[14] W. K. H. Panofsky and M. Phillips, *Classical Electricity and Magnetism*, 2nd Ed. Reading, MA: Addison-Wesley, 1962, ch. 18.
[15] A. S. Eddington, *The Mathematical Theory of Relativity*, 3rd ed. New York: Chelsea, 1975, ch. 6.
[16] C. Vassallo, "On the expansion of axial field components in terms of normal modes in perturbed waveguides," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-23, pp. 264–265, Feb., 1975.
[17] R. E. Collin, "On the incompleteness of E and H modes in waveguides," *Can. J. Phys.*, vol. 51, pp. 1135–1140, 1973.
[18] J. W. Dettman, *Mathematical Methods in Physics and Engineering*, 2nd ed. New York: McGraw-Hill, 1962, p. 63.
[19] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics*, Part II. New York: McGraw-Hill, 1953, ch. 13.
[20] W. Thomson (Lord Kelvin), *Mathematical and Physical Papers*. Cambridge, England: Cambridge Univ. Press, 1884, pp. 61–91. Actually Kelvin's idea treats the cable as a distributed resistance–capacitance line.
[21] *Reference Data for Radio Engineers*, 4th Ed., H. P. Westman, Ed. New York: International Telephone and Telegraph Corp., 1956, ch. 6.
[22] G. J. Gabriel, Part II, to be published.
[23] See [14, p. 174] for inductance.
[24] H. J. Carlin, "Distributed circuit design with transmission line elements," *Proc. IEEE*, vol. 59, pp. 1059–1081, July 1971.

Picture Coding: A Review

ARUN N. NETRAVALI, SENIOR MEMBER, IEEE, AND JOHN O. LIMB, FELLOW, IEEE

Invited Paper

Abstract—**This paper presents a review of techniques used for digital encoding of picture material. Statistical models of picture signals and elements of psychophysics relevant to picture coding are covered first, followed by a description of the coding techniques. Detailed examples of three typical systems, which combine some of the coding principles, are given. A bright future for new systems is forecasted based on emerging new concepts, technology of integrated circuits and the need to digitize in a variety of contexts.**

I. Introduction

BROADCAST television has assumed a dominant role in our everyday life to such an extent that today in the U.S. there are more homes that contain a television set than have telephone service. So it is natural that in thinking of television transmission we immediately think of the signal that is broadcast into the home. More efficient encoding of this signal would free valuable spectrum space. A difficulty in modifying the television signal that is broadcasted for local distribution is that the television receiver would most likely need to be modified or replaced.[1] The difficulty of achieving this with an invested base of over $10 billion is staggering.

There is a large amount of point-to-point transmission of picture material taking place today apart from the UHF/VHF broadcasting. For example, each of the four U.S. television networks has a distribution system spanning the whole of the continental United States; international satellite links transmit live programs around the world. Video-conferencing services are receiving increasing attention, and facsimile transmission of newspapers and printed material is becoming more widespread. Satellites are beaming to earth a continuous stream of weather photographs and earth-resource pictures, and there are a number of important military applications such as the control of remotely piloted vehicles. Efficient coding of picture material for these applications provides the opportunity for significantly decreasing transmission costs. These costs can be quite large; in comparison with a digitized speech signal at 64 kb/s, straightforward digitization of a broadcast television signal requires approximately 100 Mb/s. The aim of efficient coding is to reduce the required transmission rate for a given picture quality so as to yield a reduction in transmission costs.

A further area of application of efficient coding is where picture material needs to be stored, for example, in archiving X-ray material and in storing picture databases such as engineering drawings and fingerprints. Efficient representation will permit the storage requirements to be reduced.

Some early efforts in picture coding used analog coding techniques and attempted to reduce the required analog bandwidth, giving rise to the term "bandwidth compression".[2] Complex manipulations of the signal are today much more easily done by first sampling and digitizing the signal and then processing the signal in the digital domain rather than using analog techniques. The resulting signal may be converted back to analog form for transmission over an analog channel or be retained in digital form for transmission over a digital channel. Almost all coding methods have been oriented toward digital

Manuscript received May 11, 1979; revised October 2, 1979.
A. N. Netravali is with Bell Laboratories, Holmdel, NJ 07733.
J. O. Limb is with Bell Laboratories, Murray Hill, NJ 07974.

[1] However, there is the possibility of improving picture quality by modifying the transmitted signal such that it remains compatible with existing television receivers.

[2] Channel capacity is a function of both bandwidth and signal-to-noise ratio, thus compressing bandwidth may not reduce channel capacity if a lower noise channel is required as a result.

0018-9219/80/0300-0366$00.75

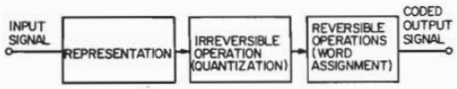


Fig. 1. Block diagram of the encoding process.

transmission for a number of reasons: it offers greater flexibility, it may be regenerated, it is easily multiplexed and encrypted, and its ubiquity is increasing [1].

Efficient coding is usually achieved in three stages (Fig. 1).

1) An initial stage in which an appropriate representation of the signal is made, for example, a set of transform coefficients for transform encoding. This operation is generally reversible.[3] Statistical redundancy may also be reduced.

2) A stage in which the accuracy of representation is reduced while still meeting the required picture quality objectives [2]. For example, dark portions of a picture may be coded more accurately than lighter portions to utilize the fact that the visual system is more sensitive to small signal changes in the darker areas. This operation is irreversible.

3) A stage in which statistical redundancy in the signal is eliminated. For example, a Huffman code [3] may be used to assign shorter code words to signal values that occur more frequently and longer code words to values that occur rarely. This operation is reversible.

In practice transmission channels are frequently prone to errors and a "catch 22" of coding is that when the signal is represented more efficiently the effect of an error becomes far more serious. Consequently, it is frequently necessary to add a controlled form of redundancy back into the signal in the form of channel encoding in order to reduce the impact of transmission errors. The typical configuration then, is shown in Fig. 2 with the coding broken down into source encoding, in which redundancy is removed from the signal for the purpose of achieving a more efficient representation, and channel coding where redundancy is reinserted into the signal in order to obtain better channel-error performance. It goes without saying that the increase in bit rate resulting from the channel coding stage should be significantly less than the decrease in bit rate resulting from the source encoding operation in order to realize a saving. In practice the application of picture coding to transmission channels is an economic tradeoff in system design, balancing picture quality, circuit complexity, bit rate, and error performance.

Where coding is used to reduce storage requirements the tradeoffs are different in that the coding operation usually need not be performed in real time and buffering may not be needed to match the output generation rate of the coder to the transmission rate of the channel. Further, the error rate encountered in the process of storage and retrieval is usually many orders of magnitude lower than the design error rate for a digital channel. As a result, for purposes of storage one can consider more complicated encoding algorithms without concern about the effects of a large error rate.

In this paper, we will be concerned primarily with describing efficient picture coding algorithms. The paper is addressed to the nonspecialist but does assume some background in digital processing techniques. The literature in this area is extensive

[3] DPCM encoding (see Section IV-B) combines stages 1 and 2.

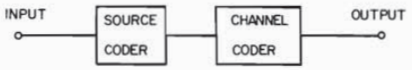


Fig. 2. Source and channel encoding.

[4], [5] and we will describe those aspects of the art which we feel are most significant. References [6]–[12] are special issues which give more detail about certain aspects of the subject. The whole topic of the efficient coding of color signals is covered in a recent paper [13] and for this reason color coding will be discussed very cursorily. One specific type of signal is the two level (black/white) waveform that results from scanning a facsimile image. This special topic is covered in [14] and is not discussed here. A recent book contains reviews of many aspects of picture coding [15].

We start by providing background on the nature and properties of the television signal source in Section II and on the human observer (who is in most applications the ultimate receiver) in Section III. In Section IV, basic waveform coding techniques are first classified and then discussed under the categories pulse-code modulation (PCM), differential PCM (DPCM), transform, hybrid, interpolative, and contour. Section V contains descriptions of state-of-the-art examples of transform encoding, frame-to-frame DPCM and frame-to-frame interpolative encoding and indicates how the techniques of the previous section have been combined in practical encoders. In Section VI issues such as the direction of new developments and the effect of new technology are discussed.

II. Source Coding and Picture Statistics

Ideally, one would like to take advantage of any structure (both geometric and statistical) in a picture signal to increase the efficiency of the encoding operation. Also the coding process should take into consideration the resolution (amplitude, spatial, and temporal) requirements of the receiver, i.e., the television display and very often the human viewer.[4] This problem of encoding can be formulated in the general framework of information theory as a source coding problem. In this section, we describe briefly the source coding problem and point out some of the difficulties in the use of results from information theory. We then present some known statistics of the picture signals and models based on these statistics.

A. Source Coding Problem

The source coding problem can be stated mathematically as follows. Given a random source waveform $L(x, y, t)$ representing, for example, the luminance information in the picture, obtain an encoding strategy such that for a given transmission bit rate it minimizes the average distortion D defined as

$$D = E[d(L, \hat{L})] \tag{1}$$

where $d(\hat{L}, L)$ is a measure of distortion between two intensity fields, L and $\hat{L}$; $\hat{L}$ being the coded representation and E denotes the statistical expectation over the ensemble of source waveforms. Design of such an encoding strategy depends obviously on the statistical description of the random source waveform, L, and on the characteristics of the distortion function d. Shannon's rate distortion theory [16], [17] provides

[4] There are many instances where pictures are processed and/or transmitted for interpretation by a machine.

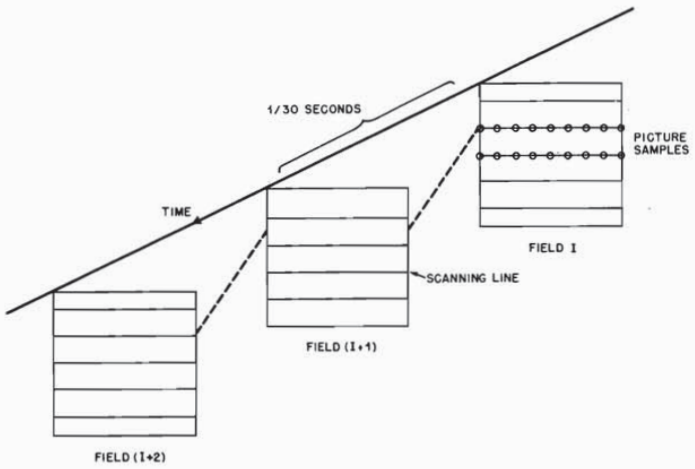


Fig. 3. Scanning process employed in a television signal.

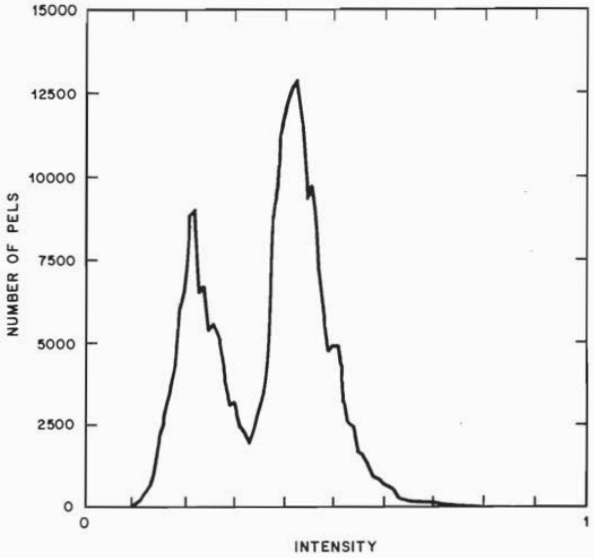


Fig. 4. Histogram of intensities of a typical image. The two peaks are in the dark and light region of the image.

the mathematical framework for analysis of this source coding problem. Let $p_1(L)$ be the probability density function of L, and $p_2(\hat{L}|L)$ be a conditional density corresponding perhaps to an encoding and decoding operation, then the rate distortion function $R(D)$ is defined as

$$R(D^*) = \min \{I(L, \hat{L})\} \tag{2}$$

where $I(L, \hat{L})$ is the average mutual information between the two random waveforms, source L and its reconstruction $\hat{L}$, and the minimum is taken over all the encoding strategies which result in average distortion D less than or equal to a given number D^*. Average mutual information $I(L, \hat{L})$ is defined by

$$I(L, \hat{L}) = -\int p_1(L)\, p_2(\hat{L}|L) \log_2 \left[\frac{p_2(\hat{L}|L)}{p_3(\hat{L})}\right] d\hat{L} \cdot dL \tag{3}$$

where $p_3(\hat{L})$ is the probability density of $\hat{L}$. Qualitatively, the mutual information represents the average uncertainty in the source output minus the average uncertainty in the source output given the coded output $\hat{L}$. The above definition of the rate distortion function becomes significant in the light of the coding theorem of Shannon, which states that for stationary sources an encoding strategy, however complex, cannot be designed to give an average distortion less than D for an average transmission rate $R(D)$; but it is possible to have an encoding strategy to give an average distortion D at a transmission bit rate arbitrarily close to $R(D)$. Thus the rate distortion function gives the minimum transmission rate to achieve an average distortion D and, therefore, provides a bound on the performance of any given encoding strategy, i.e., we can find out how far from the optimum any given practical encoding strategy is. Also it is possible to construct codes (e.g., block codes, tree codes) whose asymptotic performance in terms of rate will be close to $R(D)$; however, this information does not tell us precisely how to build practical encoders, but it is valuable in calibrating them.

In addition to the problem that rate distortion theory does not tell us how to synthesize a practical coder, it has other limitations. It is difficult to compute rate distortion functions for many realistic models of the picture source and distortion criteria. One of the combinations of source distributions and distortion criteria for which the minimization problem of (2) is solved is when the waveform $L(x, y, t)$ is taken to be a sequence of spatial images $L(x, y)$ representing a Gaussian random field, and distortion between L and $\hat{L}$ is measured by a weighted square error [18]. In this case, the optimum encoder first filters the luminance field $L(x, y)$ by the error weighting function and expands the filtered image into its Karhunen–Loeve components. (See Section IV-C.) Karhunen–Loeve components are then represented (in binary bits, for example) with equal mean-square error and transmitted. At the receiver, an estimate of the filtered luminance field is reconstructed, and it is inverse filtered to obtain an approximation of the original image. Although the optimal encoder is known explicitly in this case, the assumptions under which it is derived are not entirely appropriate for the problem of picture communication. The luminance of most picture signals does

not approximate a Gaussian process, and the weighted square error criterion (see Section III) is not appropriate if the pictures are viewed by human observers. Summarizing, there are four problems in the use of the rate distortion theory : 1) lack of good statistical models for picture signals; 2) a distortion criterion consistent with the visual processing of the human observers; 3) calculation of rate distortion functions; and 4) synthesis of an encoder to perform close to $R(D)$.

B. Picture Signal Statistics and Models

Perhaps because rate distortion theory presents many problems in its use for picture coding, many ad-hoc encoding schemes have been proposed to exploit different types of observed redundancies in the picture signal. We give a brief summary of picture signal statistics that is useful in the discussion of encoding schemes described in Section IV.

We start with the first-order statistics. We employ the conventional scanning and sampling process shown in Fig. 3 to convert the television signal from a scene into a sequence of samples. This is done by first sampling in time to get fields and then a periodic sampling of a matrix of picture elements (pels) of chosen resolution in the field. We note that the two consecutive fields are interlaced vertically in space, i.e., spatial position of a scanning line in a field is in the middle of the spatial position of scanning lines in either of its two adjacent fields. Also note that due to this interlace, distance between two horizontally adjacent pels is smaller than the distance between two vertically adjacent pels. The probability density of luminance samples thus generated is highly nonuniform, depends upon the camera settings and scene illumination, and varies widely from picture to picture. A histogram of pel intensities from a typical picture shown in Fig. 4 demonstrates that, even based on the first-order statistics, the luminance does not approximate a Gaussian process [19].

Measurements of some second-order statistics [20]–[22] show that the autocorrelation function depends upon the detail in the picture. In general, the shape of the autocorrelation function can be qualitatively related to the structure of the picture. Fig. 5 shows two pictures: a head and shoulders view of a person, and a picture containing white letters on black background. It is easy to see the relationship between these pictures and their autocorrelations shown in Fig. 5(e) and (f). Figs. 5(c) and (d) show that the autocorrelation functions decrease with increasing shift in the pels. The rate of decrease is large for shifts close to zero, but becomes smaller for large shifts. The envelope of the power spectrum shown in Fig. 6 is relatively flat to about twice the line rate (30 kHz for broadcast television), where it begins to drop at about 6 dB/octave, implying that most of the video energy is contained in the low frequencies [23], or equivalently that the neighboring pels are highly correlated. Based on these measurements, autocorrelation functions in two dimensions have been approximated [24], [25] by the functions of the form

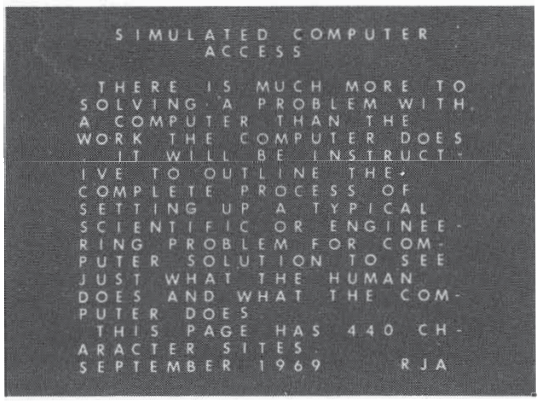
$$\exp(-k_1|\Delta x| - k_2|\Delta y|) \text{ and } \exp[-(k_1\Delta x^2 + k_2\Delta y^2)^{(1/2)}]$$

where Δx and Δy are spatial displacements and k_1 and k_2 are positive constants. Each one of these appears to be a better approximation than the other depending on the type of picture. In general, however, the second expression appears to be closer to the measured data. Using these expressions, different models have been made and used to synthesize optimal encoders [25], [26]. One of the consequences of such a high degree of correlation is that the histogram of the adjacent element difference signal, $\{L(x_i y_i) - L(x_{i-1}, y_i)\}$ is highly peaked at zero [27], [28]. Also, as measurements of Schreiber [27] and others [20], [28] indicate, most of the second-order redundancy (i.e., redundancy contained in blocks of two adjacent samples) is removed by coding adjacent element differences. Therefore, use of three previous samples for prediction does not result in significantly lower sample entropies of the prediction error histograms than the use of two previous samples. Due to the highly peaked nature of the histograms of the prediction errors, they have been modeled by the Laplacian density [29], [30]. Very few measurements [31] have been made of statistics of order higher than the second, primarily due to its variability from picture to picture, and due to the fact that a good method of utilizing such statistics for the purpose of coding does not exist.

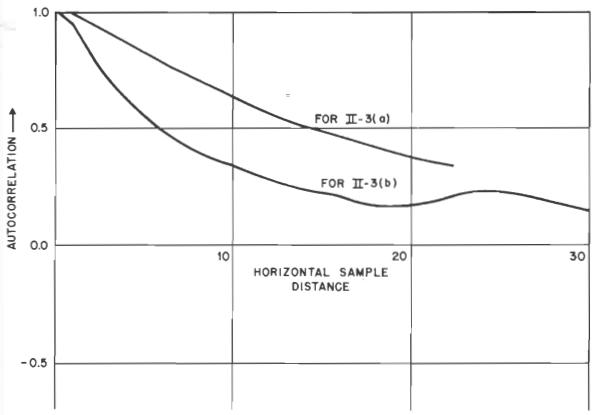
Just as the statistical measurements and models for still pictures are lacking, there are even less measurements on the luminance signal taken as a function both of space and time. Interframe statistics depend very heavily on the type of scene and, therefore, show a wide variation from scene to scene. Some early measurements [32] indicate that since television frames are taken at 30 times a second, there is a high degree of correlation from frame to frame. Thus the histogram of the frame-difference signal is highly peaked at zero. For videotelephone-type scenes, where the camera is stationary and the movement of subjects is rather limited, on the average only about 9 percent of the samples change by a significant amount (i.e., more than about 1.5 percent of the peak intensity) from frame to frame [33]. In broadcast television, where the cameras are not always stationary and there is frequently very large movement in scenes, there would be less frame-to-frame correlation than in videotelephone or videoconference scenes. More recent measurements [34] on the statistics of frame-difference signals indicate that, for scenes containing objects more or less in rectilinear motion, the power spectrum of the frame-difference signal is essentially flat at low speeds, and that the power of the frame difference signal in low frequencies increases by about 7 dB for every doubling of the speed. This is seen for a typical scene in Fig. 7. As would be expected, the spectra of frame difference signals measured in the direction of motion, show nulls at appropriate speeds, whereas spectra measured in the direction orthogonal to the direction of motion show no such nulls. Another interesting observation is that as the amount of motion increases, due to integration of the signal in the camera, the spatial correlation of picture elements increases and the temporal correlation decreases (see Fig. 7). Also there is more correlation spatially orthogonal to the direction of motion than spatially parallel or in the temporal direction. It is obvious from these measurements that there are still quite a lot of interframe statistics that are unknown.

We close this section by pointing out some recent models of picture signals which appear to be more realistic and promising. As mentioned before, the picture signal, in general, is highly nonstationary, and the local statistics vary considerably from region to region. Some of this difficulty can be overcome by considering the picture signal as the output of many sources each tuned to a certain type of statistics [35], [36]. Yan and Sakrison [35], for example, consider a two-component model in which the vertical edges (or the high-frequency components) are treated as one component and the rest (texture details) are

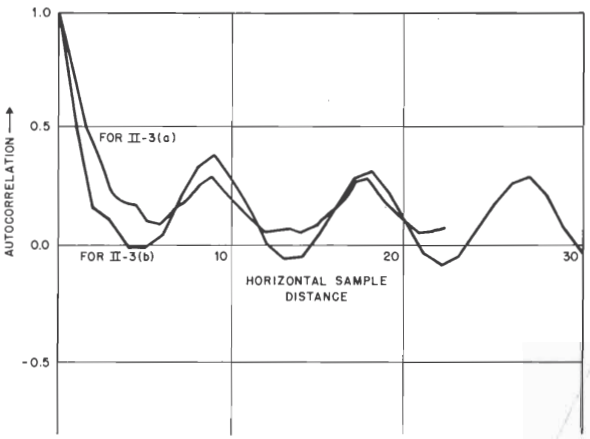
(a)



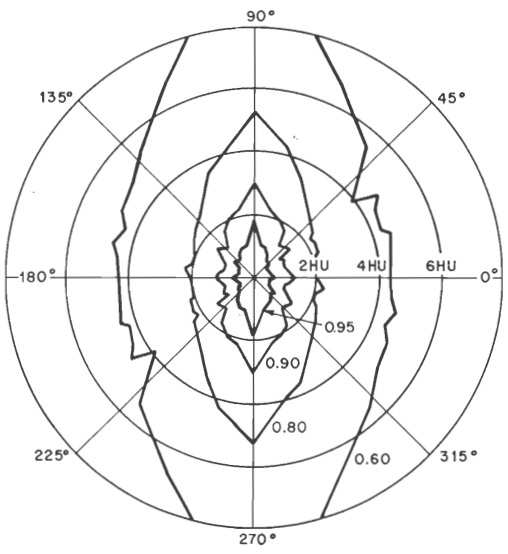
(b)



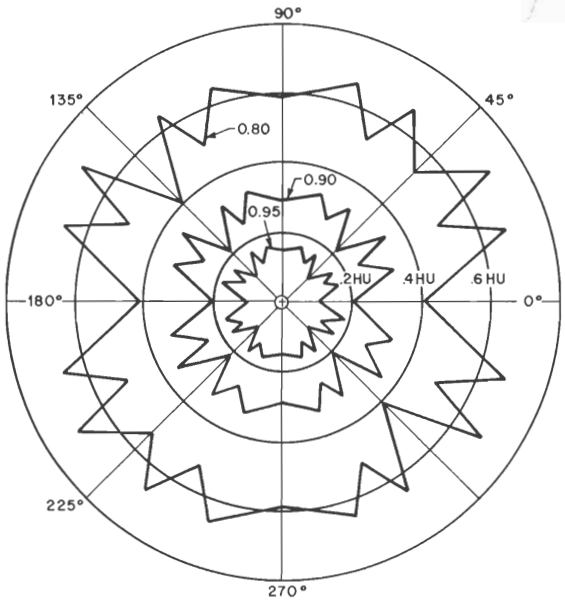
(c)



(d)



(e)



(f)

Fig. 5. (a) Head and shoulders view of a person. (b) White text on black background. (c) and (d) The autocorrelation function in horizontal and vertical direction for both scenes (a) and (b). These are for a typical videotelephone display, with 208 samples/line and 250 lines/frame and with a picture size of 5.5 in by 5 in. Horizontal sample spacing is then 0.02644 in and vertical line spacing is 0.02000 in (without regard to interlace). (e) and (f) The contours of equal autocorrelations for scenes (a) and (b). HU denotes the horizontal sample distance unit.

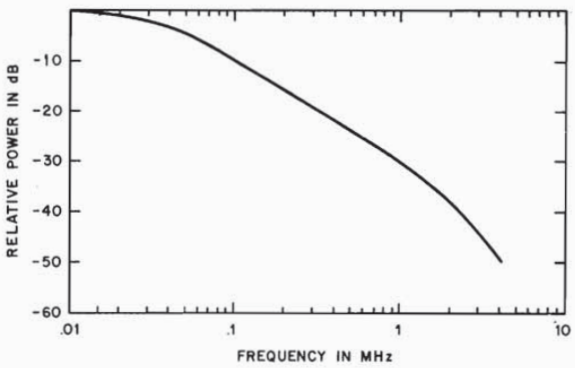


Fig. 6. Envelope of the power spectrum of a typical video signal. Note that the envelope is relatively flat, up to about twice the line rate, where it begins to drop at about 6 dB/octave (from Connor *et al.* [23]).

treated as the other component. They argue that if the edge information is subtracted from the picture signal, the rest of the signals appear to be close to a Gaussian process and, therefore, an optimal encoder, mentioned earlier, can be applied. Rate distortion theory of such two-component models may find greater use and a beginning has already been made [37]. In a different context, Lebedev and Mirkin [38], [39] develop a composite source model and describe experiments in which a picture signal is broken down into many components by using correlations at 0°, 45°, 90°, and 135° to the horizontal. They look at the picture signal as the weighted sum of these five components, weights being given by a random variable. Thus the model can be considered to be locally anisotropic, but on the average isotropic. Impressive results are claimed by Lebedev and Mirkin for image restoration using such a model. Such models have a large potential, if appropriate components could be determined and a suitable method of combining these components to form the composite picture signal could be found. A similar idea has been explored by Maxemchuk and Stuller [36], who model the image as a random field that is partitioned into independent quasi-stationary subfields. Each subfield is the output of one of six possible autoregressive sources, whose selection is governed by a space-varying probability distribution that is unknown *a priori* to the observer. The model also includes a white subsource that initiates the autoregressive sources at certain boundaries within the picture. Maxemchuk and Stuller apply this model to adaptive DPCM using a mean-square error criterion for each point and claim good results.

III. Properties of the Receiver

A. Picture Quality

Systematic distortions occur in representing a live scene by a television picture. For example, the contrast ratio in a scene (the ratio of the luminance of the lightest to the darkest parts) can frequently be 200:1 or greater whereas it is difficult to obtain a contrast ratio much greater than 50:1 under normal television viewing conditions; the color television tube, by mixing three primary colors reproduces the approximate chromaticities of the original scene, not a scene having the same spectral distribution. The fact that the viewer is usually happy to accept these approximations implies that he is not particularly sensitive to them, even when he can make a direct comparison between the original and the reproduction.

