USCIPI REPORT #700

Interframe Coding of Digital Images Using Transform and Hybrid Transform/Predictive Techniques

by

John Alan Roese

June 1976

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 404
Los Angeles, CA 90089-2564 U.S.A.

ACKNOWLEDGMENTS

The author wishes to acknowledge the sustained encouragement and assistance received from many individuals during the course of this research. The guidance and direction provided by my committee chairman, Dr. W. K. Pratt, are gratefully acknowledged. Sincere thanks are also extended to Dr. H. C. Andrews and Dr. A. Habibi for their advice and prolonged personal support. The cooperative professional spirit exhibited by Dr. G. S. Robinson during this research is sincerely appreciated.

This endeavor was made possible through the combined efforts of the Defense Advanced Research Projects Agency and the Naval Undersea Center, San Diego. I wish to thank the sponsors of this research, Col. H. M. Federhen, H. J. Whitehouse, and Dr. E. H. Wrench, for generously providing support and technical insight into the problem area. Also, the tolerant and understanding attitudes of D. A. Hanna and Dr. R. S. French were highly instrumental in the successful completion of this project.

CONTENTS

ILLUSTRATIONS

TABLES

ABSTRACT

In the design of digital image coding systems, the principal objective is to achieve high quality receiver image reconstructions with a minimum number of transmitted code bits. Bit rate reductions are achieved by exploiting statistical redundancies within an image. This is combined by transmission of only those portions of the mathematical image representation which the human observer is most sensitive to. This dissertation describes research intended to extend current image coding techniques to the coding of sequences of digital images transmitted over a digital communications channel. The emphasis is directed towards definition of an image coding system that exploits temporal as well as spatial image redundancies.

A primary objective of this investigation is to develop a class of interframe hybrid transform/predictive coders having near optimum levels of performance. The interframe hybrid coder implementations considered employ two-dimensional unitary transforms in the spatial domain coupled with first-order DPCM predictive coding in the temporal domain. Based on a statistical image representation, a model is developed for the hybrid coder transform coefficient temporal difference variance matrix. With this model, theoretical MSE performance levels for the hybrid coder with zonal coding are determined as a function of spatial subblock size.

Implementations of the interframe hybrid coder using discrete cosine and Fourier transforms are experimentally evaluated. High quality image reconstructions are demonstrated for reductions of 32:1 in average pixel bit rate. Operational considerations investigated for the hybrid interframe coder include initial conditions, spatial and temporal adaptation, reinitialization, and total transmission bit rate. Also, sensitivity of the interframe hybrid coder to channel error is studied.

Comparisons are drawn between hybrid transform/predictive and three-dimensional transform interframe coders. Theoretical zonal sampling and zonal coding MSE performance for three-dimensional cosine transform coders are evaluated for different frame storage requirements and spatial subblock sizes.

A tabular summary of experimental performance and system design parameters for the main classes of interframe coders is presented.

1. INTRODUCTION

PROBLEM DEFINITION

In the design of digital image coding systems, the principal objective is to achieve high quality receiver image reconstructions with a minimum number of transmitted code bits. Reduction in the number of transmitted code bits allows for reduced channel bandwidth requirements, higher image transmission frame rates, and lower transmitter power requirements.

Numerous single frame digital image coding systems have been developed. Most of these systems employ transform, linear predictive, or combinations of transform and linear predictive techniques to lower the average number of bits transmitted per picture element. These bit rate reductions are achieved by exploiting statistical redundancies existing within an image combined with transmitting only those portions of the mathematical image representation to which the human observer is most sensitive.

This dissertation describes research which extends these image coding techniques to the coding of frame-to-frame sequences of digital images transmitted over a digital communications channel. The emphasis is directed towards definition of an image coding system that exploits temporal as well as spatial image redundancies.

INTERFRAME CODING SURVEY

A property fundamental to the transmission of digital image sequences is that new information is transferred by changes in a relatively small number of the picture elements or pixels transmitted each frame. From a statistical viewpoint, frame-to-frame similarity represents a high level of interframe correlation between temporally adjacent pixels.

Many different image coding techniques have been investigated for interframe image coding. For example, interframe coding experiments have been performed using three-dimensional Fourier and Hadamard transforms [1]. These simulations have shown that improvements by a factor of five in average pixel bit rate can be achieved with interframe coding when compared with intraframe coding techniques.

A three-dimensional transform coder based on the Hadamard transform has been implemented [2]. This coder uses a block size of 4 X 4 pixels with the storage of 4 frames in time. It is designed for real-time bandwidth compression of standard National Television Standards Committee (NTSC) television signal transmission. In general, multidimensional transform coder implementations are characterized by serious computational complexity and frame storage requirements.

For these reasons, investigations have also been made of frame differential interframe coding techniques. The design of these systems is based on the premise that only a small percentage of pixel values actually change between successive frames. For example, in frame replenishment interframe coding, only those pixel values having changes between frames that exceed a threshold are transmitted [3]. Replenishment of the previous frame at the receiver is accomplished with the transmitted data. Experiments using typical Picturephone television images have achieved good coding results at one bit per pixel using conditional frame replenishment techniques.

A major disadvantage of frame replenishment techniques is the requirement for data buffering prior to transmission. Data buffering is needed because the data are generated at an uneven rate due to frame-to-frame variations in the number of pixels having changes exceeding a fixed threshold. Also, the buffer size determines how well motion within the image sequence can be reproduced at the receiver. It has been experimentally determined that a buffer size of ten frames is required to transmit television images containing a moderate degree of motion.

Several variations of the basic frame replenishment techniques have been proposed [4]–[6]. Two of these variations are transmission of clusters of data to reduce the buffer size requirements and transmission of only the change in pixel values for those pixels whose value changes have exceeded a fixed threshold.

A significant shortcoming of frame differential image coding techniques is that only temporal correlation in the data is exploited. Thus, the overall efficiency of frame differential coders is low since the spatial correlations within each frame are not utilized.

Another approach investigated is that of three-dimensional, third order differential pulse code modulation (DPCM) predictive coding [7]. In this technique, pixel value estimates are generated as linear combinations of adjacent previously scanned pixels in the present and previous frame. Experience gained in the operation of three-dimensional, third order DPCM coders indicates an extreme sensitivity to frame-to-frame variations in image statistics. Adaptation for local statistical differences is required for successful operation.

RESEARCH OBJECTIVES

Within the broad context of the interframe image coding problem as defined above, several specific research objectives have been identified. The primary objective of this research is the development of a class of hybrid interframe coders with near optimum levels of performance. The theoretical and experimental levels of performance achieved are compared with those of three-dimensional transform coders.

A closely related objective is to establish the degree of performance degradation that would result from simplified hybrid interframe coder implementations.

A further objective is the investigation of hybrid interframe coder sensitivity to channel error.

Since the performance of interframe coders is dependent on frame-to-frame signal correlation, effects of subject and camera motion and procedures for both spatial and transform domain motion compensation are presented.

DISSERTATION ORGANIZATION

Chapter 1 provides a brief definition and historical survey of the basic concepts involved in interframe image coding. This chapter also defines the primary objective of the dissertation to be the extension of conventional intraframe image coding concepts to include exploitation of interframe temporal correlations in sequences of digital images.

Specific research objectives are identified and a brief overview of the dissertation structure is presented.

Chapter 2 is mainly tutorial in the sense that it reviews the significant techniques used in intraframe image coding. The most commonly used transform coding and predictive coding techniques are presented in detail. This chapter concludes with a discussion of an important concept in intraframe coding--that of hybrid transform/predictive image coding.

Chapter 3 discusses the application of three-dimensional transform coding approaches to interframe image coding. Consideration is given to the development of three-dimensional Fourier and cosine transform coders. An analysis of both theoretical and experimental performance levels for the three-dimensional cosine transform coder is presented.

Chapter 4 introduces a new concept in interframe coders--that of hybrid two-dimensional transform/DPCM linear predictive coders. Two interframe hybrid coder implementations are defined. The first employs discrete cosine transforms and the second uses discrete Fourier transforms. Investigations of simplified coder implementations and effective channel bit transfer rates are also presented.

Chapter 5 investigates the sensitivity of interframe hybrid coders to the presence of channel error. Simulation results indicate image degradation effects as a function of channel error rate.

Chapter 6 identifies interframe coding image degradations resulting from camera motion. Discussions of spatial and transform domain techniques for introducing motion compensation into the interframe hybrid coders are presented.

Chapter 7 presents a tabular summary of the main classes of interframe image coders. Conclusions are given regarding the interframe hybrid and transform coders investigated.

IMAGE FIDELITY CRITERIA

To objectively evaluate the coding efficiency of the various interframe coders investigated, two image fidelity criteria measures are used. The first criterion, normalized mean square error (NMSE), is a measure of the normalized mean square error between a

single frame input to an interframe coder $f(j,k,\ell)$ and its output $\hat{f}(j,k,\ell)$ averaged over the ℓth frame. Here, j and k are spatial coordinates, ℓ is the temporal coordinate denoting frame number, and M × M is the frame size. Normalization is achieved by dividing the mean square error by the mean signal energy within the frame. The expression for the NMSE at the ℓth frame is

$$\text{NMSE} = \frac{\frac{1}{M^2}\sum_{j=0}^{M-1}\sum_{k=0}^{M-1}\left[f(j,k,\ell)-\hat{f}(j,k,\ell)\right]^2}{\frac{1}{M^2}\sum_{j=0}^{M-1}\sum_{k=0}^{M-1}\left[f(j,k,\ell)\right]^2}. \tag{1.1}$$

The second criterion, SNR, measures the ratio of peak-to-peak signal to RMS noise for the ℓth frame,

$$\text{SNR} = -10\log_{10}\left\{\frac{\frac{1}{M^2}\sum_{j=0}^{M-1}\sum_{k=0}^{M-1}\left[f(j,k,\ell)-\hat{f}(j,k,\ell)\right]^2}{Y_{MAX}^2}\right\}, \tag{1.2}$$

where Y_{MAX} represents the maximum luminance value of $f(j,k,\ell)$, typically 255. These criteria are used extensively throughout this dissertation and serve to support subjective comparisons of coder performance.

EXPERIMENTAL INTERFRAME DATA BASES

In order to experimentally evaluate the various interframe coding schemes investigated, it was necessary to generate several multiframe data bases. The data base used throughout most of this research consists of 16 images digitized from 16 sequential frames of a 16-mm, 24-frame-per-sec (fps) motion picture. This sequence of 16 frames, referred to as the "head and shoulders" data base, provides a close-in view of a subject engaged in conversation. The relatively few number of frames available corresponds to only 2/3 of a second in real time. Thus, the amount of subject motion during the 16-frame sequence, although readily apparent, is not extreme. Approximately 50 percent of each image in this data base is devoted to background material which, except for slight frame-to-frame misregistration effects, remains essentially constant. Frames 1, 8, and 16 of the "head and shoulders" data base

are illustrated in Figure 1-1 (a), (b), and (c), respectively. The "head and shoulders" data base is monochromatic and has a resolution of 256 rows by 256 columns. Each individual picture element or pixel was digitized to one of 2^8 intensity levels.

The "head and shoulders" data base is suitable for interframe coder simulations designed for picture telephone-like applications. However, other newer application areas require examination of image sequences that contain more detailed objects and which illustrate the translational and geometric distortion effects that result from camera motion.

For the above stated reasons, a second multiframe data base was generated. This second data base was digitized from a sequence of consecutive frames of a 35-mm motion picture film. The scene is an aerial view showing an approach to a complex of buildings and roads photographed from an airborne platform of known altitude, velocity, and look-down angle. The film contains geometric distortions and translational effects due to camera motion. This "moving camera" data base consists of 64 sequential frames with 512 by 512 row and column resolution. The data are monochromatic with 2^8 bits of intensity resolution per pixel. Figure 1-2 (a) through (e) shows, respectively, frames 1, 16, 32, 48, and 64 of the "moving camera" data base. From these frames, it is seen that the original images do contain highly detailed structures and that translation and distortion effects resulting from camera motion are present.

(a) Frame 1

(b) Frame 8

(c) Frame 16

Figure 1-1. "Head and Shoulders" Data Base

(a) Frame 1

(b) Frame 16

(c) Frame 32

(d) Frame 48

(e) Frame 64

Figure 1-2. "Moving Camera" Data Base

2. INTRAFRAME IMAGE CODING

The primary techniques developed for single image or intraframe image coding in the spatial domain are transform, linear predictive, and hybrid transform/linear predictive. This chapter contains operational descriptions and comparisons of these coding methods.

TRANSFORM IMAGE CODING

The basic components comprising an intraframe transform image coding system are illustrated in Figure 2-1. In this system, the input to the intraframe coder, denoted as $f(j,k)$, is a discrete two-dimensional spatial domain representation of the original image to be coded. Separable row and column one-dimensional unitary transforms are applied to the complete image or to subblock partitions of the image. From the resulting array of transform domain coefficients $F(u,v)$, a subset is selected for quantization, encoding, and subsequent transmission over a digital communications channel. At the receiver, a spatial domain image reconstruction $\hat{f}(j,k)$ is obtained by application of inverse one-dimensional transforms.

Transform coding techniques have been explored extensively both in theory and by simulation. It is shown that significant bit rate reductions can be achieved in many applications with minimal image degradation. Simulation results indicate that a bit rate reduction to 1.5 bits/pixel is obtained for monochrome image transform coding in 16 X 16 picture element blocks, whereas color images require about 2.0 bits/pixel [8]. The bit rate is reduced even further with an adaptive transform coding system.

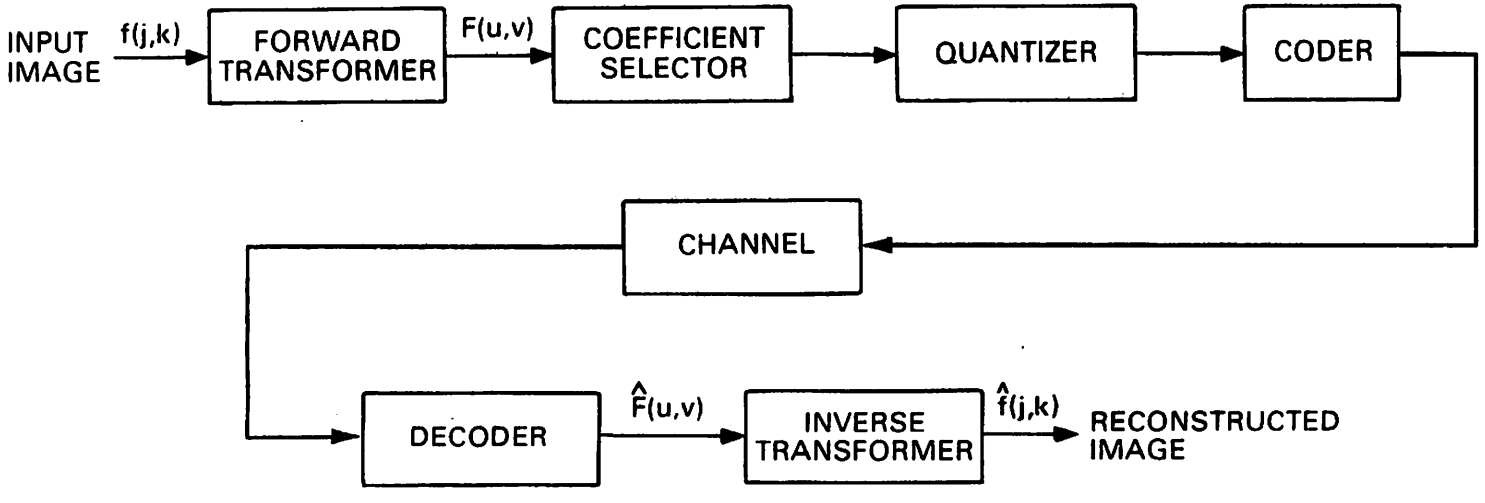


Figure 2-1. Generalized Intraframe Transform Image Coder

Unitary Transforms

This section presents a brief tutorial discussion of unitary transforms [9]. Transform image coding is based on the premise that the transform domain representation of an image has an energy distribution that is more compact and therefore more efficient to code than its spatial domain representation. For image coding applications, the transform operator is usually chosen from the class of linear unitary operators, i.e., operators [T] having the property that [T] [T]* = [T]*[T] = [I] where the symbol * denotes the conjugate transpose of the complex operator. When restricted to real valued variables, [T] is termed an orthogonal operator.

Properties of unitary transformation include: length or energy preservation of vectors undergoing transformation; transformation of sets of orthonormal basis vectors into other sets of orthonormal basis vectors; and transformation of sums of squares into sums of squares. This latter property implies that mean square error computations are equivalent in both the transform and spatial domains, and that the total energy in both domains is preserved.

Unitary transforms commonly used for image coding applications include the Karhunen-Loeve, cosine, Fourier, Slant, Hadamard, and Haar transforms [10]–[13]. Of these, only the Karhunen-Loeve transform results in completely uncorrelated components in the transform domain. However, the remaining transforms do offer the advantages of fast computational algorithms.

Equations 2.1 and 2.2 define the general forms for the two-dimensional discrete transform of an image. The two-dimensional transform of a J X K image array is itself a J X K array of transform domain coefficients. The forward transform operation can be expressed as

$$F(u, v) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} f(j, k)\, a(j, k, u, v) \ , \tag{2.1}$$

where j and k are spatial coordinates, u and v are transform domain coordinates, and a(j,k, u,v) is the forward transform kernel. The corresponding reverse transform is

$$\hat{f}(j,k) = \sum_{u=0}^{J-1} \sum_{v=0}^{K-1} F(u, v)\, a^{-1}(j, k, u, v) \ , \tag{2.2}$$

where $a^{-1}(j,k,u,v)$ is the reverse transform kernel. When the function $\hat{f}(j,k)$ is equivalent to $f(j,k)$, equation 2.2 is called an inverse transform. The transform kernel, $a(j,k,u,v)$, is called separable if it can be expressed in the form

$$a(j, k, u, v) = a_j(j, u)\, a_k(k, v) \ . \tag{2.3}$$

Computationally, transform separability means that a two-dimensional transform of an image can be performed by first applying a one-dimensional transform along the rows of $f(j,k)$

$$F(u,k) = \sum_{j=0}^{J-1} f(j, k)\, a_j(j, u) \ , \tag{2.4}$$

followed by a one-dimensional transform along the columns of $F(u,k)$ which yields

$$F(u,v) = \sum_{k=0}^{K-1} F(u, k)\, a_k(k, v) \ . \tag{2.5}$$

Often, it is useful to write separable two-dimensional transforms in matrix form. When $[f]$ is an image matrix representation of the array, $f(j,k)$, and $[F]$ is the transformed matrix representation of $F(u,v)$, then a separable two-dimensional transform may be expressed as

$$[F] = [A_j]\, [f]\, [A_k] \ , \tag{2.6}$$

where $[A_j]$ and $[A_k]$ are one-dimensional transform matrices along the image rows and columns. In the case where $[A_j]$ and $[A_k]$ have inverses, the corresponding two-dimensional inverse matrix transform is

$$[f] = [A_j]^{-1}\, [F]\, [A_k]^{-1} \ . \tag{2.7}$$

KARHUNEN-LOEVE TRANSFORM. The Karhunen-Loeve transform provides an uncorrelated transform domain representation and is a special case of an eigenvector matrix transformation. By letting $[f_j]$ be a column vector which represents the rows of image matrix $[f]$, the covariance matrix $[C_j]$ of vector $[f_j]$ becomes

$$[C_j] = E\left\{ [f_j(i) - E\{f_j(i)\}]\, [f_j(ii) - E\{f_j(ii)\}]^t \right\} , \tag{2.8}$$

where $i, ii = 0, 1, \ldots, J-1$. The eigenvectors of matrix $[C_j]$ are column vectors, $[x_{j_i}]$, which satisfy the relationship

$$[C_j]\,[x_{j_i}] = \lambda_{j_i}\,[x_{j_i}] \tag{2.9}$$

for $i = 0, 1, \ldots, J-1$, where λ_{j_i} are the eigenvalues of matrix $[C_j]$. The Karhunen-Loeve row transform matrix, $[X_j]$, is constructed from the set of j column vectors as

$$[X_j] = [x_{j_0}, x_{j_1}, \ldots, x_{j_{J-1}}] \ . \tag{2.10}$$

For eigenvalues located on the diagonal of matrix $[\lambda_j]$, the row eigenmatrix relationship is

$$[C_j]\,[X_j] = [X_j]\,[\lambda_j] \ . \tag{2.11}$$

In a similar fashion, the Karhunen-Loeve column eigenmatrix is

$$[C_k]\,[X_k] = [X_k]\,[\lambda_k] \ . \tag{2.12}$$

Using row and column transform matrices $[X_j]$ and $[X_k]$, the separable two-dimensional Karhunen-Loeve transform of image [f] can be written as

$$[F] = [X_j]\,[f]\,[X_k]^t \ . \tag{2.13}$$

When the Karhunen-Loeve transform is used for image coding, significant problems exist due principally to the excessive amounts of computation required. For example, the image covariance matrix must first be estimated. Then, the covariance matrix must be diagonalized to find its eigenvalues and eigenvectors. Finally, the transformation operation must be performed. In general, no fast computational algorithm exists for the Karhunen-Loeve transform. However, when it can be computed, the Karhunen-Loeve transformation is useful as a standard of performance for comparisons of image coding techniques [14].

DISCRETE COSINE TRANSFORM. For image arrays, f(j,k), which are real valued, the two-dimensional discrete cosine transform is

$$F(u,v) = \frac{1}{JK} \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} f(j,k) \cdot \left\{ \cos\frac{\pi}{2}\left[\frac{(2j+1)\,u}{J} + \frac{(2k+1)\,v}{K}\right] + \cos\frac{\pi}{2}\left[\frac{(2j+1)\,u}{J} - \frac{(2k+1)\,v}{K}\right] \right\} \tag{2.14}$$

The inverse two-dimensional discrete cosine transform can be written as

$$f(j,k) = \sum_{u=0}^{J-1} \sum_{v=0}^{K-1} F(u,v) \cdot \left\{ \cos\frac{\pi}{2}\left[\frac{(2j+1)\,u}{J} + \frac{(2k+1)\,v}{K}\right] + \cos\frac{\pi}{2}\left[\frac{(2j+1)\,u}{J} - \frac{(2k+1)\,v}{K}\right] \right\} \tag{2.15}$$

The multidimensional discrete cosine transform is separable. In two dimensions, the transform kernel,

$$\cos\frac{\pi}{2}\left[\frac{(2j+1)\,u}{J}+\frac{(2k+1)\,v}{K}\right]+\cos\frac{\pi}{2}\left[\frac{(2j+1)\,u}{J}-\frac{(2k+1)\,v}{K}\right],$$

can be written in the form of equation 2.3 by employing the trigonometric relationship $\cos(A+B)+\cos(A-B)=2\cos A\cdot\cos B$. Thus, the two-dimensional discrete cosine transform can be computed as two sequential one-dimensional transforms; one transform performed on the rows of the array of image intensity values, followed by another transformation along the columns. The expressions for the resulting forward and inverse one-dimensional discrete cosine transforms are

$$F(u)=\frac{1}{J}\sum_{j=0}^{J-1}f(j)\cos\frac{(2j+1)\,u\pi}{2J} \tag{2.16}$$

and

$$f(j)=\sum_{u=0}^{J-1}F(u)\cos\frac{(2j+1)\,u\pi}{2J}\,, \tag{2.17}$$

respectively. The expressions of equations 2.16 and 2.17 differ from those of Ref. [12] only by a normalization constant.

The matrix formulation for the cosine transform,

$$[\mathcal{C}]=\left[\frac{1}{\sqrt{J}}\cos\frac{(2j+1)\,u\pi}{2J}\right] \tag{2.18}$$

is real and orthogonal. Thus, the forward and inverse matrix forms for the two-dimensional cosine transform of the image [f] can be written as

$$[F]=[\mathcal{C}]\,[f]\,[\mathcal{C}]^{t} \tag{2.19}$$

and

$$[f]=[\mathcal{C}]^{t}\,[F]\,[\mathcal{C}]\,. \tag{2.20}$$

DISCRETE FOURIER TRANSFORM. The two-dimensional discrete Fourier transform of a complex square image array, f(j,k), can be expressed in series form as

$$F(u,v)=\frac{1}{JK}\sum_{j=0}^{J-1}\sum_{k=0}^{K-1}f(j,k)\exp\left[-2\pi i\left(\frac{uj}{J}+\frac{vk}{K}\right)\right]. \tag{2.21}$$

The inverse Fourier transform is

$$f(j,k) = \sum_{u=0}^{J-1} \sum_{v=0}^{K-1} F(u,v) \exp\left[2\pi i \left(\frac{uj}{J} + \frac{vk}{K}\right)\right] \quad . \tag{2.22}$$

The two-dimensional Fourier transform can also be computed as two sequential one-dimensional transforms since its transform kernel

$$\exp\left[\pm 2\pi i \left(\frac{uj}{J} + \frac{vk}{K}\right)\right]$$

is separable and symmetric. In one dimension, the discrete Fourier transform of a complex vector f(j) of length J is given by

$$F(u) = \frac{1}{J} \sum_{j=0}^{J-1} f(j) \exp\left(-\frac{2\pi i}{J} uj\right) \quad . \tag{2.23}$$

The corresponding inverse one-dimensional discrete Fourier transform is

$$f(j) = \sum_{u=0}^{J-1} F(u) \exp\left(\frac{2\pi i}{J} uj\right) \quad . \tag{2.24}$$

The discrete Fourier transform is well suited for real-time image processing applications as it can be implemented by a combination of multipliers and filters. By substituting $2ju = j^2 + u^2 - (j-u)^2$ into equation 2.23, the expression for the one-dimensional Fourier transform becomes

$$F(u) = \exp\left(-\frac{\pi i u^2}{J}\right) \frac{1}{J} \sum_{j=0}^{J-1} f(j) \exp\left(\frac{\pi i (j-u)^2}{J}\right) \exp\left(-\frac{\pi i j^2}{J}\right) \quad . \tag{2.25}$$

Equation 2.25 defines the Chirp-Z transform (CZT) representation of the discrete Fourier transform [15]. It can be interpreted as a multiplication of the vector to be transformed by a discrete chirp, periodic convolution of the multiplied vector by a discrete chirp, and post multiplication by a discrete chirp. A configuration for the CZT with parallel implementation of the complex arithmetic operations is shown in Figure 2-2 [16].

The two-dimensional Fourier transform can be expressed in matrix form by defining the symmetric unitary matrix, $[\mathcal{F}]$, as

$$[\mathcal{F}] = \left[\frac{1}{\sqrt{J}} \exp\left(-\frac{2\pi i}{J} ju\right)\right] \quad . \tag{2.26}$$

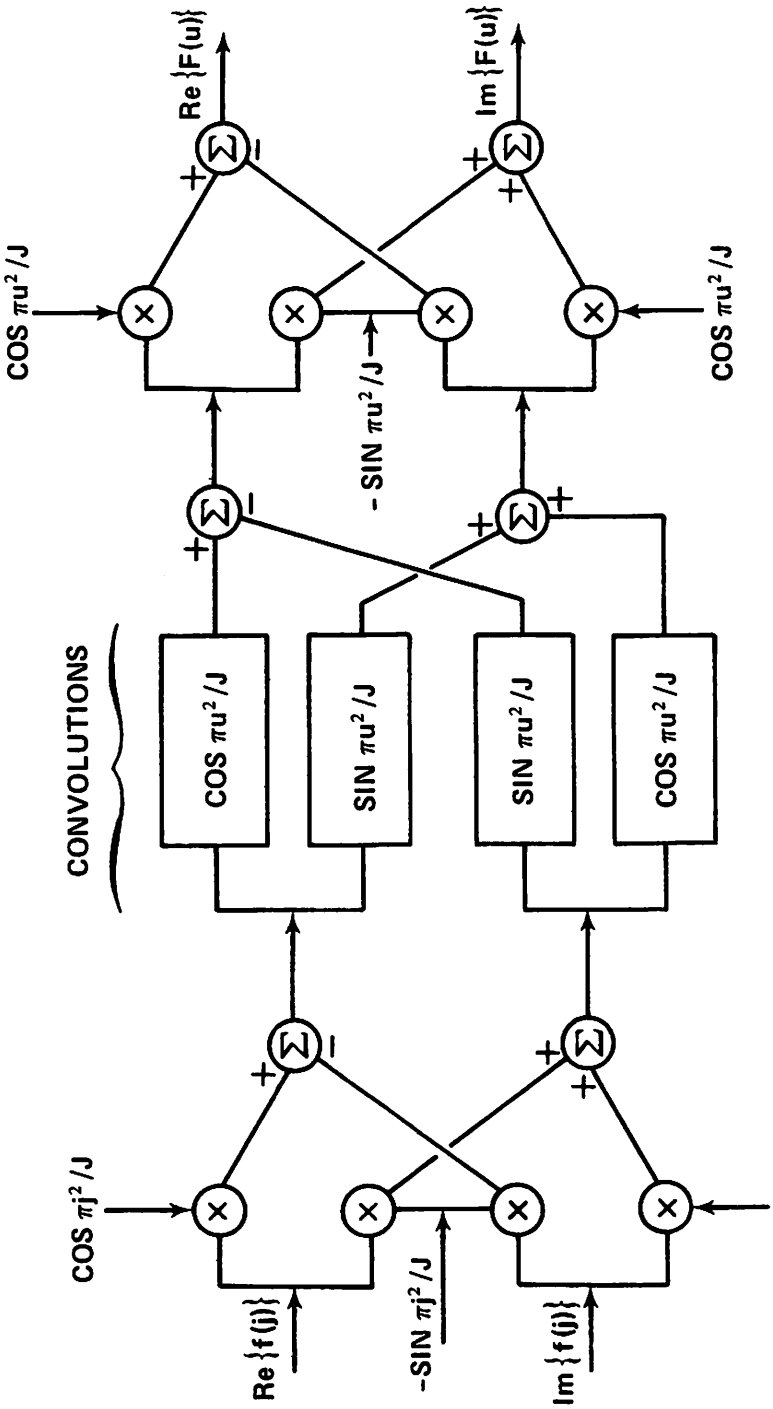


Figure 2-2. CZT Implementation for Discrete Fourier Transform

Thus, in matrix form, the forward transform of image [f] can be written as

$$[F] = [\mathcal{T}]\,[f]\,[\mathcal{T}] \quad . \tag{2.27}$$

Also, the inverse transform can be expressed in the form

$$[f] = [\mathcal{T}]^{*}[F]\,[\mathcal{T}]^{*} \quad . \tag{2.28}$$

Bit Allocation and Quantization

In transform image coding, the effective bandwidth reduction achieved is determined by the size of the subset of transform coefficients selected for transmission and the number of bits allocated to code each selected coefficient. Commonly used coefficient selection strategies include zonal sampling and threshold sampling.

In zonal sampling, a subset of the total number of transform coefficients is selected for transmission. In the simplest case of zonal sampling, the selected transform coefficients all lie within a single predetermined geometric zone and are quantized to the same number of levels. In terms of a mean square error criterion, the optimum form for this zone has been determined to be the maximum variance zone [10]. The maximum variance zone is defined by the subset of transform coefficients having the largest variances based on horizontal and vertical correlation matrix models of the image.

Extension of the above concept to include multiple zones is termed zonal coding. In zonal coding, transform coefficients within each zone are quantized to the same number of levels. Quantizer bit allocations for each zone are based on the logarithm of the transform coefficient variances. Here, coefficient selection is implicit in the number of code bits allocated for each zone. Those coefficients in zones allocated a non-zero number of bits define the subset of selected transform coefficients. Detailed descriptions of optimal bit allocation and quantization techniques used in zonal coding applicatings are presented in, respectively, Appendices A and B.

