Modification of Karhunen–Loeve transform for pattern recognition

P BARAT and AMITAVA ROY

Variable Energy Cyclotron Centre, 1/AF Bidhan Nagar, Calcutta 700 064, India
e-mail: pbarat@veccal.ernet.in

MS received 11 June 1997; revised 18 June 1998

Abstract. Application of the Karhunen–Loeve transform in pattern recognition problems necessitates knowledge of the mean vectors for the training sets of different classes for construction of the respective covariance matrices. This modality poses problems for recognizing an unknown signal in a multi-class environment and for its on-line implementation. This is because *a priori* it is unknown which mean vector of the several classes is to be subtracted from an unknown input signal. To remove this difficulty a global mean approach has been proposed. The proposed method has been applied to synthetic and experimentally observed acoustic signals successfully and its performance with respect to pattern recognition and data compression has been compared with the conventional method.

Keywords. Karhunen–Loeve transform; global mean approach; synthetic signal; acoustic signal.

1. Introduction

The Karhunen–Loeve transform (KLT) (Ahamed & Rao 1975, p. 200), is a very old technique for data compression of one- or two-dimensional signals. Its use in a practical sense was not encouraged for a long time because much computer time was necessary for its implementation, even though it was known to be one of the most efficient techniques for data compression. Recently, the availability of fast algorithms proposed by Jain (1976) and Bhama *et al* (1993) for implementing KLT in parallel computers has given it greater impetus.

Pattern recognition was and still remains a difficult problem due to the large degree of variability of the data. The standard recognition paradigm in pattern recognition is that of normalized correlation or template matching. However, this approach is only optimal in the simplistic case. Subspace methods proposed and used by Watanabe and Pakvasa (1973), Oja (1983, 1989), Oja & Kuusela (1983), Oja & Kärki (1987), Oja & Kohonen (1988), Laaksonen & Oja (1996) and Laaksonen (1997), and eigenspace decompositions presented

by Oja (1992), are particularly well-suited to such a task since they provide a compact and parametric description of the input signal vectors and also automatically identify the degrees-of-freedom of underlying statistical variability. They are usually divided into two parts; feature extraction and classification. In feature extraction, the input vectors are mapped into a vector space where its coordinates express the features of the raw data. In classification, the input vector is classified as belonging to the category whose decision region includes its feature vector. For this purpose, the feature vector space is divided into disjoint decision regions corresponding to different categories (Oja 1983, ch. 4).

The fundamental idea behind subspace methods of pattern recognition is to calculate the category of an input feature vector by projecting it onto a number of different linear subspaces and choosing the projection with the largest magnitude. Each subspace corresponds to a different category and is the subspace which gives the best approximation of the stochastic signals of that category.

The subspace, which results from a maximum mean squared criterion (Watanabe & Pakvasa 1973), is spanned by the first few eigen vectors yielded by KLT. The method has been successfully applied to pattern recognition problems (Chien & Fu 1967, 1968; Fu 1968; Tomita *et al* 1970; Rosenfeld 1976; Oja 1983).

So far, applications of KLT to pattern recognition problems were limited to classes of signals which either have zero means or which are preprocessed to have zero means. However, this modality poses problems for processing an unknown signal in a multi-class environment and for its on-line implementation. This is because *a priori* it is not known which mean vector of the several classes is to be subtracted from an unknown input signal vector. In this paper, this inherent limitation of KLT is discussed and a global mean approach is introduced to remove this difficulty. Utility of this new approach is tested by applying it to two classes of synthetic and experimentally observed acoustic signals.

Before we proceed to discuss the global mean approach in KLT, we first briefly review the basic KLT technique.

