If only the \( L \) values \( R_0 \) through \( R_{L-1} \) are required, then it can be seen by comparing (15) and (11) that the new algorithm will require fewer multiplications than the FFT method if

\[
L < 11.2 (1 + \log_2 N) \tag{16a}
\]

and will require fewer additions if

\[
L < 5.6 (1 + \log_2 N). \tag{16b}
\]

Therefore, we conclude that the new algorithm will generally be more efficient than the FFT method if

\[
N < 128 \tag{17a}
\]

or

\[
L < 10 (1 + \log_2 N). \tag{17b}
\]

**Conclusion**

A new algorithm for computing the correlation of a block of sampled data has been presented. It is a direct method which trades an increased number of additions for a decreased number of multiplications. For applications where the "cost" (e.g., the time) of a multiplication is greater than that of an addition, the new algorithm is always more computationally efficient than direct evaluation of the correlation, and it is generally more efficient than FFT methods for processing 128 or fewer data points, or for calculating only the first \( L \) "lags" for \( L < 10 \log_2 2N \).

**REFERENCES**


**Discrete Cosine Transform**

N. AHMED, T. NATARAJAN, AND K. R. RAO

Abstract—A discrete cosine transform (DCT) is defined and an algorithm to compute it using the fast Fourier transform is developed. It is shown that the discrete cosine transform can be used in the area of digital processing for the purposes of pattern recognition and Wiener filtering. Its performance is compared with that of a class of orthogonal transforms and is found to compare closely to that of the Karhunen-Loeve transform, which is known to be optimal. The performances of the Karhunen-Loeve and discrete cosine transforms are also found to compare closely with respect to the rate-distortion criterion.

Index Terms—Discrete cosine transform, discrete Fourier transform, feature selection, Haar transform, Karhunen-Loeve transform, rate distortion, Walsh-Hadamard transform, Wiener vector and scalar filtering.

**Introduction**

In recent years there has been an increasing interest with respect to using a class of orthogonal transforms in the general area of digital signal processing. This correspondence addresses itself towards two problems associated with image processing, namely, pattern recognition [1] and Wiener filtering [2].

In pattern recognition, orthogonal transforms enable a noninvertible transformation from the pattern space to a reduced dimensionality feature space. This allows a classification scheme to be implemented with substantially less features, with only a small increase in classification error.

Manuscript received January 29, 1973; revised June 7, 1973.

N. Ahmed is with the Department of Electrical Engineering and Computer Science, Kansas State University, Manhattan, Kansas.

T. Natarajan is with the Department of Electrical Engineering, Kansas State University, Manhattan, Kansas.

K. R. Rao is with the Department of Electrical Engineering, University of Texas at Arlington, Arlington, Texas 76010.

In discrete Wiener filtering applications, the filter is represented by an \((M \times M)\) matrix \( G \). The estimate \( \hat{x} \) of data vector \( x \) is given by \( Gz \), where \( z = x + n \) and \( n \) is the noise vector. This implies that approximately \( 2M^2 \) arithmetic operations are required to compute \( \hat{x} \). Use of orthogonal transforms yields a \( G \) in which a substantial number of elements are relatively small in magnitude, and hence can be set equal to zero. Thus a significant reduction in computation load is realized at the expense of a small increase in the mean-square estimation error.

The Walsh-Hadamard transform (WHT), discrete Fourier transform (DFT), the Haar transform (HT), and the slant transform (ST), have been considered for various applications [1], [2], [4]–[9] since these are orthogonal transforms that can be computed using fast algorithms. The performance of these transforms is generally compared with that of the Karhunen-Loeve transform (KLT) which is known to be optimal with respect to the following performance measures: variance distribution [1], estimation using the mean-square error criterion [2], [4], and the rate-distortion function [5]. Although the KLT is optimal, there is no general algorithm that enables its fast computation [1].

In this correspondence, a discrete cosine transform (DCT) is introduced along with an algorithm that enables its fast computation. It is shown that the performance of the DCT compares more closely to that of the KLT relative to the performances of the DFT, WHT, and HT.

**Discrete Cosine Transform**

The DCT of a data sequence \( x(m) \), \( m = 0, 1, \ldots, (M - 1) \) is defined as

\[
G_x(k) = \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} x(m) \cos \frac{(2m + 1) k \pi}{2M}, \quad k = 1, 2, \ldots, (M - 1) \tag{1}
\]

where \( G_x(k) \) is the \( k \)th DCT coefficient. It is worthwhile noting that the set of basis vectors \( \{1/\sqrt{2}, \cos ((2m + 1) k \pi)/(2M)\} \) is actually a class of discrete Chebyshev polynomials. This can be seen by recalling that Chebyshev polynomials can be defined as [3]

\[
\hat{T}_0(\xi_p) = \frac{1}{\sqrt{2}}
\]

\[
\hat{T}_k(\xi_p) = \cos (k \cos^{-1} \xi_p), \quad k, p = 1, 2, \ldots, M \tag{2}
\]

where \( \hat{T}_k(\xi_p) \) is the \( k \)th Chebyshev polynomial.