By contrast, the subset of transform coefficients selected by threshold sampling consist of those coefficients which exceed a preset threshold value. In this technique, the location of each selected transform coefficient as well as its value must be transmitted. Thus, for most image coding applications, threshold sampling is less efficient than the more commonly used zonal sampling methods.

PREDICTIVE IMAGE CODING

In predictive image coding, the high degree of correlation between a given pixel value and its nearest neighbor allows an image to be represented by coding only the difference between each pixel value and its predicted value, where the predicted value is based on previously scanned pixels. In a predictive coding system, each pixel value is subtracted from its predicted value and the resulting difference signal is quantized and coded for transmission. At the receiver, the quantized difference signal is combined with its predicted value to form the reconstructed pixel value.

There are two main types of predictive coding schemes which have been applied to image coding. These are termed Delta Modulation and Differential Pulse Code Modulation predictive coders.

Delta Modulation

The delta modulation coder is the simplest implementation of a predictive coder. It generates positive or negative pulses depending on the polarity of the difference between an input signal and its predicted value where the predicted value is based on the single previous pixel value. The resulting sequence of positive and negative pulses is then encoded in a binary representation for transmission.

A limitation of the delta modulation system is that the difference pulse is quantized to just two levels. In general, large changes in signal amplitudes cannot be followed accurately as the quantizer step sizes are normally kept small to reduce quantization errors. Modifications to the basic delta modulation system include multiple level outputs and adaptation of quantization levels to reduce slope overload [17]–[19]. However, such implementations do result in increased coder complexity.

Differential Pulse Code Modulation

The use of a multilevel quantized difference signal based on a previous pixel value or values to generate a predicted pixel value is termed a differential pulse code modulation (DPCM) coding system. First order linear prediction schemes are commonly used in DPCM

coding systems to produce the quantized difference signal which is coded and transmitted. A block diagram for a first order linear predictor DPCM is shown in Figure 2-3. In this system, quantization is performed with a quantizer designed for the probability density of the difference signal. Thus, efficient design of a DPCM system requires knowledge of the statistics of the input data difference signal.

The first-order DPCM predictor of Figure 2-3 is parametric in the quantities a_0 and a_1. The values for constants a_0 and a_1 are normally chosen to minimize the mean-square error,

$$\mathcal{E} = E\left\{[f(j) - \hat{f}(j)]^2\right\} , \tag{2.29}$$

between the input f(j) and its estimate $\hat{f}(j)$, where the first-order linear estimate of f(j) is given by

$$\hat{f}(j) = a_1 f(j-1) + a_0 . \tag{2.30}$$

Minimizing equation 2.30 to solve for a_0 and a_1 requires simultaneous solution of the two equations:

$$\frac{\partial \mathcal{E}}{\partial a_0} = 2a_0 - 2E\{f(j)\} + 2a_1 E\{f(j-1)\} = 0 \tag{2.31}$$

and

$$\frac{\partial \mathcal{E}}{\partial a_1} = 2a_1 E\{f^2(j-1)\} - 2E\{f(j) f(j-1)\} + 2a_0 E\{f(j-1)\} = 0 . \tag{2.32}$$

The general forms for the solutions are

$$a_0 = \frac{E\{f(j)\} E\{f^2(j-1)\} - E\{f(j) f(j-1)\} E\{f(j-1)\}}{E\{f^2(j-1)\} - E^2\{f(j-1)\}} \tag{2.33}$$

and

$$a_1 = \frac{E\{f(j) f(j-1)\} - E\{f(j-1)\} E\{f(j)\}}{E\{f^2(j-1)\} - E^2\{f(j-1)\}} . \tag{2.34}$$

Under the assumption of a zero-mean random process, these equations simplify to

$$a_0 = 0 \tag{2.35}$$

and

$$a_1 = \frac{E\{f(j) f(j-1)\}}{E\{f^2(j-1)\}} . \tag{2.36}$$

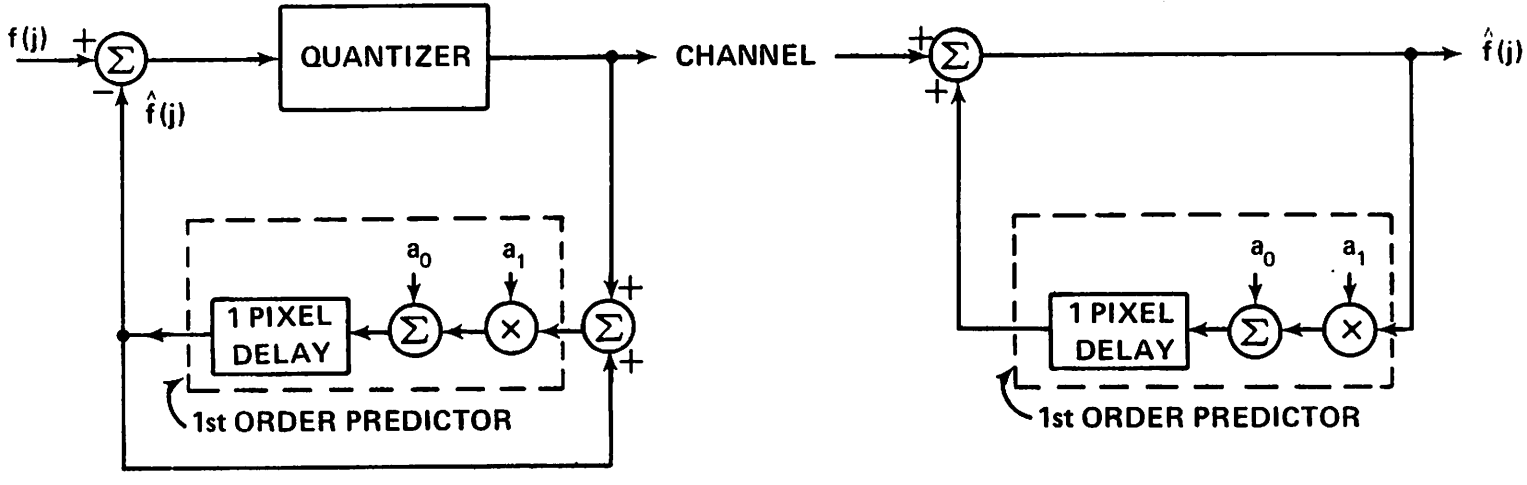


Figure 2-3. First-Order Linear Predictor DPCM Coder

Adaptive DPCM coder implementations are also possible [20]. For example, the quantizer reconstruction levels can be made adaptive by adjusting them to match the input data. Equivalently, adaptation can be achieved by simply scaling the quantizer input data to fit preset quantizer parameters.

Difference Signal Quantization

In systems which encode difference signals, maximum performance is achieved when the form of the quantizer is matched to the statistics of the difference signal. In general, differential signals are characterized by Laplacian densities [21], [22]. Thus, basic DPCM coding systems, such as that of Figure 2-3, will require quantizers tailored to Laplacian difference signal densities.

HYBRID TRANSFORM/PREDICTIVE CODING

In the previous sections, applications of both transform and DPCM predictive techniques to image coding have been discussed. A unifying theory which relates transform coding using lower triangular operators to DPCM coders has been developed for Markov data [23].

Characteristics of conventional transform coding which make this class of techniques preferable over DPCM for certain image coding applications include: higher image fidelity reconstructions at low bit rates; distribution of image coding degradations in a more visually pleasing manner; greater immunity to variations in statistics between different images; and lower sensitivity to channel error effects. However, certain other advantages occur when DPCM image coding techniques are employed. These include: superior image coding performance at higher bit rates; less complex hardware implementations with minimal delays in signal coding; and virtual elimination of the storage requirements common to transform coders [24].

Habibi has proposed a hybrid intraframe transform/DPCM coding technique which retains many of the attractive characteristics of both techniques [25]. This system employs a one-dimensional unitary transform coder cascaded with parallel DPCM coders. A block

diagram of the hybrid coder is shown in Figure 2-4. This coder performs one-dimensional transforms on successive segments along each row of the image. The bank of parallel DPCM coders is then applied to the transform coefficients in each column. Since only one-dimensional unitary transforms are performed along individual rows or row segments of the image, the resulting hybrid coder equipment complexity and total computations are considerably less than that required for two-dimensional transform coders.

The hybrid transform/DPCM coder encodes differences between vertically adjacent transform coefficients. Efficient operation is achieved when the number of bits for the quantizer in each DPCM coder is allocated on the basis of the transform coefficient variances. Since transform coefficient differences are being encoded, each DPCM quantizer is designed for difference signals having Laplacian densities.

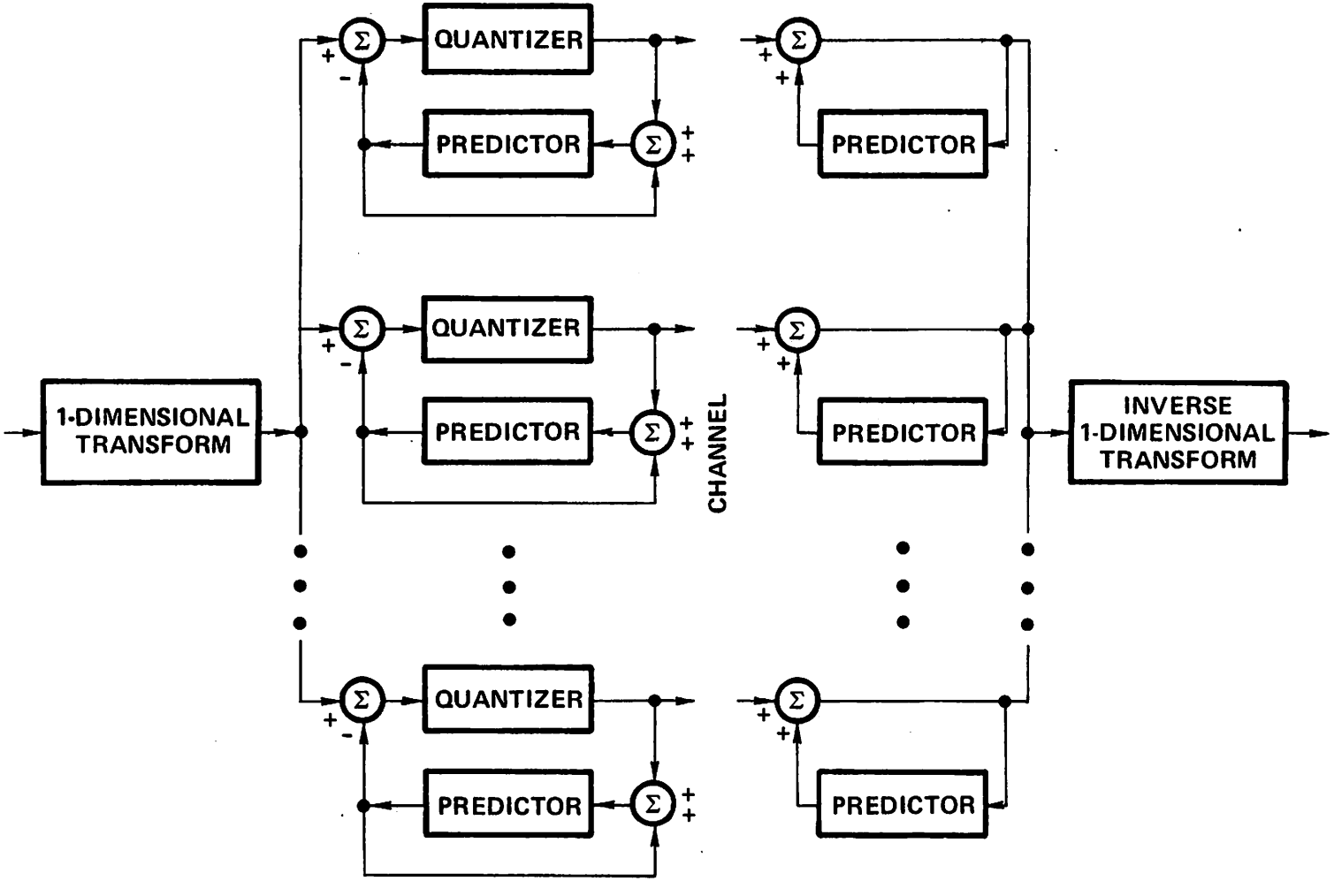


Figure 2-4. Hybrid Intraframe Transform/DPCM Coder

3. INTERFRAME TRANSFORM IMAGE CODING

In high repetition rate sequences of images, such as those encountered in television, new information is contained in only a relatively small number of picture elements or pixels. The vast majority of pixels within each frame represents background material which does not significantly change between frames. From a statistical viewpoint, the similarity of pixel values from one frame to the next implies interframe correlation between temporally adjacent pixels. Thus, the class of statistical coding techniques developed to exploit spatial correlations for coding single frames of data can, in principle, be extended to include frame-to-frame correlations, thereby achieving still further reductions in channel bandwidth requirements.

This chapter investigates a previously uninvestigated technique for three-dimensional transform interframe coding. This coding technique employs discrete cosine transforms in three dimensions, i.e., two spatial dimensions plus the temporal dimension. This approach offers near optimum levels of performance combined with an efficient computational implementation of the basic transform. In addition, implementation comparisons are made between three-dimensional cosine transform and three-dimensional Fourier transform interframe coders.

In Chapter 2, applications of the Karhunen-Loeve, cosine, and Fourier transforms to intraframe coding were described. Although optimum in the sense of completely decorrelating image data in the transform domain, it was indicated that the Karhunen-Loeve transform has primarily received only theoretical consideration, due principally to excessive computational requirements. However, other transforms which approach the optimum performance levels of the Karhunen-Loeve transform but without the attending computational problems have been intensively investigated. It should be noted that a serious drawback to the potential implementation of these transforms for interframe coding is the requirement for multiple frame storage of the transform coefficients. This characteristic of three-dimensional

transform coders severely limits their usefulness for implementation in practical interframe coding systems.

The typically large quantities of data which must be accommodated in an interframe image coding system can be conveniently represented by the three-dimensional array of pixel values shown in Figure 3-1. In this array, corresponding pixels in successive frames are treated as being temporally adjacent.

To facilitate dealing with these interframe three-dimensional data structures, a new notation is presented. The three-dimensional array denoting the sequence of length L frames of two-dimensional images will be denoted as $\{f\}$, where

$$\{f\} = \{[f_0], [f_1], \ldots, [f_{L-1}]\} \ . \tag{3.1}$$

In this definition, $[f_0], [f_1], \ldots, [f_{L-1}]$ represents an ordered sequence of equal size two-dimensional image arrays. Due to computational considerations, the individual frames are commonly partititioned into subblocks for processing. As illustrated in Figure 3-1, temporal alignment of subblocks within each image results in the formulation of multiple three-dimensional data arrays, or subcubes.

MATHEMATICAL FORMULATION

In interframe transform coding, three-dimensional discrete unitary transformations are performed on the data. Let $f(j,k,\ell)$ denote a three-dimensional array of pixel intensity values for a sequence of digital images. Also, let $F(u,v,w)$ be the three-dimensional array obtained by taking the three-dimensional transform in the (j,k,ℓ) domain. If the size of the three-dimensional image array is J X K X L, then the transform coder can be described in the general form

$$F(u,v,w) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} \sum_{\ell=0}^{L-1} f(j,k,\ell)\, \phi(u,v,w,j,k,\ell) \ . \tag{3.2}$$

The inverse operation is expressed by

$$f(j,k,\ell) = \sum_{u=0}^{J-1} \sum_{v=0}^{K-1} \sum_{w=0}^{L-1} F(u,v,w)\, \phi^{-1}(u,v,w,j,k,\ell) \ , \tag{3.3}$$

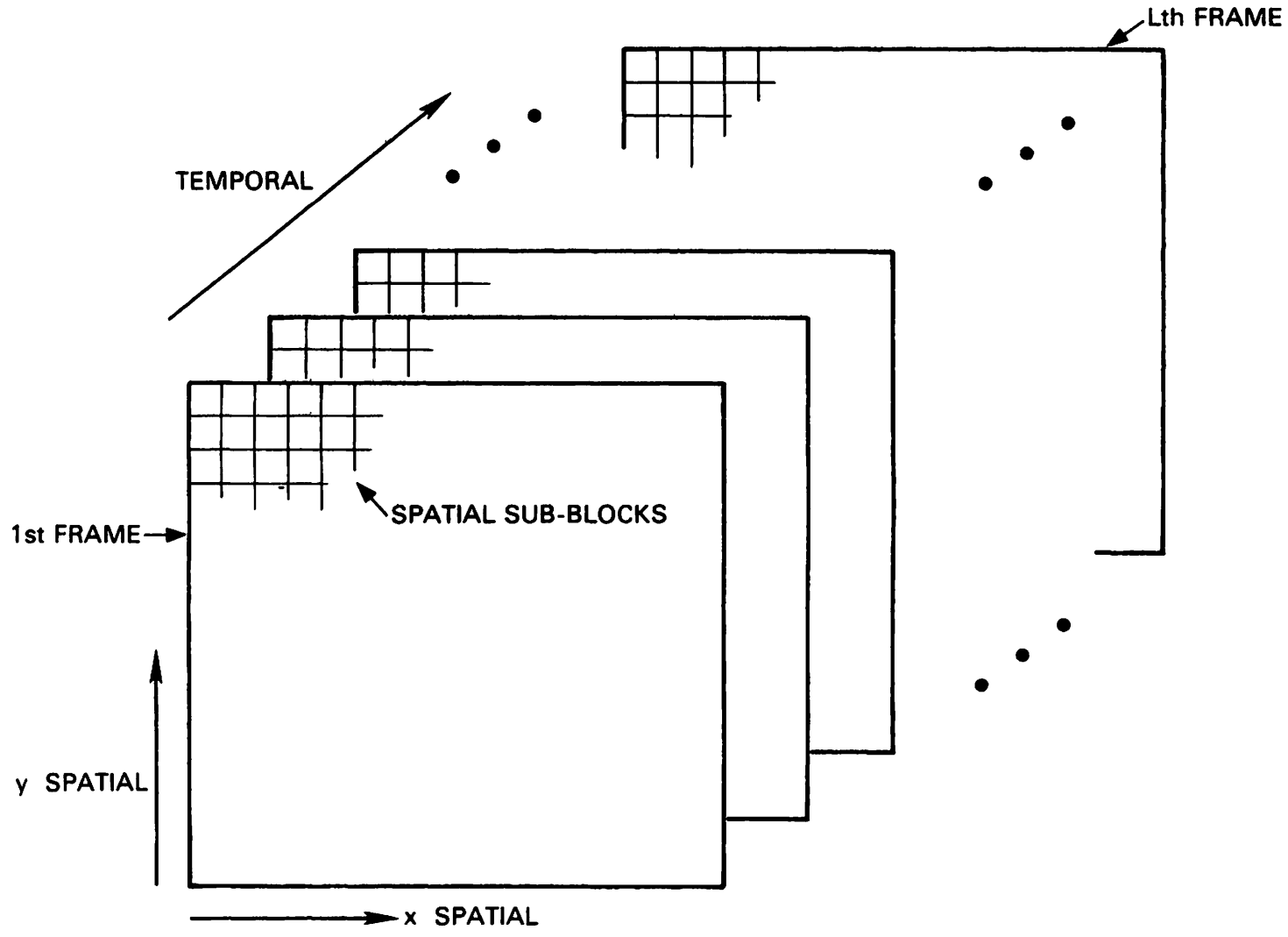


Figure 3-1. Three-Dimensional Image Data Array

where j and k are spatial coordinates, ℓ is the frame number, u,v,w are transform domain coordinates and $\phi(u,v,w,j,k,\ell)$ represents a set of three-dimensional basis functions.

For the three-dimensional discrete Fourier transform, equations 3.2 and 3.3 assume the form

$$F(u,v,w) = \frac{1}{JKL} \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} \sum_{\ell=0}^{L-1} f(j,k,\ell) \exp\left[-2\pi i \left(\frac{uj}{J} + \frac{vk}{K} + \frac{w\ell}{L}\right)\right] \tag{3.4}$$

and

$$f(j,k,\ell) = \sum_{u=0}^{J-1} \sum_{v=0}^{K-1} \sum_{w=0}^{L-1} F(u,v,w) \exp\left[2\pi i \left(\frac{uj}{J} + \frac{vk}{K} + \frac{w\ell}{L}\right)\right] . \tag{3.5}$$

Since the three-dimensional Fourier kernel,

$$\exp\left[\pm 2\pi i \left(\frac{uj}{J} + \frac{vk}{K} + \frac{w\ell}{L}\right)\right]$$

is separable, the three-dimensional transform is computable as a one-dimensional transform in the temporal direction followed by a two-dimensional transform in the spatial domain. The two-dimensional spatial transform is, in turn, computed as a one-dimensional transform along the rows followed by a one-dimensional transform along the columns. The basic one-dimensional discrete Fourier transform that must be performed is given in the previous chapter as equation 2.23. The discrete cosine transform is likewise separable and its one-dimensional form is that of equation 2.16.

TRANSFORM COEFFICIENT PROBABILITY DENSITIES

As a preliminary to the specification of parameters for quantization of transform coefficients for transmission, the general nature of the transform coefficient density functions must be established. In this section, certain assumptions are made regarding the statistical character of these density functions. Experimental verification of these assumptions are made using the 16 frame "head and shoulders" data base.

In the three-dimensional array of transform coefficients resulting from successively applying separable one-dimensional transforms to a J X K X L array of intensity values, the term F(0,0,0) corresponds to the three-dimensional dc term. Since the dc terms are positive valued, this coefficient can be considered as a sample from a distribution of

non-negative values. For this investigation, the distribution of dc terms is modeled by a scalar Rayleigh density,

$$p(x) = \frac{x}{\alpha^2} \exp(-x^2/2\alpha^2)\, U(x) \quad . \tag{3.6}$$

The remaining coefficients in the three-dimensional array are characterized by both positive and negative values. For these coefficients, a Laplacian density model is assumed, i.e.,

$$p(x) = \frac{1}{\sqrt{2}\,\sigma} \exp\left(-\frac{\sqrt{2}\,|x|}{\sigma}\right) . \tag{3.7}$$

The 16-frame "head and shoulders" data base has been used to validate the above stated assumptions. From this data base, the statistics of the dc and higher order terms resulting from successive applications of discrete one-dimensional cosine transforms were extracted. In this analysis, the sequence of frames was partitioned into 2048 subcubes of size 8 X 8 X 8.

Figure 3-2 is a histogram of the dc cosine transform coefficients computed from each subcube of the "head and shoulders" data base. This figure shows that the density function assumed for the three-dimensional cosine transform dc coefficients can be approximated by a Rayleigh model.

In a similar manner, the assumption of Laplacian densities for the higher order transform coefficients was investigated. Figure 3-3 shows the histogram for one of the low-order cosine transform coefficients adjacent to F(0,0,0). This histogram illustrates that the higher order transform coefficient densities may be reasonably modeled as Laplacian.

BANDWIDTH REDUCTION

The property of image transforms exploited for bandwidth reduction is the compression of the image energy into a relatively small number of transform domain coefficients. This property stems from the inherent correlation existing between spatially adjacent and temporally adjacent pixels in sequences of natural images.

Bandwidth reduction is achieved by selection for transmission of only those transform domain coefficients which contain the most image energy. The measure of image energy employed for coefficient selection is the variance of the transform domain coefficients. Thus, based on a three-dimensional array of variances of the transform coefficients, those coefficients which contain the greatest amounts of image energy can be selected.

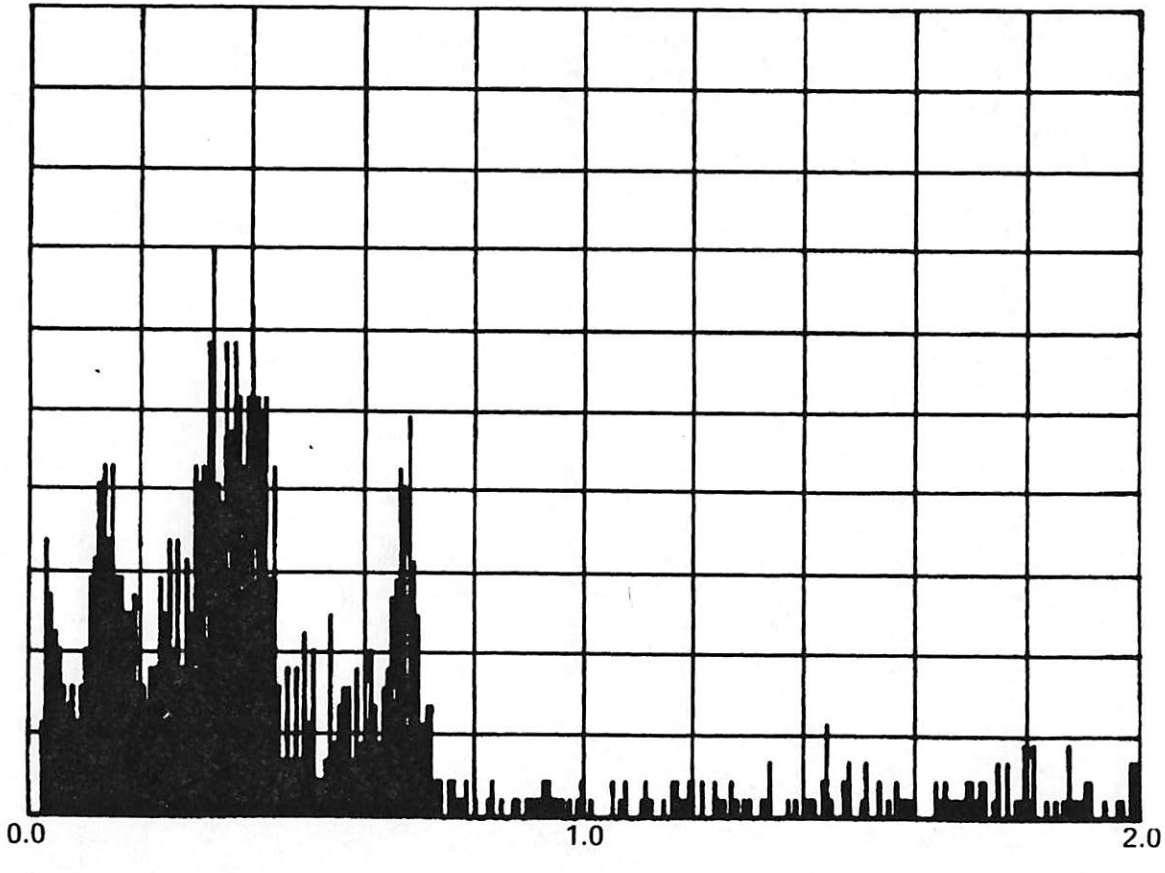


Figure 3-2. Histogram of Cosine Transform DC Coefficients for "Head and Shoulders" Data Base

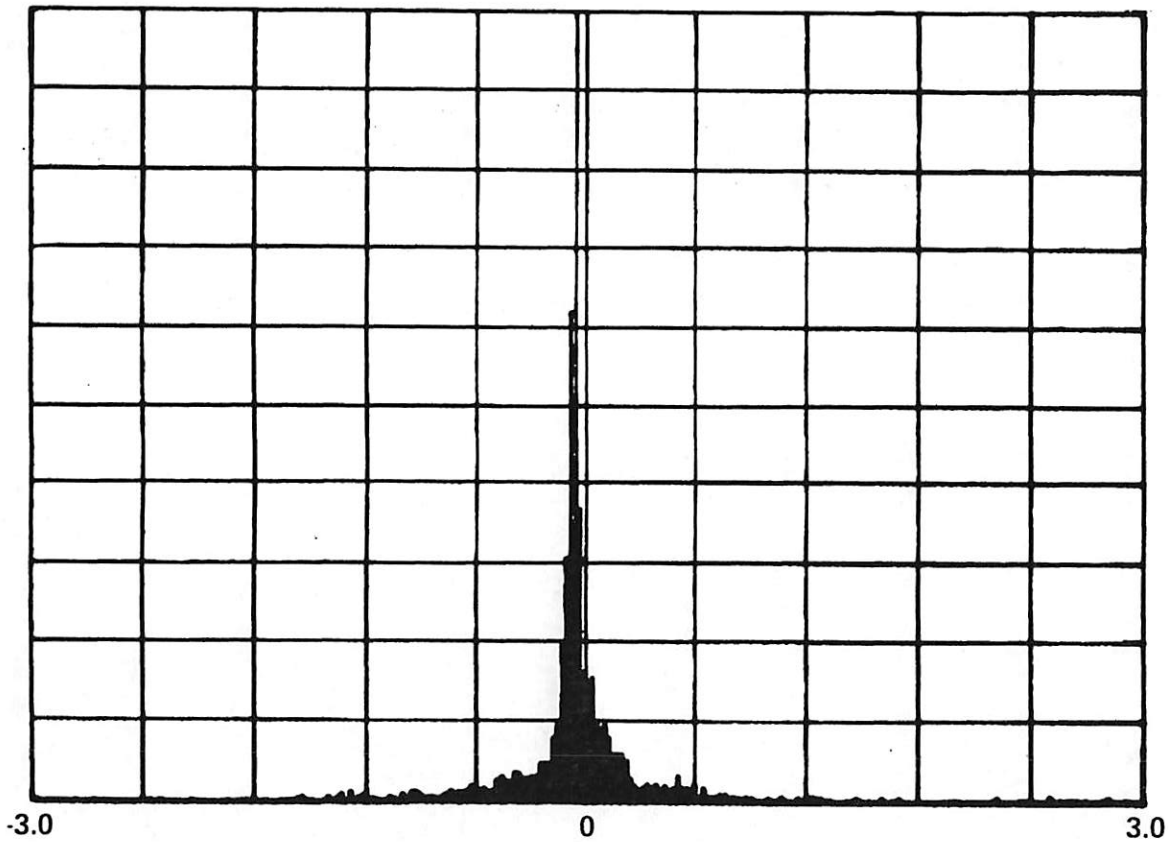


Figure 3-3. Histogram of Cosine Transform Low-Order Coefficients for "Head and Shoulders" Data Base

Transform Coefficient Variance Array

An expression for the three-dimensional array of transform coefficient variances, $\{V_F\}$, can be obtained based on an assumed model for the autocorrelation function of the transform coefficients in the image sequence. This three-dimensional autocorrelation function, $R(j_1,j_2,k_1,k_2,\ell_1,\ell_2)$, may be modeled as

$$R(j_1,j_2,k_1,k_2,\ell_1,\ell_2) = \exp(-\alpha|j_1 - j_2| - \beta|k_1 - k_2| - \gamma|\ell_1 - \ell_2|) \quad . \tag{3.8}$$

Under the assumption of spatial and temporal invariance to translation, the autocorrelation function is wide sense stationary. Thus, equation 3.8 may be rewritten as

$$R(\Delta j,\Delta k,\Delta\ell) = \exp(-\alpha|\Delta j| - \beta|\Delta k| - \gamma|\Delta\ell|) \quad . \tag{3.9}$$

The autocorrelation function of equation 3.9 is separable in the two spatial directions as well as in the temporal direction.

If the original image sequence is considered as consisting of samples from a Markov process, then horizontal, vertical, and temporal correlation matrices may be modeled as

$$\left[C_{f_j}\right] = \sigma_j^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{J-1} \\ \rho & 1 & \rho & \cdots & \rho^{J-2} \\ \rho^2 & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \rho^{J-1} & \cdots & & \cdots & 1 \end{bmatrix}, \tag{3.10a}$$

$$\left[C_{f_k}\right] = \sigma_k^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{K-1} \\ \rho & 1 & \rho & \cdots & \rho^{K-2} \\ \rho^2 & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \rho^{K-1} & \cdots & & \cdots & 1 \end{bmatrix}, \tag{3.10b}$$

and

$$\left[C_{f_\ell}\right] = \sigma_\ell^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{L-1} \\ \rho & 1 & \rho & \cdots & \rho^{L-2} \\ \rho^2 & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \rho^{L-1} & \cdots & & \cdots & 1 \end{bmatrix}, \tag{3.10c}$$

where σ_j^2, σ_k^2, and σ_ℓ^2 represent pixel intensity variances in, respectively, the horizontal, vertical, and temporal directions.