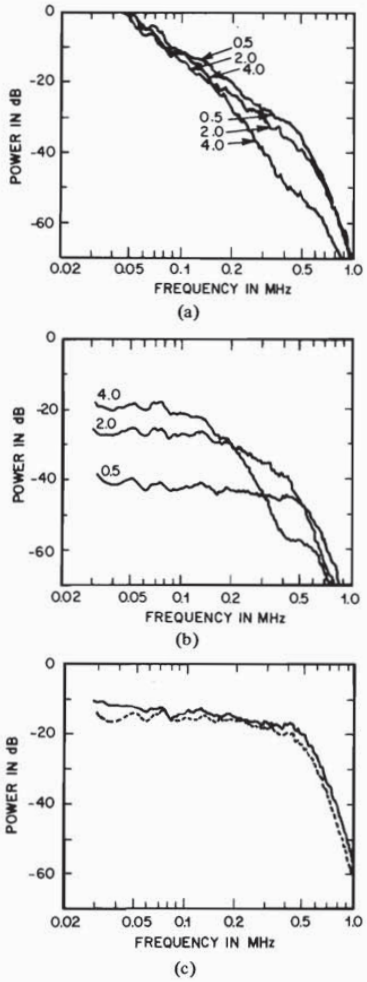


Fig. 7. (a) Power density spectra of the video signal at speeds of 0.5, 2.0 and 4.0 pels per frame (pef). This is for a video telephone type of signal containing a head and shoulders view of a person. The attenuation at high frequencies is due to the pre- and postfiltering. The effect of camera integration on the video signal at higher speeds is seen in the reduced power at high frequencies. (b) Power density spectra of the frame-difference signal at speeds of 0.5, 2.0, and 4.0 pef. Note the increase in power density at low frequencies as the speed increases and the small dip at approximately 0.45 MHz in the curve for a speed of 4 pef. (c) Comparison of power density spectra of the element-difference signal and the frame-difference signal, both recorded at a speed of 1 pef. The dashed curve is for the frame-difference signal (from Connor and Limb [34]).

Instead of seeking to make the reproduced picture as similar to the original as possible, consistent with the shortcomings of the system, one can purposely distort the picture to obtain a more pleasing effect. Examples would be filtering the signal (linear or nonlinear) in order to make it appear crisper [40]; altering hue so as to give the appearance of a healthy tan.

The task to be performed will largely determine the criteria that are used to determine picture quality. Thus a photointerpreter would attach great importance to sharpness and probably less to accurate tonal reproduction. We will be mainly concerned with an average television viewer who is performing no specific task related to the image structure in contradistinction to, say, imaging for medical diagnostics. It is convenient to start with the existing analog signal as a refer-

TABLE I

(a)	(b)	(c)
5 Excellent	5 Imperceptible	3 Much better
4 Good	4 Perceptible but not annoying	2 Better
3 Fair	3 Slightly annoying	1 Slightly better
2 Poor	2 Annoying	0 Same
1 Bad	1 Very annoying	-1 Slightly worse
		-2 Worse
		-3 Much worse

ence and measure distortion by the extent to which the distorted picture differs in appearance from the analog signal.

B. Measurement of Picture Quality

Measurements of picture quality must depend upon subjective evaluations either directly or indirectly [41]. Subjective testing is very time consuming and consequently is avoided where possible. In primary or explicit measurement of picture quality a group of subjects make subjective decisions while in secondary or implicit measurement, objective characteristics of standardized waveforms are measured and the results are then converted to quality measures through previously established relations. In the digital processing of pictures, distortions are frequently introduced that are complex in nature (e.g., they can be a complex function of the signal) such as edge noise, slope overload, and movement related distortion [42]. In such instances existing indirect methods are of little use.

Subjective evaluations are of two broad types, rating-scale methods and comparison methods. In the rating-scale method, the subject views a sequence of pictures under comfortable, natural conditions and assigns each picture to one of the several given categories. The subject may be assigning an overall quality rating to the picture using categories such as those listed in Table I(a) or he may use an impairment scale as shown in Table I(b). The results of a rating will depend upon many factors: the experience and motivation of the subjects, the range of the picture material used and the conditions under which the picture is viewed (e.g., ambient illumination, contrast ratio and viewing distance). These variables have been explored in depth and standardization is taking place at the international level [43]. This enhances the utility of the procedure making it more feasible to compare results obtained at different times and in different laboratories.

In the comparison method, the subject adds impairment of a standard type (e.g., white noise) to a reference picture until he judges the impaired and reference pictures to be of equal quality. This can be done very accurately where the two types of distortion are similar in appearance, for example, equating additive noise of differing spectral distribution. The distortion can then be assigned a quality by referring to rating scale tests on the standard impairment. One should not expect that the resulting ratings will necessarily be transitive. In a variation of this method the subject uses a comparison rating scale (Table I(c)) to compare pictures having various levels of a distortion with a reference picture. The resulting data is then processed to obtain the level which produces the "point of subjective equality" between the distorted picture and the reference.

Secondary measures of quality are more useful in the field and are usually developed after primary measurements have been established. An example is the pulse and bar waveform (Fig. 8(a)). Consider the pulse; it is chosen to yield quickly an estimate of the performance of a transmission link at the high end of the frequency band. A template has been determined such that waveform deviations reaching the edge of the template are about equiobjectionable [44]. Notice that the error tolerance varies greatly throughout the waveform (Fig. 8(b)). For other secondary measures see [45].

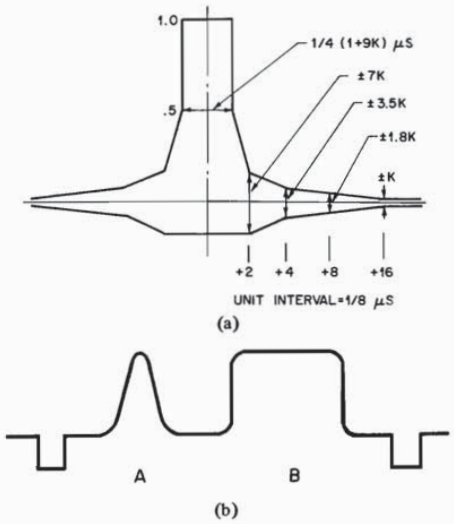


Fig. 8. (a) Test signals used for K-rating measurement method. Signal A is a sine-squared pulse of half-amplitude duration $2T$, where T equals the sampling period. Pulse B is a bar signal of width approximately half the duration of a horizontal scanning line. (b) K-rating graticule for the $2T$ pulse for the NTSC system.

C. Visibility of Distortion

Loosely speaking, the less accurately a pel needs to be reproduced, the fewer the bits required to encode that pel. The required accuracy, in turn, depends upon the visibility of the coding distortions or perturbations at, or in the neighborhood, of the pel. Certain types of distortion rarely occur in the types of coding algorithms commonly used today. For example, geometric distortions (e.g., pin-cushion distortion) would not be expected in spatial or transform domain coding. However, if more complicated algorithms are explored that are able to recognize and manipulate common shapes, then the visibility of such types of distortion would be more important.

For pel-domain coding we would like to know the amount of error tolerable at each element in the picture and how the errors in adjacent points combine. Similarly, for transform domain coding we would like to know the amount of error tolerable in each transform coefficient and how errors in different coefficients add. The amount of distortion that may be tolerated will depend upon the overall quality that is desired. For high-quality coding any differences between the coded and uncoded pictures will be subjectively small which implies that coding distortions will be close to the visual threshold. Visual threshold is the point at which a stimulus (a perturbation or distortion) just becomes visible or just ceases to be visible. It is a statistical measure and is usually defined as the amplitude of the stimulus such that it will be detected on 50 percent of occasions. It is very difficult for a subject to specify the subjective amplitude of a distortion that he sees, the exception being the visual threshold (and even at threshold the judge-

ment is often very noisy). Thus, where we are concerned with coding for high-quality images, knowledge of visual threshold will be a valuable guide in determining the relative amount of tolerable distortion at a pel. There is much work in the psychological literature describing visual threshold under a wide range of viewing conditions. On the other hand, there is little known about the visibility of transform domain perturbations (however, see [46]–[48]).

For lower quality encoding, the subjective amplitude of distortions may be quite large and we cannot assume that visibility changes with stimulus configuration in the same way for small amplitude distortions (near threshold) as it does for large amplitude distortions. Indeed, we know that the suprathreshold grating visibility function is much flatter than the corresponding threshold function [49]. However, there is relatively little known about suprathreshold visibility, but see [50]. Fortunately, most applications call for medium- or high-quality encoding where threshold visibilities are much more relevant and so we will briefly review some of the information available on threshold vision.

D. Threshold Vision

The question naturally arises as to whether the pel-domain is most appropriate for describing threshold vision or whether some transform domain is more suitable. This question has produced much heated debate in the psychological literature [51], [52] and two brief comments in favor of a pel-domain description will suffice: 1) as will be described in more detail below spatial masking is a highly local effect and depends strongly on the spatial configuration of the masking stimulus; 2) the visual field is highly inhomogeneous with a point spread function that changes rapidly with location in the visual field. Both these effects are more easily described and handled in the spatial domain.

We will now briefly review information concerning visual threshold. Three factors affecting the threshold of a stimulus can be isolated:

1) the overall luminance background against which the perturbation is presented (global effect);
2) the masking effect of suprathreshold luminance changes immediately adjacent in space and time to a perturbation (local effect);
3) contribution to threshold from perturbations in spatio-temporally adjacent elements.

These three factors have been studied separately and to date there has been little experiment to explore interactions, in particular between 2 and 3. (For example, we have little idea of how distortions sum across a luminance edge.) For simplicity, we will assume that the above factors act independently. Let us consider them in turn.

1) Global Threshold Effects: Consider that a luminance perturbation is presented against a background region of luminance L_B subtending an angle of approximately 1.5° subtended arc, the area outside the background area will be referred to as the surround having luminance L_S (Fig. 9). This particular stimulus configuration has been studied extensively [53]. We can now ask how the threshold visibility of a perturbation varies as the background and surround luminances are changed. This information will be very important in deciding how to quantize the video signal for PCM encoding or where a signal is being digitized prior to further coding (see Section IV-A). The threshold will be primarily dependent on the background luminance with the surround having much less effect. When L_S equals L_B we have a condition corresponding to that typically used to explore the Weber fraction $\Delta L/L_B$ where ΔL represents the visual threshold luminance of the perturbation presented on a background of luminance L_B. ΔL increases monotonically with L_B. At high luminances (say greater than 10 mL) the Weber fraction is fairly constant but as the luminance falls below this value $\Delta L/L_B$ starts to rise. A plot of ΔL versus L_B is shown in Fig. 10(a) (for the condition $L_S = L_B$) by the full line, with the long-dashed line representing the condition $\Delta L/L_B$ = constant (Webers law). We are interested primarily in the range of luminances encountered in a typical viewing situation and they correspond to a minimum of about 0.5 mL to a maximum of about 50 mL. If the surround is held constant at the maximum luminance, as L_B decreases, $\Delta L/L_B$ decreases much less rapidly and bottoms out at about seven times the threshold for the condition $L_B = L_S$ (Fig. 10(a)) [54]–[56]. Setting the size of the background to 1.5° diameter is rather arbitrary and local effects extend over an area that is probably significantly smaller than this value. So if we considered a background of say 0.5° diameter then again, for L_S held constant at L_{max}, $\Delta L/L_B$ would be even larger at lower values of L_B. The effect of nonuniform surrounds has been studied in some detail [53] and it has been shown that any nonuniformity in the surround is equivalent in effect to a uniform surround with a luminance referred to as the adapting luminance. The effect of a bright spot, for example, decreases rapidly as the spot increases in angular distance from the stimulus [53].

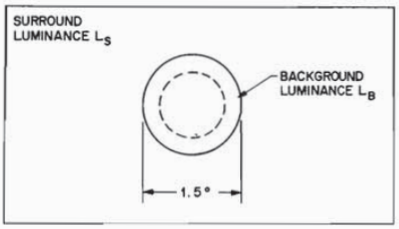


Fig. 9. Display assumed for calculating functions shown in Fig. 10(a).

We have described how visibility changes with variations in the luminance of the background and the surround. Let us now consider the picture tube and look at visibility in units of the electrical signal E applied to the picture tube. The transfer characteristic of the picture tube may be described by

$$L = kE^{\gamma} \tag{4}$$

where k and γ are constants and γ is in the range 2 to 2.5 for a typical cathode ray tube. Thus in spite of the fact that ΔL decreases as L decreases the slope of the E versus L characteristic, dE/dL increases, in partial compensation for the Weber effect. Under normal viewing conditions the situation is more complex in that ambient illumination falling on the screen further modifies the transfer characteristic shown in (4). It is impossible to obtain very low luminances and the transfer function asymptotically approaches the luminance set by the luminance of the ambient screen

$$L = kE^{\gamma} + L_{amb}. \tag{5}$$

Thus, for backgrounds having a luminance approaching the ambient screen illumination, dE/dL increases dramatically, producing a large decrease in the visibility of a perturbation when expressed as a signal rather than a luminance. For a

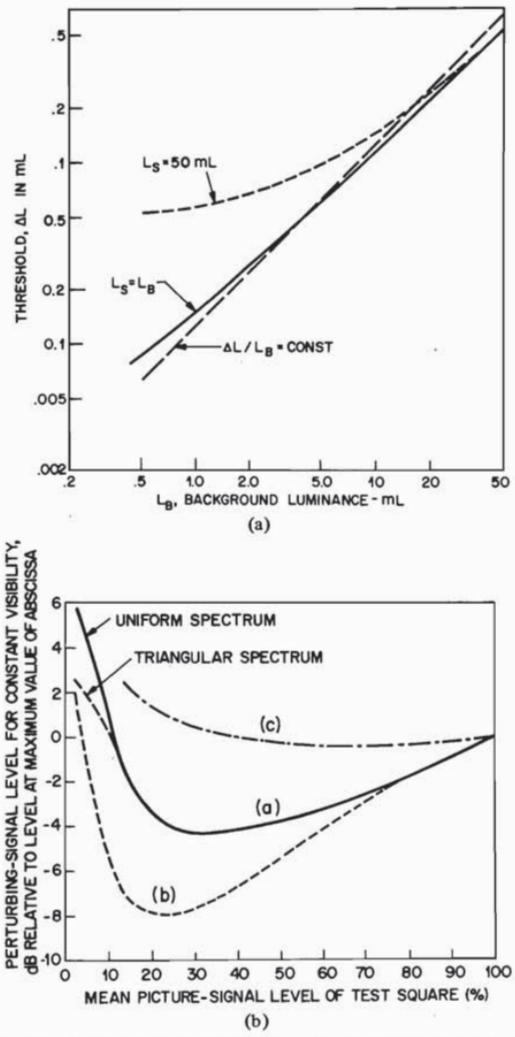


Fig. 10. (a) Threshold ΔL as a function of background luminance, L_B. The Weber law relation ($\Delta L/L_B = constant$) is shown by the straight dashed line. The other two curves are for different values of surround luminance. The short dashed curve is for surround luminance, L_S, maintained at a maximum value, while the full curve gives the result when the surround luminance is maintained equal to the background luminance. (b) Perturbing-signal levels necessary to give constant visibility: (a) random noise or pattern having low coherence between successive displays, (b) stationary patterns, (c) pattern displayed in alternate polarities (from Newell and Geddes [57]).

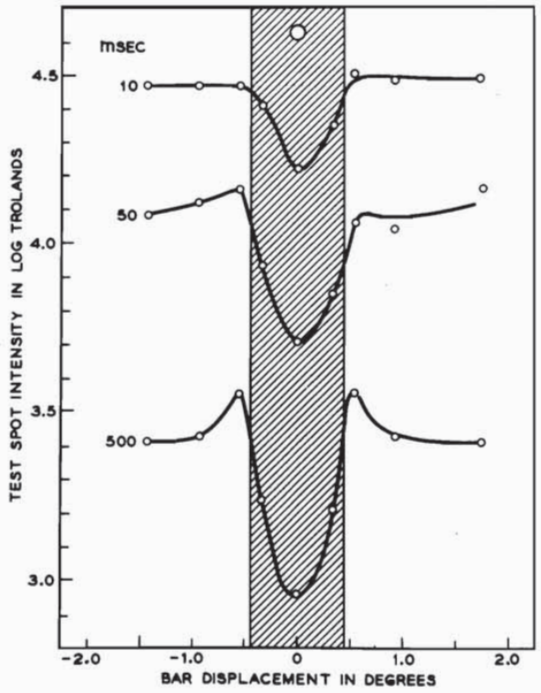


Fig. 11. Threshold luminance of the test spot as a function of bar displacement. Parameter is the exposure duration of the test spot and bar. Data for 50- and 10-ms durations have been moved up an arbitrary amount (from Novak and Sperling [62]).

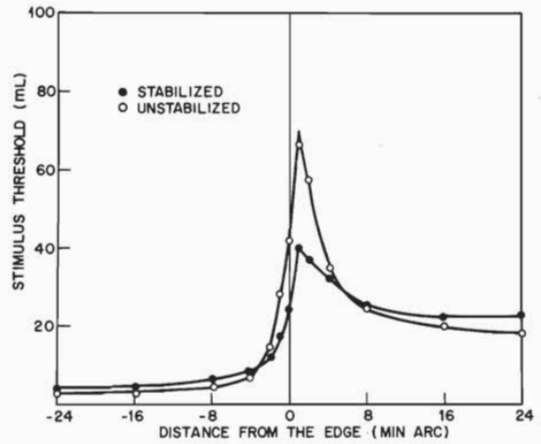


Fig. 12. Stimulus threshold versus distance from edge for both stabilized and unstabilized viewing. The edge was continuously presented and went from 1 mL on the left to 20 mL on the right. The stimulus to be detected was a thin line of 1 × 30′ of subtended arc, oriented parallel to the edge. It was presented for a duration of 100 ms (from Lukas *et al.* [59]).

given viewing condition experiments can be conducted to determine the threshold voltage versus luminance (or normalized signal voltage).

Fig. 10(b) shows results obtained for stimuli of different types presented within small test squares inserted into a fixed test pattern [57]. For a stationary pattern the amplitude changes by about 8 dB as the test square changes from light to dark. Note that stimuli that change from frame-to-frame produce quite different results, showing even less variation over the gray scale. Evidently, the amount of temporal integration taking place in the visual system increases significantly at lower luminances and as a result the perceived amplitude of the time-varying stimulus is reduced.

In summary, visual threshold at a pel is not a function of a single variable as the Weber fraction would tend to imply but depends upon the luminance in both the neighboring background (strong dependence) and surrounding areas that are further removed (weak dependence). Further confounding the situation is a change in temporal response with luminance. In terms of the signal that is transmitted, the threshold ΔE does not change as rapidly as ΔL due to the nonlinear conversion in the picture tube. However, the exact relationship will depend upon viewing conditions, in particular on the ambient illumination falling on the screen. High signal thresholds will occur

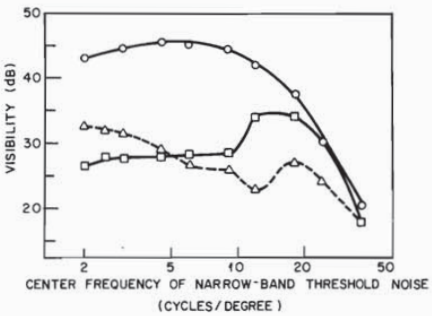


Fig. 13. Visibility of narrow-band noise of varying center frequency superposed on narrow-band background of a fixed frequency. Full curve at top is for a plain background (no noise). Lower full curve is for background centered at 4.5 cpd. Dashed curve is for background centered at 12 cpd. The presence of the background reduces overall visibility with maximum reduction occuring near the center frequency of the background (from Sakrison *et al.* [67]).

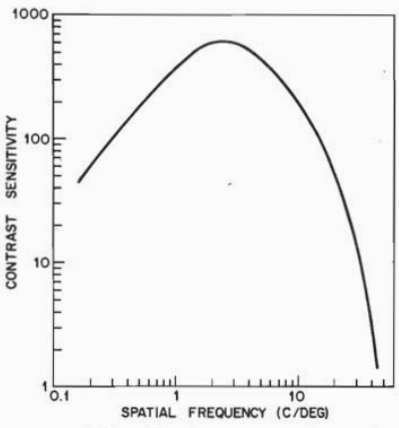


Fig. 14. Contrast sensitivity for sine-wave gratings at luminance of 500 cd/m^2 (after Campbell and Robson [69]).

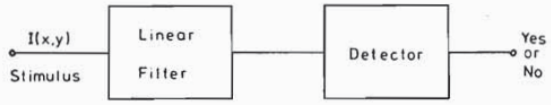


Fig. 15. Simple single-channel model of visual threshold for a stimulus presented against a plain background.

in the regions of pictures that are very dark on the one hand or very bright on the other; greatest sensitivity will occur in medium to dark-gray regions.

2) Local Thresholds Effects: Let us now consider backgrounds that are no longer uniform but contain large changes in luminance. It has been known for a long time that visual threshold increases on both sides of a change in luminance [58]. This effect is important in DPCM coding where changes in signal amplitude are directly quantized (Section IV-B). Fig. 11 shows how visual threshold changes adjacent to a vertical edge which increases in luminance by a factor of 20 from the dark side to the light side [59]. The target used to determine threshold was a line 1′ in width and 30′ in length oriented parallel to the edge. On the light side the threshold adjacent to the edge has increased between three and four times the threshold further from the edge; there is a similarly large increase adjacent to the edge on the dark side. But notice that these effects are very local and the typical spread is of the order of only 5′ of subtended arc (approximately 4 pels at a viewing distance of 8 times picture height for broadcast television). We say that the edge "masks" the perception of the signal adjacent to the edge in that the signal is perceived less accurately than would be possible in the absence of the edge. The edge effect is often confused with the Mach effect [60]; however, the Mach effect refers to a change in brightness at an edge and while the brightness increases on the light side of the edge, it decreases on the dark side. While there may well be a connection between the Mach effect and the edge threshold effect, it is certainly not straightforward [61].

There is no comprehensive model to explain threshold elevation at borders; however, certain facts are known [62]–[64]. When an edge is briefly presented ($\geqslant$10 ms), the threshold is not elevated (see Fig. 11) [62]. This experiment would tend to suggest that the edge threshold effect is related to involuntary eye movements. Recent experiments tend to confirm this view; as shown in Fig. 12 when the image of an edge is stabilized on the retina so that involuntary eye movements are eliminated, the edge effect is markedly decreased. It is unclear whether the remaining edge effect is due to residual eye movements or to neural interactions. Other work [63] shows that the length of the edge in the vicinity of the picture element being examined also increases the amount of threshold elevation.

Masking can be thought of in the frequency domain by studying the threshold of a grating in the presence of either a suprathreshold masking grating or narrow-band noise [65]–[67]. Fig. 13 shows the effect on the frequency response of narrow-band noise when presented against a masking background also of narrow-band noise having center frequencies of 4.5 cycles per degree (cpd) and 12 cpd. Although the effect of the masking stimulus is maximum near the center frequency of the mask, there is also a very large overall reduction in sensitivity.

Thus far we have treated the influence of nonuniformities as belonging to one of two classes: global or local. A theory that integrates these effects is clearly needed. The effect of nonuniformities in the range of 10′ to 60′ are not covered by either existing treatment.

3) Model of Stimulus Summation: If we knew how subthreshold perturbations combine, then we could predict the visibility of arbitrarily shaped stimuli (assuming a linear system, which is not too unreasonable under small signal conditions). This is important, for example, in interpolative coding where one would like to know whether coding distortion in a number of adjacent elements is visible or not (see Section IV-D). The eye filters the "signal" applied to it and a typical frequency response, measured by determining the threshold of spatial sine wave patterns is shown in Fig. 14. Fig. 15 gives a simple single-channel model of threshold vision that has been used with some success to determine visibility. The input stimulus $I(x, y)$ is linearly filtered, and if the resulting filtered signal exceeds some threshold, the stimulus is regarded as visible, otherwise it is not perceived. Simple as this model is, it is rather successful in predicting the visibility of certain high-frequency stimuli [68]. However, it fails at least on one particular type of stimulus, one in which luminance is varied in the horizontal direction of the display as the sum of two sine waves that are separated in frequency by more than an octave. Campbell and Robson [69] have proposed a multichannel model which consists of a number of simple models (channels)

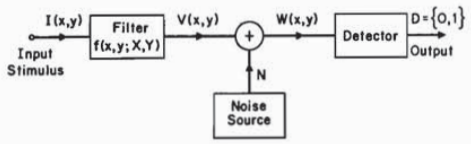


Fig. 16. Single-channel model of visual threshold incorporating visual inhomogeneity and noise (from Limb and Rubinstein [71]).

like that shown in Fig. 14 connected in parallel, with each channel tuned to a different spatial frequency. The outputs of the channels are combined such that the stimulus is detected whenever, the output of at least one channel exceeds its own separate threshold. This model predicts that sine waves that are sufficiently separated in frequency will be detected independently. Indeed experiments by Graham and Nachmias show that this is the case [52]. Sakrison *et al.* have conducted a series of experiments in an attempt to characterize channels more fully [67]. They have determined that the bandwidth of the channel having maximum sensitivity at 4.5 cpd is 2.5 cpd. Following the filtering they assume that the signal is raised to some power n and integrated (i.e., $[\Sigma|x|^n]^{1/n}$). Experiments with narrow-band noise having different probability density functions suggests that $n = 5$.

The single and multichannel models mentioned above assume linear filtering, isotropy, and homogeneity. Since the stimuli being processed are small (near threshold), the assumption of linearity is probably quite reasonable. Further, while we know that the visual system is slightly anisotropic, having poorer resolution on the diagonal than vertically or horizontally, this is a relatively small effect and is less important where we are dealing with stimuli that vary only in one dimension.

However, the visual system is by no means homogeneous. On the contrary, the visual field has high spatial resolution in the fovea (the central portion of the retinal field) and decreasing resolution toward the periphery [70]. The single-channel model may be extended to take account of visual inhomogeneity as shown in Fig. 16 [71]. The input filter now has an impulse response $f(x, y)$ that depends on the position in the input field (X, Y) at which the impulse is applied. Noise is added to the filtered signal to simulate the effect of neural noise, and the resulting signal is fed to a detector. If any single point on the waveform $W(x, y)$ exceeds a pair of symmetrical thresholds placed on either side of the background level, then we consider that the waveform is visible. This is a probabilistic event, since random noise has been added to the waveform. The input amplitude is said to be at threshold when there is more than a certain probability that one or more points on the signal W will exceed either of the symmetrical thresholds.

A series of subjective measurements were used to determine the impulse response of the visual filter of the model at varying eccentricities. As can be seen in Fig. 17, the spread of the impulse response approximately doubles in going from the fovea 0′ (Fig. 17(a)) to an eccentricity of 500′ (Fig. 17(f)). There is also an accompanying change in sensitivity. The model predicts the visibility of a number of different line stimuli in both fovea and periphery, and in all cases the predictions agree closely with experimental measurements. Because of the change in impulse response, the relative visibility of stimuli changes markedly with position in the visual field, and any model that does not incorporate this will not be accurate. These results help in accurately evaluating the visibility of small distortions but to date they have not been directly applied in an encoding algorithm.

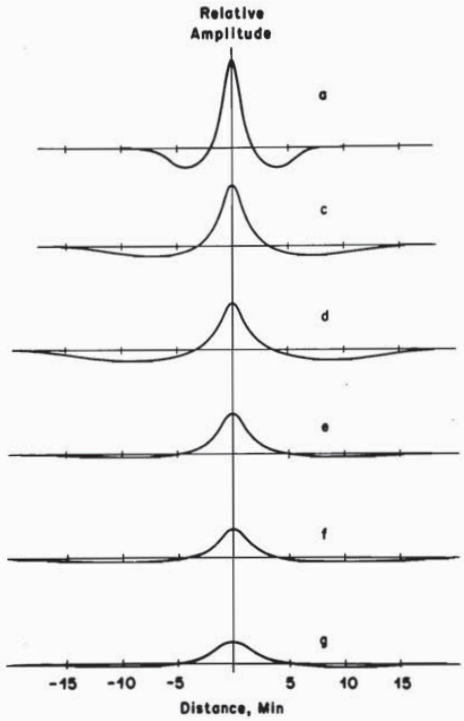


Fig. 17. Visual inhomogeneity change in shape of the line-spread function with change in position in the visual field. From top to bottom the positions are 0′, 50′, 100′, 200′, 300′, and 500′ from the point of fixation (from Limb and Rubinstein [71]).

4) Perception in Moving Areas: Two studies [72], [73] explore the spatial resolution requirements when viewing a moving object. Connor and Berrang [72] had subjects view a manikin head that was moving back and forth across a plain gray background. He concluded that over the range of speeds he studied the reduction in resolution due to integration of the image on the target of the camera dominated any reduction in resolution introduced by the visual system.

