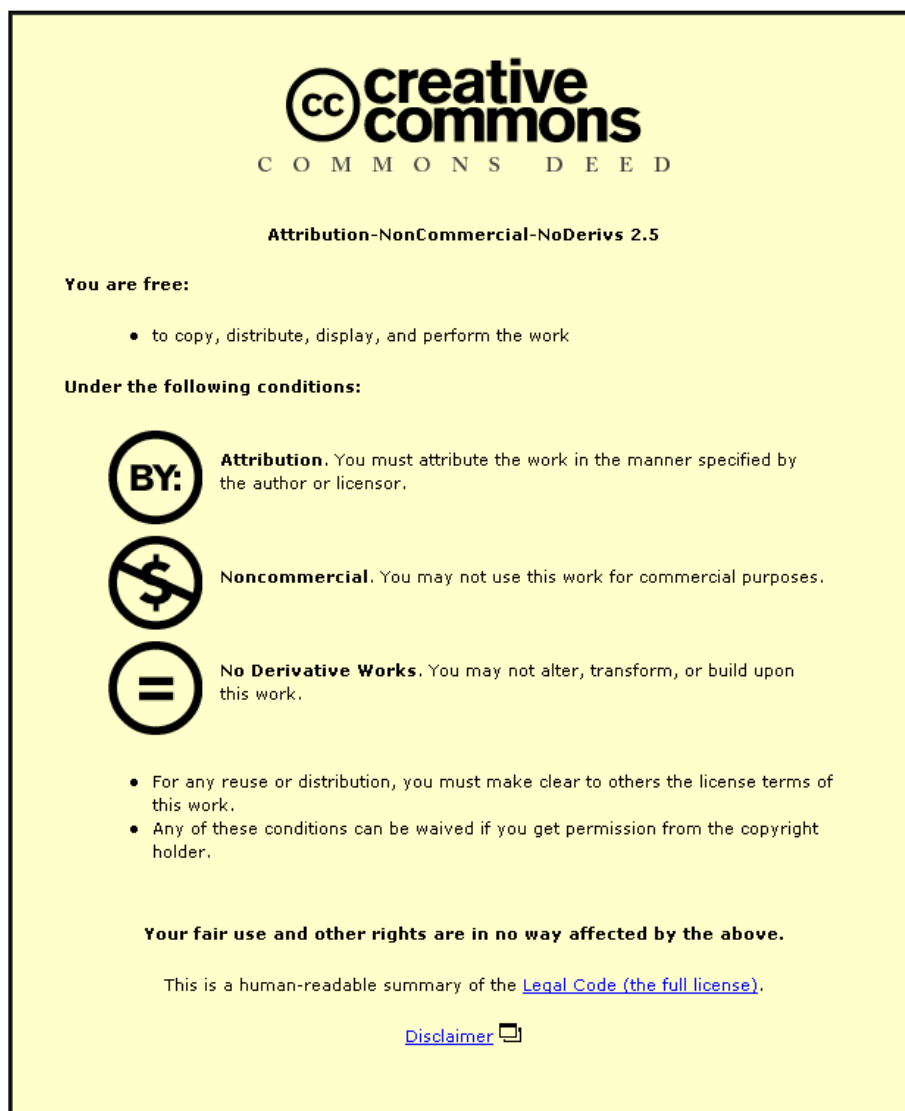


This item was submitted to Loughborough University as a PhD thesis by the author and is made available in the Institutional Repository (<https://dspace.lboro.ac.uk/>) under the following Creative Commons Licence conditions.



For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

LOUGHBOROUGH
UNIVERSITY OF TECHNOLOGY
LIBRARY

AUTHOR/FILING TITLE

KOSTIC, B

ACCESSION/COPY NO.

002623/02

VOL. NO.

CLASS MARK

LOAN COPY

JUN '86

22 NOV 1991

000 2623 02



ANALOGUE AND DIGITAL VIDEO SIGNAL PROCESSING
USING A SCRAMBLING STRATEGY

By

Branko Kostić

B.Sc. Salford University

M.Sc. Loughborough University

A Doctoral Thesis

*Submitted in partial fulfilment of the
requirements for the award of*

Doctor of Philosophy

of the Loughborough University of Technology

© by B. KOSTIĆ, AUGUST 1983.

Original
 of [illegible]
 Date
 002623/02

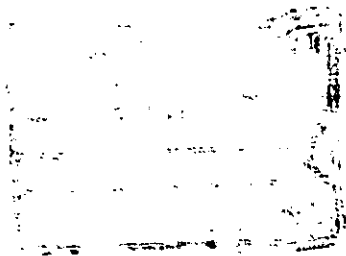
ACKNOWLEDGEMENTS

It gives me great pleasure to express my deep gratitude to my supervisor, Dr. C. Xydeas, for his guidance, encouragement and inspiration throughout this work. His invaluable suggestions and comments helped considerably to initiate and develop the work presented here. I express my special thanks to Professor R. Steele for his friendly enthusiasm, support and learned guidance during the course of this work.

I am also grateful to the Department of Electronic and Electrical Engineering of Loughborough University of Technology for providing the facilities to carry out this work.

Many thanks are also due to Mr. P. Atkinson for preparing the photographic results in this thesis. I also wish to thank all my colleagues with whom I have had many interesting and useful discussions.

I wish to express my sincere gratitude to my parents for their moral and financial support. I am also indebted to my wife, Mary, for her constant encouragement which enabled me to see this work through to completion.



LIST OF PRINCIPAL SYMBOLS AND ABBREVIATIONS

a_j	Predictor coefficients
a_v	Vertical predictor coefficient
a_h	Horizontal predictor coefficient
a_d	Diagonal predictor coefficient
BER	Bit error rate
d_i	Quantizer decision
DPCM	Differential pulse code modulation
ϵ_e^2	Mean square prediction error
f_{sc}	Sub carrier frequency
f_L	Line frequency
FSQ	Fixed Switched Quantization
NMSE	Normalised mean square error
PCM	Pulse code modulation
pdf	Probability density function
r_i	Quantizer reconstruction levels
SNR	Signal to noise ratio
TBR	Transmitted bit rate
ρ	Autocorrelation
VSQ	Variable Switched Quantization
1-D	One dimensional prediction
2-D	Two dimensional prediction

SYNOPSIS

The work described in this thesis proposes and investigates further use of scrambling in industrial analogue and digital monochrome video systems. This scrambling inevitably entails some signal modification. Providing the receiver is able to distinguish between the original and the scrambled signal, regardless of which one was transmitted, more efficient signal exploitation is possible. This more efficient signal exploitation is performed at the expense of the inherent redundancy present in the analogue and digital signals.

Analogue video signals are usually of a highly correlative nature and, this characteristic is exploited, in this thesis, by enabling them to be unwitting data carriers. The video signal is made the data carrier while the data gets a free ride. Each scan-line of the video signal is sampled, and blocks of pels are scrambled or not by modulo masking, depending on whether the data necessary for transmission is a logical '1' or '0' respectively. Prior to transmission the combined data and video sequence is converted into a continuous signal with a bandwidth that is no greater than that of the original video signal. From the knowledge of the original and the modified signal statistics, the receiver is able to perform the inverse operation of the transmitter, recovering the video signal and the data. Three novel systems are proposed for embedding data into analogue pictures. Two of these systems are capable of supporting an average of 17430 and 8713 bits per (256x256) image

respectively, with excellent recovered picture quality. The third system produced a constant number of bits per image, with a slight degradation in the recovered picture quality but, with a capability of conveying up to about 0.5 mega bits/sec of data.

The idea and technique of embedding data into analogue signals was then carried on to the digital method of coding video signals using differential pulse code modulation. However, the scrambling technique here was used to obtain a novel switched quantization scheme, with forward estimation, without the necessity of sending any side information. Scrambling was performed on the quantizer output levels by inversion. Initially, experiments were carried out using fixed length code words with one and two dimensional predictors. Blocks of quantizer output levels are scrambled, or not, depending on which quantizer was used in encoding the video signal. Hence the switching information was carried by the quantized block of error signals. This type of set-up produced only modest improvements. The quantizers were then altered by using a different number of levels and the switching information was carried one block of quantized error signals in advance. As a result, the average bit rate was reduced to about 2.7 bits/pel using a one dimensional predictor with exceptionally good subjective picture quality. When used with a two dimensional predictor, the scheme produced an average bit rate of about 1.7 bits/pel with excellent subjective picture quality.

CONTENTS

	PAGE
<i>Chapter I</i> Introduction	1
1.1. Introduction	1
1.2. Thesis Overview	3
 <i>Chapter II</i> Review	 6
2.1 Introduction	6
2.2. The Analogue Video Waveform	6
2.2.1. Monochrome TV	7
2.2.2. Colour TV	10
2.3. Digital Image Coding	14
2.3.1. Orthogonal Transform Coding	15
2.3.2. Waveform Coding	16
2.3.3. Pulse Code Modulation	17
2.3.4. Differential Pulse Code Modulation	19
2.3.5. Predictor Optimization	21
2.3.6. Quantizer Optimization	23
2.4. Psychophysical Properties of Vision	25
2.5. Assessment of Image Quality	27
 <i>Chapter III</i> Embedding Data Into Pictures By Modulo Masking	 32
3.1 Introduction	32
3.2. Data Embedding Schemes	34
3.2.1. System 1	35
3.2.1.1. The Receiver	38

3.2.1.2. Post Filtering	42
3.2.2. System 2	43
3.2.2.1. Receiver For System 2	45
3.2.2.2. Post Filtering For System 2	47
3.2.3. System 3	48
3.2.3.1. Receiver For System 3	49
3.2.3.2. Post Filtering For System 3	49
3.3. Experimental Procedure	50
3.4. Results	56
3.4.1. System 1	56
3.4.2. System 2	63
3.4.3. System 3	71
3.4.4. Average Picture Signal-To-Noise Ratios	74
3.5. Discussion	79
<i>Note on Publications</i>	81
 <i>Chapter IV Switched Quantization - Fixed Bit Rate</i>	 82
4.1. Introduction	82
4.1.1. Linear Predictors in Video Differential Coders	82
4.1.2. Quantizer in Video Differential Coders	85
4.2. Fixed Bit Rate Switched Quantization (FSQ) Scheme (1-D)	88
4.2.1. The Transmitter	89
4.2.2. The Receiver	96
4.2.3. Experimental Procedure	98
4.2.4. Results	103
4.3. Fixed Bit Rate Switched Quantization (FSQ) Scheme (2-D)	109

