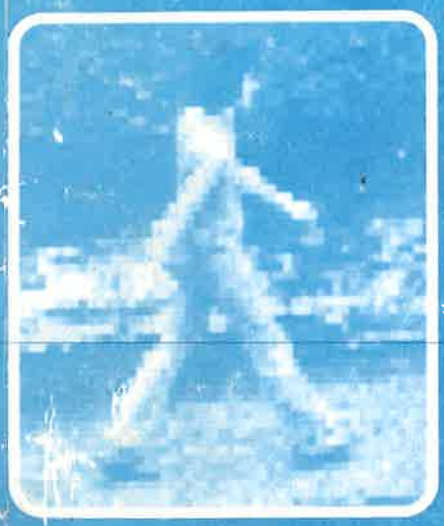
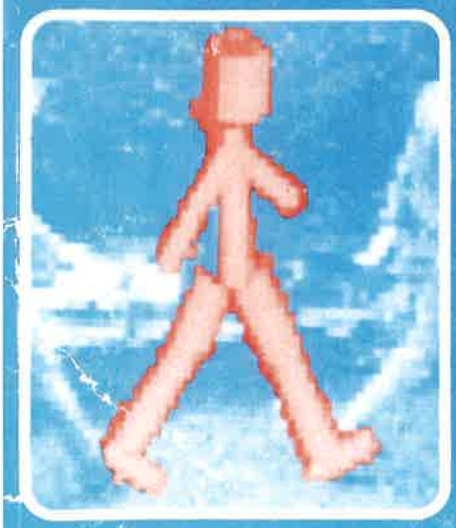
Image and Vision Computing

Volume 5 Number 4 November 1987



Linear predictive transform of monochrome images

Erlan H Feria

Digital monochromatic images are encoded using a novel minimum mean square error (MSE) linear predictive transform (LPT) coding formulation. The new formulation is appealing for two important reasons. First, it leads to simple coder implementation with a satisfactory signal-to-noise ratio (SNR). Second, it provides a general theoretical framework from which minimum MSE predictive coding and minimum MSE transform coding arise as special cases. Some specific results of this paper that illustrate the previous ideas are: a simple and generally suboptimum two-dimensional LPT coder operating at 2 bit pixel^{-1} has approximately one third the complexity of a 4 × 4 Hadamard coder while yielding a better SNR; an optimum 2D LPT coder operating at 2 bit pixel^{-1} has approximately one sixth the complexity of a 4 × 4 Karhunen–Loeve transform (KLT) coder while yielding a better SNR.

Keywords: image coding, predictive transfer coding, video processing, monochromatic images

The last two decades have seen great activity in the field of digital video processing where a major goal has been the design of easily implementable coders with satisfactory picture quality at reasonable bit rates. To address this problem three major classes of coding techniques have been offered. These are predictive coding[1], transform coding[2] and predictive transform coding[3]. Predictive coding has been found to be readily implementable[4], but its performance is often unsatisfactory at low bit rates. On the other hand, transform coding[5] yields better picture quality at low bit rates but is more difficult to implement due to the need to transform relatively large picture blocks. For example, a 4 × 4 or higher sized picture block must be transformed when using a two-dimensional (2D) transform. To obtain a compromise between the best qualities of these two classes of coders, several types of hybrid predictive transform coders have been offered such as those by Habibi[6], Haralick and Shanmugan[3,7] and Watson *et al.*[8]. Yet these schemes are still too complex when compared with simple 2D linear predictors[4] or a separable 4 × 4 Hadamard transform coder.

This paper presents a minimum mean square error (MSE) linear predictive transform (LPT) coding formulation which in its simplest version yields a good trade-off between the best qualities of predictive and transform coding. The minimum MSE LPT coding formulation also provides a general framework which contains classical minimum MSE predictive coding and classical minimum MSE transform coding as special cases. The proposed LPT formulation of this paper is applied to the coding of digital monochrome images. For these images, intrafield LPT coders that use 2D transforms to operate on 2D picture blocks have been designed. These coders have two general properties.

- They predict the transform coefficients by making linear transformations of the previously estimated pixels surrounding the block to be transformed.
- The transform and predictor matrices of the LPT coders are designed such that the coefficient errors between the transform coefficients and their estimates have the following two properties: they are uncorrelated or almost uncorrelated, and their mean value is zero. These properties allow simple unbiased scalar quantizers to be used to encode the coefficient errors.

In particular, it is found that a simple and generally suboptimum LPT coder operating at 2 bit pixel^{-1} has approximately one third the complexity of a 4 × 4 Hadamard coder while yielding a better signal-to-noise ratio (SNR). Another result is that an optimum LPT coder operating at 2 bit pixel^{-1} has approximately one sixth the complexity of a 4 × 4 KLT coder while yielding a better SNR performance.

For a comprehensive survey of the application of predictive coding, transform coding and predictive trans-

Department of Applied Sciences of the College of Staten Island, City University of New York, 130 Stuyvesant Place, Staten Island, NY 10301, USA

0262–8856/87/05267–12 $03.00

form coding to the bandwidth compression of video signals see Jayant and Noll[9] and Clarke[10]. Several standards papers in predictive coding, transform coding and predictive transform coding are referenced throughout this paper. Of special interest is a paper by Netravali *et al.*[11] in which a predictive transform scheme with 2 × 2 Hadamard coefficients is developed; however, this work used intuitive ideas to design the coder, excluding quantizer, rather than the analytical minimum MSE approach used in this paper.

The rest of this paper is organized as follows. The general structure of an LPT coder is illustrated using a 2D transform. The minimum MSE LPT coding formulation is presented and the optimum LPT coder derived. It is also shown that classical minimum MSE predictive coding and classical minimum MSE transform coding are special cases of the proposed LPT formulation. LPT coders that are easier to implement are derived using the optimum LPT coder as a basis. Then, LPT coders are illustrated via various examples. The paper ends by summarizing the most important results.

LPT CODER STRUCTURE

In this section the general structure of an LPT coder is illustrated using a 2D transform. It should be kept in mind, however, that this general structure remains essentially intact when using either simpler 1D transforms or more complex 3D transforms.

Figure 1 illustrates with 2 × 2 picture blocks the notation that will be used to represent consecutive $N \times N$ picture blocks of a picture field. From this figure note that $x_j(i)$ denotes the jth picture element in the ith picture block, moving from left to right and top to bottom in the block; the ith picture block is fully defined by the W-dimensional column vector

$$\begin{bmatrix} x_1(k-L) & x_2(k-L) \\ x_3(k-L) & x_4(k-L) \end{bmatrix} \quad \begin{bmatrix} x_1(k+1-L) & x_2(k+1-L) \\ x_3(k+1-L) & x_4(k+1-L) \end{bmatrix}$$

$$\begin{bmatrix} x_1(k+2-L) & x_2(k+2-L) \\ x_3(k+2-L) & x_4(k+2-L) \end{bmatrix}$$

$$\begin{bmatrix} x_1(k) & x_2(k) \\ x_3(k) & x_4(k) \end{bmatrix} \quad \begin{bmatrix} x_1(k+1) & x_2(k+1) \\ x_3(k+1) & x_4(k+1) \end{bmatrix}$$

$$\mathbf{x}(i) = [x_1(i), \ldots, x_4(i)]^T \qquad \text{for all } i$$

Figure 1. Notation used to represent consecutive 2 × 2 picture blocks of a picture field. The integer L represents the number of 2 × 2 picture blocks associated with each set of two consecutive scanning lines. $\mathbf{x}(k - L)$, $\mathbf{x}(k + 1 - L)$, $\mathbf{x}(k + 2 - L)$ *and* $\mathbf{x}(k)$ *are all the immediately adjacent past picture blocks to the current picture block* $\mathbf{x}(k + 1)$

$$\mathbf{x}(i) = [x_1(i), \ldots, x_W(i)]^T \qquad \text{for all } i$$

where $W = N^2$; $\mathbf{x}(k - L)$, $\mathbf{x}(k + 1 - L)$, $\mathbf{x}(k + 2 - L)$ and $\mathbf{x}(k)$ are all the immediately adjacent past picture blocks to the current picture block $\mathbf{x}(k + 1)$. The integer L denotes the number of $N \times N$ picture blocks that are associated with each set of N consecutive scanning lines.