In these studies the observer could easily track the moving object and, bearing this in mind, the results are not so surprising. Evidently, the eye acts very differently when a moving object is not tracked. Direct measurements of the temporal impulse response of the eye are at an early stage but it appears that the situation is not too dissimilar to the spatial situation. Rashbass has measured the correlation function of the temporal response to a stimulus of low spatial frequency (Fig. 18(a)) [74]. A guess at the shape of the corresponding impulse response is shown in Fig. 18(b). Measurements of the temporal impulse response of sine waves using reaction time have been made but the overall spread of the response is significantly larger than that suggested by other experiments [75].[5] It would appear that once the spatiotemporal impulse response is determined for a given viewing condition, the visibility of an arbitrarily shaped stimulus presented against a flat background could be predicted by extension of the models that are being developed.

[5] Indirect measurements of the spatiotemporal impulse response of the eye may be obtained by an inverse transform of the spatiotemporal modulation transfer function. There are a number of problems with this approach: assumptions must be made about the missing phase information; inhomogeneity will modify the results; eye movements can have a large effect.

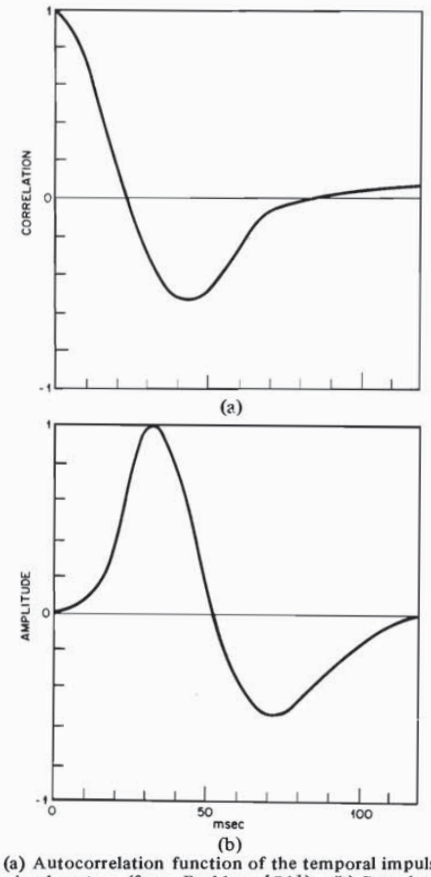


Fig. 18. (a) Autocorrelation function of the temporal impulse-response of the visual system (from Rashbass [74]). (b) Postulated form of the temporal impulse response of the visual system.

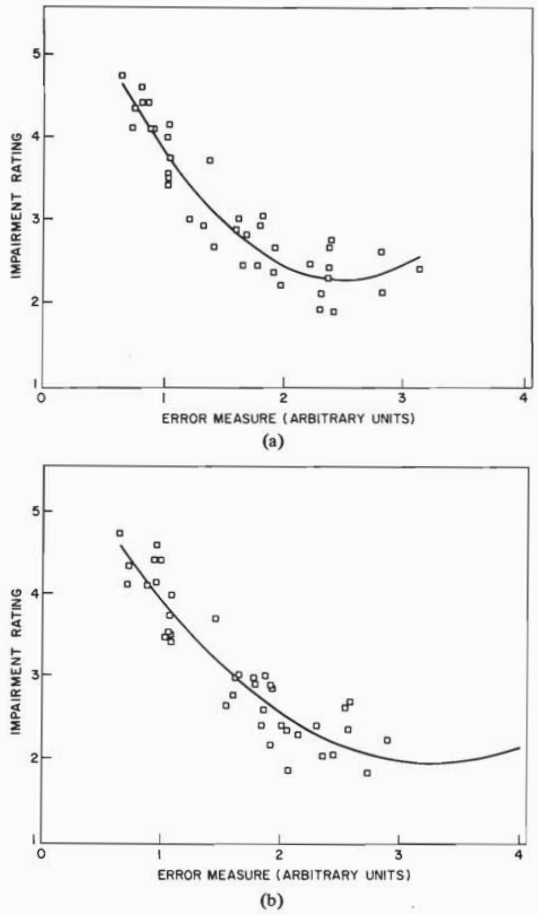


Fig. 19. (a) Impairment rating versus error measure (on a scale of 1 to 5). Error measure is global rmse (from Limb [80]). (b) Same as (a) but error measure is now the average of the two maximum local measures using a masking function. For both plots the curves are least mean-square fits to the data points.

Miyahara [73] measured the temporal and spatial resolution requirements of images that could be tracked and found similar results. However, he also studied moving objects (wind-up toys) that are not so easily tracked and found that subjects could detect loss of resolution less easily (a just detectable resolution loss equivalent to a bandwidth of 2 MHz for tracked movement versus 1.2 MHz for less easily tracked movement). Most movement displayed on a screen is not easily tracked since it is difficult to accurately predict (e.g., head or hand movements). Further, the viewer does not bother to track much of the movement he sees since it does not occupy his visual attention. We have little quantitative data about when a viewer tracks an object and how accurately he tracks it [76]. It is also unclear how the information could best be used to improve coding efficiency where a picture is being viewed by a large number of observers.

Further measurements of Miyahara [73] show smooth movement is perceived in viewing normal television (60 fields/s) up to angular speeds of 24′/s. If alternate fields are transmitted with the missing fields filled in by repeating the previously transmitted field, the speed at which degradation is *just perceptible* drops to 54′/s, and to 34′/s when every third field is transmitted. He also measured the speeds at which movement is reproduced with *acceptable* quality for normal viewing and found that the speeds are approximately 3 times the just perceptible value.

Temporal masking has been studied by Seyler and Budrikis [77] who found that spatial resolution could be reduced significantly immediately following a scene change and need only be increased slowly thereafter. As long as full resolution was restored within approximately 0.75 s the observer did not detect the reduction.

Although the psychological literature contains much work on temporal masking (e.g., [78]), apparently no attempt has been made to relate the results to picture coding.

E. Measures of Picture Quality

One factor complicating the design and evaluation of new coding algorithms is the lack of an objective measure of picture quality. This is not surprising since an accurate measure would imply an accurate visual model and, as we have seen, there is much still not understood about visual perception.

In the photographic sciences much study of picture quality criteria has been performed and the short paper by Higgins [79] summarizes some of this work.[6] However, the approach is not well suited to digital picture processing where the range of impairments encountered is much greater than in photographic systems and where much of the distortion may be de-

[6] It is unclear how and to what extent the large amount of work on task related quality measures (such as target detection and identification) relates to quality measures where the observer is viewing for entertainment.

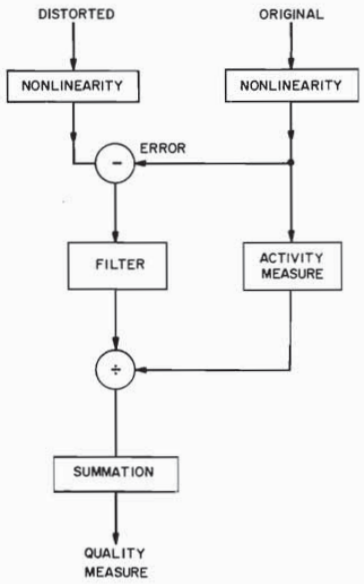


Fig. 20. Model used for study of picture quality evaluation.

pendent on the value of the signal (e.g., error resulting from nonuniform quantization). Sakrison discusses distortion measures of the observer and suggests a criterion based on a threshold model of vision [18].

A measure that is frequently used is the root mean-square error (rmse) between the original and the distorted picture. The measure is usually applied to the signal after nonlinear conversion to offset for the Weber effect (e.g., γ (power law) correction or logarithmic conversion). It is a simple matter to incorporate frequency weighting of the error to further improve performance. However, such a measure is signal independent and therefore cannot reflect the effect of masking. The chief appeal of the rmse is its simplicity.

Fig. 19 shows the performance of the rmse [80] measure. Subjects rated five different pictures to which 16 different types of distortion were added, producing 80 pictures in all. For each picture a point is shown plotting rmse against the average of the observers' ratings, made using a 5-point impairment scale. If the rmse were a good measure of picture quality, the points would lie close to a smooth monotonically decreasing curve. The extent to which the points deviate from a smooth curve may be used as an indication of the performance of the measure. As can be seen, there is quite a large spread in reported quality for pictures having similar rmse.

A study of more sophisticated models was made. Best results were obtained with the model shown in Fig. 20. We will consider briefly the operations of filtering, masking and error summation. One would expect that the best filter would have emphasized errors occurring in the mid-frequency range (say 2–8 cpd) as suggested by the frequency response curve of Fig. 14. However, for the pictures and distortions considered in this study best results were obtained for a filter with an essentially flat frequency response or with a small amount of attenuation at high frequencies; we can only postulate why this is so.

In a simulation of the masking action of vision, an activity function was first evaluated which measures the "amount of change" or "amount of busyness" in the signal in a neighborhood close to the point being processed. This value, after a nonlinear mapping, was used to normalize the error signal.

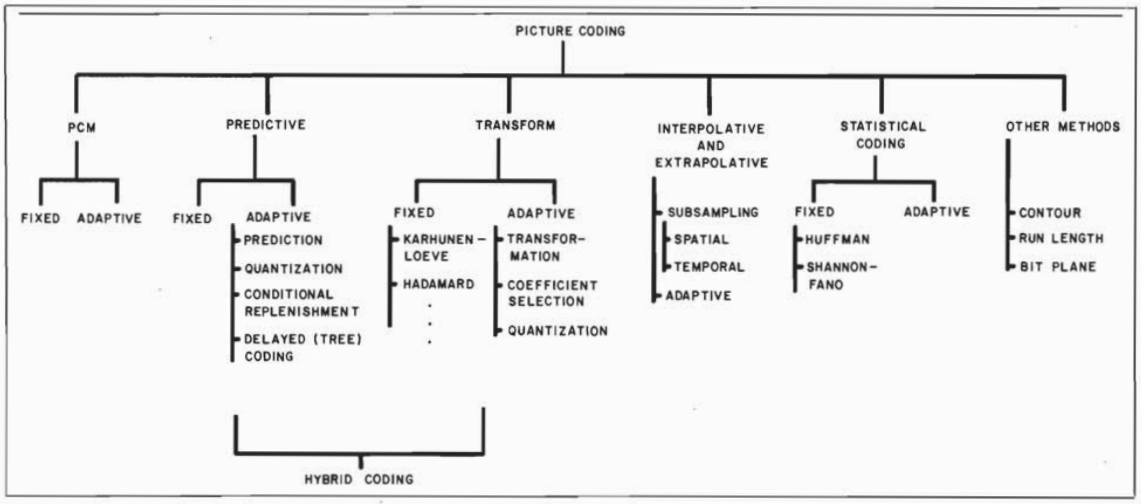
It was postulated that the observer does not necessarily sum errors over the whole picture as one does with the mse. The observer bases his estimate on the worse few local areas. Thus weighted error signal was first squared and then summed over local areas $1^\circ \times 1^\circ$ (corresponding roughly to the size of the fovea). The average of the errors in the two local areas having the greatest value of summed errors was taken as the final measure of quality. The performance of the measure is shown in Fig. 19(b). Comparison with Fig. 19(a) shows very little improvement over rmse. There is good reason for this. Unless distortions are carefully tailored to concentrate error at edges, the distortion will be most visible in flat areas of the picture and overall quality will be determined mainly by these areas. Now since mean-square-error predicts quality reasonably well for areas of just this one type, rmse will provide a good prediction of overall quality. Thus, if distortion should be greater at edges, such as in DPCM coding (see Section IV-B), the rmse is less satisfactory [80]. It is only through development of more sophisticated models that we can discover where to better "hide" coding inaccuracies with consequent improvement in coding efficiency.

IV. Coding Approaches

In this section, we first give a classification of the approaches that have been used for picture coding and then describe them in detail giving an account of the procedures for optimizing their parameters. The classification is given in Table II. General waveform coding can be classified into four major categories: PCM, predictive coding, transform coding, interpolative and extrapolative coding, and the fifth class which consists of miscellaneous schemes that do not fall into any of these four major classes. Each of these classes can be further divided based on whether the parameters of the coder are fixed or whether they change as a function of the type of data that is being coded (adaptive).

In PCM a time discrete, amplitude discrete, representation of the sample is made without removing much statistical or perceptual redundancy from the signal. The time discreteness is provided by sampling the signal generally at the Nyquist rate; amplitude discreteness is provided by using a sufficeint number of quantization levels so that the degradation due to quantization is not easily visible. In predictive coding, also known as DPCM, the sample to be encoded is predicted from the encoded values of the previously transmitted samples and only the prediction error is quantized for transmission. Such an approach can be made adaptive either by changing the prediction or quantization or by not transmitting the prediction error whenever it is below a certain threshold, as in conditional replenishment. It is also possible to delay the encoding of a sample by observing the trend of the signal as indicated by certain subsequent samples. In transform coding, an alternative representation of the signal is made first by taking linear combinations of samples in a block of data (called the coefficients) and then quantizing the selected coefficients for transmission. Several transformations (such as the simple Hadamard to a fairly complex data dependent Karhunen–Loeve) have been used. Transform coders can be made adaptive by changing the type of transformation and the criteria for selection and quantization of the coefficients. The next class of coding techniques, called interpolative and extrapolative coders, attempts to send certain samples to the receiver and either interpolate or extrapolate all the rest. Again, adaptation can be built in by vary-

TABLE II



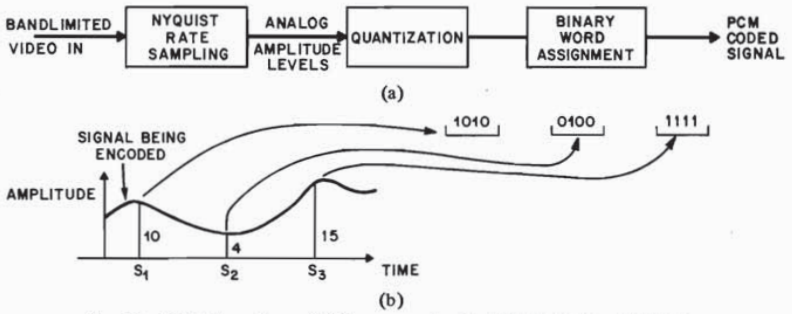


Fig. 21. PCM Encoding. (a) Components of a PCM encoder. (b) Four-bit binary representation of amplitude levels between 0 to 15.

ing the criterion for selection of the samples to be sent and the strategy for interpolating or extrapolating the remaining samples. Predictive and transform coding can be combined by techniques of hybrid coding in which the linear transformation of the signal is made first, and then predictive coding of the resulting coefficients from the adjacent (spatially or temporally) blocks is done. Besides these four categories, there are many other schemes that do not fall precisely into the four classes or are a result of unique combinations of schemes of these four classes. Some of them are applicable to the types of picture data that are not being considered in this review paper (e.g., run length coding for bilevel pictures). Notable among these schemes is the contour coding scheme, in which the picture signal is first separated into high contrast edges or contours and all the rest. Contours are typically coded by some form of run-length coding, and the other part of the signal is coded either using transform coding or predictive coding. Run-length coding of two-level signals can be extended to multilevel signals by coding the runs in several bit planes. As we explained in the introduction, the binary representation of either PCM, DPCM, or transform coded signals can be made based on the statistics of the occurrence of the various source symbols. An efficient method of assigning a comma-free code to the source symbols, such that the average number of transmitted bits is minimized, is given by the Huffman codes. The code assignment depends upon the probability of occurrence of the source symbols; therefore, this assignment can be fixed based on a long-term average statistic, or adaptive based on the short-term statistic.

A. Pulse Code Modulation

Waveform coding by PCM [81] is nothing more than a time discrete, amplitude discrete, representation of the signal. It was first applied to television signals by Goodall [82] in 1951 and continues to be used as a digitizing scheme for purposes of transmission and also for digitizing before the application of other more sophisticated coding techniques. As shown in Fig. 21 basic PCM consists of sampling the waveform (usually at the Nyquist rate) and quantizing each sample using N levels. Each level is represented by a binary word containing B bits. N is usually taken to be a power of two (i.e., $N = 2^B$). In the decoder, these binary words are converted to discrete amplitude levels, and then the time sequence of the amplitude levels is low-pass filtered. Basic PCM affords a simplicity uncommon to most other coders, but suffers from inefficiency since it does not use redundancy present in the picture signal.

Application of PCM to television [82] produces distortions which depend upon both the number of quantizing levels used and the signal-to-noise (S/N) ratio of the input signal. Several efforts have been made to optimize the placement of the quantizing levels to take advantage of the fact that the noise detection threshold ΔL increases almost linearly with increasing luminance L (known as Weber's law; see Section III-D). This implies that the visibility of unit quantization noise decreases with the luminance level, and, therefore, coarseness of the quantizer is usually made to increase with the luminance level. As shown by Kretz [56], this does improve the quality of PCM-coded pictures for a given number of bits per sample. However, the improvements are not very significant due to the

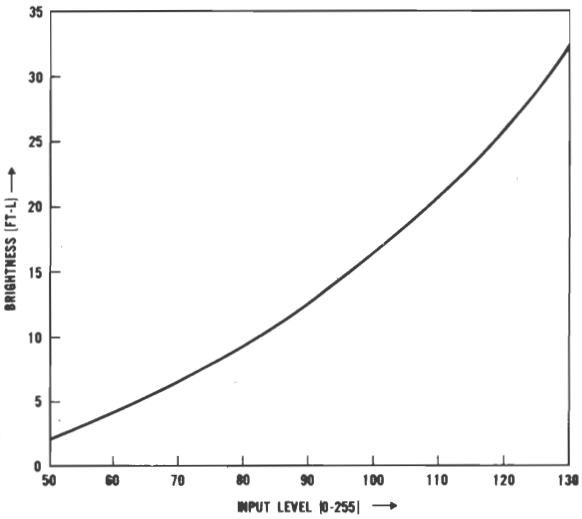


Fig. 22. Nonlinear input/output characteristic of a cathode-ray tube (from Sharma and Netravali [121]).

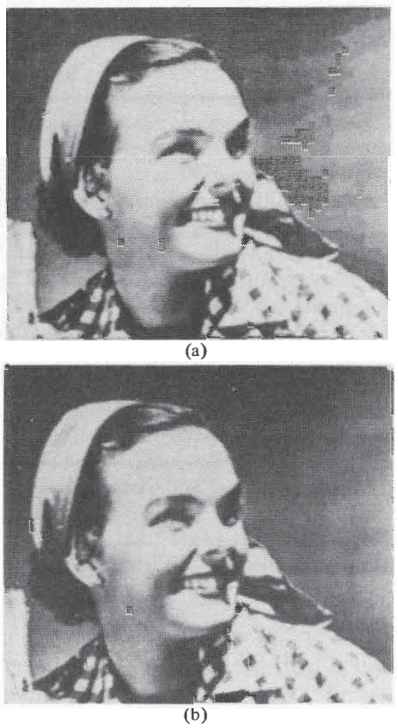
nonlinear characteristics of most cathode ray tubes used in television display. The nonlinear characteristics shown in Fig. 22, referred to as the gamma of the display, increases the amplitude of the quantization error at larger values of the luminance to some extent and, therefore, partially compensates for the Weber's law effects.

For good quality original pictures, as the number of quantization levels is decreased, quantization errors are seen as false contours in low detail areas of the picture. This is shown in Fig. 23(a) where a 5-bit PCM encoded picture is shown starting with an original signal of 50-dB S/N ratio. As seen in this picture, the quantization contours are visible in the sky. Visibility of this quantization noise can be decreased by adding some high-frequency noise to the original signal before quantizing. This noise causes the coded signal to oscillate between the quantizing levels thereby increasing the frequency content of the quantizing noise. This is seen in Fig. 23(b) where the quantizing contours are no longer visible. Noise in the input signal can be increased by deliberately adding a pseudorandom noise sequence to the input before quantization and subtracting it out at the receiver. This technique, called dither, was used successfully by Roberts [83] to produce good-quality PCM coded pictures at low bit rates. Deterministic patterns have been used to produce further improvement in two and three dimensions [84], [85].

PCM coding systems, in general, require about 128 to 256 levels (7–8 bits) for good-quality pictures under most viewing conditions. For monochrome television with a sampling rate of 8 MHz, this amounts to a bit rate of about 56 to 64 Mb/s.

B. Predictive Coding

In basic predictive coding systems [86]–[88] (Fig. 24) a prediction of the sample to be encoded is made from previously coded information that has been transmitted. The error resulting from the subtraction of the prediction from the actual value of the sample is quantized into a set of discrete amplitude levels. These levels are represented as binary words of either fixed or variable word length and sent to the channel coder for transmission. Thus the predictive coder has three basic components: 1) predictor, 2) quantizer, 3) code assigner.



(a)

(b)

Fig. 23. Effect of coarse quantization in PCM coded picture (from Connor *et al.* [23]). (a) Contouring is visible in this high input signal-to-noise ratio, five-bit PCM picture. (b) Increasing input noise level reduces the visibility of contours.

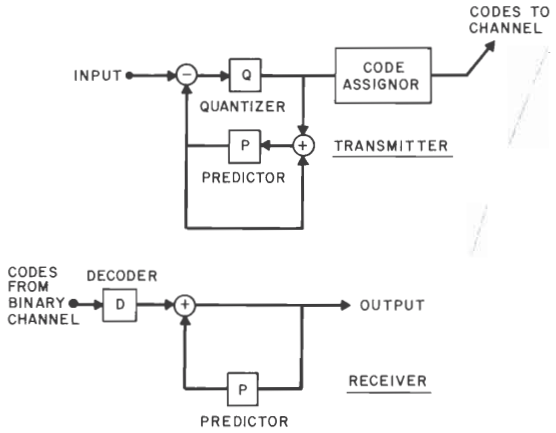


Fig. 24. Block diagram of a DPCM transmitter and receiver.

Depending upon the number of levels of the quantizer, a distinction is often made between delta modulation (DM) [86] ($N = 2$) and DPCM [87], which has N greater than two. Although DM has been used extensively in encoding other waveforms (e.g., speech), it has not found great use in encoding of pictures, due perhaps to the high sampling rates required; consequently, we will limit our discussion to DPCM encoders. In its simplest form, DPCM uses the coded value of the horizontally previous pel as the prediction. However, more sophisticated predictors use the previous line (two-dimensional predictor) as well as previous frame of information (interframe

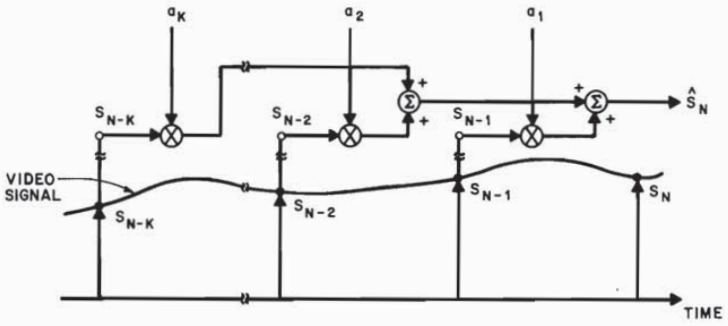


Fig. 25. An outline of linear prediction used for predictive coding. Note that quantization effects are neglected. The prediction, therefore, is made by using a weighted sum of k previous samples of the original signal, with $\{a_i\}_{,i=1,\cdots,K}$ as weights.

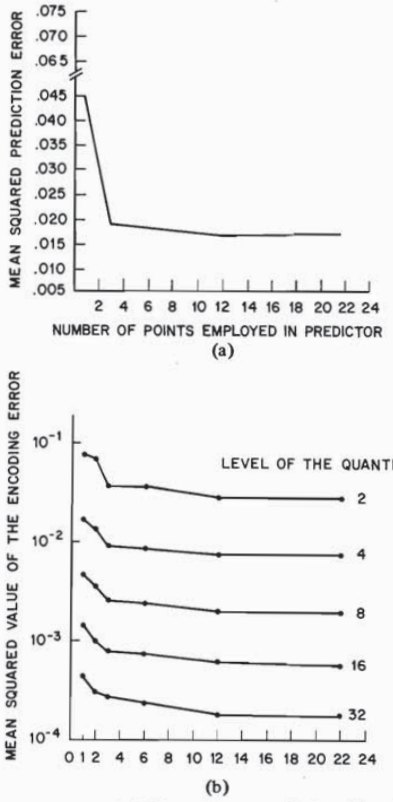


Fig. 26. Performance of different linear predictors (from Habibi [28]). (a) mse is plotted as a function of the order of the predictor (number of points employed in the predictor). (b) Mean-square error of DPCM systems as a function of the order of the predictors and using different quantization levels. Statistically optimum (Max) quantizers are used.

predictor). In this subsection we shall look at how the three components of the predictive coder have been optimized for coding of the picture signal.

1) Predictors: Predictors for DPCM coding can be classified as linear or nonlinear depending upon whether the prediction is a linear or a nonlinear function of the previously transmitted sample values. A further division can be made depending upon the location of previous elements used: one-dimensional predictors use previous elements in the same line, two-dimensional predictors use elements in the previous lines as well, whereas interframe predictors use picture elements also from the previously transmitted frames. Predictors can be fixed or adaptive. Adaptive predictors change their characteristics as a function of the data, whereas fixed predictors maintain the same characteristics independent of the data.

Linear predictors for television have been studied using the general theory of linear prediction [89]. If $\{S_i\}$ are a set of picture elements indexed according to their time occurrence, and assuming them to be identically distributed with zero mean and variance σ^2, a linear predictor for the nth element S_n using K previous elements, $S_{n-1}, S_{n-2}, \cdots, S_{n-K}$, can be written as

$$\hat{S}_n = \sum_{i=1}^{K} a_i S_{n-i}. \tag{6}$$

This is shown in Fig. 25. The coefficients $\{a_i\}_{i=1,\cdots,K}$, can be obtained by minimizing the mean-square prediction error (mse). The best coefficients are given by

$$\text{col}\,(a_1, \cdots, a_K) = R^{-1}\ \text{col}\,(R_1, R_2, \cdots, R_K) \tag{7}$$

where $R_i = \overline{(S_n \cdot S_{n-i})}$, and matrix R has (i, j)th element $\overline{(S_{n-i} \cdot S_{n-j})}$, where the bar on the top denotes an average. Using these coefficients, the mse is given by $\sigma^2 - \sum_{i=1}^{K} a_i R_i$. Thus the mean square of the input to the DPCM quantizer is reduced by $\sum_{i=1}^{K} a_i R_i$ from σ^2, the mean square of the input to a PCM quantizer.

It should be noted that the above analysis assumes stationarity and neglects the effects of quantization in the DPCM coder (i.e., in a real DPCM encoder, prediction of sample S_n can only be made by using encoded representations of the past samples $S_{n-1}, S_{n-2}, \cdots, S_{n-K}$ and not by using the original uncoded sample values). For coders which produce high quality pictures, effects of quantization are small and may be neglected. The mse has been computed by Habibi [28] for predictors which use different numbers of picture elements within a frame. His results, reproduced in Fig. 26, show that if the predictor coefficients are matched to the statistics of a picture, then for that picture the mse decreases significantly by using up to three picture elements; further decreases are rather small. However, if coefficients are not exactly matched, the decrease in mse is not significant by using three previous elements as compared to one element. Also, this improvement with the number of picture elements changes little with the coarseness of the quantizer in the DPCM loop (Fig. 26(b)).

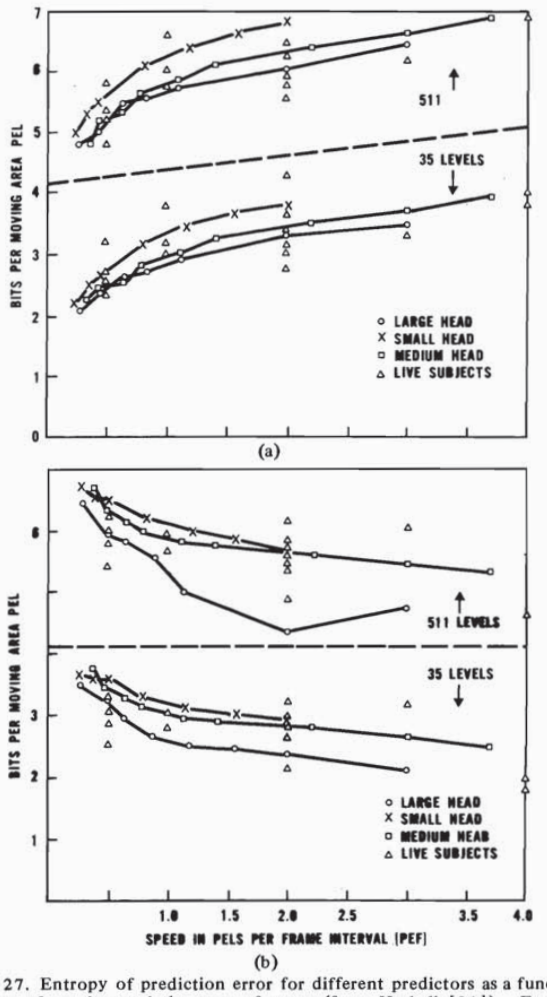


Fig. 27. Entropy of prediction error for different predictors as a function of motion and the type of scene (from Haskell [91]). Four different scenes are used. (a) Entropy of the frame difference signal in the moving area versus speed. Starting with 8-bit PCM, the differential signal could assume any of 511 different levels (due to 8 bit initial quantization). Results are also shown for coarser quantization of signals, which still gives good picture quality. Solid curves are for the mannequin head at various distances from the camera. Unconnected points are for live subjects. (b) Entropy of the element difference signal in the moving area versus speed.