2. Karhunen–Loeve transform

Most of the attention in the literature concerning feature selection by KLT has been centred around discrete KLT. In this procedure, from the set of “ n ” pattern vectors, $\{X_i\}$, each of “ m ” dimensions, forming a training set of a particular class, the covariance matrix is constructed as, $\sum_x = (1/n) \sum_{i=1}^n X_i X_i^T - \bar{X} \bar{X}^T$, where $\bar{X} = (1/n) \sum_{i=1}^n X_i$ is the mean vector for the class. Each diagonal element of $\sum_x$ represents the variance of the corresponding components of the set $\{X_i\}$. When it is diagonalised by a transformation matrix as $T^T \sum_x T = \sum_x^d$, the eigenvalues λ_j , represent the variance of the transformed domain vectors $|Y_i\rangle = T(X_i - \bar{X})$, for its j th dimension. If it is found that the trace of $\sum_x^d$ is distributed among few of the diagonal elements say m_1 out of m diagonal elements, then the variance in the rest $m_1 + 1$ to m dimensional space of $\{Y_i\}$ is almost zero and is neglected. The new vector $\tilde{X}_i$ transformed from the reduced domain by taking only 1 to m_1 components of the vector Y_i i.e. $\tilde{Y}_i$ having largest variance, is given by $\tilde{X}_i = T^T \tilde{Y}_i + \bar{X}$. The mean square error introduced in this process is $\sum_{i=m_1+1}^n \lambda_i$. It is the minimum error one can obtain in any orthogonal transform (Fukunaga 1972, ch. 8).

The transformation $Y_i = T(X_i - \bar{X})$ as well as construction of covariance matrix $\sum_x$, suggest that the evaluation of $\bar{X}$ is necessary for the application of KLT. For L class problems the covariance matrix for the KLT is constructed as $\sum_x = P_1 \sum_1 + P_2 \sum_2 + \dots + P_L \sum_L$, where $\sum_k$ is the covariance matrix for the k th class constructed from the training set of pattern vectors of that class and P_k is the *a priori* probability usually taken as $1/L$ (assuming equal number of signals in the training set for each class). The problem arises for processing an unknown signal by KLT in multi-class environment because *a priori* one does not know which mean vector is to be subtracted from the unknown signal. To avoid this problem one can think of trying the transform $Y_i = T X_i$ without subtracting the mean. However, this mode of transformation will result in a suboptimal result in KLT (Tou & Gonzalez 1974, p. 279).

3. Global mean approach

In this approach all the pattern vectors for all the classes were taken together to construct a generalized mean vector. A generalized covariance matrix $\sum'_x$ is then formed taking all the vectors for all the classes, assuming as if all of them belong to one class. For an unknown signal, this generalized mean vector is subtracted. It is then transformed to the reduced domain for the purpose of recognition by the transformation matrix, generated by diagonalising $\sum'_x$.

It is now necessary to see the difference in error ' D ', introduced for the reconstruction of pattern vectors from the transform domain, constructed by the eigen vectors of $\sum_x$ and $\sum'_x$ by retaining only first the m_1 eigen vectors out of m eigen vectors corresponding to the highest variances.

Let us designate n vectors of the k th class as $\{X_1^k, X_2^k \dots X_n^k\}$. Assume that there are L such classes and each class has n vectors. The covariance matrix of the k th class is given by

$$\sum^k = 1/n \sum_{i=1}^n X_i^k (X_i^k)^T - \left(1/n \sum_{i=1}^n X_i^k \right) \left(1/n \sum_{i=1}^n (X_i^k)^T \right).$$

Assuming the equal probability P_i for each of the L classes as $1/L$, the covariance matrix $\sum_x$ for all the classes, if it is constructed in the conventional way, is given by

$$\begin{aligned} \sum_x &= 1/L \sum_{k=1}^L \sum_{i=1}^n X_i^k (X_i^k)^T \\ &\quad - 1/L \sum_{k=1}^L \left(1/n \sum_{i=1}^n X_i^k \right) \left(1/n \sum_{i=1}^n (X_i^k)^T \right). \end{aligned} \quad (1)$$

Assuming all the signals for all the classes as belonging to one class, and taking the global mean, the covariance matrix is

$$\sum'_x = 1/Ln \sum_{i=1}^{Ln} X_i X_i^T - \left(1/Ln \sum_{i=1}^{Ln} X_i \right) \left(1/Ln \sum_{i=1}^{Ln} X_i^T \right), \quad (2)$$

where i runs from 1 to Ln for all the signals of all the classes from 1 to L . The superscript for the identification of class is omitted for convenience.