Figure 2 shows the structure of the proposed LPT coder where the defining expressions for the transform, estimator, predictor and quantizer subsystems are as follows.

Transform

$$\mathbf{c}(k + 1) = \mathbf{R}^{-1}\mathbf{x}(k + 1) \tag{1}$$

or

$$\mathbf{x}(k + 1) = \mathbf{R}\mathbf{c}(k + 1) = \sum_{i=1}^{W} \mathbf{r}_i c_i(k + 1) \tag{2}$$

where $\mathbf{R}$ is a $W \times W$ transform matrix with W W-dimensional column basis vectors $\{\mathbf{r}_i\}$, i.e.

$$\mathbf{R} = [\mathbf{r}_1, \ldots, \mathbf{r}_W]$$

and $\mathbf{c}(k + 1)$ is a W-dimensional coefficient column vector, i.e.

$$\mathbf{c}(k + 1) = [c_1(k + 1), \ldots, c_W(k + 1)]^T \tag{3}$$

Estimator

$$\hat{\mathbf{x}}(k + 1) = \mathbf{R}\hat{\mathbf{c}}(k + 1) \tag{4}$$

$$\hat{\mathbf{c}}(k + 1) = \hat{\mathbf{c}}'(k) + \hat{\delta}\mathbf{c}(k) \tag{5}$$

where $\hat{\mathbf{x}}(k + 1)$ is the estimate of the picture block $\mathbf{x}(k + 1)$; $\hat{\mathbf{c}}(k + 1)$ is the estimate of the coefficient vector $\mathbf{c}(k + 1)$; $\hat{\mathbf{c}}'(k)$ is a 'preliminary' estimate of the coefficient vector $\mathbf{c}(k + 1)$; and $\delta\hat{\mathbf{c}}(k)$ is the quantized coefficient error.

Predictor

$$\hat{\mathbf{c}}'(k) = \mathbf{P}^T\mathbf{z}(k) = [\mathbf{p}_1, \ldots, \mathbf{p}_W]^T\mathbf{z}(k) \tag{6}$$

where $\mathbf{P}$ is an $M \times W$ predictor matrix with W M-dimensional column vectors $\{\mathbf{p}_i\}$, i.e.

$$\mathbf{P} = [\mathbf{p}_i, \ldots, \mathbf{p}_W]$$

$\mathbf{z}(k)$ is an M-dimensional column vector whose elements can be any subset of all previously estimated picture elements. For example, where several coders are compared (see below), the following two subsets of past estimated pixels are used.

In the first case it is assumed that the LPT coder operates on 2 × 2 picture blocks as shown in Figure 1. The prediction of each transform coefficient $c_i(k + 1)$ for all i is made using all six past pixels immediately adjacent to the pixel block $\mathbf{x}(k + 1)$. These six past pixels are the elements of the $\mathbf{z}(k)$ column vector

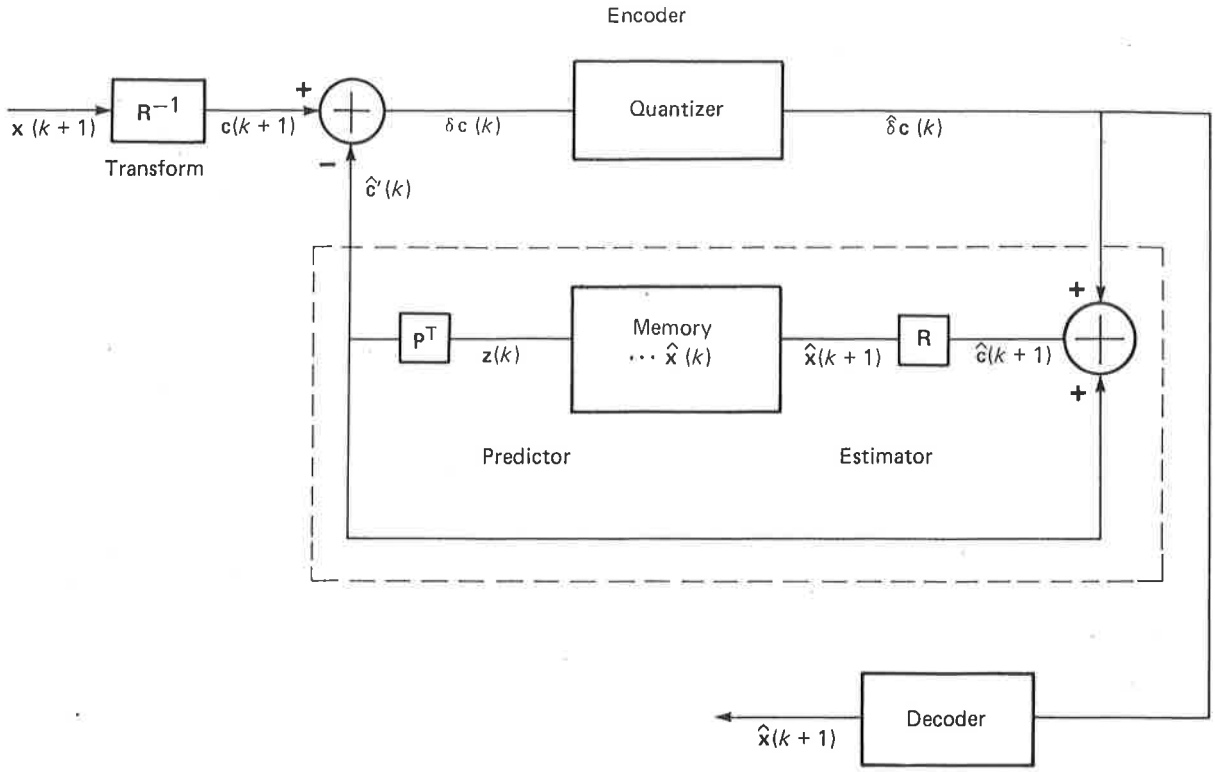


Figure 2. Linear predictive transform (LPT) coding structure. The decoder structure is similar to that of the feedback part of the encoder that is enclosed within the dashed lines

$$\mathbf{z}(k) = [x_4(k - L), x_3(k + 1 - L), x_4(k + 1 - L), x_3(k + 2 - L), x_2(k), x_4(k)]^T \quad (7)$$

In the second case the LPT coder operates on 4 × 4 picture blocks and uses all immediately adjacent past pixels, i.e.

$$\mathbf{z}(k) = [x_{16}(k - L), x_{13}(k + 1 - L), x_{14}(k + 1 - L), x_{15}(k + 1 - L), x_{16}(k + 1 - L), x_{13}(k + 2 - L), x_4(k), x_8(k), x_{12}(k), x_{16}(k)]^T \quad (9)$$

The ith element $\mathbf{p}_i^T\mathbf{z}(k)$ of the W-dimensional column vector $\hat{\mathbf{c}}'(k)$ is a preliminary estimate of the coefficient element $c_i(k + 1)$.

Quantizer

$$\hat{\delta}\mathbf{c}(k) = Q[\delta\mathbf{c}(k)] \quad (10)$$

where the operator Q represents the vector quantization of the coefficient error vector $\delta\mathbf{c}(k)$.

OPTIMUM LPT CODER

In this section a minimum MSE approach is used in the derivation of an optimum LPT coder. First, the basic mathematical constraints that are used in the generation of the optimum LPT coder are discussed. Then a mathematical statement of the minimum MSE LPT coding problem is presented and the optimum LPT coder is given in the form of a theorem. The proof of the theorem is provided in Appendix 1. The section ends by showing that classical minimum MSE predictive coding and classical minimum MSE transform coding are special cases of the proposed LPT formulation.

Constraints

There are four optimum LPT coder constraints.

