Fukunaga-Koontz Transform for Small Sample Size Problems

Abhilash A. Miranda[†] and **Paul F. Whelan**[*]

Vision Systems Group
Dublin City University, Republic of Ireland

E-mail: [†]abhilash@eeng.dcu.ie, [*]whelanp@eeng.dcu.ie

Abstract — **In this paper, we propose the Fukunaga-Koontz Transform (FKT) as applied to Small-Sample Size (SSS) problems and formulate a feature scatter matrix based equivalent of the FKT. We establish the classical Linear Discriminant Analysis (LDA) analogy of the FKT and apply it to a SSS situation. We demonstrate the significant computational savings and robustness associated with our approach using a multi-class face detection problem.**

Keywords — **Fukunaga-Koontz Transform, Linear Discriminant Analysis, Karhunen-Loève Transform, Feature Scatter Matrix**

I Introduction

In high-dimensional signal processing applications such as state-space analysis, image and array processing, etc. it is necessary to analyse the data in a low dimensional subspace. Algorithms based on the Karhunen-Loève Transform (KLT) or eigen analysis are widely used in subspace signal processing, e.g. PCA, MUSIC, ESPRIT etc. These algorithms are computationally efficient in generating a low-dimensional eigen subspace from the available signal samples to which new signal samples could be projected to process them quickly and efficiently.

a) The KLT in Two-Class Problems

In signal processing and pattern recognition, we often encounter the two-class problem where it is necessary to separate or classify the two classes of either the signal, noise or noise corrupted signal from one another. The KLT as used in two-class problems take the following general form: using the samples from one of the two classes $\mathcal{C}_1$ or $\mathcal{C}_2$, an eigen subspace is generated which is expected to capture almost all the variability of that particular class of samples, say $\mathcal{C}_1$. A new signal belonging to $\mathcal{C}_1$ is statistically expected to fit well into this eigen subspace. In many problems such as face recognition, direction-of-arrival estimation, etc., the signal belonging to $\mathcal{C}_1$ would respond very differently from that of $\mathcal{C}_2$ enabling good signal classification performance. The KLT does not tell where in the eigen subspace created using the samples from $\mathcal{C}_1$, would the projections of samples from $\mathcal{C}_2$ might fall. As a result, in many cases when there is considerable overlap between the projection of samples from $\mathcal{C}_1$ and $\mathcal{C}_2$ in the low-dimensional eigen subspace, the KLT-based signal classification might fail.

b) Organisation of the Paper

In this paper we analyse the Fukunaga-Koontz Transform (FKT) which uses the KLT to generate a shared eigenspace for both $\mathcal{C}_1$ and $\mathcal{C}_2$ where their principal eigen subspaces are orthogonal complements of each other. Since the principal eigen subspaces generated using FKT do not overlap, it could be used for effective classification in many difficult two-class problems.

Throughout signal and image processing literature, we can see numerous successful applications of FKT. In this paper, our emphasis is on the small sample size (SSS) problems which are situations where the dimensionality of the signal is larger than the number of signal samples. Such situations are very common in image processing where it is often not possible to get, say for e.g., over ten thousand image samples for 100×100 pixel images. Our motivation is that in SSS situations, the definition of FKT simply does not exist.

In this paper, we first introduce the FKT and examine the properties of its eigen subspaces and explain the process of classification using them. We then see the computational requirements of FKT in high-dimensional problems and discuss the effect of SSS problems. We go on to define the various scatter matrices associated with the available data and derive the FKT equivalent for a SSS problem using the "feature scatter matrix". Next, for addressing the computational requirements of the FKT for high-dimensional problems, we borrow the concepts from classical linear discriminant analysis (LDA). We establish the equivalence of LDA and FKT and use the results obtained from the SSS-FKT problem to derive a computationally simple FKT.

II Fukunaga-Koontz Transform

Let $\mathcal{C}_i, i = 1, 2$ be the classes in a two-class signal identification problem with *a priori* probabilities P_i and class autocorrelation matrices $\mathbf{R}_i = P_i E(\mathbf{x}\mathbf{x}^T | \mathbf{x} \in \mathcal{C}_i)$. Let $\mathbf{Q} = \mathbf{V}_Q \mathbf{\Lambda}_Q \mathbf{V}_Q^T$ denote the eigenvalue decomposition of $\mathbf{Q}$ where $\mathbf{Q} = \mathbf{R}_1 + \mathbf{R}_2$ is the sum of the class autocorrelation matrices. The Fukunaga-Koontz transform (FKT) matrix[1] is defined as [1]

$$\mathbf{S} = \mathbf{V}_Q \mathbf{\Lambda}_Q^{-1/2} \tag{1}$$

which transforms a signal $\mathbf{x}$ to $\mathbf{x_z} = \mathbf{S}^T \mathbf{x}$. From equation (1), it can be easily observed that

$$\mathbf{S}^T \mathbf{Q} \mathbf{S} = \mathbf{I}. \tag{2}$$

Certain applications of the FKT can be seen in [2], [3] and [4] and [5] discusses some of the statistical properties of the FKT.

