs better than Eigenfaces for face recognition. For both recognition and reconstruction problems, the two methods proposed here perform intermediately between Eigenfaces and Fisherfaces. For face recognition this is noteworthy since previous authors [3] have used Eigenfaces for face recognition. We have shown here that better results can be acheived with a much cheaper method. Work is in progress in applying the new techniques to larger databases of face images.

References

[1] L. Sirovich and M. Kirby, "A low–dimensional procedure for charecterization of human faces," *J. Opt. Soc. Amerr. A*, vol. 4, no. 3, pp. 519-524, 1987.

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

[3] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. for Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.

[4] P. Belhumeur, J. Hespanha and D. Kriegman, "Eigenfaces vs Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.

[5] B. Moghaddam, W. Wahid and A. Pentland, "Beyond Eigenfaces: Probabilistic matching for face recognition," 3^{rd} *IEEE Conf. Automatic Face and Gesture Recognition*, April 1998.

[6] A. Pentland, B. Moghaddam and T. Starner, , "View–based and modular eigenspaces for face recognition," *IEEE Conf. on Comp. Vision and Pattern Recognition*, 1994.

[7] R. Chellappa, C. Wilson and S. Sirohey, "Human and Machine Recognition of Faces: A Survey," *Proc. IEEE*, vol. 83, no. 5, pp. 705-740, 1995

[8] H. A. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp 23-38, January 1998.

[9] R. Duda and P. Hart, *Pattern Classification and Scene Analysis* Wiley-Interscience, 1973.