Constraint 1. The basis vectors $\{\mathbf{r}_i\}$ of the transform matrix $\mathbf{R}$ will be constrained to be orthonormal, i.e.

$$\mathbf{r}_i^T\mathbf{r}_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad (11)$$

or equivalently, the matrix $\mathbf{R}$ is unitary, i.e.

$$\mathbf{R}^{-1} = \mathbf{R}^T \quad (12)$$

One reason for this constraint is to give equal weight to the energy associated with each coefficient error. This in turn simplifies the analysis, as can be seen in Appendix 1. A second reason for this constraint is that it results in uncorrelated coefficient errors as shown in Appendix 1. This in turn implies that we can approximate the vector quantizer (Equation (10)) operating on the coefficient error vector $\delta\mathbf{c}(k)$ by the simple scalar quantizers

$$Q[\delta\mathbf{c}(k)] = \{Q_1[\delta c_1(k)], \ldots, Q_W[\delta c_W(k)]\} \quad (13)$$

where $Q_j[\delta c_j(k)]$ represents the scalar quantization of the jth coefficient error $\delta c_j(k)$. Note that the scalar quantizers are not strictly optimum since the picture elements are

well known to have jointly nonGaussian distributions.

Constraint 2. The optimum transform and predictor matrices must yield coefficient error components with zero mean value, i.e.

$$E[\delta c_i(j)] = 0 \qquad \text{for all } i \tag{14}$$

The objective of this constraint is to simplify the design of the scalar quantizer since it then follows that we do not need to be concerned with coefficient error elements that have a DC component. It should be noted that the above unitary constraint (Equation (12)) and the zero mean constraint (Equation (14)) imply the following linear relation between corresponding column vectors in the transform and predictor matrices **R** and **P**

$$\mathbf{v}_W^T\mathbf{r}_i - \mathbf{v}_M^T\mathbf{p}_i = 0 \qquad \text{for all } i \tag{15}$$

where $\mathbf{v}_W$ and $\mathbf{v}_M$ are unit column vectors with W and M elements, respectively. This linear relation can be readily derived as follows: first using the unitary constraint (Equation (12)) and Equations (1) and (6) the following expressions are obtained for the ith coefficient $c_i(k+1)$ and its prediction $\hat{c}'_i(k)$

$$c_i(k+1) = \mathbf{r}_i^T\mathbf{x}(k+1)$$

$$c_i(k) = \mathbf{p}_i^T\mathbf{z}(k)$$

Second, using the above relations in the coefficient error expression

$$\delta c_i(k) = c_i(k+1) - \hat{c}'_i(k)$$

we obtain

$$\delta c_i(k) = \mathbf{r}_i^T\mathbf{x}(k+1) - \mathbf{p}_i^T\mathbf{z}(k) \qquad \text{for all } i \tag{16}$$

Finally, taking the expected value of Equation (16), using the constraint of equation (14), and assuming that the expected value of each picture sample is constant, we obtain the desired result (Equation (15)).

Constraint 3. The quantizer will be assumed to work as follows. J arbitrary coefficient error components are not quantized, i.e.

$$\hat{\delta}c_i(k) = \delta c_i(k) \tag{17}$$

for J arbitrary components of $\delta\mathbf{c}(k)$ where $J<\mathrm{W}$. The remaining $W - J$ coefficient error components are set to zero by the quantizer, i.e.

$$\hat{\delta}c_i(k) = 0 \tag{18}$$

for the remaining $W - J$ components of $\delta\mathbf{c}(k)$. The basic advantage of this constraint is that it makes the design of the optimum transform and predictor matrices a mathematically tractable problem. Clearly, after we design our transform and predictor matrices the above quantizer constraint is removed and standard linear or nonlinear quantizers are designed for the LPT coder. This approach is analogous to what is done in the design of KLT coders[5].

Constraint 4. The mean square error (MSE)

$$E\{[\mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1)]^T[\mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1)]\} \tag{19}$$

between the current picture block $\mathbf{x}(k+1)$ and its estimate $\hat{\mathbf{x}}(k+1)$ will be used as the performance criterion that the optimum transform and predictor matrix should minimize. The selection of this criterion arises due mainly to its well known mathematical tractability properties.

Statement of the problem and its solution

The problem is that given the previously stated constraints of Equations (11)–(18), the transform and predictor matrices **R** and **P** that minimize the MSE constraint of Equation (19) must be determined. The solution to the previously stated optimum LPT coder problem is given below in the form of a theorem whose proof is given in Appendix 1.

Theorem 1

The transform and predictor matrices **R** and **P** that minimize the MSE constraint of Equation (19) subject to the constraints of Equations (11)–(18) are obtained from the evaluation of the following matrix equations

$$\{E[\mathbf{x}(k+1)\mathbf{x}^T(k+1)] - \mathbf{A}\}\mathbf{r}_i = \lambda_i\mathbf{r}_i \qquad \text{for all } i \tag{20}$$

with

$$\mathbf{A} = \begin{bmatrix} E[\mathbf{x}(k+1)\mathbf{z}^T(k)] & \begin{matrix}1/2\\ \vdots\\ 1/2\end{matrix} \end{bmatrix} \begin{bmatrix} E[\mathbf{z}(k)\mathbf{z}^T(k)] & \begin{matrix}1/2\\ \vdots\\ 1/2\end{matrix} \\ 1/2 \ldots 1/2 & 0 \end{bmatrix}^{-1} \begin{bmatrix} E[\mathbf{z}(k)\mathbf{x}^T(k+1)] \\ 1/2 \quad \ldots \quad 1/2 \end{bmatrix} \tag{21}$$

and

$$\begin{bmatrix} \mathbf{p}_i \\ u_i \end{bmatrix} = \begin{bmatrix} E[\mathbf{z}(k)\mathbf{z}^T(k)] & \begin{matrix}1/2\\ \vdots\\ 1/2\end{matrix} \\ 1/2 \ldots 1/2 & 0 \end{bmatrix}^{-1} \begin{bmatrix} E[\mathbf{z}(k)\mathbf{x}^T(k+1)] \\ 1/2 \quad \ldots \quad 1/2 \end{bmatrix} \mathbf{r}_i \qquad \text{for all } i \tag{22}$$

where $\{\mathbf{r}_i\}$ and $\{\mathbf{p}_i\}$ are the columns of the optimum transform and predictor matrices **R** and **P**; $E[\mathbf{x}(k+1)\mathbf{x}^T(k+1)]$ is the second-order statistics matrix of the picture block $\mathbf{x}(k+1)$; $E[\mathbf{z}(k)\mathbf{z}^T(k)]$ is the second-order statistics matrix of the past picture elements $\mathbf{z}(k)$; $E[\mathbf{z}(k)\mathbf{x}^T(k+1)]$ is the correlation statistics matrix associated with the block $\mathbf{x}(k+1)$ and its surrounding picture elements $\mathbf{z}(k)$; λ_i is a Lagrange multiplier associated with the orthonormal constraint of Equation (11); and u_i is a Lagrange multiplier associated with

the zero mean constraint of Equation (15). The matrix inversion shown in Equations (21) and (22) is assumed to exist.

In addition, the minimum MSE obtained with these matrices is given by

$$\begin{aligned}\min_{\mathbf{R},\mathbf{P}} E\{[\mathbf{x}(k+1)-\hat{\mathbf{x}}(k+1)]^{\mathrm{T}}[\mathbf{x}(k+1)-\hat{\mathbf{x}}(k+1)]\} &= \sum_{i=J+1}^{W} E[\delta c_i^2(k)] \\ &= \sum_{i=J+1}^{W} \mathbf{r}_i^{\mathrm{T}}\{E[\mathbf{x}(k+1)\,\mathbf{x}^{\mathrm{T}}(k+1)]-\mathbf{A}\}\mathbf{r}_i \\ &= \sum_{i=J+1}^{W} \lambda_i \end{aligned} \tag{23}$$

where $\lambda_{J+1}, \ldots, \lambda_W$ are the smallest $W - J$ eigenvalues of the eigensystem of Equation (20). Note from Equation (23) that the variance of $\delta c_i(k)$ is given by the eigenvalue λ_i, i.e.

$$E[\delta c_i^2(k)] = \lambda_i \tag{24}$$

Finally, the optimum transform and predictor matrices result in uncorrelated coefficient errors.

Illustrative example

For an LPT coder operating on 2 × 2 picture blocks the optimum transform matrix, optimum predictor matrix and standard deviation of each coefficient error are presented. In the prediction of each transform coefficient the 2 × 2 LPT coder uses all six immediately adjacent past pixels to $\mathbf{x}(k+1)$ (see Figure 1 and Equation (7)). The second-order statistics of two images are used in the design of the optimum transform and predictor matrices.

