IMAGE and VIDEO COMPRESSION for MULTIMEDIA ENGINEERING

Fundamentals, Algorithms, and Standards

Yun Q. Shi
Huifang Sun

IMAGE and VIDEO COMPRESSION for MULTIMEDIA ENGINEERING

Fundamentals, Algorithms, and Standards

IMAGE PROCESSING SERIES

Series Editor: Phillip A. Laplante

Forthcoming Titles

Adaptive Image Processing: A Computational Intelligence Perspective
Ling Guan, Hau-San Wong, and Stuart William Perry

Shape Analysis and Classification: Theory and Practice
Luciano da Fontoura Costa and Roberto Marcondes Cesar, Jr.

IMAGE and VIDEO COMPRESSION for MULTIMEDIA ENGINEERING

Fundamentals, Algorithms, and Standards

Yun Q. Shi
New Jersey Institute of Technology
Newark, NJ

Huifang Sun
Mitsubishi Electric Information Technology Center
America Advanced Television Laboratory
New Providence, NJ

CRC Press
Boca Raton London New York Washington, D.C.

Library of Congress Cataloging-in-Publication Data

Shi, Yun Q.
Image and video compression for multimedia engineering : fundamentals, algorithms, and standards / Yun Q. Shi, Huifang Sun.
p. cm.
Includes bibliographical references and index.
ISBN 0-8493-3491-8 (alk. paper)
1.Multimedia systems. 2. Image compressions. 3. Video compression.
I. Sun, Huifnag. II. Title.
QA76.575.S555 1999
006.7—dc21 99-047137
CIP

International Standard Book Number 0-8493-3491-8
Library of Congress Card Number 99-047137
Printed in the United States of America 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

Preface

It is well known that in the 1960s the advent of the semiconductor computer and the space program swiftly brought the field of digital image processing into public focus. Since then the field has experienced rapid growth and has entered into every aspect of modern technology. Since the early 1980s, digital image sequence processing has been an attractive research area because an image sequence, as a collection of images, may provide more information than a single image frame. The increased computational complexity and memory space required for image sequence processing are becoming more attainable. This is due to more advanced, achievable computational capability resulting from the continuing progress made in technologies, especially those associated with the VLSI industry and information processing.

In addition to image and image sequence processing in the digitized domain, facsimile transmission has switched from analog to digital since the 1970s. However, the concept of high definition television (HDTV) when proposed in the late 1970s and early 1980s continued to be analog. This has since changed. In the U.S., the first digital system proposal for HDTV appeared in 1990. The Advanced Television Standards Committee (ATSC), formed by the television industry, recommended the digital HDTV system developed jointly by the seven Grand Alliance members as the standard, which was approved by the Federal Communication Commission (FCC) in 1997. Today's worldwide prevailing concept of HDTV is digital. Digital television (DTV) provides the signal that can be used in computers. Consequently, the marriage of TV and computers has begun. Direct broadcasting by satellite (DBS), digital video disks (DVD), video-on-demand (VOD), video games, and other digital video related media and services are available now, or soon will be.

As in the case of image and video transmission and storage, audio transmission and storage through some media have changed from analog to digital. Examples include entertainment audio on compact disks (CD) and telephone transmission over long and medium distances. Digital TV signals, mentioned above, provide another example since they include audio signals. Transmission and storage of audio signals through some other media are about to change to digital. Examples of this include telephone transmission through local area and cable TV.

Although most signals generated from various sensors are analog in nature, the switching from analog to digital is motivated by the superiority of digital signal processing and transmission over their analog counterparts. The principal advantage of the digital signal is its robustness against various noises. Clearly, this results from the fact that only binary digits exist in digital format and it is much easier to distinguish one state from the other than to handle analog signals.

Another advantage of being digital is ease of signal manipulation. In addition to the development of a variety of digital signal processing techniques (including image, video, and audio) and specially designed software and hardware that may be well known, the following development is an example of this advantage. The digitized information format, i.e., the bitstream, often in a compressed version, is a revolutionary change in the video industry that enables many manipulations which are either impossible or very complicated to execute in analog format. For instance, video, audio, and other data can be first compressed to separate bitstreams and then combined to form a signal bitstream, thus providing a multimedia solution for many practical applications. Information from different sources and to different devices can be multiplexed and demultiplexed in terms of the bitstream. Bitstream conversion in terms of bit rate conversion, resolution conversion, and syntax conversion becomes feasible. In digital video, content-based coding, retrieval, and manipulation and the ability to edit video in the compressed domain become feasible. All system-timing signals in the digital systems can be included in the bitstream instead of being transmitted separately as in traditional analog systems.

The digital format is well suited to the recent development of modern telecommunication structures as exemplified by the Internet and World Wide Web (WWW). Therefore, we can see that digital computers, consumer electronics (including television and video games), and telecommunications networks are combined to produce an information revolution. By combining audio, video, and other data, multimedia becomes an indispensable element of modern life. While the pace and the future of this revolution cannot be predicted, one thing is certain: this process is going to drastically change many aspects of our world in the next several decades.

One of the enabling technologies in the information revolution is digital data compression, since the digitization of analog signals causes data expansion. In other words, storage and/or transmission of digitized signals require more storage space and/or bandwidth than the original analog signals.

The focus of this book is on image and video compression encountered in multimedia engineering. Fundamentals, algorithms, and standards are the three emphases of the book. It is intended to serve as a senior/graduate-level text. Its material is sufficient for a one-semester or one-quarter graduate course on digital image and video coding. For this purpose, at the end of each chapter there is a section of exercises containing problems and projects for practice, and a section of references for further reading.

Based on this book, a short course entitled "Image and Video Compression for Multimedia," was conducted at Nanyang Technological University, Singapore in March and April, 1999. The response to the short course was overwhelmingly positive.

Contents

Section I Fundamentals

Chapter 4 Transform Coding

Chapter 5 Variable-Length Coding: Information Theory Results (II)

Chapter 6

Section I

Chapter 7

Section III Motion Estimation and Compression

Section IV Video Compression

Chapter 15 Fundamentals of Digital Video Coding

Chapter 16 Digital Video Coding Standards — MPEG-1/2 Video

Chapter 17 Application Issues of MPEG-1/2 Video Coding

3 Differential Coding

Instead of encoding a signal directly, the *differential coding* technique codes the difference between the signal itself and its prediction. Therefore it is also known as *predictive coding*. By utilizing spatial and/or temporal interpixel correlation, differential coding is an efficient and yet computationally simple coding technique. In this chapter, we first describe the differential technique in general. Two components of differential coding, prediction and quantization, are discussed. There is an emphasis on (optimum) prediction, since quantization was discussed in Chapter 2. When the difference signal (also known as prediction error) is quantized, the differential coding is called differential pulse code modulation (DPCM). Some issues in DPCM are discussed, after which delta modulation (DM) as a special case of DPCM is covered. The idea of differential coding involving image sequences is briefly discussed in this chapter. More detailed coverage is presented in Sections III and IV, starting from Chapter 10. If quantization is not included, the differential coding is referred to as information-preserving differential coding. This is discussed at the end of the chapter.

3.1 INTRODUCTION TO DPCM

As depicted in Figure 2.3, a source encoder consists of the following three components: transformation, quantization, and codeword assignment. The transformation converts input into a format for quantization followed by codeword assignment. In other words, the component of transformation decides which format of input is to be encoded. As mentioned in the previous chapter, input itself is not necessarily the most suitable format for encoding.

Consider the case of monochrome image encoding. The input is usually a 2-D array of gray level values of an image obtained via PCM coding. The concept of spatial redundancy, discussed in Section 1.2.1.1, tells us that neighboring pixels of an image are usually highly correlated. Therefore, it is more efficient to encode the gray difference between two neighboring pixels instead of encoding the gray level values of each pixel. At the receiver, the decoded difference is added back to reconstruct the gray level value of the pixel. Since neighboring pixels are highly correlated, their gray level values bear a great similarity. Hence, we expect that the variance of the difference signal will be smaller than that of the original signal. Assume uniform quantization and natural binary coding for the sake of simplicity. Then we see that for the same bit rate (bits per sample) the quantization error will be smaller, i.e., a higher quality of reconstructed signal can be achieved. Or, for the same quality of reconstructed signal, we need a lower bit rate.