By application of the separable matrix form of the discrete cosine transform, $[\mathcal{C}]$, the corresponding transform domain horizontal, vertical, and temporal correlation matrices are

$$\left[C_{F_u}\right] = [\mathcal{C}] \left[C_{f_j}\right] [\mathcal{C}]^t , \tag{3.11a}$$

$$\left[C_{F_v}\right] = [\mathcal{C}] \left[C_{f_k}\right] [\mathcal{C}]^t , \tag{3.11b}$$

and

$$\left[C_{F_w}\right] = [\mathcal{C}] \left[C_{f_\ell}\right] [\mathcal{C}]^t . \tag{3.11c}$$

Horizontal, vertical, and temporal transform domain variance vectors, $[V_F]$, can be defined as consisting of the transform domain variance values which lie along the main diagonals of the correlation matrices of equation 3.11, i.e.,

$$\left[V_{F_u}\right]^t = \left[C_{F_u}(0, 0), C_{F_u}(1, 1), \ldots, C_{F_u}(J-1, J-1)\right] , \tag{3.12a}$$

$$\left[V_{F_v}\right]^t = \left[C_{F_v}(0, 0), C_{F_v}(1, 1), \ldots, C_{F_v}(K-1, K-1)\right] , \tag{3.12b}$$

and

$$\left[V_{F_w}\right]^t = \left[C_{F_w}(0, 0), C_{F_w}(1, 1), \ldots, C_{F_w}(L-1, L-1)\right] . \tag{3.12c}$$

The variance vectors, $[V_F]$, contain transform variances ordered by decreasing magnitude. Figure 3-4 is a graph of cosine transform variance components of $[V_F]$ versus transform component number for an assumed correlation of 0.95. For comparison purposes, variance vector components for the Karhunen-Loeve, Slant, Fourier, Hadamard, and Haar transforms are also included [26]. From this figure, it can be observed that the cosine transform variances lie extremely close to the Karhunen-Loeve variance function which, for the assumed Markov model, provides the optimal image energy compression in the transform domain.

In order to perform a statistical evaluation for the performance of the assumed model of the three-dimensional cosine transform coder, it is necessary to express the three-dimensional array of transform variances, $\{V_F\}$, in terms of the variance vectors, $\left[V_{F_u}\right]$, $\left[V_{F_v}\right]$, and $\left[V_{F_w}\right]$. This expression is

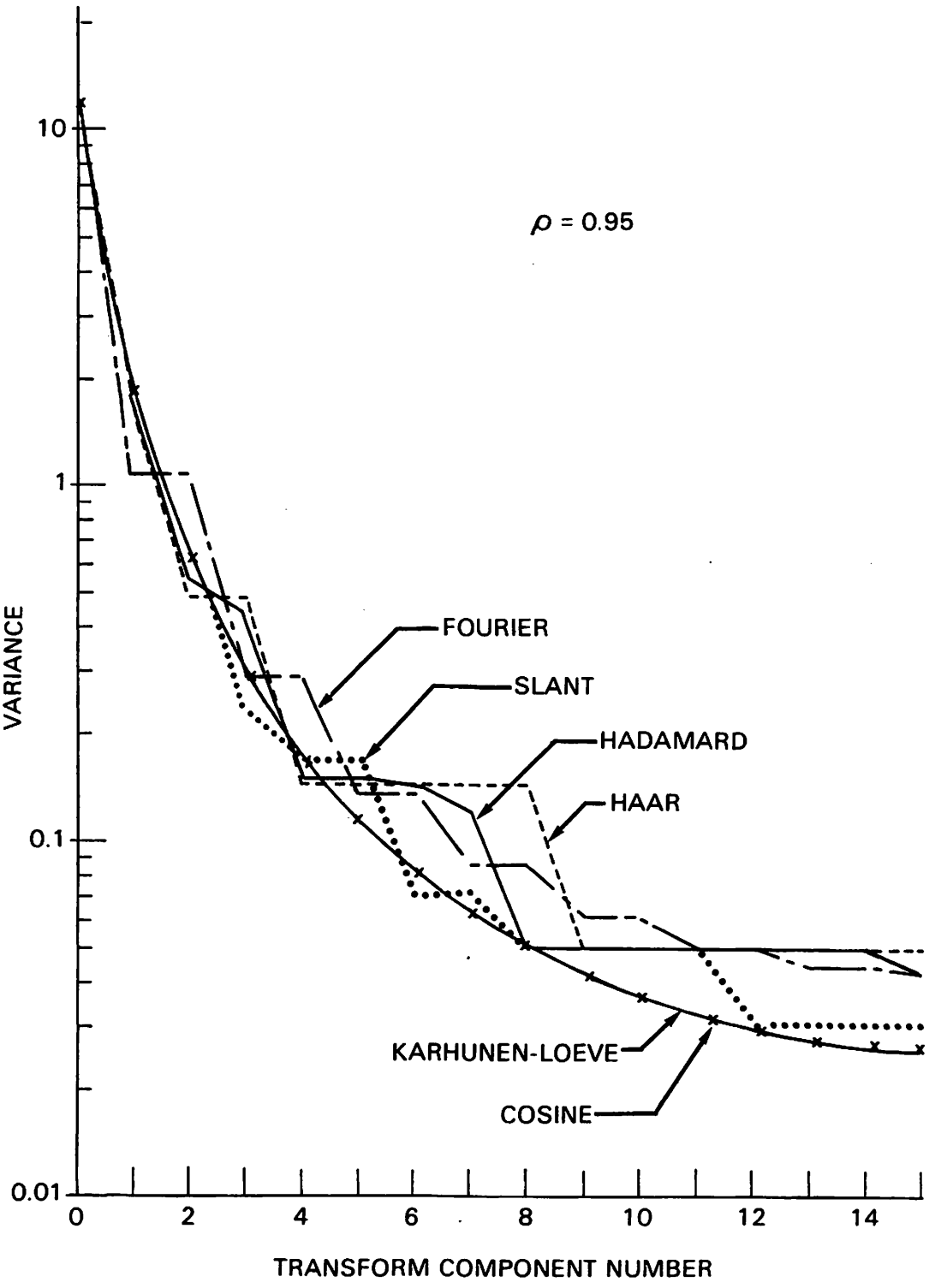


Figure 3-4. Comparison of Transform Domain Variance Functions

$$\{V_F\} = \{v_{F_w}(0) \cdot [V_F], v_{F_w}(1) \cdot [V_F], \ldots, v_{F_w}(L-1)\,[V_F]\} \quad , \tag{3.13}$$

where the two-dimensional variance matrix $[V_F]$ is computed as

$$[V_F] = \left[V_{F_u}\right] \left[V_{F_v}\right]^t \quad , \tag{3.14}$$

and the scalars $v_{F_w}(\,)$ are the elements of variance vector $\left[V_{F_w}\right]$.

Zonal Sampling of Transform Coefficients

The first coefficient selection technique considered is zonal sampling. In this method, a sample bandwidth reduction is achieved by defining a zonal partition of the transform space. Coefficients within the zone define the subset of transform coefficients selected for coding and transmission. Selection of a subset of transform coefficients results in a loss of image fidelity at the receiver since not all coefficients are available for image reconstruction.

The zonal sampling coefficient selection process can be conveniently represented by a three-dimensional binary selection operator in the transform domain S(u,v,w). The form for the reconstructed image $\hat{f}(j,k,\ell)$ is

$$\hat{f}(j,k,\ell) = \sum_{u=0}^{J-1} \sum_{v=0}^{K-1} \sum_{w=0}^{L-1} F(u,v,w)\, S(u,v,w)\, \phi^{-1}(j,k,\ell,u,v,w) \tag{3.15}$$

where S(u,v,w) is unity for values of u,v,w corresponding to transform coefficients selected for coding and transmission and zero, otherwise.

Figure 3-5 conceptually illustrates the structure of the resulting three-dimensional array of selected transform coefficients. In this figure, the single dc transform coefficient is shown in the upper left hand front corner with the remaining solid regions indicating higher order selected coefficients.

The mean square sampling error for a sequence of length L frames is

$$\mathcal{E}_s = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} \sum_{\ell=0}^{L-1} E\left\{ [f(j,k,\ell) - \hat{f}(j,k,\ell)]^2 \right\} \tag{3.16}$$

By substitution of equations 3.3 and 3.15 into equation 3.16, the mean square sampling error becomes

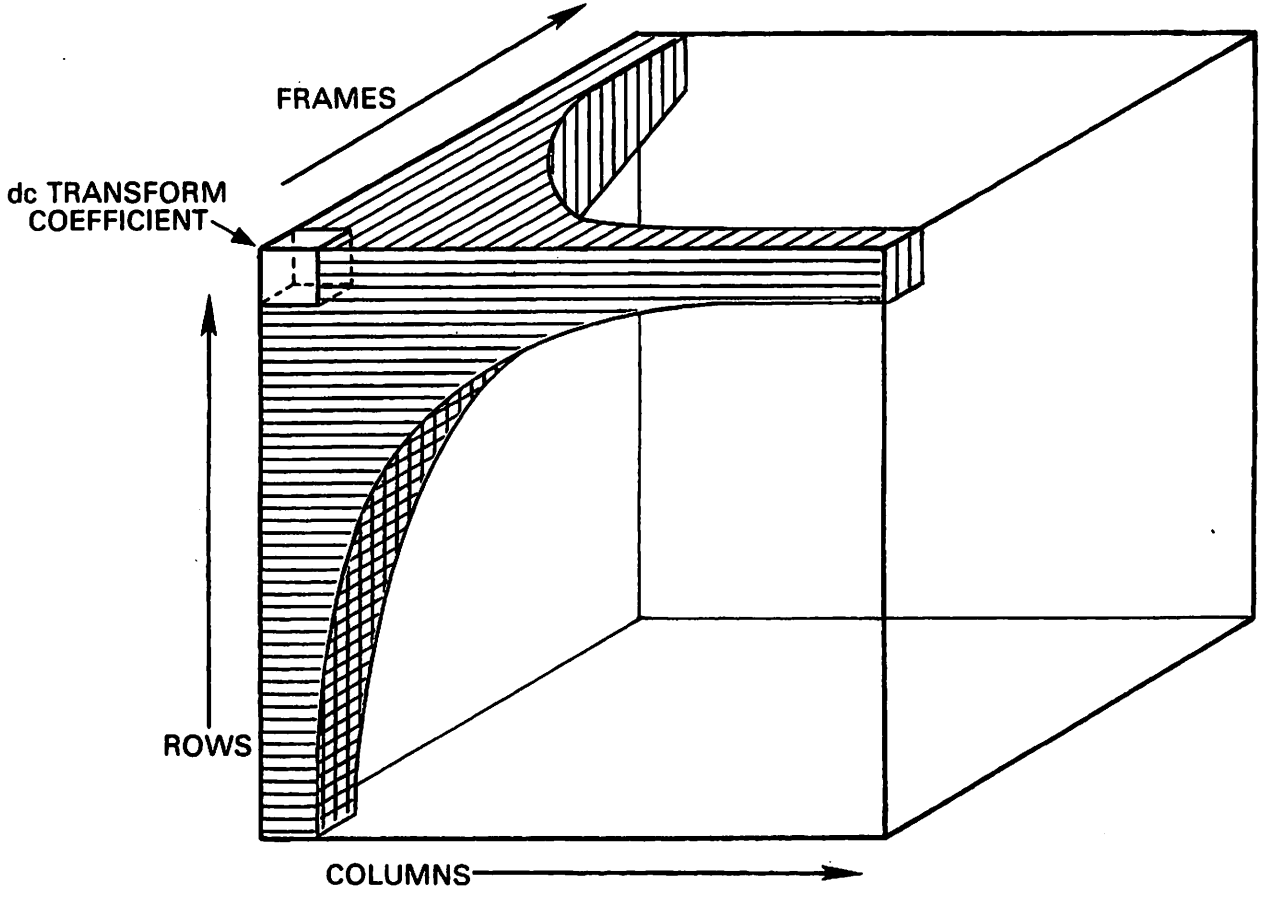


Figure 3-5. Three-Dimensional Array of Selected Transform Coefficients

$$\mathcal{E}_S = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} \sum_{\ell=0}^{L-1} E\left\{ \left[\sum_{u=0}^{J-1} \sum_{v=0}^{K-1} \sum_{w=0}^{L-1} F(u,v,w)\, \phi^{-1}(j,k,\ell,u,v,w) \cdot [1 - S(u,v,w)] \right]^2 \right\} . \quad (3.17)$$

Chen [26] shows that, for transforms having the orthonormality property, the mean square sampling error reduces to the form

$$\mathcal{E}_S = \sum_{u=0}^{J-1} \sum_{v=0}^{K-1} \sum_{w=0}^{L-1} E\left\{ \left[F(u,v,w)\, [1 - S(u,v,w)] \right]^2 \right\} \quad (3.18a)$$

or

$$\mathcal{E}_S = \sum_{\substack{u=0 \\ S(u,v,w)=0}}^{J-1} \sum_{v=0}^{K-1} \sum_{w=0}^{L-1} E\left\{ F^2(u,v,w) \right\} . \quad (3.18b)$$

Thus, the mean square error for coefficient selection by zonal sampling is merely the sum of the second moments of the discarded transform coefficients.

The results of a theoretical performance evaluation for the three-dimensional cosine transform coder using maximum variance zonal sampling with no quantization error are presented in Figure 3-6. In the generation of this figure, the original image was described statistically as a Markov process having the form of equation 3.10. Assumed correlation coefficients were 0.95 for the vertical, horizontal, and temporal directions. A 32:1 sample reduction was used, based on those transform coefficients having the largest variances. The fidelity criterion employed to evaluate the three-dimensional coder performance is the percentage mean square sampling error.

The curve exhibiting the highest percentage mean square sampling error in Figure 3-6 represents a temporal sequence of length one (L = 1). This is the case for which the three-dimensional interframe cosine transform collapses into a two-dimensional intraframe coder. The remaining curves illustrate the effects of increasing image sequence length. In all cases, the effect of additional frame sequence length provided further reductions in the sampling error. In addition, increased spatial subblock size (J X K) also resulted in reduced sampling error.

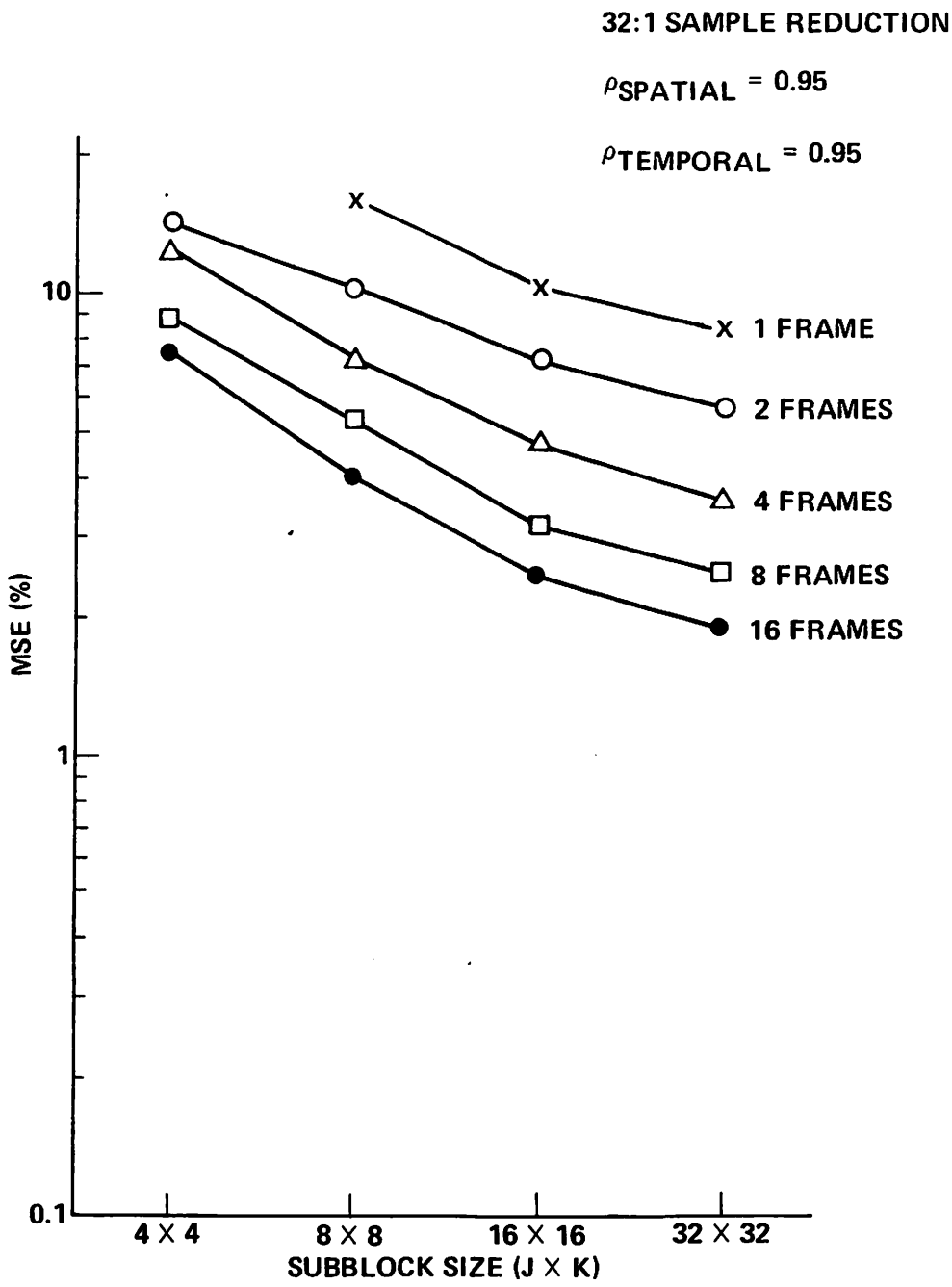


Figure 3-6. Theoretical Performance Evaluation for Three-Dimensional Cosine Transform Coder with Zonal Sampling

Zonal Coding of Transform Coefficients

A more efficient technique for transform coefficient selection and coding is zonal coding. Here, the number of code bits allocated is different for the transform coefficients in each zone. Transform coefficients with the least transform domain energy are assigned zero bits and are not coded for transmission to the receiver.

Optimal mean square error bit assignment algorithms for Gaussian distributed variables have been developed by several researchers. (A discussion of optimal bit assignment procedures is given in Appendix A.) Application of an optimum bit assignment algorithm to the three-dimensional array of transform coefficients results in a new three-dimensional array of bit assignments which defines the number of quantization levels for each transform coefficient.

For transform coefficients having a non-zero bit assignment, an optimal quantization process is performed in accordance with the number of bit coding levels specified in the bit allocation array. (Optimum mean square error quantization techniques are presented in Appendix B.) Computation of the bit allocation array allows the mean square error to be expressed for the zonal coding case. This expression is

$$\mathcal{E}_c = \sum_{u=0}^{J-1} \sum_{v=0}^{K-1} \sum_{w=0}^{L-1} \left[E\left\{F^2(u,v,w)\right\} - \sum_{n=1}^{2^{b}(u,v,w)} R_n^2(u,v,w) \; P\left\{D_n(u,v,w) \leq F(u,v,w) < D_{n+1}(u,v,w)\right\} \right], \qquad (3.19)$$

where D_n, D_{n+1}, and R_n are optimum MSE quantizer decision and reconstruction levels, b is the number of bits allocated to the transform coefficient at location u,v,w, and $P\{\ \}$ is the probability that F(u,v,w) lies between decision levels D_n and D_{n+1} [27].

As shown previously the probability densities for the dc and higher order terms of the three dimensional cosine transform can be approximately modeled as Rayleigh and Laplacian, respectively. For this case, equation 3.19 assumes the form

$$\mathcal{E}_c = 2\alpha^2 - \sum_{n=1}^{2^{b(0,0,0)}} R_n^2(0,0,0)\left[\exp\left(\frac{-D_n^2(0,0,0)}{2\alpha^2}\right) - \exp\left(\frac{-D_{n+1}^2(0,0,0)}{2\alpha^2}\right)\right]$$

$$+ \sum_{\substack{u=0 \\ (u,v,w)\neq(0,0,0)}}^{J-1} \sum_{v=0}^{K-1} \sum_{w=0}^{L-1} \left\{\sigma^2(u,v,w) - 1/2 \sum_{n=1}^{2^{b(u,v,w)}} R_n^2(u,v,w)\right.$$

$$\left. \cdot \left[\exp\left(\frac{-\sqrt{2}\, D_n(u,v,w)}{\sigma(u,v,w)}\right) - \exp\left(\frac{\sqrt{2}\, D_{n+1}(u,v,w)}{\sigma(u,v,w)}\right)\right]\right\} \quad (3.20)$$

Figure 3-7 illustrates a theoretical performance evaluation of the three-dimensional cosine transform coder with zonal coding. The three-dimensional bit allocation array used resulted in a 32:1 bit rate reduction, i.e., from 8 to 0.25 bits/pixel. The original image used to obtain the performance results of Figure 3-7 was the same statistical model of a Markov process previously used in the zonal sampling performance evaluation.

This figure illustrates percentage MSE as a function of spatial subblock size for different temporal sequence lengths. The relative efficiency in terms of theoretical coding performance between zonal sampling and zonal coding is evident by direct comparison of Figures 3-6 and 3-7. This comparison shows that for the same 32:1 effective reduction in bandwidth, zonal coding techniques achieved superior levels of performance. It can also be seen that the results of Figure 3-7 demonstrate a continued improvement in coding performance for increasing subblock size and for sequences of increasing length.

EXPERIMENTAL EVALUATION OF CODER PERFORMANCE

For the results presented in this section, a three-dimensional cosine transform coder was employed using zonal coding with separate Max quantizers for the Rayleigh distributed dc coefficient and the higher order Laplacian distributed coefficients. The same bit assignment was used for each subcube of the image data array. Thus, no attempt was made in the three-dimensional cosine transform coder to adapt to local variations within the data.

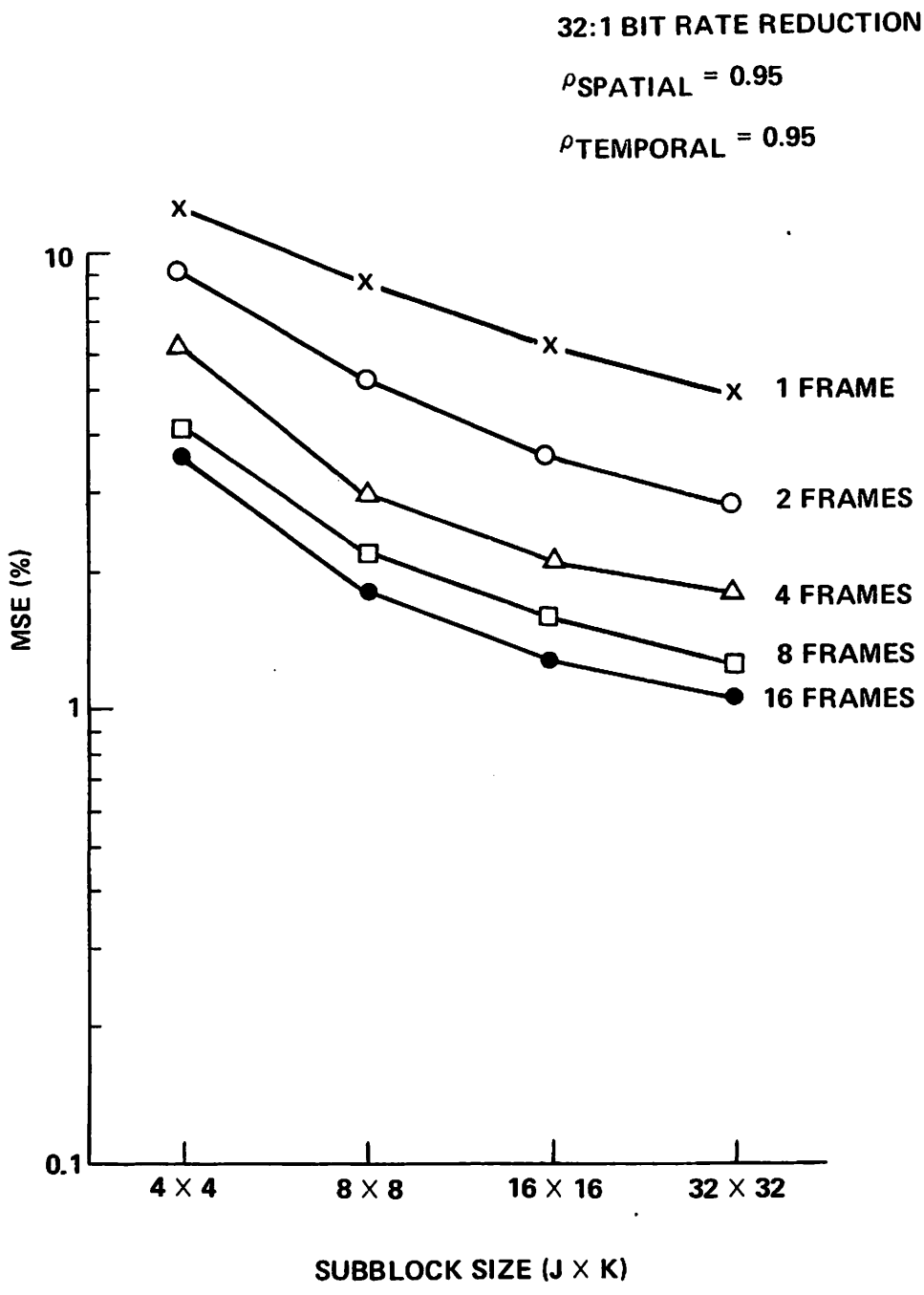


Figure 3-7. Theoretical Performance Evaluation for Three-Dimensional Cosine Transform Coder with Zonal Coding

The simulation evaluates performance of the three-dimensional cosine transform coder as a function of frame number. Figure 3-8 illustrates the results of this experiment for average pixel bit rates of 0.1, 0.25, 0.5 and 1.0 bits/pixel/frame. Photographs corresponding to frame numbers 1, 8, and 16 are shown in, respectively, Figure 3-9 (a), (b), and (c), to illustrate the performance of the three-dimensional cosine transform coder at the 0.25 bits/pixel/frame rate.

RESTORATION OF QUANTIZED SIGNALS

The Max quantizer employed in the previous section is based on the assumption of uncorrelated Laplacian samples. Under this assumption, each higher order coefficient was separately quantized with no consideration given to possible correlations with the distributions of neighboring higher order coefficients.

Recent results by Huhns [28] show that MSE improvements are achieved by optimum nonlinear restoration of quantized samples from multivariate correlated distributions. However, the degree of improvement depends on the residual correlation in the transform domain variables. Huhns indicates that the three-dimensional discrete cosine transform is highly efficient in terms of decorrelating samples in the transform domain. Consequently, although optimum restoration techniques could, in principle, be applied to the three-dimensional cosine coder, this approach is not pursued as only slight improvements in coding performance would be realized.

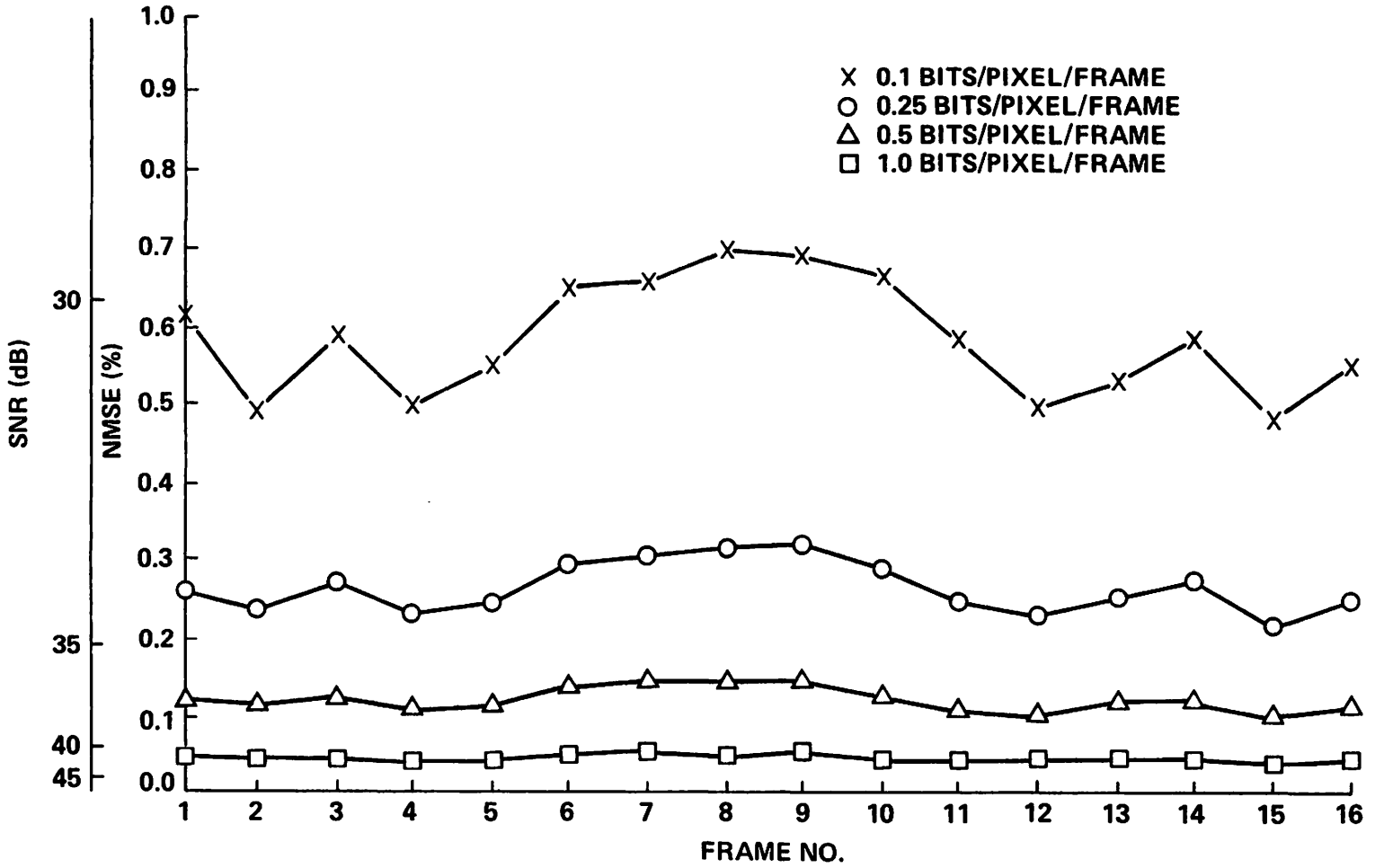


Figure 3-8. Coding Performance as a Function of Frame Number for the Three-Dimensional Cosine Transform Coder

(a) Frame 1

(b) Frame 8

(c) Frame 16

Figure 3-9. Coding Performance of Three-Dimensional Cosine Transform Coder at 0.25 Bits/Pixel/Frame

4. INTERFRAME HYBRID TRANSFORM/PREDICTIVE IMAGE CODING

In the previous chapter, interframe coding techniques using three-dimensional unitary transforms were presented. The performance of these interframe three-dimensional transform coders greatly exceeds that of conventional intraframe coders. However, when implementation considerations are taken into account, three-dimensional transform coders are often undesirable due to excessive storage requirements. In general, for a subcube size of L in the temporal domain, these coders require memory sufficient to store L frames of transform coefficient values. For example, the experimental results reported for the three-dimensional cosine transform coder used image sequences of length 16 and required storage of 16 frames of transform coefficients.