4.3.1.	The Transmitter	112
4.3.2.	The Receiver	115
4.3.3.	Experimental Procedure and Results	116
4.4.	Discussion	123
<i>Chapter V</i>	Switched Quantization - Variable Bit Rate	129
5.1.	Introduction	129
5.1.1.	Video Adaptive Quantizers	130
5.2.	Variable Bit-Rate Switched Quantization (VSQ) Scheme (1-D)	132
5.2.1.	The Transmitter	133
5.2.2.	The Receiver	141
5.2.3.	Experimental Procedure and Results	144
5.3.	Variable Bit-Rate Switched Quantization (VSQ) Scheme (2-D)	154
5.3.1.	The Transmitter	155
5.3.2.	The Receiver	160
5.3.3.	Experimental Procedure and Results	162
5.4.	Discussion	170
<i>Chapter VI</i>	Conclusion and Suggestion For Future Work	172
6.1.	Introduction	172
6.2.	Embedding Data Into Pictures By Modulo Masking	173
6.3.	Switched Quantization - Fixed Bit Rate (FSQ)	174
6.4.	Switched Quantization - Variable Bit Rate (VSQ)	176
References		179

CHAPTER I

DIGITAL TELEVISION

1.1 Introduction

The transmission and storage of pictorial information is of practical importance in a wide variety of applications including, digital television, picturephone, biomedical, industrial quality control and others. In certain applications, such as television, the amount of existing analogue equipment is so enormous that an immediate conversion to digital transmission is unlikely. For other applications such as space exploration, digital image transmission is mandatory because of the desired error rates and limited power capability of a space craft. For the majority of applications however, a combination of digital methods with analogue and optical methods may be employed to achieve more efficient, good quality image communications. With recent advances in very large scale integrated circuit (VLSI) technology it is anticipated that the combination of analogue and digital signal processing will enter consumer homes in the next few years by the introduction of digital television sets (not to be confused with digital television transmission, a different technology).

In digital television transmission the received signal is already digitized. It is then processed and recovered into an analogue signal for transfer to the picture tube. To achieve this practically the present day TV signals would have to be digitally transmitted requiring a large bandwidth that is too large for current technology

to handle economically. Instead the analogue signal would be transmitted and demodulated, at the receiver, using conventional analogue circuitry, and the demodulated signal would then be digitized. The complete digitization of the demodulated signal offers a great many advantages some of which are:

- (a) Digital TV sets should provide better quality images.
- (b) It is possible to lock onto a synchronization signal to suppress interference from unwanted signals.
- (c) The capacity of digital TV sets to store entire pictures can allow a faster rescanning of each picture on the screen to suppress the flicker effect.
- (d) The ability to zoom in on any part of a picture.

In addition to this it would be easier to make more efficient use of the existing analogue signals. Such a scheme is proposed and investigated using industrial quality monochrome analogue pictures, in the first part of this thesis.

Digital television is also likely to be of great importance in the development of direct digital TV transmission. It will permit direct processing of the transmitted signals. However, the central problem in digital image transmission is channel capacity or achieving storage reduction while maintaining an acceptable fidelity or image quality. The possibility of bandwidth reduction is indicated by two observations. First, there is a large amount of statistical redundancy or correlation in normal images. For example, two points that are spatially close together tend to have nearly the same brightness level.

Second, there is a large amount of psychovisual redundancy in most images. That is, a certain amount of information is irrelevant and may be eliminated without causing a loss in subjective image quality. Therefore, the digital transmission of television signals would be greatly facilitated if improved coding methods could be found to obtain more economical numerical descriptions of picture signals. Many methods have been devised that reduce the channel capacity by removing the redundancy and irrelevant information from the images. In the second part of this thesis attention is focused on one particular part of the communication system namely, the source encoding aspect of a digital transmission system.

1.2 Thesis Overview

In this section a brief outline of the work described in each chapter of this thesis is given. The results presented in this thesis have all been obtained using computer simulation.

Chapter II is also an introductory chapter in which a review is provided of some of the coding techniques as applied to video signals. The survey presented is not intended to be exhaustive but rather to provide the background material to the work described in later chapters. It begins with a brief description of the formation and format of the analogue video signals. Following that an outline is given on some of the techniques used in digital encoding, decoding, of video signals with emphasis on digital waveform encoding. The remainder of that chapter includes some important characteristics of psychovisual

properties of the human visual perception system with methods for assessing picture quality.

Chapter III describes a technique for embedding data into analogue video pictures. This means that as well as receiving the analogue signal one is able to receive digital data, by using the analogue signal to convey the digital data information at the same time. The technique utilizes a scrambling, or signal modification, strategy to achieve this simultaneous video and data transmission. The purpose is not to describe numerous scrambling techniques, but rather to suggest a method whereby more efficient exploitation of the analogue signal is possible. Three systems are proposed for embedding data into industrial quality analogue pictures. Two of these systems are able to convey the data information in a variable rate bit stream and one is able to convey this information in a constant rate bit stream.

Chapter IV describes a novel switched quantization scheme, with feed forward estimation, using an intraframe DPCM encoder without the need of transmitting any side information. It is first used in a differential encoder with a first order predictor and a 3 bit/pel fixed code assignment. It is then used in a differential encoder with a 2-D, third order predictor and a 2 bit/pel fixed code assignment. Using a scrambling, or signal modification, technique the encoded blocks carried their own information as to which quantizer (of two quantizers) was used to encode the transmitted block of pels. The performance of both systems, using 1-D and 2-D predictors, is illustrated by

means of signal to noise ratio (SNR) calculations and recovered images.

In chapter V, the novel switched quantization scheme used in chapter IV is modified, so that better quality images are obtained at a reduced average bit rate, compared to those in chapter IV. It is first used in a differential encoder with a first order predictor and then after further modification with a 2-D third order predictor. Using the scrambling or signal modification technique, the encoded blocks carried the information, as to which quantizer (of two quantizers using a different number of levels) was used to encode the transmitted block of pels, one block of pels in advance. Although this type of system set-up is more complex, compared to that of chapter IV, it does provide impressive SNR values and vastly improved recovered image quality, than those presented in chapter IV.

Finally, in chapter VI, the main results reported in this thesis are analyzed and criticized.

CHAPTER II

REVIEW

2.1 Introduction

The object of a television system is to convey to the viewer visual information. Television signals, analogue or digital, contain redundancy, which increases the channel requirement for transmission in its original form. Advances in VLSI and computer technology, for mass storage and digital processing, have paved the way for implementing advanced image processing techniques to improve the efficiency of transmission and storage of images. Before describing and examining the schemes proposed in this thesis, presented in this chapter are some coding techniques as applied to video signals. This will provide the background material to the work portrayed in later chapters.

2.2 The Analogue Video Waveform

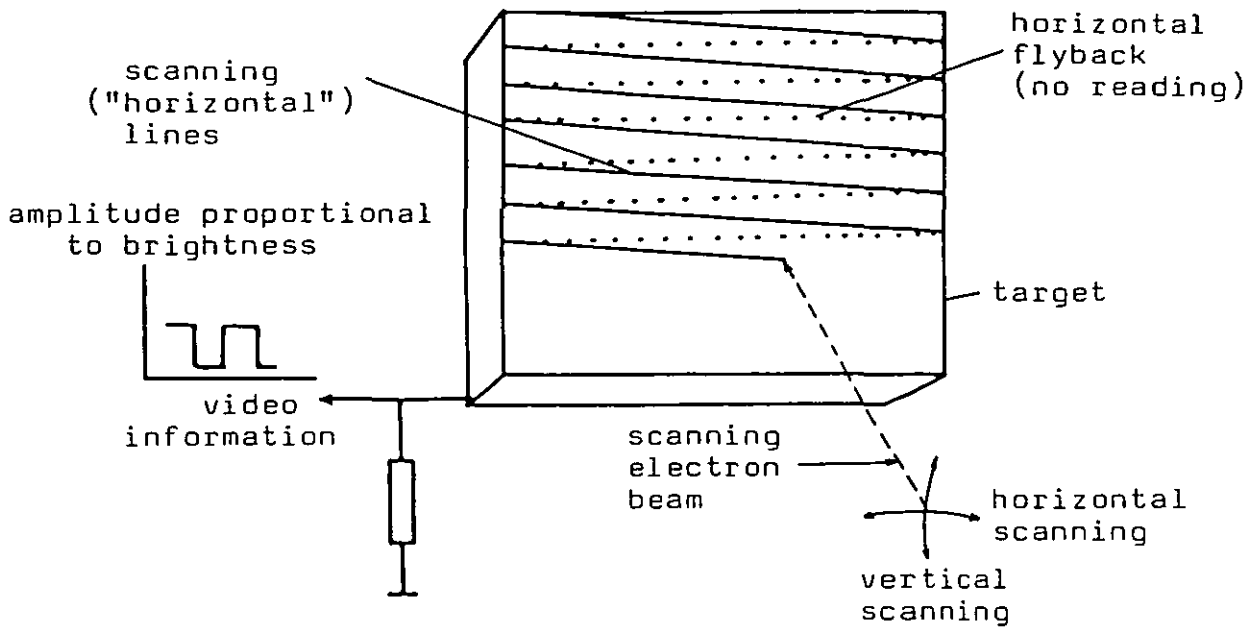
Presented in this section is a short and quantitative description of the formation and format of the analogue video waveform. For a more complete analysis, interested readers can refer to the many text books on the subject. The two books by Carnt and Townsend⁽¹⁾ cover all major colour TV systems with a large part devoted to colour representation. A useful introductory book, due to its concise block-diagram approach, is presented by Sims⁽²⁾. The book by Pearson⁽³⁾ differs from most other text books because of its emphasis on signal formation and analysis as a three-dimensional sampling process, aspects of digital transmission reception, and subjective assessment of pictures.