3.1.1 Simple Pixel-to-Pixel DPCM

Denote the gray level values of pixels along a row of an image as z_i, $i = 1,\cdots,M$, where M is the total number of pixels within the row. Using the immediately preceding pixel's gray level value, z_{i-1}, as a prediction of that of the present pixel, $\hat{z}_i$, i.e.,

$$\hat{z}_i = z_{i-1} \tag{3.1}$$

we then have the difference signal

$$d_i = z_i - \hat{z}_i = z_i - z_{i-1} \tag{3.2}$$

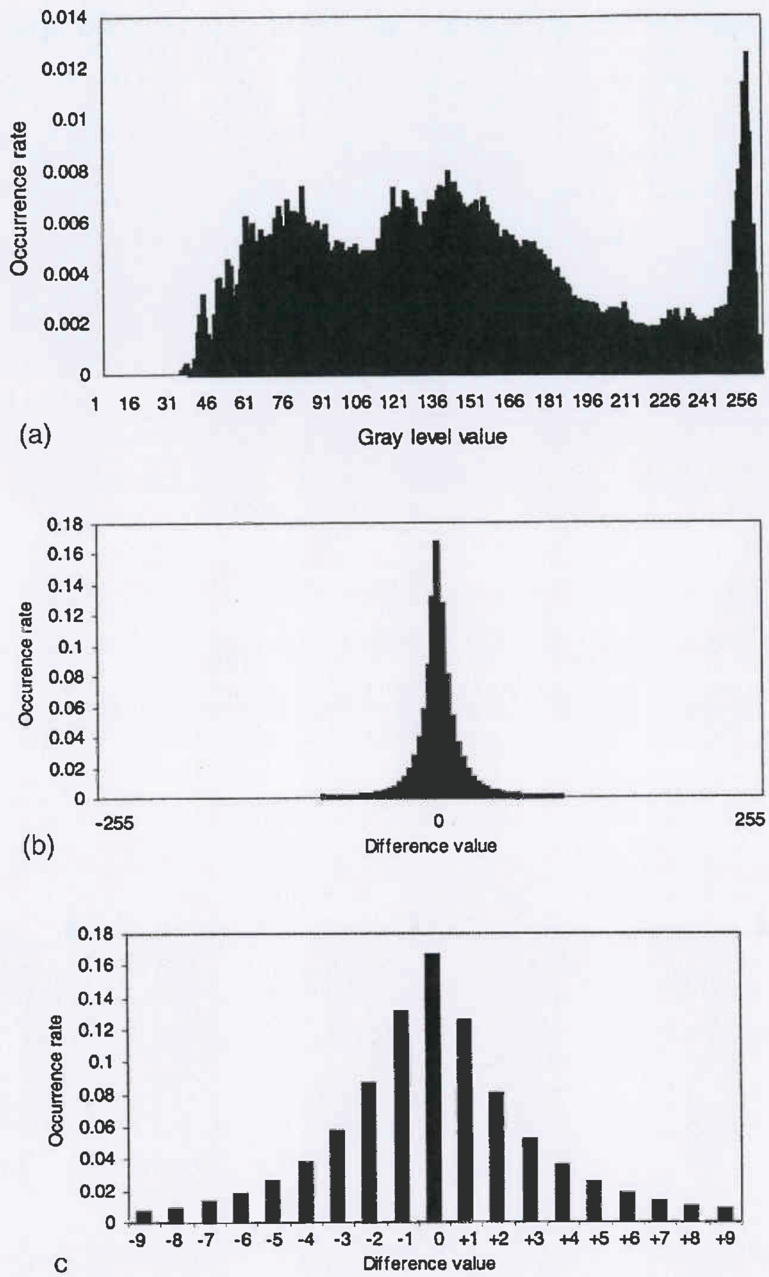

FIGURE 3.1 (a) Histogram of the original "boy and girl" image. (b) Histogram of the difference image obtained by using horizontal pixel-to-pixel differencing. (c) A close-up of the central portion of the histogram of the difference image.

Assume a bit rate of eight bits per sample in the quantization. We can see that although the dynamic range of the difference signal is theoretically doubled, from 256 to 512, the variance of the difference signal is actually much smaller. This can be confirmed from the histograms of the "boy and girl" image (refer to Figure 1.1) and its difference image obtained by horizontal pixel-to-pixel differencing, shown in Figure 3.1(a) and (b), respectively. Figure 3.1(b) and its close-up (c) indicate that by a rate of 42.44% the difference values fall into the range of −1, 0, and +1. In other words, the histogram of the difference signal is much more narrowly concentrated than that of the original signal.

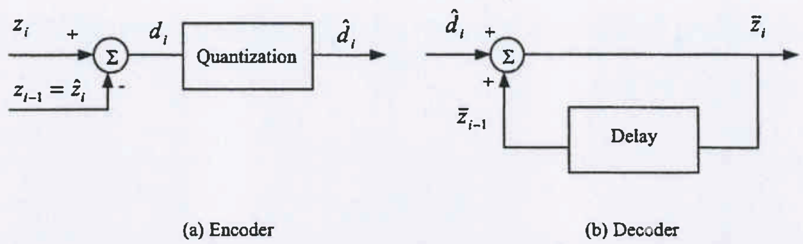

FIGURE 3.2 Block diagram of a pixel-to-pixel differential coding system.

A block diagram of the scheme described above is shown in Figure 3.2. There z_i denotes the sequence of pixels along a row, d_i is the corresponding difference signal, and $\hat{d}_i$ is the quantized version of the difference, i.e.,

$$\hat{d}_i = Q(d_i) = d_i + e_q \tag{3.3}$$

where e_q represents the quantization error. In the decoder, $\bar{z}_i$ represents the reconstructed pixel gray value, and we have

$$\bar{z}_i = \bar{z}_{i-1} + \hat{d}_i \tag{3.4}$$

This simple scheme, however, suffers from an accumulated quantization error. We can see this clearly from the following derivation (Sayood, 1996), where we assume the initial value z_0 is available for both the encoder and the decoder.

$$\begin{aligned} \text{as} \quad i = 1, \quad d_1 &= z_1 - z_0 \\ \hat{d}_1 &= d_1 + e_{q,1} \\ \bar{z}_1 = z_0 + \hat{d}_1 &= z_0 + d_1 + e_{q,1} = z_1 + e_{q,1} \end{aligned} \tag{3.5}$$

Similarly, we can have

$$as \quad i = 2, \quad \bar{z}_2 = z_2 + e_{q,1} + e_{q,2} \tag{3.6}$$

and, in general,

$$\bar{z}_i = z_i + \sum_{j=1}^{i} e_{q,j} \tag{3.7}$$

This problem can be remedied by the following scheme, shown in Figure 3.3. Now we see that in both the encoder and the decoder, the reconstructed signal is generated in the same way, i.e.,

$$\bar{z}_i = \bar{z}_{i-1} + \hat{d}_i \tag{3.8}$$

and in the encoder the difference signal changes to

$$d_i = z_i - \bar{z}_{i-1} \tag{3.9}$$

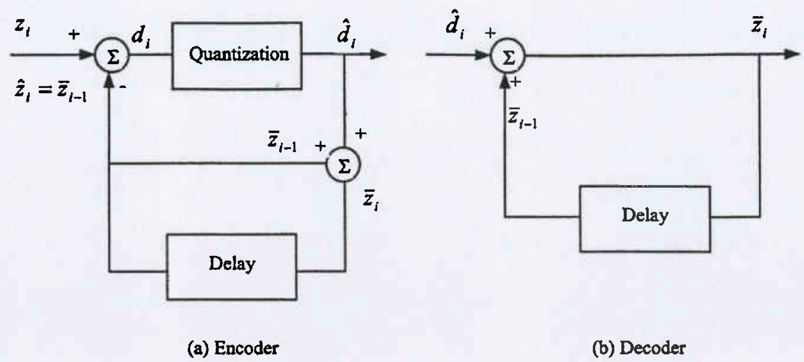

FIGURE 3.3 Block diagram of a practical pixel-to-pixel differential coding system.