To alleviate problems associated with three-dimensional transform coders, a new class of interframe hybrid coders has been proposed [7], [29]. These interframe hybrid coders employ two-dimensional unitary transforms within each frame coupled with first order linear predictive coding between frames. Thus, interframe hybrid coders utilize both spatial and temporal correlations while, at the same time greatly reducing memory storage and computational requirements.

SYSTEM DEFINITION

A block diagram of the basic interframe hybrid coder is shown in Figure 4-1. In this coding system, a two-dimensional unitary transform is performed on each partition or spatial subblock of the image data. One of the bank of parallel DPCM linear predictive coders is then applied to each set of transform coefficients in the temporal direction. The resulting sequences of transform coefficient differences are quantized and coded for transmission. Image reconstruction occurs at the receiver where the transform coefficient differences are

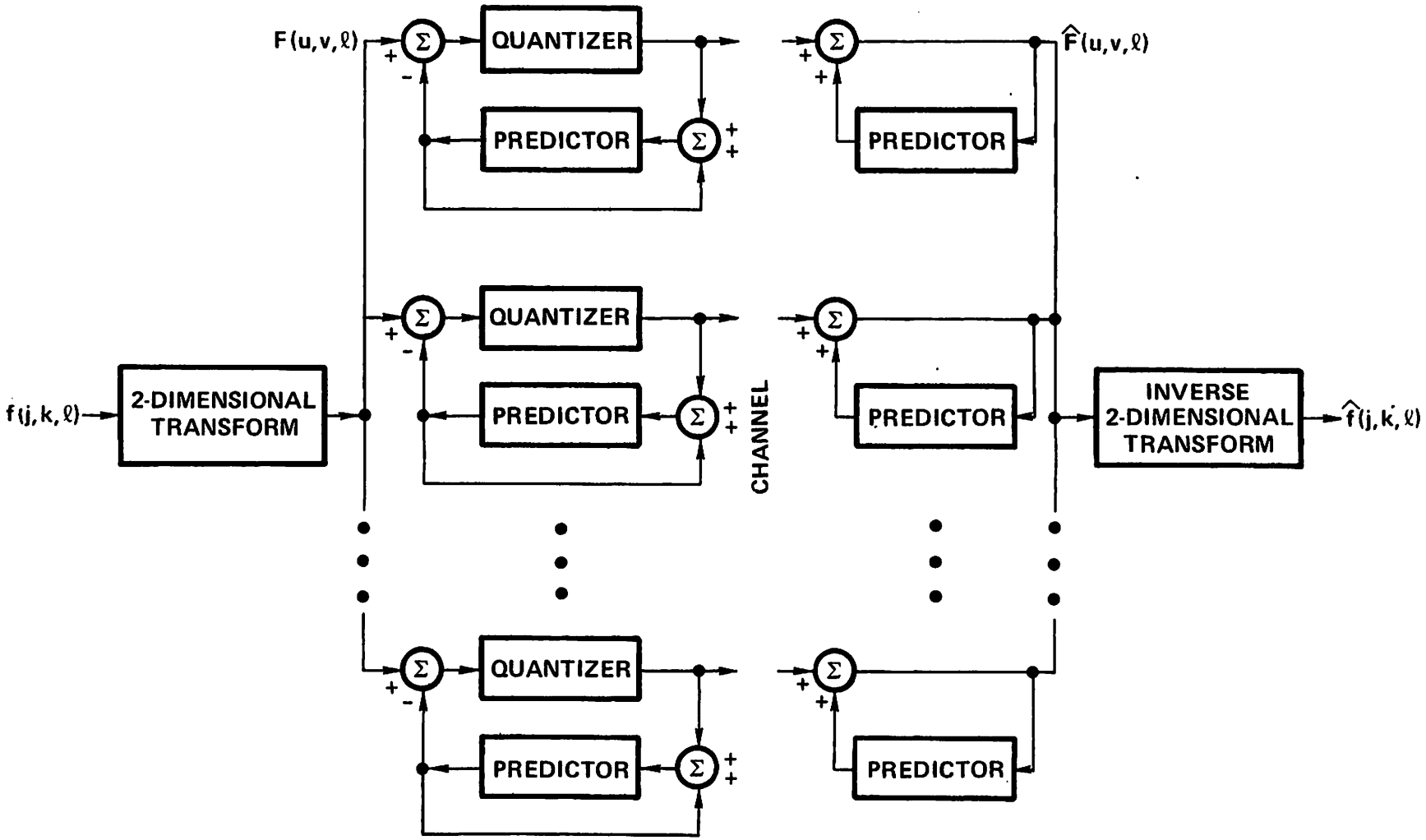


Figure 4-1. Hybrid (Two-Dimensional Transform)/DPCM Coder

decoded and a replica of each transmitted image is reconstructed using the appropriate two-dimensional inverse transformation.

Advantages of hybrid interframe coders include the requirement for only a single frame of storage for first order linear prediction and fewer computations than three-dimensional transform coders. An additional benefit of the interframe hybrid coder's use of DPCM rather than transform coding in the temporal direction is a less complicated hardware implementation.

The basic implementations investigated for the interframe hybrid coder were two-dimensional discrete cosine transformations and two-dimensional discrete Fourier transforms in the spatial domain with DPCM temporal coding. For notational convenience, the interframe hybrid coders employing two-dimensional cosine transforms and two-dimensional Fourier transforms are denoted as, respectively, CCD and FFD.

Since the discrete Fourier transform is expressible in equivalent real and imaginary or amplitude and phase components, two versions of the hybrid FFD coder are possible. The cosine transform, however, is real and only a single formulation for the hybrid CCD coder exists.

A general expression for the spatial domain processing of the interframe hybrid coder can be obtained by letting $f(j,k,\ell)$ denote a three-dimensional array of amplitude values for a digital image sequence of length L frames. Also, let $F(u,v,\ell)$ be the three-dimensional array obtained by taking the two-dimensional transform in the j,k domain for each frame.

Mathematically, this transformation pair can be described in a general form by

$$F(u,v,\ell) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} f(j, k, \ell)\, \phi(u, v, j, k) \tag{4.1}$$

and

$$f(j,k,\ell) = \sum_{u=0}^{J-1} \sum_{v=0}^{K-1} F(u, v, \ell)\, \phi^{-1}(u, v, j, k) \ , \tag{4.2}$$

where j and k are spatial coordinates within a subblock of size J X K, u and v are transform domain coordinates, ℓ is the temporal coordinate indicating frame number, and $\phi(u,v,j,k)$ is a set of two-dimensional orthogonal basis matrices.

In the case of the two-dimensional discrete Fourier transform, equations 4.1 and 4.2 assume the form

$$F(u,v,\ell) = \frac{1}{JK} \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} f(j,k,\ell) \exp\left[-2\pi i \left(\frac{uj}{J} + \frac{vk}{K}\right)\right] \tag{4.3}$$

and

$$f(j,k,\ell) = \sum_{u=0}^{J-1} \sum_{v=0}^{K-1} F(u,v,\ell) \exp\left[2\pi i \left(\frac{uj}{J} + \frac{vk}{K}\right)\right] , \tag{4.4}$$

where $F(u,v,\ell)$ is the array of two-dimensional Fourier transform coefficients for the ℓth frame. Since the two-dimensional Fourier transform is separable, spatial domain transformations are accomplished by sequentially applying a one-dimensional transform to each row and then reapplying the transform to each column. For the discrete Fourier transform, the form for the one-dimensional forward transform is given in equation 2.23.

For image processing applications, $f(j,k)$ is a positive real function representing brightness of the spatial sample within a frame. The two-dimensional Fourier transform of a real-valued function has the property of conjugate symmetry, i.e.,

$$F^*(u,v) = F(J-u, K-v) , \tag{4.5}$$

where $u,v, = 1,2,\ldots,\frac{J}{2} - 1$. The Fourier transform consists of $2J^2$ components, i.e., the real and imaginary or magnitude and phase components of each spatial frequency. However, as a result of the conjugate symmetry property, only J^2 components are required to completely define the Fourier transform [30].

In the Fourier transform, a shift in the spatial domain variables results in a multiplication of the Fourier transform of the unshifted image by a phase factor. If the input image $f(j,k,\ell)$ is shifted by the amount j_0 in the j-direction and k_0 in the k-direction between frames ℓ_1 and ℓ_2, then the phase corrected Fourier transform of the image in frame ℓ_2 is

$$F_{PC}(u,v,\ell_2) = F(u,v,\ell_2) \exp\left[2\pi i \left(\frac{uj_0}{J} + \frac{vk_0}{K}\right)\right] . \tag{4.6}$$

This shifting property is potentially useful for detecting and compensating for effects of motion between frames since many types of motion, such as panned motion, produce significant changes in phase components and small changes in amplitude components. Thus, compensation for camera motion may be implemented directly in the array of phase components by application of appropriate phase correction factors.

In a like manner, equations 4.1 and 4.2 have the following form for the discrete two-dimensional cosine transform:

$$F(u,v,\ell) = \frac{1}{JK} \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} f(j,k,\ell) \cdot \left\{ \cos\frac{\pi}{2}\left[\frac{(2j+1)\,u}{J} + \frac{(2k+1)\,v}{K}\right] + \cos\frac{\pi}{2}\left[\frac{(2j+1)\,u}{J} - \frac{(2k+1)\,v}{K}\right]\right\} \quad (4.7)$$

and

$$f(j,k,\ell) = \sum_{u=0}^{J-1} \sum_{v=0}^{K-1} F(u,v,\ell) \cdot \left\{ \cos\frac{\pi}{2}\left[\frac{(2j+1)\,u}{J} + \frac{(2k+1)\,v}{K}\right] + \cos\frac{\pi}{2}\left[\frac{(2j+1)\,u}{J} - \frac{(2k+1)\,v}{K}\right]\right\} . \quad (4.8)$$

The two-dimensional cosine transform is also separable. The basic one-dimensional cosine transform that must be performed on individual image rows and columns is given by equation 2.16.

Once the transform domain coefficients have been computed, the interframe hybrid coder uses first-order linear predictive DPCM techniques to code differences between temporally adjacent transform coefficients. The defining equation for the first-order linear minimum mean square error estimate of input signal $f(j,k,\ell)$, as given in equation 2.30, is

$$\hat{f}(j,k,\ell) = a_1 f(j,k,\ell-1) + a_0 \quad (4.9)$$

Under the assumption of $f(j,k,\ell)$ being a zero mean sequence of transform coefficients, the quantities a_0 and a_1 used to form the estimate $\hat{f}(j,k,\ell)$ are given by equations 2.35 and 2.36. These equations define the form for the predictors used in each of the parallel DPCM coders shown in Figure 4-1.

TRANSFORM COEFFICIENT DIFFERENCE PROBABILITY DENSITIES

In order to properly encode temporal sequences of transform coefficient differences for each image subblock, the underlying structure of their probability density functions must be determined. Such a determination will allow selection of quantizer configurations which are optimally matched to the statistics of the difference signal.

Previous research based on frame difference intensity values has shown that, for a large class of natural images, the underlying density is double-sided exponential or Laplacian. However, the question of the statistics of interframe transform coefficient differences has received little attention. To determine these densities, the 16 frame "head and shoulders" data base was transformed using two-dimensional cosine transforms. A subblock size of 8 X 8 was employed to generate 1024 subblocks of cosine transform coefficients in each frame.

The proposed model for the transform coefficient differences is that of a Laplacian density. The scalar form for this density is given by equation 3.7. Figures 4-2 and 4-3 show normalized histograms of the dc and a typical lower order cosine transform coefficient difference sequence. For these figures, a histogram generation technique was used which tabulated the composite transform coefficient differences for all of the spatial subblocks. This approach was employed because of the small number (15) of coefficient differences available for the temporal sequences within an individual subblock. The plots of Figures 4-2 and 4-3 validate the assumed scalar Laplacian form for the density of cosine transform coefficient differences in the temporal direction.

An investigation of multidimensional Laplacian densities was not undertaken as the two-dimensional cosine transform is known to be extremely efficient in terms of decorrelating transform domain coefficients. Thus, little residual correlation would be anticipated between temporal sequences of cosine transform coefficient differences.

BANDWIDTH REDUCTION

The primary goal in the design of image coders is to reduce channel bandwidth requirements. The basic rationale for the inclusion of transform coding in interframe hybrid coders is that multidimensional transforms of an image exhibit energy distributions which readily lend themselves to image coding. This property stems from correlations existing between pixels and results in the compression of transform domain energy into a relatively small number of transform domain coefficients.

As an illustration of this property in two-dimensions, separable one-dimensional cosine transforms were applied to the 16th frame of the "head and shoulders" data base. A

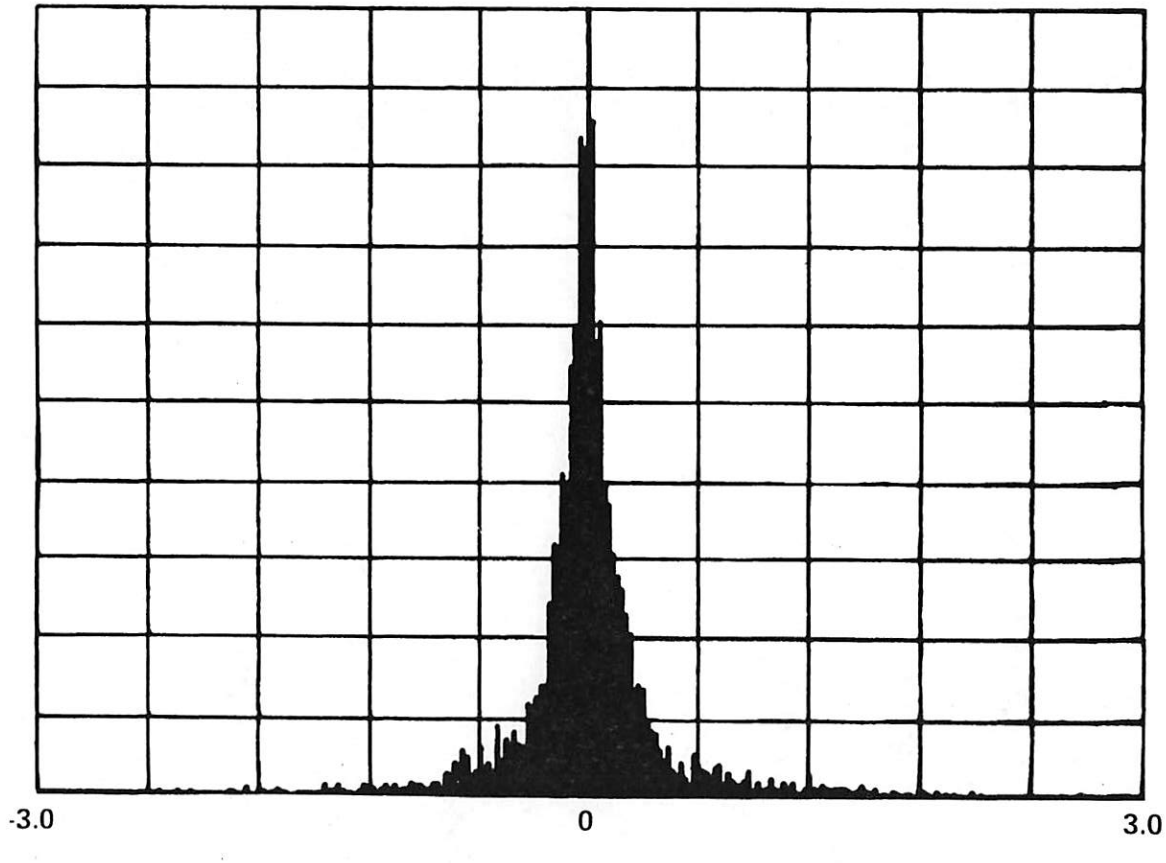


Figure 4-2. Histogram of Laplacian Density for DC Cosine Transform Coefficient Differences

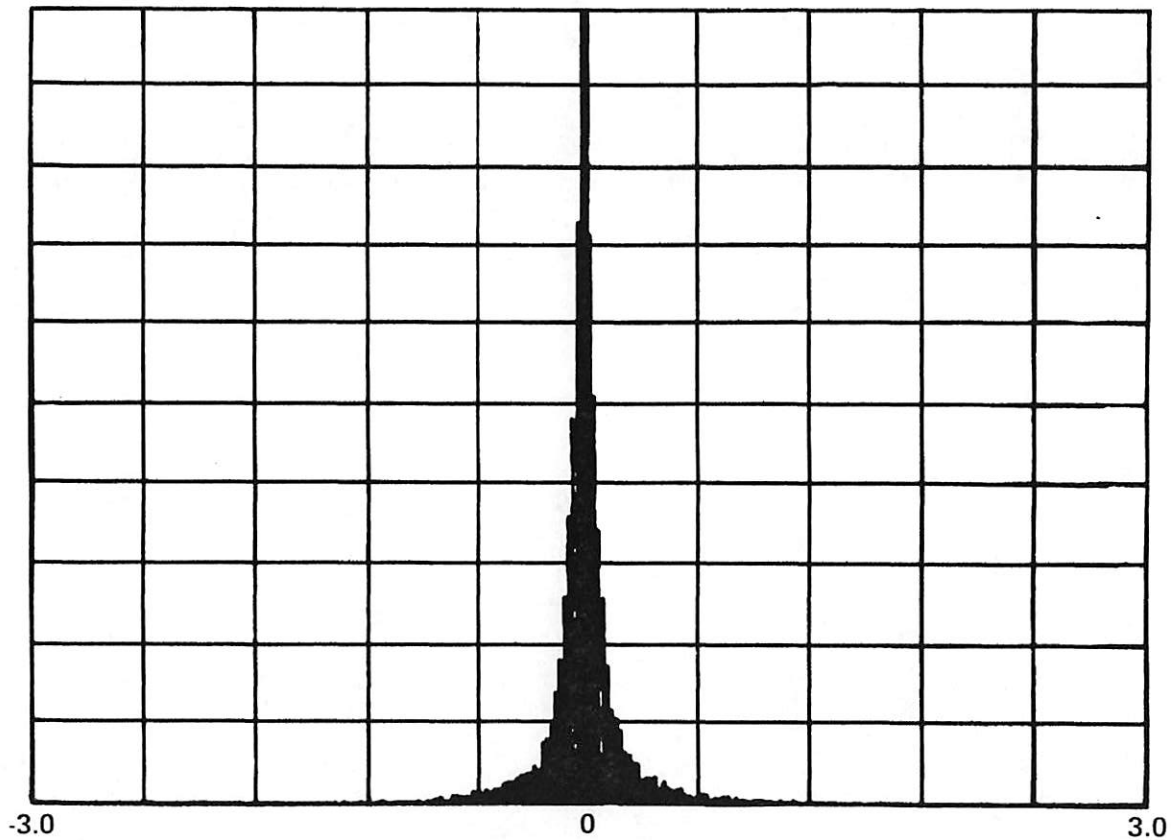


Figure 4-3. Histogram of Laplacian Density for Typical Low-Order Cosine Transform Coefficient Differences

subblock size of 64 X 64 pixels was used to partition the 256 X 256 image into 16 subimages. The results are illustrated in Figure 4-4. In this figure, the general pattern of energy about the dc and low frequency terms in the upper left hand corner of each subblock is clearly visible although there are noticeable variations between subblocks. The logarithm of the magnitude of the transform coefficients was displayed in this figure to compensate for the large dynamic range of the transform domain coefficients.

The hybrid interframe coder employs DPCM techniques in the temporal direction to code differences between temporally adjacent transform coefficients. The coding strategy employed is based on the matrix of transform coefficient difference variances. In the following sections, a model of the transform coefficient difference variance matrix for the interframe hybrid coder is developed. The performance of the interframe hybrid coder for modeled images is analyzed using a zonal coding strategy.

Transform Coefficient Difference Variance Matrix

Use of zonal coding schemes requires knowledge of the two-dimensional distribution of the transform coefficient difference variances. To model the distribution of the transform coefficient difference variances, it is necessary to formulate an expression for the transform coefficient difference variance matrix, $[V_D]$. The variance matrix represents the variance for each temporal sequence of transform coefficient differences within an image or image subblock.

An expression for variance matrix $[V_D]$ can be derived by modeling the transform coefficient differences which result from DPCM coding in the temporal direction. For the class of interframe hybrid coders, the two-dimensional transform domain representation for two temporally adjacent images or image subblocks is computed. Under the assumption of a wide sense stationary process, the two-dimensional autocorrelation function may be expressed as

$$R(\Delta j, \Delta k) = \exp(-\alpha |\Delta j| - \beta |\Delta k|) \ . \qquad (4.10)$$

Assuming that the images may be treated as samples from a Markov process, the respective horizontal and vertical correlation matrices, $\left[C_{f_j}\right]$ and $\left[C_{f_k}\right]$, may be modeled by equations 3.10a and 3.10b. For the case of the discrete cosine transform, the corresponding horizontal and vertical transform domain correlation matrices, $\left[C_{F_u}\right]$ and

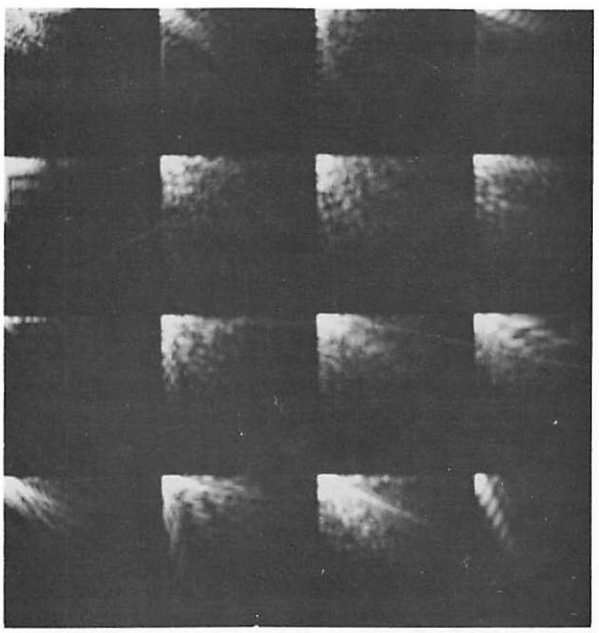


Figure 4-4. Two-Dimensional Cosine Transform Domain Representation of Frame No. 16 of "Head and Shoulders" Data Base

$[C_{F_v}]$, are given by equations 3.11a and 3.11b. In a similar fashion, the transform domain horizontal and vertical variance vectors, $\left[V_{F_u}\right]$ and $\left[V_{F_v}\right]$, are computed by, respectively, equations 3.12a and 3.12b. Thus, for the interframe hybrid coder, the two-dimensional matrix of transform coefficient variances, $[V_F]$, is easily computed as

$$[V_F] = \left[V_{F_u}\right] \left[V_{F_v}\right]^t . \tag{4.11}$$

The relationship between the matrix of transform coefficient variances $[V_F]$ and the matrix of transform coefficient temporal difference variances $[V_D]$, for first order DPCM predictive coding, is given by Habibi [20]. This relationship is

$$[V_D] = [V_F] \{1 - \exp(-2\gamma)\} , \tag{4.12}$$

where $\exp(-\gamma)$ is the temporal correlation of the transform coefficients. Equation 4.12 allows the matrix $[V_D]$ to be modeled for interframe hybrid coders using unitary transforms and first-order DPCM predictive coding with specified horizontal, vertical, and temporal correlations.

Zonal Coding of Transform Coefficient Differences

In zonal coding, the number of code bits allocated for quantization is different for each zone of the transform coefficient difference variance matrix, $[V_D]$. Those transform coefficient difference variances having zero allocated bits define the subset of transform coefficient differences which are discarded. The bit assignment algorithm used is known to be optimal in a mean square error sense for Gaussian distributed variables. However, good experimental results have been obtained for sequences having underlying Laplacian densities. This is due in part to the similar nature of the Gaussian and Laplacian densities, especially at points far from the mean.

Figure 4-5 illustrates the zonal coding bit assignment for a 16 X 16 matrix of cosine transform coefficient difference variances computed for the Markov model of equation 4.10, with horizontal and vertical correlations of 0.95. With this bit assignment, a bit rate reduction of 32:1 is achieved. The measure of bit rate reduction is the ratio of the average number of bits allocated for image coding compared to an 8 bits/pixel representation of the original image. Thus, a 32:1 bit rate reduction implies that, on the average, 0.25 bits/pixel/frame are available for coding each temporal sequence of transform coefficient differences.

5	4	3	2	2	2	2	1	1	1	1	1	1	1	1	1
4	2	2	1	1	0	0	0	0	0	0	0	0	0	0	0
3	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 4-5. Zonal Coding Bit Assignment Array for Markov Model of Cosine Transform Coefficient Difference Sequences with a 32:1 Bit Rate Reduction

Equation 3.19 presented the general form for mean square error due to zonal coding. For the interframe hybrid coder, this expression is defined in terms of quantizer decision and reconstruction levels, bit allocation and the probability distribution of the transform coefficient differences.

The form for the probability density of the dc and higher order cosine transform coefficient differences was experimentally determined to be Laplacian. Thus, the mean square error expression for hybrid interframe coders has the form

$$\mathcal{E}_c = \sum_{u=0}^{J-1} \sum_{v=0}^{K-1} \left\{ \sigma_D^2(u,v) - \frac{1}{2} \sum_{n=1}^{2^{b(u,v)}} R_n^2(u,v) \cdot \left[\exp\left(\frac{-\sqrt{2}\, D_n(u,v)}{\sigma_D(u,v)} \right) - \exp\left(\frac{-\sqrt{2}\, D_{n+1}(u,v)}{\sigma_D(u,v)} \right) \right] \right\} \quad ,(4.13)$$

where σ_D^2 is the variance of the transform coefficient differences at location (u,v), D_n, D_{n+1}, and R_n are optimal Laplacian quantizer decision and reconstruction levels and b is the number of assigned bits.

The theoretical performance of the interframe hybrid coder was evaluated using separable cosine transforms and first order DPCM predictive coding. The original image was modeled as a Markov process having the form of equation 4.10 with horizontal, vertical and temporal correlations of 0.95. In these simulations, zonal coding with Max quantizer decision and reconstruction levels was used to achieve bit rate reductions of 32:1. The bit assignment was that of Figure 4-5. The results of the hybrid interframe coder theoretical performance evaluation are presented in Figure 4-6.

OPERATIONAL MODES

There are several possible modes of operation for interframe hybrid coders. Definition of an operational mode includes specification of initial conditions assumed for the transmitter and receiver and the extent to which the coder is spatially and temporally adaptive.

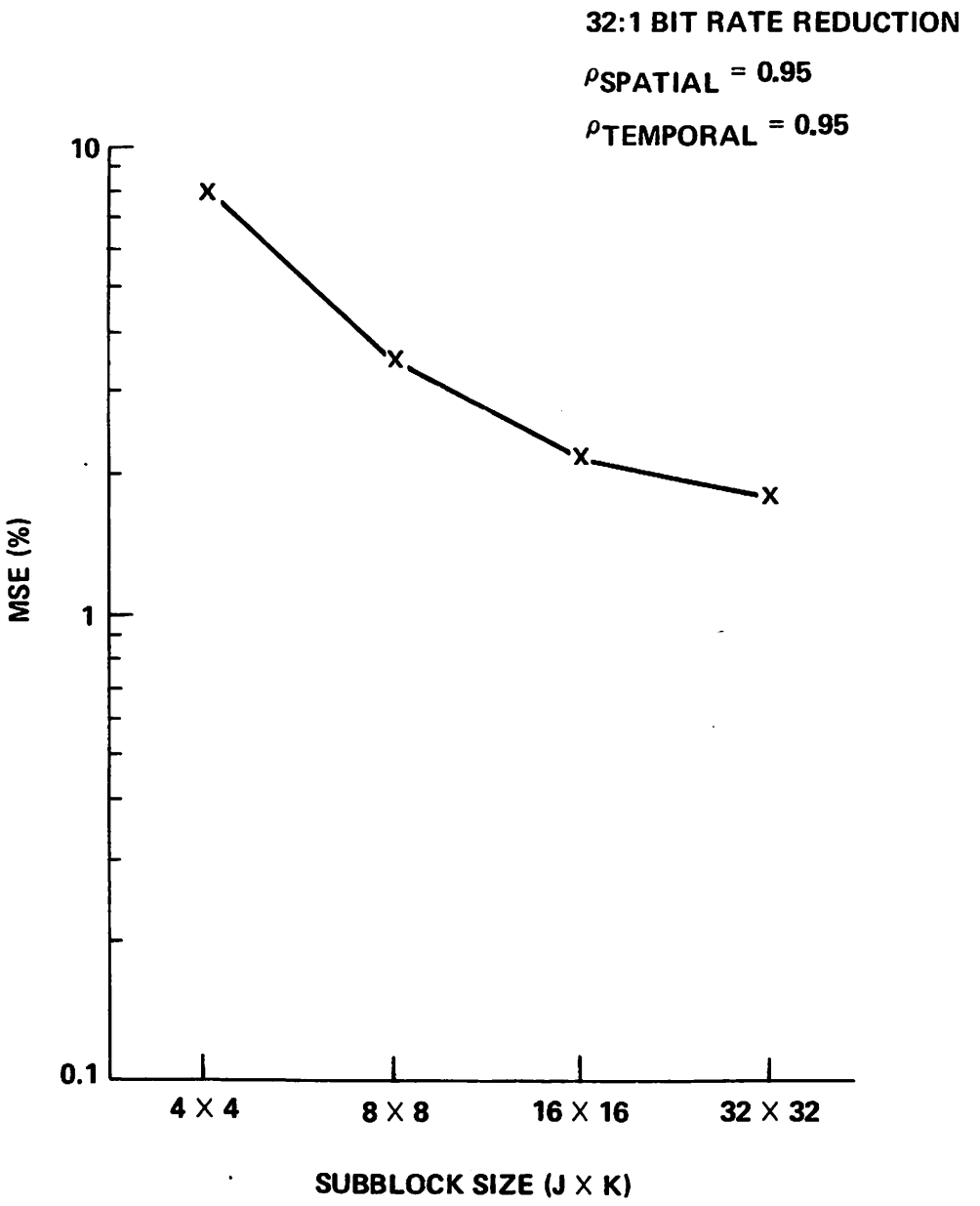


Figure 4-6. Theoretical Performance Evaluation for Hybrid CCD Coder with Zonal Coding.

Initial Conditions

Selection of initial conditions includes specification of the assumed knowledge of transform coefficient statistics at both the transmitter and receiver and the availability of the first frame at the receiver. Three proposed sets of initial conditions are summarized in Table 4-1.

In the no *a priori* information available case of Table 4-1, the first frame is not assumed to be available at the receiver. Consequently, the initial difference signal to be quantized, coded, and transmitted by the DPCM coder is the transform domain representation of the first frame. The remaining statistical parameters are also not assumed to be available initially. Values for these parameters must be continuously refined with the coding of subsequent frames. The resulting mode of operation is an extreme case requiring several frames for the coder to settle. This situation is of little practical value as at least simple initial estimates of the first frame and the necessary statistical parameters can normally be obtained.

The second set of initial conditions is intended for operation under assumptions of limited or imperfect initial knowledge of the first frame and statistics of the frame sequence. For example, the interframe hybrid coder might employ a current mean estimate of the first frame to be coded. Alternately, an intraframe coded version of the first frame could be used if available. In a similar fashion, the statistical parameters could be based on statistical models or on mean values. This set of initial conditions is perhaps the most realistic from the viewpoint of system implementation. Actual specification of the initial conditions will, of course, depend on hardware design configurations for the hybrid coder and its intended operational environment.