2.2.1. Monochrome TV

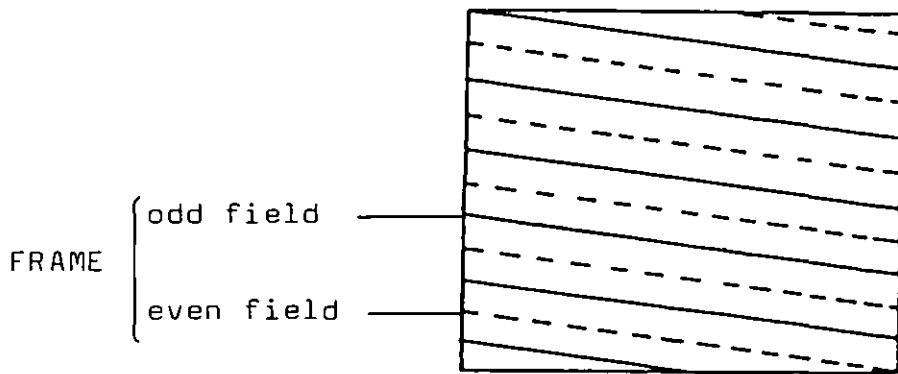
The first point to note is that a camera is required to convert a two dimensional image into a single dimension time waveform. The two-dimensional electronic image of the scene to be transmitted is formed by a distribution of electronic charges whose density is proportional to the brightness of the light falling onto the opto-electronic image of the picture tube⁽⁴⁾. This distribution of charges is 'read' by a scanning beam of electrons. This reading, i.e. the information retrieving process, consists of the electronic beam scanning the target left to right and from top to bottom. The visual information is then extracted as a series of vertically spaced and nearly horizontal lines, in much the same way as words are arranged and read on this page. When the scanning beam is returning (flyback) from left to right and from the bottom to the top of the picture, the information retrieval process is inhibited, i.e. 'blanked'. The process is illustrated in fig.2.1.a.

In system I, adopted in the U.K., the blanking level is also the black level, but in some other systems, such as M, adopted in the U.S.A., they can differ. The time interval during which the television waveform is kept in the blanking level is known as the horizontal and vertical blanking interval.

One complete scanning of the image, from top to bottom, constitutes one 'field'. After a field is produced, the scanning beam is returned to the top of the picture to produce another field, and so on. This process is equivalent to sampling a two-dimensional image in the



(a)



(b)

Fig.2.1. Scanning in TV

(a) principle (b) interlaced scanning

vertical direction by a series of horizontal lines that covers the entire image. The higher the number of horizontal lines, the higher is the spacial resolution in the vertical direction.

The frequency of repetition of the field is determined by the need to provide continuity of brightness in the television picture, i.e. pictures without 'flicker' effects. In system I, adopted in the U.K., the field frequency is set to 50 Hz. However, continuity of movement can be achieved if a sequence of frames is presented to the eye at about half that rate. Therefore, to accommodate efficiently the requirements of continuity of movement and brightness, each frame is made up of two interlaced fields, fig.2.1.b. The two fields are said to be interlaced as the lines that constitute each field are spacially interlaced. The scanning process that generates such fields is said to be line interlaced scanning. In the non-interlaced scanning, the fields and frames coincide. This separation of frames into fields follows closely the process in motion pictures where 24 frames per second are recorded in a film, to provide continuity of movement, and each frame is then projected twice (i.e. at a rate of 48 projections per second) to give continuity of brightness. To produce 50 frames per second in TV or 48 different frames per second in motion pictures would necessitate redundancy for the average moving scene. In motion pictures, this redundancy would be reflected by the doubling of the films physical length and in TV it would double the video signal bandwidth⁽⁴⁾ (other parameters kept constant).

As the reconstruction of the picture in the receiver requires that the horizontal lines, the fields and the frames be synchronized, horizontal and vertical synchronizing pulses (in short sync pulses) are added to the video waveform. To distinguish the sync pulses from the visual data, they are added towards the blacker than black levels. To distinguish the horizontal sync pulses from the vertical sync pulses, their durations are made different. Horizontal sync pulses are also provided during the vertical flyback so that horizontal line oscillator in the receiver is always in synchronism with the transmitter oscillator. As the timing of the horizontal lines relative to the vertical flyback differs by half a line period between two consecutive fields due to interlacing, equalizing pulses are provided during the vertical sync pulse duration. The complete video signal waveform is known as the composite video signal, and is shown in fig.2.2.

2.2.2. Colour TV

The fundamental feature of colour that is exploited for the development of chromatic television is that most colours can be represented as a linear combinations of three primary colours. The main features that allows the colour TV system to be compatible with a monochrome TV system are :

- (a) the brightness in the colour system and the luminance signal is the result of a linear combinations of the luminances of the three primaries
- (b) the colour information can be transmitted with a bandwidth that is much smaller than that of the

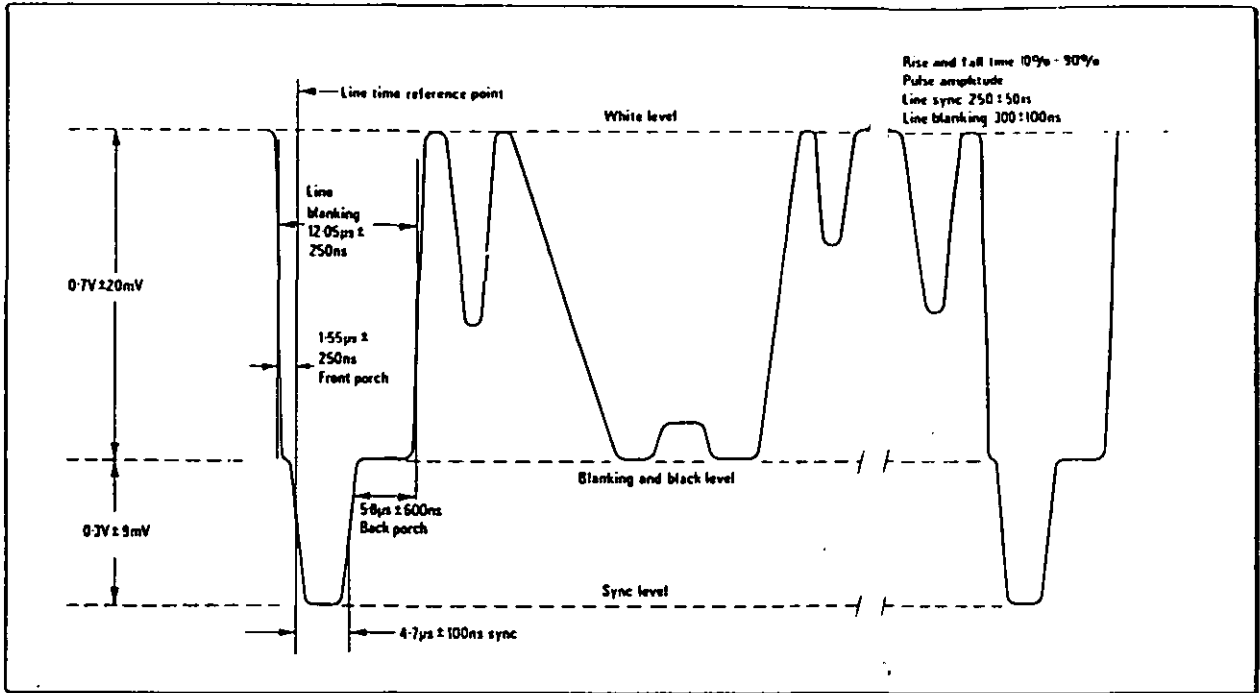


Fig.2.2(a) The waveform of a typical line showing synchronising signals (ref.5)