III Classification using FK Transform

a) *Orthogonality of FK Principal Subspaces*

The class autocorrelation matrices for the FK transformed signals $\mathbf{x_z}$ are $\mathbf{R}_{\mathbf{z}_i} = P_i E(\mathbf{x_z}\mathbf{x_z}^T | \mathbf{x_z} \in \mathcal{C}_i) = P_i E(\mathbf{S}^T \mathbf{x}\mathbf{x}^T \mathbf{S} | \mathbf{x} \in \mathcal{C}_i) = \mathbf{S}^T \mathbf{R}_i \mathbf{S}$, or

$$\mathbf{R}_{\mathbf{z}_i} = \mathbf{S}^T \mathbf{R}_i \mathbf{S}. \tag{3}$$

With the aid of equation (2), it is easy to see that

$$\mathbf{R}_{\mathbf{z}_1} + \mathbf{R}_{\mathbf{z}_2} = \mathbf{S}^T (\mathbf{R}_1 + \mathbf{R}_2) \mathbf{S} = \mathbf{S}^T \mathbf{Q} \mathbf{S} = \mathbf{I} \tag{4}$$

Let the columns of $\mathbf{V}_{\mathbf{z}_i}$ be the eigenvectors of $\mathbf{R}_{\mathbf{z}_i}$ arranged in descending order of their corresponding eigenvalues and let $\hat{\mathbf{V}}_{\mathbf{z}_i}$ denote the first few columns [2] of $\mathbf{V}_{\mathbf{z}_i}$ corresponding to the principal eigenvectors of $\mathbf{R}_{\mathbf{z}_i}$.

It has been proven[1] that the principal eigenvectors of $\mathbf{R}_{\mathbf{z}_1}$ are the least principal eigenvectors of $\mathbf{R}_{\mathbf{z}_2}$ and vice-versa. We evaluate this fact in the form of the following theorem that we propose as follows.

Theorem III.1. *The FK autocorrelation matrices $\mathbf{R}_{\mathbf{z}_1}$ and $\mathbf{R}_{\mathbf{z}_2}$ share the same eigenspace such that their principal eigen spaces are orthogonal complements of each other.*

Proof. Let $\mathbf{e}$ be an eigenvector of $\mathbf{R}_{\mathbf{z}_1}$ with eigenvalue λ. This implies $\mathbf{R}_{\mathbf{z}_1}\mathbf{e} = \lambda\mathbf{e}$. But according to equation (4), $\mathbf{R}_{\mathbf{z}_2}\mathbf{e} = (\mathbf{I} - \mathbf{R}_{\mathbf{z}_1})\mathbf{e} = \mathbf{e} - \mathbf{R}_{\mathbf{z}_1}\mathbf{e} = \mathbf{e} - \lambda\mathbf{e} = (1 - \lambda)\mathbf{e}$. This shows that if $\mathbf{e}$ is an eigenvector of $\mathbf{R}_{\mathbf{z}_1}$ with eigenvalue λ, then $\mathbf{e}$ is also an eigenvector of $\mathbf{R}_{\mathbf{z}_2}$ with eigenvalue $(1 - \lambda)$. Since autocorrelation matrices are positive definite, the eigenvalues of $\mathbf{R}_{\mathbf{z}_1}$ and $\mathbf{R}_{\mathbf{z}_2}$ are all positive. For eigenvalues λ and $1 - \lambda$ to be simultaneously positive, $1 \geq \lambda \geq 0$ for all eigenvalues of $\mathbf{R}_{\mathbf{z}_1}$ and $\mathbf{R}_{\mathbf{z}_2}$. Let us retain M principal normalized eigenvectors of $\mathbf{R}_{\mathbf{z}_1}$ and form $\hat{\mathbf{V}}_{\mathbf{z}_1} = \left[\mathbf{v}_{\mathbf{z}_{1_1}} \ldots \mathbf{v}_{\mathbf{z}_{1_M}}\right]$ corresponding to the eigenvalues $\hat{\boldsymbol{\lambda}}_{\mathbf{z}_1} = (\lambda_{\mathbf{z}_{1_1}} \ldots \lambda_{\mathbf{z}_{1_M}})$ such that $1 \geq \lambda_{\mathbf{z}_{1_1}} \geq \ldots \geq \lambda_{\mathbf{z}_{1_M}} > 0$. Then from KLT, the best approximation of a vector $\mathbf{x} \in \mathcal{C}_1$ in the least square sense using any linear combination of M vectors is obtained using $\hat{\mathbf{V}}_{\mathbf{z}_1}$.

