**Face recognition methods:**

**Eigen faces (via the Principal Component Analysis approach):**

As correlation methods are computationally expensive and require great amounts of storage, it is natural to pursue dimensionality reduction schemes. A technique now commonly used for dimensionality reduction in computer vision—particularly in face recognition—is principal components analysis (PCA). PCA techniques, also known as Karhunen-Loeve methods, choose a dimensionality reducing linear projection that maximizes the scatter of all projected samples. More formally, let us consider a set of *N* sample images {**x**1, **x**2, …, **x***N*} taking values in an *n*-dimensional image space, and assume that each image belongs to one of *c* classes {*X*1,*X*2,…,*Xc*}. Let us also consider a linear transformation mapping the original *n*-dimensional image space into an *m*-dimensional feature space, where *m* < *n*. The new feature vectors ****are defined by the following linear transformation:



where *n* is the number of sample images, and **is the mean image of all samples, then after applying the linear transformation *WT* , the scatter of the transformed feature vectors {**y**1,**y**2,…,**y***N}* is *WTSW T* . In PCA, the projection *Wopt* is chosen to maximize the determinant of the total scatter matrix of the projected samples, i.e.,



where **** is the set of *n*-dimensional eigenvectors of *ST* corresponding to the *m* largest eigenvalues. Since these eigenvectors have the same dimension as the original images, they are referred to as Eigenpictures in [6] and Eigenfaces in [7], [8]. If classification is performed using a nearest neighbor classifier in the reduced feature space and *m* is chosen to be the number of images *N* in the training set, then the Eigenface method is equivalent to the correlation method in the previous section. A drawback of this approach is that the scatter being maximized is due not only to the between-class scatter that is useful for classification, but also to the within-class scatter that, for classification purposes, is unwanted information. Recall the comment by Moses et al. [9]: Much of the variation from one image to the next is due to illumination changes. Thus if PCA is presented with images of faces under varying illumination, the projection matrix *Wopt* will contain principal components (i.e., Eigenfaces) which retain, in the projected feature space, the variation due lighting. Consequently, the points in the projected space will not be well clustered, and worse, the classes may be smeared together. It has been suggested that by discarding the three most significant principal components, the variation due to lighting is reduced. The hope is that if the first principal components capture the variation due to lighting, then better clustering of projected samples is achieved by ignoring them. Yet, it is unlikely that the first several principal components correspond solely to variation in lighting; as a consequence, information that is useful for discrimination may be lost.

**Fisher Faces (via Linear Discriminant Analysis approach):**

One can perform dimensionality reduction using linear projection and still preserve linear separability. This is a strong argument in favor of using linear methods for dimensionality reduction in the face recognition problem, at least when one seeks insensitivity to lighting conditions. Since the learning set is labeled, it makes sense to use this information to build a more reliable method for reducing the dimensionality of the feature space. Fisher’s Linear Discriminant (FLD) [5] is an example of a *class specific method*, in the sense that it tries to “shape” the scatter in order to make it more reliable for classification. This method selects *W* in [1] in such a way that the ratio of the between-class scatter and the within class scatter is maximized. Let the between-class scatter matrix be defined as



where *µi* is the mean image of class *Xi* , and *Ni* is the number of samples in class *Xi* . If *SW* is nonsingular, the optimal projection *Wopt* is chosen as the matrix with orthonormal columns which maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples, i.e.,



where { wi|i= 1, 2, …, m} is the set of generalized eigenvectors of SB and SW corresponding to the m largest generalized eigenvalues i.e.: 

Note that there are at most *c* - 1 nonzero generalized eigenvalues, and so an upper bound on *m* is *c* - 1, where *c* is the number of classes. See [4]. To illustrate the benefits of class specific linear projection, we constructed a low dimensional analogue to the classification problem in which the samples from each class lie near a linear subspace. Fig. 2 is a comparison of PCA and FLD for a two-class problem in which the samples from each class are randomly perturbed in a direction perpendicular to a linear subspace. For this example, *N* = 20, *n* = 2, and *m* = 1. So, the samples from each class lie near a line passing through the origin in the 2D feature space. Both PCA and FLD have been used to project the points from 2D down to 1D. Comparing the two projections in the figure, *PCA actually smears the classes together* so that they are no longer linearly separable in the projected space. It is clear that, although PCA achieves larger total scatter, FLD achieves greater between-class scatter, and, consequently, classification is simplified. In the face recognition problem, one is confronted with the difficulty that the within-class scatter matrix **is always singular. This stems from the fact that the rank of *SW* is at most *N* - *c*, and, in general, the number of images in the learning set *N* is much smaller than the number of pixels in each image *n*. This means that it is possible to choose the matrix *W* such that the within-class scatter of the projected samples can be made exactly zero. In order to overcome the complication of a singular *SW*, we propose an alternative to the criterion in (4). This method, which we call Fisherfaces, avoids this problem by projecting the image set to a lower dimensional space so that the resulting within-class scatter matrix *SW* is nonsingular. This is achieved by using PCA to reduce the dimension of the feature space to *N* - *c,* and then applying the standard FLD defined by (4) to reduce the dimension to *c* - 1. More formally, *Wopt* is given by:



****

Note that the optimization for *Wpca* is performed over *n* ´ (*N* - *c*) matrices with orthonormal columns, while the optimization for *Wfld* is performed over (*N* - *c*) ´ *m* matrices with orthonormal columns. In computing *Wpca* , we have thrown away only the smallest *c - 1* principal components. There are certainly other ways of reducing the within class scatter while preserving between-class scatter. For example, a second method which we are currently investigating chooses *W* to maximize the between-class scatter of the projected samples after having first reduced the within class scatter. Taken to an extreme, we can maximize the between-class scatter of the projected samples subject to the constraint that the within-class scatter is zero, i.e.:



where W is the set of *n* x *m* matrices with orthonormal columns contained in the kernel of *SW.*

**Applied generalizations – Eigen, Fisher objects, databases for Eigen, Fisher objects and CBIR access methods:**