Database

One of the two monochromatic images that are used to generate second-order statistics is the 'boy picture' shown in Figure 3a. The other image is another 'boy picture' with considerably more detail than that shown in Figure 3a. The two images were developed via an NTSC monochromatic TV camera. The pictures were sampled at 525 lines per frame, 450 samples per line (7.2 MHz sampling rate) and 8 bit per sample. They

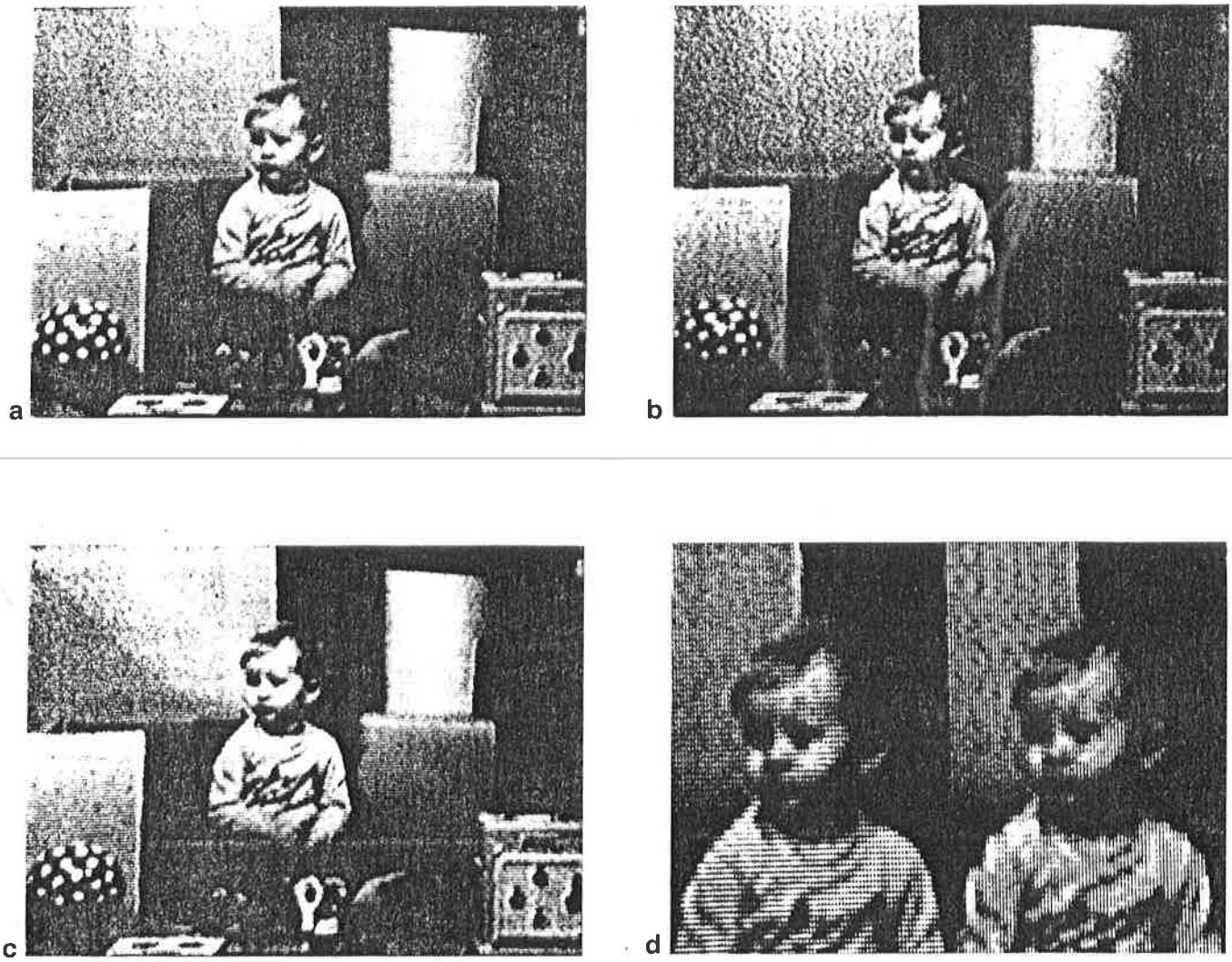


Figure 3. **a,** *original image (8 bit per sample);* **b,** *4 × 4 Hadamard image (2 bit per sample);* **c,** *simple 2 × 2 LPT image (2 bit per sample);* **d,** *enlargement of simple 2 × 2 LPT and 4 × 4 Hadamard images*

are also interlaced. In addition, the synch. and blanking information of the pictures are not used in the acquisition of second-order statistics nor in the evaluation of coder performance.

Second-order statistics

The second-order statistics derived with our database are

$$E[\mathbf{x}(k+1)\,\mathbf{x}^{\mathrm{T}}(k+1)] = \begin{bmatrix} 21236 & 21213 & 21213 & 21193 \\ 21213 & 21236 & 21193 & 21213 \\ 21213 & 21193 & 21236 & 21213 \\ 21193 & 21213 & 21213 & 21236 \end{bmatrix}$$

$$E[\mathbf{z}(k)\mathbf{z}^{\mathrm{T}}(k)] = \begin{bmatrix} 21236 & 21213 & 21173 & 21137 & 21213 & 21173 \\ 21213 & 21236 & 21213 & 21173 & 21193 & 21156 \\ 21173 & 21213 & 21236 & 21213 & 21156 & 21127 \\ 21137 & 21173 & 21213 & 21236 & 21123 & 21098 \\ 21213 & 21193 & 21156 & 21123 & 21236 & 21213 \\ 21173 & 21156 & 21127 & 21098 & 21213 & 21236 \end{bmatrix}$$

$$E[\mathbf{x}(k)\mathbf{z}^{\mathrm{T}}(k)] = \begin{bmatrix} 21193 & 21213 & 21193 & 21156 & 21213 & 21193 \\ 21156 & 21193 & 21213 & 21193 & 21173 & 21156 \\ 21156 & 21173 & 21156 & 21127 & 21193 & 21213 \\ 21127 & 21156 & 21173 & 21156 & 21156 & 21173 \end{bmatrix}$$

Optimum transform and predictor matrices

The optimum transform and predictor matrices are

$$\mathbf{R} = \begin{bmatrix} 0.2523 & 0.7182 & -0.0130 & 0.6484 \\ 0.4316 & 0.2680 & -0.7157 & -0.4792 \\ 0.4304 & 0.3008 & 0.6981 & -0.4867 \\ 0.7515 & -0.5673 & 0.0156 & 0.3363 \end{bmatrix} \tag{25}$$

$$\mathbf{P} = \begin{bmatrix} 1.1591 & 0.6210 & -0.0102 & 0.0034 \\ -0.5670 & -0.8238 & -0.5169 & -0.1474 \\ -0.8582 & 0.1369 & 0.6617 & 0.1038 \\ -0.1001 & 0.0221 & -0.0178 & 0.0366 \\ -0.5043 & -0.8136 & 0.5412 & -0.1762 \\ -0.9955 & 0.1377 & -0.6430 & 0.1612 \end{bmatrix} \tag{26}$$

Standard deviation matrix

The standard deviation of the four coefficient errors is given by

$$\mathrm{SD} = \begin{bmatrix} (\lambda_1)^{\frac{1}{2}} & (\lambda_2)^{\frac{1}{2}} \\ (\lambda_3)^{\frac{1}{2}} & (\lambda_4)^{\frac{1}{2}} \end{bmatrix} = \begin{bmatrix} 9.19 & 2.66 \\ 2.53 & 1.30 \end{bmatrix} \tag{27}$$

Two special cases

It is now shown that both classical minimum MSE predictive coding and classical minimum MSE transform coding are special cases of the LPT formulation.

Predictive coding

The general structure of a predictive coder[9] is given in Figure 4. The design equation for the predictor matrix **P** is in turn given by

$$\begin{bmatrix} \mathbf{p}_i \\ u_i \end{bmatrix} = \begin{bmatrix} E[\mathbf{z}(k)\mathbf{z}^{\mathrm{T}}(k)] & \begin{matrix} 1/2 \\ \vdots \\ 1/2 \end{matrix} \\ 1/2 \;\ldots\; 1/2 & 0 \end{bmatrix}^{-1} \begin{bmatrix} E[\mathbf{z}(k)x^{\mathrm{T}}(k+1)] \\ 1/2 \quad \ldots \quad 1/2 \end{bmatrix}$$

for all i

where $\mathbf{p}_i$, u_i, $\mathbf{z}(k)$, $x(k+1)$ and the expectations are as defined for Equation (22). Note that $x(k+1)$ is now a scalar rather than a vector since predictive coding encodes one sample at a time.