The first term of $\sum'_x$ can be written as

$$\begin{aligned} & 1/L \left[1/n \sum_{i=1}^n X_i^1 (X_i^1)^T + 1/n \sum_{i=1}^n X_i^2 (X_i^2)^T + \cdots 1/n \sum_{i=1}^n X_i^L (X_i^L)^T \right] \\ &= 1/L \sum_{k=1}^L 1/n \sum_{i=1}^n X_i^k (X_i^k)^T. \end{aligned}$$

The second term of $\sum'_x$ can be written as

$$= -1/L^2 \left[\sum_{k=1}^L 1/n \sum_{i=1}^n X_i^k \sum_{l=1}^L 1/n \sum_{i=1}^n (X_i^l)^T \right] = -1/L^2 \sum_{k=1}^L \sum_{l=1}^L \mathbf{M}_k (\mathbf{M}_l)^T,$$

where $\mathbf{M}_k = (1/n) \sum_{i=1}^n X_i^k$ is the mean vector for the k th class, then

$$\sum'_x - \sum_x = 1/L \sum_{k=1}^L \mathbf{M}_k (\mathbf{M}_k)^T - (1/L)^2 \sum_{k=1}^L \sum_{l=1}^L \mathbf{M}_k (\mathbf{M}_l)^T = \mathbf{C}. \quad (3)$$

The mean vector $\mathbf{M}_k$ is a column vector of m components as $[M_{k1}, M_{k2}, M_{k3}, \dots, M_{km}]^T$. The i th diagonal element of the matrix $\mathbf{C}$ can be written as $1/L \sum_{k=1}^L M_{ki}^2 - (1/L \sum_{k=1}^L M_{ki})^2$. Now $1/L \sum_{k=1}^L M_{ki} = \mu_i$ is the mean of all the i th components of the mean vectors.

$$\begin{aligned} 1/L \sum_{k=1}^L M_{ki}^2 - \left(1/L \sum_{k=1}^L M_{ki} \right)^2 &= 1/L \sum_{k=1}^L M_{ki}^2 - 2\mu_i 1/L \sum_{k=1}^L M_{ki} + \mu_i^2 \\ &= 1/L \sum_{k=1}^L [M_{ki}^2 - 2\mu_i M_{ki} + \mu_i^2] \\ &= 1/L \sum_{k=1}^L [M_{ki} - \mu_i]^2. \end{aligned} \quad (4)$$

Thus the diagonal elements of the matrix $\mathbf{C}$ represent the variance of the respective components of the mean vector and each element is a positive quantity. So its trace is also a positive quantity. As the trace of $\mathbf{C}$ is positive the trace of $\sum'_x$ will be greater than that of $\sum_x$. The commutator $[\sum'_x, \sum_x] \neq 0$ so they cannot be diagonalised simultaneously.

Let

$$\begin{aligned} \sum'_x \phi'_i &= \lambda'_i \phi'_i, \\ \sum_x \phi_j &= \lambda_j \phi_j, \\ \mathbf{C} \phi_k^c &= \lambda_k^c \phi_k^c. \end{aligned} \quad (5)$$

It can be shown that $D = \sum_{i=m_1+1}^m (\lambda'_i - \lambda_i) = \sum_{i=m_1+1}^m [\phi_i C (\phi'_i)^T / \phi_i (\phi'_i)^T]$ assuming $\phi_i (\phi'_i)^T \neq 0$ and assuming $\sum'_x$, $\sum_x$ and $\mathbf{C}$ are real symmetric matrices, so that their eigenvalues are all real.