Similarly, suppose we retain $N - M$ principal normalized eigenvectors of $\mathbf{R}_{\mathbf{z}_2}$ and form $\hat{\mathbf{V}}_{\mathbf{z}_2} = \left[\mathbf{v}_{\mathbf{z}_{2_1}} \ldots \mathbf{v}_{\mathbf{z}_{2_{N-M}}}\right]$ corresponding to the eigenvalues $\hat{\boldsymbol{\lambda}}_{\mathbf{z}_2} = (\lambda_{\mathbf{z}_{2_1}} \ldots \lambda_{\mathbf{z}_{2_{N-M}}})$ such that $1 \geq \lambda_{\mathbf{z}_{2_1}} \geq \ldots \geq \lambda_{\mathbf{z}_{2_{N-M}}} > 0$. Then from KLT, the best approximation of a vector $\mathbf{x} \in \mathcal{C}_2$ in the least square sense using any combination of $N - M$ vectors is obtained using $\mathbf{V}_{\mathbf{z}_2}$.

Because, an eigenvector of $\mathbf{R}_{\mathbf{z}_1}$ with eigenvalue λ is also an eigenvector of $\mathbf{R}_{\mathbf{z}_2}$ with eigenvalue $1 - \lambda$, we can write $\hat{\boldsymbol{\lambda}}_{\mathbf{z}_2}$ as: $\hat{\boldsymbol{\lambda}}_{\mathbf{z}_2} = \Big((1 - \lambda_{\mathbf{z}_{1_{M+1}}}) \ldots (1 - \lambda_{\mathbf{z}_{1_N}})\Big)$. Hence, $\hat{\boldsymbol{\lambda}}_{\mathbf{z}_2}$ corresponds to the set of least important $N - M$ orthonormal eigenvectors of $\mathbf{V}_{\mathbf{z}_1}$ which we denote [2] as $\tilde{\mathbf{V}}_{\mathbf{z}_1}$. So we can write

$$\hat{\mathbf{V}}_{\mathbf{z}_2} = \tilde{\mathbf{V}}_{\mathbf{z}_1}. \tag{5}$$

But because of the orthogonality of the eigenvectors in $\tilde{\mathbf{V}}_{\mathbf{z}_1}$ and $\hat{\mathbf{V}}_{\mathbf{z}_1}$, we have [3]

$$(span(\hat{\mathbf{V}}_{\mathbf{z}_1}))^{\perp} = span(\tilde{\mathbf{V}}_{\mathbf{z}_1}). \tag{6}$$

Using equation (5), we write the above equation as

$$(span(\hat{\mathbf{V}}_{\mathbf{z}_1}))^{\perp} = span(\hat{\mathbf{V}}_{\mathbf{z}_2}) \tag{7}$$

We note that the columns of $\hat{\mathbf{V}}_{\mathbf{z}_1}$ is a set of bases for $span(\hat{\mathbf{V}}_{\mathbf{z}_1}) \subset \mathbb{R}^N$. Similarly, the columns of $\hat{\mathbf{V}}_{\mathbf{z}_2}$ is a set of bases for $span(\hat{\mathbf{V}}_{\mathbf{z}_2}) \subset \mathbb{R}^N$. From linear algebra, we understand that this is possible if and only if $span(\hat{\mathbf{V}}_{\mathbf{z}_1})$ and $span(\hat{\mathbf{V}}_{\mathbf{z}_2})$ are *complementary subspaces* of $\mathbb{R}^N$:

$$\Big(span(\hat{\mathbf{V}}_{\mathbf{z}_1}) \oplus span(\hat{\mathbf{V}}_{\mathbf{z}_2})\Big) \subset \mathbb{R}^N. \tag{8}$$

---

[1] $\mathbf{\Lambda}_A = \text{diag}(\lambda_{A_1} \ldots \lambda_{A_N})$ is the diagonal eigenvalue matrix of $\mathbf{A}$, where $\lambda_{A_i} \geq \lambda_{A_j}$ for $i < j$ and $\mathbf{\Lambda}_A^p = \text{diag}(\lambda_{A_1}^p \ldots \lambda_{A_N}^p)$ for $p \in \mathbb{R}$.

[2] $\mathbf{A} = [\hat{\mathbf{A}}\check{\mathbf{A}}]$ implies column partitioning of $\mathbf{A}$ and for eigenvector matrix $\mathbf{V}_A = [\hat{\mathbf{V}}_A \check{\mathbf{V}}_A]$, columns of $\hat{\mathbf{V}}_A$ are principal and columns of $\check{\mathbf{V}}_A$ are least principal eigenvectors

[3] $span(\mathbf{A})$ is the vector space generated by all the linear combinations of the columns of $\mathbf{A}$
$\mathcal{V}^{\perp}$ is the orthogonal complement of vector space $\mathcal{V}$
$\mathcal{V} \oplus \mathcal{U}$ means the direct sum of vector spaces $\mathcal{V}$ and $\mathcal{U}$