The maximum information set of initial conditions is primarily intended for experimental evaluation of the interframe hybrid coder's performance. Here, it is assumed that the first frame is initially available at the receiver. In addition, frame sequence statistical parameters are assumed to be available prior to coding. These statistical measures are the mean value and correlation for each temporal sequence of transform coefficients, i.e.,

$$\bar{F}(u,v) = \frac{1}{L} \sum_{\ell=0}^{L-1} F(u, v, \ell) \tag{4.14}$$

Table 4-1. Initial Conditions for the Interframe Hybrid Coder

INITIAL CONDITIONS	FIRST FRAME	FRAME SEQUENCE STATISTICS
NO *A PRIORI* INFORMATION	NOT AVAILABLE	NOT AVAILABLE
LIMITED INFORMATION	MEAN VALUE OR INTRAFRAME CODED RECONSTRUCTION	MEAN VALUES OR STATISTICAL MODELS
MAXIMUM INFORMATION	AVAILABLE	COMPUTED ON FRAME SEQUENCE

and

$$\rho(u,v) = \frac{\frac{1}{L-1}\sum_{\ell=1}^{L-1} F(u,v,\ell)\,F(u,v,\ell-1)}{\frac{1}{L}\sum_{\ell=0}^{L-1} F^2(u,v,\ell)}, \tag{4.15}$$

and the variance of the estimated temporal transform coefficient differences

$$\sigma_D^2(u,v) = \frac{1}{L-1}\sum_{\ell=1}^{L-1} \hat{F}_D^2(u,v,\ell) - \left(\frac{1}{L-1}\sum_{\ell=1}^{L-1} \hat{F}_D(u,v,\ell)\right)^2. \tag{4.16}$$

In equation 4.16, the sequence of L-1 estimated transform coefficient differences, $\hat{F}_D(u,v,\ell)$, is generated by

$$\hat{F}_D(u,v,\ell) = F(u,v,\ell) - \rho(u,v) \cdot F(u,v,\ell-1) \tag{4.17}$$

for $\ell = 1,2,\ldots,L-1$, where the term $\rho(u,v) \cdot F(u,v,\ell-1)$ represents the effect of DPCM first order linear prediction with perfect quantization.

In the interframe hybrid coder, the transform coefficient temporal correlation is the gain coefficient in the DPCM predictor feedback loop for each sequence of transform coefficients. Also, the mean value provides biasing to achieve a zero-mean input sequence of transform coefficients. The computed transform coefficient difference variances are used to generate optimal bit assignments. Finally, the standard deviation of each sequence of transform coefficient differences is employed as a pre- and post-scaling factor for normalization of the quantizer inputs.

Spatial Adaptation

The implementation of the interframe hybrid coder used for experimental performance evaluations is spatially adaptive in the sense that transform coefficient difference statistics for each spatial subblock are separately estimated. Thus, the resulting two-dimensional bit assignments, predictor feedback loop gain coefficients, and scaling factors are, in general, different for different subblocks. Although the subblock bit allocations are different, the total number of bits available for coding within all subblocks is constrained to be equal. An even higher level of spatial adaptation would be to make the total bits available to each

subblock a variable based on subblock-to-subblock variations in image energy. For the experimental results reported in this chapter, only the simpler form of adaptive interframe coding is employed.

Local adaptation to the measured statistics of image subblocks will normally produce improved coding results when compared with non-adaptive implementations. However, adaptation does result in increased implementation complexity. In a later section of this chapter, experimental results are presented for a simplified non-spatially adaptive version of the interframe hybrid coder.

Temporal Adaptation

Analogous to the case of spatial adaptation, improvements in performance can be achieved if the interframe hybrid coder adapts to temporal signal variations. Although desirable, temporally adaptive coder implementations also result in increased implementation complexity.

Figure 4-7 illustrates the operation of a temporally adaptive interframe hybrid coder implementation. At the start of the frame sequence, a set of initial conditions is specified for the coder. Given these initial conditions, it is anticipated that the coder will operate in a stable manner for a sequence of length M frames. It is assumed here that after M frames, the statistics used by the coder will no longer be representative and an update of statistical parameters at both the transmitter and receiver will be required. Figure 4-7 indicates that current estimates of the statistical parameters are accumulated from the most recent N frames coded where, in general, $N \ll M$. This sequence of periodic statistical parameter updating continues until the total sequence of frames is encoded or the coder is reinitialized.

For efficient operation with temporal adaptation, the statistical parameters transmitted during each update are the transform coefficient mean, correlation, and difference variance for each transmitted sequence. However, the computation of these parameters is costly in terms of equipment complexity. Also, their periodic transmission impacts the effective coder transmission bit rate.

Operation of the interframe hybrid coder without temporal adaptation will result in a reduced level of performance as temporal variations in the image sequence are ignored. To effect a temporally non-adaptive mode of operation, fixed estimates of individual transform

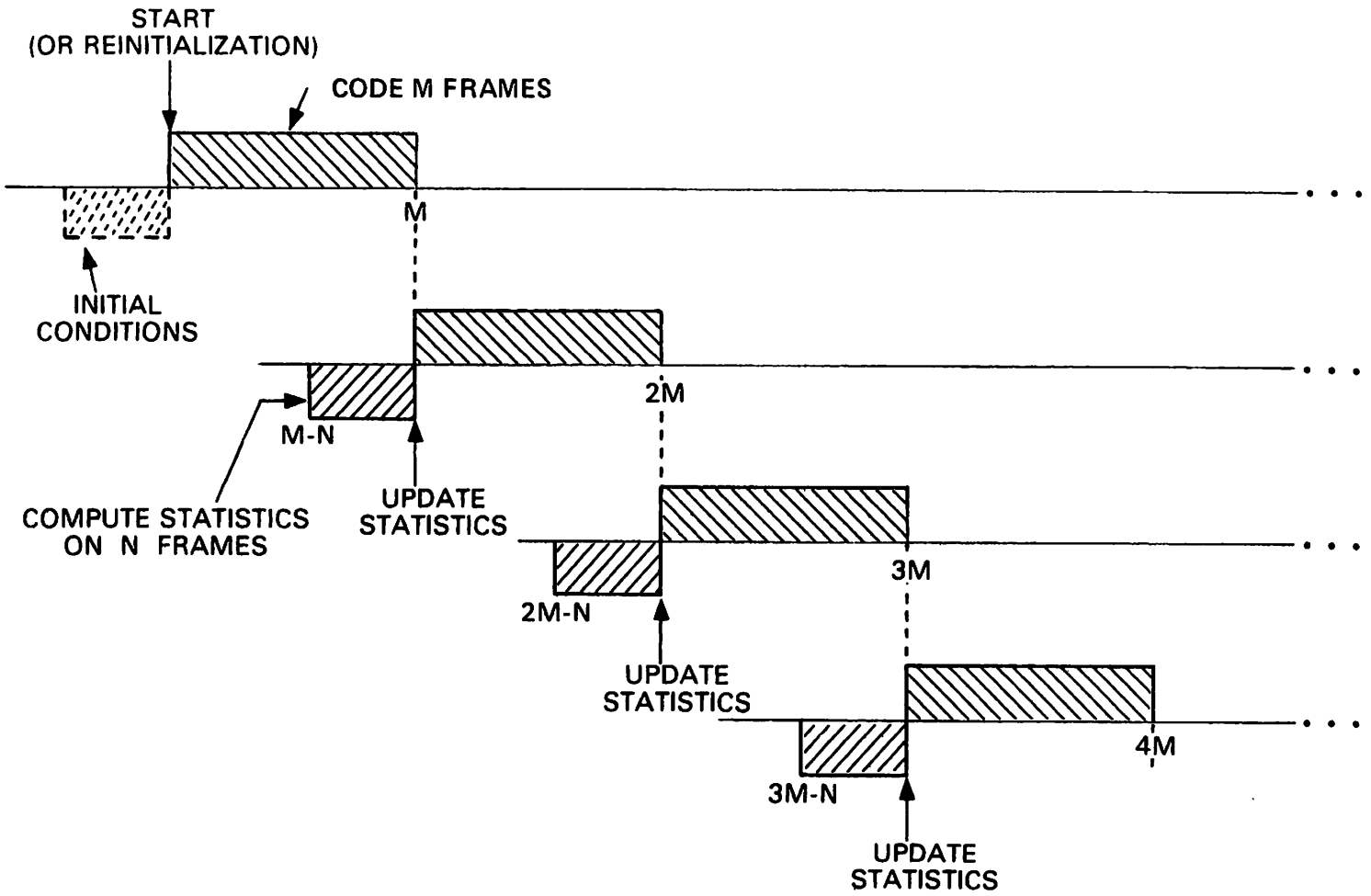


Figure 4-7. Image Coding for Temporally Adaptive Interframe Hybrid Coder

coefficient difference means, correlations and variances are required. For example, based on the Laplacian nature of the transform coefficient difference probability density, a zero mean estimate is reasonable. Also, an estimate for interframe transform coefficient correlation can be made based on knowledge of anticipated image transmission frame rates and camera platform motion. Finally, a non-temporally varying bit assignment or set of bit assignments can be employed.

Reinitialization

When operated in the presence of channel errors, the DPCM transmitter and receiver first order prediction signals will differ. Correction of predictor discrepancies can be accomplished by periodically interjecting corrected signals into the transmitter and receiver prediction loops of each DPCM coder. This situation is illustrated in Figure 4-8 where switches A and B can be set to allow the DPCM predicted signals to be internally generated or externally introduced. As indicated, corrected prediction signals are directly available at the transmitter. However, these signals must be transmitted to the receiver.

For low probabilities of channel error, the rate of occurrence of bit transmission errors is sufficiently small so that frequent coder reinitialization is not required. However, high channel error probabilities can cause the transmitter and receiver predictors to diverge rapidly with corresponding decreases in the fidelity of reconstructed images. Operation of the interframe hybrid coder in a high channel error environment clearly necessitates frequent coder reinitialization.

TOTAL TRANSMISSION BIT RATE

In this section, a measure is defined for the total number of bits which must be transmitted in order for the receiver to reconstruct sequences of coded images. The total transmission bit rate is the sum of the average bit rate per pixel plus overhead pixel bit rates which are decreasing functions of the number of frames transmitted before parameter updating and reinitialization is required. A working knowledge of the total transmission bit rate is important as it determines the true extent of bandwidth reduction achievable by the interframe coding system.

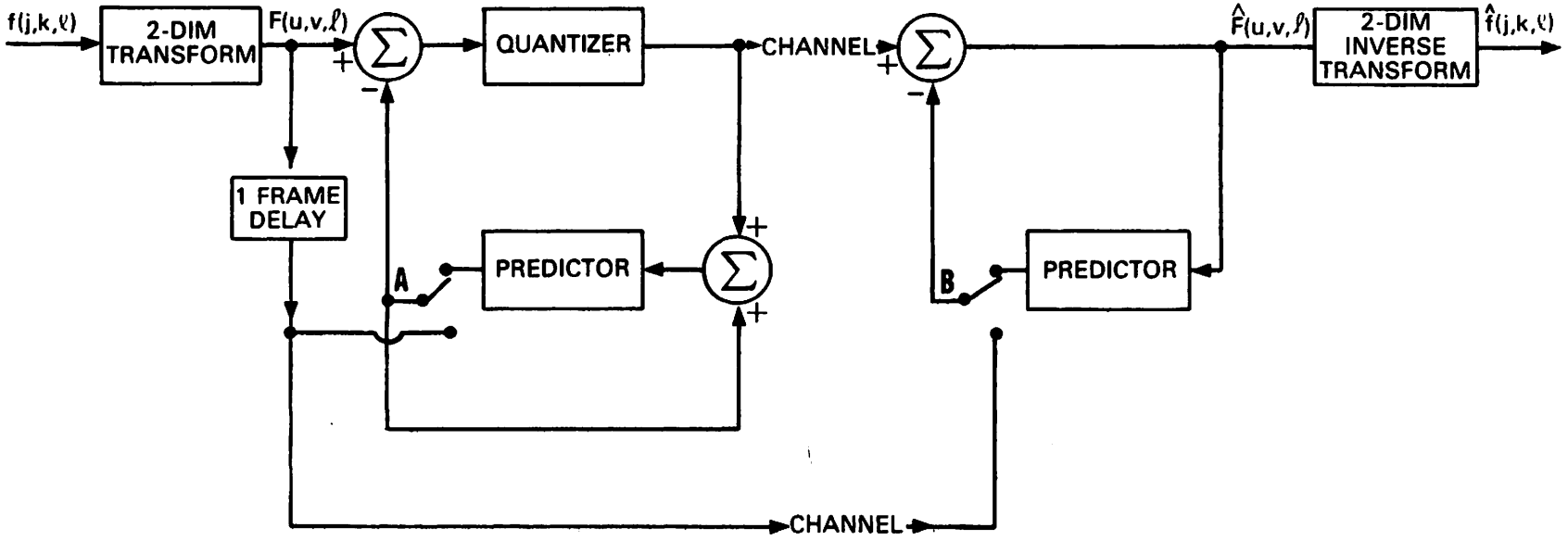


Figure 4-8. DPCM Coder with Reinitialization

The relationship between total transmission bit rate, BR_T, average pixel bit rate, BR_{AP}, and the parameter updating and reinitialization overhead bit rates, BR_{PU} and BR_R, can be simply expressed as

$$BR_T = BR_{AP} + \frac{BR_{PU}}{M} + \frac{BR_R}{L} , \tag{4.18}$$

where M and L are the transmitted frame sequence lengths before, respectively, parameter updating and reinitialization. In equation 4.18, the overhead bit rates are shown as decreasing functions of the number of transmitted frames as they represent information which is transmitted only when a parameter update or coder reinitialization occurs.

When a parameter update occurs, the overhead information to be transmitted consists of a binary map indicating which transform coefficient difference sequences are to be coded for transmission and arrays of transform coefficient means, correlations, and difference variances corresponding to 1's in the binary map. Other selections are possible which will also provide the required overhead information. For example, the binary map could be eliminated leaving the required zonal definition operation to be performed on the complete variance, correlation, and mean arrays at the receiver. Since these arrays are relatively sparse, the former approach was selected for this analysis.

Reductions in the average pixel bit rate, BR_{AP}, have traditionally received the most theoretical attention. A low value for the average pixel bit rate has great importance for interframe coding systems as this quantity represents the average number of bits transmitted on a per frame basis and, consequently, is independent of the length of the frame sequence.

Figure 4-9 illustrates total transmission bit rate for a 32:1 reduction in average pixel bit rate with L = M. In this figure, total transmission bit rate curves for spatially adaptive and spatially non-adaptive operational modes are shown as a function of increasing frame sequence length. A 256 X 256 image size was assumed with 16 X 16 subblocks for the spatially adaptive operational mode. These curves were generated under the additional assumption of format simplifications for the transmission of the overhead data. For example, the first frame transform coefficients could be transmitted after coding by conventional intraframe coding techniques at an average reduced rate of 1.5 bits/pixel. Also, the sparseness of the binary bit map lends itself to run-length coding techniques and could be coded at 0.5 bits/pixel. Finally, the arrays of variance, correlation, and mean values could be reduced to 8, 6, and 8 bits, respectively.

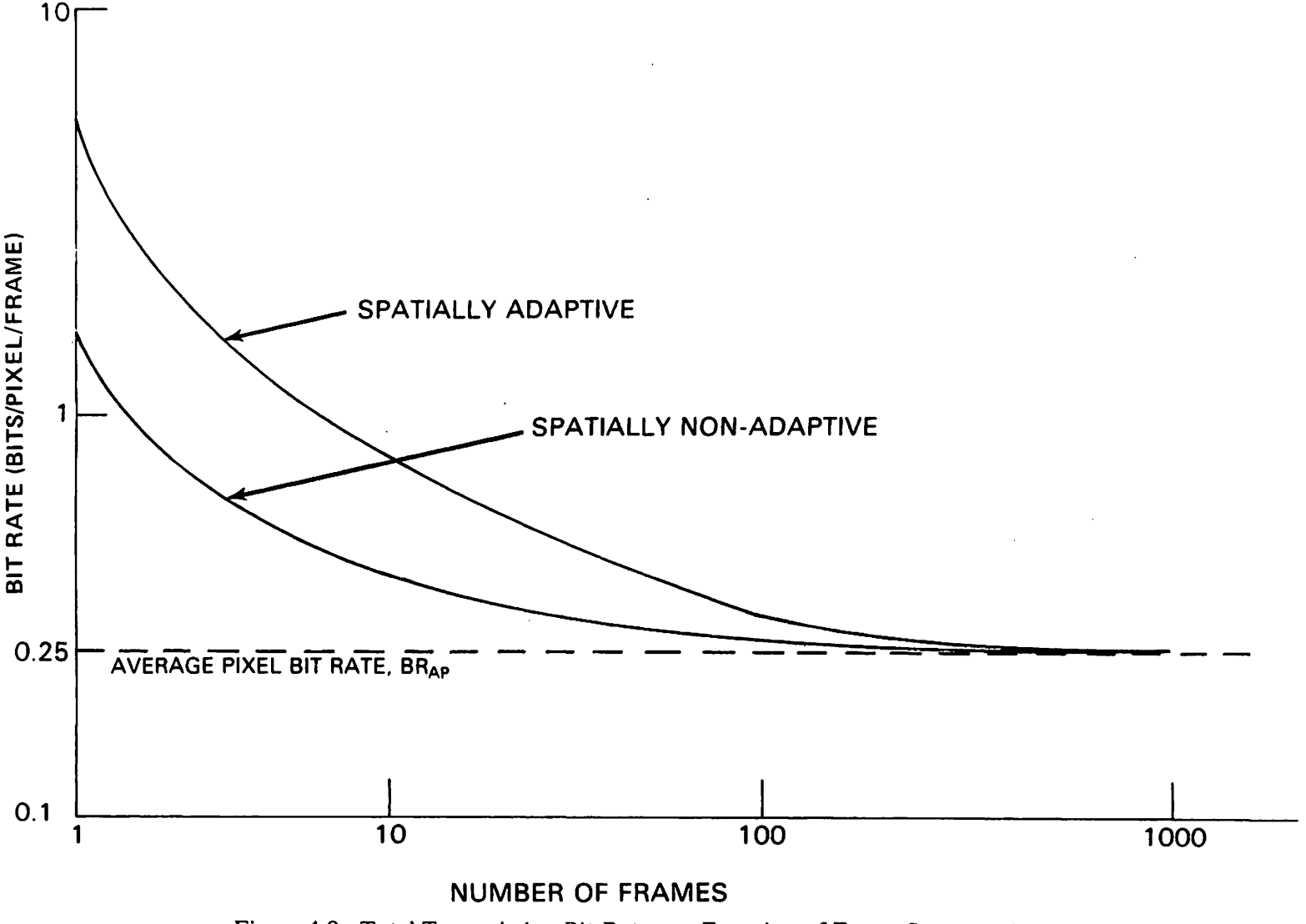


Figure 4-9. Total Transmission Bit Rate as a Function of Frame Sequence Length

Under these assumptions, the expression for the total transmission bit rate is

$$BR_T = 0.25 + \frac{1.5 + NB(0.5 + \Delta(8 + 6 + 8))}{L} \text{ bits/pixel/frame} . \quad (4.19)$$

In equation 4.19, the number 0.25 is the average pixel bit rate, NB is the proportion of updated image subblocks, L is the number of frames, and Δ is the fraction of the transform coefficient difference sequences for which mean, correlation, and variance values are transmitted. For a 32:1 reduction in average bit rate, the value of Δ has been experimentally determined to be approximately 5/32.

The upper curve of Figure 4-9 illustrates total transmission bit rate for the spatially adaptive operational mode. In this case, the proportion of image subblocks updated, NB, is 1. The curve corresponding to the spatially non-adaptive operational mode was generated with NB = 1/256 since only a single bit map and set of mean, correlation, and variance arrays are transmitted. These curves show that for frame sequences of length 50 or more, the dominant factor in the total transmission bit rate is the average pixel bit rate, BR_{AP}.

EXPERIMENTAL EVALUATION OF CODER PERFORMANCE

An extensive set of computer simulations are performed to experimentally evaluate the coding performance of the interframe hybrid coders. The hybrid coders considered are parametric in many variables including choice of separable cosine or Fourier transforms, zonal sampling or zonal coding with companding or Max quantizers, operational modes, spatial subblock size, and average pixel bit rate. In evaluation of coder performance levels, the normalized mean square error (NMSE) and signal-to-noise ratio criteria (SNR) defined in the first chapter are employed in conjunction with subjective visual evaluations.

Two main classes of simulations are performed. The first evaluated coding performance versus subblock size at different average pixel bit rates. All NMSE and SNR calculations are made on the 16th frame after coding stability has been achieved. The second set of simulations illustrates NMSE and SNR as a function of frame number for various average pixel bit rates. Both the two-dimensional cosine and Fourier transform implementation are evaluated in the second series of experiments.

Table 4-2 is presented as an aid to evaluate channel bandwidth reduction factors achieved for the various average pixel bit rates employed in these simulations. In this

Table 4-2. Bit Rate Reductions for Various Average Pixel Bit Rates

Average pixel bit rate (bits/pixel/frame)	Bit Rate Reduction
1.0	8:1
0.5	16:1
0.25	32:1
0.1	80:1

table, original images are assumed to be represented by 8 bits/pixel/frame prior to coding and transmission.

All simulation results in this section are obtained using the 16-frame "head and shoulders" data base with no channel error. The interframe hybrid coders operate under maximum information available initial conditions, spatial adaptation between subblocks, and zonal coding with companding quantizers.

Coding Performance Versus Subblock Size

This series of simulations is designed to determine coding performance sensitivity to variations in spatial subblock size. Figure 4-10 illustrates NMSE and SNR values versus subblock size for the interframe hybrid CCD coder on frame number 16 of the "head and shoulders" data base. The average pixel bit rate values used were in the interval 0.1 to 1.0 bits/pixel/frame.

Figure 4-10 also illustrates continued improvement in coding performance for increasing subblock size. The implication of this series of simulations for potential hardware implementation of interframe hybrid coders is that the largest possible subblock size consistent with storage and complexity limitations should be employed.

Coding Performance versus Number of Frames

Simulations are also performed to define coding performance as a function of frame number. For purposes of comparing levels of performance for the hybrid CCD and FFD coders, separate tests are made using both coder implementations. Subjectively, the

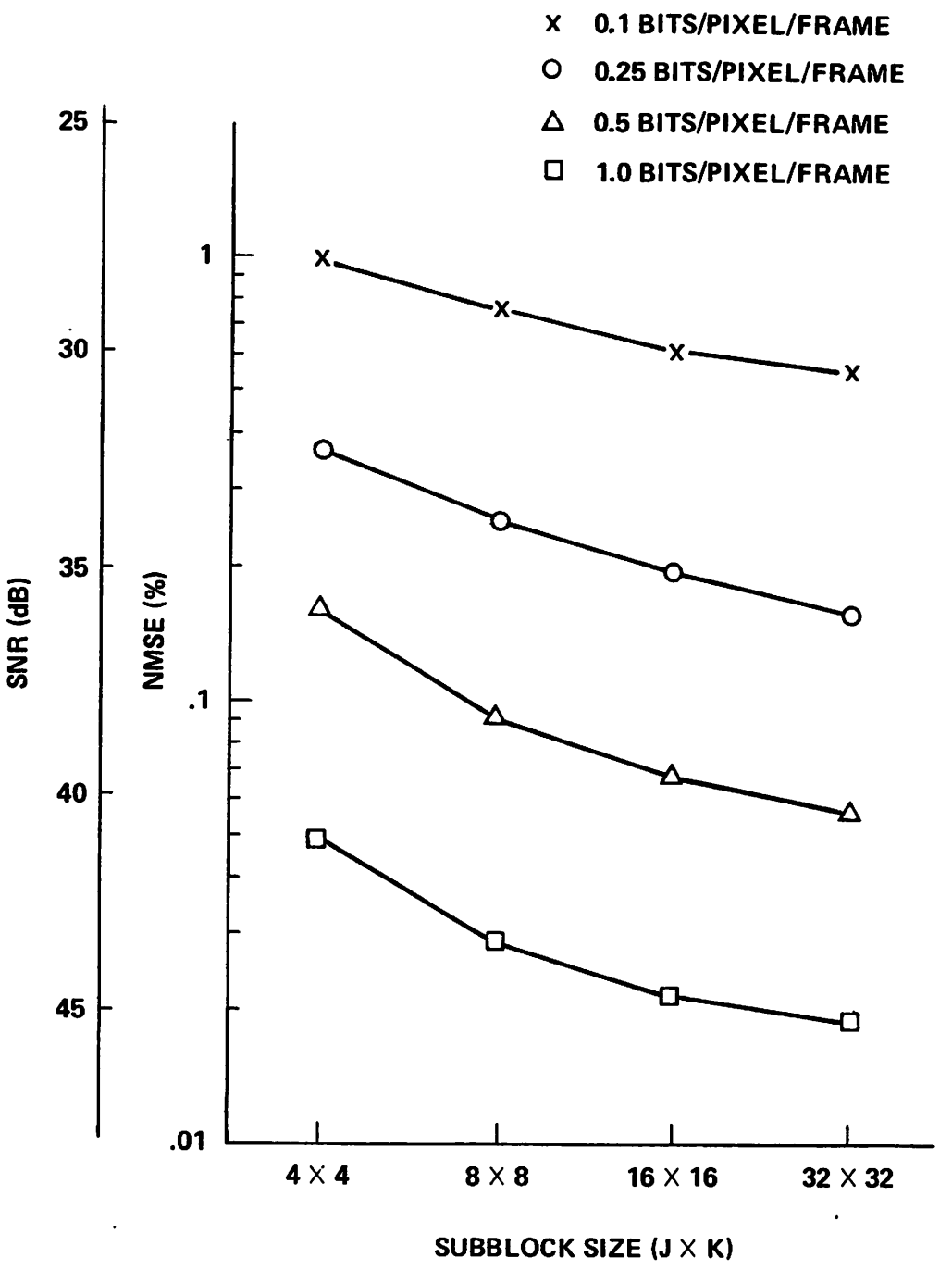


Figure 4-10. Experimental Performance Evaluation of CCD Coder as a Function of Subblock Size

coding performance measures applied to the various coder implementations under evaluation agree with photographs of selected coded frames which accompany the performance plot of each simulation.

Figure 4-11 illustrates NMSE and SNR measures as a function of frame number for the hybrid CCD coder implementation at average pixel bit rates from 0.1 to 1.0 bits/pixel/frame with 16 X 16 subblocks. These graphs indicate that, even with the lowest bit rate, stability in coder performance is achieved within the first 8 frames. For the higher bit rates, performance stability occurs much earlier in the frame sequence. For the maximum information available initial conditions used in these simulations, the first frame was assumed available and consequently zero NMSE is indicated for the initial frame.

Photographs corresponding to frame number 16 at the various average bit rates are illustrated in Figure 4-12 (a) through (d) and show virtually no image degradation for bit rates as low as 0.5 bits/pixel/frame. Some effects of the blocking effect due to the 16 X 16 subblock partitioning of the images is apparent at the 0.25 bits/pixel/frame rate. Also, regions outlining the subject's head begin to show degradation at this bit rate due to head motion and the relatively few coefficients assigned to transmit high frequency coefficients. The observed image degradations are similar in nature but more pronounced at the 0.1 bits/pixel/frame bit rate.

In a like manner, the hybrid FFD coder is evaluated under the same test conditions. Here, the real and imaginary component version of the FFD coder is employed with the bits available for coding equally divided between the two components. The plots of NMSE and SNR for the hybrid FFD coder are shown in Figure 4-13. By way of comparison, the general nature of the data in Figures 4-13 and 4-11 is similar except that, for all average pixel bit rates examined, NMSE values for the FFD coder are higher, and in some cases double, than for the CCD coder. This result is subjectively borne out by comparison of the photographs (a) through (d) in Figures 4-12 and 4-14. These photographs indicate that the fidelity of the reconstructed image is less for the FFD coder than for the CCD coder at each pixel bit rate considered.