Two-dimensional predictors have also been used for DPCM coding. In general, the improvement in mse by using two-dimensional prediction is small [29]. However, subjective evaluation indicates that the rendition of vertical edges due to two-dimensional prediction is significantly improved [90]. Also, by proper choice of the coefficients, it is possible to get improved prediction as well as quick decay of the effects of the transmission errors in the reconstructed picture.

Predictors for frame-to-frame coding have used a combination of elements from the present frame as well as previous frame. For scenes with low detail and small motion, frame difference prediction appears to be the best. In scenes with higher detail and motion, field difference prediction does better than frame difference prediction [91]. As the motion in the scene is increased further, intrafield predictors do better [92]. This is largely due to two reasons: 1) for higher motion, there is less correlation between the present pel and either the previous field or the frame pels, 2) due to the integration of the signal in the video camera, the spatial correlation of the television signal in the direction of movement is increased [34]. For the same reason, predictions such as element or line difference of frame or field differences perform better than frame or field difference for higher motion. A typical variation of the entropy of the prediction error for elements which are significantly different from the corresponding elements of the previous frame (or "moving area pels") is shown in Fig. 27 as a function of different predictors and motion of the object in the scene [91].

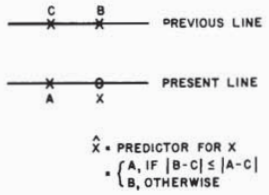


Fig. 28. Graham's rule for adaptive prediction.

We mentioned before that the picture signal is highly nonstationary, and that it is advantageous to change the prediction based on the local properties of the picture signal. For the intraframe situation, one popular method has been to compute some measures of directional correlation based on the local neighboring transmitted pels and use it to select a predictor along the direction of largest correlation. The set of predictors from which a predictor is selected are usually linear and are chosen such that each one of them will give small prediction error if the signal was correlated in a certain manner. Examples of this type of approach are the predictors used by Graham [93], Zschunke [94], and Dukhovich and O'Neal [95]. In Graham's predictor, either the previous line or the previous element is used for prediction, and the switching is done by the surrounding line and element differences as shown in Fig. 28. Several extensions have been made of this basic philosophy. However, the results have not been very encouraging in terms of the mse or the entropy of the prediction error. In some cases the rendition of certain types of edges can be remarkably improved by these adaptive predictors. Another variation [96] in adaptive prediction is to use a weighted sum of several predictors, where the weights are switched from element to element and are chosen by observing certain characteristics of already transmitted neighboring pels. As distinct from the earlier "switched" predictors, they may not use weights of 1 or 0. As an example, assume that pel A is already transmitted (Fig. 28). The prediction error of A using all the predictors in a set of predictors is then evaluated. The predictor that gives the least prediction error is used for the prediction of X. The same calculation can be performed at the receiver and, therefore, the predictor switching information does not need to be transmitted. Such techniques have been successful in color component coding [96], and are being considered for gray-level signals [97].

Generalizations of Graham's rule have been made for frame-to-frame prediction, where one selects either a previous frame or an intraframe predictor, depending on surrounding information [95]. However, more successful adaptive predictors for frame-to-frame coding are the ones that take into account motion of objects. These are based on the notion that, if there are objects moving in the field of view of a television camera and if an estimate of their translation is available, then more efficient predictive coding can be performed by taking differ-

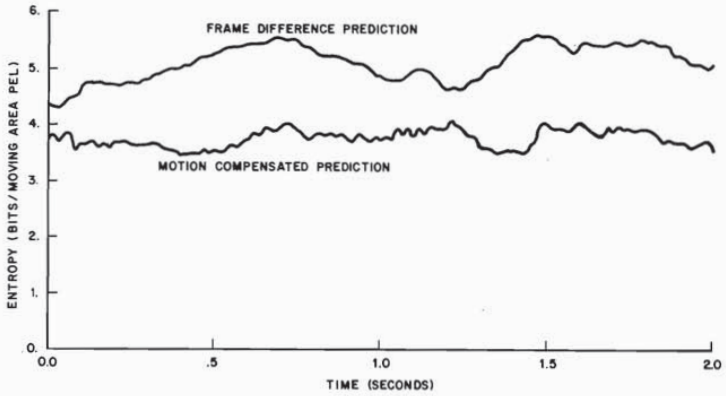


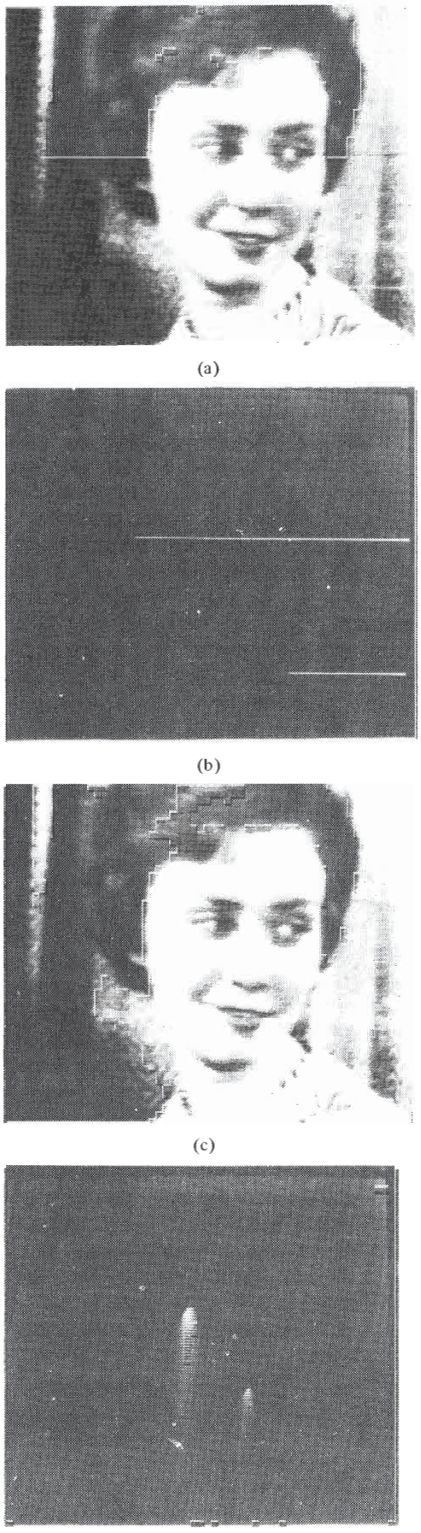
Fig. 29. Improvement in entropy of the prediction error due to motion compensation for a typical scene containing fairly active head and shoulders movement (from Netravali and Robbins [102]).

ences of elements with respect to elements in the previous frame that are appropriately spatially translated. Such prediction has been called motion compensated prediction [98], [99]. Its success obviously depends upon the amount of translational motion of objects in real television scenes and the ability of an algorithm to estimate translation with the accuracy that is desirable for good prediction. One set of techniques developed by Limb and Murphy [100], and Rocca [101] obtain an estimate of translation in a block of pels, whereas techniques developed by Netravali *et al.* [102], [103], [104], recursively adjust the translational estimate at every pel or at every small block of pels. Another approach to motion compensation is adaptive linear prediction by using elements in both the present and the previous frame (or field), which surround the element being encoded, and adapting the coefficients to minimize an intensity error function [105]. Such an approach is implementationally difficult and requires transmission of coefficients of the predictors.

Performance of motion-compensated prediction for video signals with initial quantization of 8 bit/sample is superior to the frame difference predictor by about one or two bits (entropy of the prediction error) depending upon the type of motion in the scene. A typical set of results for motion-compensated prediction is shown in Fig. 29 for a scene containing head and shoulders view of a person involved in an active conversation and occupying about 15–51 percent area of the picture. Four frames of this scene are shown in [102]. In terms of the total coder bit rate, the decrease due to motion compensation might be 20–70 percent [103]. With the cost of processing coming down, future coder implementations will certainly take advantage of this significant bit-rate reduction.

2) *Effect of Transmission Errors:* One of the questions that is often overlooked in the design of predictors is the propagation of transmission errors in the reconstructed pictures. Since prediction of the next sample is made in the DPCM coder using certain previously transmitted samples, a transmission error in any of the previous samples could affect the present sample which in turn could affect all the future samples. The amount and spread of this effect depends upon the type of predictor used. In the case of optimum linear one-dimensional predictors, it is known [106] that for most types of correlations generally found in pictures, the predictor is stable, which implies that the effect of transmission error decays. The previous element predictor is unstable, and transmission errors result in horizontal streaks in the reconstructed pictures. However, this predictor can be stabilized by providing leak, i.e., using a fraction (<1) of the previous element value for prediction. For two-dimensional predictors the situation is somewhat difficult. Optimum linear predictors for many commonly used correlation functions may be unstable [106]. A good criterion for stability is not yet known [107], [108]. However, as pointed out earlier, there are some stable two-dimensional predictors for which the pattern of distortion produced by transmission errors is much less annoying than the one-dimensional predictors [90], [107]. Many methods of error concealment and/or correction exist. In some, an erroneous scan line is substituted by either previous line or an average of the surrounding lines [109]. Other methods depend upon the error pattern and the statistics of the video signal for error correction [110], [111]. In the case of adaptive predictors, a transmission error can have two types of effects—one due to the use of the wrong value for the prediction, and the other due to selecting a wrong predictor. Some improvement in the performance may be obtained by providing a "leak" in the rule of adaptation as suggested by Maxemchuk and Stuller [112]. However, in general, criteria for stability of such predictors and methods of transmission error correction are much less understood. The only recourse appears to be the use of error detecting and correcting codes. Some of the patterns produced by transmission errors for a previous element predictor and a spatial average predictor are shown in Fig. 30.

3) *Quantization:* DPCM schemes achieve compression, to a large extent, by not quantizing the prediction error as finely as the original signal itself. Several methods of optimizing quantizers have been studied, but quantizer design still remains an art and somewhat ad hoc. Most of the work on systematic procedures for quantizer optimization has been for the previous element DPCM coding, in which approximate horizontal slope of the input signal is quantized. Although obvious extensions can be made to the case of two-dimensional and interframe predictors, they have not yet received enough attention. For this reason, we shall discuss in detail quantizers for the previous element DPCM and point out how extensions to the other cases can be made. Three types of degradations can be seen due to improper design of the quantizer of a DPCM coder. These are referred to as granular noise, edge busyness and slope overload as shown in Fig. 31. If the inner levels (for small magnitudes of differential signal) of the quantizer are



(a) (b) (c) (d)

Fig. 30. Effect of transmission errors in DPCM coders (from Connor *et al.* [23]). (a) Previous element prediction. (b) Error patterns for previous element prediction. (c) Spatial average prediction. (d) Error patterns for spatial average prediction.

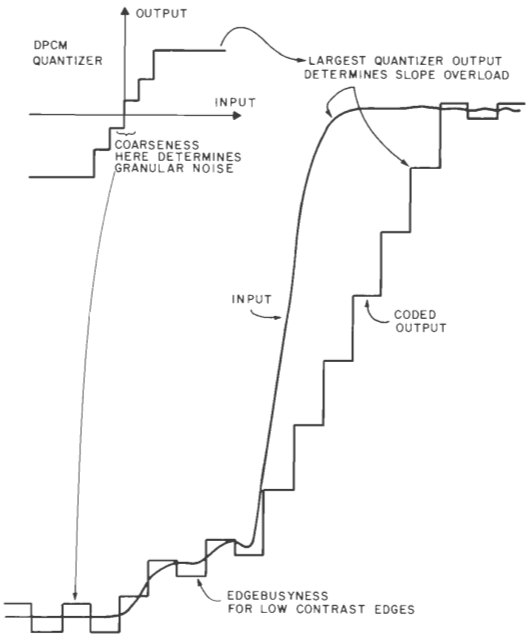


Fig. 31. An intuitive classification of quantizing distortion due to DPCM coding. Three classes of noise are identified: granular noise, edge busyness, and slope overload.

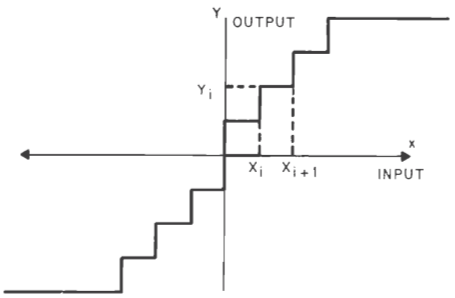


Fig. 32. Characteristics of a quantizer. x is the input and y is the output. $\{X_i\}$ and $\{Y_i\}$ are decision and representative levels, respectively. Inputs between X_i and X_{i+1} are represented by Y_i.

too coarse, then the flat areas are coarsely quantized and have the appearance of random noise added to the picture. On the other hand, if the dynamic range (i.e., largest representative level) of the quantizer is small, then for every high contrast edge it takes several samples for the output to follow the input, resulting in slope overload, which appears similar to low-pass filtering of the image. For edges whose contrast changes somewhat gradually, the quantizer output oscillates around the signal value and may change from line to line, or frame to frame, giving the appearance of a "busy edge." Quantizers can be designed purely on a statistical basis or by using certain psychovisual measures. They may be adaptive or fixed. We shall discuss each of these cases in the following sections.

a) Nonadaptive quantization: Optimum quantizers that are statistically based have been derived using the work of Max [113]. Considering Fig. 32, if x is the input to the DPCM quantizer with probability density $p(x)$, then quantizer parameters can be obtained to minimize the following measure of

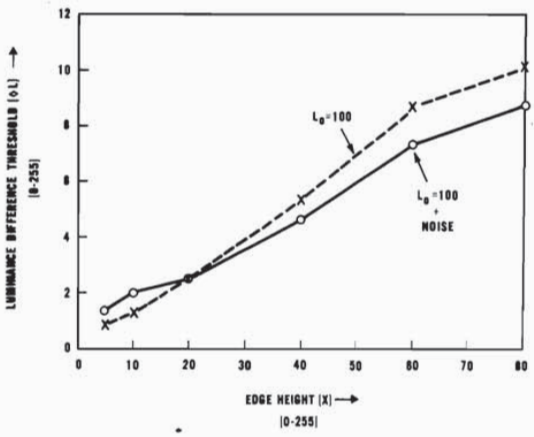


Fig. 33. Amplitude threshold versus slope for a single luminance edge which rises within one element to its peak value. Edge height and luminance difference threshold are plotted in units 0 to 255 (8-b number) for two cases 1) without any additive noise and 2) with additive noise (from Sharma and Netravali [121]).

quantization error:

$$D = \sum_{i=1}^{N} \int_{X_i}^{X_{i+1}} f(x - Y_i) \cdot p(x)\, dx \tag{8}$$

where $X_1 < X_2 < \cdots < X_{N+1}$, and $Y_1 < Y_2 < \cdots < Y_N$ are decision and representative levels, respectively, and $f(\cdot)$ is a nonnegative function. We assume, as shown in Fig. 32, that all inputs $X_i < x \leqslant X_{i+1}$ to the quantizer are represented as Y_i. Necessary conditions for the optimality with respect to X_i and Y_i for a fixed number of levels N are given by:

$$f(X_j - Y_{j-1}) = f(X_j - Y_j), \quad j = 2, \cdots, N \tag{9}$$

and

$$\int_{X_j}^{X_{j+1}} \frac{df(x - Y_j)}{dx}\,(x - Y_j) \cdot p(x)\, dx = 0, \quad j = 1, \cdots, N. \tag{10}$$

assuming that $f(\cdot)$ is differentiable. In the case of the mean-square error criterion, $f(z) = z^2$, and the above equations reduce to

$$X_j = (Y_j + Y_{j-1})/2 \tag{11}$$

and

$$Y_j = \int_{X_j}^{X_{j+1}} x p(x)\, dx \Big/ \int_{X_j}^{X_{j+1}} p(x)\, dx. \tag{12}$$

Several algorithms and approximations exist for the solution of these equations. The probability density, $p(x)$, for the case of a previous element differential coder can be approximated by a Laplacian density, and the optimum quantizer [114] for this case is companded, i.e., the step size $(X_j - X_{j-1})$ increases as x increases in magnitude. Since most of the time the picture signal has slope x close to zero, this results in fine spacing of levels near $x = 0$ and rather coarse spacing of levels as $|x|$ increases. This results in overspecification of the low detailed areas of the picture and consequently a small amount of granular noise but poor reproduction of detailed areas or edges. Due to the small dynamic range of the quantizer, this is visible

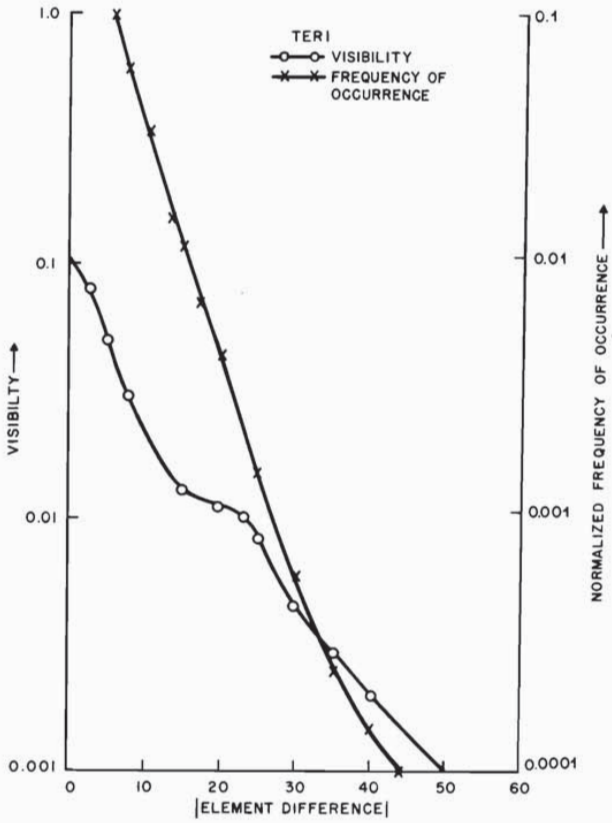
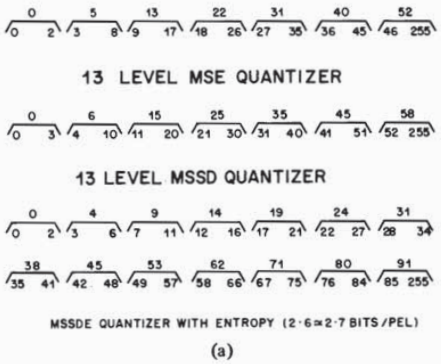


Fig. 34. Weighting function for quantization noise visibility as related to the magnitude of the element difference for a typical head and shoulders view. Also shown is a histogram of the magnitude of the element differences (from Netravali [117]).

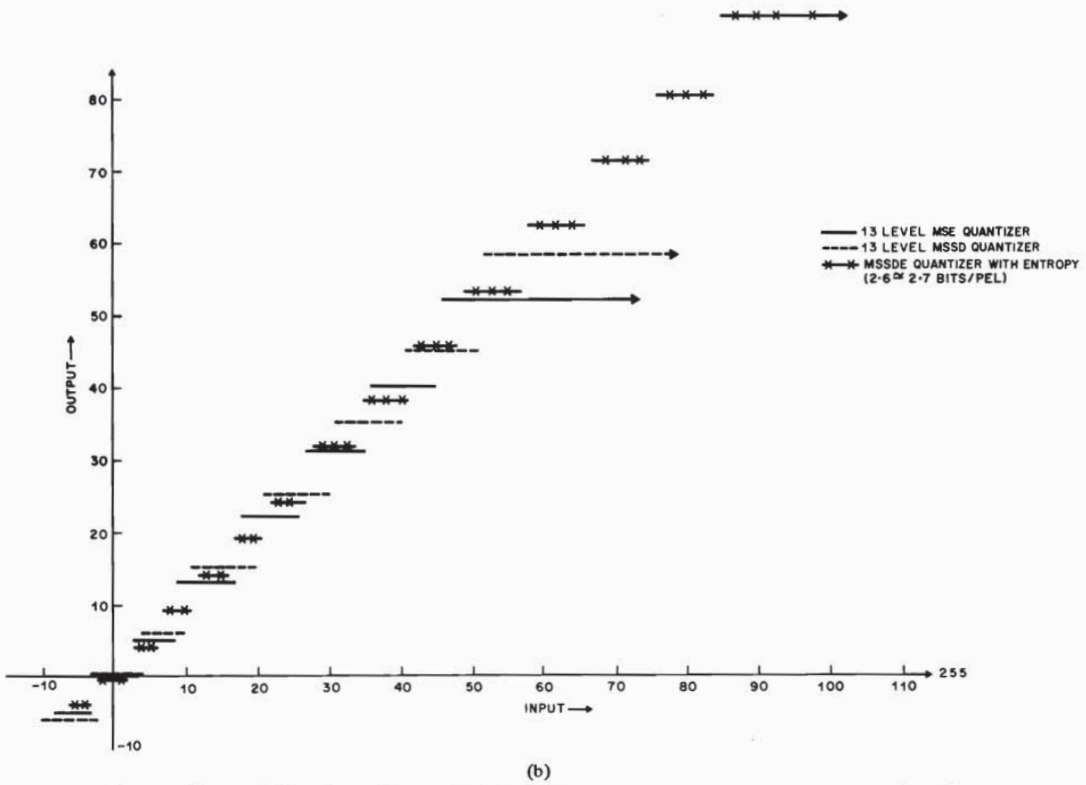
as edge busyness for high contrast, vertical edges. Thus for subjective reasons, the qauntizers should not be designed based on the mean square error criterion.

Minimization of an error measure for a fixed number of levels is relevant for DPCM systems which represent the quantizer outputs with fixed-length binary words, since in that case the output bit rate depends on the logarithm of the number of levels of the quantizer. However, since the distribution of occurrence of different quantizer levels is highly skewed, there is a considerable advantage to using variable length words [3] to represent quantizer outputs. In such a case, since the average bit rate is lower bounded by the entropy of the quantizer output, it is more relevant to minimize a measure of distortion subject to the entropy of the quantizer output being less than a given number. It is known that [115], for random variables with Laplacian density, quantizers optimum with respect to this criterion are uniform. Although such quantizers may provide adequate picture quality for small step sizes, they may be complex to implement because of the large number of levels [116].

It had been realized for some time that for a better picture quality, quantizers should be designed on the basis of psychovisual criteria. However, a debate [117]–[119] continues on what is a good criterion to use, and expectedly so, considering the complexities of the human visual system (see Section III-E). We will describe two approaches and point the reader to other works. In the first method, a quantizer is designed such that the quantization error is at the threshold of visibility



(a)



(b)

Fig. 35. Characteristics of quantizers optimized under different criteria (from Netravali [117]). (a) Input and output characteristics of quantizers optimized under three criteria are shown: (1) minimum mean-square error (mse); (2) minimum weighted mean-square error (mssd); (3) mssd with a constraint on the entropy (mssde) of the quantizer output. For (2) and (3), weights are subjectively derived. (b) Quantizer levels are shown in units of 0 to 255 (eight bits).

(using a certain model of the error visibility) while minimizing the number of quantizer levels or the entropy of the quantizer output. We mentioned in the previous section that the visibility of small perturbations in intensity is masked by spatial variations of intensity. Threshold at an edge is the value by which the magnitude of an edge can be perturbed such that the perturbation is just visible. A typical relationship [120], [121] for an edge of one pel width, and luminance $L_0 = 100$ (on a scale of 0–255) and variable height, is given in Fig. 33. Knowing this relationship[7] between the edge threshold and slope of a single edge the quantization error can be constrained to be below the amplitude threshold for all slopes of the signal. Optimum quantizers satisfying such a constraint turn out coincidentally to be the same whether the number of levels is minimized or the entropy is minimized [121]. For a videotelephone type signal, 27 levels are required for no visible error at normal (six times the picture height) viewing distance [121].

Another method of designing psychovisual quantizers is to minimize a weighted mean-square quantization error, where the weights are derived from subjective tests. Such optimization would be similar to the mean-square error quantization, where the probability density is replaced by a weighting function. Limb [120] used a product of the probability density

[7] Real pictures instead of a single edge, can be used to determine this relationship [122].

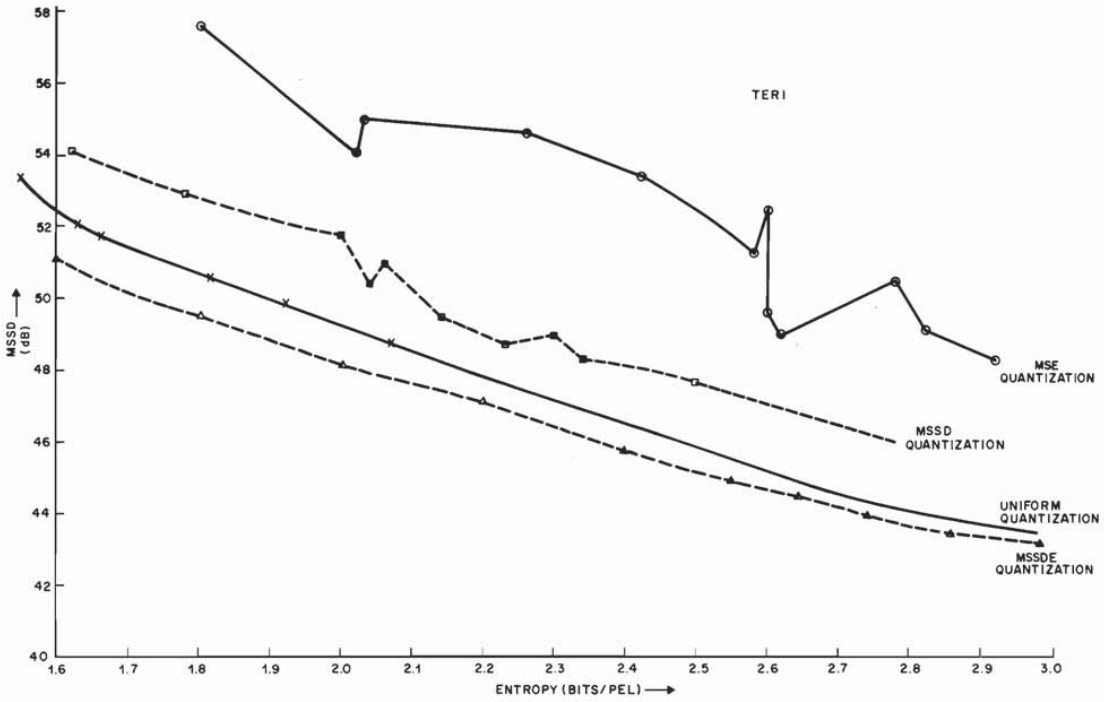


Fig. 36. Performance of quantizers optimized under different criteria for a typical head and shoulders picture "Teri." Mean-square weighted quantization noise (mssd) is plotted with respect to entropy. It is assumed that mssd is a reasonable measure of picture quality (from Netravali [117]).

and a function of the amplitude threshold as the weighting function, whereas Candy and Bosworth, [123] and Netravali [117] have used complex pictures as stimuli and derived a weighting function by measuring visibility of noise added to regions of the picture as determined by the edge contrast. A typical weighting function for a head and shoulders type of picture is shown in Fig. 34. Such weighting functions are picture dependent, but the variation for a class of pictures (e.g., head and shoulders view) may not be significant. There has been and perhaps will continue to be considerable debate on the choice of weighting functions. Although the optimum choice of the weighting function is unknown, optimum quantizers designed on the basis of the above weighting functions have been quite successful. In general, quantizers which minimize the weighted mean square error for a fixed number of levels are less companded than the minimum mean-square error quantizers and reproduce edges faithfully. Figs. 35 and 36 give characteristics of the quantizers optimized under different criteria and the performance of these quantizers in terms of entropy of the quantizer output and a measure of picture quality. Other criteria for quantizer design and their comparisons have been considered in [117], [118], [121], [124], [125].