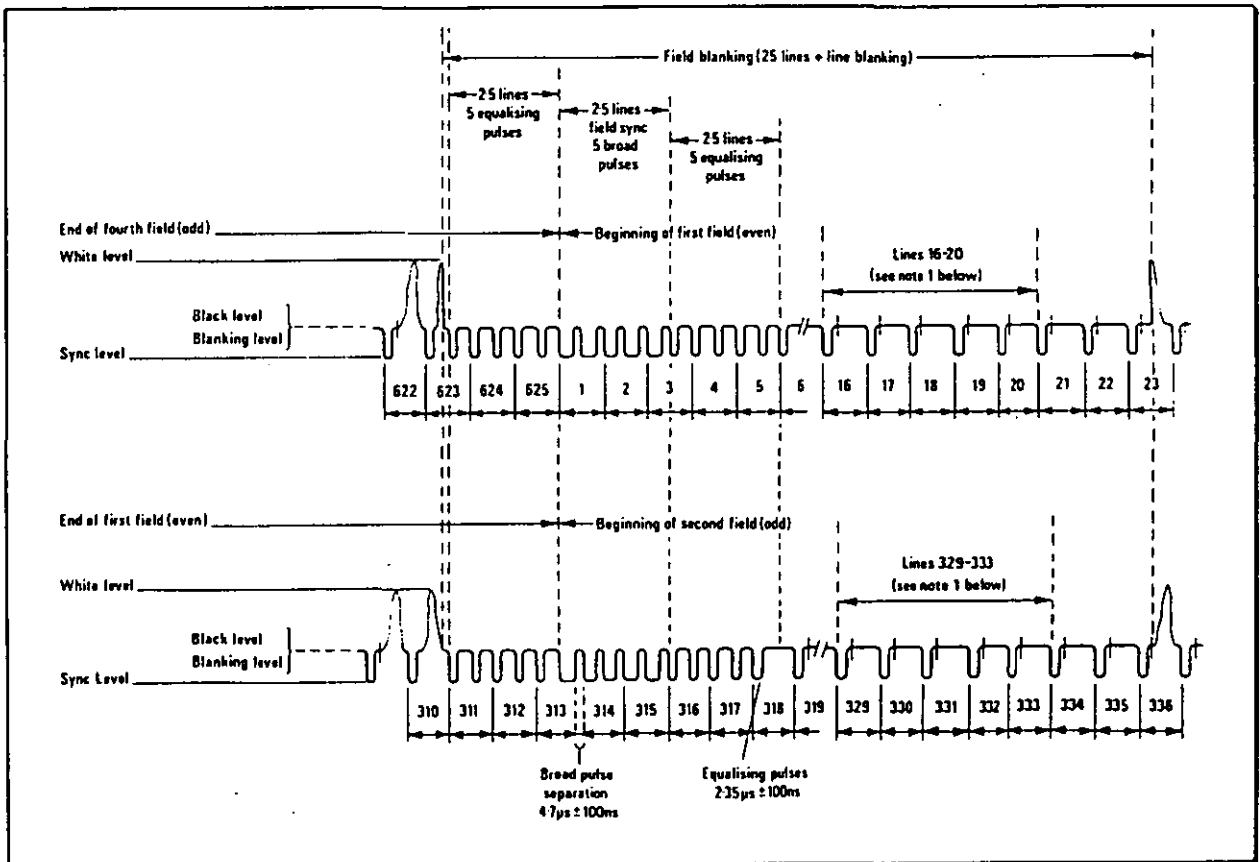


Fig.2.2(b) Vertical synchronising and blanking waveforms for a typical signal (ref.5)

luminance signal

(c) the power spectrum of the monochrome TV signal presents gaps as the spectral components are clustered around the harmonics of the line frequency, forming a line spectrum. The colour information can be inserted in these gaps.

(d) apart from the luminance of the colour, there are only two more features that need to be transmitted to characterize the colour: the hue and saturation.

These can be transmitted as one vector whose amplitude and phase (relative to a reference frequency) convey the saturation and hue data, respectively.

Taking advantage of these features, in the colour TV the colour source is first split optically into its three primary components, from each of which electrical signals are derived following a process similar to that for monochrome TV. The three colour components are usually referred to in literature as the R, G and B signals, after the colours used as the primaries in TV, i.e. red, green and blue. From these three colour components, three other signals are derived for transmission, by a linear matrixing process: a luminance (Y) and two colour-difference or chrominance signals (U and V). These are linear combinations of the differences R-Y, B-Y and G-Y with the weights for the combination depending upon the particular standard of colour transmission. For the PAL system I, the luminance and chrominance signals are⁽⁵⁾:

$$\left. \begin{aligned} Y &= 0.299R + 0.587G + 0.114B \\ U &= \frac{B - Y}{2.03} \\ V &= \frac{R - Y}{1.14} \end{aligned} \right\} \text{---} \quad (2.1)$$

Generally when no colour is present in the picture the colour difference signals are zero and only the Y signal is finite and conveys the brightness information. When colour is present in the picture the colour signals R, G and B become unequal, but Y, indicative of the brightness and the producer of black and white pictures on monochrome TV receivers, retains the proportions of 0.299, 0.587 and 0.114 of the R, G and B signals.

The chrominance signals modulate one colour subcarrier in quadrature. Quadrature modulation is accomplished by U modulating the subcarrier signal, in amplitude with suppressed carrier, at a certain phase and V does the same with a subcarrier of exactly the same frequency as for U, but with its phase altered by $\pi/2$ (thus quadrature). The resulting signal produces a subcarrier modulated in phase and amplitude. The phase carries the information about the hue of the colour, whereas the amplitude carries the saturation information.

The frequency of subcarrier is carefully chosen so that the chrominance signal spectrum falls within the gaps left vacant by the luminance spectrum. In the NTSC system, this happens if the subcarrier frequency is equal to an odd multiple of half the line frequency (half line off-set). However, in the PAL system the V component phase is

switched by 180° at each spectral line of the V signal relative to that of the U signal. To avoid the coincidence of the V component spectrum lines with those of the Y spectrum, the offset employed in the PAL colour system is close to a quarter of the line frequency. Unlike in the NTSC system, the spectrum of the two chrominance signals in the PAL system do not coincide with each other. The subcarrier frequency for PAL-I system is:

$$f_{sc} = (284 + \frac{1}{4}) f_L + 25 \text{ Hz} \quad (2.2)$$

The 25 Hz added to the quarter line offset frequency is intended to minimize the visibility of the subcarrier in the monochrome reception. With

$$f_L = 15.625 \text{ KHz}$$

$$f_{sc} = 4.43361875 \text{ MHz}$$

2.3. Digital Image Coding

The digital transmission of pictorial information was first accomplished with telegraph equipment over 50 years ago. The current importance of and interest in this area are reflected in the large number of publications and special issues of journals directed towards image coding. For a general introductory survey of digital picture coding, a paper by Habibi and Robinson⁽⁶⁾ offers an overview of this field of research that is suitable for the non-specialized reader. The proceedings of a conference on bandwidth compression of picture signals have been published

as a book⁽⁷⁾ and it has a number of review/survey type of papers on the most important aspects of picture bandwidth compression: the human observer, some of the theory behind data compression and major categories of video coding techniques.

The paper by Limb, Rubinstein and Thompson⁽⁸⁾ presents a comprehensive review of coding colour TV signals, giving also the background of colour signals representation and colour perception. Detailed examples of three typical systems, combining some of the coding principles, are given in a picture coding review by Netravali and Limb⁽⁹⁾. There is also a large number of papers on the subject of picture coding in special issues of the IEEE journals⁽¹⁰⁻¹⁵⁾ and books⁽¹⁶⁻²⁰⁾.

Digital methods for coding picture signals can be broadly classified into two categories:

- (a) orthogonal transform coding systems
- (b) waveform coding techniques

2.3.1. Orthogonal Transform Coding

Transform coding was first applied to coding of one-dimensional signals and then, many years later, to the coding of pictures. In orthogonal transform coding the video signal is transformed into a new domain where the correlation between coefficients is much smaller than that of the samples in the time domain sequence. The objective of transform coding is to produce images of good subjective quality using the least number of bits.

Figure 2.3 shows the block diagram of a transform coding system. Many transformations have been devised that produce

coefficients with less correlation than the image itself.

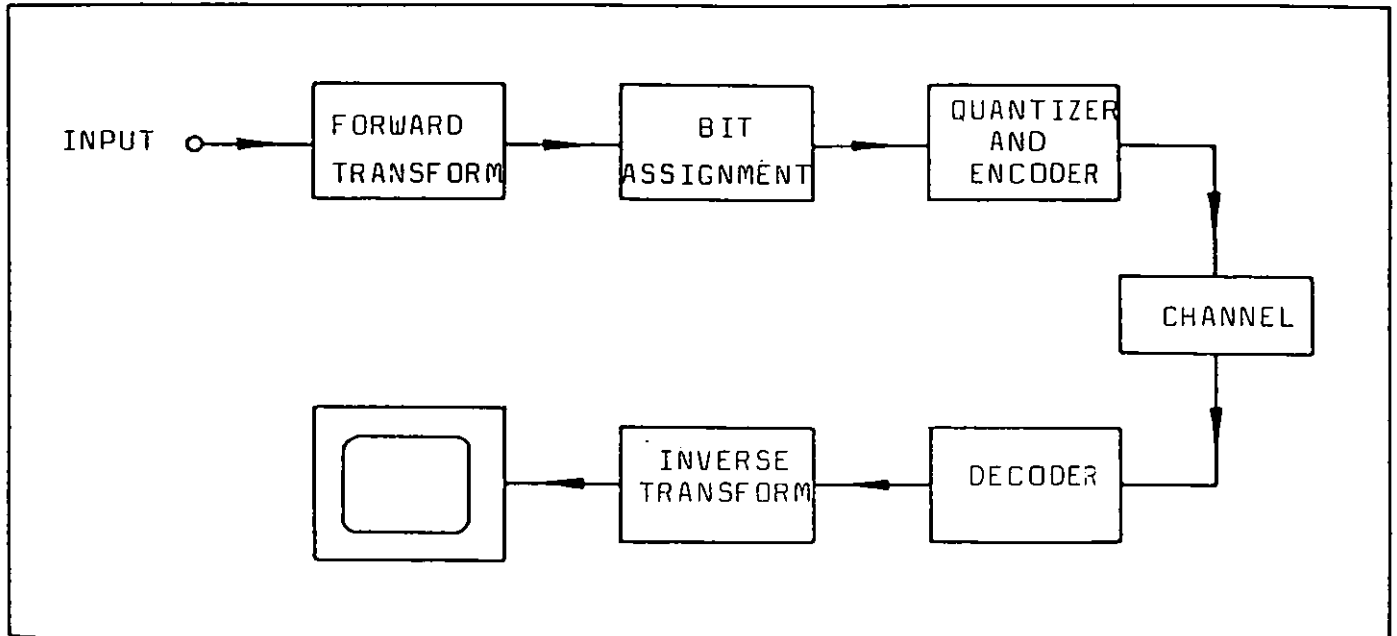


Figure 2.3 Block diagram of a transform coding system

Among these are the Kahunen-Loeve transform⁽²¹⁾, discrete cosine transform⁽²²⁾, Hadamard transform⁽²³⁾, Slant transform⁽²⁴⁾ and Haar transform⁽²⁵⁾. Transform coding techniques are able to achieve good quality image reproduction at about 0.5 bits/pel⁽²⁶⁾ for still pictures. The main difficulty regarding transform coding resides in the fact that it generally requires very complex equipment.

2.3.2. Waveform Coding

In video waveform coding systems the objective is to reconstruct the television picture from its time samples. The video waveform is sampled, encoded and transmitted. An encoding scheme widely used for the transmission of digital signals is pulse code modulation (PCM).