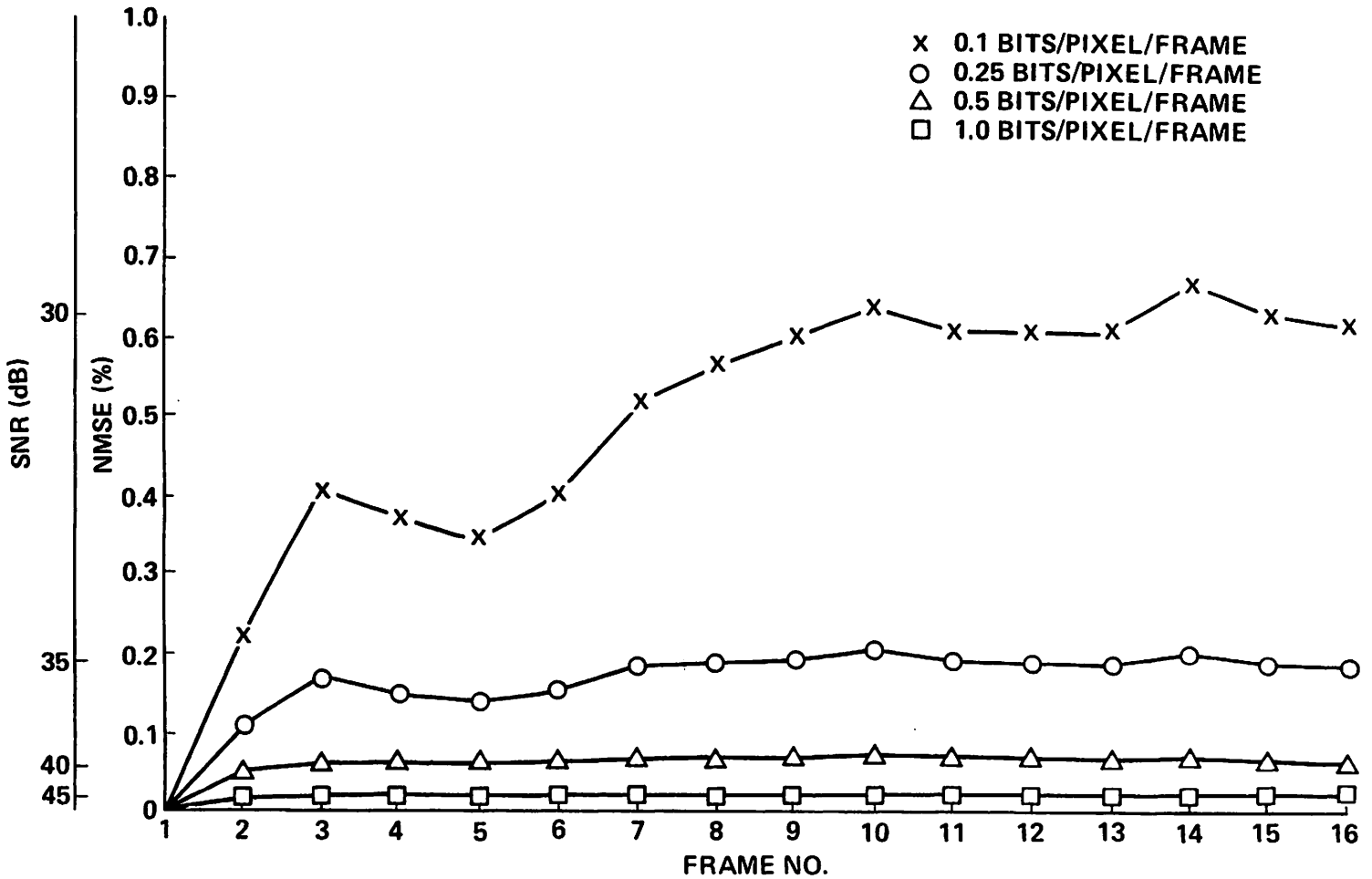


Figure 4-11. Coding Performance as a Function of Frame Number for the Hybrid CCD Coder

(a) 1.0-bits/pixel/frame

(b) 0.5 bits/pixel/frame

(c) 0.25 bits/pixel/frame

(d) 0.1 bits/pixel/frame

Figure 4-12. Coding Performance of Hybrid CCD Coder

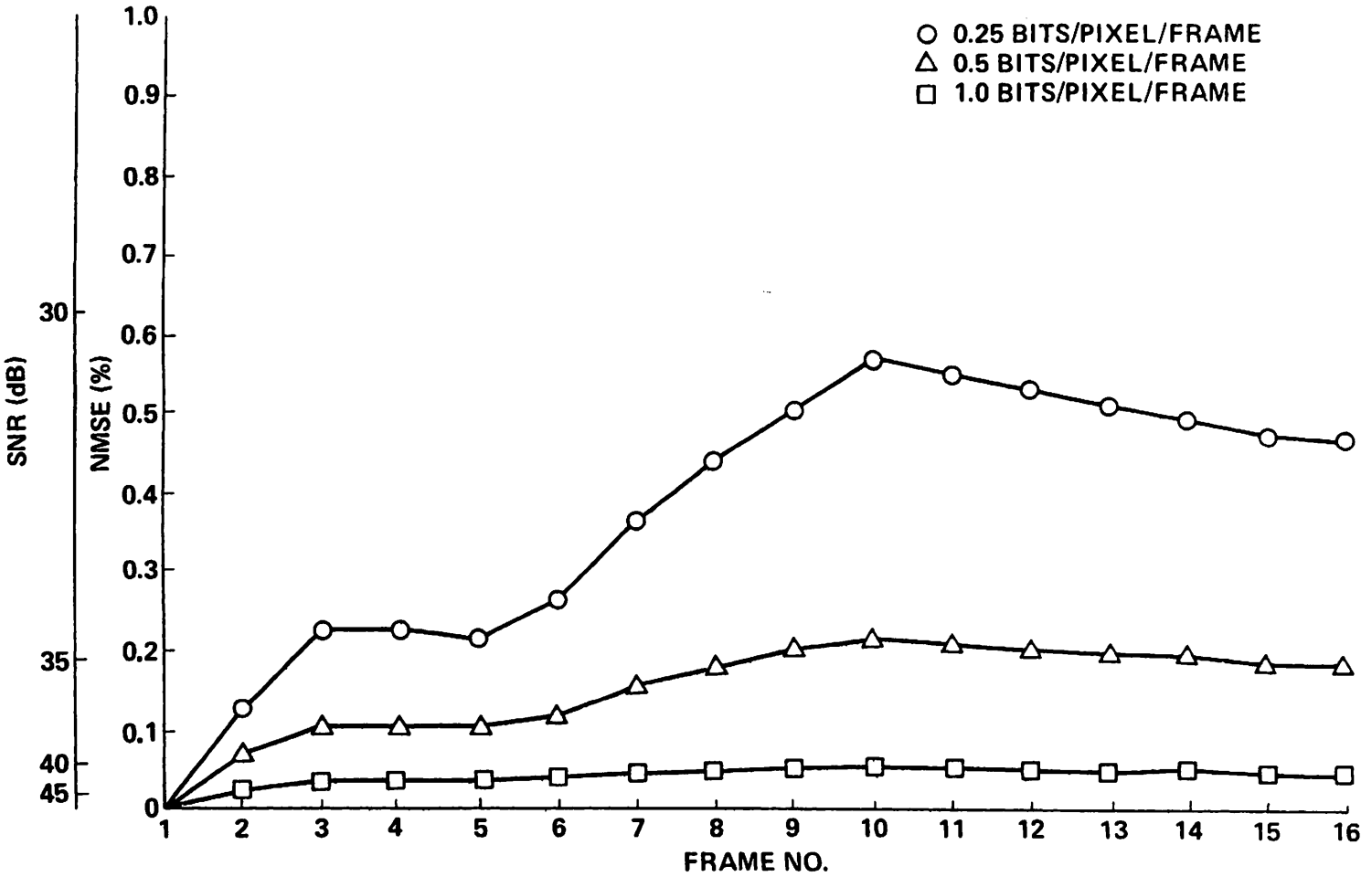


Figure 4-13. Coding Performance as a Function of Frame Number for the Hybrid FFD Coder

(a) 1.0 bits/pixel/frame

(b) 0.5 bits/pixel/frame

(c) 0.25 bits/pixel/frame

(d) 0.1 bits/pixel/frame

Figure 4-14. Coding Performance of Hybrid FFD Coder

CHANNEL BIT TRANSFER RATE

In systems designed to code sequences of digital images, a fundamental tradeoff exists between the frame repetition rate and the average pixel bit rate per frame for a given channel bit transfer rate. The bit transfer rate (BTR) is defined as the product of the average pixel bit rate per frame multiplied by the frame rate and has units of bits/pixel/sec. For comparison, Table 4-3 provides BTR values for several combinations of frame rate and average pixel bit rate and gives the corresponding channel bandwidth requirements based on a 256 X 256 image size.

A series of simulations was performed with the hybrid CCD coder in which the BTR across the channel was fixed. The 16-frame "head and shoulders" data base which was extracted from a 24 fps motion picture sequence was used in these simulations. By employing frame skipping techniques, temporal subsampling was used to simulate short 12, 8, and 6 fps sequences from the original 16-frame, 24-fps data base. Average pixel bit rates in the interval 0.083 to 1.333 bits/pixel/frame were used in conjunction with the four frame rates mentioned above to perform simulations with fixed BTR values of 8, 6, 4, and 2 bits/pixel/sec. Table 4-4 summarizes the frame rate, average pixel bit rate, and BTR values for each computer simulation performed. The results of the BTR simulations are shown in Figures 4-15 through 4-18 and illustrate NMSE as a function of frame number for BTR values of 8, 6, 4, and 2 bits/pixel/sec, respectively.

For all cases examined, the interpretation of the experimental BTR results is that frame-to-frame reductions in correlation due to temporal subsampling, i.e., reduced frame rates, are completely compensated for by the increased number of bits available for coding. However, no consideration is given to image degradations which occur at the receiver as a result of not transmitting every available frame. Subjectively, reduced frame rates tend to produce jerky transitions between image updates. This is most apparent for rapidly moving objects in the field of view and is of lesser consequence for slowly changing scenes. Such receiver image degradations may be acceptable for the intended system application or, if not, frame-to-frame smoothing techniques may be required.

Table 4-3. BTR Values and Channel Bandwidth Requirements for 256 X 256 Images

BTR (BITS/PIXEL/SEC)

AVERAGE PIXEL BIT RATE (BITS/PIXEL/FRAME)	FRAME RATE (FPS) 1	3	6	8	12	24
0.1	0.1	0.3	0.6	0.8	1.2	2.4
0.25	0.25	0.75	1.5	2.	3.	6.
0.5	0.5	1.5	3.	4.	6.	12.
1.0	1.	3.	6.	8.	12.	24.
8.0	8.	24.	48.	64.	96.	192.

CHANNEL BANDWIDTH (K BITS/SEC)

	FRAME RATE (FPS) 1	3	6	8	12	24
0.1	6	20	39	52	78	156
0.25	16	49	98	131	196	393
0.5	32	98	196	262	392	786
1.0	64	196	392	524	786	1.57×10^3
8.0	524	1.57×10^3	3.14×10^3	4.19×10^3	6.29×10^3	1.26×10^4

(256 × 256 IMAGE SIZE)

Table 4-4. Bit Transfer Rates for BTR Simulations with Hybrid CCD Coder

AVERAGE PIXEL BIT RATE (BITS/PIXEL/FRAME)	FRAME RATE (FPS) 6	8	12	24
0.083	—	—	—	2
0.167	—	—	2	4
0.25	—	2	—	6
0.333	2	—	4	8
0.5	—	4	6	—
0.667	4	—	8	—
0.75	—	6	—	—
1.0	6	8	—	—
1.333	8	—	—	—

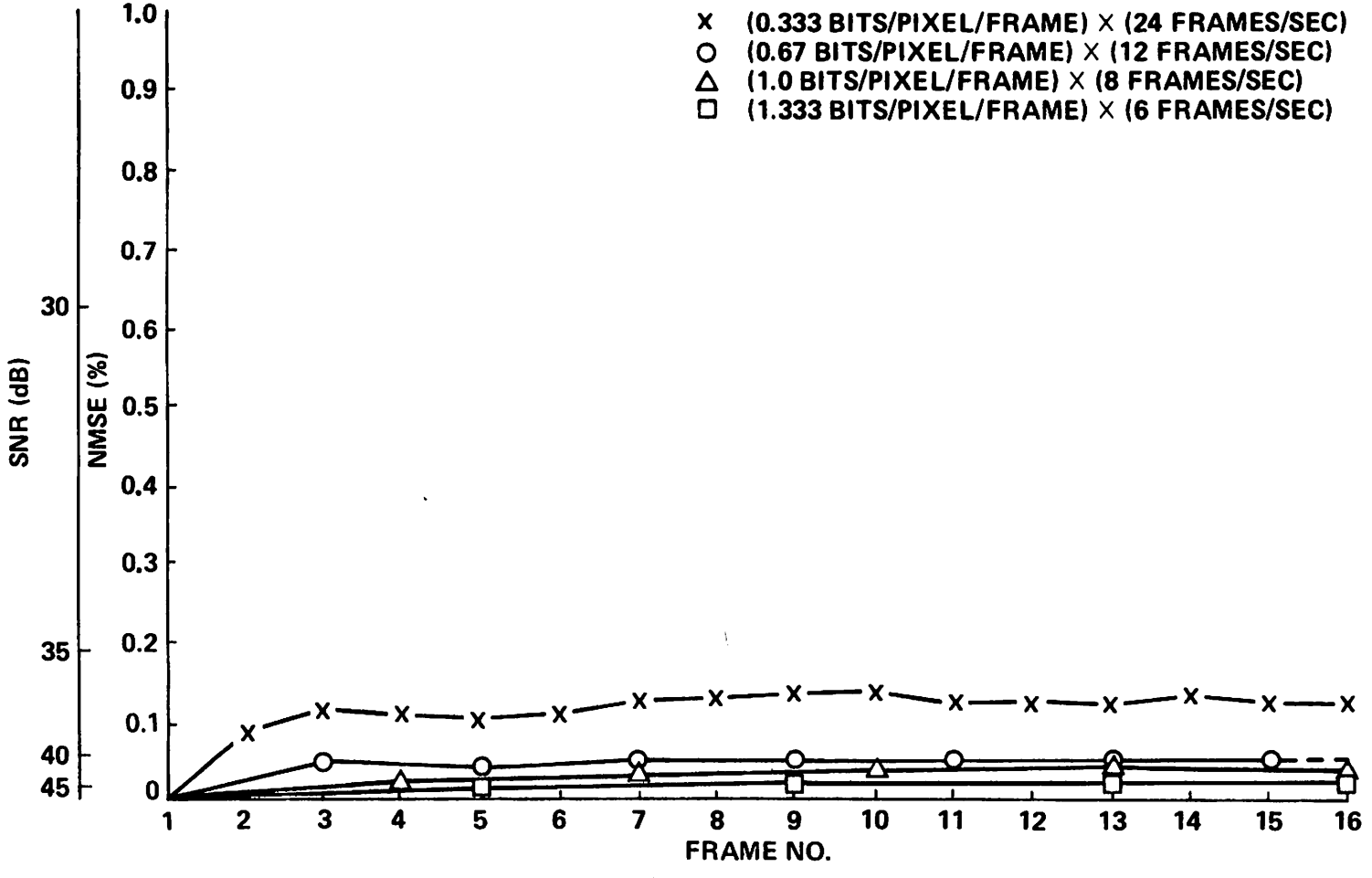


Figure 4-15. Hybrid CCD Coder at Bit Transfer Rate of 8 Bits/Pixel/Sec

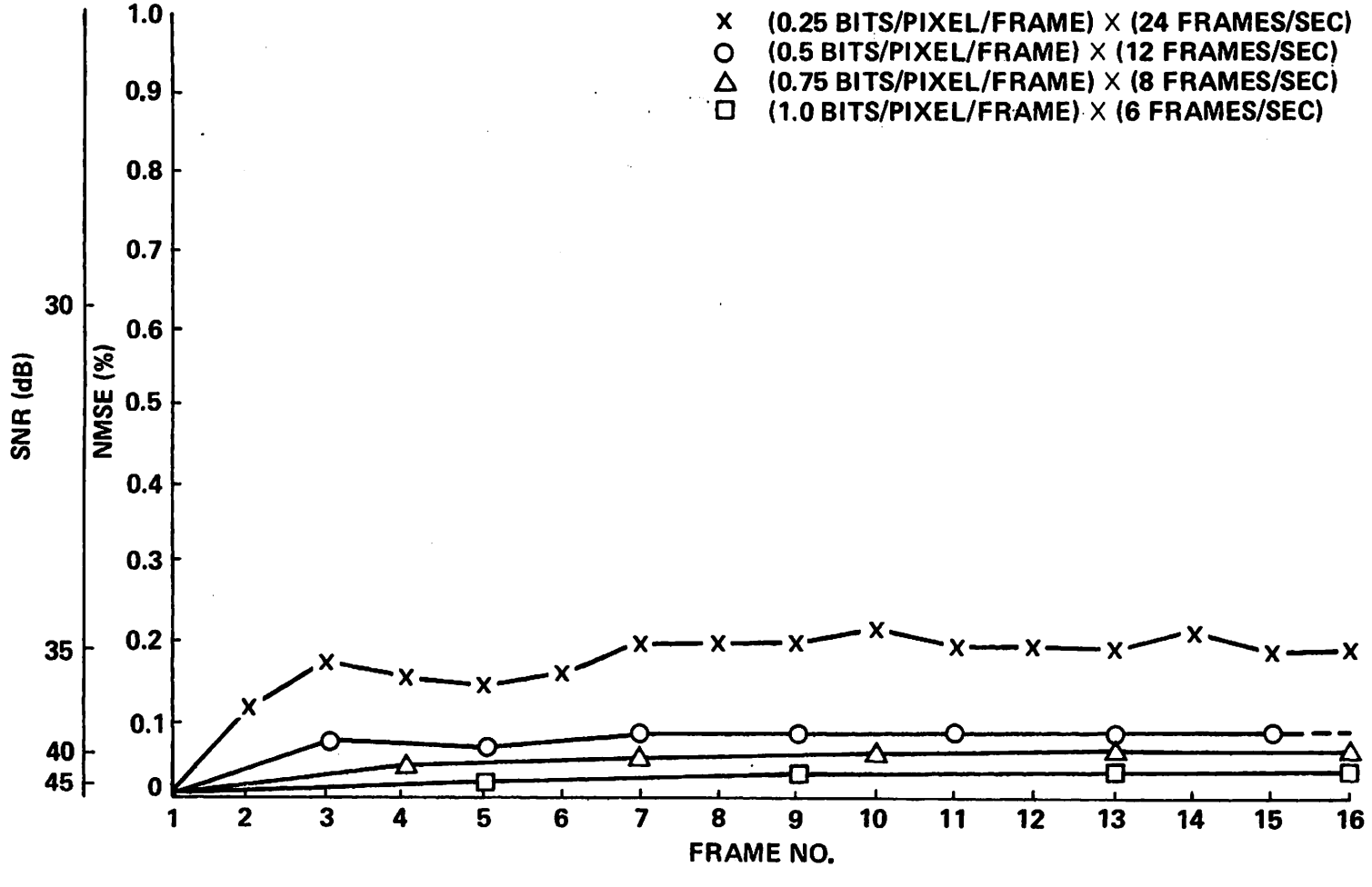


Figure 4-16. Hybrid CCD Coder at Bit Transfer Rate of 6 Bits/Pixel/Sec

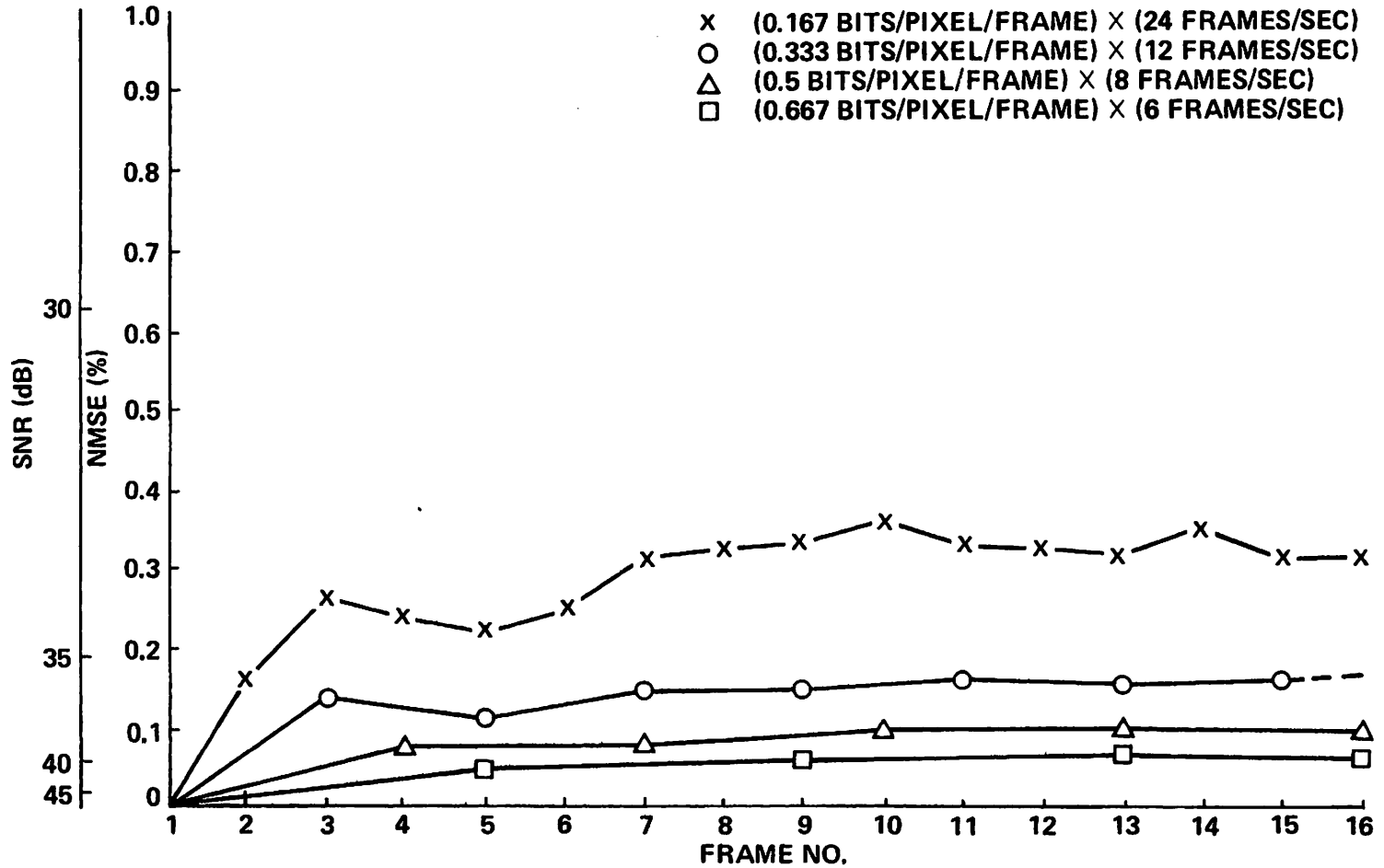


Figure 4-17. Hybrid CCD Coder at Bit Transfer Rate of 4 Bits/Pixel/Sec

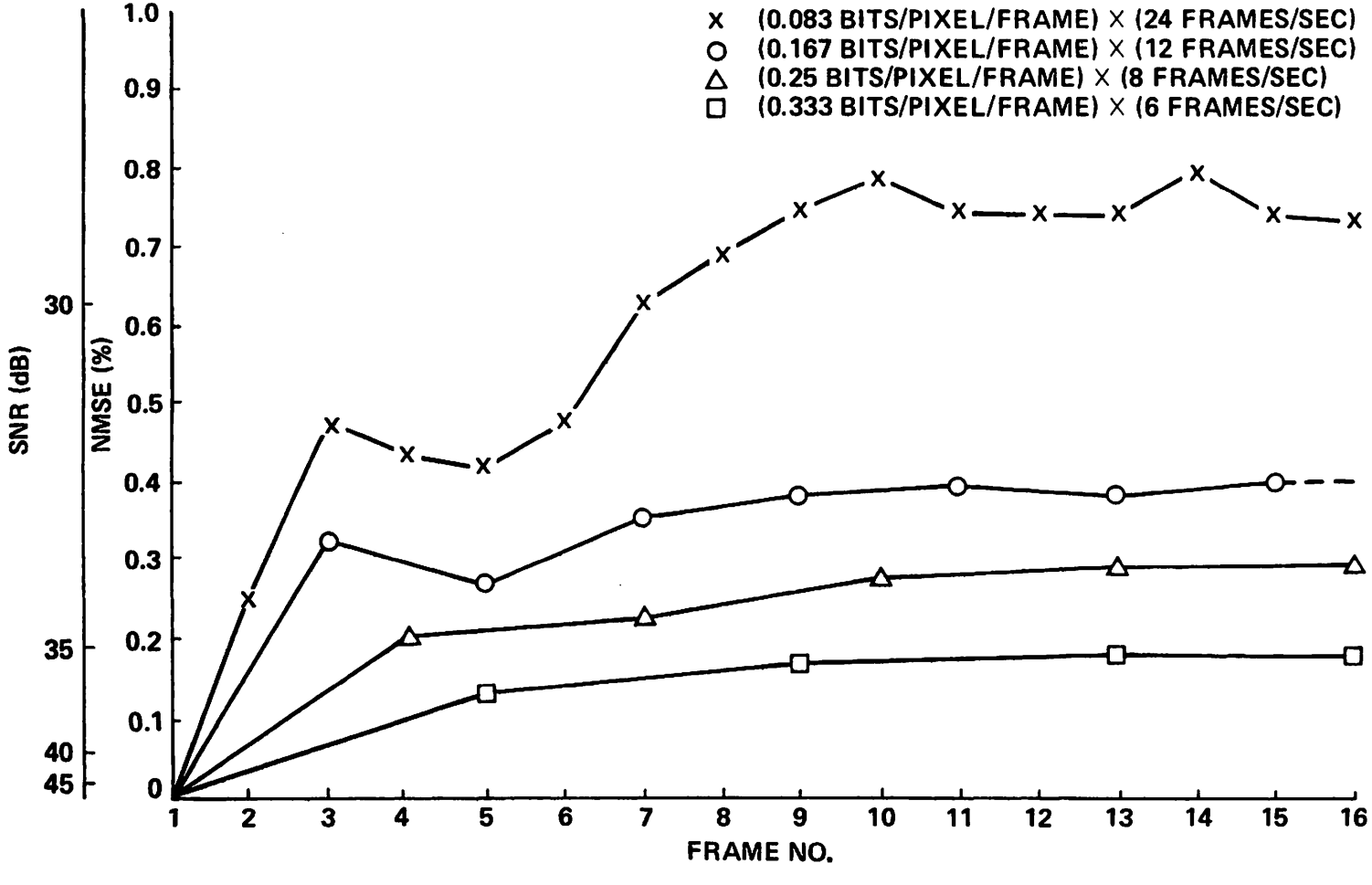


Figure 4-18. Hybrid CCD Coder at Bit Transfer Rate of 2 Bits/Pixel/Sec

SIMPLIFIED CODER IMPLEMENTATION

Throughout this investigation of interframe hybrid coders, the emphasis is placed on coder implementations designed to achieve high levels of coding performance. For example, the experimental results presented thus far have been obtained with spatially adaptive implementations of the hybrid CCD and FFD coders. However, applications for hybrid interframe coders exist in which hardware implementation constraints may make spatially adaptive coder implementations undesirable.

In this section, an experimental investigation is made of coding performance for a spatially non-adaptive implementation of the interframe hybrid CCD coder. This simplified hybrid CCD coder is non-adaptive in the sense that a single set of statistics and, consequently, the same bit allocation array is applied to each spatial image subblock. Sixteen frames of the "moving camera" data base were used for this set of coding simulations. Based on a 16 X 16 spatial subblock size, the resulting fixed bit assignment array is shown as Figure 4-19.

In Figure 4-20, the coding performance of the simplified hybrid CCD coder is given as a function of frame number at a bit rate of 0.25 bits/pixel. In this simulation, the simplified coder used a computed set of fixed statistics and the bit allocation array of Figure 4-19. For comparison purposes, the coding performance of the spatially adaptive hybrid CCD coder is also plotted in Figure 4-20.

3	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0
3	2	2	1	1	1	0	0	0	0	0	0	0	0	0	0
3	2	2	2	1	1	0	0	0	0	0	0	0	0	0	0
2	2	2	2	1	1	1	0	0	0	0	0	0	0	0	0
2	2	1	1	1	1	0	0	0	0	0	0	0	0	0	0
2	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 4-19. Spatially Non-Adaptive Bit Allocation Array for "Moving Camera" Data Base with 16 X 16 Subblock Size

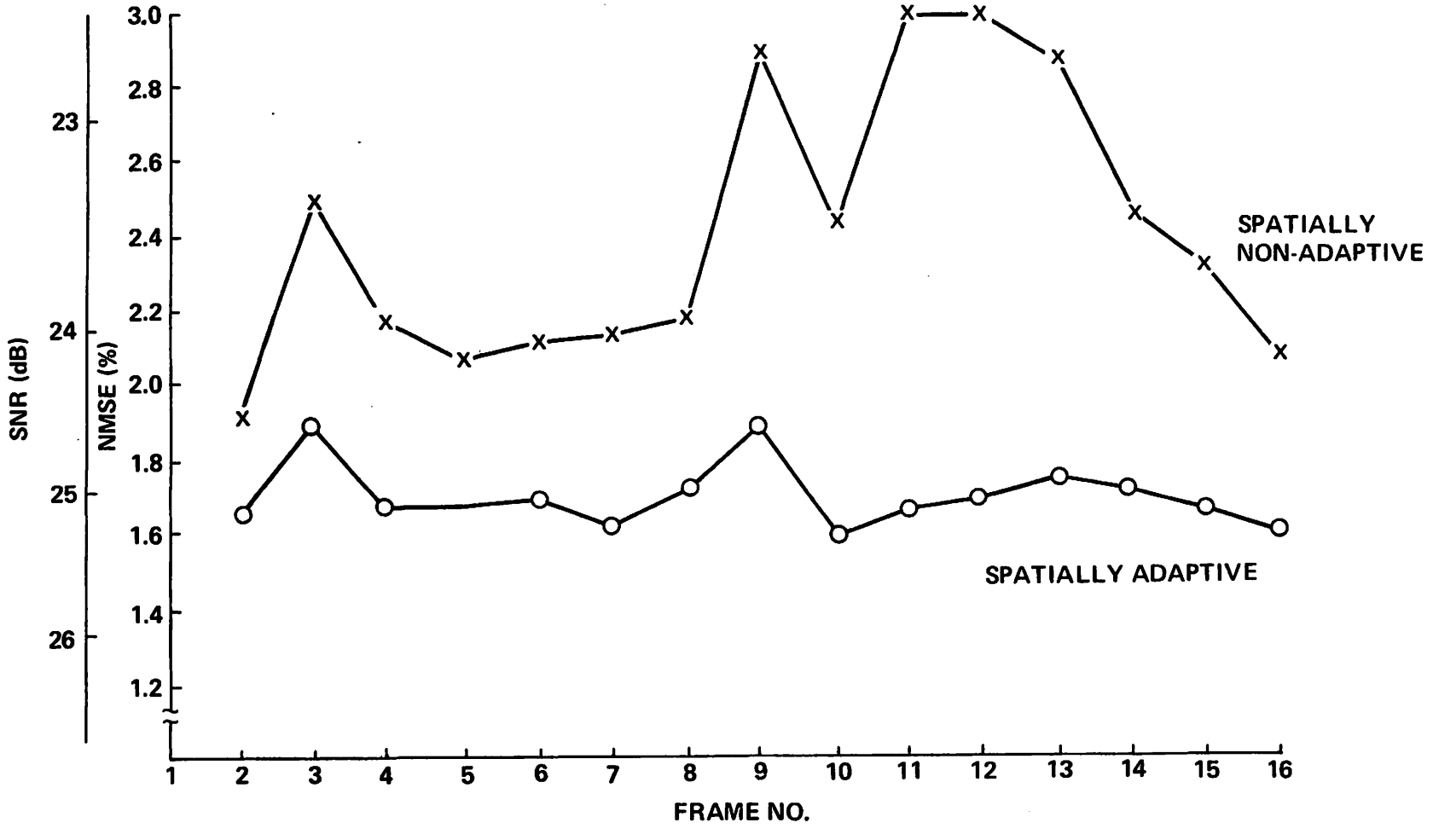


Figure 4-20. Coding Performance Comparison for Spatially Adaptive and Non-Adaptive Hybrid CCD Coders

5. CHANNEL ERROR EFFECTS

In the preceding chapters, implementations for interframe three-dimensional transform and hybrid coders are defined. Also, simulations are performed to illustrate the performance characteristics of these coders. However, these simulations assume an error-free channel for image transmission.

For all practical digital communication systems, this assumption is unrealistic. In the case of interframe hybrid CCD and FFD coders, periodic reinitialization of the DPCM transmitter and receiver prediction signals can be used to correct for the cumulative effects of channel error. Sensitivity of the coders to channel error determines the frequency at which reinitialization is required.

This chapter investigates the effects of channel errors for the interframe hybrid CCD and FFD coders. In this analysis, computer simulations are performed with various probabilities of channel error. Degradations in receiver image reconstructions, including the effects of channel error, are evaluated in terms of NMSE and SNR criteria. In addition, subjective evaluations are made from photographs of the transmitted images.

BINARY SYMMETRIC CHANNEL MODEL

A binary symmetric channel is simulated to evaluate the effects of channel error. The basic form for this model is shown in Figure 5-1. This model assumes that the channel operates on each binary digit independently, changing each binary digit from 0 to 1 or from 1 to 0 with probability P_{ce} and leaving the digit unchanged with probability $1-P_{ce}$. At the receiver, each image is reconstructed from the string of transmitted binary digits, including errors.

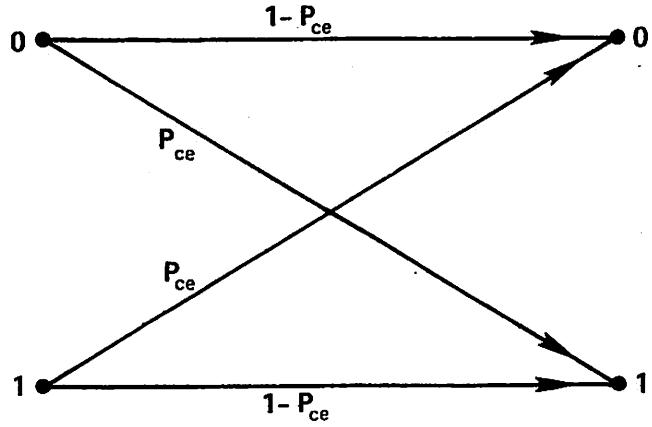


Figure 5-1. Binary Symmetric Channel Model

DPCM coding of transform coefficients in the presence of channel errors results in degraded image reconstructions at the receiver. In addition, the recursive nature of the DPCM predictor causes propagation of transform coefficient errors to succeeding frames. In the CCD and FFD coders, the feedback loop gain coefficients are normalized transform coefficient correlations. Since the correlations, $\rho(u,v)$, computed for each set of transform coefficient differences are nominally less than unity, the effect of an isolated transform coefficient error tends to decay with the coding of additional frames. Thus, the rate at which the channel error decays is dependent on the size of $\rho(u,v)$, with smaller correlations producing more rapid decays.