Thus, the previously reconstructed $\bar{z}_{i-1}$ is used as the predictor, $\hat{z}_i$, i.e.,

$$\hat{z}_i = \bar{z}_{i-1}. \tag{3.10}$$

In this way, we have

$$\begin{aligned} as \quad i = 1, \quad d_1 &= z_1 - z_0 \\ \hat{d}_1 &= d_1 + e_{q,1} \\ \bar{z}_1 = z_0 + \hat{d}_1 &= z_0 + d_1 + e_{q,1} = z_1 + e_{q,1} \end{aligned} \tag{3.11}$$

Similarly, we have

$$\begin{aligned} as \quad i = 2, \quad d_2 &= z_2 - \bar{z}_1 \\ \hat{d}_2 &= d_2 + e_{q,2} \\ \bar{z}_2 = \bar{z}_1 + \hat{d}_2 &= z_2 + e_{q,2} \end{aligned} \tag{3.12}$$

In general,

$$\bar{z}_i = z_i + e_{q,i} \tag{3.13}$$

Thus, we see that the problem of the quantization error accumulation has been resolved by having both the encoder and the decoder work in the same fashion, as indicated in Figure 3.3, or in Equations 3.3, 3.9, and 3.10.

3.1.2 General DPCM Systems

In the above discussion, we can view the reconstructed neighboring pixel's gray value as a prediction of that of the pixel being coded. Now, we generalize this simple pixel-to-pixel DPCM. In a general DPCM system, a pixel's gray level value is first predicted from the preceding reconstructed pixels' gray level values. The difference between the pixel's gray level value and the predicted value is then quantized. Finally, the quantized difference is encoded and transmitted to the receiver. A block

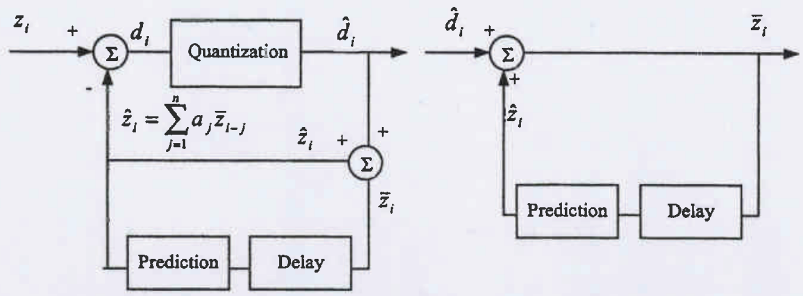

FIGURE 3.4 Block diagram of a general DPCM system.

diagram of this general differential coding scheme is shown in Figure 3.4, where the codeword assignment in the encoder and its counterpart in decoder are not included.

It is noted that, instead of using the previously reconstructed sample, $\bar{z}_{i-1}$, as a predictor, we now have the predicted version of z_i, $\hat{z}_i$, as a function of the n previously reconstructed samples, $\bar{z}_{i-1}, \bar{z}_{i-2}, \cdots, \bar{z}_{i-n}$. That is,

$$\hat{z}_i = f\left(\bar{z}_{i-1}, \bar{z}_{i-2}, \cdots, \bar{z}_{i-n}\right) \tag{3.14}$$

Linear prediction, i.e., that the function f in Equation 3.14 is linear, is of particular interest and is widely used in differential coding. In linear prediction, we have

$$\hat{z}_i = \sum_{j=1}^{n} a_j \bar{z}_{i-j} \tag{3.15}$$

where a_j are real parameters. Hence, we see that the simple pixel-to-pixel differential coding is a special case of general differential coding with linear prediction, i.e., $n = 1$ and $a_1 = 1$.

In Figure 3.4, d_i is the difference signal and is equal to the difference between the original signal, z_i, and the prediction $\hat{z}_i$. That is,

$$d_i = z_i - \hat{z}_i \tag{3.16}$$

The quantized version of d_i is denoted by $\hat{d}_i$. The reconstructed version of z_i is represented by $\bar{z}_i$, and

$$\bar{z}_i = \hat{z}_i + \hat{d}_i \tag{3.17}$$

Note that this is true for both the encoder and the decoder. Recall that the accumulation of the quantization error can be remedied by using this method.

The difference between the original input and the predicted input is called prediction error, which is denoted by e_p. That is,

$$e_p = z_i - \hat{z}_i \tag{3.18}$$

where the e_p is understood as the prediction error associated with the index i. Quantization error, e_q, is equal to the reconstruction error or coding error, e_r, defined as the difference between the original signal, z_i, and the reconstructed signal, $\bar{z}_i$, when the transmission is error free:

$$\begin{aligned} e_q &= d_i - \hat{d}_i \\ &= (z_i - \hat{z}_i) - (\bar{z}_i - \hat{z}_i) \\ &= z_i - \bar{z}_i = e_r \end{aligned} \tag{3.19}$$

This indicates that quantization error is the only source of information loss with an error-free transmission channel.

The DPCM system depicted in Figure 3.4 is also called closed-loop DPCM with feedback around the quantizer (Jayant, 1984). This term reflects the feature in DPCM structure.

Before we leave this section, let us take a look at the history of the development of differential image coding. According to an excellent early article on differential image coding (Musmann, 1979), the first theoretical and experimental approaches to image coding involving linear prediction began in 1952 at the Bell Telephone Laboratories (Oliver, 1952; Kretzmer, 1952; Harrison, 1952). The concepts of DPCM and DM were also developed in 1952 (Cutler, 1952; Dejager, 1952). Predictive coding capable of preserving information for a PCM signal was established at the Massachusetts Institute of Technology (Elias, 1955).

The differential coding technique has played an important role in image and video coding. In the international coding standard for still images, JPEG (covered in Chapter 7), we can see that differential coding is used in the lossless mode and in the DCT-based mode for coding DC coefficients. Motion-compensated (MC) coding has been a major development in video coding since the 1980s and has been adopted by all the international video coding standards such as H.261 and H.263 (covered in Chapter 19), MPEG 1 and MPEG 2 (covered in Chapter 16). MC coding is essentially a predictive coding technique applied to video sequences involving displacement motion vectors.

3.2 OPTIMUM LINEAR PREDICTION

Figure 3.4 demonstrates that a differential coding system consists of two major components: prediction and quantization. Quantization was discussed in the previous chapter. Hence, in this chapter we emphasize prediction. Below, we formulate an optimum linear prediction problem and then present a theoretical solution to the problem.

3.2.1 FORMULATION

Optimum linear prediction can be formulated as follows. Consider a discrete-time random process z. At a typical moment i, it is a random variable z_i. We have n previous observations $\bar{z}_{i-1}, \bar{z}_{i-2}, \cdots, \bar{z}_{i-n}$ available and would like to form a prediction of z_i, denoted by $\hat{z}_i$. The output of the predictor, $\hat{z}_i$, is a linear function of the n previous observations. That is,

$$\hat{z}_i = \sum_{j=1}^{n} a_j \bar{z}_{i-j} \tag{3.20}$$

with a_j, $j = 1,2,\cdots,n$ being a set of real coefficients. An illustration of a linear predictor is shown in Figure 3.5. As defined above, the prediction error, e_p, is

$$e_p = z_i - \hat{z}_i \tag{3.21}$$

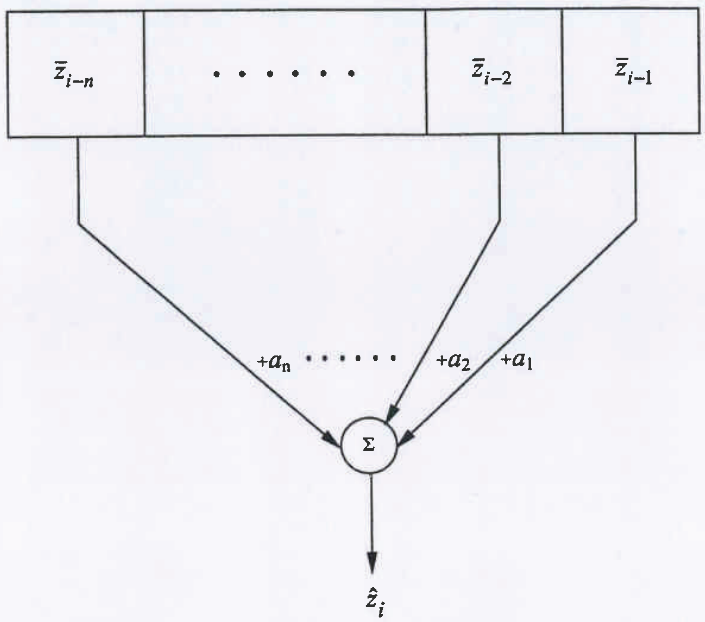

FIGURE 3.5 An illustration of a linear predictor.