Expanding ϕ'_i and ϕ_i in terms of the eigenfunction of $\mathbf{C}$ as

$$\begin{aligned}\phi'_i &= \sum_{k=1}^m \xi'_{ik} \phi_k^c \quad \text{and} \quad \phi_i = \sum_{l=1}^m \xi_{il} \phi_l^c, \\ D &= \sum_{i=m_1+1}^m \left[\left(\sum_{k=1}^m \xi_{ik} \xi'_{ik} \lambda_k^c \right) / \left(\sum_{k=1}^m \xi_{ik} \xi'_{ik} \right) \right], \quad [\lambda_k^c \geq 0].\end{aligned}\quad (6)$$

If λ_k^c (Min), λ_k^c (Max) are the minimum and maximum value of λ_k^c then

$$(m - m_1) \lambda_k^c (\text{Min}) \leq D \leq (m - m_1) \lambda_k^c (\text{Max}).$$

So the difference in the absolute error D depends entirely on the eigenvalues of the matrix $\mathbf{C}$ and is a positive quantity. Thus the absolute error of reconstruction in the global mean approach is larger than that in the conventional approach. However, as the trace of $\sum'_x$ is greater than that of $\sum_x$, the magnitude of the relative errors (obtained by normalizing the respective absolute errors by the respective traces) can be larger in the conventional approach. As $\sum'_x$ and $\sum_x$ cannot be diagonalised simultaneously, it is not possible to judge this effect analytically. These effects have been judged through the analysis of synthetic and experimentally observed acoustic signals which will be discussed in § 4.

In this context it is necessary to introduce another concept which is a measure for representing the information content of the individual eigen vectors in the transformed domain.

As the eigenvalues of the covariance matrix represent the variances, it is evident that the greater the value of the variance, the more is the information about X contained in the corresponding component of the transformed domain vector Y . In formal terms, a greater information compression will be achieved if the covariance matrix has a very few eigenvalues of large magnitude and all the other eigenvalues are small, than if the eigenvalues are distributed uniformly over the axes defined by the eigenfunctions of the covariance matrix.

In order to assess the degree of information compressibility, it is essential to define an information compression indicator, which will in turn measure the dispersion of information over the axes of the transformed coordinate system.

Consider a function of the form

$$H = \sum_{q=1}^m \left[\left(\lambda_q / \sum_{k=1}^m \lambda_k \right) \times \log \left(\lambda_q / \sum_{k=1}^m \lambda_k \right) \right].$$

This function will attain a maximum value when all λ_q 's are identical. In this particular case the approximation X by m_1 terms in the reduced domain will be associated with the largest error, or, in different words, with the greatest uncertainty about the true vector X . On the other hand, zero entropy implies that all the information about X is contained in one term of the reduced domain. Evidently, the magnitude of the entropy function reflects the uncertainty about X if instead of X we work its representative $\tilde{X}$. Thus the criterion H can therefore be considered a measure of representative uncertainty, and is referred to as representation entropy.

4. Analysis

For the analysis, synthetic signals and experimentally observed acoustic signals were used.

The synthetic signals of two classes were the Sine and Square wave forms of fixed frequency and phase. The amplitude of these signals of the two classes were selected randomly between 0.8 and 1.0. Subsequently, at each time step, the computed magnitude of the signals was further modulated randomly within a limit of $\pm 5\%$ of its computed magnitude. 128 discrete data points were taken to represent one such signal. These signals of Sine and Square waves for two different classes are designated as class-1 and class-2 respectively. One thousand signals for each class were used for the analysis.

Acoustic signals were obtained by exciting a broad-band acoustic probe, of central frequency of 2.5 MHz and of crystal diameter 12.6 mm, by a pulser receiver unit. This excited probe was then used to generate acoustic pulses inside aluminium (Al) and copper (Cu) plates of 10 mm thickness each. The same probe was used to receive scattered acoustic signals from the incident acoustic pulse from these materials. The scattered signals were captured by a digital oscilloscope at a sampling rate of 40 ns after suitable amplification. 200 such signals each from the Cu and Al plates were obtained by holding the probe at different spatial locations of the plates and were used for the analysis. The fluctuation of these signals were due to microstructural variation in the samples. 128 data points were taken for each signal.