Quantizer design for intraframe coders which use a more sophisticated predictor has been either on the basis of minimum mean square error or by trial and error. However, the methods of psychovisual quantizers discussed above can be extended easily for this case. For interframe coders, the situation is somewhat complicated since the quantization error occurs predominantly in the moving areas of the picture and its visibility depends upon the spatial as well as temporal variations in the scene. Because of the lack of suitable models for

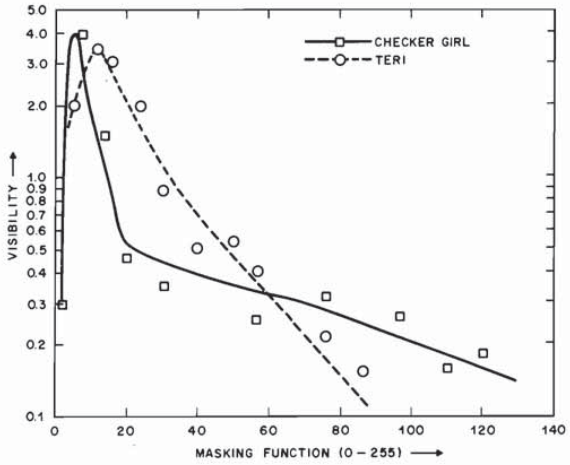


Fig. 37. Relationship between visibility of noise and a measure of spatial detail (called Masking function), which uses a combination of element and line differences in a 3 X 3 neighborhood. Dependence of the visibility on the picture content is shown for two head and shoulders views, checkered girl (Fig. 38(a)) and "Teri" (from Netravali and Prasada [131]).

both the distribution of the prediction error and the visibility of quantizing error, the problem of designing both minimum mean-square and psychovisual quantizers has not been tackled. Quantizers used in practice are based on trial and error.

b) Adaptive quantization: Due to the variation of the picture statistics and the required fidelity of reproduction in dif-

Fig. 38. Original picture (a) and its four segments (b), (c), (d), (e) designed on the basis of spatial detail (two-dimensional) and noise visibility. Segment (b) has highest noise visibility, whereas segment (e) has the least noise visibility (from Netravali and Prasada [131]).

ferent regions of the picture, adapting the DPCM quantizer is advantageous. In general, one would like to segment a picture into several subpictures such that both the perception of the quantization noise and the statistical properties of the differential signal within the subpicture are uniform and stationary. However, this is an extremely difficult task since the perception of noise and the statistics may not be sufficiently related to each other. Several approximations to this goal have been made, some purely statistical [126], [127] and some based on certain psychovisual criteria [128]-[133]. One can work in the frequency domain [128], [129] and split the signal into two frequency bands to exploit detail sensitivity of the eye. In such a case, signal in the low-frequency band is sampled at a low rate, but is quantized finely because of the high sensitiv-

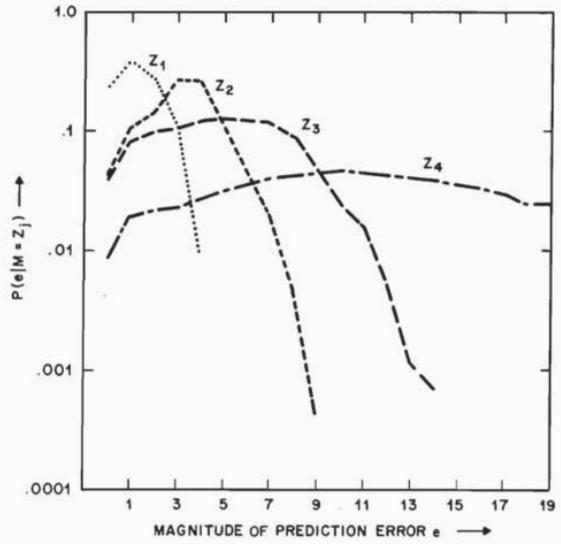


Fig. 39. Plots of conditional histograms of prediction errors conditioned on the four segments $\{Z_1, Z_2, Z_3, Z_4\}$ of Fig. 38 (from Netravali and Prasada [131]).

LEVEL NO.	CODE WORD LENGTH	CODE
1	12	100101010101
2	10	1001010100
3	8	10010100
4	6	100100
5	4	1000
6	4	1111
7	3	110
8	2	01
9	2	00
10	3	101
11	4	1110
12	5	10011
13	7	1001011
14	9	100101011
15	11	10010101011
16	12	100101010100

Fig. 40. A typical variable-length code for a DPCM coded signal, with 16 quantizer levels.

ity of the eye to the noise in the low frequency region. The high frequency component of the video signal is sampled at a higher rate, but is quantized coarsely due to reduced sensitivity to noise in highly detailed regions. Pel domain approaches [130]–[132] start out with a measure of spatial detail and obtain experimentally the relationship between the visibility of noise and the measure of spatial detail. A typical relationship for the head and shoulders pictures is shown in Fig. 37, where a weighted sum of slopes in a 3 X 3 neighborhood is used as a measure of spatial detail. Many other measures of spatial detail may be used. As an example see Musmann and Erdmann [132]. This relationship is then used to segment a picture into subpictures, such that a unit noise would be approximately equally visible in the entire subpicture. Such a segmentation, using the relationship shown in Fig. 37, is shown in Fig. 38. Fig. 38(a) shows an original picture and Figs. 38(b) through 38(e) show four subpictures which are in the order of increasing spatial detail or decreasing noise visibility. Thus flat areas of the picture are covered by the first subpicture (i.e., Fig. 38(b)) and severe edges are included in the last subpicture (i.e., Fig. 38(e)). Quantizers for each of the subpictures could be designed in a way similar to the earlier nonadaptive techniques. One side benefit of such a scheme is that the prediction error (i.e., the differential signal) statistics shows a marked change from one segment to the other. In Fig. 39, the prediction error histograms are shown for four subpictures of the picture in Fig. 38(a). It is evident that since these histograms are quite different from segment to segment, a different variable length code may be used to represent the quantizer outputs from each segment. This together with the adaptive quantizer is capable of reducing the bit rate by about 30 percent over the previous element DPCM coder [131]. However, much of this reduction is a result of coarser quantization of different segments rather than use of different variable codes for different segments [130]. It is also interesting to note that the advantage of adaptation increases with the number of segments used, up to four, after which further increase is small. Another method of adapting the quantizer, called delayed coding [133]–[135], is to treat the operations of quantization as that of following a path through the encoding tree determined by the quantizer steps. The path is chosen to minimize the weighted quantization error at several samples, including some subsequent samples. Since the quantization steps are not varied sample to sample, it is not necessary to send the adaptation information to the receiver. In summary, for intraframe coding, although the optimum rules for adaptation are not known, rules based on intuition, and trial and error have shown significant reduction in bit rate over nonadaptive systems. In the case of interframe coders, no systematic investigation of adaptive quantizers has been made, primarily due to lack of understanding of statistics as well as the perceptual basis for adapting the quantizers. Quantizers have also been adapted in an ad hoc manner to facilitate matching of the variable bit rate coder to the constant bit rate channel. We shall discuss this in detail next.

4) Code Assignment: We mentioned eariler that the frequency of occurrence of the quantizer output levels is not uniform for intraframe as well as interframe predictive coders and, therefore, lends itself to representation using code words of variable lengths (e.g., Huffman codes). The average bit rate for such a code is very close to and is lower bounded by the entropy of the quantizer output signal. A typical variable length code for a previous element DPCM coder with 16 quantizer levels is shown [136] in Fig. 40. Inner levels occur much more often and, therefore, are represented by a smaller length word. This type of optimum code requires exact knowledge of the probabilities of the quantized output. There are other procedures which work with approximate probabilities and remain reasonably efficient even when the probabilities start to depart from their approximate values [137]. The optimum code is not too sensitive to changes in the probability distribution [138] and, therefore, it appears that for picture-type statistics, optimum codes based on an average of many pictures, remain quite efficient compared with optimum codes designed on the basis of knowledge of the statistics of the individual pictures. For most pictures, the entropy of the quantizer out-

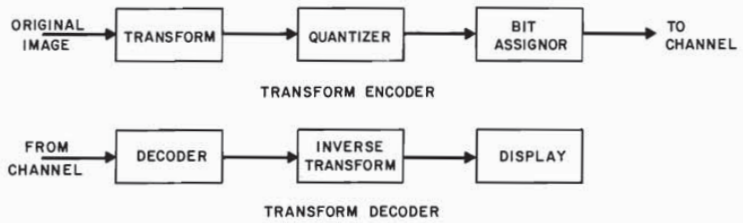


Fig. 41. Block diagram of a transform encoder/decoder pair.

put (and consequently the average bit rate using Huffman coding) is about one bit per pel less than the corresponding bit ($= \log_2 N$, N: number of quantizer levels) for a fixed length code [139]. Stated differently, an optimum variable length code results in an improvement of about 6-dB S/N ratio with no change in the average transmission bit rate. A similar conclusion was arrived at quantitatively by O'Neal [140], who proved that for a Laplacian density (a good approximation for the probability density of the differential signal), optimum variable length code improves the S/N ratio by about 5.6 dB if the number of quantizer levels is large.

One of the problems with the use of variable length codes is that the output rate from the source coder changes with local picture content. In order to send such a signal over a constant bit rate channel, the source coder output has to be held temporarily in a buffer which can accept inputs at a nonuniform rate and can be read out to the channel at a uniform rate. Since in any practical system a finite length buffer is used, there are problems of buffer overflow and underflow, which depend upon the size of the buffer, type of the variable length code used, and the channel rate. By using a channel whose rate is higher than the entropy of the quantizer output, probability of buffer overflow is reduced [141]. Also, by designing codes which minimize the probability of the length of the code words exceeding a certain number, probability of buffer overflow can be reduced [136]. However, since the buffer overflow cannot always be prevented, strategies have been designed by which, as the buffer begins to fill, the output bit rate of the coder is gradually reduced. This is particularly important in the case of frame-to-frame coders since the motion in a television scene is bursty. Several techniques of reducing the input to the buffer have been tried, e.g., subsampling, coarser quantization [142]. These techniques allow graceful degradation of picture quality when the source coder is overloaded. Some of these will be discussed in the next subsection.

C. Transform Coding

A straightforward application of rate-distortion theory to picture coding indicates that we assign a probability measure on the picture ensemble and then use a variable length code to give smaller length codes to those pictures that are more likely. However, the number of possible pictures is so large that it is impossible to assign a probability measure on it, much less construct a code book. One alternative is to divide a picture into subpictures (called blocks) of smaller size and then do "block-coding" using a code book. However, here again, subpictures of reasonable size (e.g., 4 elements X 4 lines) result in a large code book. So instead, we take the picture elements in the block and transform them into certain coordinates which can be treated independently and which are related to the probability measure on the original subpicture. Thus, in transform coding, we divide a picture into subpictures and then transform each of these subpictures into a set of "more independent" coefficients. The coefficients are then quantized and coded for transmission. At the receiver, the received bits are decoded into transform coefficients. An inverse transformation is applied to recover intensities of picture elements. These operations are shown in Fig. 41. Much of the compression is a result of dropping coefficients from the transmission that are small and coarsely quantizing the others as required by the picture quality. It is seen from this figure that important parameters that determine the performance of a transform coder are: size and shape of the subpictures, type of transformation used, selection of the coefficients to be transmitted and quantization of them, and the bit assignor which assigns a binary word for each of the quantizer outputs.

In this section, we describe both fixed and adaptive (data dependent) transform coding schemes and their performance. It should be noted that due to the inherent complexity of implementation, most of the transform coding literature describes computer simulations on still pictures, rather than real-time hardware implementations. Also, most of them use sequentially scanned (rather than interlaced) pictures. This poses problems in comparative judgment of picture quality since the coding degradations which do not change with respect to time (as in computer simulation of still pictures) are much less visible than the coding degradations which change with respect to time (as in real-time hardware).

Although transform coding is a natural outgrowth of the principles of rate-distortion theory, it was first applied to one-dimensional signals and then after many years, applied to the coding of pictures [143]-[148] by many workers independently. Since then, a significant effort has been devoted to it. Wintz [148] contains an excellent survey of transform coding work.

1) Transformations: The primary purpose of the transformation is to convert statistically dependent picture elements into "somewhat independent" coefficients. Most of the transformations that are used are linear and unitary. Transformation and separate coding of each subpicture neglects the redundancies that exist between the subpictures, and therefore, purely on a statistical basis, it is advantageous to have a large subpicture. However, for implementational simplicity as well as to exploit local changes in picture statistics and visual fidelity, a smaller subpicture is desirable. Of course, for practical reasons, an important consideration is that, since most of the compression results from dropping the coefficients with small energy, it is desirable to have a transform which compacts most of the image energy in a few coefficients as possible. Another consideration is the ease of performing the transformation itself.

a) Optimum transform: A truly optimum transform would result in the best picture quality using the least number of bits, but this criterion is difficult to specify quantitatively. A simpler criterion is to require that the transform coefficients be statistically independent, but this requires knowledge of

higher order (higher than 2) statistics of the images, which we do not yet have. So we seek a transformation that results in uncorrelated coefficients. Considering pels in a subpicture as an N-component vector X, we would like a transform "A" in the form of a $N \times N$ matrix that results in uncorrelated coefficient vector Y having N components, i.e.,

$$Y = AX \tag{13}$$

Since we are interested in dropping some of the coefficients, we may also want the error (mean-square) in reconstruction of X, due to dropping of some coefficients be the least. Also we would like image energy to be compacted into as few coefficients as possible. Fortunately, there is one transform that satisfies all these criteria, and it is known as the Hotelling transform [149] or Karhunen–Loeve transform[8] (KLT) [150]. Although KLT has been known for some time, its use for the problems of information transmission was made much later [151], [152]. The optimum transformation can be computed from the covariance of the pel vector X:

$$C_X = E\{(X - E(X)) \cdot (X - E(X))^T\} \tag{14}$$

where E is the statistical expectation and superscript T denotes transpose. Rows of the optimum matrix A are normalized eigenvectors of the matrix C_X, i.e., they are solutions of the equation

$$C_X X = \lambda_i X \tag{15}$$

The coefficients Y that are obtained by such transformation, have a covariance matrix given by

$$C_Y = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdots & \cdots & \cdots & \cdots & \lambda_N \end{bmatrix} \tag{16}$$

where $\lambda_1, \cdots, \lambda_N$ are the eigenvalues of the matrix C_X. C_X being a covariance matrix has all its eigenvalues nonnegative, and if we order them according to their magnitude, maximum energy is compacted in the first K coefficients ($K < N$) corresponding to the eigenvectors of the K largest eigenvalues. If only the K coefficients corresponding to the largest eigenvalues are sent, then at the receiver there will be some reconstruction error whose mean square is given by

$$\sum_{j=K+1}^{N} \lambda_j. \tag{17}$$

This is usually small, since only the smallest eigenvalues are included in the above summation.

An alternative way of expressing the above is to use a set of basis vectors $\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_N$ which are orthonormal and then represent the pel vector X as a linear sum of these basis vectors, i.e.,

$$X = \sum_{i=1}^{N} a_i \boldsymbol{\phi}_i. \tag{18}$$

It is well known that the optimum a_i which minimize the error between X and its representation (i.e., $\Sigma_{i=1}^{N} a_i \boldsymbol{\phi}_i$) are given by

$$a_i = X^T \boldsymbol{\phi}_i. \tag{19}$$

Now instead of transmitting picture element intensities (or components of X), we transmit projections of X on $\boldsymbol{\phi}_i$, i.e., $\{a_i\}$. In order to achieve the maximum degree of compression, we would like to send as few $\{a_i\}$ as possible, without a significant error in the reconstruction of X at the receiver. The basis vectors which compact maximum energy in the fewest $\{a_i\}$ turn out to be the eigenvectors of the covariance matrix C_X. Thus the KLT provides us with the optimum basis functions for representing a subpicture. In the above, we have converted a two-dimensional subpicture into a vector X. The above analysis can be easily extended to the case where the image is treated as a matrix and an optimum transformation is sought to convert it into a coefficient matrix. All the above results hold and a KLT can again be defined. However, no generality is lost by treating the subpicture intensities as vectors.

Although the optimum transformation is explicitly known, its use in practice presents many problems. First of all, the covariance function of an image is not stationary, and therefore, one must either choose a different covariance matrix matched to different regions of the picture or use an average (with a consequent loss in performance). The second problem is in the computation of the eigenvectors of C_X. In many cases [147] this matrix C_X turns out to be singular, and then some eigenvectors cannot be uniquely defined. Another problem is the difficulty of implementing the optimum transform. In general, it requires N^2 multiplications by constants which may be complicated (i.e., not simple powers of 2 for easy binary arithmetic). These problems have prevented any hardware implementation of the optimum transform. It is mainly studied in computer simulations to obtain a bound on the mean square error performance using any other transform.

b) Suboptimum transforms: Many other transforms have been invented which produce less correlated coefficients than the image itself and which are easier to implement. Some of the popular transforms are: discrete Fourier transformation [144]

$$A = \text{Matrix}\ \{a_{ij}\} \tag{20}$$

where

$$a_{ij} = \frac{1}{\sqrt{N}} \exp\left[-2\pi\sqrt{-1} \cdot (ij)\right].$$

Discrete cosine transform [153]

$$a_{ij} = \frac{2K(i)}{\sqrt{N}} \cos\left[(2j+1)\, i\pi/2N\right] \tag{21}$$

where

$$K(i) = \begin{cases} 1/\sqrt{2}, & \text{for } i = 1 \\ 1, & \text{for } i = 2, \cdots, N \\ 0, & \text{otherwise.} \end{cases}$$

Hadamard transform [134], [138].

Symmetric Hadamard transformation of order $N = 2^n$ is defined by

$$a_{ij} = \frac{1}{\sqrt{N}} \cdot (-1)^{b(i,j)} \tag{22}$$

[8] Strictly speaking, KLT was derived for the case of continuous waveforms rather than vectors obtained by sampling a continuous waveform. However, it is used widely for continuous as well as discrete cases.

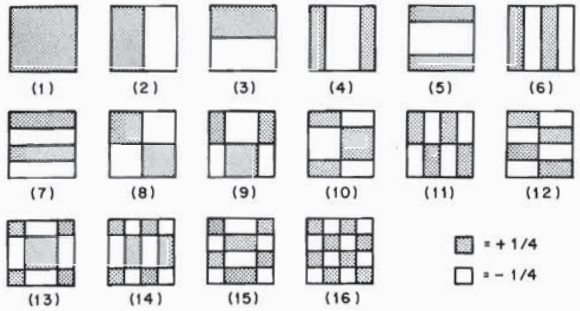


Fig. 42. Basis functions of a 4 X 4 Hadamard transform. Note that, in general, the sequency increases as the coefficient number (from Landau and Slepian [147]).

where

$$b(i, j) = \sum_{l=0}^{n-1} i_l j_l.$$

The terms i_l and j_l are the bit states of the binary representations of i and j, respectively. Thus a 4 X 4 Hadamard transform matrix can be written as:

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}. \tag{23}$$

Many other transforms have been used [154], [155]. Slant transforms, in particular, contain basis vectors which have a piecewise linear variation of their components. One obvious difference between these transforms and the KLT is that they are not dependent on image statistics. Another difference is in the ease of implementation. All the above transforms are unitary, i.e., $A^{T^*} = A^{-1}$ (where A^{T^*} is the complex conjugate of A^T), and therefore inverse transformation which is done at the receiver is as easy to implement as the transformation itself. The discrete Fourier transform can be implemented by using the fast Fourier transforms (FFT) techniques [156]. Instead of the N^2 multiplications required for the KLT, $2N^1 \log_2 N^1$ multiplications and additions are required. The discrete cosine transform can be implemented by a technique developed by Chen *et al.* [157] using $(3N/2 (\log_2 N - 1) + 2)$ real additions and $\{N \log_2 N - (3N/2) + 4\}$ real multiplications. The Hadamard transform is the simplest to implement since the Hadamard matrix consists of ±1's and, therefore, only additions and no multiplications are required.

Although not optimum, these transforms possess good energy compaction properties. In general, the bases that they represent correspond to some most likely subpictures as well as some very unlikely subpictures. The basis functions of a 4 X 4 Hadamard transform [147] are shown in Fig. 42. As is easy to see, the first basis function is a most likely subpicture, whereas the last basis function being a very chaotic subpicture is highly unlikely. The coefficients using these transforms, although not uncorrelated as in the case of the KLT, are "more independent" than the pels. Precisely how far from optimum are these transforms? That is a difficult question since it requires knowledge of the image statistics and judgment of picture quality using each of the individual transform codes. However, some answers can be given for simplified image models. If the image is assumed to be stationary, with separable correlation which is exponential in both the horizontal and vertical directions, then the mean-square error for encoding at a fixed bit rate varies with respect to the transform used and the block size as shown in Fig. 43. Although this model of the correlation is less typical for pictures, results of Fig. 43 are quite typical. It is interesting to observe that for small subpictures the mean-square error performance of different transforms is similar to that of the KLT. As the block size increases the Hadamard transform does not perform as well, but the Fourier transform gets better. In fact, it can be proved that the asymptotic performance of the Fourier transform is the same as that of the KLT. The discrete cosine transform, on the other hand, remains close to the optimum for all block sizes. Fig. 44 shows the basis functions of the discrete cosine transform and the KLT. Due to the similarity of the two basis functions, discrete cosine and KLT performances are very close to each other. It can be proved that for a first-order Markov process with exponential correlation, for most correlation coefficients [158], the best fast[9] transform is the discrete cosine transform. In fact, for many specific types of image models, the eigenvectors of the covariance matrix can be explicitly evaluated to construct the so called fast KLT [158]–[160]. It should be remembered, however, that the above remarks are strictly valid only for data from the stationary Markov processes with exponential correlation. For real images, these assumptions are not valid, particularly if the block size is small; and therefore, the simulations using these unitary transforms show results which in many cases, are quite inferior to those obtained by the KLT.

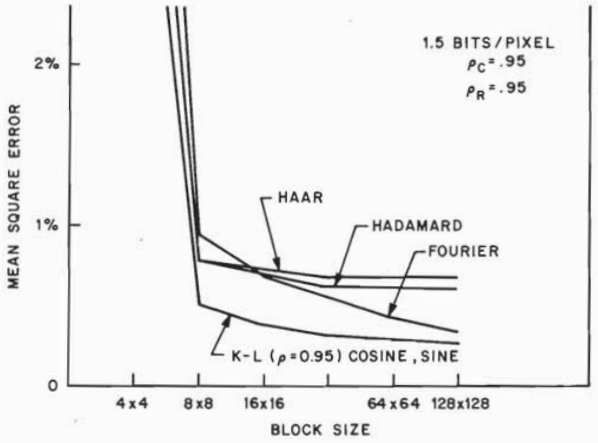


Fig. 43. Mean-square error performance of different transforms for two-dimensional Markov image source with horizontal correlation factor of ρ_C and vertical correlation factor of ρ_R.

Computer simulations on real pictures show that the mse produced by transform coding (for a given bit rate) improves with the size of the subpicture. However, the improvement is not significant as subpicture size is increased beyond 16 X 16. In fact the penalty in mean-square error by using a block of 8 X 8 is rather small compared to blocks of 16 X 16 and larger. Subjective quality of pictures, however, does not appear to improve with the size of the block beyond 4 X 4 [148]. Also, for real images, performances of KLT, Hadamard, and Fourier transforms are roughly the same for subpictures of 4 X 4; but

[9] One which can be computed by an algorithm similar to the FFT.

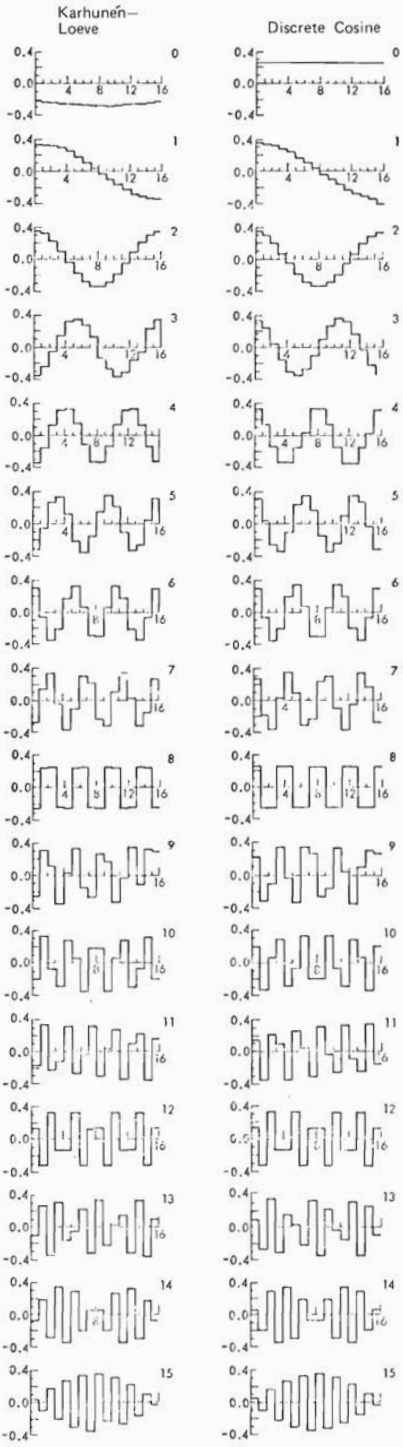


Fig. 44. Basis functions for KLT and discrete cosine transforms for a Markov image source. Note that except for scale factors, discrete cosine bases are very similar to the KLT bases.

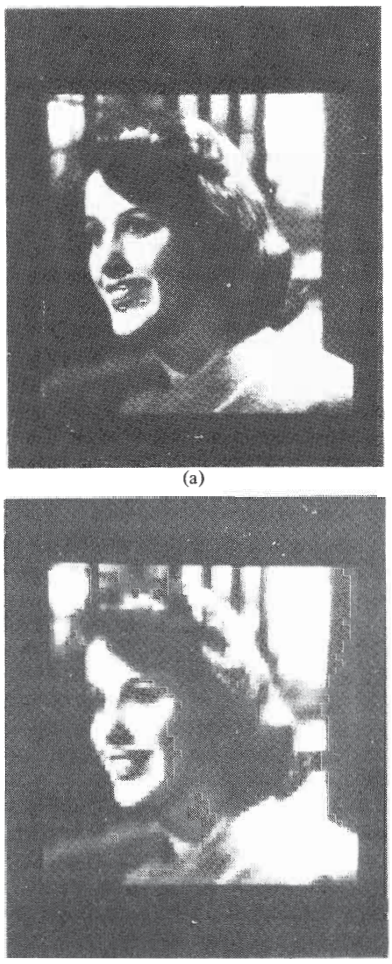
for 8 X 8 and 16 X 16 blocks, KLT is better than Fourier by about 0.1 or 0.2 bit/pel (comparing bit rate for a given mean-square error), which in turn is better than Hadamard by about 0.1 or 0.2 bit/pel. Also, a subpicture consisting of pels from many scan lines, (i.e., a two-dimensional block), does better than the subpicture consisting of elements from the same scanning line, (i.e., a one-dimensional block). However, the improvement is rather small [161], [162], about 0.2 bit/pel. This observation is analogous to the one for predictive encoders, where use of two-dimensional predictors does not improve the performance substantially over that obtained by one-dimensional predictors.

We have so far concerned ourselves with subpictures which are either one or two dimensional. It is possible to use three-dimensional subpictures, using the temporal dimension as the third dimension [163]–[165]. Block sizes of 4 X 4 X 4 (elements X line X frames) have been used in real-time implementation, and 16 X 16 X 16 blocks have been used in computer simulations [163]. As expected, the performance of the transform coders is considerably improved by using the three-dimensional blocks. However, this improvement is achieved at the expense of a significant increase in the amount of storage, which at this time appears impractical.