The mean square prediction error, MSE_p, is

$$MSE_p = E\left[\left(e_p\right)^2\right] = E\left[\left(z_i - \hat{z}_i\right)^2\right] \tag{3.22}$$

The optimum prediction, then, refers to the determination of a set of coefficients a_j, $j = 1,2,\cdots,n$ such that the mean square prediction error, MSE_p, is minimized.

This optimization problem turns out to be computationally intractable for most practical cases due to the feedback around the quantizer shown in Figure 3.4, and the nonlinear nature of the quantizer. Therefore, the optimization problem is solved in two separate stages. That is, the best linear predictor is first designed ignoring the quantizer. Then, the quantizer is optimized for the distribution of the difference signal (Habibi, 1971). Although the predictor thus designed is suboptimal, ignoring the quantizer in the optimum predictor design allows us to substitute the reconstructed $\bar{z}_{i-j}$ by z_{i-j} for $j = 1,2,\cdots,n$, according to Equation 3.20. Consequently, we can apply the theory of optimum linear prediction to handle the design of the optimum predictor as shown below.

3.2.2 Orthogonality Condition and Minimum Mean Square Error

By taking the differentiation of MSE_p with respect to coefficient a_js, one can derive the following necessary conditions, which are usually referred to as the *orthogonality condition*:

$$E\left[e_p \cdot z_{i-j}\right] = 0 \quad for \quad j = 1,2,\cdots,n \tag{3.23}$$

The interpretation of Equation 3.23 is that the prediction error, e_p, must be orthogonal to all the observations, which are now the preceding samples: z_{i-j}, $j = 1,2,\cdots,n$ according to our discussion in Section 3.2.1. These are equivalent to

$$R_z(m)=\sum_{j=1}^{n} a_j R_z(m-j) \quad for \quad m=1,2,\cdots,n \tag{3.24}$$

where R_z represents the autocorrelation function of z. In a vector-matrix format, the above orthogonal conditions can be written as

$$\begin{bmatrix} R_z(1) \\ R_z(2) \\ \vdots \\ \vdots \\ R_z(n) \end{bmatrix} = \begin{bmatrix} R_z(0) & R_z(1) & \cdots & \cdots & R_z(n-1) \\ R_z(1) & R_z(0) & \cdots & \cdots & R_z(n-2) \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ R_z(n-1) & R_z(n) & \cdots & \cdots & R_z(0) \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_n \end{bmatrix} \tag{3.25}$$

Equations 3.24 and 3.25 are called Yule-Walker equations.

The minimum mean square prediction error is then found to be

$$MSE_p = R_z(0) - \sum_{j=1}^{n} a_j R_z(j) \tag{3.26}$$

These results can be found in texts dealing with random processes, e.g., in (Leon-Garcia, 1994).

3.2.3 Solution to Yule-Walker Equations

Once autocorrelation data are available, the Yule-Walker equation can be solved by matrix inversion. A recursive procedure was developed by Levinson to solve the Yule-Walker equations (Leon-Garcia, 1993). When the number of previous samples used in the linear predictor is large, i.e., the dimension of the matrix is high, the Levinson recursive algorithm becomes more attractive. Note that in the field of image coding the autocorrelation function of various types of video frames is derived from measurements (O'Neal, 1966; Habibi, 1971).

3.3 SOME ISSUES IN THE IMPLEMENTATION OF DPCM

Several related issues in the implementation of DPCM are discussed in this section.

3.3.1 Optimum DPCM System

Since DPCM consists mainly of two parts, prediction and quantization, its optimization should not be carried out separately. The interaction between the two parts is quite complicated, however, and thus combined optimization of the whole DPCM system is difficult. Fortunately, with the mean square error criterion, the relation between quantization error and prediction error has been found:

$$MSE_q \approx \frac{9}{2N^2} MSE_p \tag{3.27}$$

where N is the total number of reconstruction levels in the quantizer (O'Neal, 1966; Musmann, 1979). That is, the mean square error of quantization is approximately proportional to the mean square error of prediction. With this approximation, we can optimize the two parts separately, as mentioned in Section 3.2.1. While the optimization of quantization was addressed in Chapter 2, the optimum predictor was discussed in Section 3.2. A large amount of work has been done on this subject. For instance, the optimum predictor for color image coding was designed and tested in (Pirsch and Stenger, 1977).

3.3.2 1-D, 2-D, AND 3-D DPCM

In Section 3.1.2, we expressed linear prediction in Equation 3.15. However, so far we have not really discussed how to predict a pixel's gray level value by using its neighboring pixels' coded gray level values.

Recall that a practical pixel-to-pixel differential coding system was discussed in Section 3.1.1. There, the reconstructed intensity of the immediately preceding pixel along the same scan line is used as a prediction of the pixel intensity being coded. This type of differential coding is referred to as 1-D DPCM. In general, 1-D DPCM may use the reconstructed gray level values of more than one of the preceding pixels within the same scan line to predict that of a pixel being coded. By far, however, the immediately preceding pixel in the same scan line is most frequently used in 1-D DPCM. That is, pixel A in Figure 3.6 is often used as a prediction of pixel Z, which is being DPCM coded.

Sometimes in DPCM image coding, both the decoded intensity values of adjacent pixels within the same scan line and the decoded intensity values of neighboring pixels in different scan lines are involved in the prediction. This is called 2-D DPCM. A typical pixel arrangement in 2-D predictive coding is shown in Figure 3.6. Note that the pixels involved in the prediction are restricted to be either in the lines above the line where the pixel being coded, Z, is located or on the left-hand side of pixel Z if they are in the same line. Traditionally, a TV frame is scanned from top to bottom and from left to right. Hence, the above restriction indicates that only those pixels, which have been coded and available in both the transmitter and the receiver, are used in the prediction. In 2-D system theory, this support is referred to as recursively computable (Bose, 1982). An often-used 2-D prediction involves pixels A, D, and E.

Obviously, 2-D predictive coding utilizes not only the spatial correlation existing within a scan line but also that existing in neighboring scan lines. In other words, the spatial correlation is utilized both horizontally and vertically. It was reported that 2-D predictive coding outperforms 1-D predictive coding by decreasing the prediction error by a factor of two, or equivalently, 3dB in *SNR*. The improvement in subjective assessment is even larger (Musmann, 1979). Furthermore, the transmission error in 2-D predictive image coding is much less severe than in 1-D predictive image coding. This is discussed in Section 3.6.

In the context of image sequences, neighboring pixels may be located not only in the same image frame but also in successive frames. That is, neighboring pixels along the time dimension are also involved. If the prediction of a DPCM system involves three types of neighboring pixels: those along the same scan line, those in the different scan lines of the same image frame, and those

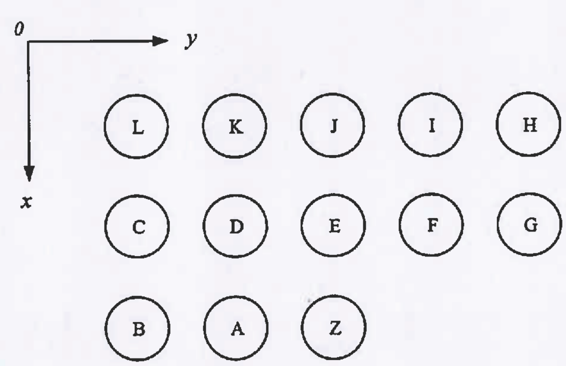

FIGURE 3.6 Pixel arrangement in 1-D and 2-D prediction.

in the different frames, the DPCM is then called 3-D differential coding. It will be discussed in Section 3.5.

3.3.3 Order of Predictor

The number of coefficients in the linear prediction, n, is referred to as the order of the predictor. The relation between the mean square prediction error, MSE_p, and the order of the predictor, n, has been studied. As shown in Figure 3.7, the MSE_p decreases quite effectively as n increases, but the performance improvement becomes negligible as $n > 3$ (Habibi, 1971).

3.3.4 Adaptive Prediction

Adaptive DPCM means adaptive prediction and adaptive quantization. As adaptive quantization was discussed in Chapter 2, here we discuss adaptive prediction only.