Figure 1 shows the normalized variance distribution of these signals when covariance matrices $\sum_x$ and $\sum'_x$, were diagonalised for synthetic and acoustic signals. In these figures only those variances which are of appreciable value are shown. It is interesting to see that

Figure 1. Normalized variance distribution (a) for $\sum_x$ and (b) for $\sum'_x$.

Table 1. The eigenvalues and their respective sums for $\sum_x$ and $\sum'_x$.

λ_1	λ_2	$\sum_{j=3}^{128} \lambda_j$	$\sum_{j=3}^{128} \bar{\lambda}_j$	λ'_1	λ'_2	$\sum_{j=3}^{128} \lambda'_j$	$\sum_{j=3}^{128} \bar{\lambda}'_j$	$\sum_{j=3}^{128} (\lambda'_j - \lambda_j)$	$\sum_{j=3}^{128} (\bar{\lambda}'_j - \bar{\lambda}_j)$
For synthetic signals									
0.215	0.110	0.1048	0.2440	39.25	0.144	0.1049	0.0026	0.0001	−0.2414
For acoustic signals									
2.754	1.467	1.943	0.3152	20.90	2.585	3.292	0.1229	1.349	−0.1923

the normalized variance distribution converges faster when the global mean approach is used to construct the covariance matrices.

The eigenvalues and their respective sums of $\sum_x$ and $\sum'_x$ for synthetic and acoustic signals are given in table 1. From table 1 it is seen that $\sum_{j=3}^{128} (\lambda'_j - \lambda_j)$ is positive for both the cases. Hence, the absolute error for the reconstruction of signals using the first two eigenfunctions, is greater in the global mean approach than that in the conventional approach. However, when the individual eigenvalues of the respective matrices were normalized to its respective traces, it was found that $\sum_{j=3}^{128} (\bar{\lambda}'_j - \bar{\lambda}_j)$ is negative. Table 1 also shows the percentage mean square error introduced in the reconstruction of original signals taking only two eigenfunctions corresponding to the largest variances is 0.26 for synthetic

Figure 2. Projection of class-1 and class-2 signals in dominant subspace in (a) conventional approach, and (b) global mean approach.

Figure 3. Projection of the acoustic signals obtained from Al and Cu plates in dominant subspace in (a) conventional approach and (b) global mean approach.

signals and 12.29 for acoustic signals in case of the global mean approach whereas the corresponding values are 24.40 and 31.52 in case of the conventional approach and which are appreciably more. The largest eigenvalues of $\sum'_x$ are much larger in magnitude compared to the magnitudes of the eigenvalues of $\sum_x$. This is because of the fact that the incorporation of an entirely new class of signals in another class in case of the global mean approach, causes higher magnitude variances of the entire ensemble of the signals. The representation entropy H in case of the global mean approach is 5.475×10^{-2} for synthetic signals and 0.933 for acoustic signals which are less than the corresponding values of 2.161 and 1.844 obtained by the conventional approach.

In order to understand the performance of the proposed global mean approach, 200 signals of each class in case of synthetic and acoustic signals were taken. These 200 signals were transformed to the eigenspace spanned by the eigen vectors, obtained by diagonalising $\sum_x$ and $\sum'_x$. The projection of the transformed domain signals onto the

Figure 4. Percentage error in reconstruction of individual signals of class-1 and class-2 in conventional approach and global mean approach.

