M.I.T Media Laboratory Perceptual Computing Section Technical Report No. 335
Appears in: *IEEE International Conference on Image Processing, Washington DC, October 1995.*

A Subspace Method for Maximum Likelihood Target Detection

Baback Moghaddam and Alex Pentland
Perceptual Computing Section, The Media Laboratory,
Massachusetts Institute of Technology
20 Ames Street, Cambridge MA 02139, U.S.A.

Abstract

We present an unsupervised technique for visual target modeling which is based on density estimation in high-dimensional spaces using an eigenspace decomposition. A computationally efficient and optimal estimator for a multivariate Gaussian distribution is derived. This density estimate is then used to formulate a *maximum likelihood* estimation framework for visual search and target detection. Our learning technique is applied to the probabilistic visual modeling and subsequent detection of facial features and is shown to be superior to matched filtering.

1 Introduction

The standard detection paradigm in image processing is that of normalized correlation or template matching. However this approach is only optimal in the simplistic case of a *deterministic* signal embedded in additive white Gaussian noise. When we begin to consider a target *class* detection problem — *e.g.,* finding a generic human face or a human hand in a scene — we must incorporate the underlying probability distribution of the object. Subspace methods and eigenspace decompositions are particularly well-suited to such a task since they provide a compact and *parametric* description of the object's appearance and also automatically identify the *degrees-of-freedom* of the underlying statistical variability.

In particular, the eigenspace formulation leads to a powerful alternative to standard detection techniques such as template matching or normalized correlation. The reconstruction error (or residual) of the eigenspace decomposition (referred to as the "distance-from-feature-space" or DFFS [7, 6]) is an effective indicator of similarity. This residual error is easily computed using the projection coefficients and the original signal energy. This detection strategy is equivalent to matching with a linear combination of *eigentemplates* and allows for a greater range of distortions in the input signal (including lighting, and moderate rotation and scale). In a statistical signal detection framework, the use of eigentemplates has been shown to yield superior performance in comparison with standard matched filtering [3][6].

2 Density Estimation

In this paper we show that the DFFS target detection technique introduced by [7] and used by [6] is a partial component of a complete estimate of the target's probability density function. Our approach is based on estimating the probability density or *likelihood* function of the target. However, instead of applying estimation techniques directly in the original high-dimensional space of the training imagery, we use an eigenspace decomposition to yield a computationally feasible estimate. Specifically, given a set of training images $\{\mathbf{x}^t\}_{t=1}^{N_T}$, from the target class Ω, we wish to estimate the class membership or *likelihood* function for this data — *i.e.,* $P(\mathbf{x}|\Omega)$. Before introducing these estimators we briefly review eigenvector decomposition as commonly used in principal component analysis (PCA) [2].

2.1 Principal Component Imagery

Given a set of m-by-n images $\{I^t\}_{t=1}^{N_T}$, we can form a training set of vectors $\{\mathbf{x}^t\}$, where $\mathbf{x} \in \mathcal{R}^{N=mn}$, by lexicographic ordering of the pixel elements of each image I^t. The basis functions in a Karhunen-Loève Transform (KLT) [4] are obtained by solving the eigenvalue problem

$$\Lambda \;=\; \Phi^T \Sigma \Phi \tag{1}$$

where Σ is the covariance matrix of the data, Φ is the eigenvector matrix of Σ and Λ is the corresponding diagonal matrix of eigenvalues. In PCA, a partial KLT is performed to identify the largest-eigenvalue eigenvectors and obtain a principal component feature vector $\mathbf{y} = \Phi_M^T\, \tilde{\mathbf{x}}$, where $\tilde{\mathbf{x}} = \mathbf{x} - \bar{\mathbf{x}}$ is the mean-normalized image vector and Φ_M is a submatrix of Φ containing the principal eigenvectors. PCA can be seen as a linear transformation $\mathbf{y} = \mathcal{T}(\mathbf{x}) : \mathcal{R}^N \rightarrow \mathcal{R}^M$ which extracts a lower-dimensional subspace of the KL basis corresponding to the maximal eigenvalues. This corresponds to an orthogonal decomposition of the vector space $\mathcal{R}^N$ into two mutually exclusive and complementary subspaces: the principal subspace (or feature space) $F = \{\Phi_i\}_{i=1}^M$ containing the principal components and its orthogonal complement $\bar{F} = \{\Phi_i\}_{i=M+1}^N$, as illustrated in Figure 1.