From equations (7) and (8) , we deduce that $\hat{\mathbf{V}}_{\mathbf{z}_1}$ and $\hat{\mathbf{V}}_{\mathbf{z}_2}$ span two spaces which are orthogonal to each other and their direct sum spans $\mathbb{R}^N$. But the spaces spanned by $\hat{\mathbf{V}}_{\mathbf{z}_1}$ and $\hat{\mathbf{V}}_{\mathbf{z}_2}$ are the principal subspaces of $\mathbf{R}_{\mathbf{z}_1}$ and $\mathbf{R}_{\mathbf{z}_2}$ respectively and hence the proof.

b) Classification Schemes using $\hat{\mathbf{V}}_{\mathbf{z}_i}$

From linear algebra[6], the best reconstruction of $\mathbf{x} \in \mathcal{C}_i$ in the subspace spanned by $\hat{\mathbf{V}}_{\mathbf{z}_i}$ is given by

$$\hat{\mathbf{x}}_{\mathbf{z}_i} = \hat{\mathbf{V}}_{\mathbf{z}_i} \hat{\mathbf{V}}_{\mathbf{z}_i}^T \mathbf{S}^T \mathbf{x}. \tag{9}$$

By the orthogonal complementary property of the principal eigenspaces, viz., $\hat{\mathbf{V}}_{\mathbf{z}_i} = \hat{\mathbf{V}}_{\mathbf{z}_j}^{\perp}; i, j \in \{1, 2\}; i \neq j$; we could use several schemes to perform classification. Either we could use the neighbourhood of the coordinates of the projection as a classification scheme. Or, for $\mathbf{x}_\mathbf{z} = \mathbf{S}^T \mathbf{x}$, we could evaluate the reconstruction error $||\mathbf{x}_\mathbf{z} - \hat{\mathbf{x}}_{\mathbf{z}_i}||$ based classification criterion i.e,

$$\text{if } ||\mathbf{x}_\mathbf{z} - \hat{\mathbf{x}}_{\mathbf{z}_i}|| < ||\mathbf{x}_\mathbf{z} - \hat{\mathbf{x}}_{\mathbf{z}_j}||, \text{ classify } \mathbf{x} \in \mathcal{C}_i \tag{10}$$

IV Limitations of FK Transform

Though robust features can be extracted using it, there are two fundamental limitations of FKT.

First, there are certain very expensive operations that have to be performed, for e.g., the complete eigenvalue decomposition $\mathbf{Q} = \mathbf{V}_Q \mathbf{\Lambda}_Q \mathbf{V}_Q^T$ of the sum of the class autocorrelation matrices $\mathbf{Q}$ and then the computing of the FKT matrix $\mathbf{S} = \mathbf{V}_Q \mathbf{\Lambda}_Q^{-1/2}$. This has to be followed by the computing of the FKT class autocorrelation matrices $\mathbf{R}_{\mathbf{z}_i} = \mathbf{S}\mathbf{R}_i\mathbf{S}^T$ and then its complete eigenvalue decomposition $\mathbf{R}_{\mathbf{z}_i} = \mathbf{V}_{\mathbf{R}_{\mathbf{z}i}} \mathbf{\Lambda}_{\mathbf{R}_{\mathbf{z}i}} \mathbf{V}_{\mathbf{R}_{\mathbf{z}i}}^T$ to obtain the principal eigenvectors of $\mathbf{V}_{\mathbf{R}_{\mathbf{z}i}}$ onto which we need to project $\mathbf{x}_\mathbf{z} = \mathbf{S}^T \mathbf{x}$ to perform classification. These computations are prohibitive for high-dimensional signals.

Secondly for an SSS problem, $\mathbf{Q}$ is rank-deficient and hence some of its eigenvalues occurring as diagonal elements of $\mathbf{\Lambda}_Q$ would be zeros. Consequently, the inverse $\mathbf{\Lambda}_Q^{-1}$ does not exist and we cannot define the FKT matrix $\mathbf{S} = \mathbf{V}_Q \mathbf{\Lambda}_Q^{-1/2}$.

V Finding S using Feature Scatter

a) Scatter Matrices

Assuming that the mixture mean of the data is subtracted from all the signal samples in the database, we could compute the scatter matrices[7] from the *zero mean* data as follows. The 'between class' scatter matrix signals is $\boldsymbol{\zeta}_b = \sum_{i=1,2} P_i \boldsymbol{\mu}_i \boldsymbol{\mu}_i^T$, where $\boldsymbol{\mu}_i$ is the mean of the class $\mathcal{C}_i$. The 'within class' scatter matrix is defined as

$$\begin{aligned} \boldsymbol{\zeta}_w &= \sum_{i=1,2} P_i E((\mathbf{x} - \boldsymbol{\mu}_i)(\mathbf{x} - \boldsymbol{\mu}_i)^T | (\mathbf{x} \in \mathcal{C}_i)) \\ &= \sum_{i=1,2} P_i E(\mathbf{x}\mathbf{x}^T | x \in \mathcal{C}_i) - \sum_{i=1,2} P_i \boldsymbol{\mu}_i \boldsymbol{\mu}_i^T \\ &= (\mathbf{R}_1 + \mathbf{R}_2) - \sum_{i=1,2} P_i \boldsymbol{\mu}_i \boldsymbol{\mu}_i^T \\ &= \mathbf{Q} - \boldsymbol{\zeta}_b. \end{aligned} \tag{11}$$