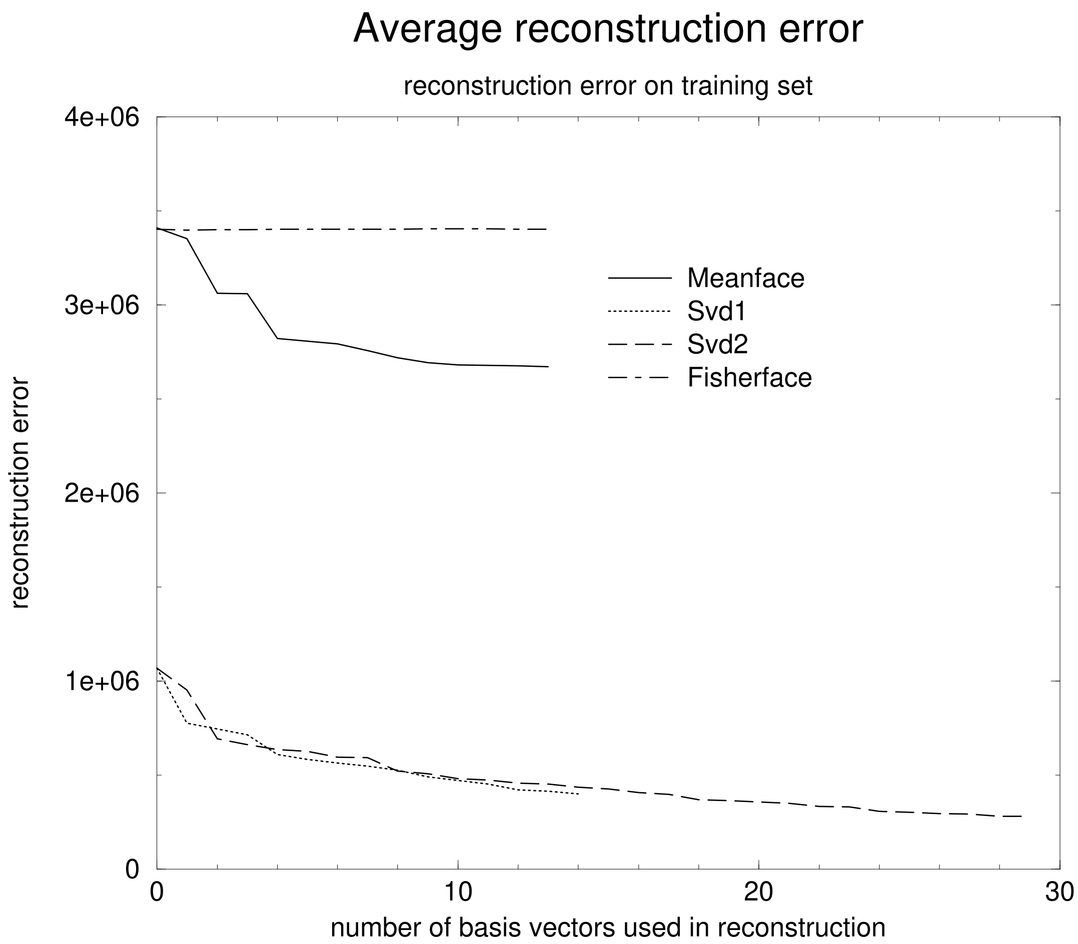

Figure 2: Average reconstruction error as a function of number of basis vectors used for the different algorithms considered in this paper. The algorithms have projection subspaces of different maximum dimension so the curves stop for different values of k.

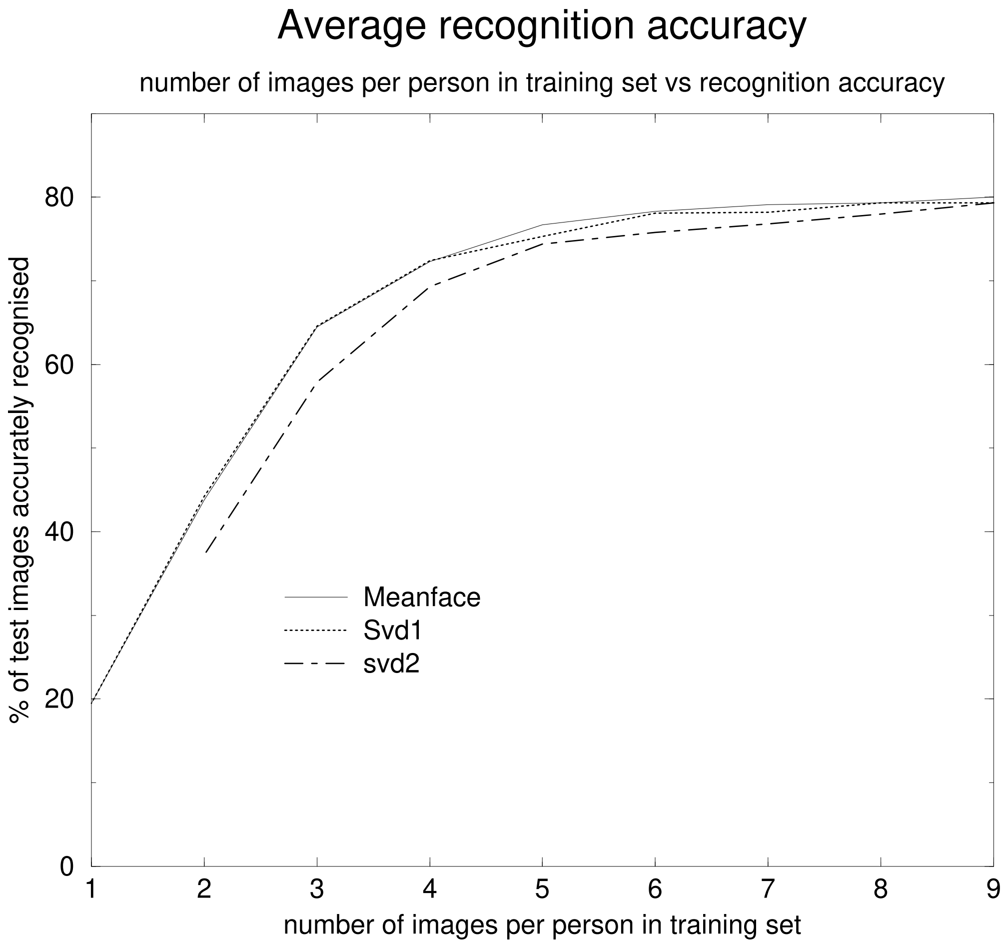

Figure 3: Percentage of figures accurately recognised by Meanfaces and Svdfaces.The number of training images used person is varied along the x-axis. Svd1 refers to Svdfaces with only the dominant singular vector per class and Svd2 refers to Svdfaces with 2 dominant vectors per class.