In a partial KL expansion, the residual reconstruction error is defined as

$$\epsilon^2(\mathbf{x}) \;=\; \sum_{i=M+1}^{N} y_i^2 \;=\; ||\tilde{\mathbf{x}}||^2 \;-\; \sum_{i=1}^{M} y_i^2 \tag{2}$$

and can be easily computed from the first M principal components and the L_2-norm of the mean-normalized image $\tilde{\mathbf{x}}$. Consequently the L_2 norm of every element $\mathbf{x} \in \mathcal{R}^N$ can be decomposed in terms of its projections in these two subspaces. We refer to the component in the orthogonal subspace $\bar{F}$ as the "distance-from-feature-space" (DFFS)

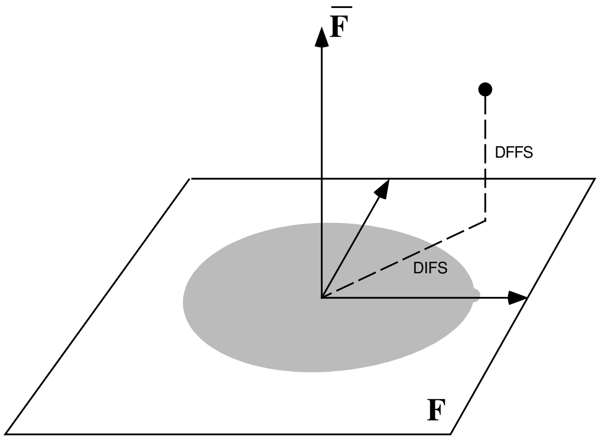


Figure 1: The principal subspace F and its orthogonal complement $\bar{F}$ for a Gaussian density.

which is a simple Euclidean distance and is equivalent to the residual error $\epsilon^2(\mathbf{x})$ in Eq.(2). The component of $\mathbf{x}$ which lies *in* the feature space F is referred to as the "distance-in-feature-space" (DIFS) but is generally not a distance-based norm, but can be interpreted in terms of the probability distribution of y in F.

2.2 Gaussian Subspace Densities

We begin by considering an optimal approach for estimating high-dimensional Gaussian densities. We assume that we have (robustly) estimated the mean $\bar{\mathbf{x}}$ and covariance Σ of the distribution from the given training set $\{\mathbf{x}^t\}$. [1] Under this assumption, the likelihood of a input pattern $\mathbf{x}$ is given by

$$P(\mathbf{x}|\Omega) = \frac{\exp\left[-\frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}})^T\Sigma^{-1}(\mathbf{x}-\bar{\mathbf{x}})\right]}{(2\pi)^{N/2}\ |\Sigma|^{1/2}} \tag{3}$$

The sufficient statistic for characterizing this likelihood is the *Mahalanobis* distance

$$d(\mathbf{x}) = \tilde{\mathbf{x}}^T\Sigma^{-1}\tilde{\mathbf{x}} \tag{4}$$

where $\tilde{\mathbf{x}} = \mathbf{x} - \bar{\mathbf{x}}$. Using the eigenvectors and eigenvalues of Σ we can rewrite Σ^{-1} in the diagonalized form

$$\begin{aligned} d(\mathbf{x}) &= \tilde{\mathbf{x}}^T\Sigma^{-1}\tilde{\mathbf{x}} \\ &= \tilde{\mathbf{x}}^T\left[\Phi\Lambda^{-1}\Phi^T\right]\tilde{\mathbf{x}} \\ &= \mathbf{y}^T\Lambda^{-1}\mathbf{y} \end{aligned} \tag{5}$$

where $\mathbf{y} = \Phi^T\tilde{\mathbf{x}}$ are the new variables obtained by the change of coordinates in a KLT. Because of the diagonalized form, the *Mahalanobis* distance can also be expressed in terms of the sum

$$d(\mathbf{x}) = \sum_{i=1}^{N}\frac{y_i^2}{\lambda_i} \tag{6}$$

[1] In practice, a full rank N-dimensional covariance Σ can not be estimated from N_T independent observations where typically $N_T << N$. But as we shall see our estimator does not require the full covariance, but only its first M principal eigenvectors where $M < N_T << N$.