The mixture scatter matrix is defined as

$$\boldsymbol{\zeta}_m = \boldsymbol{\zeta}_w + \boldsymbol{\zeta}_b \tag{12}$$

which means that for a mixture mean subtracted dataset, we have

$$\boldsymbol{\zeta}_m = \mathbf{Q}. \tag{13}$$

and by substituting equation (13) in equation (2), we get

$$\mathbf{S}^T \boldsymbol{\zeta}_m \mathbf{S} = \mathbf{I} \tag{14}$$

b) Small Sample Size Eigenvalue Decomposition

Let $\mathbf{H}_m$ denote the matrix whose columns correspond to samples from some database with zero mixture mean. The *signal mixture scatter matrix* for these samples is defined as

$$\boldsymbol{\zeta}_m = \mathbf{H}_m \mathbf{H}_m^T \tag{15}$$

and the *feature scatter matrix* for these samples is defined as

$$\boldsymbol{\chi}_m = \mathbf{H}_m^T \mathbf{H}_m. \tag{16}$$

We write the eigenvalue decomposition of the feature scatter matrix as follows

$$\boldsymbol{\chi}_m = \mathbf{V}_{\chi_m} \mathbf{\Lambda}_{\chi_m} \mathbf{V}_{\chi_m}^T \tag{17}$$

where $\mathbf{\Lambda}_{\chi_m}$ is diagonal, and $\mathbf{V}_{\chi_m}$ is orthogonal. Post-multiplying equation (17) by $\mathbf{V}_{\chi_m}$ followed by pre-multiplying by $\mathbf{H}_m$, we see that

$$\begin{aligned} &\left(\mathbf{H}_m \mathbf{H}_m^T\right) \left(\mathbf{H}_m \mathbf{V}_{\chi_m}\right) = \left(\mathbf{H}_m \mathbf{V}_{\chi_m}\right) \mathbf{\Lambda}_{\chi_m} \\ \text{or} \quad &\boldsymbol{\zeta}_m \left(\mathbf{H}_m \mathbf{V}_{\chi_m}\right) = \left(\mathbf{H}_m \mathbf{V}_{\chi_m}\right) \mathbf{\Lambda}_{\chi_m}. \end{aligned} \tag{18}$$

Hence, we can write equation (18) in the form

$$\boldsymbol{\zeta}_m \mathbf{E}_m = \mathbf{E}_m \mathbf{\Lambda}_{\chi_m} \tag{19}$$

where

$$\mathbf{E}_m = \mathbf{H}_m \mathbf{V}_{\chi_m} \tag{20}$$

which means that the columns of $\mathbf{E}_m$ constitute the orthogonal eigenvectors of $\boldsymbol{\zeta}_m$.

c) Singular Eigenvalues in SSS Problems

Since the eigenvectors corresponding to singular values of $\boldsymbol{\chi}_m$ will also be eigenvectors corresponding to the singular values of $\boldsymbol{\zeta}_m$, we do not need to include those eigenvectors in $\mathbf{V}_{\chi_m}$. As will be discussed later, all eigenvectors in the nullspace of $\boldsymbol{\zeta}_m$ could be safely eliminated.

We have to emphasize here that the feature scatter matrix eigenvalue decomposition is computationally simpler than finding the complete eigenvalue decomposition of $\boldsymbol{\zeta}_m$ and subsequent elimination of its eigenvectors corresponding to singular eigenvalues.

d) Small Sample Size FKT

Due to the equivalence of $\boldsymbol{\zeta}_m$ and $\mathbf{Q}$ in equation (13), we understand that the columns of $\mathbf{E}_m$ are the eigenvectors of $\mathbf{Q}$ corresponding to the non-singular eigenvalues of $\mathbf{Q}$ found as the diagonal elements of $\boldsymbol{\Lambda}_{\chi_m}$. Hence we define the small sample size equivalent of the FKT matrix of equation (1) as

$$\hat{\mathbf{S}} = \mathbf{E}_m \boldsymbol{\Lambda}_{\chi_m}^{-1/2}. \tag{21}$$

VI LDA equivalent of FK Transform

a) Equivalence of $\mathbf{S}$ *and* $\boldsymbol{\Theta}$

In classical linear discriminant analysis(LDA), the quantities $\mathsf{trace}(\boldsymbol{\zeta}_b)$ and $\mathsf{trace}(\boldsymbol{\zeta}_w)$ measure the separation 'between classes' and 'within classes', respectively. The direction of the set of optimal discriminants $\boldsymbol{\Theta}$ should satisfy[7]:

$$\boldsymbol{\Theta} = \left(\max_{\boldsymbol{\Theta}} \mathsf{trace}(\boldsymbol{\Theta}^T\boldsymbol{\zeta}_b\boldsymbol{\Theta})\right) \cap \left(\min_{\boldsymbol{\Theta}} \mathsf{trace}(\boldsymbol{\Theta}^T\boldsymbol{\zeta}_w\boldsymbol{\Theta})\right). \tag{22}$$