Now, in (2), \( \xi_p \) is chosen to be the \( p \)th zero of \( \hat{T}_M(\xi) \), which is given by [3]

\[
\xi_p = \frac{(2p - 1) \pi}{2M}, \quad p = 1, 2, \ldots, M. \tag{3}
\]

Substituting (3) in (2), one obtains the set of Chebyshev polynomials

\[
\hat{T}_0(p) = \frac{1}{\sqrt{2}}
\]

\[
\hat{T}_k(p) = \cos \frac{(2p - 1) k \pi}{2M}, \quad k, p = 1, 2, \ldots, M. \tag{4}
\]

From (4) it follows that the \( \hat{T}_k(p) \) can equivalently be defined as

\[
T_0(m) = \frac{1}{\sqrt{2}}
\]

\[
T_k(m) = \cos \frac{(2m + 1) k \pi}{2M}, \quad k = 1, 2, \ldots, (M - 1), \quad m = 0, 1, \ldots, M - 1. \tag{5}
\]

Comparing (5) with (1) we conclude that the basis member \( \cos ((2m +
\[ X(m) = \frac{1}{\sqrt{2}} X(0) + \sum_{k=1}^{M-1} \frac{G_x(k) \cos \left( \frac{(2m+1)\pi k}{2M} \right)}{2M}, \quad m = 0, 1, \ldots, (M-1). \] (6)

We note that applying the orthogonal property \[ \Lambda^T \Lambda = \frac{M}{2} \mathbb{I} \] (8)

where \( \Lambda^T \) is the transpose of \( \Lambda \) and \( \mathbb{I} \) is the \((M \times M)\) identity matrix.

**Motivation**

The motivation for defining the DCT is that it can be demonstrated that its basis set provides a good approximation to the eigenvectors of the class of Toeplitz matrices defined as

\[
\psi = \begin{bmatrix}
1 & \rho & \rho^2 & \ldots & \rho^{M-1} \\
\rho & 1 & \rho & \ldots & \rho^{M-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho^{M-1} & \rho^{M-2} & \ldots & \ldots & 1 \\
\end{bmatrix}, \quad 0 < \rho < 1.
\] (9)

For the purposes of illustration, the eigenvectors of \( \psi \) for \( M = 8 \) and \( \rho = 0.9 \) are plotted (see Fig. 1) against

\[
\left\{ \frac{\cos \left( \frac{(2m+1)\pi k}{2M} \right)}{\sqrt{2}}, k = 1, 2, \ldots, 7, m = 0, 1, \ldots, 7 \right\}
\] (10)

which constitute the basis set for the DCT. The close resemblance (aside from the 180° phase shift) between the eigenvectors and the set defined in (10) is apparent.

**Algorithm**

It can be shown that (1) can be expressed as

\[
G_x(k) = \frac{2}{M} \text{Re} \left\{ e^{-i(2\pi k)/(2M)} \sum_{m=0}^{M-1} X(m) i^{km} \right\},
\]

\[
k = 1, 2, \ldots, (M-1)
\] (11)

where

\[
W = e^{-i2\pi/2M}, \quad i = \sqrt{-1},
\]

\[
X(m) = 0, \quad m = M, (M+1), \ldots, (2M-1)
\]

and \( \text{Re} \{ \cdot \} \) implies the real part of the term enclosed. From (11) it follows that all the \( M \) DCT coefficients can be computed using a 2M-point fast Fourier transform (FFT). Since (1) and (6) are of the same form, FFT can also be used to compute the IDCT. Similarly, if a discrete sine transform were defined, then the \( \text{Re} \{ \cdot \} \) in (11) would be replaced by \( \text{Im} \{ \cdot \} \), which denotes the imaginary part of the term enclosed.

**Computational Results**

In image processing applications, \( \psi \) in (9) provides a useful model for the data covariance matrix corresponding to the rows and columns of an image matrix \([6, \; 7]\). The covariance matrix in the transform domain is denoted by \( \Psi \) and is given by

\[
\Psi = \Lambda \psi \Lambda^*\]

(12)

where \( \Lambda \) is the matrix representation of an orthogonal transformation and \( \Lambda^* \) is the complex conjugate of \( \Lambda \). From (12) it follows that \( \Psi \) can be computed as a two-dimensional transform of \( \psi \).

**Feature Selection**

A criterion for eliminating features (i.e., components of a transform vector), which are least useful for classification purposes, was developed by Andrews [1]. It states that features whose variances (i.e., main diagonal elements of \( \Psi \)) are relatively large should be retained. (Fig. 2 should be retained.) Fig. 2 shows the various variances ranked in decreasing order of magnitude. From the information in Fig. 2 it is apparent that relative to the set of orthogonal transforms shown, the DCT compares most closely to the KLT.