2) Quantization: The next step in transform coding is the selection and quantization of certain coefficients. In this subsection, we only discuss nonadaptive techniques, deferring adaptive techniques until later. One method of choosing the coefficients for transmission, is to evaluate the coefficient variances on a set of "average" pictures, and then discard all the coefficients whose variance is lower than a certain value. Such a scheme is called "zonal filtering" or "zonal subsampling." Obviously its performance will vary depending upon the type of picture being encoded. Coding degradations can be large if the picture contains large components of the type of basis functions that are discarded from transmission. At the receiver the discarded coefficients are usually set to zero. Using the Hadamard basis of Fig. 42 and a typical picture [147] (Fig. 45(a)), the measured variances of each of the coefficients are shown in Table III. Landau and Slepian [147] found that for a 4 X 4 transform by discarding coefficients 11, 12, . . . , 16 from the transmission, very little picture degradation is seen for most pictures. Most of the picture energy is contained in the first coefficient. Indeed, if all the coefficients except the first are dropped from the transmission, then the reconstructed picture is quite reasonable as shown in Fig. 45(b). The block structure and the loss of resolution is clearly seen in this picture.

Having decided which coefficients to transmit, we must design a quantizer for each of them. This could be done by dividing a given total number of bits among all the coefficients. In order to minimize the mean-square error for a given total number of bits for Gaussian variables, the optimum assignment of bits is done by making the average quantization error of each coefficient the same [152], [166]. This requires that bits be assigned to the coefficients in proportion to the logarithm of their variances. An improvement in the bit assignment technique is given by O'Neal and Natarajan [25].

Optimum quantizers for the transform coefficients have been designed mainly on a statistical basis. Approximations to the histograms of the coefficients have been made. The first coefficient, which is an average of the picture elements within a block, is usually modeled by a Rayleigh density, whereas all the other coefficients are modeled with Gaussian densities assuming a given variance. Since each coefficient is made of linear combinations of the picture elements, as the block size is increased by the central limit theorem, the histograms of the



(a) (b)

Fig. 45. (a) Original picture. (b) Reconstructed picture by using only the first coefficient in a 4 × 4 transform shown in Fig. 42 (from Landau and Slepian [147]).

coefficients are indeed roughly bell shaped. If each coefficient is quantized independently, optimum quantization can be achieved by minimizing the mean-square error as done by Max [113].

For better subjective picture quality, the quantizers should be designed to optimize the picture quality for a given bit rate. This problem is difficult since there are no accurate measures of picture quality. Landau and Slepian [147] found by trial and error that among all the quantization scales that they tried, a companding characteristic of the form $K\sqrt{x}$ was best subjectively. Mounts *et al.* [47] gave a systematic procedure for designing better subjectively matched quantizers for a Hadamard transform coder which used a $2 \times 2 \times 2$ block. Their technique is similar to the one used for predictive coding which was described in Section IV-B. It takes complex pictures as stimuli and obtains weighting functions which give visibility of a unit quantization noise as a function of the amplitude of the coefficient. Quantizers are designed to minimize the weighted quantization noise. They compared these optimum quantizers to the quantizers which minimize the mean-square quantization errors and found them to be superior [47].

TABLE III

COEFFICIENT NUMBER	COEFFICIENT VARIANCE
1	1.00
2	0.098
3	0.087
4	0.035
5	0.038
6	0.051
7	0.048
8	0.034
9	0.024
10	0.024
11	0.020
12	0.022
13	0.019
14	0.015
15	0.016
16	0.014

3) Adaptive Transform Coding: The parameters of a transform coder can be matched to the statistics of the subpicture being coded. Since the picture statistics may be highly nonstationary, adaptation can increase the coding efficiency significantly. Two types of adaptations are possible: one in which changes in parameters are based on the previously transmitted data, and the other in which some future data is used to compute the parameter changes. In the latter case, there is some overhead due to transmission of the adaptation information. The latter appear to be more suited for transforms with large blocks, since it would require less percent overhead information, and nonstationarity between the blocks would be more significant for larger blocks. Nearly all parameters of transform coders, such as dropping and quantizing the coefficients and bit assignment for the quantized coefficients, have been adapted to the local image statistics. Very little effort has been made, however, in adapting parameters based on psychovisual criteria, mainly because these criteria are not well understood, especially in the transform domain.

The transformation can be made adaptive by computing the covariance of each subpicture and the associated KLT. However, this is highly impractical since due to the problems of singularity, the KLT of each block is difficult to derive, and the overhead due to the transmission of basis vectors of KLT may be excessive. A more practical approach is taken by Tasto and Wintz [167], who classify a 6×6 block into one of three categories based on the block statistics. For each category, an average covariance matrix and the corresponding set of eigenvectors are used for transforming the picture elements of the subpicture. The overhead information for adaptation is only 2 bits per block, which is small. Tasto and Wintz report improvement in coding efficiency of as much as 30 to 50 percent over the nonadaptive case.

The procedure given by Tasto and Wintz is still perhaps too complex to be practical. Also, it is not clear how much of the advantage of adaptation is due to the adaptive transform and how much is due to the adaptive quantization. It appears that most of the advantages of adaptation can be derived by using a simple unitary transform followed by adaptive selection of the coefficients for transmission and adaptive quantization. This approach is more widely used. Threshold sampling [168],

[169], one of the popular techniques in this class, selects a threshold and transmits all the coefficients that are above this threshold. Thus the number, as well as the type of coefficients are adapted to the local structure of the subpicture. However, the overhead for adaptation information is rather large. Efforts have been made to decrease this overhead [169], [170].

Another type of adaptation is to construct a measure of spatial activity in the subpicture and then to adapt the coefficient selection. This would make use of the lower sensitivity to amplitude variations in regions of high spatial detail. Claire [171] and Gimlett [172] have proposed a definition of "activity index" using a weighted sum of the absolute values of the transform coefficients, and assigned more bits for coding those subpictures having a higher activity index. Netravali *et al.* [173] eliminated the adaptation overhead information by first selecting certain coefficients for transmission (by zonal subsampling) and then choosing a certain order for coefficient transmission. Having transmitted the first k coefficients for each block, quantization of the $(k+1)$th coefficient was adapted using a weighted sum of the first k coefficients. They demonstrated this technique on a $2 \times 2 \times 2$ Hadamard transform in which quantization of the first coefficient was made coarse if the weighted sum of the magnitudes of all the other coefficients was above a certain threshold. Quantization scales and the threshold for adaptation were obtained by subjective measurements to optimize the picture quality. They reported gains of only about 15 percent by such adaptation. Chen and Smith [174] describe a scheme which requires two passes through the image; one pass to collect the relevant statistics, and the other to code the image using the statistics. They compute the "ac" energy of the pels in each subpicture and then classify it into one of four classes. Thresholds for classification, as well as the bits assigned to the coding of each coefficient are selected specifically for each image. They report good picture quality at 1 bit/pel and satisfactory quality at 0.5 bit/pel on still pictures. Instead of using a weighted sum of the magnitudes of the coefficients as measures of activity, variances of the coefficients may be used to define spatial activity, which is then used for adaptive sampling and quantization of the coefficients. Tescher *et al.* [175], [176] estimate the variance of the transform coefficient and then assign bits in proportion to the logarithm of the estimated variance. Overall, this results in assigning more bits for transform blocks with large coefficients. Most of the methods described above require an excessive amount of computation and have not been implemented in real-time.

In spatiotemporal (three-dimensional) transform blocks, spatial as well as temporal activity can be used for adaptive selection and quantization of the coefficients. Knauer [163] describes an interesting technique using $4 \times 4 \times 4$ blocks in which certain coefficients are taken as measures of temporal activity. For large temporal changes, coarse sampling and quantization of the "spatial coefficients" is used, whereas for small temporal changes, spatial coefficients are transmitted with higher fidelity. Such techniques attempt to exploit the decrease in the sensitivity of the human eye to perceive spatial detail accompanied by large temporal changes. We close this section by pointing out that although many methods of adaptation of spatial transform coders have been attempted, they are either too difficult or ad hoc. However, we suspect that a truly optimum adaptive coder is not too far from some of the coders that have been simulated. Much more needs to be learned however in adapting spatiotemporal transform coders.

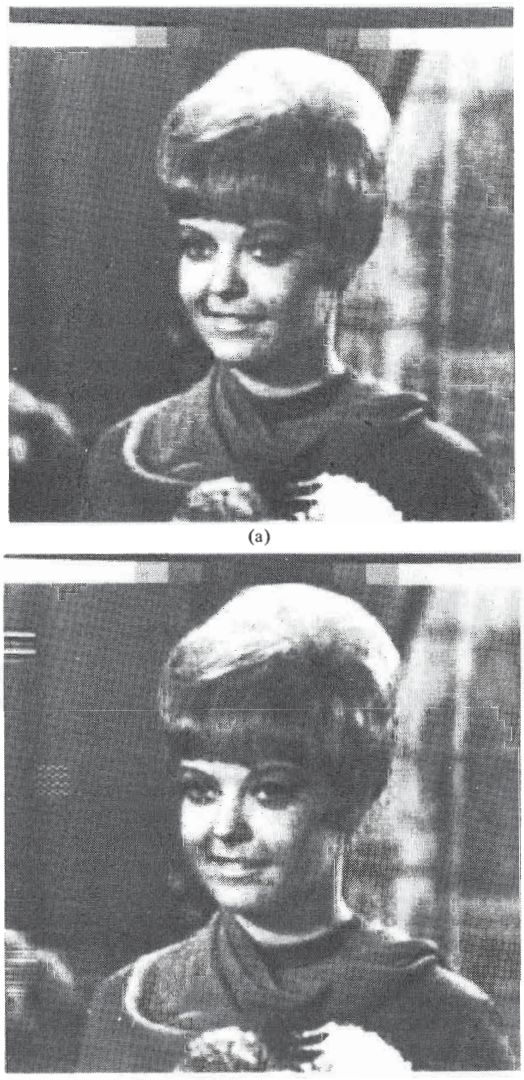


Fig. 46. Reconstructed pictures obtained after dividing the picture into 16×6 subpictures, coding with 1.5 bit/pel, and making random bit errors at a rate of 10^{-4}. Both zonal (a) and threshold (b) coding are considered (from Pratt [210]).

Here due to the extra degree of freedom, greater payoff may be possible.

4) Effect of Channel Errors: One of the advantages of the nonadaptive transform coders is that the effect of bit errors in the channel does not spread beyond the block. If a coefficient is decoded erroneously at the receiver due to a transmission error, on inverse transformation only pels in the same block are affected. The exact nature of the degradation depends upon the type of transform used and the coefficients in error. In general, errors in lower frequency (or sequency) coefficients are more visible than higher frequency coefficients due to the fact that above a certain frequency the sensitivity of the eye decreases with frequency. Also, probability of errors taking place in the lower frequency coefficients is higher due to larger number of bits that are required to code them. As the block size decreases the averaging property (i.e., spreading of error

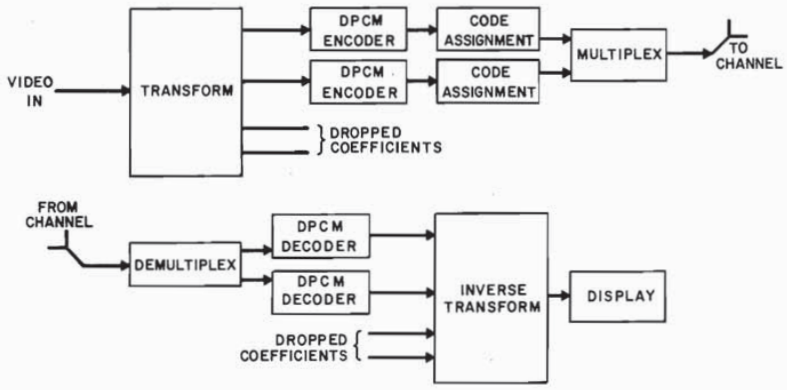


Fig. 47. Block diagram of a hybrid transform/DPCM encoder.

over the entire block) decreases, and the transmission errors appear as blotches in the picture as shown in Fig. 46. It appears that the visibility of the transmission errors in predictive coding with two-dimensional predictors is comparable to that for the nonadaptive transform coders, (compare Figs. 46 and 30). Error concealment techniques have been developed in a way analogous to those for DPCM coding systems [177]. Adaptive transform coding may result in the spread of the effects of transmission errors, unless one is careful in not using information from one block to the next either for adaptation or for decoding at the receiver. In general if each subpicture is allowed to generate a fixed number of bits and adaptation is done using information within the subpicture, then the effects of transmission errors can be limited to subpictures, but this might impose a severe penalty on the coding efficiency especially for small block sizes. Nevertheless, adaptive coders have been described which are not overly sensitive to the channel errors [174].

D. Hybrid Coding

We have pointed out before that transform coding systems are inherently more complex in terms of both the storage of data and the number of operations per pel. Although the use of large block sizes removes statistical redundancy quite effectively, it has two distinct disadvantages: 1) it requires storage of large amounts of data both at the transmitter and the receiver, and consequently produces a delay in transmission, and 2) the accuracy with which different regions of the image need to be coded may vary widely within the block, and this makes adaptive coding (e.g., quantization) more difficult to accomplish. Hybrid coding is a partial answer to this problem. In hybrid coding, we consider small blocks, evaluate the coefficients, and perform DPCM of the coefficients using coefficients of the previously transmitted blocks as predictors. A block diagram of a hybrid coder is shown in Fig. 47. Specifically, three types of schemes have been studied: 1) a one-dimensional block along the scan line with DPCM in the vertical direction; 2) a small two-dimensional block, and DPCM using coefficients of horizontally previous block for prediction; 3) a two-dimensional block, and DPCM in the temporal direction. The first two schemes are intrafield methods, whereas the third scheme is interframe. The saving in storage due to hybrid coding is considerable in the third scheme compared to a transform coding scheme which uses a three-dimensional block. All the three variations have received attention in the literature [165], [173], [178]–[184]. Habibi showed that the theoretical performance of hybrid coders (of the first and second type) was quite superior. He made some approximations to the statistics of the difference signals (i.e., the coefficient minus its prediction) and calculated S/N ratio as a function of bit rate and block size for different transforms. He also showed with computer simulations that the performance of the hybrid coders in the presence of transmission errors was reasonable, although not as good as that of the nonadaptive transform coders. It is certain that the hybrid coders may require some type of protection against transmission errors as required in predictive coders (e.g., leak in prediction, etc.). Netravali *et al.* [173] used a small two-dimensional block and showed that if the optimum transform (i.e., KLT) is not used, there is correlation between coefficients of the same block; and therefore, a better predictor can be designed by using not just the corresponding coefficient of the previous block, but by using coefficients of the previous block, as well as, those of the present block which are available to the receiver. They showed that such a predictor was 25 percent more efficient in terms of bits/pel for the same picture quality, than a predictor which used the corresponding coefficient of the previous block. Experimental and theoretical (assuming simple Markov source approximation for images) performances have been obtained by simulation by Roese *et al.* [165] for transform coding using three-dimensional blocks, and hybrid coding using two-dimensional blocks and DPCM in temporal direction. They show that the hybrid coder is quite efficient and does as well as a three-dimensional transform coder which uses four frames of storage. They also claim that the hybrid coder does not show ill effects at bit-error rates of 10^{-4}. Just as in pel domain, predictor in interframe hybrid coding can be adapted by estimating motion and using a coefficient from displaced block in the previous frame for prediction [185], [186]. However, motion compensation in transform domain does not appear to be as efficient as in pel domain.

E. Interpolative Coding

In interpolative coding, a subset of picture elements are transmitted and the remaining pels are interpolated. As an example, Fig. 48 shows a case in which there is a 2:1 subsampling of picture elements along each scan line. The subsampled elements are staggered from one line to the next and

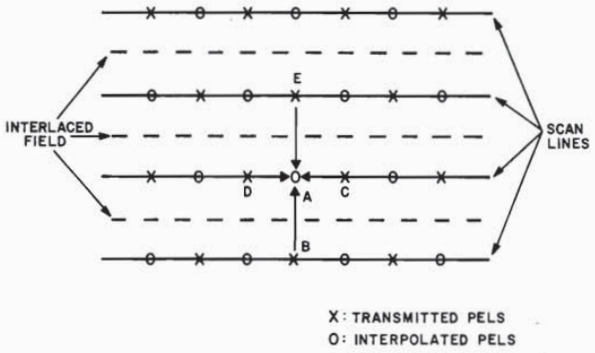


Fig. 48. An example of interpolative coding using 2:1 horizontal subsampling, which is staggered from one line to the next.

are interpolated by a four-way average as shown by arrows (e.g., element A is interpolated by averaging elements B, C, D, E).

Interpolative coding techniques have been studied extensively in the past for digital coding of pictures [187], [188]. In general, they can be divided into two classes: fixed, and adaptive. In fixed interpolative methods, a fixed set of picture elements are selected for transmission and the rest are interpolated. Various types of samples can be chosen for transmission: as examples, one may transmit every alternate sample, one sample out of four, every alternate scan line, every alternate field or frame. The picture quality that results is a function of the number and type of samples that are dropped and the method of interpolation. Most interpolation methods have used weighted averages either using straight lines or higher degree polynomials [189]–[191]. It appears that interpolation using straight lines is quite effective and not much is gained by interpolation using polynomials of higher degree. Switched interpolation may be more effective than the fixed interpolation. As an example, in Fig. 48, pel A may be reconstructed by the following rule:

$$\hat{A} = \begin{cases} 0.5\,(C+D), & \text{if } |C-D| \leqslant |B-E| \\ 0.5\,(B+E), & \text{otherwise.} \end{cases} \tag{24}$$

Such interpolations switch between horizontal, vertical, temporal or other directional averages depending upon the type of local correlation existing in the data as measured by the transmitted pels. Another method of interpolation is due to Gabor and Hill [192], which, for example, drops alternate fields and attempts to construct them by making movement of edges temporarily continuous, i.e., placing the edges in the dropped field at places dictated by their uniform motion between the adjacent transmitted field. Such a technique may be useful, but in practice it has been found to be rather difficult to implement, since it requires definition of edges and their motion. Several techniques have been used to transmit the pels from which other pels are derived. Predictive coding appears to be more common [193] due to its flexibility and simplicity.

Adaptive interpolative coding consists of three parts: 1) choosing certain points for transmission, 2) constructing the interpolation of nontransmitted pels, and 3) evaluating the interpolation error; if the error is below a certain threshold, then choosing fewer points for transmission; if it goes above threshold, then choosing a larger number of points for transmission. The fewest number of points that are needed to keep the interpolation error below the threshold is normally the desired characteristic. As an example, consider the video waveform in a scan line as shown in Fig. 49. Starting with sample A, one can go up to sample E and interpolate, using a straight line, all the in-between samples B, C, and D, keeping the interpolation error below the threshold. However, when samples A and F are chosen for transmission, then a simple straight line interpolation between samples A and F may result in excessive interpolation error for pels B, C, D, and E. Thus pels A and E should be chosen for transmission and pels B, C, and D should be interpolated. Several variations exist depending on: 1) method used for transmission of pels such as A and E; 2) technique of interpolation and 3) method of judging the quality of interpolation. It is also possible to treat this as a problem in waveform approximation in which a value at certain points (called knots) is transmitted and an approximation to the entire curve is made from the knot position and the transmitted value which may not be image intensity at the knot position. Criteria that are used for the decision to interpolate have traditionally been based on mean-square error [190]; but recently more sophisticated criteria, which may be closely related to human visual perception, have been used [193], [194]. Also adaptive two-dimensional interpolation has been used, which is helpful in decreasing the number of pels transmitted [189]. It should be remembered that, unlike in fixed interpolative coding, adaptive interpolative coding requires two kinds of information: 1) addresses of picture elements chosen for transmission[10] and 2) intensity values of picture elements chosen for transmission.

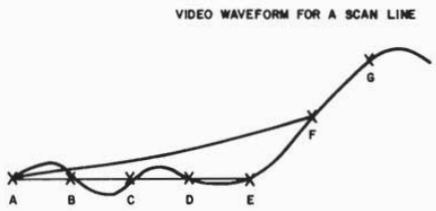


Fig. 49. An illustration of adaptive interpolation technique.

Performance of the interpolative coders depends upon the technique used for transmitting samples and for interpolating. It appears that in terms of the number of bits required for a given quality, interpolative coders show an improvement over nonadaptive DPCM and transform coding; but they appear to be inferior to both adaptive DPCM as well as transform techniques. However, they might be useful for applications where low bit rate is required, and where it may not be possible to encode each sample.

F. Contour Coding

We close this section by mentioning contour coding and its variations. In contour coding, a picture is separated into two parts: edges and all the rest. Such a separation is hypothesized to do well, since by isolating edges and reproducing them well, the pictures look sharp. Spatial continuity of edge points permits the use of algorithms developed for two-level (facsimile) data. Early schemes [196], [197] separated the picture into

[10]See [195] for a technique which does not require explicit address transmission.

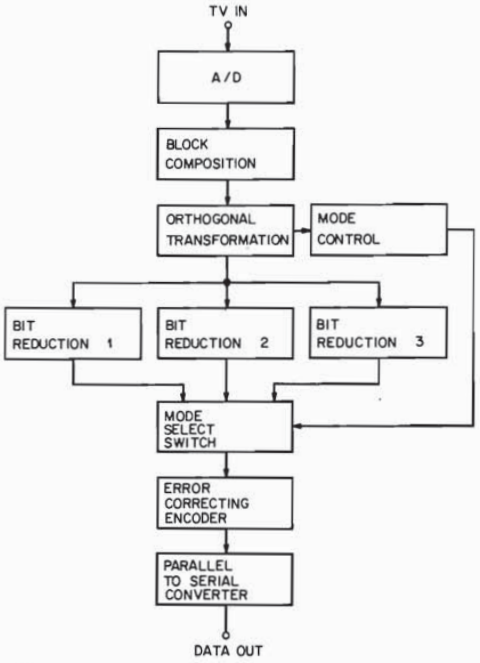


Fig. 50. Block diagram of adaptive orthogonal transform coding system (from Ohira, Hayakawa, and Matsumoto [199]).

TABLE IV

V \ H	1	2	3	4	5	6	7	8
1	8	6	5	4	0	0	0	0
2	6	4	4	3	3	2	0	0
3	6	4	4	3	0	2	0	0
4	6	4	0	3	4	6	6	4

MODE 1

V \ H	1	2	3	4	5	6	7	8
1	8	6	5	5	4	0	0	0
2	6	4	0	0	0	2	2	0
3	6	4	0	0	0	4	4	0
4	6	4	0	5	5	6	6	5

MODE 2

V \ H	1	2	3	4	5	6	7	8
1	8	6	5	5	5	6	6	5
2	6	4	3	2	2	2	0	0
3	6	4	0	0	0	0	0	0
4	6	4	0	0	3	4	5	0

MODE 3

"highs" and "lows," then used a contour-tracing type algorithm for coding "highs." The "lows" part of the signal was subsampled and coded with higher amplitude accuracy. Yan and Sakrison [35] used a similar approach, in which the "lows" were coded by transform coding. Since the "lows" did not contain much high frequency information, a large number of coefficients could be dropped. They coded still pictures using this technique and claimed good quality pictures at 0.83 bit/pel.

V. Description of Three Systems

The previous section describes many different coding techniques in isolation. In this section we consider three specific coders in some detail in order to show how the different techniques may be effectively combined. Each coder has been pursued to the point that the performance in terms of picture quality and transmission error response is known. These coders are designed to operate at transmission rates consistent with the digital PCM hierarchy. In North America and Japan the rates of interest are 1.5, 6.3, and 44 Mb/s.

A. Intraframe Transform Coder

Although there have been many computer simulations of transform coding algorithms, there has been relatively little real-time exploration. The coder of Ohira *et al.* has been fully implemented and is evolving in sophistication [198], [199]. The block diagram of the encoder is shown in Fig. 50; the decoder performs essentially the reverse operations.

The coder is designed to accept a composite baseband NTSC signal which is then encoded in composite form. The A/D converter samples at 9.7 MHz and gives 8 bits/sample. The signal is then formed into two-dimensional blocks 8 elements by 4 lines (from the same field) in such a way that adjacent blocks across the picture are offset vertically by 1 line. It is assumed that the offset simplifies the hardware design by enabling pipelining of operations and by reducing buffering requirements. In addition block-to-block distortions may be less visible with the offset structure. A 32nd-order transform is performed by first calculating an 8th-order slant transform [200] in the horizontal direction and expanding this to a 32nd-order by two applications of the well-known relation [148],

$$[H_{2n}] = \frac{1}{2}\begin{bmatrix} H_n & H_n \\ H_n & -H_n \end{bmatrix} \tag{25}$$

where H_n is a Hadamard matrix of order n.

The algorithm is adaptive in that one of three different quantization bit assignments is chosen depending upon the values of the raw transform coefficients. The three bit assignments are shown in Table IV and in each case 97 bits are allocated per block with an additional 2 bits used to specify the mode. As can be seen from the table, the low-sequency bit assignment is the same for each mode assuring high quality transmission of these components. Mode 1 is designed to handle regions of a contrasty nature with medium-frequency detail. Mode 2 is designed for regions having saturated colors and hence those coefficients that represent frequencies around the color subcarrier are emphasized. The third mode is designed to handle high-frequency vertical detail. In order to select the appropriate mode the group of coefficients important to each mode is examined. Any large amplitude component in the group will favor the selection of the corresponding mode. However, mode 1 is given priority over mode 2 and mode 2 is given priority over mode 3. Both the selection criteria and the bit allocations were determined experimentally.

The first coefficient (dc term) is linearly quantized and the remaining coefficients are nonlinearly quantized to take ad-

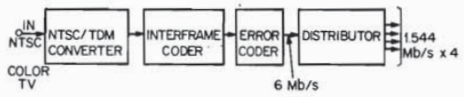


Fig. 51. Block diagram of NETEC-6 interframe coder (from Iinuma *et al.* [203]).

vantage of the highly peaked nature of the probability density functions of the coefficients.

Transmission errors occurring on certain bits of the coded signal can produce very visible error patches. The more significant bits of the lower-sequency coefficients are most sensitive. In order to reduce the visibility of the degradation due to transmission errors, the eleven most sensitive bits are protected by a (15, 11) Hamming code. The result is that errors do not start to become objectionable until the error rate is as high as 10^{-3}. The spatial appearance of the errors is shown in photographs of the television screen, however, no description of the temporal appearance of the errors is given.

Evidently, at a transmission rate of 32 Mb/s a certain amount of degradation is noticeable in still areas. In the absence of adaptation (i.e., using just a single optimized mode) the quality has been compared with a 6-bits PCM signal [201] which would be inadequate for broadcast quality television. Adaptation improved the quality of unusual pictures such as those containing saturated colors or high-frequency vertical components but did not change the quality significantly for most material [199].

B. Conditional Frame-to-Frame Coder

A conditional frame-to-frame coder divides each frame into a changed part and an unchanged part. Only the changed part is coded and transmitted [202]. The NETEC-6 is an example of this technique, operating at 6 Mb/s [203].

The coder consists of 3 basic blocks, as shown in Fig. 51. The NTSC/TDM converter takes the composite color signal and demodulates it into a luminance signal Y and two chrominance signals R-Y and B-Y. One chrominance signal per line is time compressed by a factor of 5 and inserted into the blanking interval of the luminance signal. The R-Y and B-Y signals are inserted on alternate luminance lines. The result is that each line contains 420 luminance samples, 84 chrominance samples and 6 samples are used for synchronization. The resulting composite TDM signal is then processed in the frame-to-frame coder as a single signal.

A block diagram of the frame-to-frame coder is shown in Fig. 52 and is essentially a frame-to-frame differential quantizer. In stationary parts of the picture, frame-to-frame differences will be extremely small and the representation of the picture stored in the frame memory need not be updated. When significant frame differences do occur, they tend to be clustered, as would be expected from differences generated by moving objects. The significance determination circuit is used to determine whether or not the frame difference signal at a pel is significant. If the difference is greater than a threshold and it is not an isolated change (it is an isolated change if none of its neighbors are significant changes) then the frame-to-frame difference is quantized in the 64-level quantizer and transmitted to the receiver. A ±27-level nonuniform quantizer is used with the step size varying between 1 and 5/256th. This may seem to be a large number of levels relative to the number required to provide high-quality intraframe encoding. How-

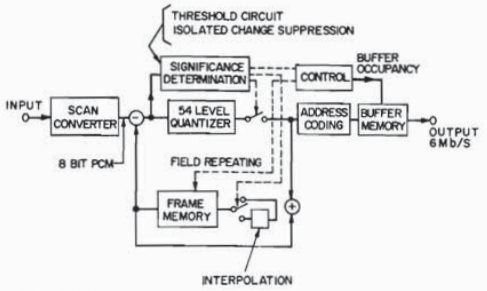


Fig. 52. Block diagram of conditional replenishment interframe coder (from Iinuma *et al.* [203]).

ever, fast-moving high-contrast edges can produce large difference signals and if these are not quantized reasonably accurately moving edges will appear jagged. A simple two-length variable-length code is used with the inner levels having a length of 4 bits and the rarely used outer levels having 6 b. An address is attached to indicate to the receiver where the change should be inserted. It would require approximately 9 bits to indicate the horizontal address of each changed pel and means are needed to reduce this figure. Since changes are highly clustered one method is to give a start-of-cluster address and a word to indicate the end of a cluster [202]. In the NETEC-6 coder addressing is achieved by first dividing the signal into 8 element blocks, 4 elements in width and 2 lines high. A word is then used to indicate whether the block has significant changes occurring within it; if so, a bit per pel is used to indicate those elements within the block that have changed significantly.