The predictive coding structure and its corresponding design equation are now shown to be special cases of the LPT formulation. To obtain the predictive coding structure, i.e Figure 4, we simply replace the transform matrix **R** in Figure 2 by a unity gain. Also, each vector variable in Figure 2, except for the past estimate vector $\mathbf{z}(k)$, is replaced by a scalar variable corresponding to that used in the predictive coding structure. The design equation for the predictive coding matrix **P** given above is readily obtained from the LPT design Equation (22) by replacing the vector $\mathbf{x}^{\mathrm{T}}(k+1)$ by the scalar $x(k+1)$ and the eigenvector $\mathbf{r}_i$ by a unity gain.

Transform coding

Figure 5 shows the general structure of a transform coder[9]. The corresponding design equation for the transform matrix **R** is the eigensystem

$$E[\mathbf{x}(k+1)\mathbf{x}^{\mathrm{T}}(k+1)]\mathbf{r}_i = \lambda_i \mathbf{r}_i$$

where $\mathbf{r}_i$, λ_i, and the expectation are as defined for Equation (20).

Transform coding can be easily derived from the LPT formulation. First the transform coding structure of Figure 5 is obtained from the LPT coding structure of Figure 2 by simply eliminating the encoder feedback enclosed with dashed lines. On the other hand, the design equation for the transform matrix **R** given above is easily derived from Equation (20) by dropping the matrix **A**. Note that the matrix **A** is a function of the correlation between presently processed and formally processed picture data. In transform coding this correlation information is not used in the coder design.

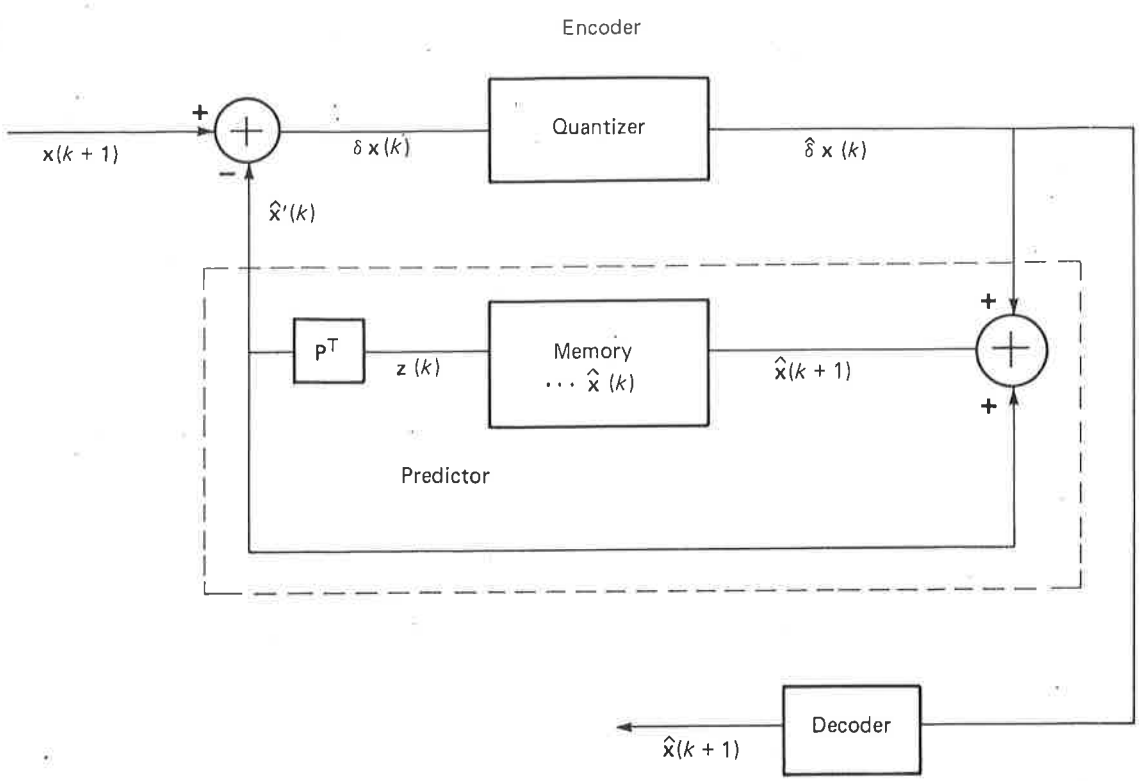


Figure 4. Linear predictive coding structure. The decoder structure is similar to that of the feedback part of the encoder that is enclosed within the dashed lines

SUBOPTIMUM LPT CODER

In this section an LPT coder that is easier to implement and generally suboptimum is derived under several constraints. This coder is used in the next section to derive an even simpler LPT coder.

Constraints

The five suboptimum LPT coder constraints are as follows.

Constraint 1. The transform matrix **R** is assumed to be fixed. In this paper the transform matrix will be assumed to be the Hadamard transform matrix which is separable and can be implemented with at most $W\log_2 W$ additions or subtractions.

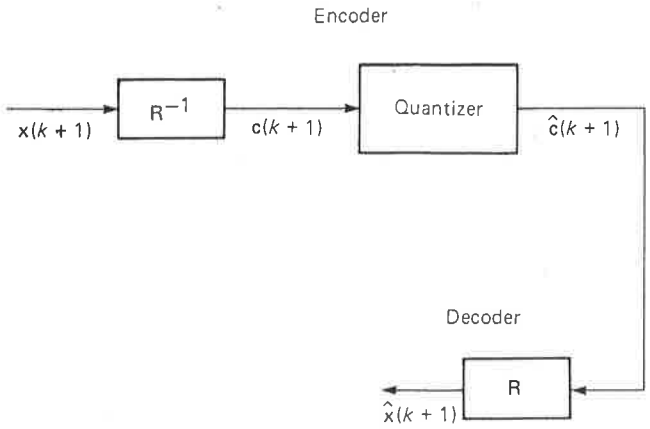


Figure 5. Linear transform coding structure

Constraints 2–5. Similar to constraints 1–4 for the optimum LPT coder except that **R** is fixed (see Equations (11)–(18)).

Statement of the problem and its solution

Given the previously stated constraint of Equations (11)–(18) and assuming that the transform matrix **R** is fixed, determine the predictor matrix **P** that minimizes the MSE (Equation (19)). Note that in general the coefficient error will be correlated. The solution of the suboptimum LPT coder problem is now given in the form of a theorem whose proof closely follows the one given in Appendix 1 for Theorem 1. For conciseness the proof for this theorem is not given here.

Theorem 2

The predictor matrix **P** that minimizes the MSE (Equation (19)) subject to the constraints of Equations (11)–(18) and fixed **R** is given by

$$\begin{bmatrix} \mathbf{p}_i \\ u_i \end{bmatrix} = \begin{bmatrix} E[\mathbf{z}(k)\mathbf{z}^{\mathrm{T}}(k)] & \begin{matrix} 1/2 \\ \vdots \\ 1/2 \end{matrix} \\ \begin{matrix} 1/2 & \dots & 1/2 \end{matrix} & 0 \end{bmatrix}^{-1} \begin{bmatrix} E[\mathbf{z}(k)\mathbf{x}^{\mathrm{T}}(k+1)] \\ \begin{matrix} 1/2 & \dots & 1/2 \end{matrix} \end{bmatrix} \mathbf{r}_i$$

for all i (28)

where $\{\mathbf{p}_i\}$, $\{u_i\}$, $E[\mathbf{z}(k)\mathbf{z}^{\mathrm{T}}(k)]$ and $E[\mathbf{z}(k)\mathbf{x}^{\mathrm{T}}(k+1)]$ are as defined above for Equation (22); and $\{\mathbf{r}_i\}$ are the basis vectors of the assumed transform matrix **R**.