We propose the following modification to the above criterion: using equation (12), we understand that we will get the same set of optimal discriminants above if we choose:

$$\boldsymbol{\Theta} = \left(\max_{\boldsymbol{\Theta}} \mathsf{trace}(\boldsymbol{\Theta}^T\boldsymbol{\zeta}_m\boldsymbol{\Theta})\right) \cap \left(\min_{\boldsymbol{\Theta}} \mathsf{trace}(\boldsymbol{\Theta}^T\boldsymbol{\zeta}_w\boldsymbol{\Theta})\right). \tag{23}$$

It is widely known and has been proven in reference [7] that the equation (22) is invariant under any non-singular linear transformation $\boldsymbol{\Theta}$. Hence, we can write the equivalent form of (23) above:

$$\begin{aligned}\boldsymbol{\Theta} &= \min_{\boldsymbol{\Theta}} \mathsf{trace}\left((\boldsymbol{\Theta}^T\boldsymbol{\zeta}_m\boldsymbol{\Theta})^{-1}(\boldsymbol{\Theta}^T\boldsymbol{\zeta}_w\boldsymbol{\Theta})\right) \\ &= \min_{\boldsymbol{\Theta}} \mathsf{trace}(\boldsymbol{\zeta}_m^{-1}\boldsymbol{\zeta}_w).\end{aligned} \tag{24}$$

According to KLT, the set of discriminants $\boldsymbol{\Theta}$ can be obtained by solving the eigenvalue problem:

$$\boldsymbol{\zeta}_m^{-1}\boldsymbol{\zeta}_w\boldsymbol{\theta} = \gamma\boldsymbol{\theta} \tag{25}$$

where γ is the eigenvalue corresponding to the eigenvector $\boldsymbol{\theta}$ of $\boldsymbol{\zeta}_m^{-1}\boldsymbol{\zeta}_w$. The eigenvalue problem in equation (25) can be written as

$$\boldsymbol{\Theta}^T\boldsymbol{\zeta}_w\boldsymbol{\Theta} = \boldsymbol{\Gamma} \tag{26}$$

and

$$\boldsymbol{\Theta}^T\boldsymbol{\zeta}_m\boldsymbol{\Theta} = \mathbf{I} \tag{27}$$

where $\boldsymbol{\Gamma}$ is a diagonal matrix whose diagonal entries are the eigenvalues γ and $\mathbf{I}$ is the identity matrix.

Comparing equation (14) with equation (27), we observe the equivalence

$$\boldsymbol{\Theta} \equiv \mathbf{S} \tag{28}$$

of the set of discriminant vectors $\boldsymbol{\Theta}$ and the FKT matrix $\mathbf{S}$.

b) Feature Scatter based FKT

If we post-multiply equation (19) by $\boldsymbol{\Lambda}_{\chi_m}^{-1/2}$ and compare it with equation (21), we find that

$$\boldsymbol{\zeta}_m\hat{\mathbf{S}} = \hat{\mathbf{S}}\boldsymbol{\Lambda}_{\chi_m} \tag{29}$$

so that $\hat{\mathbf{S}}$ constitutes the orthonormal eigenvector matrix of ζ_m; orthonormal because $\hat{\mathbf{S}}^{-1} = \hat{\mathbf{S}}^T$. From equation (29), now we are able to find the inverse of $\boldsymbol{\zeta}_m = \hat{\mathbf{S}}\boldsymbol{\Lambda}_{\chi_m}\hat{\mathbf{S}}^T$ by just inverting the diagonal entries of $\boldsymbol{\Lambda}_{\chi_m}$ so that

$$\begin{aligned}\boldsymbol{\zeta}_m^{-1} &= \hat{\mathbf{S}}\boldsymbol{\Lambda}_{\chi_m}^{-1}\hat{\mathbf{S}}^T \\ &= \left(\mathbf{E}_m\boldsymbol{\Lambda}_{\chi_m}^{-1/2}\right)\boldsymbol{\Lambda}_{\chi_m}^{-1}\left(\mathbf{E}_m\boldsymbol{\Lambda}_{\chi_m}^{-1/2}\right)^T\end{aligned}$$

which can be written as

$$\boldsymbol{\zeta}_m^{-1} = \mathbf{E}_m\boldsymbol{\Lambda}_{\chi_m}^{-2}\mathbf{E}_m^T. \tag{30}$$

For SSS problems, equation (30) gives the solution of the inverse of the mixture scatter matrix $\boldsymbol{\zeta}_m$ by ignoring the space spanned by its singular eigenvectors. The inverse $\boldsymbol{\zeta}_m^{-1}$ thus computed could be used in equation (25) to obtain the set of discriminant vectors $\boldsymbol{\Theta}$ or equivalently the FKT matrix $\mathbf{S}$. This simply establishes the existence of LDA equivalent of SSS-FKT.