A frame-to-frame coder generates data at a very uneven rate and it is necessary to smooth the data to a constant rate for transmission over the channel. Even with a large buffer, the coder may generate information at a short-term average rate that is greater than the channel rate and there will be a tendency for the buffer to overflow [141], [204], [205]. Feedback from the buffer is used in a number of ways to reduce the data generation rate. They are:

control of threshold level in significant-change determination circuit;
control of threshold level in adaptive interpolation;
field repeating;
suspension of replenishment.

Perhaps the most effective means of reducing the data rate is to use a more severe criterion in the definition of a significant change. The threshold is raised from a minimum value of 2/256th to a maximum value of 5/256th. The high thresholds result in a small amount of degradation in the moving areas that appears like viewing a picture through a dirty window. Adaptive interpolation [130], [193], [194], [203] is a means of reducing the spatial resolution gradually and in so doing, reducing the number of elements to be coded in the moving area. In this instance the amplitude of a significant frame difference is compared with the average of the frame difference for the adjacent two elements. If they differ by less than an adaptive threshold then the frame difference is not transmitted and at the receiver the element value is replaced by the average of the values of the adjacent two elements. If, on the other hand, the frame differences differ by more than the threshold then the frame difference signal is transmitted normally. As the

threshold increases, fewer and fewer frame differences are transmitted for the alternate elements, until finally when the threshold is very large the number of frame differences to be transmitted in the moving area is halved.

Field repeating further reduces the spatial resolution since alternate fields are derived from the previous field by using an average of the two adjacent lines. Field repeating also causes a slight jerkiness of rapidly moving objects in the scene. Should the data rate continue to rise in spite of these measures, rather than risk buffer overload with attendant loss of synchronization between coder and decoder, replenishment is halted and the picture freezes until the buffer empties sufficiently for updating to continue. This is a last resort control and is usually invoked only under conditions of extreme movement.

In an encoder of this type a transmission error can have a large influence on the decoded signal, particularly if an address word is affected. Further, once an error is made, it tends to remain from frame-to-frame. To prevent this a refresh signal is sent in which a complete line is PCM coded with 8-bits accuracy and transmitted at the rate of two lines per frame. Thus an error would be visible, on average, for about 4 seconds before being overwritten by the refresh signal. To reduce the effect of transmission errors, a rate (11, 12) convolutional code is used; this code will reduce a random error rate of 10^{-5} to 10^{-8}.

There is little standardization on procedures for measuring performance of frame-to-frame coders and agreement on standard test material would be a very desirable first step. Like nearly all frame-to-frame coders, very high quality performance is achieved by this coder when there is little movement and quality decreases as the percentage of changed area in the picture increases. High-quality pictures are obtained up to the point where field repeating is introduced; approximately 10 percent of the picture can change before this control mechanism is applied. Since this coder is designed primarily for conferencing applications where the camera is usually stationary and there is rarely large movements by the conferees, a high quality picture is obtained for more than 98 percent of the time.

Another example of a frame-to-frame DPCM coder that has been implemented is the NETEC 22H coder [206]. Rather than decomposing the color signal into Y, I, and Q components a luminance-like and a chrominance-like signal is generated by adding and subtracting alternate scan lines of the composite color signal. The basic encoding algorithm consists of intraframe DPCM encoding of the frame difference signal. The intraframe prediction algorithm changes depending on the state of the buffer and the amount of frame difference. The coder uses many of the techniques already described for the NETEC-6 coder. An additional method for reducing bit-rate is to nonlinearly process the frame difference signal prior to quantizing so that small values only are attenuated. This is equivalent to temporal low-pass filtering of the signal. The level at which the attenuation is switched out is increased as the buffer fills, producing more visible temporal smearing.

C. Interpolative Frame-to-Frame Coder

The third coder to be described is a frame-to-frame coder that uses a combination of DPCM and conditional interpolation [207]. Although the coder is intended for a color broadcast television signal, coding of the luminance signal only has been studied. The coder has been designed to operate at

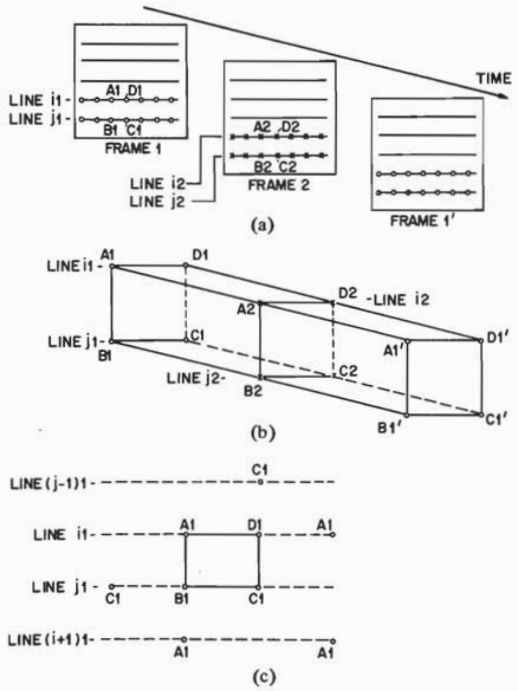


Fig. 53. Arrangement of pels that make up a 3-dimensional 2 X 2 X 2 block for interpolative coder (from Limb and Bowen [207]).

10 Mb/s for the luminance signal. Allowing an additional 5 Mb/s for the chrominance signals would give a total bit rate of 15 Mb/s, permitting three signals to be transmitted over the third level of the North American digital hierarchy (45 Mb/s). No hardware has been built; however, the study was performed on a special simulation facility which allowed the viewing of coded sequences at 30 frames/s using pictures of approximately half the vertical and horizontal resolution of normal television. Thus real-time evaluation of the coded picture quality could be made and the effects of transmission errors could be observed. It should be noted that the visibility of coding distortions varies slightly between the lower resolution picture of the simulator and the broadcast television picture.[11]

One of the basic considerations in designing the coding algorithm was to limit the tendency for error propagation, which is inherent in any coder using variable length codes and/or DPCM coding techniques. Therefore, the basic coding scheme was limited to groups of four lines, consisting of pairs of consecutive lines in consecutive frames as shown in Fig. 53(a). The group of four lines is subdivided into blocks of eight pels, as shown in Fig. 53(b). The key elements in a block, on which all of the other elements depend are the $A1$'s. These elements are DPCM coded along a line using the previous $A1$ as a predictor. The coding of the remaining elements (B's, C's, D's) uses the four surrounding $A1$'s, two of which are in the next pair of lines which would not normally be coded with the current block (Fig. 53(c)). For this reason, the DPCM coding of the $A1$'s proceeds one line ahead of the coding of the remaining elements in the block. Since the $A1$ elements are critical to the coding (in a

[11] From comparison between the two it is known that relative to the broadcast television picture the picture of the simulator is more sensitive to spatial types of distortion but less sensitive to temporal types of distortion.

DATA GENERATED BY 4 LINE GROUP, ARRANGED FOR TRANSMISSION

WORD FOR BIT COUNT PLUS PROTECTION 45 BITS	A1 DATA PLUS BIT FOR INTERPOLATE/CODE OF FRM2 AVG = 524 BITS	B1, C1, D1 DATA	PROTECTION FOR 1000 BITS 24 BITS	B1, C1, D1 DATA	A2, B2, C2, D2 DATA

MIN BITS = 568, MAX BITS = 12738, AVG BITS = 1590

Fig. 54. Data generated by 4 line group, arranged for transmission.

visual as well as informational sense), they are quite finely coded (32 levels).

Once the $A1$ elements have been coded, the other three elements, $B1$, $C1$, and $D1$ are processed starting with $C1$. An average is found of the surrounding $A1$'s and compared to $C1$. If the resulting error is less than a threshold, the average value is used, otherwise the error value is transmitted using a coarser DPCM coder (21 levels). When quantizing the error, quantizer levels less than the threshold will not be used. They are removed from the scale and the level containing the threshold becomes the first level and all other higher levels are adjusted downward accordingly [130]. As explained later, the threshold is determined from the coded $A1$ elements and thus need not be explicitedly transmitted. The third and fourth points to be processed are $B1$ and $D1$ for which a switched predictor is used. An estimated value for $B1$ is found from the average of either its horizontal ($C1$'s) or vertical ($A1$'s) neighbors, whichever pair has the least difference. Note that these neighboring points have been encoded at this point. If the interpolation error is less than a second threshold, the interpolated value is used, otherwise the error value is transmitted using the coarse DPCM quantizer as described for element $C1$. The method used for $D1$ is identical to that used for $B1$.

The corresponding part of the block in frame 2 ($A2$, $B2$, $C2$, $D2$) is then processed. A first test is made to see if it is permissible to interpolate all four elements from frame 1 and frame 1′. If the error of an individual pel, or if the average error of all four pels, is greater than a third threshold, then the block must be coded. First, the $A2$ pel is coded by transmitting a quantized correction to the value of $A1$ (fine quantizer). ($A1$ is more precisely known than the previous $A2$ pel from the same frame.) The other three pels ($B2$, $C2$, $D2$) are coded in precisely the same manner as the corresponding three pels in frame 1. If all four points in frame 2 can be interpolated, no additional data is required other than a single bit to designate coding or interpolation.

High detail areas have a masking effect so that a viewer is more tolerant of coding errors in these areas (Section III-D2). Therefore, the intraframe and interframe thresholds are made dependent upon an activity function derived from the amount of activity in the area being coded and designed to reflect the amount of local masking. In order to prevent buffer overflow during periods of peak activity the threshold range is made a function of the buffer state. When the buffer fills thresholds are raised and more picture elements are interpolated which in turn results in the use of shorter code words.

The data is not transmitted as generated but in the order shown in Fig. 54. This enables the $A1$ picture elements, which are by far the most sensitive to error, to be protected. The first 1000 bits in a coded line (which with high probability includes all the $A1$ data) is protected with a BCH (1002, 1023) distance 6 code. One error would be corrected and virtually all other errors occurring in the preceding 1000 bits would be detected at error rates of 10^{-4} or less.

When an error in the $A1$ pels is detected, the whole line is replaced by repeating the previous good line. Most errors occurring in the B, C, and D data are detected because the length of the decoded line will not match the length specified by the line synchronization word. In this case, the B, C, D information is set to zero, and the subjective effect is a very small loss in picture sharpness on two lines. If the error in the *BCD* data is not detected the two lines will appear a little noisy for a portion of the line.

With the threshold adjusted to give a high-quality picture in open loop operation (that is, no feedback from the buffer), the scene with normal movement resulted in a bit rate of 0.96 bits per pel. If the second frame were coded in the same manner as the first frame (i.e., no frame to frame coding) the bit-rate will have increased by 45 percent. With larger amounts of moving area the advantage of frame-to-frame coding would have been even less.

When feedback from the buffer is used to maintain the output bit rate at 1.5 bit/pel average sequences with an average amount of movement are very easily handled and an excellent picture quality results. Noisy sequences are also handled very well; careful comparison between the coded and uncoded sequences shows a small increase in the noise level. When highly detailed scenes are panned slowly, there are small increases in noise level resulting in a picture quality which is probably marginal for broadcast television purposes. With fast panning there is a very noticeable increase in noise level, and the picture quality is definitely unacceptable. In an attempt to handle the material containing fast pans, the basic algorithm was modified to incorporate conditional field interpolation [208]. In conditional field interpolation, the even frame is interpolated from the adjacent two lines in both the preceding and succeeding field. Only when this interpolated value differs by more than a threshold quantity from the value of the uncoded sample is additional information transmitted to improve the interpolation. The quality of the fast pan scene increased greatly as a result of using field interpolation to the point where this extreme condition is handled with only a small increase in noise level.

At transmission error rates of 10^{-6}, very few undetected errors occur in the picture. At an error rate of 10^{-4}, approximately five errors are injected per frame, and approximately one quarter of these are uncorrected. Because of line substitution, the errors do not result in any streaking, but produce a small increase in noise level in highly detailed areas of moving objects. The overall picture quality remains high, and the picture would still be usable at even lower error rates.

VI. Discussion and Conclusions

The field of picture coding is now a rather mature one, stretching back approximately 25 years. With a small stretch of the imagination one can divide this period into three broad segments. Until about 1968, basic systems were constructed in a reasonably straightforward manner. Little attention was paid to such practical matters as error performance and buffer overflow. During the next six years the basic designs were embellished greatly and optimization of the various components of the systems was attempted. For example, multidimensional optimum predictors were explored, quantizing characteristics were optimized under various assumptions, and the behavior of various orthogonal transformers was analyzed. During the last three or four years attention has turned to making coding algorithms adaptive both to the local statistics of the picture

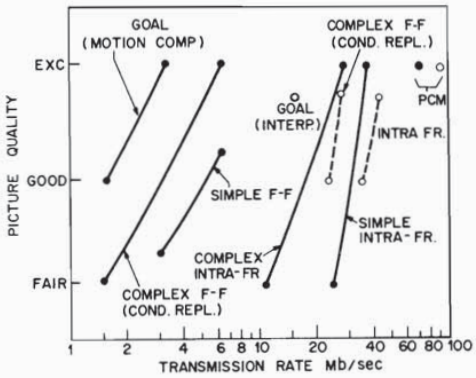


Fig. 55. An attempt to indicate the relative performance of coding techniques in terms of picture quality versus transmission rate. Crosses and dashed lines indicate results for broadcast television signals. Dots and full lines indicate results for video conferencing. Abbreviations F-F, frame-to-frame; Cond. Repl., conditional replenishment; Intra-Fr., intra-frame; Interp., interpolation; Motion Comp., motion compensation.

signal and to the local properties of the visual system. This has resulted in a reexamination of the models that have been used, implicitly or explicitly, to describe the picture signal source and the human viewer. We will continue to see refinement of source and receiver models and, looking beyond this, the incorporation of the knowledge of these models into encoder designs.

A new phase appears to be emerging in which more sophisticated parameters are being extracted from pictures. Rather than just calculating activity functions or measuring global speed, one could divide the picture into distinct objects or sections of objects, identify textured regions and categories, and accurately track connected regions. Perhaps 3-*D* representations of objects within a scene could be stored so that camera or object movements can be transmitted with just a few parameters. Already computer simulation of a scheme in which a local measure of velocity is calculated in order to reduce prediction error is yielding reductions in bit rate of a third to a half that required for conditional replenishment schemes [102]. To achieve this level of sophistication in real time (approximately 100 ns/pel), high speed calculations and large amounts of storage are required. However, with the prospect of 32-bit processors on a chip (albeit a large one) and with 64-kb RAM's already beginning to appear, image segmentation, identification and manipulation in real time may soon become economically possible. (It is primarily because of the speed requirements that microprocessors have had so little impact to date on the implementation of real-time coders.)

In Fig. 55 we attempt to show the performance that coders currently achieve in terms of picture quality as a function of bit rate. Also shown are expectations for the future, based on the results of recent studies. The data of Fig. 55 are the personal interpretation of the authors and no doubt show their bias. The measure of picture quality is purposely vague as is the use of the words "simple" and "complex." A complex coder would employ more sophisticated algorithms and use more adaptation; the required hardware would be more extensive.

The figure is divided into two parts, broadcast television, represented by crosses and dashed lines, and video conferencing television, represented by dots and full lines. This distinction is made necessary by the fact that the quality requirements for the two situations are rather different. A picture judged excellent for video conferencing purposes may be regarded as only mediocre or even poor for broadcast television. For broadcast television, PCM encoding requires approximately 90 Mb/s for NTSC standards. This can be reduced to the range 30 to 50 Mb/s by intraframe coding techniques depending upon the quality and coding method. Coding of color components yields results that are perhaps 10 to 20 percent lower than coding of the composite signal; based on systems so far constructed, DPCM methods give a small improvement over transform coding. Using complex frame-to-frame techniques involving conditional replenishment, bit rates in the range 20 to 25 Mb/s are possible. Notice that the improvement in bit-rate obtained by going to frame-to-frame coding is no more than a factor of 2. It is quite likely that in the near future we might be able to achieve bit rates of 15 Mb/s using, for example, interpolative coding.

Turning to video conferencing results, we immediately see that much bigger reductions are possible, primarily due to the less stringent requirements on motion rendition. For this reason frame-to-frame coding gives a much larger reduction in bit rates relative to intraframe coding. Intraframe coding yields bit rates of 10 to 35 Mb/s depending upon the complexity of the algorithm and the picture quality required. For example, the complex intraframe coder might contain adaptation of the predictor and quantizer with adaptive variable-length encoding, all under buffer control. Frame-to-frame coding yields transmission rates in the range 1.5 to 6 Mb/s. In the near future, we expect to see good-to-excellent results at bit rates in the range 1.5 to 3 Mb/s using for example, motion compensation.

Consider now the applications of picture coding. This is essentially a question of economics. The *raison d'etre* for coding in the first place is to reduce overall transmission or storage costs, thus coders will be used whenever the cost of encoding is attractively less than the cost of providing additional channel capacity. On local transmission links coders will need to be inexpensive, while globe spanning transmissions can support expensive terminals. While technology will no doubt reduce the cost of transmission channels, for the same reasons the cost of coders will also fall and probably more steeply. Today it is not so much a question of whether digital encoding of a television signal would be adopted for a digital transmission facility but rather how sophisticated that coder will be. Up to this point there has been a big distinction between intraframe and frame-to-frame coding, but now that a whole frame can be stored on one IC board, we can expect this distinction to fade.

One of the primary areas of potential application has been for the network transmission of broadcast television signals. Applications in this area have been rather slow. This is partly due to the fact that broadcasters demand that digital encoding not degrade the quality of any studio signal [209]. However, many efficient coders have been designed such that they meet this requirement under normal operation, but for extreme types of material which may be encountered very rarely noticeable degradation may result. More importantly perhaps, digital transmission facilities to carry the digitally encoded signal are not readily available. Overlaying digital facilities on existing analog systems is not efficient. It is difficult to achieve a transmission rate of 44 Mb/s on the existing North American TD-2 broadcast system. Thus, if we assume that a broadcast

quality signal can be obtained at a transmission rate of 22 Mb/s, then it is not obvious that two digital signals can be transmitted in place of one analog signal. Digital satellite links operating in the range 60 to 90 Mb/s offer new possibilities for digital transmission.

Looking to the future we could expect television scenes with perhaps twice the horizontal and twice the vertical resolution of the current standard. This need becomes more obvious with the proliferation of home projection television systems where large screens magnify many shortcomings in the current television standard. An upgrading of the picture could be easily handled within a digital framework. Further, addition of high-quality stereo sound also would be easily handled (of course, we have not addressed here the question of the transmission from the broadcast station to the residence).

Transmission costs are still high enough that we believe they severely limit potential applications of video systems. In many cases the function of the system can be performed by a picture of much lower quality than broadcast television. A good example is video conferencing where very rarely is there a large amount of movement by any of the participants and even when there is, a small amount of blurring or noisiness would not be bothersome. Surveillance applications is another area where high standards are not as important, usable pictures at bit rates as low as 1–5 Mb/s can be obtained. As both transmission and coding costs decrease, we can expect to see an increased demand for these types of services and a significant increase in the application of picture coding in these areas.

Acknowledgment

The authors would like to thank B. G. Haskell, N. S. Jayant, J. B. O'Neal, B. Prasada, C. B. Rubinstein, W. F. Schreiber, and R. Steele for comments on early drafts of the paper.

References

[1] M. R. Aaron, "Digital communications—The silent (R) evolution," *IEEE Commun. Mag.*, vol. 17, pp. 16–26, Jan. 1979.
[2] R. E. Graham, "Subjective experiments in visual communication," *IRE Commun. Rec.*, pp. 100–106, 1958.
[3] D. A. Huffman, "A method for the construction of minimum redundancy codes," *Proc. IRE*, vol. 40, pp. 1098–1101, Sept. 1952.
[4] L. Wilkens, and P. A. Wintz, "Data compression and image processing," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 180–198, Mar. 1971.
[5] W. K. Pratt, "Bibliography on digital image processing and related topics," Univ. of Southern California, USCEE Rep. 453, Sept. 1973.
[6] C. C. Cutler (Guest Ed.), *Proc. IEEE Special Issue on Redundancy Reduction*, vol. 55, Mar. 1967.
[7] T. S. Huang, and O. J. Tretiak, *Picture Bandwidth Compression*. New York: Gordon and Breach, 1972.
[8] M. R. Aaron, Guest Ed., *IEEE Trans. Commun. Tech. Special Issue on Signal Processing for Digital Communications*, vol. COM-19, pt 1, Dec. 1971.
[9] H. C. Andrews, and L. H. Enloe, (Guest Eds.), *Proc. IEEE Special Issue on Digital Picture Processing*, vol. 60, pp. 763–922, July 1972.
[10] N. S. Jayant, Ed. *Waveform quantization and coding.* New York: IEEE Press Selected Reprint Series, 1976.
[11] A. Habibi, Guest Ed., *IEEE Trans. Commun. Special Issue on Image Bandwidth Compression*, vol. COM-25, Nov. 1977.
[12] M. R. Aaron, and A. Lender, (Guest Eds.), *IEEE Trans. Commun. Special Issue on Digital Signal Processing*, vol. COM-26, May 1978.
[13] J. O. Limb, and C. B. Rubinstein, "Digital coding of color video signals—A review," *IEEE Trans. Commun.*, vol. COM-25, pp. 1349–1385, Nov. 1977.
[14] T. S. Huang, "Coding of two-tone images," *IEEE Trans. Commun.*, vol. COM-25, pp. 1406–1424, Nov. 1977.
[15] W. K. Pratt, *Image Transmission Techniques.* New York: Academic Press, 1979.
[16] C. E. Shannon, *The Mathematical Theory of Communication.* Urbana, IL: Univ. Illinois Press, 1949.
[17] —, "Coding theorem for a discrete source with a fidelity criterion," in *Information and Decision Processes*, R. E. Machol, Ed. New York: McGraw-Hill, 1960, pp. 93–126.
[18] D. J. Sakrison, "On the role of the observer and a distortion measure in image transmission," *IEEE Trans. Commun.*, vol. COM-25, pp. 1251–1267, Nov. 1977.
[19] B. R. Hunt, and T. M. Cannon, "Nonstationary assumptions for Gaussian models of images," *IEEE Trans. Syst. Man, Cybern.*, pp. 876–882, Dec. 1976.
[20] E. R. Kretzmer, "Statistics of television signals," *Bell Syst. Tech. J.* vol. 31, pp. 751–763, July 1952.
[21] J. Capon, "Bounds on the entropy of the television signals," Res. Lab. Electron., M.I.T. Tech. Rep. 296, May 1955.
[22] R. J. Allerot, Bell Laboratories, unpublished memo., 1971.
[23] D. J. Connor, R. C. Brainard, and J. O. Limb, "Intraframe coding for picture transmission," *Proc. IEEE*, vol. 60, pp. 779–791, July 1972.
[24] A. Habibi, and P. A. Wintz, "Linear transformations for encoding two-dimensional sources," School Elec. Eng., Purdue Univ., Lafayette, IN, Rep. TR-EE 70-2, Mar. 1970.
[25] J. B. O'Neal and T. R. Natarajan, "Coding isotropic images," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 697–707, Nov. 1977.
[26] A. K. Jain and S. H. Wang, "Stochastic image models and hybrid coding," State Univ. New York, Buffalo, Rep. Oct. 1977.
[27] W. F. Schreiber, "The measurement of third-order probability distribution of television signals," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 94–105, Sept. 1956.
[28] A. Habibi, "Comparison of N-th-order DPCM encoder with linear transformations and block quantization techniques," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 948–956, Dec. 1971.
[29] J. B. O'Neal, "Predictive quantizing differential pulse code modulation for the transmission of television signals," *Bell Syst. Tech. J.*, vol. 45, pp. 689–722, May–June 1966.
[30] K. H. Powers and H. Staras, "Some relations between television picture redundancy and bandwidth requirements," *Trans. AIEE*, vol. 76, pp. 492–496, Sept. 1957.
[31] J. O. Limb, "Entropy of quantized television signals," *Proc. Inst. Elec. Eng.*, vol. 115, pp. 16–20, Jan. 1969.
[32] A. J. Seyler, "Probability distributions of television frame differences," *Proc. IREE (Australia)*, pp. 355–366, Nov. 1965.
[33] B. G. Haskell, F. W. Mounts, and J. C. Candy, "Interframe coding of videotelephone pictures," *Proc. IEEE*, vol. 60, pp. 792–800, July 1972.
[34] D. J. Connor, and J. O. Limb, "Properties of frame-difference signals generated by moving images," *IEEE Trans. Commun.*, vol. COM-22, pp. 1564–1575, Oct. 1974.
[35] J. K. Yan and D. J. Sakrison, "Encoding of images based on two-component source model," *IEEE Trans. Commun.*, vol. COM-25, pp. 1315–1323, Nov. 1977.
[36] N. F. Maxemchuk and J. A. Stuller, "An adaptive intraframe DPCM codec based upon nonstationary image model," *Bell Syst. Tech. J.*, vol. 58, no. 6, pp. 1395–1413, July–Aug. 1979.
[37] D. Anastassiou and D. J. Sakrison, "New bounds to R(D) of additive sources and application to image encoding," 1978, preprint of a paper.
[38] D. S. Lebedev and L. I. Mirkin, "Two-dimensional smoothing of images using composite source model," preprint of a paper.
[39] —, "Digital nonlinear smoothing of images," preprint of a paper.
[40] E. F. Brown, "Television: The subjective effects of filter ringing transients," *J. Soc. Motion Pict. Telev. Eng.*, vol. 78, no. 4, pp. 249–255, Apr. 1969.
[41] D. E. Pearson, "Techniques for scaling television picture quality" in *Picture Bandwidth Compression.* T. S. Huang and O. J. Tretiak, Eds. New York: Gordon and Breach, 1972, pp. 47–95.
[42] F. Kretz, "Introduction to visual picture quality in coding schemes," NATO Advanced Study Institute on Digital Image Processing and Analysis, Bonas, France, June 1976.
[43] C. C. I. R., "Method for the subjective assessment of the quality of television pictures," 13th Plenary Assembly, Rec. 500, vol. 11, pp. 65–68, 1974.
[44] R. E. Mallon, "Application of K-rating to USA NTSC systems," *J. Soc. Motion Pict. Telev. Eng.*, vol. 79, no. 1, pp. 16–19, Jan. 1970.
[45] NTC Rep. no. 7, "Video facility testing, An engineering report of the network transmission committee," The Public Broadcasting Service, June 1975.
[46] S. C. Knauer, "Criteria for building 3D vector sets in interlaced video systems," in *Nat. Telecommunications Conf.*, pp. 44.5-1–44.5-6, 1976.
[47] F. W. Mounts, A. N. Netravali, and B. Prasada, "Design of quantizers for real-time hadamard-transform coding of pictures,"