We now seek to *estimate* $d(\mathbf{x})$ using only the M principal projections. Therefore, we formulate an estimator for $d(\mathbf{x})$ as follows

$$\begin{aligned} \hat{d}(\mathbf{x}) &= \sum_{i=1}^{M}\frac{y_i^2}{\lambda_i} + \frac{1}{\rho}\left[\sum_{i=M+1}^{N} y_i^2\right] \\ &= \sum_{i=1}^{M}\frac{y_i^2}{\lambda_i} + \frac{\epsilon^2(\mathbf{x})}{\rho} \end{aligned} \tag{7}$$

where the term in the brackets is the DFFS $\epsilon^2(\mathbf{x})$, which as we have seen can be computed using the first M principal components. We can therefore write the form of the likelihood estimate based on $\hat{d}(\mathbf{x})$ as the product of two marginal and independent Gaussian densities

$$\begin{aligned} \hat{P}(\mathbf{x}|\Omega) &= \left[\frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{M}\frac{y_i^2}{\lambda_i}\right)}{(2\pi)^{M/2}\prod_{i=1}^{M}\lambda_i^{1/2}}\right]\cdot\left[\frac{\exp\left(-\frac{\epsilon^2(\mathbf{x})}{2\rho}\right)}{(2\pi\rho)^{(N-M)/2}}\right] \\ &= P_F(\mathbf{x}|\Omega)\ \hat{P}_{\bar{F}}(\mathbf{x}|\Omega) \end{aligned} \tag{8}$$

where $P_F(\mathbf{x}|\Omega)$ is the true marginal density in F-space and $\hat{P}_{\bar{F}}(\mathbf{x}|\Omega)$ is the estimated marginal density in the orthogonal complement $\bar{F}$-space. The optimal value of ρ can now be determined by minimizing a suitable cost function $J(\rho)$. From an information-theoretic point of view, this cost function should be the Kullback-Leibler *divergence* [1] between the true density $P(\mathbf{x}|\Omega)$ and its estimate $\hat{P}(\mathbf{x}|\Omega)$

$$J(\rho) = \mathrm{E}\left[\log\frac{P(\mathbf{x}|\Omega)}{\hat{P}(\mathbf{x}|\Omega)}\right] \tag{9}$$

Using the diagonalized forms of the *Mahalanobis* distance $d(\mathbf{x})$ and its estimate $\hat{d}(\mathbf{x})$ and the fact that $\mathrm{E}[y_i^2] = \lambda_i$, it can be easily shown that

$$J(\rho) = \frac{1}{2}\sum_{i=M+1}^{N}\left[\frac{\lambda_i}{\rho} - 1 + \log\frac{\rho}{\lambda_i}\right] \tag{10}$$

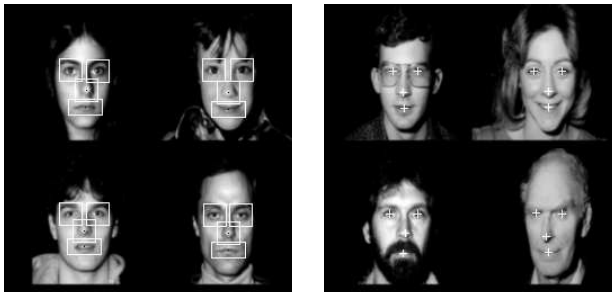
The optimal weight ρ^* can be then found by minimizing this cost function with respect to ρ. Solving the equation $\frac{\partial J}{\partial \rho} = 0$ yields

$$\rho^* = \frac{1}{N-M}\sum_{i=M+1}^{N}\lambda_i \tag{11}$$

which is simply the arithmetic average of the eigenvalues in the orthogonal subspace $\bar{F}$. In addition to its optimality, ρ^* also results in an *unbiased* estimate of the *Mahalanobis* distance — *i.e.,* $\mathrm{E}[\hat{d}(\mathbf{x};\rho^*)] = \mathrm{E}[d(\mathbf{x})]$. What this derivation shows is that once we select the M-dimensional principal subspace F (as indicated, for example, by PCA), the optimal density estimate $\hat{P}(\mathbf{x}|\Omega)$ has the form of Eq.(8) with ρ given by Eq.(11).