VII Advantages of the Feature-Scatter based FKT

a) Analysis of Discriminant Information

The feature scatter approach eliminates the nullspace $\mathcal{N}(\boldsymbol{\zeta}_m)$ of the mixture pattern scatter matrix completely. In order to analyse the loss of discriminant information caused by avoiding $\mathcal{N}(\boldsymbol{\zeta}_m)$, we re-write the eigenvalue problem in equation (25) as $\boldsymbol{\zeta}_w\boldsymbol{\theta}_j = \gamma_j\boldsymbol{\zeta}_m\boldsymbol{\theta}_j$ and hence, search for the optimal discriminants $\boldsymbol{\theta}_j$ in either of these two cases:

$$\mathsf{Case\ 1:}\quad \{\boldsymbol{\theta}_j \in \mathcal{N}(\boldsymbol{\zeta}_m); \boldsymbol{\theta}_j \in \mathcal{N}(\boldsymbol{\zeta}_w)\} \tag{31}$$

where we have an no changes in the quantities $\mathsf{trace}(\boldsymbol{\Theta}^T\boldsymbol{\zeta}_w\boldsymbol{\Theta}) = \sum \boldsymbol{\theta}_j^T\boldsymbol{\zeta}_w\boldsymbol{\theta}_j$ and $\mathsf{trace}(\boldsymbol{\Theta}^T\boldsymbol{\zeta}_w\boldsymbol{\Theta}) = \sum \boldsymbol{\theta}_j^T\boldsymbol{\zeta}_m\boldsymbol{\theta}_j$.

$$\mathsf{Case\ 2:}\quad \{\boldsymbol{\theta}_j \in \mathcal{N}(\boldsymbol{\zeta}_m); \boldsymbol{\theta}_j \notin \mathcal{N}(\boldsymbol{\zeta}_w)\} \tag{32}$$

where we have an increase in the quantity $\mathsf{trace}(\boldsymbol{\Theta}^T\boldsymbol{\zeta}_w\boldsymbol{\Theta}) = \sum \boldsymbol{\theta}_j^T\boldsymbol{\zeta}_w\boldsymbol{\theta}_j$ and no change in $\mathsf{trace}(\boldsymbol{\Theta}^T\boldsymbol{\zeta}_m\boldsymbol{\Theta}) = \sum \boldsymbol{\theta}_j^T\boldsymbol{\zeta}_m\boldsymbol{\theta}_j$.

Since neither of the two cases further minimises the quantity $\mathsf{trace}(\boldsymbol{\zeta}_m^{-1}\boldsymbol{\zeta}_w)$ of the condition in equation (24), we do not lose any discriminant information by using the feature scatter based approach which eliminates $\mathcal{N}(\boldsymbol{\zeta}_m)$.

b) Computing the SSS-FKT

From the section (a) above, we understand that there is no loss in ignoring the null-space $\mathcal{N}(\boldsymbol{\zeta}_m)$. This would mean that one can proceed with the direct implementation of the FKT which overcomes the limitation of the SSS problem and finds a lower order matrix $\hat{\mathbf{S}}$ which reduces the computation in finding the SSS-FKT class autocorrelation matrices

$$\hat{\mathbf{R}}_{\mathbf{z}_i} = \hat{\mathbf{S}}^T\mathbf{R}_i\hat{\mathbf{S}}. \tag{33}$$

where we use SSS-FKT matrix $\hat{\mathbf{S}}$ in equation (3) instead of the complete FKT matrix $\mathbf{S}$.

It is now required to perform complete eigenvalue decomposition of $\hat{\mathbf{R}}_{\mathbf{z}_1}$ and $\hat{\mathbf{R}}_{\mathbf{z}_1}$ and as explained in section III, arbitrary number of their dominant eigenvectors have to be selected to form the matrices $\hat{\mathbf{V}}_{\mathbf{z}_1}$ and $\hat{\mathbf{V}}_{\mathbf{z}_2}$.

Classification of a mixture mean subtracted signal $\mathbf{x}$ can be performed by first computing its best reconstruction according to equation (9)

$$\hat{\mathbf{x}}_{\mathbf{z}_i} = \hat{\mathbf{V}}_{\mathbf{z}_i}\hat{\mathbf{V}}_{\mathbf{z}_i}^T\hat{\mathbf{S}}^T\mathbf{x} \tag{34}$$

where we use SSS-FKT matrix $\hat{\mathbf{S}}$ instead of complete FKT matrix $\mathbf{S}$ and the reconstruction error scheme explained in III(b) can be used to perform signal classification.