In the implementation of the interframe hybrid CCD and FFD coders, the distribution of the bits available to code each transform coefficient difference is uneven. Consequently, the effect of a given channel error probability, P_{ce}, is uneven for different transform coefficients. Also, the occurrence of a channel error in the most significant digit is usually more detrimental than an error occurring in one of the least significant digits. Finally, an error resulting in an incorrect low-order or dc transform coefficient is generally more visually degrading to the fidelity of a reconstructed image than an error in a higher order transform coefficient.

EXPERIMENTAL EVALUATION OF CHANNEL ERROR EFFECTS

Degradations in terms of NMSE and SNR resulting from channel errors having probabilities, P_{ce}, of 10^{-3} and 10^{-2} are shown in Figures 5-2 through 5-5 for the CCD and FFD coders. For these simulations, the CCD and FFD coders were operated under the same conditions as discussed earlier. The results illustrated are for average pixel bit rates of 1.0 and 0.25 bits/pixel/frame. The 16 frame "head and shoulders" data base was used without coder reinitialization. For comparison purposes, graphs showing coder performance levels with $P_{ce} = 0$ are also included. Simulations were also run with a channel error probability of 10^{-4}. However, the coding results obtained with $P_{ce} = 10^{-4}$ were essentially indistinguishable from the case with $P_{ce} = 0$ and are not included. Photographs showing the visual effects of channel noise on frame number 16 of the "head and shoulders" data base are shown in Figures 5-6 (a) through (d) and 5-7 (a) through (d) for the two coders.

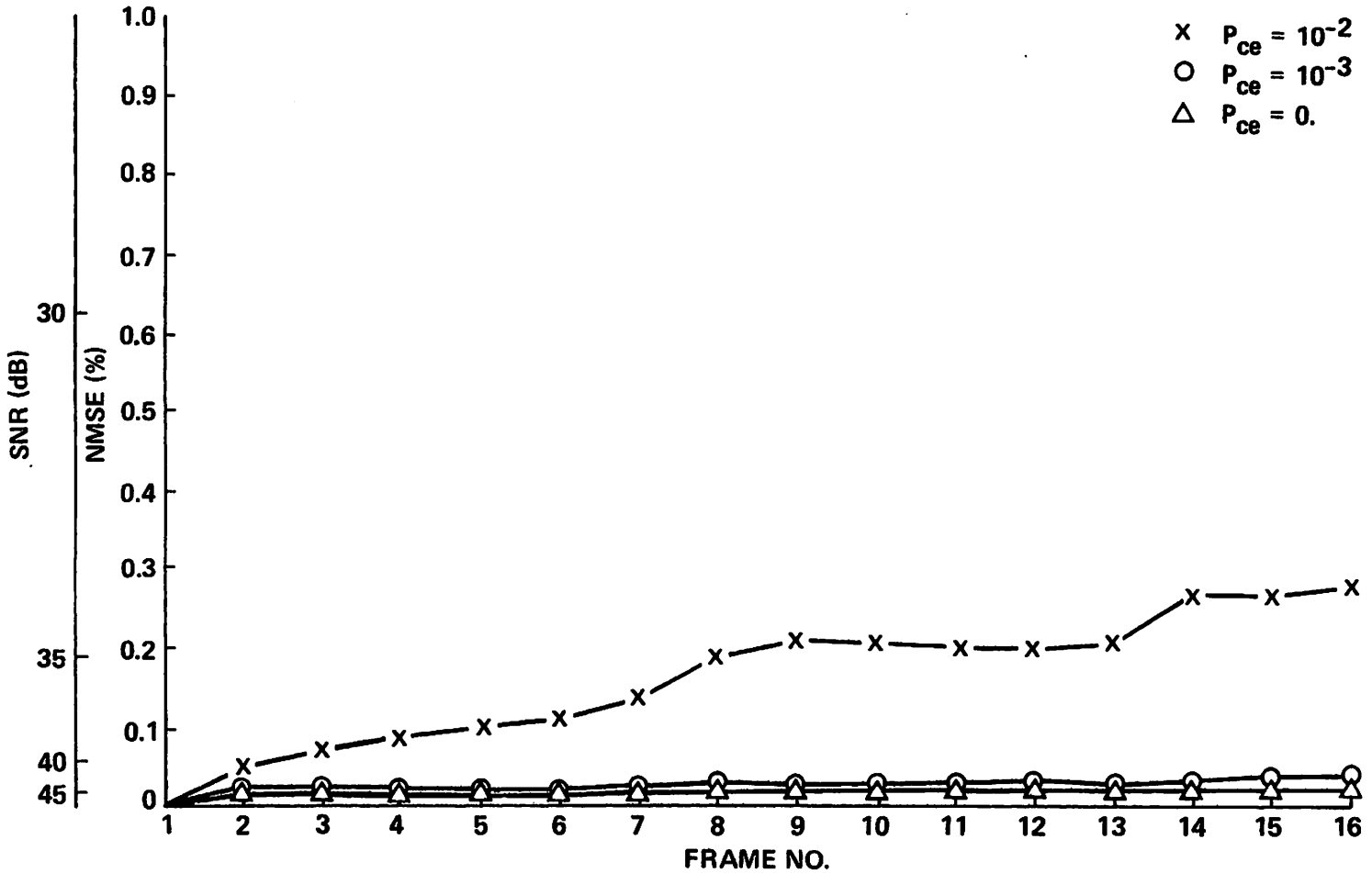


Figure 5-2. Effects of Channel Error for CCD Coder at 1.0 Bits/Pixel/Frame Using "Head and Shoulders" Data Base

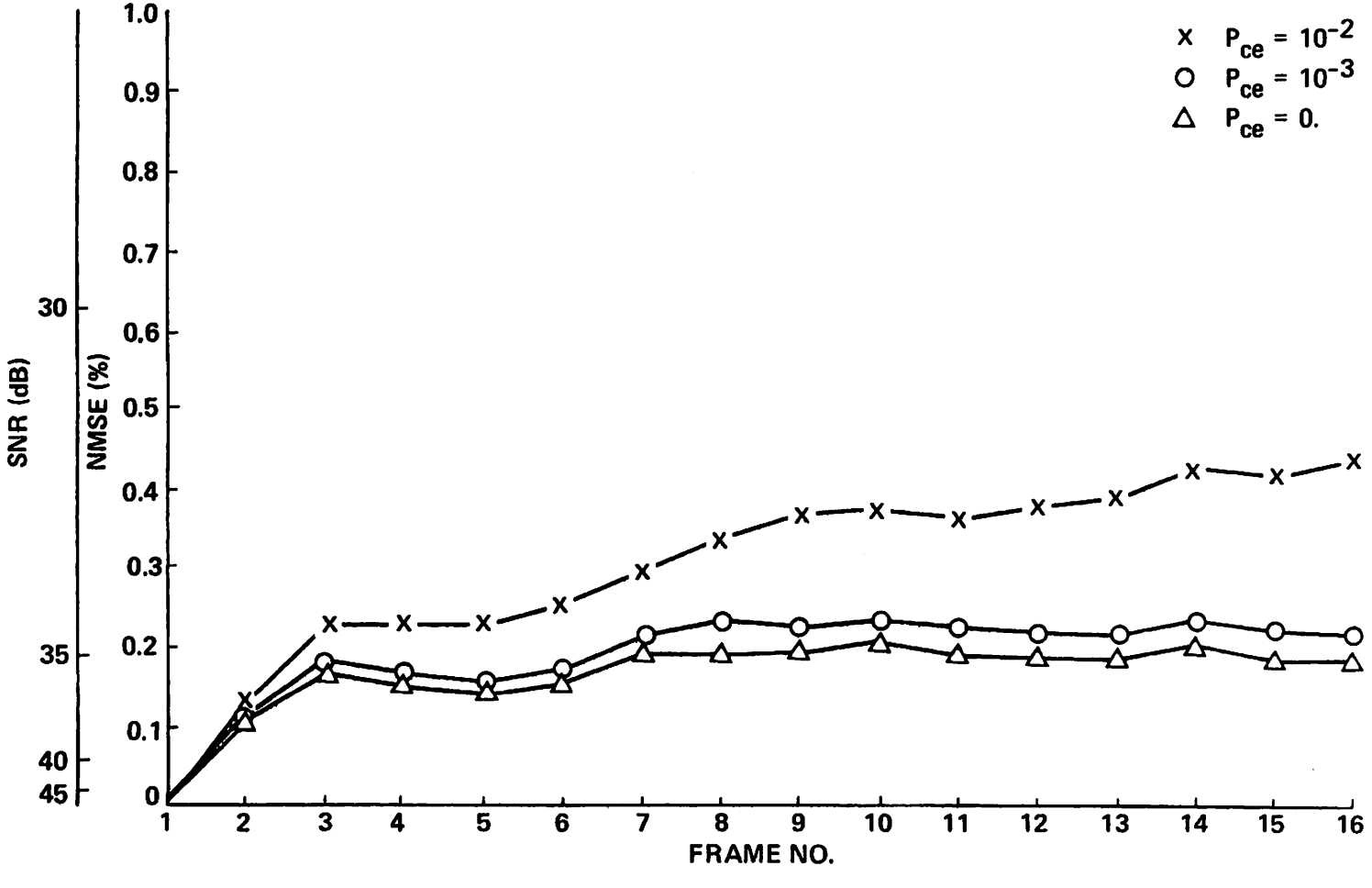


Figure 5-3. Effects of Channel Error for CCD Coder at 0.25 Bits/Pixel/Frame Using "Head and Shoulders" Data Base

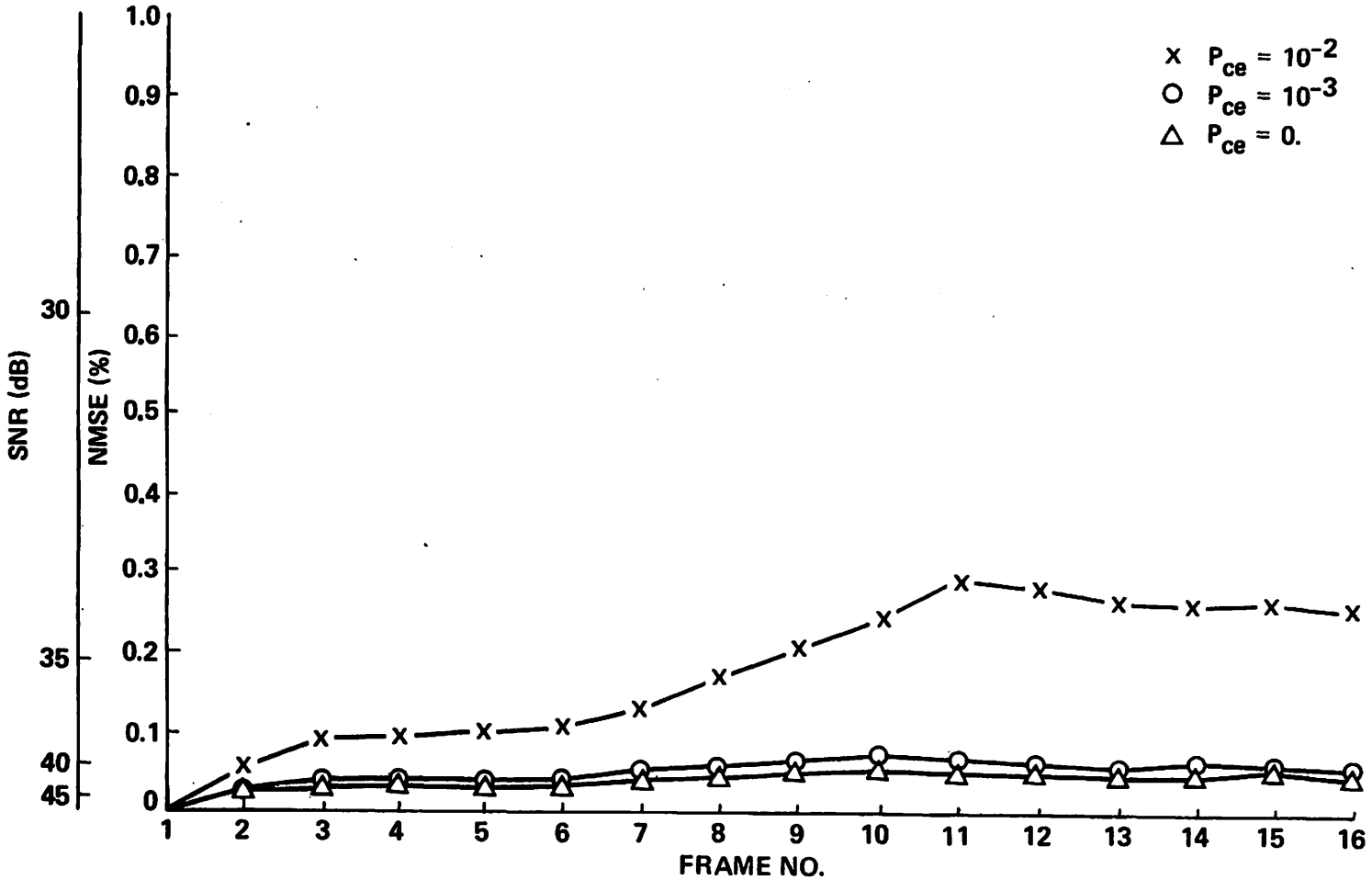


Figure 5-4. Effects of Channel Error for FFD Coder at 1.0 Bits/Pixel/Frame Using "Head and Shoulders" Data Base

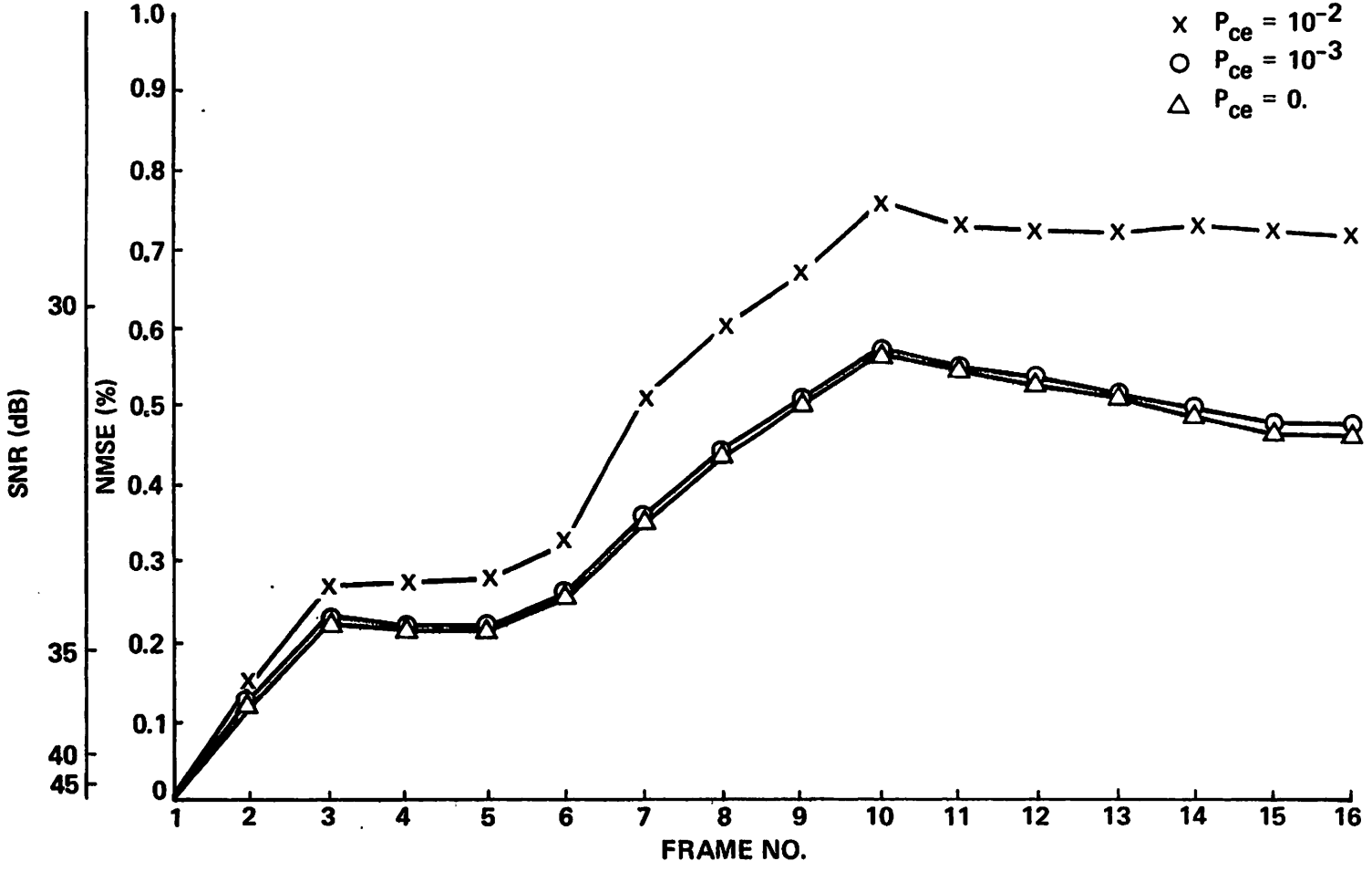


Figure 5-5. Effects of Channel Error for FFD Coder at 0.25 Bits/Pixel/Frame Using "Head and Shoulders" Data Base

(a) 1.0 bits/pixel, $P_{ce} = 10^{-3}$

(b) 1.0 bits/pixel, $P_{ce} = 10^{-2}$

(c) 0.25 bits/pixel, $P_{ce} = 10^{-3}$

(d) 0.25 bits/pixel, $P_{ce} = 10^{-2}$

Figure 5-6. Performance of CCD Coder with Channel Error for Frame No. 16 of "Head and Shoulders" Data Base

(a) 1.0 bits/pixel, $P_{ce} = 10^{-3}$

(b) 1.0 bits/pixel, $P_{ce} = 10^{-2}$

(c) 0.25 bits/pixel, $P_{ce} = 10^{-3}$

(d) 0.25 bits/pixel, $P_{ce} = 10^{-2}$

Figure 5-7. Performance of FFD Coder with Channel Error for Frame No. 16 of "Head and Shoulders" Data Base

The generally stable character of the NMSE and SNR curves for channel error probabilities of 10^{-3} illustrates that the rate of DPCM feedback loop decay is sufficient to avoid excessive error buildup. This is not true, however, for the simulations performed with the relatively high channel error probability of 10^{-2}. In this case, degradations in coding performance continue to increase with the coding of subsequent frames. This indicates accumulation of coding errors at a rate exceeding the coder's capacity to absorb them with normal DPCM feedback loop decay.

6. CAMERA MOTION COMPENSATION

The effectiveness of any interframe image coding scheme is dependent on the extent to which the amplitudes of temporally adjacent pixels are correlated. This is evidenced by the simulation results attained using the "head and shoulders" data base. In the reconstructed images, the greatest losses in image fidelity occur in regions of maximum change between frames. Generally, these regions are along the transitions between the background and the subject's head and shoulders. Conversely, excellent quality image reconstruction is achieved for the stationary background areas.

By contrast, the "moving camera" data base was originally photographed from a moving platform. Thus, as a result of camera motion, this entire image sequence is spatially shifted from frame to frame. In terms of image coding, the effect of spatial shifts induced by camera motion will, in general, degrade the performance of interframe coding algorithms. This is because correlations between temporally adjacent pixels across the entire image are effectively reduced.

The intent of this chapter is two-fold. First, an experimental validation is presented which demonstrates an improved level of interframe coder performance under the assumption of motion correction. Second, two mathematical approaches for compensation of camera motion effects are presented. Both compensation techniques involve preprocessing of the image data prior to quantization for transmission. The first method considered is to compensate each frame in the spatial domain based on frame registration techniques. Following spatial domain compensation, the interframe transform and hybrid transform/DPCM coding techniques developed in earlier chapters are directly applied. The second method employs the shifting property of the Fourier transform. In the Fourier transform, shifts in the spatial domain variables result in a multiplication of the Fourier transform of the variables by a phase factor. Thus, perturbations resulting from camera motion are used to

generate phase correction factors. These phase correction factors are applied to the Fourier transform coefficients prior to coding in the temporal domain.

INTERFRAME CODER PERFORMANCE WITH MOTION CORRECTION

In order to validate the basic concept that motion correction results in improved coder performance, a series of computer simulations is performed with a motion corrected subset of the "moving camera" data base. A 16-frame sequence, selected from near the end of the 64-frame "moving camera" data base, is used for this series of simulations. The "moving camera" data base represents a scenario in which the platform is moving closer to the structures within the field of view; thus, the selected 16-frame sequence illustrates the maximum effects of camera motion.

Since the original 64-frame "moving camera" data base is digitized at a resolution of 512 X 512 pixels, it is possible to partition off spatially displaced 256 X 256 pixel size subimages within each of the 16 frames selected. The ability to define subimages within each 512 X 512 resolution frame is essential for this set of simulations as it permits use of a very simple motion correction technique.

In these simulations, motion correction is performed by measuring the extent of horizontal and vertical motion of various features of the image between the first and sixteenth frames. The measured horizontal and vertical displacements in terms of pixel columns and rows displaced per frame are, respectively, 0.25 and 2 pixels.

Figure 6-1 shows that motion correction is achieved by fixing the coordinates of the 16th frame to be $x_{16} = 128$ and $y_{16} = 256$. With the use of the measured number of pixel rows and columns shifted per frame, the corresponding coordinates x_1 and y_1 of the first frame are determined. Thus, given the original 512 X 512 resolution sequence, it is possible to generate the desired motion corrected 256 X 256 resolution image sequency by a scheme that requires only a simple moving partition.

With the 16-frame motion corrected data base and the corresponding data base without motion correction, a set of stimulations is performed with the interframe hybrid CCD coder. The results for both the motion corrected and non-motion corrected data bases are shown in Figure 6-2. This figure indicates that even with the simple techniques

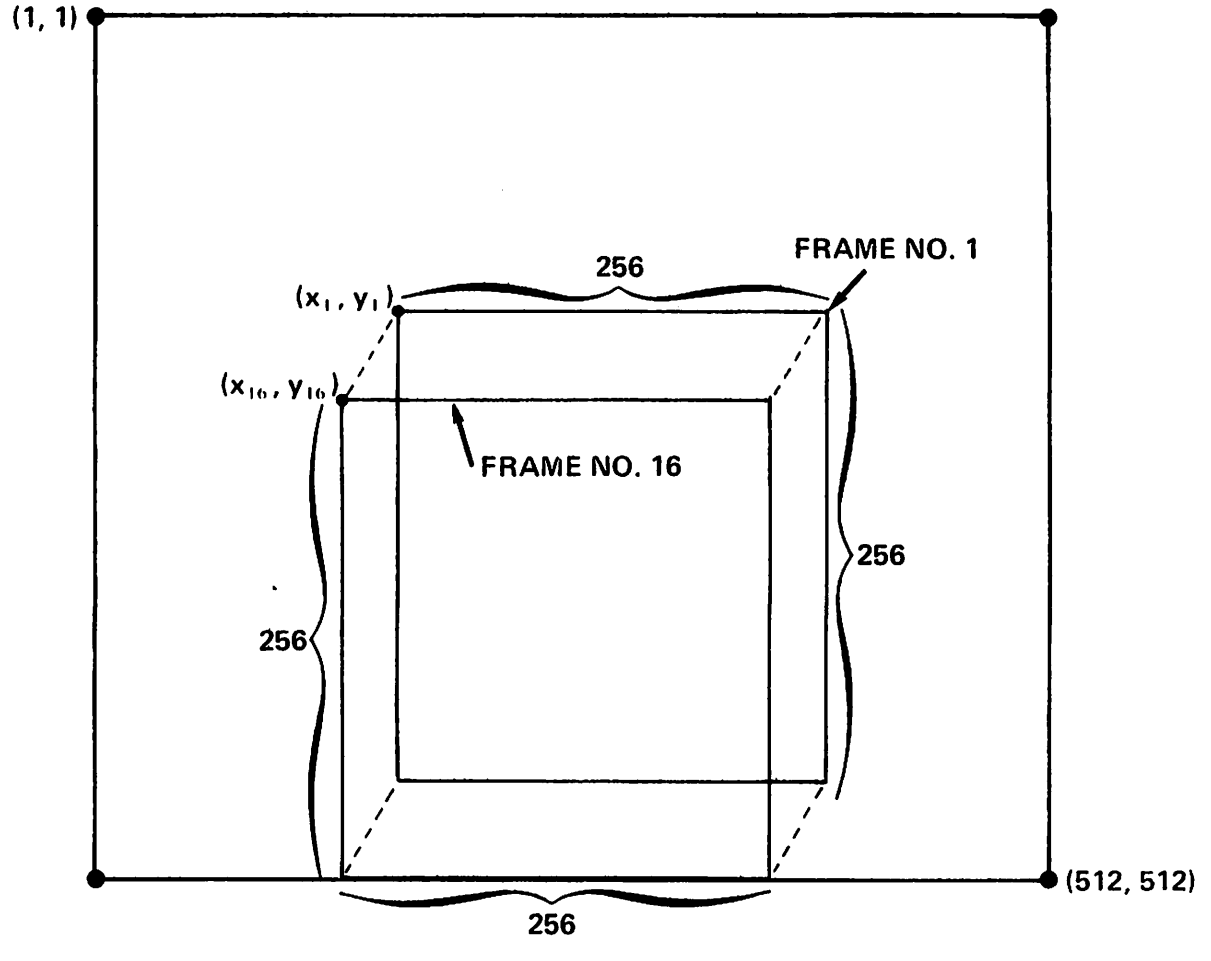


Figure 6-1. Generation of 16 Frame Data Base With Motion Correction

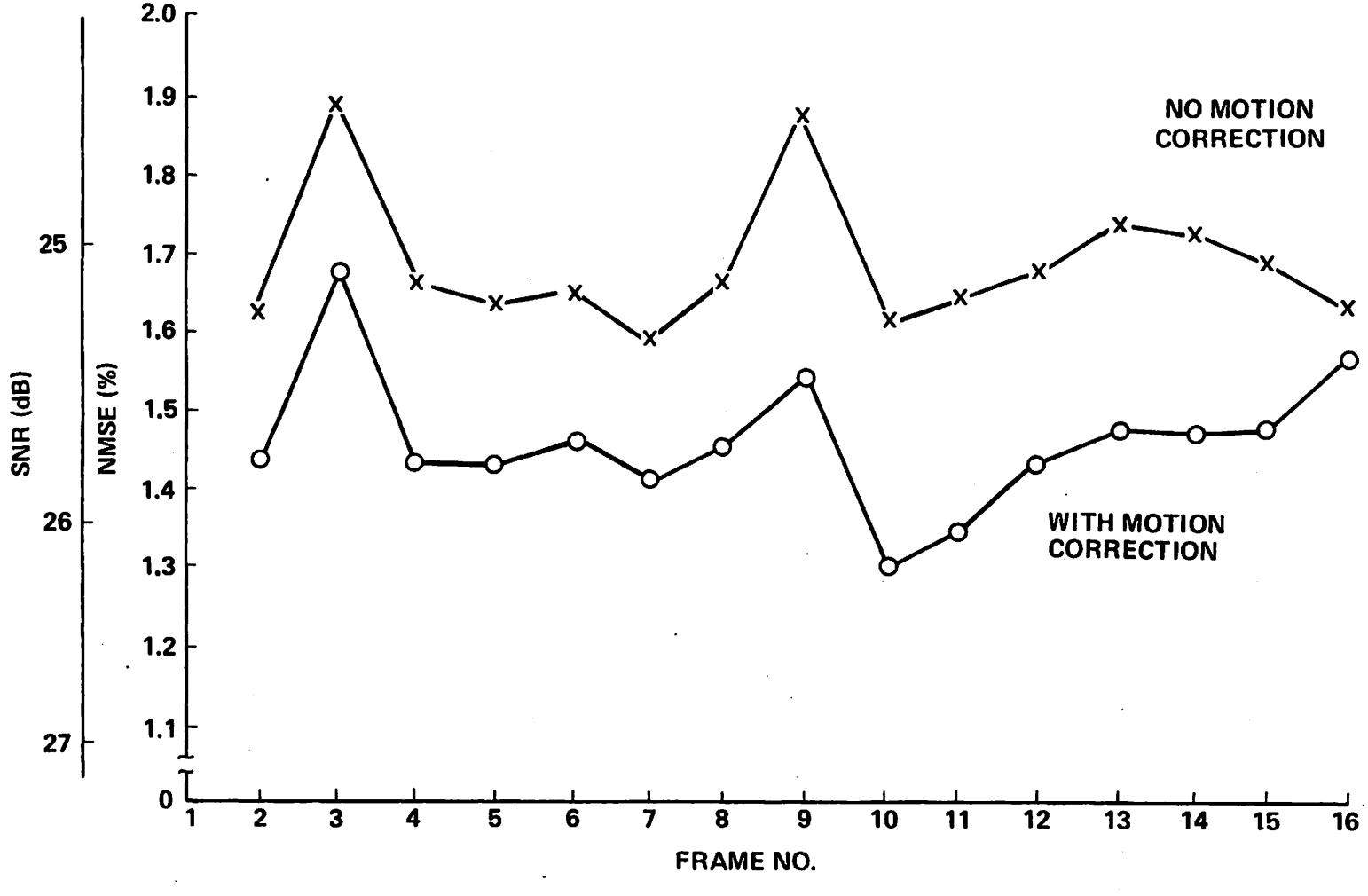


Figure 6-2. Performance of CCD Coder on "Moving Camera" Data Base Showing Effects of Motion Correction

employed in this simulation up to a 20 percent decrease in NMSE occurs as a direct result of motion correction.

SPATIAL DOMAIN MOTION COMPENSATION

One approach to motion compensation is to perform successive row and column shifts of one entire image relative to another to determine the position of maximum correlations [31]. This is a standard technique used to spatially register images and thereby compensate for translational displacements.

The simplest measure of correlation between two images, $f_1(j,k)$ and $f_2(j,k)$, is

$$r(u,v) = \frac{\sum_{j=0}\sum_{k=0} f_1(j,k)\, f_2(j-u,k-v)}{\left[\sum_{j=0}\sum_{k=0} f_1^2(j,k)\right]^{1/2}\left[\sum_{j=0}\sum_{k=0} f_2^2(j-u,k-v)\right]^{1/2}}, \tag{6.1}$$

where j,k are coordinates bounded by the spatial intersection of the images $f_1(j,k)$ and $f_2(j,k)$. Difficulties with this approach include the requirement to evaluate r(u,v) for many relative positionings between the two images and the generally ill-defined nature of the peak correlation function.

Pratt [32] proposes an image registration technique designed to provide an improved definition of the correlation peak. This method employs a statistical measure of correlation

$$R(u,v) = \frac{\sum_{j=0}\sum_{k=0} g_1(j,k)\, g_2(j-u,k-v)}{\left[\sum_{j=0}\sum_{k=0} g_1^2(j,k)\right]^{1/2}\left[\sum_{j=0}\sum_{k=0} g_2^2(j-u,k-v)\right]^{1/2}}, \tag{6.2}$$

where the quantities $g_1(j,k)$ and $g_2(j,k)$ result from convolution of the original images, $f_1(j,k)$ and $f_2(j,k)$, with filter functions selected to maximize the ratio $R(u',v')/R(u,v)$ for all $u \neq u'$, $v \neq v'$. Computation of the matrix form of the convolution operators requires calculation of the eigenvalues and eigenvectors of the covariance matrix for each of the images undergoing registration. Since the resulting image covariance matrices are large,

calculation of the required eigenvalues and eigenvectors is numerically difficult. Pratt suggests simplifying assumptions which greatly reduce the computational requirements while introducing minimal degradations in performance.