In addition, the minimum MSE obtained with the optimum predictor matrix is given by

$$\min_{\mathbf{P}} E\{[\mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1)]^T[\mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1)]\}$$
$$= \sum_{i=J+1}^{W} E[\delta c_i^2(k)]$$
$$= \sum_{i=J+1}^{W} \mathbf{r}_i^T\{E[\mathbf{x}(k+1)\,\mathbf{x}^T(k+1)] - \mathbf{A}\}\mathbf{r}_i \quad (29)$$

where **A** is given by Equation (21).

Illustrative example

Again the LPT coder is assumed to act on 2 × 2 picture blocks and to use all six immediately adjacent past pixels to $\mathbf{x}(k+1)$ in the prediction of each transform coefficient. Also, the database used to design the coder is the same as that introduced in the 'illustrative example' above.

Transform matrix

The selected transform matrix is the easily implementable Hadamard transform matrix

$$\mathbf{R} = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & -1/2 & 1/2 \end{bmatrix} \quad (30)$$

Predictor matrix

The following predictor matrix was derived when we assumed the suboptimum transform matrix Equation (30):

$$\mathbf{P} = \begin{bmatrix} 1.3050 & 0.1128 & 0.1083 & -0.0444 \\ -0.8229 & -0.7817 & -0.0256 & -0.0170 \\ -0.7554 & 0.7790 & -0.1612 & 0.0185 \\ -0.0850 & 0.0329 & 0.0575 & 0.0238 \\ -0.7689 & -0.0621 & -0.8025 & -0.0321 \\ -0.8728 & -0.0809 & 0.8236 & 0.0513 \end{bmatrix} \quad (31)$$

Standard deviation matrix

The standard deviation matrix of the suboptimum LPT coefficient errors is given by

$$\mathrm{SD} = \begin{bmatrix} 8.63 & 3.39 \\ 3.37 & 1.53 \end{bmatrix} \quad (32)$$

SIMPLE LPT CODER

This section describes the construction of a simple LPT coder using the suboptimum LPT coder as a basis.

Illustrative example

The simple LPT coder is illustrated using the suboptimum 2 × 2 LPT coder of the previous section. The transform and predictor matrices of the suboptimum 2 × 2 LPT coder are given by Equations (30) and (31).

Transform matrix

The transform matrix of the simple LPT coder is the same as that used for the suboptimum 2 × 2 LPT coder (Equation (30)).

Predictor matrix

The predictor matrix is derived by roughly approximating the predictor matrix of the suboptimum 2 × 2 LPT coder (Equation (31)) with an easily implementable matrix. The simple predictor matrix is

$$\mathbf{P} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 \\ -1 & 0 & 1 & 0 \end{bmatrix} \quad (33)$$

and satisfies the zero mean constraint of Equation (15).

Standard deviation

The standard deviations of the simple LPT coefficient errors are given by

$$\mathrm{SD} = \begin{bmatrix} 9.60 & 3.84 \\ 3.84 & 1.67 \end{bmatrix} \quad (34)$$

RESULTS

Coders

The following coders are compared:

- optimum 2 × 2 KLT coder
- 4 × 4 Hadamard coder
- optimum 4 × 4 KLT coder
- simple 2 × 2 LPT coder of the illustrative example above (see Equations (30), (33) and (34))
- optimum 2 × 2 LPT coder illustrated above (see Equations (25)–(27))
- optimum 4 × 4 LPT coder using all immediately adjacent past pixels (see Equation (9))

Database

The database used to obtain the second-order statistics of all coders is similar to that used in the illustrative examples above.

Quantizer

An optimum nonuniform quantizer[12] based on Laplacian

density functions for the coefficients of the Hadamard and KLT coders (except for the DC coefficients which are not quantized) and the coefficient errors of the LPT coder is employed[13].

Bit rate

The bit rate used with each coder is 2 bit pixel^{-1}.

Bit assignment

An optimum bit assignment was obtained for each coder using as subjective criterion the blocking effect produced on our database as viewed on a standard 23 cm studio monitor. The approach employed to search for the best bit assignment was to perturb the bit assignment[13], $\{<m_i>\}$, with

$$m_i = \frac{V}{W} + \frac{(\log_2 s_i^2)}{2} - \frac{\sum_{j=1}^{W} \log_2 s_j^2}{2W} \qquad \text{for all } i \quad (35)$$

where $<m_i>$ is the nearest integer to m_i and denotes the number of bits assigned to the ith coefficient $c_i(k+1)$ of the Hadamard or KLT coders (in the case of the LPT coders $<m_i>$ denotes the number of bits assigned to the ith coefficient error $\delta c_i(k)$); V is the total number of bits assigned to each coefficient (or coefficient error) block; W is the number of samples in each coefficient (or coefficient error) block; and s_i is the standard deviation of the ith coefficient $c_i(k+1)$ (or the ith coefficient error $\delta c_i(k)$). After the set of integers $\{<m_i>\}$ are found they are judiciously adjusted such that

$$\sum_{i=1}^{W} <m_i> = V$$

It should be noted that the DC coefficients of the Hadamard and KLT coders were always assigned 8 bit, although in the actual simulation the quantizer did not quantize the DC coefficients.

Standard deviation and bit matrices

In Figure 6 the standard deviation and optimum bit assignment matrices of all the coders are given.

Coder performance

The performance of each coder is evaluated using both objective and subjective criteria.

Objective criterion

The objective criterion used was the decibels SNR which is defined as

$$\text{SNR} = 10 \log_{10} \left\{ \frac{255^2}{\sum_{i=1}^{L^2} \sum_{j=1}^{W} [x_j(i) - \hat{x}_j(i)]^2/(WL^2)} \right\}$$

where L is as defined for Figure 1 and W is the number of pixels in a picture block. Table 1 presents the SNR that is obtained when the database is processed with each coder. Note that the coders are organized in this table according to their SNR. The following results are noted.

- The optimum 2 × 2 LPT coder results in a SNR at least 0.5 dB better than the optimum 4 × 4 KLT coder.
- The optimum 2 × 2 LPT coder is only 1 dB away from the optimum 4 × 4 LPT coder which is 10 times more complex.
- The simple 2 × 2 LPT coder yields an SNR superior to that of the Hadamard coder.

Subjective criterion

The blocking effect that is observed in the processed images is used as the subjective criterion. The following results were obtained.

- The optimum 2 × 2 KLT and 4 × 4 Hadamard coders yield processed images with quite noticeable blocking effects. Figure 3b shows the processed image produced by the 4 × 4 Hadamard coder.
- The optimum 4 × 4 KLT coder results in a smaller blocking effect than the optimum 2 × 2 KLT and 4 × 4 Hadamard coders.
- All the LPT coders result in improved picture quality with considerably less blocking effects (see Figure 3c where the simple 2 × 2 LPT coder was used to generate the processed image; a split screen enlargement of Figures 3b and c is given in Figure 3d).

Coder complexity

The coder complexity is measured by the number of multiplications, additions and subtractions required by each coder in the evaluation of each picture block estimate $\mathbf{x}(k+1)$. Table 1 presents the encoder complexity associated with each coder, from which the following results are highlighted.

- The optimum 2 × 2 LPT encoder requires approximately one sixth of the operations needed by an optimum 4 × 4 KLT encoder.
- The simple 2 × 2 LPT encoder complexity is close to one third that of the 4 × 4 Hadamard encoder.

CONCLUSIONS

In this paper a minimum MSE predictive transform coding formulation has been developed and found to be appealing for two fundamental reasons.