Similar to the discussion on adaptive quantization, adaptive prediction can be done in two different ways: forward adaptive and backward adaptive prediction. In the former, adaptation is based on the input of a DPCM system, while in the latter, adaptation is based on the output of the DPCM. Therefore, forward adaptive prediction is more sensitive to changes in local statistics. Prediction parameters (the coefficients of the predictor), however, need to be transmitted as side information to the decoder. On the other hand, quantization error is involved in backward adaptive prediction. Hence, the adaptation is less sensitive to local changing statistics. But, it does not need to transmit side information.

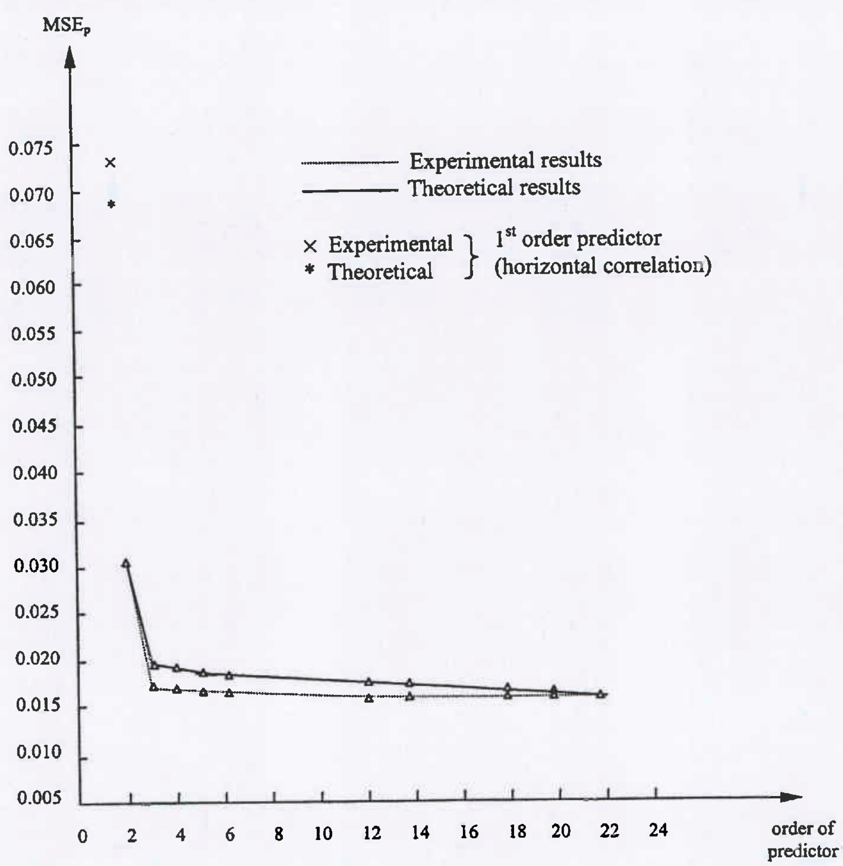

FIGURE 3.7 Mean square prediction error vs. order of predictor. (Data from Habibi, 1971.)

In either case, the data (either input or output) have to be buffered. Autocorrelation coefficients are analyzed, based on which prediction parameters are determined.

3.3.5 Effect of Transmission Errors

Transmission errors caused by channel noise may reverse the binary bit information from 0 to 1 or 1 to 0 with what is known as *bit error probability*, or *bit error rate*. The effect of transmission errors on reconstructed images varies depending on different coding techniques.

In the case of the PCM-coding technique, each pixel is coded independently of the others. Therefore bit reversal in the transmission only affects the gray level value of the corresponding pixel in the reconstructed image. It does not affect other pixels in the reconstructed image.

In DPCM, however, the effect caused by transmission errors becomes more severe. Consider a bit reversal occurring in transmission. It causes error in the corresponding pixel. But, this is not the end of the effect. The affected pixel causes errors in reconstructing those pixels towards which the erroneous gray level value was used in the prediction. In this way, the transmission error propagates.

Interestingly, it is reported that the error propagation is more severe in 1-D differential image coding than in 2-D differential coding. This may be explained as follows. In 1-D differential coding, usually only the immediate preceding pixel in the same scan line is involved in prediction. Therefore, an error will be propagated along the scan line until the beginning of the next line, where the pixel gray level value is reinitialized. In 2-D differential coding, the prediction of a pixel gray level value depends not only on the reconstructed gray level values of pixels along the same scan line but also on the reconstructed gray level values of the vertical neighbors. Hence, the effect caused by a bit reversal transmission error is less severe than in the 1-D differential coding.

For this reason, the bit error rate required by DPCM coding is lower than that required by PCM coding. For instance, while a bit error rate less than $5 \cdot 10^{-6}$ is normally required for PCM to provide broadcast TV quality, for the same application a bit error rate less than 10^{-7} and 10^{-9} are required for DPCM coding with 2-D prediction and 1-D prediction, respectively (Musmann, 1979).

Channel encoding with an error correction capability was applied to lower the bit error rate. For instance, to lower the bit error rate from the order of 10^{-6} to the order of 10^{-9} for DPCM coding with 1-D prediction, an error correction by adding 3% redundancy in channel coding has been used (Bruders, 1978).

3.4 DELTA MODULATION

Delta modulation (DM) is an important, simple, special case of DPCM, as discussed above. It has been widely applied and is thus an important coding technique in and of itself.

The above discussion and characterization of DPCM systems are applicable to DM systems. This is because DM is essentially a special type of DPCM, with the following two features.

1. The linear predictor is of the first order, with the coefficient a_1 equal to 1.
2. The quantizer is a one-bit quantizer. That is, depending on whether the difference signal is positive or negative, the output is either $+\Delta/2$ or $-\Delta/2$.

To perceive these two features, let us take a look at the block diagram of a DM system and the input-output characteristic of its one-bit quantizer, shown in Figures 3.8 and 3.9, respectively. Due to the first feature listed above, we have:

$$\hat{z}_i = \bar{z}_{i-1} \tag{3.28}$$

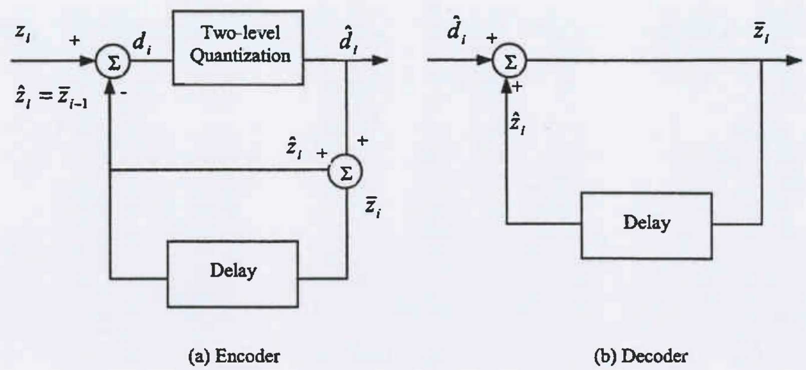

FIGURE 3.8 Block diagram of DM systems.

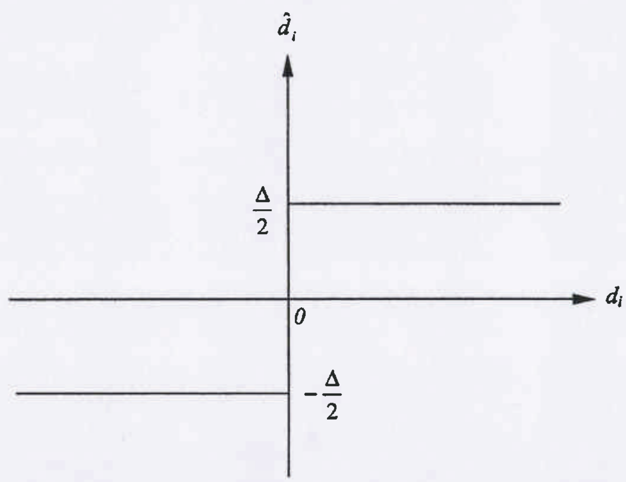

FIGURE 3.9 Input-output characteristic of two-level quantization in DM.