subspace spanned by the eigen functions corresponding to the first two eigenvalues of the highest magnitude are shown in figures 2 and 3 for synthetic and acoustic signals respectively. These figures emphasize that the intrinsic discriminating features between the classes which were initially present when conventional KLT were used, did not degrade when the global mean approach was adopted. In fact it is easily evident from figures 2 and 3 that signals of the two classes are better separated when global mean approach is used. This is because the representation entropy H is much less in the global mean approach than in the conventional approach. Taking again only first two prominent components of the reduced domain vector $\tilde{Y}_i$, $\tilde{X}_i$ was reconstructed for 200 signals for each class using conventional and global mean approach for synthetic signals only. The percentage error for each signal X_i was computed as $[\sum_{k=1}^m \{(X_i)_k - (\tilde{X}_i)_k\}^2]^{1/2} / [X_i(X_i)^T]^{1/2} \times 100$ (the suffix k represents the k th component of the vector X_i). Figure 4 gives the percentage error

for these signals of different classes. It is interesting to see that the error did not enhance in any case when the global mean approach was used for the construction of the covariance matrix.

5. Conclusion

The proposed global mean approach removes the inherent difficulty of applying KLT in recognizing mean patterns in multi-class environment. The discriminating behavior in the transformed domain is not degraded from the original method. Thus KLT with global mean approach offers a practical solution for dimensionality reduction in pattern recognition problems.

References

- Ahamed N, Rao K R 1975 *Orthogonal transforms for digital signal processing* (Berlin: Springer-Verlag)
- Bhama S, Singh H, Phadte N D 1993 Parallelism for the faster implementation of the K-L transform for image compression. *Pattern Recog. Lett.* 14: 651–659
- Chien Y T, Fu K S 1967 On the generalized Karhunen–Loeve expansion. *IEEE Trans. Inf. Theory* IT-13: 518–520
- Chien Y T, Fu K S 1968 Selection and ordering of feature observations in a pattern recognition system. *Inf. Control* 12: 395–414
- Fu K S 1968 *Sequential methods in pattern recognition and machine learning* (New York: Academic Press)
- Fukunaga K 1972 *Introduction to statistical pattern recognition* (New York: Academic Press)
- Jain A K 1976 A fast Karhunen–Loeve transform for a class of random processes. *IEEE Trans. Commun.* COM-24: 1023–1029
- Laaksonen J, Oja E 1996 Subspace dimension selection and averaged learning subspace method in handwritten digit classification. *Proc. of the Int. Conf. on Artificial Neural Networks*, Bochum, Germany, pp 227–232
- Laaksonen J 1997 Local subspace classifier and local subspace SOM. *Proc. of the Workshop on Self-Organizing Maps*, Otaniemi, Finland, pp 32–37
- Oja E 1983 *Subspace methods of pattern recognition* (New York: Research Studies Press)
- Oja E 1989 Neural networks, principal components, and subspaces. *Int. J. Neural Syst.* 1: 61–68
- Oja E 1992 Principal components, minor components, and linear neural networks. *Neural Networks* 5: 927–935
- Oja E, Kärki T 1987 Properties of the ALSM subspace classifier for texture analysis. *Proceedings of the 5th Scandinavian Conf. on Image Analysis*, Stockholm, pp 419–426
- Oja E, Kohonen T 1988 The subspace learning algorithm as a formalism for pattern recognition and neural networks. *IEEE Int. Conf. on Neural Networks*, San Diego, California, pp I277–I284
- Oja E, Kuusela M 1983 The ALSM algorithm an improved subspace method of classification. *Pattern Recog.* 16: 421–427
- Rosenfeld A 1976 *Digital picture processing* (New York: Academic Press)
- Tomita S, Noguchi S, Oizumi J 1970 On evaluation of hand-written characters by Karhunen–Loeve orthonormal system. *Proc. 3rd. Hawaii Int. Conf. on Syst. Sci.*, Part I, pp 501–504
- Tou J T, Gonzalez R C 1974 *Pattern recognition principles* (London: Addison-Wesley)
- Watanabe S, Pakvasa N 1973 Subspace method of pattern recognition. *Proceedings of the first Int. Conf. on Pattern Recognition*, Washington, DC, pp 25–32