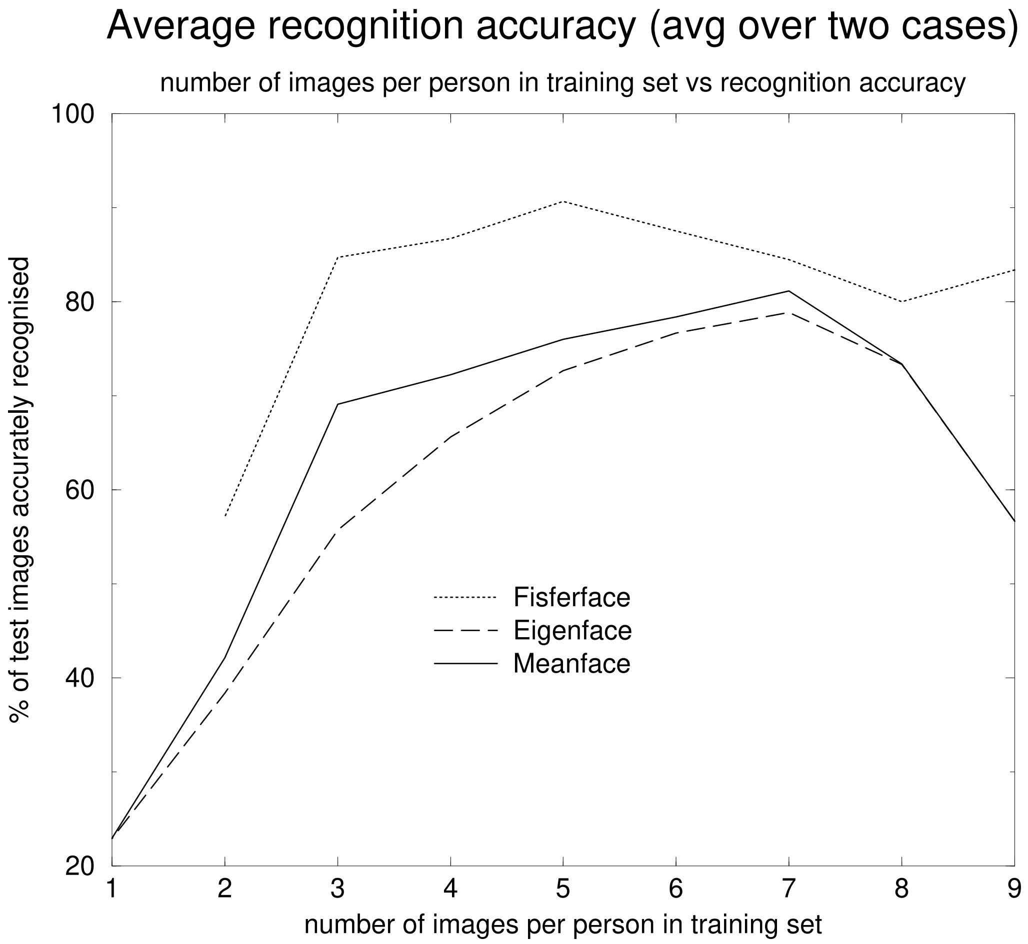

Figure 4: Percentage of figures accurately recognised by Eigenfaces, Fisherfaces and Meanfaces. The number of training images used per person is varied along the x-axis. The non-monotonic nature of curves is due to fewer number of partitions over which the average is computed.