Next, we see that there are only two reconstruction levels in quantization because of the second feature. That is,

$$\hat{d}_i = \begin{cases} +\Delta/2 & if \quad z_i > \bar{z}_{i-1} \\ -\Delta/2 & if \quad z_i < \bar{z}_{i-1} \end{cases} \tag{3.29}$$

From the relation between the reconstructed value and the predicted value of DPCM, discussed above, and the fact that DM is a special case of DPCM, we have

$$\bar{z}_i = \hat{z}_i + \hat{d}_i \tag{3.30}$$

Combining Equations 3.28, 3.29, and 3.30, we have

$$\bar{z}_i = \begin{cases} \bar{z}_{i-1} + \Delta/2 & if \quad z_i > \bar{z}_{i-1} \\ \bar{z}_{i-1} - \Delta/2 & if \quad z_i < \bar{z}_{i-1} \end{cases} \tag{3.31}$$

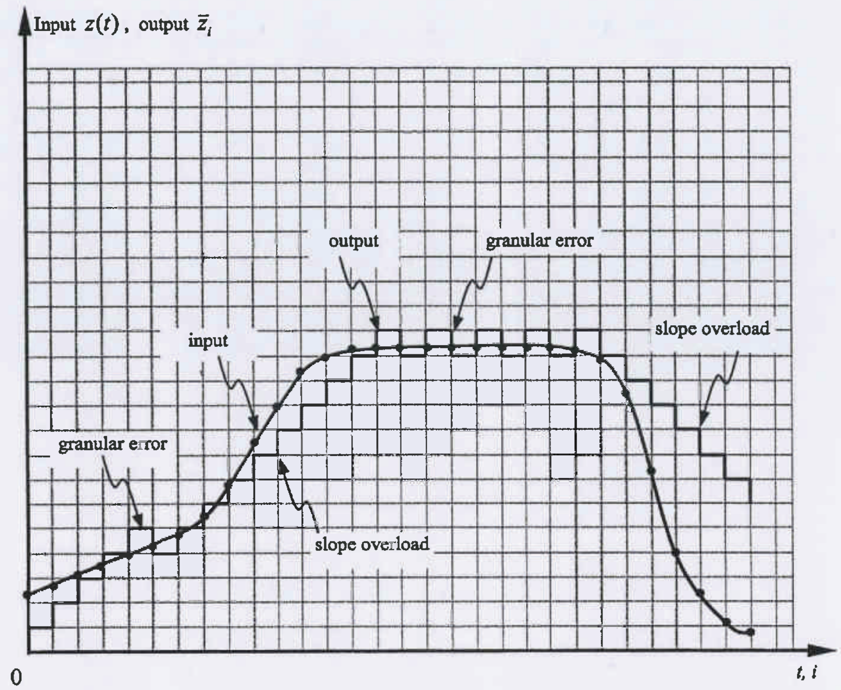

FIGURE 3.10 DM with fixed step size.

The above mathematical relationships are of importance in understanding DM systems. For instance, Equation 3.31 indicates that the step size Δ of DM is a crucial parameter. We notice that a large step size compared with the magnitude of the difference signal causes granular error, as shown in Figure 3.10. Therefore, in order to reduce the granular error we should choose a small step size. On the other hand, a small step size compared with the magnitude of the difference signal will lead to the overload error discussed in Chapter 2 for quantization. Since in DM systems it is the difference signal that is quantized, the overload error in DM becomes *slope overload* error, as shown in Figure 3.10. That is, it takes a while (multiple steps) for the reconstructed samples to catch up with the sudden change in input. Therefore, the step size should be large in order to avoid the slope overload. Considering these two conflicting factors, a proper compromise in choosing the step size is common practice in DM.

To improve the performance of DM, an oversampling technique is often applied. That is, the input is oversampled prior to the application of DM. By oversampling, we mean that the sampling frequency is higher than the sampling frequency used in obtaining the original input signal. The increased sample density caused by oversampling decreases the magnitude of the difference signal. Consequently, a relatively small step size can be used so as to decrease the granular noise without increasing the slope overload error. Thus, the resolution of the DM-coded image is kept the same as that of the original input (Jayant, 1984; Lim, 1990).

To achieve better performance for changing inputs, an adaptive technique can be applied in DM. That is, either input (forward adaptation) or output (backward adaptation) data are buffered and the data variation is analyzed. The step size is then chosen accordingly. If it is forward adaptation, side information is required for transmission to the decoder. Figure 3.11 demonstrates step size adaptation. We see the same input as that shown in Figure 3.10. But, the step size is now not fixed. Instead, the step size is adapted according to the varying input. When the input changes with a large slope, the step size increases to avoid the slope overload error. On the other hand, when the input changes slowly, the step size decreases to reduce the granular error.

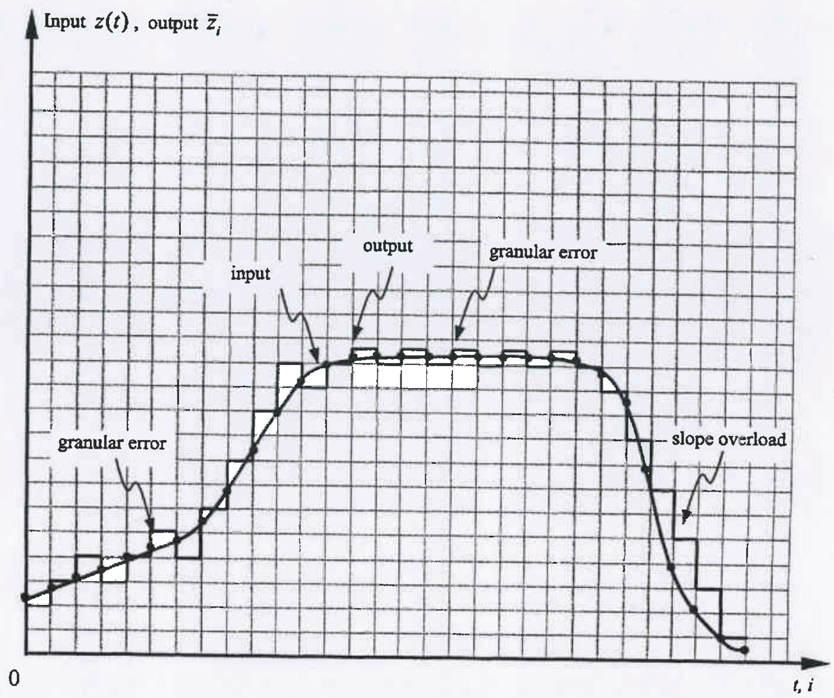

FIGURE 3.11 Adaptive DM.

3.5 INTERFRAME DIFFERENTIAL CODING

As was mentioned in Section 3.3.2, 3-D differential coding involves an image sequence. Consider a sensor located in 3-D world space. For instance, in applications such as videophony and videoconferencing, the sensor is fixed in position for a while and it takes pictures. As time goes by, the images form a temporal image sequence. The coding of such an image sequence is referred to as interframe coding. The subject of image sequence and video coding is addressed in Sections III and IV. In this section, we briefly discuss how differential coding is applied to interframe coding.

3.5.1 CONDITIONAL REPLENISHMENT

Recognizing the great similarity between consecutive TV frames, a conditional replenishment coding technique was proposed and developed (Mounts, 1969). It was regarded as one of the first real demonstrations of interframe coding exploiting interframe redundancy (Netravali and Robbins, 1979).

In this scheme, the previous frame is used as a reference for the present frame. Consider a pair of pixels: one in the previous frame, the other in the present frame — both occupying the same spatial position in the frames. If the gray level difference between the pair of pixels exceeds a certain criterion, then the pixel is considered a *changing* pixel. The present pixel gray level value and its position information are transmitted to the receiving side, where the pixel is replenished. Otherwise, the pixel is considered *unchanged*. At the receiver its previous gray level is repeated. A block diagram of conditional replenishment is shown in Figure 3.12. There, a frame memory unit in the transmitter is used to store frames. The differencing and thresholding of corresponding pixels in two consecutive frames can then be conducted there. A buffer in the transmitter is used to smooth the transmission data rate. This is necessary because the data rate varies from region to region within an image frame and from frame to frame within an image sequence. A buffer in the receiver is needed for a similar consideration. In the frame memory unit, the replenishment is

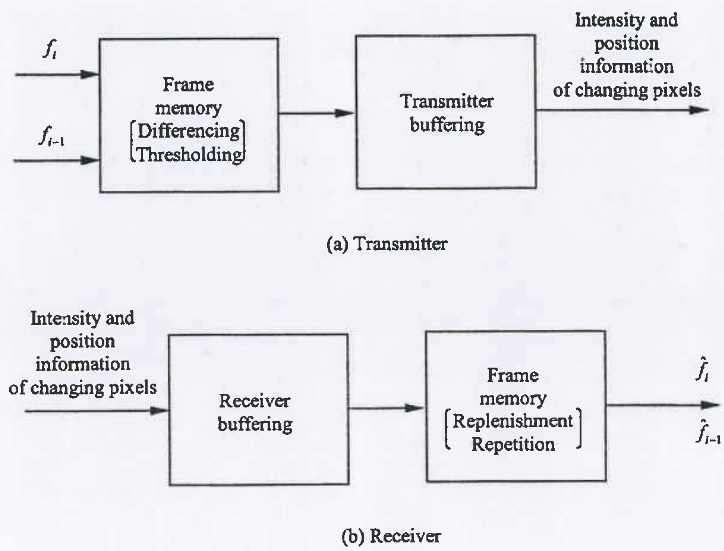

FIGURE 3.12 Block diagram of conditional replenishment.

carried out for the changing pixels and the gray level values in the receiver are repeated for the unchanged pixels.