2.3.3. Pulse Code Modulation

Pulse code modulation was invented by Reeves⁽²⁷⁾ and has been considerably refined and analyzed since its invention⁽²⁸⁻³¹⁾. In its simplest form, PCM consists of sampling the amplitude of a band-limited analogue signal and then quantizing this signal into the nearest one of N different levels. The rate of sampling must exceed the Nyquist rate, i.e. twice the highest frequency in the analogue signal. For image signals the number of quantization levels must be large enough so that the quantization noise does not cause any contouring effects on the reconstructed image. The samples are then binary encoded to produce a PCM signal that is both amplitude and time quantized. An image produced in this way can, after passing through a transmission channel, be perfectly restored, provided the channel has sufficient bandwidth and that the interference does not exceed half the peak-to-peak value of the binary PCM signal at the output of the encoder. The receiver simply decides at each clock instant which binary level is present. Having reformed the binary signal, the receiver decodes the binary numbers into a voltage sample. To illustrate the whole concept of PCM, a block diagram is presented in fig.2.4.

One of the most important applications of pulse code modulation to digital television is the analogue to digital conversion^(32,33) in which the analogue video signal is converted into a series of fixed length words prior to encoding. For monochrome images, 50 or more quantization levels are required to prevent grey scale

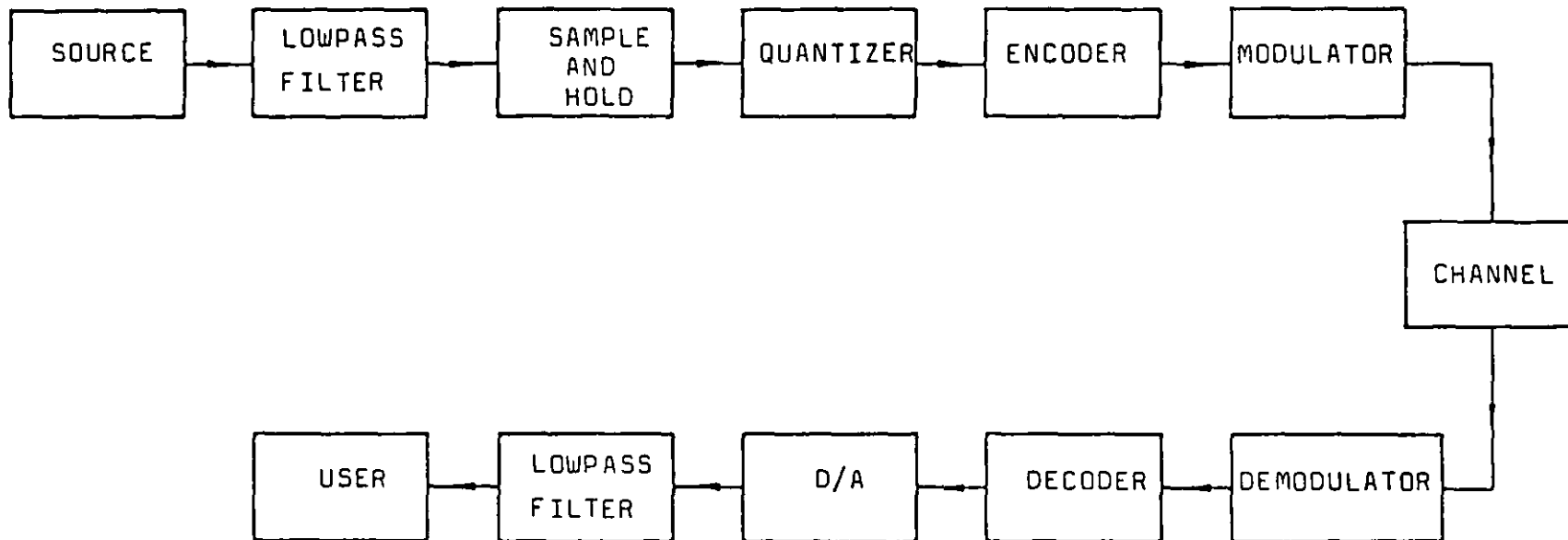


Fig.2.4. Block diagram of PCM system

contouring. Consequently, in a PCM monochrome image coder, the image luminance is usually quantized at between 6 and 8 bits per picture element (pel or pixel). In terms of SNR (measured as the peak-to-peak signal to RMS noise) a value from 38 dB to 50 dB is required to obtain satisfactory subjective image quality.

Since PCM quantization ignores the correlation that exists between samples in the video signal, more efficient digital encoders can be designed by considering this and other factors.

2.3.4 Differential Pulse Code Modulation

The general concept of linear predictive coding has developed from Cutlers⁽³⁴⁾ invention of differential pulse code modulation, abbreviated DPCM. His original patent proposed that integrators be employed to predict the present sample based on the previous sample value along a line, and that the difference between the present sample and its estimate be quantized and coded for transmission. Later efforts in linear predictive coding are due to numerous workers⁽³⁵⁻³⁹⁾ who improved the basic system and extended it to two and three-dimensions.

As seen earlier in PCM, the representation of a sample is made without removing much statistical or perceptual redundancy from the picture signal. Predictive coding techniques make use of the picture signal statistics and more recently, human visual characteristics^(40,41) to reduce the bit rate. It is common knowledge that television signals are highly correlated, both spatially, and temporally.

This linear statistical dependence indicates that linear prediction of sample values based on the values of the neighbouring pels will result in prediction errors that have a smaller variance than the original pels. Due to the smaller variance of the signal to be quantized, coded and transmitted, the amplitude range of the quantizer and thus the number of quantization levels can be reduced. Hence a fewer number of bits per pel can be assigned than for the PCM system for the same SNR.

A block diagram of the DPCM codec is shown in fig.2.5. The input signal X_i are 8-bit PCM samples. For

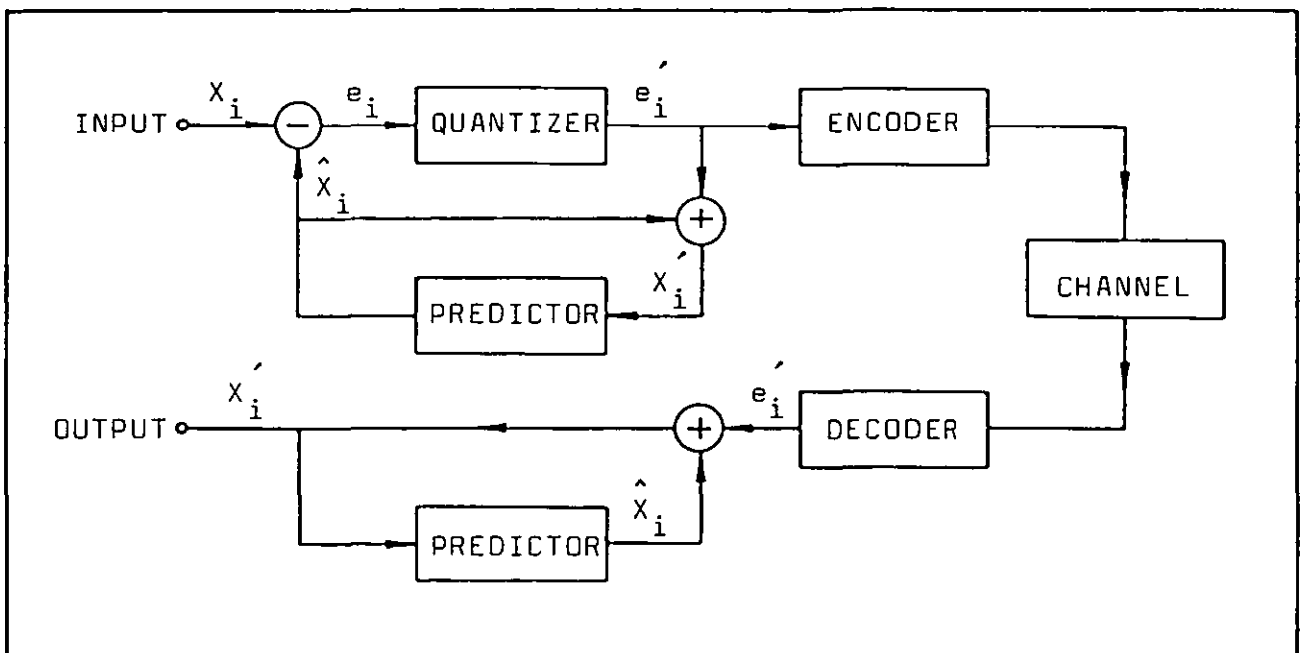


Figure 2.5. Block diagram of a DPCM system

every input sample X_i the linear predictor generates a prediction value $\hat{X}_i$, which is calculated from N previous samples according to the relation:

$$\hat{X}_i = \sum_{j=1}^N a_j X_{i-j} \quad (2.3)$$

Only previously transmitted samples are used for the prediction so that the receiver is also able to calculate $\hat{X}_i$. The predictor coefficients a_j are optimized to yield a prediction error with minimum variance from :

$$e_i = X_i - \hat{X}_i \quad (2.4)$$

The prediction error is then quantized to produce e'_i and coded for transmission. At the receiver the output X'_i is reconstructed and can be shown to differ from the input signal X_i by the quantization error q_i :

$$X'_i = X_i + q_i \quad (2.5)$$

2.3.5. Predictor Optimization

The values of a_j must be chosen such that the mean square error is minimized. Hence

$$\begin{aligned} \epsilon_e^2 &= E[e_i^2] \\ &= E[(X_i - \hat{X}_i)^2] \\ &= E\left[\left(X_i - \sum_{j=1}^N a_j X_{i-j}\right)^2\right] \end{aligned} \quad (2.6)$$