It is interesting to consider the minimal cost $J(\rho^*)$

$$J(\rho^*) = \frac{1}{2}\sum_{i=M+1}^{N}\log\frac{\rho^*}{\lambda_i} \tag{12}$$



(a) (b)

Figure 2: (a) Examples of facial feature training templates and (b) the resulting typical detections.

from the point of view of the $\bar{F}$-space eigenvalues $\{\lambda_i : i = M+1, \cdots, N\}$. It is easy to show that $J(\rho^*)$ is minimized when the the $\bar{F}$-space eigenvalues have the *least* spread about their mean ρ^*. This suggests a strategy for selecting the principal subspace: choose F such that the eigenvalues associated with its orthogonal complement $\bar{F}$ have the least absolute deviation about their mean. In practice, the higher-order eigenvalues typically decay and stabilize near the observation noise variance. Therefore this strategy is usually consistent with the standard PCA practice of discarding the higher-order components since these tend to correspond to the "flattest" portion of the eigenvalue spectrum.

3 Maximum Likelihood Detection

The density estimate $\hat{P}(\mathbf{x}|\Omega)$ can be used to form probabilistic *saliency* maps for target detection. This is done by computing the likelihood estimate at each spatial location in the input image

$$S(i,j;\ \Omega) = \hat{P}(\mathbf{x}^{ij}|\Omega) \tag{13}$$

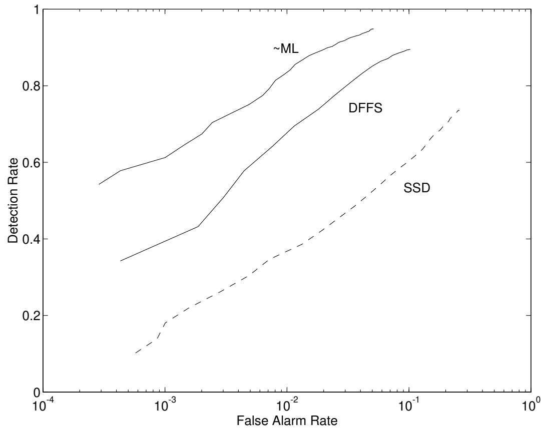
where $\mathbf{x}^{ij}$ is the observation vector obtained by vectorizing the local subimage centered at location (i,j) in the input image. The ML estimate of the spatial position of the target is then obtained by

$$(i^*, j^*) \ = \ \arg\max_{i,j} \ S(i,j;\ \Omega) \tag{14}$$

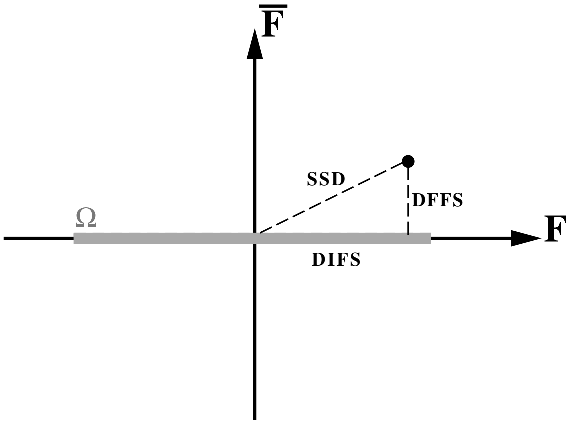
4 Experiments

The eigentemplate approach to the detection of facial features in "mugshots" was proposed in [6], where the DFFS metric was shown to be superior to standard template matching for target detection. The detection task was the estimation of the position of facial features (the left and right eyes, the tip of the nose and the center of the mouth) in frontal view photographs of faces at fixed scale. Figure 2 shows examples of facial feature training templates and the resulting detections on the MIT Media Laboratory's database of 7,562 "mugshots".