With conditional replenishment, a considerable savings in bit rate was achieved in applications such as videophony, videoconferencing, and TV broadcasting. Experiments in real time, using the head-and-shoulder view of a person in animated conversation as the video source, demonstrated an average bit rate of 1 bit/pixel with a quality of reconstructed video comparable with standard 8 bit/pixel PCM transmission (Mounts, 1969). Compared with pixel-to-pixel 1-D DPCM, the most popularly used coding technique at the time, the conditional replenishment technique is more efficient due to the exploitation of high interframe redundancy. As pointed in (Mounts, 1969), there is more correlation between television pixels along the frame-to-frame temporal dimension than there is between adjacent pixels within a signal frame. That is, the temporal redundancy is normally higher than spatial redundancy for TV signals.

Tremendous efforts have been made to improve the efficiency of this rudimentary technique. For an excellent review, readers are referred to (Haskell et al., 1972, 1979). 3-D DPCM coding is among the improvements and is discussed next.

3.5.2 3-D DPCM

It was soon realized that it is more efficient to transmit the gray level difference than to transmit the gray level itself, resulting in interframe differential coding. Furthermore, instead of treating each pixel independently of its neighboring pixels, it is more efficient to utilize spatial redundancy as well as temporal redundancy, resulting in 3-D DPCM.

Consider two consecutive TV frames, each consisting of an odd and an even field. Figure 3.13 demonstrates the small neighborhood of a pixel, Z, in the context. As with the 1-D and 2-D DPCM discussed before, the prediction can only be based on the previously encoded pixels. If the pixel under consideration, Z, is located in the even field of the present frame, then the odd field of the present frame and both odd and even fields of the previous frame are available. As mentioned in Section 3.3.2, it is assumed that in the even field of the present frame, only those pixels in the lines above the line where pixel Z lies and those pixels left of the Z in the line where Z lies are used for prediction.

Table 3.1 lists several utilized linear prediction schemes. It is recognized that the case of *element difference* is a 1-D predictor since the immediately preceding pixel is used as the predictor. The

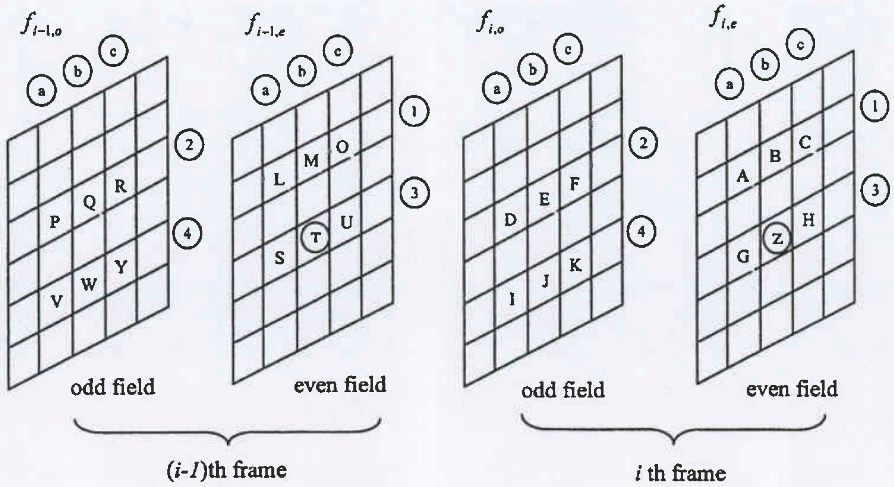

FIGURE 3.13 Pixel arrangement in two TV frames. (After Haskell, 1979.)

field difference is defined as the arithmetic average of two immediately vertical neighboring pixels in the previous odd field. Since the odd field is generated first, followed by the even field, this predictor cannot be regarded as a pure 2-D predictor. Instead, it should be considered a 3-D predictor. The remaining cases are all 3-D predictors. One thing is common in all the cases: the gray levels of pixels used in the prediction have already been coded and thus are available in both the transmitter and the receiver. The prediction error of each changing pixel Z identified in thresholding process is then quantized and coded.

An analysis of the relationship between the entropy of moving areas (bits per changing pixel) and the speed of the motion (pixels per frame interval) in an image containing a moving mannequin's head was studied with different linear predictions, as listed in Table 3.1 in Haskell (1979). It was found that the element difference of field difference generally corresponds to the lowest entropy, meaning that this prediction is the most efficient. The frame difference and element difference correspond to higher entropy. It is recognized that, in the circumstances, transmission error will be propagated if the pixels in the previous line are used in prediction (Connor, 1973). Hence, the linear predictor should use only pixels from the same line or the same line in the previous frame when bit reversal error in transmission needs to be considered. Combining these two factors, the element difference of frame difference prediction is preferred.

TABLE 3.1
Some Linear Prediction Schemes. (After Haskell, 1979).

	Original signal (Z)	Prediction signal ($\hat{Z}$)	Differential signal (d_z)
Element difference	Z	G	$Z-G$
Field difference	Z	$\frac{E+J}{2}$	$Z-\frac{E+J}{2}$
Frame difference	Z	T	$Z-T$
Element difference of frame difference	Z	$T+G-S$	$(Z-G)-(T-S)$
Line difference of frame difference	Z	$T+B-M$	$(Z-B)-(T-M)$
Element difference of field difference	Z	$T+\frac{E+J}{2}-\frac{Q+W}{2}$	$\left(Z-\frac{E+J}{2}\right)-\left(T-\frac{Q+W}{2}\right)$

3.5.3 Motion-Compensated Predictive Coding

When frames are taken densely enough, changes in successive frames can be attributed to the motion of objects during the interval between frames. Under this assumption, if we can analyze object motion from successive frames, then we should be able to predict objects in the next frame based on their positions in the previous frame and the estimated motion. The difference between the original frame and the predicted frame thus generated and the motion vectors are then quantized and coded. If the motion estimation is accurate enough, the motion-compensated prediction error can be smaller than 3-D DPCM. In other words, the variance of the prediction error will be smaller, resulting in more efficient coding. Take motion into consideration — this differential technique is called motion compensated predictive coding. This has been a major development in image sequence coding since the 1980s. It has been adopted by all international video coding standards. A more detailed discussion is provided in Chapter 10.

3.6 INFORMATION-PRESERVING DIFFERENTIAL CODING

As emphasized in Chapter 2, quantization is not reversible in the sense that it causes permanent information loss. The DPCM technique, discussed above, includes quantization, and hence is lossy coding. In applications such as those involving scientific measurements, information preservation is required. In this section, the following question is addressed: under these circumstances, how should we apply differential coding in order to reduce the bit rate while preserving information?

Figure 3.14 shows a block diagram of information-preserving differential coding. First, we see that there is no quantizer. Therefore, the irreversible information loss associated with quantization does not exist in this technique. Second, we observe that prediction and differencing are still used. That is, the differential (predictive) technique still applies. Hence it is expected that the variance of the difference signal is smaller than that of the original signal, as explained in Section 3.1. Consequently, the higher-peaked histograms make coding more efficient. Third, an efficient lossless coder is utilized. Since quantizers cannot be used here, PCM with natural binary coding is not used here. Since the histogram of the difference signal is narrowly concentrated about its mean, lossless coding techniques such as an efficient Huffman coder (discussed in Chapter 5) is naturally a suitable choice here.