Expanding eq.(2.6) into :

$$\epsilon_e^2 = E \left[x_i^2 \right] - 2 \sum_{j=1}^N a_j E \left[x_i x_{i-j} \right] + \sum_{j=1}^N \sum_{\ell=1}^N a_j a_\ell E \left[x_{i-j} x_{i-\ell} \right] \quad (2.7)$$

or in matrix form :

$$\epsilon_e^2 = \epsilon_x^2 - 2A^T G + A^T R A \quad (2.8)$$

where

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a_N \end{bmatrix} \quad G = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \rho_N \end{bmatrix} \quad R = \begin{bmatrix} \rho_0 & \rho_1 & \dots & \dots & \dots & \rho_{N-1} \\ \rho_1 & \rho_0 & & & & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & & \cdot & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot \\ \rho_{N-1} & \dots & \dots & \dots & \dots & \rho_N \end{bmatrix}$$

The elements of G and R are the values of the autocovariances function of the input signal x_i . To find the predictor coefficients a_j that satisfy eq.(2.8) the partial derivatives of ϵ_e^2 with respect to each a_j are set to zero. Thus

$$\left. \frac{\partial \epsilon_e^2}{\partial A} \right|_{A=A_{opt}} = 0$$

or

$$- 2G + 2AR = 0 \quad (2.9)$$

Solving eq.(2.9) for A_{opt} gives :

$$A_{opt} = R^{-1}G \quad (2.10)$$

Therefore knowing the autocovariances of the input signal the optimum predictor coefficients a_j based on a minimum mean square error criterion can be calculated. For this case the value of ϵ_e^2 is :

$$\epsilon_{e(min)}^2 = \epsilon_x^2 - G^T R^{-1}G \quad (2.11)$$

or

$$\epsilon_{e(min)}^2 = \epsilon_x^2 - A_{opt}^T G \quad (2.12)$$

It should be noted that the above analysis assumes stationarity and neglects the effect of quantization in the DPCM coder. For coders which produce high quality pictures, effects of quantization are small and may be neglected.

2.3.6. Quantizer Optimization

Quantization is the other important operation which determines the encoding performance. In most cases quantizer optimization has been based on a minimum mean square error criterion. Such quantizers were originally studied by Panter and Dite⁽⁴²⁾ and an algorithm for their design was presented by Max⁽⁴³⁾. Nitadori⁽⁴⁴⁾ and O'Neal⁽⁴⁵⁾ found that for speech and television, well designed DPCM systems had quantizer input signals whose probability density functions were approximately Laplacian.

It is well known that the statistics of picture signals are nonstationary and that the required fidelity of reproduction demanded by the human eye varies from picture element to picture element (pel or pixel). Consequently, for more efficient digital coding it is desirable to design the coding strategies to those properties of the picture signal which determine the sensitivity of human observers to quantization noise. The design of quantizers based on the visibility threshold of quantization errors has been developed by Sharma and Netravali⁽⁴¹⁾. In their model, the visibility threshold, defined as the amount of quantization error at a picture element which is masked by the signal slope at the point (represented by the signal error) is obtained by carrying out subjective tests. Masking being defined as the influence of the signal surrounding a point in a picture on the perception of the signal at that point. The method by which the decision and reconstruction levels are determined is illustrated in fig.2.6. Hence the reconstruction levels are given by the intersection of the dotted 45° lines and the abscissa. The decision levels can be obtained by dropping perpendiculars from the intersections of the dotted 45° lines and the visibility threshold curve to the abscissa. The procedure is continued until the last reconstruction level exceeds the amplitude range of the prediction error. This technique has been extended to quantize colour signals by Pearson⁽⁴⁶⁾, achieving a rate of 5 bits per pel with good picture quality.

It is generally agreed that visual masking is an

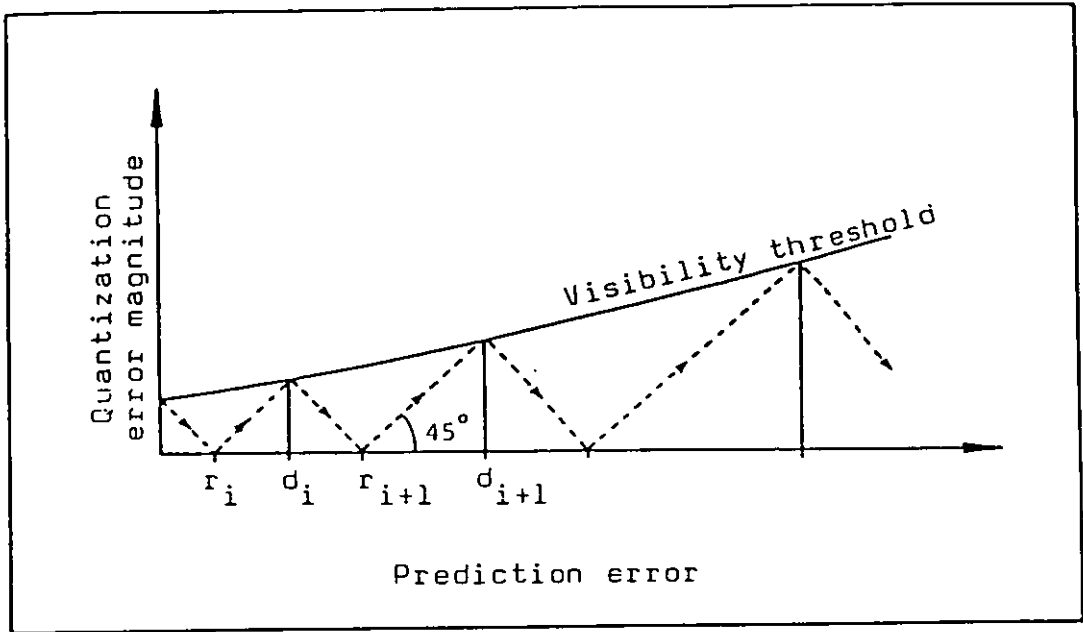


Fig.2.6. Symmetrical visibility threshold quantizer design procedure⁽⁴¹⁾.

extremely nonlinear process. The observed complexities indicate that it involves much or all of the visual system. No single theoretical formulation has been able to account for the different cases of masking. Due to these difficulties, a very wide variety of masking situations have been studied and reported in the literature^(40,41,47-49). Therefore the design of quantizers still remains an art and somewhat ad hoc.

2.4. Psychophysical Properties of Vision

Since the end user of an image processing system is the human observer, an understanding of the eye response^(3,50-54) to a visual stimulation is clearly useful in the design of such systems. This knowledge can be used to develop conceptual models of the human visual process. These models are not only useful in the design of image processing systems, but also in the construction of measures for image fidelity and intelligibility. Some of

the more important characteristics of achromatic vision are contrast resolution, spatial resolution and temporal resolution.

Contrast resolution relates to the ability of the eye to detect small changes of luminance. Against a uniform background of luminance L the just noticeable difference in luminance ΔL increases proportionally with L . This is known as Weber's law and can be formulated as :

$$\frac{\Delta L}{L} = K \quad (2.13)$$

where K is a constant in the range 0.01 to 0.02. However, this result does not hold at very low and very high values of L ^(54,55). Nevertheless Weber's law hints that encoding distortions can be larger in areas of increasing brightness. It was also noted⁽⁵⁵⁾ that the visibility of the changes depends greatly on the adaptation of the eye to the background field. As the eye takes a finite time to respond to excitation changes (persistence of vision), such as detail and movement, picture content has a strong influence on the perception of any distortion or noise. Therefore one can expect that there should be an interaction between the spatial and temporal responses of the eye. With reservation, some of the more important properties of the eye to monochromatic visual stimulation are^(3,55,56):

- (a) noise is less visible in pictures with higher amount detailed areas
- (b) random noise is less visible, and less annoying, than correlated noise and regular patterns are more

noticeable than random ones.

- (c) sharp changes in luminance levels in space or time reduce the visibility of small changes in their vicinity (masking).
- (d) the higher the amount of movement (high temporal frequencies), the less noticeable are the details (high spacial frequencies)
- (e) at high spacial frequencies, contrast resolution is reduced
- (f) at high luminance levels, the eye is more sensitive to flicker but less to luminance changes (spacial)
- (g) flicker is less noticeable in high detailed areas

In conclusion, it is never too often to remind ourselves that not enough is known about monochromatic vision (and colour vision), and specially how different types of distortions and noise effect our perception of TV pictures. The brief overview of psychophysical properties of vision given here should be sufficient to justify some of the video processing techniques used in this thesis.

2.5. Assessment of Image Quality

Two of the commonly used criteria of image quality are image fidelity and intelligibility. Image fidelity characterizes the deviation of a processed image from some standard images, while image intelligibility denotes the ability of man or machine to extract relevant information from the image. It is clearly desirable to formulate quantitative measures of image fidelity and intelligibility as a basis for the design and evaluation of image processing systems. Much progress has been made in this direction but

the measures that have been developed so far do not correlate well with the subjective rating of the image, and this is mainly due to our still poor understanding of the human visual system.

As a criterion of performance the SNR is often used and sometimes defined as :

$$\text{SNR} = 20 \log \frac{\sum_{i=1}^N X_i}{\sum_{i=1}^N (X_i - \hat{X}_i)} \quad (2.14)$$

or

$$\text{SNR} = 20 \log \frac{V_{pp}}{\left[\frac{1}{N} \sum_{i=1}^N (X_i - \hat{X}_i)^2 \right]^{1/2}} \quad (2.15)$$

or

$$\text{SNR} = -10 \log (\text{NMSE}) \quad (2.16)$$

where

$$\text{NMSE} = \frac{\sum_{i=1}^N (X_i - \hat{X}_i)^2}{\sum_{i=1}^N X_i}$$

and X_i - original image samples

$\hat{X}_i$ - processed image samples

V_{pp} - value of the peak-to-peak signal of the original image samples

As mentioned earlier, these SNR measures do not correlate well with subjective evaluation of the image.

However, it does provide an indication of the performance of a video system, and, coupled with informal subjective testing, it is a useful tool for the measurement of image fidelity.

Experiments to judge image quality using human observers have been conducted by many workers⁽⁵⁷⁻⁶⁰⁾. In some instances, untrained, 'non-expert' observers are used so that judgement represents image quality as perceived by the average viewer. In others, tests are conducted with trained 'expert' observers experienced in image processing work and allegedly better able to provide a critical judgement of picture quality.