In general, statistical and standard correlation image registration techniques are of limited use for interframe coding systems. One disadvantage is that transmission of full resolution images would require transmitter processing of even larger images to ensure spatial image intersections of the proper size. Another consideration is the large number of computations required to fully define the correlation measure. Also, conventional registration techniques are not intended to facilitate local image misregistration effects, such as geometric distortions from near-field magnification or translational differences between near- and far-field objects in the field of view.

TRANSFORM DOMAIN MOTION COMPENSATION

As an alternative to motion compensation in the spatial domain using image registration techniques, a new technique is presented which allows motion compensation to be performed in the transform domain. Transform domain motion compensation is of interest as it does not result in reduced image field sizes and can easily be implemented in an adaptive manner to accommodate local image misregistration effects.

In Chapter 4, it is stated that the discrete Fourier transform possesses the shifting property, i.e., shifts in the spatial domain variables result in a multiplication of the Fourier transform of the unshifted image by an array of phase factors, $\Delta\phi(u,v,j_0,k_0)$. Thus, if an image in the spatial domain, $f(j,k,\ell)$ undergoes a translation of j_0 in the j-direction and k_0 in the k-direction between frames ℓ_1 and ℓ_2, the phase corrected Fourier transform of the shifted image in frame ℓ_2 is

$$F_{PC}(u,v,\ell_2) = F(u, v, \ell_2)\,\Delta\phi(u, v, j_0, k_0) \ , \tag{6.3}$$

where

$$\Delta\phi(u,v,j_0,k_0) = \exp\left(\frac{2\pi i}{N}(uj_0 + vk_0)\right) \ . \tag{6.4}$$

Thus, for translational shifts in the j and k directions, the shifting property of the Fourier transform results in changes to only the phase components.

For the spatially adaptive implementation of the FFD coder, the basic approach suggested for motion compensation is to compute an estimate of the frame-to-frame phase component difference arrays, $\Delta\phi(u,v,j_0,k_0)$, for each subblock. Given the set of subblock phase difference arrays, camera motion compensation can be performed by first computing the subblock two-dimensional Fourier transform, $F(u,v,\ell)$, and then multiplying by the corresponding phase correction array. The resulting subblock Fourier transform will then be corrected for camera motion relative to a given initial frame.

This implementation has the advantage of being locally adaptive to differing effects of motion that may occur at various subblock locations within the image. Corresponding disadvantages include the computational and storage requirements necessary to apply phase correction to each individual subblock. From an operational point of view, camera motion compensation must be initialized at a selected frame with phase component corrections being applied for a limited number of successive frames before being reinitialized.

In many practical situations, the geometry of the camera motion is known or can be approximated. For example, in the case of the 64-frame "moving camera" data base, the images were photographed with a camera mounted in an airborne vehicle where the camera orientation was forward-looking with a fixed downward tilt. During the 64-frame sequence, the airborne platform exhibited relatively small effects due to roll although a slight turning motion is evident. The combined effect of this geometric configuration on the generation of the 64-frame sequence is illustrated in Figure 6-3. This figure shows that the extent of vertical displacement for this sequence is greater at the bottom of the image, which shows the near-field objects, than at the top. Also, Figure 6-3 indicates the extent of top and bottom horizontal displacements. As illustrated, the magnitude of the horizontal displacements is considerably less than that of vertical distortions resulting from magnification of objects as the camera approaches. Distortion due to magnification is, in general, a nonlinear function whose effects are most pronounced for the near-field objects in the image foreground.

Based on the motion effects illustrated for the 64-frame "moving camera" data base, simplifying approximations can be introduced to reduce the computational requirements associated with estimation of subblock phase correction arrays. Since the horizontal displacements remain essentially constant for horizontal bands extending across the image, an average or typical phase correction array can be computed for each horizontal row of

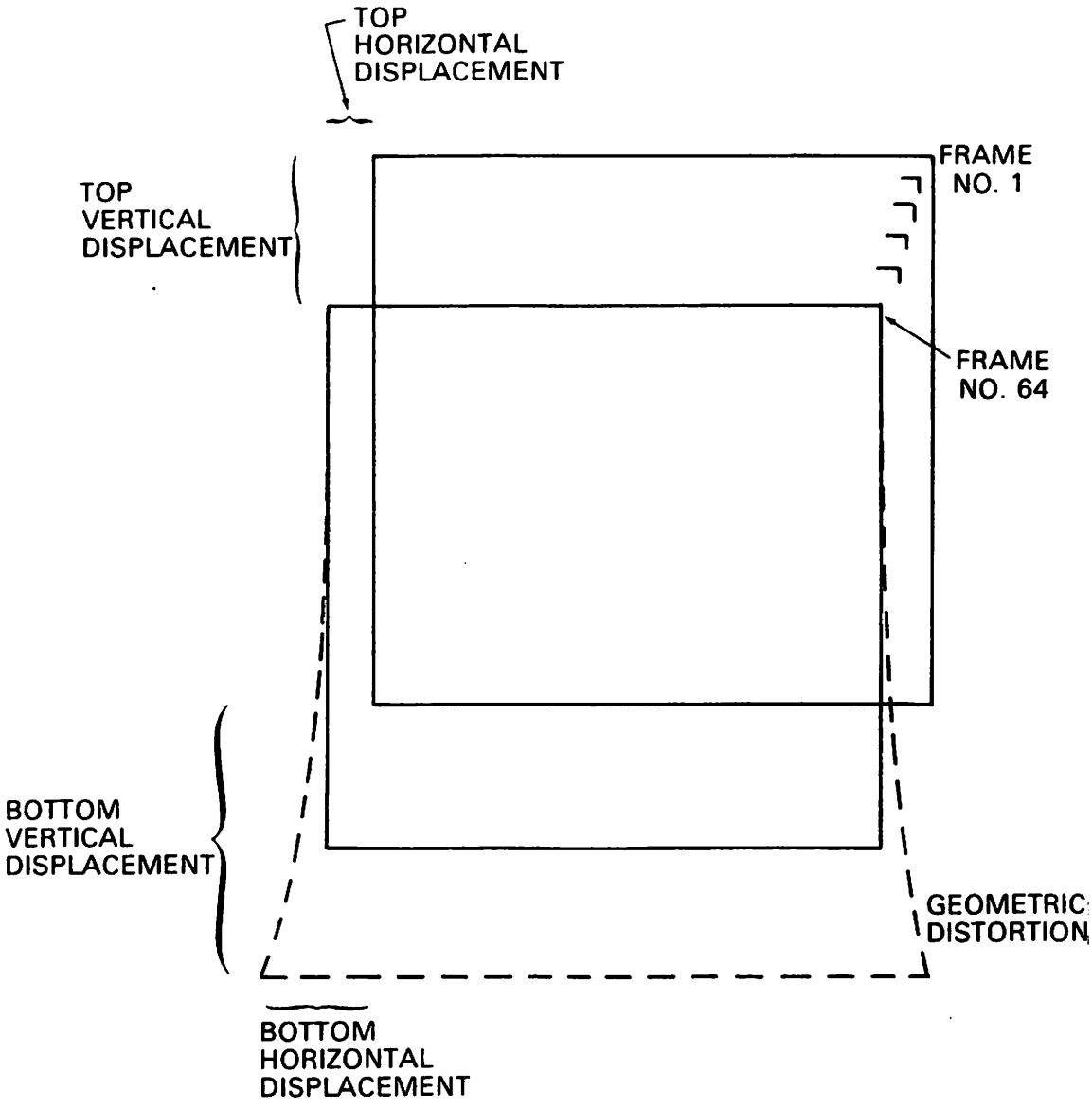


Figure 6-3. Distortions Due to Camera Motion for 64 Frame "Moving Camera" Data Base

subblocks. For 256 X 256 images with 16 X 16 subblock size, this approximation results in a 16:1 reduction in both computation and storage of phase correction arrays.

An extension of this approach is to compute a single, fixed phase correction array for use with all subblocks. Although minimizing storage requirements, a fixed array would not be locally adaptive and its use would result in motion compensation of reduced accuracy when compared with adaptive phase correction techniques.

7. SUMMARY AND CONCLUSIONS

This chapter presents a tabular summary comparing the performance and operational characteristics of the main classes of interframe coders. Although specific requirements, such as hardware implementation complexity, may make one approach preferable over another, the comparison based on the total of all criteria employed strongly favors the choice of hybrid transform/DPCM techniques for interframe coding applications.

In addition, conclusions with regard to operational characteristics of interframe hybrid coders and performance comparisons between interframe hybrid and intraframe hybrid coders are presented.

COMPARISON OF INTERFRAME CODING TECHNIQUES

In this section, comparisons are drawn between hybrid transform/DPCM, three-dimensional transform, three-dimensional DPCM, and frame replenishment interframe coders. The comparison criteria used are: coding efficiency, storage requirements, immunity to channel error, and hardware implementation complexity. A summary of the comparison results is given in Table 7-1.

In this table, coding efficiency for the interframe coders is evaluated experimentally in terms of NMSE for frame number 16 of the "head and shoulders" data base. The coding performance given is for an average bit rate of 1.0 bits/pixel/frame. This bit rate was selected as it is the lowest bit rate at which three-dimensional DPCM coders can operate. Although not implemented, the conditional frame replenishment coder was included for comparison purposes. It is anticipated that the coding efficiency for this coder would be low relative to the remaining four interfame coders. Comparison of the NMSE values indicates that the spatially adaptive hybrid CCD and FFD coder implementations are superior

Table 7-1. Comparison of Interframe Coding Techniques

CRITERIA \ INTERFRAME CODER	HYBRID CCD	HYBRID FFD	THREE-DIMENSIONAL COSINE TRANSFORM	THREE-DIMENSIONAL DPCM	CONDITIONAL FRAME REPLENISHMENT
CODING EFFICIENCY (NMSE)	0.021%	0.025%	0.045%	0.168%	LOW
MEMORY REQUIREMENTS (FRAMES×BITS/PIXEL ×PIXEL/FRAMES)	$1\times8\times256^2$ bits	$1\times8\times256^2$ bits	$L\times8\times256^2$ bits	$1^+\times8\times256^2$ bits	$(1\text{-}10)\times8\times256^2$ bits
CHANNEL ERROR IMMUNITY	GOOD	GOOD	VERY GOOD	POOR	VERY GOOD
HARDWARE COMPLEXITY (SIZE, WEIGHT AND POWER)	MODERATE	MODERATE	HIGH	SIMPLE	SIMPLE-MODERATE

in terms of NMSE coding efficiency. This conclusion is supported by subjective comparison of the coded frames illustrated in Chapters 3 and 4.

A significant disadvantage of all interframe coding systems is the requirement for storage of one or more previously scanned data frames. Of the interframe systems considered, the three-dimensional cosine transform implementation is the least attractive in terms of required storage of previous frames. This is evident when the three-dimensional cosine transform coder is compared with coder implementations which use first-order DPCM predictive coding in the temporal domain. Under the assumption of first-order predictors, only the single previous frame needs to be stored. A special case occurs for the three-dimensional DPCM implementation where, in addition to the previous frame, it is also necessary to retain the previously scanned line of the current frame.

In terms of immunity to channel error, the least sensitive coders are those employing conditional frame replenishment and three-dimensional transforms. Conversely, the DPCM encoder is the most vulnerable to channel error due to its transmission of simple pixel amplitude differences at a low bit rate. The hybrid transform/DPCM coders transmit differences in transform coefficients instead of pixel amplitudes and are less sensitive to channel error than strictly predictive coders.

The implementation complexity criterion is intended to be a coarse measure of weight, size, and power requirements for each coder. The inherently simple operation of DPCM coders combined with essentially a single frame of storage favors the DPCM interframe coder. At the other extreme, the multiple frame storage requirements of three-dimensional transform coders severely limit their usefulness.

In summary, the two hybrid transform/DPCM coders represent attractive compromises as they combine single frame storage requirements with the simplicity of DPCM operation in the temporal domain.

CONCLUSIONS

Based on theoretical and experimental results obtained during the course of this investigation, certain major conclusions have been reached. The first is that exploitation of temporal correlation in addition to spatial correlation is a viable technique for coding sequences of digital images. This fact is demonstrated by a comparison of the coding

performance for the interframe hybrid CCD coder and the intraframe cosine transform/ DPCM coder. The "head and shoulders" and motion corrected "moving camera" data bases are used for these comparisons. Figure 7-1 illustrates the results for the intraframe and interframe coders. The average pixel bit rates used give subjectively equivalent image reconstructions although the nature of the image degradations differs for the two coders. For the "head and shoulders" data base image, inclusion of temporal correlation resulted in an 8:1 reduction in average pixel bit rate. Comparison of coding performance using the "moving camera" data base shows an average pixel bit rate improvement of 4:1.

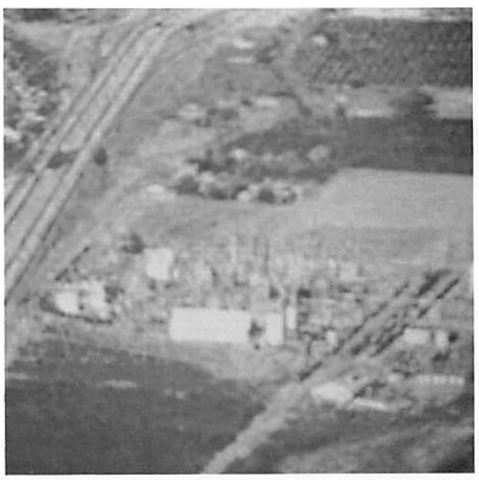
This leads to a related conclusion which is that performance of the interframe hybrid coders is heavily data dependent. In the case of the "head and shoulders" data base, superior coding performance is achieved since interframe subject movement is restricted to a relatively small portion of the image. Coding performance with the "moving camera" data base is degraded from the previous case because of the higher spatial frequency content of the coded image combined with geometric distortions and residual frame-to-frame pixel amplitude variations across the entire image. Since performance of the hybrid interframe coders is dependent on temporal correlation, reduced levels of performance are to be anticipated for image sequences distorted by the effects of camera motion.

(a) 2.0 bits/pixel

(b) 0.8 bits/pixel

Hybrid Intraframe Coder

(c) 0.25 bits/pixel/frame



(d) 0.2 bits/pixel/frame

Hybrid Interframe Coder

Figure 7-1. Comparison of Hybrid Intraframe and Hybrid Interframe Cosine Transform/DPCM Coders

APPENDIX A. BIT ASSIGNMENT PROCEDURES

In zonal coding systems, the allocation of bits is based on the variances of the transform coefficients or transform coefficient differences. An optimum mean square quantization error bit assignment algorithm for Gaussian densities has been developed by several researchers [33]--[36].

In the general three-dimensional case, the bit assignment rule is

$$b(u,v,w) = \text{integer}\left\{ \frac{B}{JKL} + 2\log_{10}[\sigma^2(u,v,w)] - \left[\frac{2}{JKL} \sum_{u=0}^{J-1} \sum_{v=0}^{K-1} \sum_{w=0}^{L-1} \log_{10} \sigma^2(u,v,w) \right] + 0.5 \right\}, \quad (A.1)$$

where $b(u,v,w)$ is the number of bits assigned at location, u,v,w, $\sigma^2(u,v,w)$ is the transform coefficient or coefficient difference variance, and

$$B = \sum_{u=0}^{J-1} \sum_{v=0}^{K-1} \sum_{w=0}^{L-1} b(u,v,w) \quad (A.2)$$

is the total number of code bits available. Simplified one- and two-dimensional bit assignments required in other coder implementations follow directly from equations A.1 and A.2.

Certain computational difficulties can arise in the application of this bit assignment rule. For example, negative bit assignments often result which must be set to zero. Also, the sum of the individual bit assignments may not be equal to B. In this case, the computed bit assignments are systematically increased or decreased until equation A.2 is satisfied.

An alternate bit assignment procedure developed by G. S. Robinson sequentially assigns individual code bits. This technique minimizes the total quantization error by equalizing the error contribution of each quantized transform coefficient. The error contributions for individual transform coefficients are proportional to $\sigma^2(u,v,w)/2^{2b(u,v,w)}$

where b(u,v,w) is the current bit allocation at location u,v,w [37]. Based on this relationship, the error equalization bit assignment procedure iteratively selects and then reduces the largest variance array value by assigning one additional bit to that location. Since the error contribution for a given transform coefficient decreases with the allocation of additional bits, repeated application produces an increasingly even distribution of the total quantization error.

This procedure is terminated when the sum of the allocated code bits equals B. Although it requires a separate search of the adjusted transform coefficient variance array before assigning each code bit, this technique avoids the computational difficulties mentioned for the bit assignment rule of equation A.1.

Tests performed with both bit assignment procedures using the same variance array result in nearly identical bit assignments.

APPENDIX B. IMAGE QUANTIZATION

The material in this appendix is summarized from Digital Image Processing by W. K. Pratt [27]. In image coding, the quantization operation requires conversion from essentially analog sequences of values into discrete representations. For intraframe or interframe transform coders, the sequence of values to be quantized represents spatially and temporally adjacent transform coefficients. However, for hybrid transform/DPCM intraframe or interframe coders, differences between spatially and temporally adjacent transform coefficients are quantized. In either case, the quantization process is illustrated graphically in Figure B-1 where the values to be quantized are compared to a set of decision levels, d_n. Each input value falls between a pair of decision levels, d_{n-1} and d_n, and is quantized to a fixed reconstruction level, r_n, which corresponds to the interval bounded by d_{n-1} and d_n. The quantized value is then assigned a binary code for transmission. In the quantization process, it is assumed that the values to be quantized, f, are samples from a scalar random process with known probability density, p(f), which are bounded by upper and lower limits a_U and a_L; thus, for a given number of quantization levels, N, the quantization process requires specification of a set of decision levels, d_n, and reconstruction levels, r_n, such that for

$$d_{n-1} \leqslant f < d_n \ , \tag{B.1}$$

f is quantized by assigning to it the reconstruction value r_n.

The criterion most often used to select values for the decision and reconstruction levels is to minimize the mean square error between the original sample value, f, and its quantized representation, f_q. For N quantization levels, the mean square quantization error, $\mathcal{E}_q$, may be written as

$$\mathcal{E}_q = E\left\{(f - f_q)^2\right\} = \sum_{n=0}^{N-1} (f-r_n)^2 \, p(f) \, df \ . \tag{B.2}$$

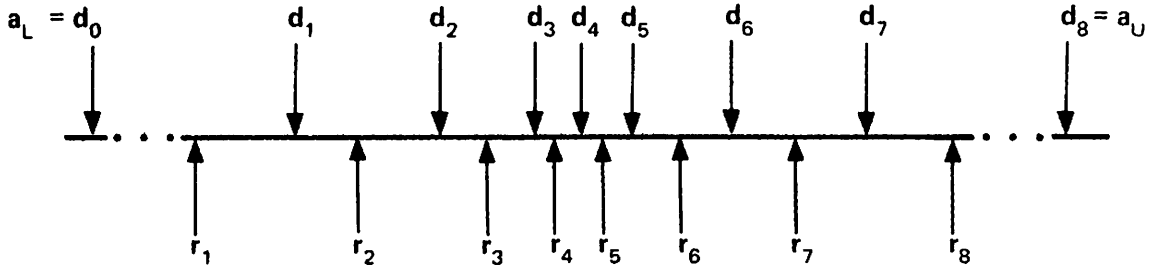


Figure B-1. Typical Quantization Decision and Reconstruction Levels for N = 8 with p(f) Symmetric

Historically, there have been numerous approaches taken to finding sets of decision and reconstruction values which minimize the MSE expression of equation B.2. One solution, originally proposed by Panter and Dite [38], assumes that the density p(f) may be reasonably approximated as a constant value over each quantization interval. For a large number of quantization levels, this approximation is acceptable as the resulting quantization intervals are small. However, it is shown in Chapters 3 and 4 that interframe coders can operate with extremely small bit allocations. Thus, direct application of the Panter and Dite quantizer is generally not appropriate for use with interframe coders.

Another approach to selection of minimum mean square quantization error decision and reconstruction levels is proposed by Max [39]. This approach produces exact solutions for the decision and reconstruction level values minimizing equation B.2, but without the restrictive assumption of constant levels for p(f) within each quantization interval. Consequently, the general Max solution is well suited for interframe coders as it is accurate for small as well as large numbers of quantization levels.

An alternate approach which is also applicable to interframe image coding is to use a companding operation [38], [40]. Companding is defined as: *com*pressing of the original non-uniformly distributed signal by applying a nonlinear transformation, performing a uniform linear quantization, and then ex*panding* the quantized output by means of the inverse nonlinear transformation.

MAX QUANTIZER

The Max quantization approach is to derive an exact solution to equation B.2. Minimization of the mean square quantization error, $\mathcal{E}_q$, is accomplished by setting the partial derivations of the mean square quantization error, with respect to the decision and reconstruction levels, to zero. Upon simplification, equations

$$d_n = \frac{r_n + r_{n+1}}{2} \tag{B.3}$$

and

$$\int_{d_{n-1}}^{d_n} (f - r_n)\, p(f)\, df = 0 \tag{B.4}$$

result for n = 0, 1, . . . , N. For a fixed number of quantization intervals, N, the decision and reconstruction levels are obtained by recursive solution of equations B.3 and B.4. In general, the set of simultaneous equations resulting from solution of these equations cannot be solved explicitly.

For the investigation of hybrid coders, it is necessary to generate normalized Max decision and reconstruction levels for transform coefficient differences having Laplacian densities. These values are generated with iterative techniques and represent an improvement in accuracy over those previously available [41]. The computed Laplacian density Max decision and reconstruction levels are listed in Table B-1.

COMPANDING QUANTIZER

Companding quantizers are also suitable for use with intraframe and interframe coders. Figure B-2 illustrates the basic companding process in which input samples f, having a nonuniform density function, undergo a nonlinear transformation Z(f) to achieve a uniform probability density function. For the Laplacian case, the nonlinear transform is

$$Z(f) = \frac{E_0 \, [1 - \exp(-mf/E_0)]}{1 - \exp(-m)} \quad , \tag{B.5}$$

where E_0 is the maximum value of f, $m = (2E_0/3\sigma)^{1/2}$, and σ^2 is the variance of the transform coefficients or coefficient differences. Given that p(Z(f)) is uniform, a linear quantization can be performed as indicated in Figure B-2. The expression for the inverse nonlinear transformation is

$$f_q = - \frac{E_0}{m} \ln \left\{ 1 - \frac{Z_q(f)}{E_0} \, [1 - \exp(-m)] \right\} \quad , \tag{B.6}$$

where $Z_q(f)$ is the quantized value of Z(f).

Table B-1. Laplacian Decision and Reconstruction Levels for Max Quantizer

No. of Bits	Decision Levels d_n	Reconstruction Levels r_n
1	$-\infty$	-0.7071
	0	0.7071
	∞	
2	$-\infty$	-1.8340
	-1.1269	-0.4198
	0	0.4198
	1.1269	1.8340
	∞	
3	$-\infty$	-3.0867
	-2.3796	-1.6725
	-1.2527	-0.8330
	-0.5332	-0.2334
	0	0.2334
	0.5332	0.8330
	1.2527	1.6725
	2.3796	3.0867
	∞	
4	$-\infty$	-4.4311
	-3.7240	-3.0169
	-2.5971	-2.1774
	-1.8776	-1.5778
	-1.3444	-1.1110
	-0.9198	-0.7287
	-0.5667	-0.4048
	-0.2644	-0.1240
	0	0.1240
	0.2644	0.4048
	0.5667	0.7287
	0.9198	1.1110
	1.3444	1.5778
	1.8776	2.1774
	2.5971	3.0169
	3.7240	4.4311
	∞	

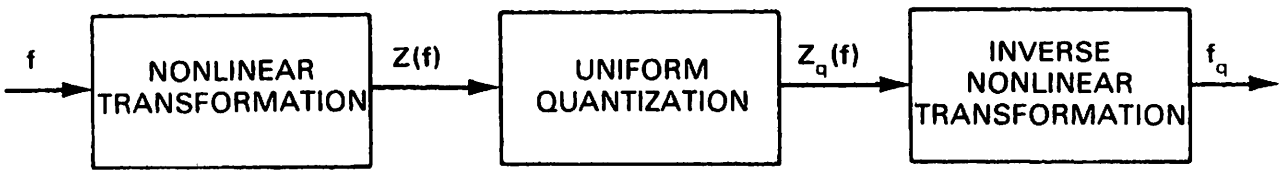


Figure B-2. Uniform Quantizer with Companding

REFERENCES

[1] A. G. Tescher, "The Role of Phase in Adaptive Image Coding," USCIPI Rept. 510, University of Southern California, Image Processing Institute, January 1974.

[2] S. C. Noble, S. C. Knauer, and J. I. Giem, "A Real-Time Hadamard Transform System for Spacial and Temporal Redundancy Reduction in Television," *Proc. Internatl. Telemetering Conf.* (Washington, D.C.), vol. 9, p. 496, October 1973.

[3] F. W. Mounts, "A Video Encoding System Using Conditional Picture-Element Replenishment," *Bell System Tech. J.*, vol. 48, no. 7, pp. 2545–2555, September 1969.

[4] J. C. Candy et al., "Transmitting Television as Clusters of Frame to Frame Differences," *Bell System Tech. J.*, vol. 50, no. 6, pp. 1889–1917, July–August 1971.

[5] T. O. Limb, "Buffering of Data Generated by the Coding of Moving Images," *Bell System Tech. J.*, vol. 51, no. 1, pp. 239–261, January 1972.

[6] B. G. Haskell, "Buffer and Channel Shaping by Several Interframe Picturephone Coders," *Bell System Tech. J.*, vol. 51, no. 1, pp. 261–291, January 1972.

[7] J. A. Roese, et al., "Interframe Transform Coding and Predictive Coding Methods," *Proc. 1975 Internatl. Conf. on Communications* (San Francisco), vol. 2, pp. 17–21, June 1975.

[8] W. K. Pratt, "Spatial Transform Coding of Color Images," *IEEE Trans. Communication Technology*, vol. COM-19, no. 6, pp. 980–992, December 1971.

[9] W. K. Pratt and H. C. Andrews, "Transform Image Coding," USCEE Rept. 387, March 1970.

[10] P. A. Wintz, "Transform Picture Coding," *Proc. IEEE*, vol. 60, no. 7, pp. 809–820, July 1972.

[11] W. K. Pratt, L. R. Welch, and W. Chen, "Slant Transforms for Image Coding," *IEEE Trans. Communications*, vol. COM-22, no. 8, pp. 1075–1093, August 1974.

[12] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete Cosine Transform," *IEEE Trans. Computers*, vol. C-23, no. 1, pp. 90–93, January 1974.

[13] W. K. Pratt, J. Kane and H. C. Andrews, "Hadamard Transform Image Coding," *Proc. IEEE*, vol. 57, pp. 58–68, January 1969.

[14] H. C. Andrews, *Computer Techniques in Image Processing.* New York: Academic Press, 1970.

[15] L. R. Rabiner, et al., "The Chirp-Z Transform Algorithm," *IEEE Trans. Audio and Electroacoustics*, vol. AD-17, pp. 86–92, June 1969.

[16] A. Habibi, et al., "Real-Time Image Redundancy Reduction Using Transform Coding Techniques," *Proc. Internl. Conf. on Communications* (Minneapolis), vol. 10, paper 18 A, pp. 18A-1–18A-8, June 1974.

[17] M. R. Winkler, "High Information Delta Modulation," *IEEE Internatl. Conv. Rec.*, part 8, pp. 260–265, 1963.

[18] M. R. Winkler, "Pictorial Transmission with HIDM," *IEEE Internatl. Conv. Rec.*, part 1, pp. 285–291, 1965.

[19] N. S. Joyant, "Adaptive Delta Modulation with One Bit Memory," *Bell System Tech. J.*, vol. 45, pp. 321–342, March 1970.

[20] A. Habibi, "Comparison of n^{th}-order DPCM Encoder with Linear Transformation and Block Quantization Techniques," *IEEE Trans. Communication Technology*, vol. COM-19, no. 6, part I, pp. 948–956, December 1971.

[21] A. J. Seyler, "Probability Distributions of Television Frame Differences," *Proc. IEEE* (Australia), pp. 355–366, November 1966.

[22] L. E. Franks, "A Model for the Random Video Process," *Bell System Tech. J.*, vol. 45, pp. 609–630, April 1966.

[23] A. Habibi and R. S. Hershel, "A Unified Representation of Differential Pulse Code Modulation (DPCM) and Transform Coding Systems," *IEEE Trans. Communications*, vol. COM-22, no. 5, pp. 692–696, May 1974.

[24] A. Habibi and G. S. Robinson, "A Survey of Digital Picture Coding," *IEEE Computer*, vol. 7, no. 5, pp. 22–34, May 1974.

[25] A. Habibi, "Hybrid Coding of Pictorial Data," *IEEE Trans. Communications*, vol. COM-22, no. 5, pp. 614–624, May 1974.

[26] W. Chen, "Slant Transform Image Coding," USCEE Rept. 441, University of Southern California, Electronic Sciences Laboratory, May 1973.

[27] W. K. Pratt, *Digital Image Processing*, in preparation.

[28] M. N. Huhns, "Optimum Restoration of Quantized Correlated Signals," USCIPI Rept. 600, University of Southern California, Image Processing Institute, August 1975.

[29] J. A. Roese and G. S. Robinson, "Combined Spatial and Temporal Coding of Digital Image Sequences," *Proc. SPIE*, vol. 66, pp. 172–180, August 1975.

[30] G. S. Robinson, "Orthogonal Transform Feasibility Study," NASA Final Rept., Contract NASA-CR-115314, N72-13143, November 1971.

[31] P. F. Anuta, "Digital Registration of Multispectral Video Imagery," *SPIE J.*, vol. 7, pp. 168–175, September 1969.

[32] W. K. Pratt, "Correlation Techniques of Image Registration," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-10, no. 3, pp. 353–358, May 1974.

[33] J. J. Y. Huang and P. M. Schultheiss, "Block Quantization of Correlated Gaussian Random Variables," *IEEE Trans. Communication Systems*, vol. CS-11, pp. 289–296, September 1963.

[34] P. J. Ready and P. A. Wintz, "Multispectral Data Compression Through Transform Coding and Block Quantization," School of Electrical Engineering, Purdue University, Technical Report TR-EE 72-29, May 1972.

[35] A. J. Kurtenback and P. A. Wintz, "Quantizing for Noisy Channels," *IEEE Trans. Communication Technology*, vol. COM-17, pp. 291–302, April 1969.

[36] A. Habibi and P. A. Wintz, "Image Coding by Linear Transformation and Block Quantization," *IEEE Trans. Communication Technology*, vol. COM-19, no. 1, pp. 50–62, February 1971.

[37] S. J. Campanella and G. S. Robinson, "A Comparison of Orthogonal Transformations for Digital Speech Processing," *IEEE Trans. Communication Technology*, vol. COM-19, no. 6, pp. 1045–1050, December 1971.

[38] P. F. Panter and W. Dite, "Quantization Distortion in Pulse-Count Modulation with Nonuniform Spacing of Levels," *Proc. IRE*, vol. 39, pp. 44–48, January 1951.

[39] J. Max, "Quantizing for Minimum Distortion," *IEEE Trans. Information Theory*, vol. IT-6, pp. 7–12, March 1960.

[40] B. Smith, "Instantaneous Companding of Quantized Signals," *Bell System Tech. J.*, vol. 46, pp. 653–709, May 1967.

[41] M. D. Paez and T. H. Glisson, "Minimum Mean-Square-Error Quantization in Speech PCM and DPCM Systems," *IEEE Trans. Communications*, vol. COM-20, pp. 225–230, April 1972.