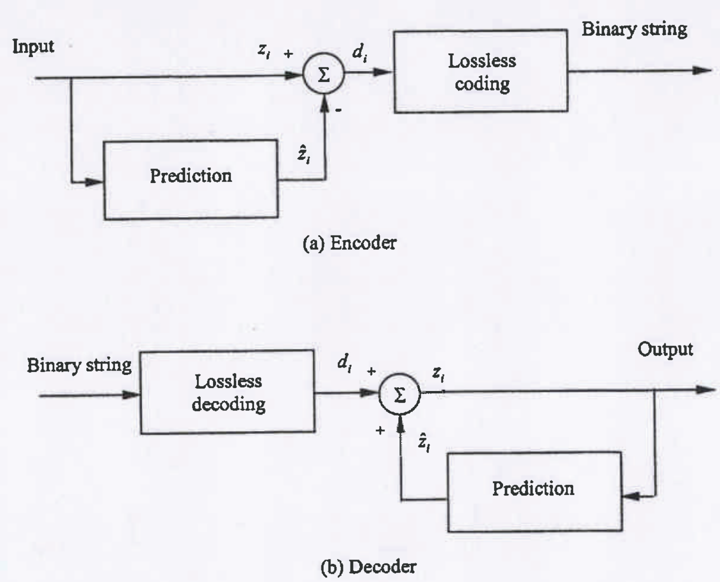

FIGURE 3.14 Block diagram of information-preserving differential coding.

As mentioned before, input images are normally in a PCM coded format with a bit rate of eight bits per pixel for monochrome pictures. The difference signal is therefore integer-valued. Having no quantization and using an efficient lossless coder, the coding system depicted in Figure 3.14, therefore, is an information-preserving differential coding technique.

3.7 SUMMARY

Rather than coding the signal itself, differential coding, also known as predictive coding, encodes the difference between the signal and its prediction. Utilizing spatial and/or temporal correlation between pixels in the prediction, the variance of the difference signal can be much smaller than that of the original signal, thus making differential coding quite efficient.

Among differential coding methods, differential pulse code modulation (DPCM) is used most widely. In DPCM coding, the difference signal is quantized and codewords are assigned to the quantized difference. Prediction and quantization are therefore two major components in the DPCM systems. Since quantization was addressed in Chapter 2, this chapter emphasizes prediction. The theory of optimum linear prediction is introduced. Here, optimum means minimization of the mean square prediction error. The formulation of optimum linear prediction, the orthogonality condition, and the minimum mean square prediction error are presented. The orthogonality condition states that the prediction error must be orthogonal to each observation, i.e., to the reconstructed sample intensity values used in the linear prediction. By solving the Yule-Walker equation, the optimum prediction coefficients may be determined.

In addition, some fundamental issues in implementing the DPCM technique are discussed. One issue is the dimensionality of the predictor in DPCM. We discussed 1-D, 2-D, and 3-D predictors. DPCM with a 2-D predictor demonstrates better performance than a 1-D predictor since 2-D DPCM utilizes more spatial correlation, i.e., not only horizontally but also vertically. As a result, a 3-dB improvement in *SNR* was reported. 3-D prediction is encountered in what is known as interframe coding. There, temporal correlation exists. 3-D DPCM utilizes both spatial and temporal correlation between neighboring pixels in successive frames. Consequently, more redundancy can be removed. Motion-compensated predictive coding is a very powerful technique in video coding related to differential coding. It uses a more advanced translational motion model in the prediction, however, and it is covered in Sections III and IV.

Another issue is the order of predictors and its effect on the performance of prediction in terms of mean square prediction error. Increasing the prediction order can lower the mean square prediction error effectively, but the performance improvement becomes insignificant after the third order.

Adaptive prediction is another issue. Similar to adaptive quantization, discussed in Chapter 2, we can adapt the prediction coefficients in the linear predictor to varying local statistics.

The last issue is concerned with the effect of transmission error. Bit reversal in transmission causes a different effect on reconstructed images depending on the type of coding technique used. PCM is known to be bit-consuming. (An acceptable PCM representation of monochrome images requires six to eight bits per pixel.) But a one-bit reversal only affects an individual pixel. For the DPCM coding technique, however, a transmission error may propagate from one pixel to the other. In particular, DPCM with a 1-D predictor suffers from error propagation more severely than DPCM with a 2-D predictor.

Delta modulation is an important special case of DPCM in which the predictor is of the first order. Specifically, the immediately preceding coded sample is used as a prediction of the present input sample. Furthermore, the quantizer has only two reconstruction levels.

Finally, an information-preserving differential coding technique is discussed. As mentioned in Chapter 2, quantization is an irreversible process: it causes information loss. In order to preserve information, there is no quantizer in this type of system. To be efficient, lossless codes such as Huffman code or arithmetic code should be used for difference signal encoding.

3.8 EXERCISES

3-1. Justify the necessity of the closed-loop DPCM with feedback around quantizers. That is, convince yourself why the quantization error will be accumulated if, instead of using the reconstructed preceding samples, we use the immediately preceding sample as the prediction of the sample being coded in DPCM.

3-2. Why does the overload error encountered in quantization appear to be the slope overload in DM?

3-3. What advantage does oversampling bring up in the DM technique?

3-4. What are the two features of DM that make it a subclass of DPCM?

3-5. Explain why DPCM with a 1-D predictor suffers from bit reversal transmission error more severely than DPCM with a 2-D predictor.

3-6. Explain why no quantizer can be used in information-preserving differential coding, and why the differential system can work without a quantizer.

3-7. Why do all the pixels involved in prediction of differential coding have to be in a recursively computable order from the point of view of the pixel being coded?

3-8. Discuss the similarity and dissimilarity between DPCM and motion compensated predictive coding.

REFERENCES

Bose, N. K. *Applied Multidimensional System Theory*, Van Nostrand Reinhold, New York, 1982.

Bruders, R., T. Kummerow, P. Neuhold, and P. Stamnitz, Ein versuchssystem zur digitalen ubertragung von fernsehsignalen unter besonderer berucksichtigung von ubertragungsfehlern, Festschrift 50 Jahre Heinrich-Hertz-Institut, Berlin, 1978.

Connor, D. J. *IEEE Trans. Commun.*, com-21, 695-706, 1973.

Cutler, C. C. U. S. Patent 2,605,361, 1952.

DeJager, F. *Philips Res. Rep.*, 7, 442-466, 1952.

Elias, P. *IRE Trans. Inf. Theory,* it-1, 16-32, 1955.

Habibi, A. Comparison of nth-order DPCM encoder with linear transformations and block quantization techniques, *IEEE Trans. Commun. Technol.*, COM-19(6), 948-956, 1971.

Harrison, C. W. *Bell Syst. Tech. J.*, 31, 764-783, 1952.

Haskell, B. G., F. W. Mounts, and J. C. Candy, Interframe coding of videotelephone pictures, *Proc. IEEE,* 60, 7, 792-800, 1972.

Haskell, B. G. Frame replenishment coding of television, in *Image Transmission Techniques*, W. K. Pratt (Ed.), Academic Press, New York, 1979.

Jayant, N. S. and P. Noll, *Digital Coding of Waveforms,* Prentice-Hall, Upper Saddle River, NJ, 1984.

Kretzmer, E. R. Statistics of television signals, *Bell Syst. Tech. J.*, 31, 751-763, 1952.

Leon-Garcia, A. *Probability and Random Processes for Electrical Engineering*, 2nd ed., Addison-Wesley, Reading, MA, 1994.

Lim, J. S. *Two-Dimensional Signal and Image Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1990.

Mounts, F. W. A video encoding system with conditional picture-element replenishment, *Bell Syst. Tech. J.*, 48, 7, 1969.

Musmann, H. G. Predictive Image Coding, in *Image Transmission Techniques*, W. K. Pratt (Ed.), Academic Press, New York, 1979.

Netravali, A. N. and J. D. Robbins, Motion-compensated television coding. Part I, *Bell Syst. Tech. J.*, 58, 3, 631-670, 1979.

Oliver, B. M. *Bell Syst. Tech. J.*, 31, 724-750, 1952.

O'Neal, J. B. *Bell Syst. Tech. J.*, 45, 689-721, 1966.

Pirsch, P. and L. Stenger, *Acta Electron.*, 19, 277-287, 1977.

Sayood, K. *Introduction to Data Compression*, Morgan Kaufmann, San Francisco, CA, 1996.