There are two common types of subjective evaluation : absolute and comparative. In the former, observers are shown an image and asked to judge its quality according to some predefined rating scale. The viewer may be shown a set of reference images or may have to rely on his previous viewing experience. Comparative evaluation involves the viewer ranking a set of images from 'best' to 'worst' for a particular group of images. Table 2.1 summarises the commonly used rating scales for subjective evaluation of images.

Subjective rating results are normally calculated as a mean opinion score defined as :

$$C = \frac{\sum_{i=1}^N n_i C_i}{\sum_{i=1}^N n_i} \quad (2.18)$$

Table 2.1. Rating scales for subjective evaluation
of images⁽⁵⁷⁾

I. Quality scale

- 5. Excellent
- 4. Good
- 3. Fair
- 2. Poor
- 1. Bad

II. Impairment scale

- 5. Imperceptible
- 4. Perceptible, but not annoying
- 3. Slightly annoying
- 2. Annoying
- 1. Very annoying

III. Comparsion scale

- +3 Much better
- +2 Better
- +1 Slightly better
- 0 The same
- 1 Slightly worse
- 2 Worse
- 3 Much worse

where n_i is the number of images judgedd to be in the i^{th} category and C_i is the numerical category value. At least twenty subjects are considered necessary to ensure statistical confidence in such subjective image quality experiments.

The results of the subjective testing are influenced by the types of images presented to the viewer and the viewing conditions. If the images are familiar to the observer, he is more likely to be more critical of impairments because of preconceived notions of the image structure. Furthermore, test viewing conditions should be designed to match 'typical' viewing conditions as closely as possible. In the U.K., the viewing distance is specified as 6 times the picture height, the peak luminance as 50 cd/m^2 and the number of observers (who should be non-experts) around 20-25.

CHAPTER III

EMBEDDING DATA INTO PICTURES BY MODULO MASKING

3.1. Introduction

There is a consistent trend to digitize communication networks yet one may anticipate that for many years, analogue signals, such as speech, facsimile, and television, will continue to be transmitted. By their very nature, analogue speech and picture signals are highly correlated and this characteristic can be exploited to enable them to be unwitting data carriers. The process of inserting data, such as computer traffic, into the speech and picture signals causes the transmitted signals to be different from the original signals. This signal modification far from being undesirable, may be welcomed, making it fatiguing or difficult for an eavesdropper to comprehend the information being transmitted. What must be ensured is that the speech or picture can be recovered at the receiver with an acceptably small perceptual degradation, and that the data be regenerated with a bit error rate (BER) below a specified level.

Steele and Vitello⁶¹ embedded data into speech by using scrambling techniques. The principle of their scheme is as follows⁶². A correlative source has N of its symbols scrambled by a symbol from a data source. One may view a data symbol as being composed of n bits, from which 2^n scrambling keys can be formulated. The receiver is required to deduce the scrambling key in order to recover the original block of N symbols, and in determining the key

the receiver de facto knows the n-bit data symbol. Other methods of embedding data with analogue signals have been proposed. Wong, Steele and Xydeas⁶³ have inserted data onto the phase of speech signals, while Brown and Netravali⁶⁴ have proposed a method for inserting data into pictures using predictive encoding techniques. Feher et al has described methods for incorporating data with microwave analogue signals, known as data-above-voice (DAV) and data-under-voice (DUV)⁶⁵, and data-above-video (DAVID)⁶⁶. Recently, Lockhart and Al-Jalili⁶⁷ have suggested a scheme for superimposing data on amplitude-modulated signals by arranging for the data to conjugate the complex zeros of the modulated signal. Consideration has also been given^{68,69} to simultaneous transmission of high-speed data and analogue samples.

Starting with the concepts of Steele and Vitello^{61,62} three methods of embedding data into monochrome pictures conceived, and are investigated in this chapter have been . Specifically, contiguous blocks of N pels along the scan lines (one dimensional processing) are used to support data. One bit may be inserted into a block by scrambling the luminance level of only one pel, rather than N pels. The method of scrambling is modulo masking, where the luminance level of the pel to be scrambled is modulo added to a fixed number(s). Another feature of this data embedding strategy is picture modification prior to data insertion, and post filtering to enhance the recovered picture after data has been extracted. In the experimental procedure six different monochrome pictures were used, each picture having 256 lines

with 256 pels per line. It was possible to embed data into these pictures, ranging from approximately 8 kbits to 21 kbits per picture, to regenerate the bits without a single error, to avoid an expansion of the picture band-width when data was inserted, and to suffer only mild degradation in picture quality.

Section 3.2 describes the basic procedures for embedding data into pictures, and for separating the data and the pictures at the receiver. Three schemes are discussed. The data rate in System 1 and 2 vary with picture statistics, with the former having for a given picture the greater transmitted bit rate (TBR), and the poorer received picture quality. System 3 is a constant bit rate system that has the highest TBR, the lowest picture quality, and a robustness to data errors that is better than that of System 1 and 2. The experimental arrangement for evaluating the systems is described in section 3.3. The performance of the three data embedding schemes is discussed in section 3.4, while the final section is concerned with an assessment of the various proposals.

3.2. Data Embedding Schemes

The video signal, bandlimited to f_c Hz, is sampled at a rate $f_s > 2f_c$, and uniformly quantized by an 8-bit quantizer to yield a sequence of luminance samples. Data is then embedded into this sequence, subject to criteria to be described. The combined picture and data multilevel samples are then converted into a continuous analogue signal by an interpolating filter which also band limits this signal to f_c .

Described in this section are the data embedding and extraction algorithms of three different schemes, which are simply referred to as Systems 1, 2 and 3. Later in section 3.4 it is shown that these schemes have different performance in terms of transmitted bit rate, robustness to transmission errors, and quality of recovered images.

3.2.1. System 1

Commencing with the first three pels on the first scan line having luminance levels x_1, x_2 , and x_3 , where the subscripts specify the pel location, form :

$$A_1 = \frac{(x_1 - x_2)^2 + (x_2 - x_3)^2}{2} \quad (3.1)$$

The parameter A_1 may be considered as a measure of local picture activity, and if :

$$A_1 < T_{a,1}; \text{ data embedded} \quad (3.2)$$

where $T_{a,1}$ is a system threshold, it is considered that the block of $N=3$ pels has a sufficiently low activity for it to support one bit of data. Prior to data insertion the value of x_2 is modified to :

$$\hat{x}_2 = \frac{x_1 + x_2 + x_3}{3} \quad (3.3)$$

and then data is inserted into the block of three pels by either modulo masking scrambling the luminance level of $\hat{x}_2$, or by a similar scrambling technique, depending on whether the data is a logical 1 or 0, respectively.

Note at this juncture that the representation of modulo addition is

$$u = v \oplus w \bmod z \quad (3.4)$$

which means v is added to w , modulo z , to give a number u ; and

$$v \triangleq u \ominus w \bmod z \quad (3.5)$$

to be the modulo z inverse subtraction enabling v to be recovered from u knowing w .

The data is embedded into the block of pels by scrambling the luminance level $\hat{x}_2$ according to

$$\tilde{x}_2 = \begin{cases} \hat{x}_2 \oplus 128 \bmod 256 ; \text{ data logical 1} \\ \hat{x}_2 - 70 , \text{ if } \hat{x}_2 > 69 ; \text{ data logical 0} \\ \hat{x}_2 + 70 , \text{ if } \hat{x}_2 \leq 69 ; \text{ data logical 0} \end{cases} \quad (3.6)$$

Therefore, if a logical 1 is present the change in luminance level is 128, whereas if a logical 0 is to be transmitted the luminance level undergoes the smaller change of 70 quantized levels. In section 3.4 it will be shown that by these two scrambling strategies the receiver is able to differentiate the absence or presence of data, and if data is present, whether it is logical 1 or logical 0. It is not claimed that the luminance changes of 128 and 70 quantized levels are optimum, but they do enable good results to be achieved, both in terms of zero bit error rate and recovered picture quality for the six pictures used in the

experiments.

Thus when data is embedded into the first block of pels, the centre pel is initially smoothed to $\hat{x}_2$, and then significantly changed to $\tilde{x}_2$. Consequently the data-conveying block is composed of luminance levels

$$S_T = x_1, \tilde{x}_2, x_3 \quad (3.7)$$

Should the local activity of the luminance levels x_1, x_2, x_3 , be excessive, i.e.,

$$A_1 \geq T_{a,1} ; \text{ data not embedded} \quad (3.8)$$

no data is transmitted. However, to assist the receiver in its task of deciding if the block of pels contains data, the activity A_1 of this block is lowered by altering the centre pel to $\hat{x}_2$ using eq.(3.3), and formulate the output sequence

$$S_T = x_1, \hat{x}_2, x_3 \quad (3.9)$$

Summarizing, when the activity factor A_1 is below the threshold $T_{a,1}$ one bit of data is permitted to be embedded into the block of three pels. This results in an increase in the activity factor of the output sequence S_T . However, when A_1 is too high, no data is inserted into the pels, and the activity of the output sequence S_T is lowered by smoothing x_2 according to eq.(3.3). Thus the centre pel x_2 is always changed, either by luminance averaging followed

by scrambling, see eqs.(3.3) and (3.6), or only by luminance averaging.

Having processed the block of the first three pels on the first scan-line, the operation continues on the next block of three pels on the same line. The procedure continues until every block of three pels on the first line has been examined and data has been inserted where the criteria permits. The second scan-line in the picture is next processed in an identical way to that of the first line, and this data embedding technique continues on subsequent lines until data is embedded, where allowable, in the entire frame. The multi-level samples are then converted into a continuous analogue signal, bandlimited to f_c , by an interpolating filter.