**Wiener Filtering**

The role of orthogonal transforms played in filtering applications is illustrated in Fig. 3 [2]. Z is an \((M \times 1)\) vector which is the sum of a
vector $X$ and a noise vector $N$. $X$ is considered to belong to a random process whose covariance matrix is given by $\Psi$ which is defined in (9). The Wiener filter $G$ is in the form of an $(M \times M)$ matrix. $A$ and $A^{-1}$ represent an orthonormal transform and its inverse, respectively, while $\hat{X}$ denotes the estimate of $X$, using the mean-square error criterion.

We restrict our attention to the case when $G$ is constrained to be a diagonal matrix $Q$. This class of Wiener filters is referred to as scalar filters while the more general class (denoted by $G$) is referred to as vector filters. The additive noise (see Fig. 3) is considered to be white, zero mean, and uncorrelated with the data. If the mean-square estimation error due to scalar filtering is denoted by $e_Q$, then $e_Q$ can be expressed as [4]

$$e_Q = 1 - \frac{1}{M} \sum_{m=1}^{M} \frac{\Psi_{X}(s,s)}{\Psi_{X}(s,s) + \Psi_{N}(s,s)}$$

(13)

where $\Psi_{X}$ and $\Psi_{N}$ denote the transform domain covariance matrices of the data and noise, respectively. Table I lists the values of $e_Q$ for different values of $M$ for the case $p = 0.9$ and a signal-to-noise ratio of unity. From Table I it is evident that the DCT comes closest to the KLT which is optimal. This information is presented in terms of a set of performance curves in Fig. 4.

**ADDITIONAL CONSIDERATIONS**

In conclusion, we compare the performance of the DCT with KLT, DFT, and the identity transforms, using the rate-distortion criterion [5]. This performance criterion provides a measure of the information rate $R$ that can be achieved while still maintaining a fixed distortion $D$, for encoding purposes. Considering Gaussian sources along with the mean-square error criterion, the rate-distortion performance measure is given by [5]

$$R(A, D) = \frac{1}{2M} \sum_{j=1}^{M} \max \left\{ 0, \ln \left( \frac{a_j}{D} \right) \right\}$$

(14a)
Minimization of Linear Sequential Machines

CHI-TSONG CHEN

Abstract—An algorithm is presented to minimize linear sequential machines to reduced form.

Index Terms—Algorithm, linear machine, minimization, reduced machine, reduction, sequential machine.

I. INTRODUCTION

The minimization of a linear sequential machine to a reduced form is an important topic and is discussed in many texts [1]–[4]. The minimization procedure presented in [1]–[4] is as follows. Let \( \{A, B, C, D\} \) be an \( n \)-dimensional linear machine over \( GF(p) \), and let \( r \), with \( r < n \), be the rank of the diagnostic matrix \( K = [C'A'C' \cdots (A')^{n-r}C']' \). where the prime stands for the transpose. Define an \( r \times n \) matrix \( T \) consisting of the first \( r \) linearly independent rows of \( K \), and an \( n \times r \) matrix \( R \) denoting the right inverse of \( T \) so that \( TR = I \). Then the linear machine \( \{TAR, TR, CR, D\} \) is a reduced form of \( \{A, B, C, D\} \). In this correspondence, an algorithm will be introduced to find a special set of \( r \) linearly independent rows in \( K \). A reduced machine can then be read out from this algorithm without the need of inverting any matrix. Furthermore, the reduced machine will be in a canonical form.

II. ALGORITHM

Let \( C \) be a \( q \times n \) matrix, and let \( c_i \) be the \( i \)th row of \( C \). The diagnostic matrix of \( \{A, B, C, D\} \) will be arranged in the following order:

\[
P = \begin{bmatrix}
c_1 \\
c_1A \\
\vdots \\
c_1A^{r-1} \\
c_q \\
c_qA \\
\vdots \\
c_qA^{r-1}
\end{bmatrix}
\]  

(1)

Now an algorithm will be introduced to find the first \( r \) linearly independent rows in \( P \). This is achieved by a series of elementary transformations. Let

\[
K_r K_{r-1} \cdots K_2 K_1 P \equiv KP
\]  

(2)

where \( K_i \) are lower triangular matrices with all diagonal elements unity, and are obtained in the following manner. Let \( p_{ij}(j) \) be the first nonzero element from the left in the first row of \( P \), where \( j \) denotes the position. Then \( K_i \) is chosen so that all, except the first two, elements of the \( j \)th column of \( K_i P \) are zero. Let \( p_{ij}(l) \) be the first nonzero element from the left of the second row of \( K_i P \). Then \( K_2 \) is chosen so that all, except the first two, elements of the \( l \)th column of \( K_2 K_1 P \) are zero. Proceed in this manner until all linearly independent rows of \( P \) are found. Note that in this process, if one row is identically zero, then proceed to the next nonzero row. By multiplying these \( K_i \), we obtain \( K = K_r K_{r-1} \cdots K_2 K_1 \). Note that \( KP \) will be final of form

Manuscript received February 21, 1973; revised May 21, 1973.

The author is with the Department of Electrical Sciences, State University of New York, Stony Brook, N.Y. 11790.

REFERENCES