- The LPT formulation leads to the design of simple coder implementations with satisfactory SNR performance.
- The new formulation provides a general theoretical framework from which classical minimum MSE

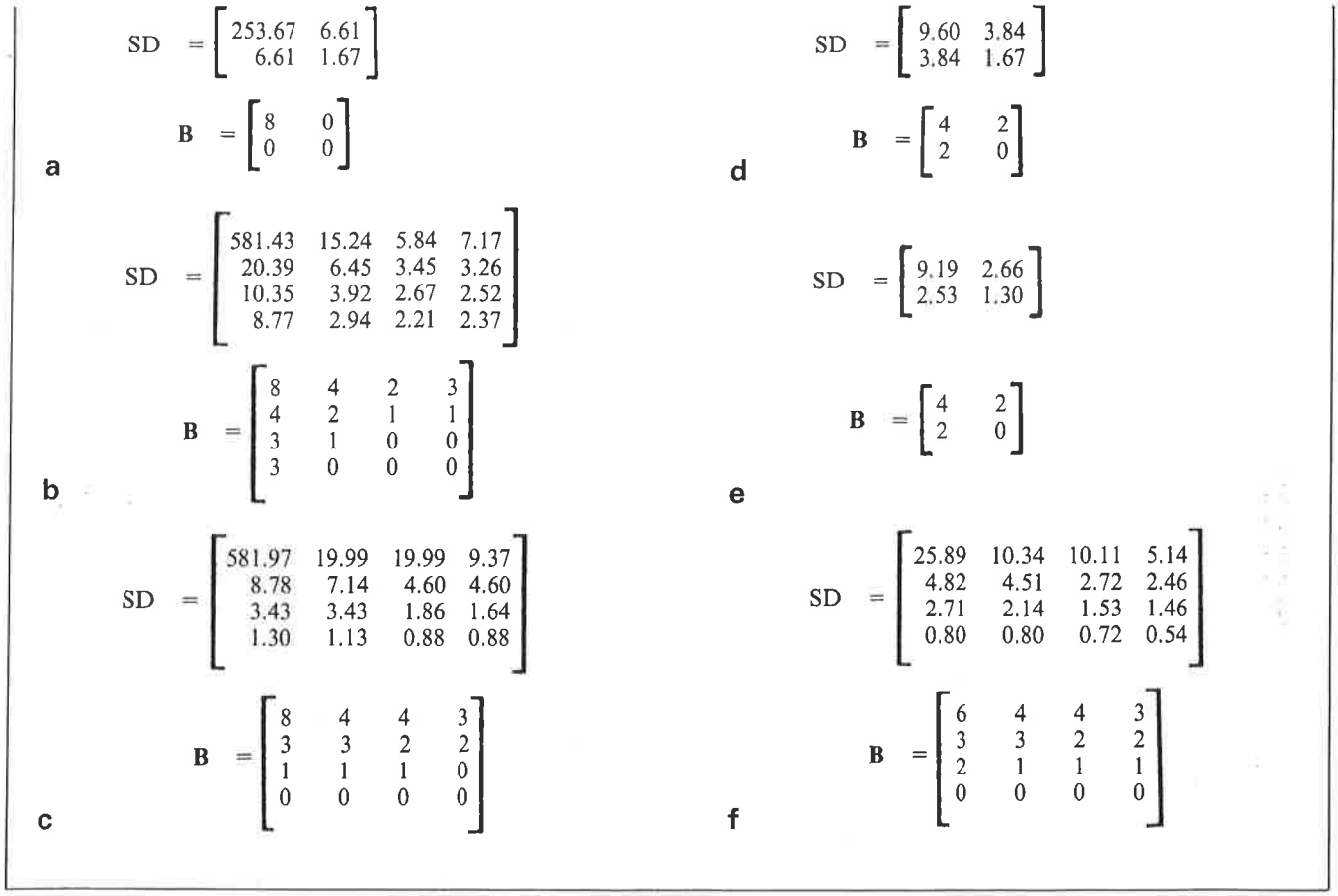


*Figure 6. Standard deviation matrix (SD) and optimum bit assignment matrix (**B**) for the coders:* **a**, *optimum 2 × 2 K–L coder case;* **b**, *4 × 4 Hadamard coder case;* **c**, *optimum 4 × 4 K–L coder case;* **d**, *simple 2 × 2 LPT coder case;* **e**, *optimum 2 × 2 LPT coder;* **f**, *optimum 4 × 4 LPT coder*

predictive coding and classical MSE transform coding arise as special cases.

Obvious extensions of the present work that should be explored are: the application of LPT coders to speech waveform coding; the design of LPT coders for composite and component colour images; the design of adaptive LPT coders; and the design of interframe/interfield LPT coders etc.

REFERENCES

1 **Graham, R E** 'Predictive quantizers of television signals' *IRE Wescon Convention Record* Vol 2 Part 4 (1958) pp 147–157

2 **Andrews, H C, Kane, J and Pratt, W K** 'Hadamard transform image coding' *Proc. IEEE* Vol 57 (Jan. 1969) pp 58–68

3 **Haralick, R M, Shanmugan, K, Young, J and Goel, D** 'A comparative study of transform data compression techniques for imagery' *Proc. National Electronics Conf.* Vol 27 (1972) pp 89–94

4 **Lei, R, Scheinberg, N and Schilling, D L** 'Adaptive delta modulation for video encoding' *IEEE Trans. Commun.* Vol C–25 (Nov. 1977) pp 1302–1314

5 **Wintz, P A** 'Transform picture coding' *Proc. IEEE* Vol 60 (July 1972) pp 809–820

Table 1. Performance and complexity of encoders

Coder	SNR (dB)		Encoder complexity per block	
	Other boy	Boy of Figure 3a	Multiplication	Addition or subtraction
Optimum 2 × 2 KLT coder	34.03	36.28	16	12
4 × 4 Hadamard coder	38.69	39.81	None	64
Simple 2 × 2 LPT coder	39.31	40.63	None	23
Optimum 4 × 4 KLT coder	40.74	42.27	256	240
Optimum 2 × 2 LPT coder	41.58	42.83	40	40
Optimum 4 × 4 LPT coder	42.95	44.11	416	416

6 **Habibi, A** 'Hybrid coding of pictorial data' *IEEE Trans. Commun.* Vol C–22 (May 1974) pp 614–624
7 **Haralick, R M and Shanmugan, K** 'Comparative study of a discrete linear basis for image data compression' *IEEE Trans. Syst. Man and Cybern.* Vol SMC–4 (Jan. 1974) pp 16–27
8 **Watson, L T, Haralick, R M and Zuniga, O A** 'Constrained transformed coding and surface fitting' *IEEE Trans. Commun.* Vol C–31 (May 1983) pp 717–726
9 **Jayant, N S and Noll, P** *Digital coding of waveforms* Prentice-Hall, Englewood Cliffs, NJ, USA (1984)
10 **Clarke, R J** *Transform coding of images* Academic Press, London, UK (1985)
11 **Netravali, A N, Prasada, B and Mounts, F W** 'Some experiments in adaptive and predictive Hadamard transform coding of pictures' *Bell Syst. Tech. J.* (Oct. 1977)
12 **Max, J** 'Quantization for minimum distortion' *IRE Trans. Inform. Theory* Vol IT–6 (March 1960) pp 7–12
13 **Kamangar, F A and Rao, K R** 'Interfield hybrid coding of component color television signals' *IEEE Trans. Comm.* Vol C–29 (Dec. 1981) pp 1740–1753
14 **Wylie, C R** *Advanced engineering mathematics* McGraw-Hill, New York, USA (1975)