Bell Syst. Tech. J., vol. 56, no. 1, pp. 21–48, Jan. 1977.
[48] J. D. Olsen and C. M. Heard, "A comparison of the effects of two transform domain encoding approaches," in *Application of Digital Image Processing (SPIE)*, vol. 119, pp. 137–146, 1977.
[49] A. Watanabe *et al.*, "Spatial sinewave responses of the human visual system," *Vision Res.*, vol. 8, pp. 1245–1263, Sept. 1968.
[50] J. R. Hamerly, R. F. Quick, and T. A. Reichert, *A Study of Grating Contrast Judgement*, vol. 17. New York: Pergamon, 1977, pp. 201–207.
[51] M. B. Sachs, J. Nachmias, and J. G. Robson, "Spatial frequency channels in human vision," *J. Opt. Soc. Am.*, vol. 61, pp. 1176–1186, Sept. 1971.
[52] N. Graham and J. Nachmias, "Detection of grating patterns containing two spatial frequencies: A comparison of single channel and multiple channel models," *Vision Res.*, vol. 11, pp. 251–259, Mar. 1971.
[53] P. Moon and D. E. Spencer, "The visual effect of nonuniform surrounds," *J. Opt. Soc. Amer.*, vol. 35, pp. 233–248, Mar. 1945.
[54] P. Mertz, "Perception of television random noise," *J. Soc. Motion Pict. Telev. Eng.*, vol. 54, pp. 8–34, Jan. 1950.
[55] J. O. Limb, "Vision oriented coding of visual signals," Ph.D. thesis, Univ. Western Australia, Nedlands, W.A., Australia, 1966.
[56] F. Kretz, "Subjectively optimal quantization of pictures," *IEEE Trans. Commun.*, vol. COM-23, pp. 1288–1292, Nov. 1975.
[57] G. F. Newell and W. K. E. Geddes, "The visibility of small luminance perturbations in television displays," Res. Dep., British Broadcasting Corp., Rep. No. T.106, 1963.
[58] A. Fiorentini, M. Jeanne, and G. T. DiFranchi, "Measurements of differential threshold in the presence of a spatial illumination gradient," *Atti Fond. Ronchi*, vol. 10, pp. 371–379.
[59] F. X. J. Lukas, U. Tulunay-Keesey, and J. O. Limb, "Thresholds at luminance edges under stabilized and unstabilized viewing conditions," presented at ARVO Spring Conf., May 1978.
[60] T. N. Cornsweet, *Visual Perception*. New York: Academic Press, 1970, p. 276.
[61] T. N. Cornsweet and D. Y. Teller, "Relation of increment threshold to brightness and luminance," *J. Opt. Soc. Am.*, vol. 55, no. 10, pp. 1303–1308, Oct. 1965.
[62] S. Novak and G. Sperling, "Visual thresholds near a continuously visible or a briefly presented light-dark boundary," Opt. Acta, vol. 10, pp. 187–191, Apr. 1963.
[63] W. H. Payne, "Foveal border-contrast," *Vision Res.*, vol. 10, pp. 513–518, June 1970.
[64] A. Vassilev, "Contrast sensitivity near borders: significance of test stimulus form, size, and duration," *Vision Res.*, vol. 13, pp. 719–730, Apr. 1973.
[65] F. E. Campbell, A. Pantle, and R. Sekuler, "Size-detecting mechanisms in human vision," *Science* vol. 162, pp. 1146–1148, Dec. 6, 1968.
[66] B. Julez and C. F. Stromeyer, "Spatial-frequency masking in vision; critical bands and spread of masking," *J. Opt. Soc. Am.*, vol. 62, no. 10, pp. 1221–1232, Oct. 1972.
[67] D. J. Sakrison, M. Halter, H. Mostafavi, "Properties of the human visual system as related to the encoding of images," in *New Directions in Signal Processing in Communication and Control*. Groningen, The Netherlands: Noordhoff International, NATO Advanced Studies Institute Series.
[68] F. W. Campbell, R. H. S. Carpenter, and J. Z. Levinson, "Visibility of aperiodic patterns compared with that of sinusoidal gratings," *J. Physiol*, vol. 204, pp. 283–298, 1969.
[69] F. W. Campbell and J. G. Robson, "Application of Fourier analysis to the visibility of gratings," *J. Physiol.*, vol. 197, pp. 551–566, 1968.
[70] A. J. VanDoorn, J. J. Koenderink, and M. A. Bouman, "The influence of the retinal inhomogeneity on the perception of spatial patterns," *Kybernetik*, vol. 10, pp. 223–230, Apr. 1972.
[71] J. O. Limb and C. B. Rubinstein, "A model of threshold vision incorporating inhomogeneity of the visual field," *Vision Res.*, vol. 17, pp. 571–584, no. 4, 1977.
[72] D. J. Connor and J. E. Berrang, "Resolution loss in video images," *Natl. Telecommunications Conf. Rec.*, pp. 54–60, Dec. 1974.
[73] M. Miyahara, "Analysis of perception of motion in television signals and its application to bandwidth compression," *IEEE Trans. Commun.*, vol. COM-23, pp. 761–766, July 1975.
[74] C. Rashbass, "The visibility of transient changes of luminance," *J. Physiol.*, vol. 210, pp. 165–186, 1970.
[75] U. Lupp and R. P. Dooley, Ed., "Advances in the psychophysical and visual aspects of image evaluation," in *Summary Proc. SPSE Tech. Section Conf.* (Oct. 24–25, 1977, Rochester, NY), 1977.
[76] A. L. Yarbus, *Eye Movements and Vision*. New York: Plenum Press, 1967, ch. 3.
[77] A. J. Seyler and Z. L. Budrikis, "Detail perception after scene changes in television image presentations," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 31–43, Jan. 1965.
[78] N. Weisstein, "Metacontrast," in *Handbook of Sensory Physiology*, D. Jameson and L. M. Hurvich, Ed. Berlin, Germany: Springer-Verlag, 1972.
[79] G. C. Higgins, "Image quality criteria," *J. Appl.* Photographic *Eng.*, vol. 3, pp. 53–60, Spring 1977.
[80] J. O. Limb, "Distortion criteria of the human viewer," *IEEE Trans. Sys., Man, Cybern.*, Dec. 1979.
[81] B. M. Oliver, J. R. Pierce, and C. E. Shannon, "The philosophy of PCM," *Proc. IRE*, vol. 36, pp. 1324–1331, Oct. 1948.
[82] W. M. Goodall, "Television by pulse code modulation," *Bell Syst. Tech. J.*, vol. 30, pp. 33–49, Jan. 1951.
[83] L. Roberts, "Picture coding using pseudorandom noise," *IEEE Trans. Inform. Theory*, vol. IT-8, pp. 145–154, Feb. 1962.
[84] J. O. Limb, "Design of dither waveforms for quantized visual signals," *Bell Syst. Tech. J.*, vol. 48, pp. 2555–2582, Sept. 1969.
[85] B. Lippel, and M. Kurland, "The effect of dither on luminance quantization of pictures," *IEEE Trans. Commun.*, vol. 6, pp. 879–888, 1971.
[86] F. deJager, "Delta modulation, A method of PCM transmission using a 7-unit code," *Philips Res. Rep.*, pp. 442–466, Dec. 1952.
[87] C. C. Cutler, "Differential quantization of communication signals," U.S. Patent 2 605 361, July 1952.
[88] P. Elias, "Predictive coding," *IRE Trans. Inform. Theory*, vol. IT-1, pp. 16–33, Mar. 1955.
[89] C. W. Harrison, "Experiments with linear prediction in television," *Bell Syst. Tech. J.*, vol. 31, pp. 764–783, July 1952.
[90] D. J. Connor, R. F. W. Pease, and W. G. Scholes, "Television coding using two-dimensional spatial prediction," *Bell Syst. Tech. J.*, vol. 50, pp. 1049–1061, Mar. 1971.
[91] B. G. Haskell, "Entropy measurements for nonadaptive and adaptive, frame-to-frame, linear predictive coding of video telephone signals," *Bell Syst. Tech. J.*, vol. 54, no. 6, pp. 1155–1174, Aug. 1975.
[92] J. O. Limb, R. F. W. Pease, and K. A. Walsh, "Combining intra-frame and frame-to-frame coding for television," *Bell Syst. Tech. J.*, vol. 53, no. 6, pp. 1137–1173, Aug. 1974.
[93] R. E. Graham, "Predictive quantizing of television signals," IRE Wescon Conv. Rec. vol. 2, pt. 4, pp. 147–157, 1958.
[94] W. Zschunke, "DPCM picture coding with adaptive prediction," *IEEE Trans. Commun.*, vol. COM-25, no. 11, pp. 1295–1302, Nov. 1977.
[95] I. Dukhovich and J. B. O'Neal, "A three-dimensional spatial nonlinear predictor for television," *IEEE Trans. Commun.*, vol. COM-26, no. 5, pp. 578–583, May 1978.
[96] A. N. Netravali and C. B. Rubinstein, "Luminance adaptive coding of the *th* chrominance signals," *IEEE Trans. Commun.*, vol. COM-27, pp. 703–710, Apr. 1979.
[97] H. Yasuda and H. Kawanishi, "Predictor adaptive DPCM," in *Proc. SPIE Conf. Applications of Digital Image Processing*, vol. 149, pp. 189–195, Aug. 1978.
[98] F. Rocca, "Television bandwidth compression utilizing frame-to-frame correlation and movement correlation," in *Symp. Return Bandwidth Compression* (MIT), April 1969. New York: Gordon and Breach, 1972.
[99] B. G. Haskell and J. O. Limb, "Predictive video encoding using measured subject velocity," U.S. Patent 3 632 865, Jan. 1972.
[100] J. O. Limb and J. A. Murphy, "Estimating the velocity of moving images from television signals," *Computer Graphics and Image Processing*, vol. 4, 1975, pp. 311–327.
[101] C. Cafforio and F. Rocca, "Methods for measuring small displacements of television images," *IEEE Trans. Inform. Theory*, vol. 11-22, pp. 573–579, Sept. 1976.
[102] A. N. Netravali and J. D. Robbins, "Motion compensated television coding—Part I," *Bell Syst. Tech. J.*, pp. 631–670, Mar. 1979.
[103] J. D. Robbins and A. N. Netravali, "Interframe television coding using movement compensation," in *Proc. Inter. Conf. Communications* (Boston, MA), pp. 23.4.1–23.4.5, June 1979.
[104] J. A. Stuller, A. N. Netravali, and J. D. Robbins, "Interframe television coding using gain and movement compensation," *Bell Syst. Tech. J.*, to appear.
[105] B. G. Haskell, "Frame replenishment coding of television," *Image Transmission Techniques*, W. K. Pratt, Ed. New York: Academic, 1979.
[106] J. Woods, "Stability of DPCM coders for television," *IEEE Trans. Commun.*, vol. COM-23, pp. 845–846, Aug. 1975.
[107] P. Jung and R. Lippmann, "Error response of DPCM decoders," *Sonderdruck aus Nachrichtentechnische Zeitschrift*, Bd. 28, pp. 431–536, 1974.
[108] N. F. Maxemchuk and A. N. Netravali, "Performance of DPCM coders for picture signals in the presence of transmission errors," Proc. Inter. Conf. Communications, vol. 1, pp. 12B.3.1–12B.3.6, June 1978.
[109] E. G. Bowen, and J. O. Limb, "Subjective effects of substituting lines in a video-telephone signal," *IEEE Trans. Commun.*, pp. 1208–1211, Oct. 1976.
[110] R. Lippmann, "Influence of channel errors on DPCM picture

coding," *Acta Electron.*, vol. 19, pp. 289–294, 1977.
[111] D. M. Fenwick, R. Steele, and N. Vassanji, "Partial correction of DPCM video signals using a walsh corrector," *Radio Electron. Eng.*, vol. 48, no. 6, pp. 271–276, June 1978.
[112] N. F. Maxemchuk and J. A. Stuller, "Reduction of transmission error in adaptively predicted, DPCM encoded picture," *Bell Syst. Tech. J.*, vol. 58, no. 6, pp. 1413–1425, July–Aug. 1979.
[113] J. Max, "Quantizing for minimum distortion," *IEEE Trans. Inform. Theory*, vol. IT-6, pp. 7–12, Mar. 1960.
[114] B. Smith, "Instantaneous companding of quantized signals," *Bell Syst. Tech. J.*, vol. 36, no. 3, pp. 653–709, Jan. 1957.
[115] T. Berger, "Optimum quantizers and permutation codes," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 754–755, Nov. 1972.
[116] A. N. Netravali and R. Saigal, "Optimum quantizer design using a fixed-point algorithm," *Bell Syst. Tech. J.*, vol. 55, no. 9, pp. 1423–1435, Nov. 1976.
[117] A. N. Netravali, "On quantizers for DPCM coding of picture signals," *IEEE Trans. Inform. Theory*, vol. IT-23, no. 3, pp. 360–370, May, 1977.
[118] F. Kretz, "Degradation des signaux d'images et qualite subjective en codage numerique: visibilite du flottement de contour annales telecommunications," vol. 31, no. 9–10, pp. 333–356, Sept. 1976.
[119] J. O. Limb, and C. B. Rubinstein, "On the design of quantizers for DPCM coders: A functional relationship between visibility, probability, and masking," *IEEE Trans. Commun.*, vol. COM-26, pp. 573–578, May 1978.
[120] J. O. Limb, "Source receiver encoding of television signals," *Proc. IEEE*, vol. 55, pp. 364–380, Mar. 1967.
[121] D. K. Sharma and A. N. Netravali, "Design of quantizers for DPCM coding of picture signals," *IEEE Trans. Commun.*, vol. COM-25, pp. 1267–1274, Nov. 1978.
[122] W. Thoma, "Optimizing the DPCM for video signals using a model of *th* human visual system," *Proc. 1974, IEEE Inter. Zurich Sem. Digital Communication*, pp. C3(1)–C3(7), March 1974.
[123] J. C. Candy and R. H. Bosworth, "Methods for designing differential quantizers based on subjective evaluations of edge busyness," *Bell Syst. Tech. J.*, vol. 51, pp. 1495–1516, Sept. 1972.
[124] H. G. Musmann, "Codierung von videosignalen," *Nachrichtentech. Z.*, vol. 24, pp. 114–116, Feb. 1971.
[125] G. Bostelmann, "A high quality DPCM system for videotelephone signals using 8 Mbit per second," *Nachrichtentech. Z.*, vol. 27, pp. 115–117, Feb. 1974.
[126] T. Kummerow, "DPCM system with two-dimensional predictor and controlled quantizer," *NTG-Tagung Signalverarbeitung Erlangen*, pp. 425–439, 1973.
[127] H. Inose and Ishizaka, "Adaptive coding using two-dimensional prediction for image signals," *Trans. Instit. of Electron. Commun. Eng., Japan*, pp. 527–534, 1974.
[128] E. R. Kretzmer, "Reduced-alphabet representation of television signals," *IRE Natl. Conv. Rec.*, vol. 4, pt. 4, pp. 140–147, 1956.
[129] W. F. Schreiber, C. F. Knapp, and N. D. Kay, "Synthetic highs–An experimental TV bandwidth reduction system," *J. SMPTE*, vol. 68, pp. 525–537, Aug. 1959.
[130] J. O. Limb, "Adaptive encoding of picture signals," presented at Symp. Picture Bandwidth Compression, MIT, Cambridge, MA, Apr. 1969.
[131] A. N. Netravali and B. Prasada, "Adaptive quantization of picture signals using spatial masking," *Proc. IEEE*, vol. 65, pp. 536–548, Apr. 1977.
[132] H. G. Musmann and W. D. Erdmann, German Patent P2740945.6, 1977.
[133] G. W. Aughenbaugh, J. D. Irwin, and J. B. O'Neal, "Delayed differential pulse code modulation," in *Proc. 2nd Annu. Princeton Conf.*, pp. 125–130, Oct. 1970.
[134] C. C. Cutler, "Delayed encoding: Stabilizer for adaptive coders," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 898–907, Dec. 1971.
[135] J. A. Stuller and B. Kurz, "Interframe sequential picture coding," *IEEE Trans. Commun.*, vol. COM-25, pp. 485–495, May 1977.
[136] P. Pirsch, "Block coding of color video signals," in *Proc. Nat. Telecommunications Conf.*, pp. 10:5:1–10:5:5, 1977.
[137] E. N. Gilbert and E. F. Moore, "Variable-length binary encodings," *Bell Syst. Tech. J.*, vol. 38, pp. 933–937, July 1959.
[138] J. O. Limb, "Efficiency of variable-length binary codes," in *Proc. UMR, Mervin J. Kelly Commun. Congr.* (Rolla, MO) pp. 13-1–13-3-9, Oct. 1970.
[139] M. C. Chow, "Variable-length redundancy removal coders for differentially coded video telephone signals," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 923–926, Dec. 1971.
[140] J. B. O'Neal, "Entropy coding in speech and television differential PCM systems," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 758–761, Nov. 1971.
[141] Z. L. Budrikis, J. L. Hullet, and D. Q. Phiet, "Transient-mode buffer stores for nonuniform code TV," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 913–922, Dec. 1971.
[142] S. K. Goyal, and J. B. O'Neal, "Entropy coded differential pulse-code modulation systems for television," *IEEE Trans. Commun.*, vol. COM-21, pp. 660–666, June 1973.
[143] H. Enomoto and K. Shibata, "Features of hadamard transformed television signals," *1965 Nat. Conf. IECE, Japan*, paper no. 881.
[144] W. K. Pratt and H. C. Andrews, "Fourier transform coding of images," in *Proc. Hawaii Inter. Conf. System Sciences*, Jan. 1968.
[145] A. Habibi and P. A. Wintz, "Optimum linear transformation for encoding two-dimensional data," presented at *Symp.* Picture Bandwidth Compression, M.I.T., Cambridge, MA, 1969.
[146] J. W. Woods and T. S. Huang, "Picture bandwidth compression by linear transformation and block quantization," *Symp.* Picture Bandwidth Compression, M.I.T., Cambridge, MA, 1969.
[147] H. J. Landau and D. Slepian, "Some computer experiments in picture processing for bandwidth reduction," *Bell Syst. Tech. J.*, vol. 50, pp. 1525–1540, May–June 1971.
[148] P. A. Wintz, "Transform picture coding," *Proc. IEEE*, vol. 60, no. 7, pp. 809–820, July 1972.
[149] H. Hotelling, "Analysis of a complex of statistical variables into principle components," *J. Educ. Psychol.*, vol. 24, pp. 417–441, 1933.
[150] H. Karhunen, "Uber Lineare Methoden in der Wahrscheinlichkeitsrechnung," *Ann. Acad. Sci. Fenn.*, Ser. A.I. 37, (Helsinki, Finland), 1947.
[151] H. P. Kramer and M. V. Mathews, "A linear encoding for transmitting a set of correlated signals," *IRE Trans. Inform. Theory*, vol. IT-2, pp. 41–46, Sept. 1956.
[152] J. J. Y. Huang and P. M. Schultheiss, "Block quantization of correlated Gaussian random variables," *IEEE Trans. Commun. Syst.*, vol. CS-11, pp. 289–296, Sept. 1963.
[153] N. Ahmed, T. Natarajan, and K. R. Rao, "On image processing and a discrete cosine transform," *IEEE Trans. Comput.*, vol. C-23, pp. 90–93, Jan. 1974.
[154] W. K. Pratt, W. Chen, and L. R. Welch, "Slant transform image coding," *IEEE Trans. Commun.*, vol. COM-22, pp. 1075–1093, Aug. 1974.
[155] R. M. Haralick and K. Shanmugam, "Comparative study of discrete linear basis for image data compressions," *IEEE Trans. Syst., Man, Cybern.*, vol. 4, pp. 16–27, Jan. 1974.
[156] R. C. Singleton, "An algorithm for computing the mixed radix fast Fournier transform," *IEEE Trans. Audio Electroacoust.*, vol. AU-17, pp. 93–103, June 1969.
[157] W. Chen, C. H. Smith, and S. Fralick, "A fast computational algorithm for the discrete cosine transform," *IEEE Trans. Commun.*, pp. 1004–1009, Sept. 1977.
[158] A. Jain, "A fast Karhunen-Loeve transform for a class of stochastic process," *IEEE Trans. Commun.*, vol. COM-24, pp. 1023–1029, Sept. 1976.
[159] ——, "A fast Karhunen-Loeve transform for finite discrete images," in *Proc. Nat. Electronics Conf.*, Oct. 1974.
[160] R. M. Haralick, N. Griswold, and N. Kaltingabulwanich, "A fast two-dimensional Karhunen-Loeve transform," *Proc. SPIE*, vol. 66, pp. 144–159, Aug. 1975.
[161] D. J. Sakrison and V. R. Algazi, "Comparison of line-by-line and two-dimensional encoding of random images," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 386–398, July 1971.
[162] M. Tasto and P. A. Wintz, "Picture bandwidth compression by adaptive block quantization," Purdue Univ. Lafayette, IN, Tech. rep. TR-EE-70-14, June 1970.
[163] S. C. Knauer, "Real-time video compression algorithm for hadamard transform processing," *Proc. SPIE*, vol. 66, pp. 58–69, Aug. 1975.
[164] K. Shibata, "Three-dimensional orthogonal transform coding of NTSC color signals," presented at Picture Coding Symp., Tokyo, Japan, 1977.
[165] J. A. Roese, W. K. Pratt, and G. S. Robinson, "Interframe cosine transform image coding," *IEEE Trans. Commun.*, vol. COM-25, pp. 1329–1339, Nov. 1977.
[166] A. Segall, "Bit allocation and encoding for vector sources," *IEEE Trans. Inform. Theory*, vol. IT-22, pp. 162–169, Mar. 1976.
[167] M. Tasto and P. A. Wintz, "Image coding by adaptive block quantization," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 957–971, Dec. 1971.
[168] W. K. Pratt, J. Kane, and H. C. Andrews, "Hadamard transform coding of images," *Proc. IEEE*, pp. 58–68, Jan. 1969.
[169] G. M. Dillard, "Application of ranking techniques for data compression for image transmission," in *Proc. Nat. Telecommunications Conf.*, vol. 1, pp. 22.18–22.22, 1975.
[170] G. B. Anderson and T. S. Huang, "Piece wise Fourier transformation for picture bandwidth compression," *IEEE Trans. Commun. Technol.*, vol. COM-19, pp. 133–140, Apr. 1971.
[171] E. T. Claire, "Bandwidth compression in image transmission," in *Proc. Int. Communications Conf.*, pp. 39.2–39.13, 1972.

[172] T. I. Gimlett, "Use of activity classes in adaptive transform image coding," *IEEE Trans. Commun.*, vol. COM-23, pp. 785-786, July 1975.
[173] A. N. Netravali, B. Prasada, and F. W. Mounts, "Some experiments in adaptive and predictive Hadamard transform coding of pictures," *Bell Syst. Tech. J.*, vol. 56, pp. 1531-1547, Oct. 1977.
[174] W. H. Chen and C. N. Smith, "Adaptive coding of monochrome and color images," *IEEE Trans. Commun.*, vol. COM-25, pp. 1285-1292, Nov. 1977.
[175] A. G. Tescher, "The role of phase in adaptive image coding," Ph.D. dissertation, Dep. Elec. Eng., Univ. Southern California, Los Angeles, Jan. 1974.
[176] A. G. Tescher and R. V. Cox, "An adaptive transform coding algorithm," *Proc. Int. Communication Conf.*, pp. 47.20-47.25, 1976.
[177] W. C. Wong and R. Steele, "Partial correction of transmission errors in walsh transform image without recourse to error correction coding," *Electron. Lett.*, vol. 14, pp. 298-300, May 1978.
[178] D. O. Reudink, unpublished work, Bell Lab., 1971.
[179] C. Reader, "Orthogonal transform coding of still and moving pictures," Ph.D. dissertation, Univ. Sussex, Sussex, England, 1974.
[180] A. Habibi, "Hybrid coding of pictorial data," *IEEE Trans. Commun.*, vol. COM-22, pp. 614-624, May 1974.
[181] M. Ishii, "Picture bandwidth compression by DPCM in the Hadamard transform domain," *Fujitsu Sci., Tech. J.*, pp. 51-65, Sept. 1974.
[182] J. A. Heller, "A real-time Hadamard transform video compression system using frame-to-frame differencing," presented at Nat. Telecommunication Conf., vol. NTC-74, San Diego, 1974.
[183] R. W. Means, H. J. Whitehouse, and J. M. Speiser, "Television encoding using a hybrid discrete cosine transform and differential pulse code modulator in real-time," *Proc. Nat. Telecommunications Conf.*, San Diego, pp. 61-66, Dec. 1974.
[184] H. W. Jones, "A conditional replenishment Hadamard Video compressor," *SPIE*, vol. 119, Applications of Digital Image Processing, pp. 91-98, 1977.
[185] J. A. Stuller and A. N. Netravali, "Transform domain motion estimation," *Bell Syst. Tech. J.*, pp. 1673-1702, Sept. 1979.
[186] A. N. Netravali and J. A. Stuller, "Motion compensated transform coding," *Bell Syst. Tech. J.*, pp. 1703-1718, Sept. 1979.
[187] C. M. Kurtman, "Redundancy reduction—A practical method of data compression," *Proc. IEEE*, vol. 55, pp. 253-263, Mar. 1967.
[188] D. Hochman, H. Katzman, and D. R. Weber, "Application of redundancy reduction to television bandwidth compression," *Proc. IEEE*, vol. 55, pp. 263-266, Mar. 1967.
[189] L. Ehrman, "Analysis of some redundancy removal bandwidth compression techniques," *Proc. IEEE*, vol. 55, pp. 278-287, Mar. 1967.
[190] L. D. Davisson, "Data compression using straight line interpolation," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 390-394, May 1968.
[191] T. Pavlidis, "Optimal piecewise approximation of functions of one and two dimensional variables," *IEEE Trans. Comput.*, vol. C-24, pp. 98-102, Jan. 1975.
[192] D. Gabor and P. C. J. Hill, "Television band compression by contour interpolation," *Proc. Inst. Elec. Eng.*, vol. 108B, pp. 303-313, May 1961.
[193] J. O. Limb, "Picture coding: The use of a viewer model in source encoding," *Bell Syst. Tech. J.*, vol. 52, pp. 1271-1302, Oct. 1973.
[194] A. N. Netravali, "Interpolative picture coding using a subjective criterion," *IEEE Trans. Commun.*, vol. COM-25, pp. 503-508, May 1977.
[195] K. Murakami, K. Tachibana, H. Fujishita, and K. Omura, "Variable sampling rate coder," Tech. Rep., Osaka Univ. Osaka, Japan, vol. 26, pp. 499-507, Oct. 1976.
[196] W. F. Schreiber, "The mathematical foundation of the synthetic highs system," Res. Lab. Electron., Quarterly Progress report 68, M.I.T., Cambridge, MA, Jan. 1963.
[197] D. N. Graham, "Image transmission by two-dimensional contour coding," *Proc. IEEE*, vol. 55, pp. 336-346, Mar. 1967.
[198] T. Ohira, M. Hayakawa, and K. Matsumoto, "Orthogonal transform coding system for NTSC color television signals," *IEEE Trans. Commun.*, vol. COM-26, pp. 1454-1463, Oct. 1978.
[199] T. Ohira, M. Hayakawa, and K. Matsumoto, "Adaptive orthogonal transform coding system for NTSC color television signals," *Proc. Nat. Telecommunication Conf.* 1978, pp. 10.6.1-10.6.5, Dec. 1978.
[200] H. Enomoto, and K. Shibata, "Orthogonal transform coding system for television signals," *IEEE Trans. Electromagn. Compat.*, vol. EMC-13, pp. 11-17, Aug. 1971.
[201] T. Ohira, M. Hayakawa, and K. Matsumoto, "Orthogonal transform coding system for NTSC color television signal," in *Proc. Int. Communications Conf.*, pp. 4B.3-86-4B.3-90, 1977.
[202] J. C. Candy, M. A. Franke, B. G. Haskell, and F. W. Mounts, "Transmitting Television as clusters of frame-to-frame differences," *Bell Syst. Tech. J.*, vol. 50, pp. 1889-1919, July-Aug. 1971.
[203] K. Iinuma *et al.*, "NETEC-6: Interframe encoder for color television signals," *NEC Res. Devel.*, no. 44, pp. 92-96, Jan. 1977.
[204] J. O. Limb, "Buffering of data generated by the coding of moving images," *Bell Syst. Tech. J.*, vol. 51, no. 1, pp. 239-259, Jan. 1972.
[205] B. G. Haskell, "Buffer and channel sharing by several interframe picturephone colders," *Bell Syst. Tech. J.*, vol. 51, no. 1, pp. 261-289, Jan. 1972.
[206] T. Ishiguro, K. Iinuma, Y. Iijima, T. Koga, S. Azami, and T. Mune, "Composite interframe coding of NTSC color television signals," in *Proc. Int. Communications Conf.*, pp. 6.4-1-6.4-5, 1977.
[207] J. O. Limb and E. G. Bowen, "An interframe coder for broadcast television," in preparation.
[208] R. F. W. Pease, "Conditional vertical subsampling—A technique to assist in the coding of television signals," *Bell Syst. Tech. J.*, vol. 51, no. 4, pp. 787-802, Apr. 1972.
[209] C.C.I.R., "The specification of performance for transmission circuits which employ digital methods," 13th Plenary Assembly, rep. 645, vol. 12, pp. 206-208, 1974.
[210] W. K. Pratt, *Digital Image Processing*. New York: Wiley, 1978.