We have compared the detection performance of three different detectors on approximately 7,000 test images from this database: a sum-of-square-differences (SSD) detector based on the average facial feature (in this case the left eye), an eigentemplate or DFFS detector and a ML detector based on $S(i,j;\Omega)$ as defined in section 2.2. Figure 3 shows the *receiver operating characteristic* (ROC) curves for these detectors, obtained by varying the detection threshold independently for each detector. The DFFS and ML detectors were computed based on a 5-dimensional principal subspace. Since the projection coefficients were unimodal a Gaussian distribution was used in modeling the true distribution for the ML detector as in section 2.2. Note that the ML detector exhibits the best detection vs. false-alarm tradeoff and yields the highest detection rate (of 95%). Indeed, at the *same* detection rate the ML detector has a false-alarm rate which is nearly 2 orders of magnitude lower than the SSD.



(a)



(b)

Figure 3: (a) Performance of a SSD, DFFS and ML detectors, (b) geometric interpretation of the 3 detectors.

Figure 3(b) provides the geometric intuition regarding the operation of these detectors. The SSD detector's threshold is based on the *radial* distance between the average template (the origin of this space) and the input pattern. This leads to hyperspherical detection regions about the origin. In contrast, the DFFS detector measures the orthogonal distance to F, thus forming planar acceptance regions about F. Consequently to accept valid object patterns in Ω which are very different from the mean, the SSD detector must operate with high thresholds which consequently lead to many false alarms. But at the same time, the DFFS detector can not discriminate between the object class Ω and non-Ω patterns in F. The solution is provided by the ML detector which incorporates both the $\bar{F}$-space component (DFFS) and the F-space likelihood (DIFS). The probabilistic interpretation of Figure 3(b) is as follows: SSD assumes a *single* prototype (the mean) in additive white Gaussian noise whereas the DFFS assumes a *uniform* density in F. The ML detector, on the other hand, uses an estimate for the *complete* probability density function — *i.e.,* DIFS + DFFS.

5 Discussion

We have described an eigenspace density estimation technique for unsupervised visual learning which exploits the *intrinsic* low-dimensionality of the training imagery to form a computationally simple estimator for the complete likelihood function of the object. Our estimator is based on a subspace decomposition and can be evaluated using only the M-dimensional principal component vector. We have derived the form for an optimal estimator and its associated expected cost for the case of a Gaussian density. In contrast to previous work on learning and characterization — which uses PCA primarily for dimensionality reduction and/or feature extraction — our method uses the eigenspace decomposition as an integral part of estimating *complete* density functions in high-dimensional image spaces. These density estimates were then used in a maximum likelihood formulation for target detection, in which substantial performance gain was noted over standard matched filtering. Recently we have developed and tested a *multiscale* version of our algorithm and extended it to the case of non-Gaussian and multimodal distributions using a Gaussian Mixture model [5].

References

[1] Cover, M. and Thomas, J.A., *Elements of Information Theory*, John Wiley & Sons, New York, 1994.

[2] Jolliffe, I.T., *Principal Component Analysis*, Springer-Verlag, New York, 1986.

[3] Kumar, B., Casasent, D., and Murakami, H., "Principal Component Imagery for Statistical Pattern Recognition Correlators," *Optical Engineering,* vol. 21, no. 1, Jan/Feb 1982.

[4] Loève, M.M., *Probability Theory,* Van Nostrand, Princeton, 1955.

[5] Moghaddam, B. and Pentland, A., "Probabilistic visual learning for object detection," *Proc. of International Conference on Computer Vision*, Cambridge, MA, June 1995.

[6] Pentland, A., Moghaddam, B. and Starner, T., "View-based and modular eigenspaces for face recognition," *Proc. of IEEE Conf. on Computer Vision & Pattern Recognition*, June 1994.

[7] Turk, M., and Pentland, A., "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience,* Vol. 3, No. 1, 1991.