BIBLIOGRAPHY

Ahmen, N, Natarajan, T and Rao, K R 'On image processing and a discrete cosine transform' *IEEE Trans. Comput.* Vol C–23 (Jan. 1974) pp 90–93
Anderson, G B and Huang, T S 'Piecewise Fourier transformation for picture bandwidth compression' *IEEE Trans. Commun. Tech.* Vol C–19 (April 1971) pp 133–140
Anderson, G B and Huang, T S 'Piecewise Fourier transformation for picture bandwidth compression correction' *IEEE Trans. Commun. Tech.* Vol C–20 (June 1972) pp 448–491
Andrews, H C and Patterson, C L III 'Singular value decompositions and digital image processing' *IEEE Trans. Acoust. Speech Signal Process.* Vol ASSP–24 (Feb. 1976) pp 26–53
Devarajan, V and Rao, K R 'Predictor adaptive DPCM coders for NTSC composite TV signals' *IEEE Trans. Commun.* Vol C–28 (July 1980) pp 1079–1084
Habibi, A and Wintz, P 'Image coding by linear transformations and block quantizations' *IEEE Trans. Commun. Tech.* Vol C–19 (Feb. 1971) pp 50–62
Haralick, R M and Zuniga, O 'Adaptive image data compression' *IEEE Conf. Patt. Recogn. Image Process., Chicago, IL, USA* (Aug. 1979)
Hotelling, H 'Analysis of a complex of statistical variables into principal components' *J. Educ. Psychology* Vol 24 (1933) pp 417–441, 498–520
Karhunen, H 'Uber lineare Methoden in der Wahrscheinlich-Keitsrechnung' *Ann. Acad. Sci. Fenn.* Ser. AI 37, Helsinki, Finland (1947) (English translation available as 'On linear methods in probability theory' *The Rand Corp., Doc. T–131* (11 Aug. 1960))
Loeve, M 'Fonctions aleatoires de seconde ordre' in **Levy, P** *Processes stochastiques et mouvement Brownien* Hermann, Paris, France (1948)
Netravali, A N and Limb, J O 'Picture coding: a review' *Proc. IEEE* Vol 68 (March 1980) pp 366–406
Pratt, W K, Chen, W H and Welch, L R 'Slant transform image coding' *IEEE Trans. Commun.* Vol C–22 (Aug. 1974) pp 1075–1093
Rao, K R, Narasimhan, M A and Revuluri, K 'Image data processing by Hadamard-Haar transforms' *IEEE Trans. Comput.* Vol C–23 (Sept. 1975) pp 888–896
Sawada, K and Kotera, H '32 Mbit/sec transmission of NTSC color TV signals by composite DPCM coding' *IEEE Trans. Commun.* Vol C–26 (Oct. 1978) pp 1432–1438
Scheinberg, N, Barba, J, Feria, E and Schilling, D L 'Composite NTSC color video bandwidth compressor' *IEEE Trans. Commun.* Vol C–32 (Dec. 1984) pp 1331–1335
Tasto, M and Wintz, P 'Image coding of adaptive block quantizations' *IEEE Trans. Commun. Tech.* Vol C–19 (Dec. 1971) pp 957–972
Zschunke, W 'DPCM picture coding with adaptive prediction' *IEEE Trans. Commun.* Vol C–25 (Nov. 1977) pp 1295–1302

APPENDIX 1: PROOF OF OPTIMUM LPT CODER DESIGN EQUATIONS

This appendix shows that the optimum transform and predictor matrices are found from Equations (20)–(22) and that the minimum MSE is given by Equation (23). It is also shown that the optimum transform and predictor matrices yield uncorrelated coefficient errors. The proof consists of eight steps.

Step 1. Making use of Equations (2), (4)–(6) and the quantizer constraints of Equations (17) and (18) gives

$$\mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1) = \sum_{i=J+1}^{W} r_i \delta c_i(k) \tag{36}$$

where J represents the number of coefficient error components unaffected by the quantizer.

Step 2. Using Equation (36) in Expression (19) gives

$$\begin{aligned} \text{MSE} &= E\{[\mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1)]^{\mathrm{T}}[\mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1)]\} \\ &= \sum_{i=J+1}^{W} \sum_{i=J+1}^{W} \mathbf{r}_i^{\mathrm{T}} \mathbf{r}_j E[\delta c_i(k) \delta c_j(k)] \end{aligned} \tag{37}$$

Step 3. Using the orthonormal constraint of Equation (11) we find

$$\text{MSE} = \sum_{i=J+1}^{W} E[\delta c_i^2(k)] \tag{38}$$

Step 4. Making use of Equation (16) in Equation (38) gives

$$\sum_{i=J+1}^{W} E[\delta c_i^2(k)] = \sum_{i=J+1}^{W} \mathbf{q}_i^{\mathrm{T}} E[\mathbf{y}(k+1)\, \mathbf{y}^{\mathrm{T}}(k+1)]\, \mathbf{q}_i \tag{39}$$

where

$$\mathbf{q}_i^{\mathrm{T}} = [-\ \mathbf{p}_i^{\mathrm{T}}\ \mathbf{r}_i^{\mathrm{T}}] \tag{40a}$$

and

$$\mathbf{y}^T(k+1) = [\mathbf{z}^T(k)\ \mathbf{x}^T(k+1)] \tag{40b}$$

Step 5. Lagrange multipliers are used to formulate the minimization of Equation (39) with respect to the vectors $\mathbf{q}_{J+1}, \ldots, \mathbf{q}_W$ and subject to the constraints of Equations (11) and (15), i.e. the following minimization has to be performed

$$\min_{\mathbf{q}_{J+1}\cdots\mathbf{q}_W} \left\{ \sum_{i=J+1}^{W} \{\mathbf{q}_i^T E[\mathbf{y}(k+1)\,\mathbf{y}^T(k+1)]\,\mathbf{q}_i - \lambda_i(\mathbf{r}_i^T\mathbf{r}_i - 1) - u_i(\mathbf{v}_W^T\mathbf{r}_i - \mathbf{v}_M^T\mathbf{p}_i)\} \right\} \tag{41}$$

where λ_i and u_i are Lagrange multipliers.

Step 6. Using standard minimization techniques[14] the minimization of Expression (41) is performed yielding the desired Equations (20)–(22).

Step 7. The minimum MSE Expression (23) is now obtained. First $\mathbf{q}_i^T E[\mathbf{y}(k+1)\mathbf{y}^T(k+1)]\,\mathbf{q}_i$ is expanded to obtain

$$\begin{aligned}&\mathbf{q}_i^T E[\mathbf{y}(k+1)\,\mathbf{y}^T(k+1)]\,\mathbf{q}_i \\ &= \mathbf{r}_i^T E[\mathbf{x}(k+1)\mathbf{x}^T(k+1)]\,\mathbf{r}_i + \mathbf{p}_i^T E\,[\mathbf{z}(k)\mathbf{z}^T(k)]\mathbf{p}_i \\ &- \mathbf{p}_i^T E[\mathbf{z}(k)\,\mathbf{x}^T(k+1)]\mathbf{r}_i - \mathbf{r}_i^T E[\mathbf{x}(k+1)\,\mathbf{z}^T(k)]\mathbf{p}_i\end{aligned} \tag{42}$$

Second, Equation (22) is solved for $E[\mathbf{z}(k)\mathbf{x}^T(k+1)]\,\mathbf{r}_i$ to yield

$$E[\mathbf{z}(k)\mathbf{x}^T(k+1)]\mathbf{r}_i = \begin{bmatrix} E[\mathbf{z}(k)\mathbf{z}^T(k)] & \vdots & 1/2 \\ & & 1/2 \end{bmatrix} \begin{bmatrix} \mathbf{p}_i \\ u_i \end{bmatrix} \tag{43}$$

Third, Equation (43) is used in Equation (42) to yield

$$\begin{aligned}&\mathbf{q}_i^T E[\mathbf{y}(k+1)\,\mathbf{y}^T(k+1)\mathbf{q}_i \\ &= \mathbf{r}_i^T\{E[\,\mathbf{x}(k+1)\,\mathbf{x}^T(k+1)] - \mathbf{A}\}\mathbf{r}_i\end{aligned} \tag{44}$$

where $\mathbf{A}$ is given by Equation (21). Last, making use of the eigensystem in Equation (20) and the orthonormal constraint of Equation (11) in Equation (44) gives

$$\mathbf{q}_i^T E[\mathbf{y}(k+1)\mathbf{y}^T(k+1)]\,\mathbf{q}_i = \lambda_i \tag{45}$$

which implies the desired result (Equation (23)).

Step 8. Next, it is shown that the optimum transform and predictor matrices result in uncorrelated coefficient errors. First, using an approach similar to that of steps 4 and 7 it is shown that

$$\begin{aligned}&E[\delta c_j(k)\delta c_i(k)] \\ &= \mathbf{r}_j^T\{E[\mathbf{x}(k+1)\,\mathbf{x}^T(k+1)] - \mathbf{A}\}\mathbf{r}_i\end{aligned} \tag{46}$$

where $\mathbf{r}_i$, $E[\mathbf{x}(k+1)\,\mathbf{x}^T(k+1)]$ and $\mathbf{A}$ are as defined for Equation (20). Second, using Equation (20) in Equation (46) it is found that

$$E[\delta c_j(k)\delta c_i(k)] = \lambda_i \mathbf{r}_j^T\mathbf{r}_i \qquad \text{for all } i \text{ and } j \tag{47}$$

Last, using the orthonormal condition of Equation (11) in Equation (47) gives the desired result

$$E[\delta c_j(k)\delta c_i(k)] = 0 \qquad \text{for all } i \neq j \tag{48}$$