VIII Experiments with SSS-FKT

In order to verify the robustness of the SSS-FKT, we chose the AT&T Database of Faces[4] to conduct experiments on all ${}^{15}C_2 = 105$ combinations $\{\{\mathcal{C}_i, \mathcal{C}_j\} : i, j \in \{1, \ldots, 15\}; i \neq j\}$ of images of 15 individuals viz., $\mathcal{C}_1, \ldots, \mathcal{C}_{15}$ of the database. Each class consists of 10 various poses of the sample image. By combining two classes at a time, a set of 5 poses from each class was used for building the SSS-FKT. The remaining set of 5 poses from each class was used for testing. The size of each image in the database was 92×112 pixels, with 256 grey levels per pixel.

Fig. 1: Sample images from the 15 classes used in our experiments. A total of ${}^{15}C_2 = 105$ two-class tests were performed to evaluate the robustness of the SSS-FKT.

Stage 1: Each of the images were first converted to a 1×10304 dimensional array by lexicographically scanning the image. The mixture mean calculated from all 10 images (5 poses each from a class) was subtracted from the data before any further processing. The matrix of the zero mixture mean data matrix $\mathbf{H}_m$ of equation (15) was used to calculate the feature scatter matrix. The eigenvalue decomposition of feature scatter matrix was used to obtain the SSS-FKT matrix $\hat{\mathbf{S}}$ according to equations (17) to (21). Any singular eigenvectors could be safely eliminated at this stage as explained in section V(c). The SSS-FKT class autocorrelation matrices $\hat{\mathbf{R}}_{\mathbf{z}_i}$ are constructed next using the equation (33). It can be verified at this stage that even the SSS-FKT autocorrelation matrices $\hat{\mathbf{R}}_{\mathbf{z}_i}$ satisfies equation (4).

We are now able to validate the computational benefits discussed in sections V(c) and VII(b). The mixture scatter matrix $\boldsymbol{\zeta}_m$ in this problem has an order 10304×10304. It would be prohibitive, if not impossible, to perform eigenvalue decomposition of $\boldsymbol{\zeta}_m = \mathbf{Q}$ to obtain the FKT matrix according to equation (1). On the other hand, the SSS-FKT requires the eigenvalue decomposition of the feature scatter matrix $\boldsymbol{\chi}_m$ which is of the order 10×10 due to the 10 various poses of face images we consider. Again, at this stage any singular

[4] AT&T Laboratories, Cambridge

eigenvectors of $\boldsymbol{\chi}_m$ could be discarded as explained in VII(b) further reducing the subsequent computations. Building of the SSS-FKT completes the first stage of the problem and the time elapsed for all the processes till we find the eigenvectors $\hat{\mathbf{V}}_{\mathbf{z}_i}$ of $\mathbf{R}_{\mathbf{z}i}$ is not more than 7 seconds on a standard PC for all the 105 tests that we conducted.

Stage 2: In the second stage of the classification problem, the matrix $\hat{\mathbf{V}}_{\mathbf{z}_i}$ is used to reconstruct any mixture mean subtracted image according to equation (34). Now, the simple criterion in equation (10) could be used to perform classification using the reconstructed image.

On testing with the mixture mean subtracted images of the remaining poses of a particular class which were originally not used for creating the SSS-FKT matrix, we were able to achieve an average classification error rate of 4% for the 105 tests that we performed on the different classes $\mathcal{C}_1, \ldots, \mathcal{C}_{15}$ of the database. The computations required for this stage is much simpler than that required for the first stage and we were able to perform classification for a set of 5 test poses from each of the classes in an average time of 0.7 seconds on a standard PC for all the 105 two-class tests that we conducted.

IX Conclusion

In this paper, we have defined the FKT for SSS problems. We derived the LDA equivalent of the FKT and used the concept to prove that by ignoring the nullspace of the mixture scatter matrix to obtain the SSS-FKT, we do not eliminate any useful discriminant information. Also, we have implemented the technique for face detection, which is a high-dimensional SSS problem, and we achieved very good classification results. Though we started with a two-class problem, we foresee that the LDA equivalent of FKT can extend the FKT to multiclass problems.

References

[1] K. Fukunaga and W. Koontz. Application of the karhunen-loève expansion to feature selection and ordering. *IEEE Transactions on Computers*, C-19(4):311–318, 1970.

[2] M.N. Wernick and G.M. Morris. Image classification at low light levels. *Journal of the Optical Society of America A*, 3(12):2179–2187, 1986.

[3] A. Mahanalobis, S.R.F. Sims, and A. Van Nevel. Signal-to-clutter measure for measuring automatic target recognition performance using complimentary eigenvalue distribution analysis. *Optical Engineering*, 42(4):1144–1151, 2003.

[4] J.R. Leger and S.H. Lee. Image classification by an optical implementation of the fukunaga-koontz transform. *Journal of the Optical Society of America*, 5(2):556, 1982.

[5] X. Huo. A statistical analysis of fukunaga koontz transform. *IEEE Signal Processing Letters*, 11(2):123–126, 2004.

[6] C.D. Meyer. *Matrix Analysis and Applied Linear Algebra.* SIAM, Philadelphia, 2000.

[7] K. Fukunaga. *Statistical Pattern Recognition.* Academic Press, San Diego California, 1990.