The transmitted bit rate TBR of the embedded data in the analogue video signal is a variable, and for television the TBR has a maximum of one N^{th} of the pels per frame multiplied by the number of frames per second, and TBR has a minimum of zero, viz :

$$0 < \text{TBR} < f_F M_1 M_2 / N \quad (3.10)$$

where f_F is the frame rate, M_1 is the number of pels per line and is assumed to be an integer multiple of N , M_2 is the number of lines/frame, and $N=3$. In the case of a single frame, f_F is unity, and TBR becomes the number of bits in the frame.

3.2.1.1. The Receiver

The continuous combined picture and data signal is

samples at f_s Hz to yield the received sequence. From this sequence :

$$S_R = x'_1, x'_2, x'_3 \quad (3.11)$$

where a prime above the symbol implies its presence at the receiver :

$$\lambda_2 = \frac{x'_1 + x'_3}{2} \quad (3.12)$$

is formed. A distance measure :

$$D = |x'_2 - \lambda_2| \quad (3.13)$$

is made, and compared with two distance thresholds $T_{d,1}$ and $T_{d,2}$. If :

$$D \leq T_{d,1} ; \text{ no data present} \quad (3.14)$$

it is concluded that no data is present, i.e. x'_2 is considered to be the receiver version of $\hat{x}_2$ given by eq.(3.3). From eq.(3.3), the original intensity of the centre pel was

$$x_2 = 3\hat{x}_2 - x_1 - x_3$$

and consequently the luminance of the recovered pel at the receiver is formulated as

$$x_2 = 3x'_2 - x'_1 - x'_3 \quad (3.15)$$

The regenerated luminance sequence of pels for the first-block on the first scan-line is, therefore,

$$L_{1,1} = x'_1, x_2, x'_3 \quad (3.16)$$

Whenever Inequality (3.14) is inapplicable D is tested against $T_{d,2}$, viz :

$$D > T_{d,2} ; \text{ data of logical 1} \quad (3.17)$$

and if D does exceed this high threshold, where

$$T_{d2} > T_{d1}$$

it is concluded that the original x_2 has experienced a modulo 256 addition with the introduced level 128, and x'_2 must therefore be supporting data of logical 1. Consequently x'_2 must be descrambled (see eq.(3.5)) to give

$$x'_2 = x'_2 \ominus 128 \text{ mod } 256 \quad (3.18)$$

i.e., 128 is modulo 256 subtracted from x'_2 . The value of x'_2 is further smoothed according to

$$x_2 = \frac{x'_1 + x'_2 + x'_3}{3} \quad (3.19)$$

and the regenerated sequence of pels is given in eq.(3.16),

where x_2 is determined using eq.(3.19). If Inequalities (3.14) and (3.17) have not been satisfied D must be confined to

$$T_{d,1} \leq D \leq T_{d,2} ; \text{ data of logical 0} \quad (3.20)$$

and it is deemed that data has been transmitted of logical 0, and to descramble x_2' proceed as follows. It is known that the transmitter either added or subtracted 70 from $\hat{x}_2$, see eq.(3.6), and arbitrarily assume that 70 was subtracted. Consequently :

$$x_2' = x_2 + 70$$

and formulate a new distance measure

$$D_1' = |x_2' - \lambda_2|$$

If this x_2' is correct the distance measure D_1' will be small, and if the no data test of eq.(3.14) is repeated, which in this case is

$$D_1' \leq T_{d,1}$$

then it is known that the x_2' level is correct. Should D_1' exceed $T_{d,1}$ the original addition of 70 to x_2' was wrong, in which case

$$x_2' = x_2 - 70$$

Thus,

$$x'_2 = \begin{cases} x'_2 + 70 & \text{if } |x'_2 - \lambda_2| \leq \tau_{d,1} \\ x'_2 - 70 & \text{if } |x'_2 - \lambda_2| > \tau_{d,1} \end{cases} \quad (3.21)$$

The luminance level x'_2 is then smooth according to eq.(3.19) to yield x_2 , and it is this value of x_2 that appears in the regenerated sequence of pels specified by eq.(3.16).

Having processed the first block of pels on the first scan line, the examination of subsequent blocks on the same line is continued, and then on the next line, and so on, until the entire frame has been scrutinized. By this procedure the data is removed from the frame and the luminance levels of the middle pels in each block are appropriately modified to be in close agreement with their original values.

3.2.1.2 Post Filtering

The sequence of output pels for the first and second blocks is given by eq.(3.16) and

$$L_{2,1} = x'_4, x_5, x'_6 \quad (3.22)$$

respectively, where x_5 is the modified value of the received pel x'_5 . The value of x_5 is found in a similar way to that of x_2 , see eqs.(3.15), (3.19) and (3.21) where subscript 2 is replaced by 5. It was found experimentally that the effect of channel dispersion on a block of pels whose centre pel had been subjected to modulo masking scrambling as a means of data intertention, was to cause

distortion in neighbouring pels. The greatest distortion was inflicted on the last pel in the block, as it immediately followed the scrambled pel whose luminance level had been significantly altered by the modulo addition. Thus, for the first block of data, x'_3 contains an error far greater than the error in the next pel x'_4 belonging to the second block. The luminance errors in the picture are therefore reduced by replacing the luminance level of the pel at the end of the blocks by an average luminance value computed using the last two pels in the previous block, and the first pel in the current block. This procedure is only applied if the subsequent block is also used to convey data. Thus, if the first three blocks are employed as a data carrier, x'_3 and x'_6 are smoothed to,

$$x_3 = \frac{x_2 + x'_3 + x'_4}{3} \quad (3.23)$$

and

$$x_6 = \frac{x_5 + x'_6 + x'_7}{3} \quad (3.24)$$

it will be observed that the post filtering is performed across the blocks. Finally, an interpolating filter provides the continuous picture signal, bandlimited to f_c Hz, for display on the picture monitor.

3.2.2. System 2

The block size N for this system is 5. Again the blocks are one dimensional along the scan-lines, and the system commences by considering the first five pels along

the first scan-line. The activity factor

$$A_2 = \frac{(x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - x_4)^2 + (x_4 - x_5)^2}{4} \quad (3.25)$$

is formed, x_1, x_2, x_3, x_4 and x_5 are the luminance levels of the first five pels. In order to decide if data is to be embedded into this block of pels A_2 is tested against a system threshold $T_{a,2}$. If :

$$A_2 < T_{a,2} ; \text{ data embedded} \quad (3.26)$$

the activity is considered to be sufficiently low to accommodate one bit of data. Accordingly, either luminance level x_2 or x_4 is scrambled depending on the logical value of the data bit, viz :

$$\tilde{x}_2 = x_2 \oplus 128 \text{ mod } 256 ; \text{ data of logical 0} \quad (3.27)$$

or

$$\tilde{x}_4 = x_4 \oplus 128 \text{ mod } 256 ; \text{ data of logical 1} \quad (3.28)$$

The output luminance sequence is therefore,

$$S_T \begin{cases} x_1, \tilde{x}_2, x_3, x_4, x_5 ; \text{ data of logical 0} \\ x_1, x_2, x_3, \tilde{x}_4, x_5 ; \text{ data of logical 1} \end{cases} \quad (3.29)$$

Should

$$A_2 \geq T_{a,2} ; \text{ no data embedded} \quad (3.30)$$

the activity in the block is deemed to be too high to support data, and in general the pels are conveyed to the interpolating filter without luminance modification, i.e.,

$$S_T = x_1, x_2, x_3, x_4, x_5 ; \text{ no data transmitted} \quad (3.31)$$

However, there is a small probability that the data detection process at the receiver will make a bit error by asserting the presence of a bit when no bit was transmitted. To reduce the likelihood of this generation of false bits at the receiver the data is decoded locally at the transmitter in the manner used by the receiver (described in the next sub-section), and if no data is embedded in a block it is observed whether the local decoder erroneously identifies the existence of a data bit. If a false bit is locally identified the data bearing luminance is smoothed with its adjacent pels, see eq.(3.3) thereby reducing the probability that a false bit will be generated at the receiver. However, for the six pictures used in these experiments no such smoothing was required, i.e., the local decoding was not utilized as not a single false bit error occurred.

The transmitted bit rate of the data embedded in the video signal is again a variable, and eq.(3.10) applies with $N=5$.

3.2.2.1. Receiver For System 2

The received sequence for the first block of pels is

$$S_R = x'_1, x'_2, x'_3, x'_4, x'_5 \quad (3.32)$$

and forming

$$\lambda_2 = \frac{x'_1 + x'_3}{2}$$

eq.(3.12) and

$$\lambda_4 = \frac{x'_3 + x'_5}{2} \quad (3.33)$$

for use in the data identification process. To decide if S_R is supporting a logical zero the distance measure $|x'_2 - \lambda_2|$ is tested against a system threshold Γ_1 . Thus, if

$$D_2 = |x'_2 - \lambda_2| > \Gamma_1, \text{ data of logical 0} \quad (3.34)$$

the luminance level x'_2 of the second pel is considered to have been scrambled and hence it must be supporting a logical 0. The luminance of the second pel is then unscrambled,

$$X_2 = x'_2 \ominus 128 \bmod 256$$

i.e., as in eq.(3.18), and because the dispersive properties of the transmission channel may inflict a significant error in X_2 relative to x_2 , X_2 is smoothed to give

$$\hat{X}_2 = \frac{x'_1 + X_2 + x'_3}{3} \quad (3.35)$$

If data is detected in x'_2 the search for data is stopped and the system proceeds on to the second block of five pels. However, if no data was detected in the second pel the fourth pel is inspected by performing the test :

$$D_4 = |x'_4 - \lambda_4| > \Gamma_2, \text{ data of logical 1} \quad (3.36)$$

which identifies the presence of a logical 1 as x'_4 is significantly different from the average value of x'_3 and x'_5 . Descrambling of x'_4 ensues :

$$X_4 = x'_4 \ominus 128 \bmod 256 \quad (3.37)$$

and as in the case of X_2 , X_4 is smoothed to

$$\hat{X}_4 = \frac{x'_3 + X_4 + x'_5}{3} \quad (3.38)$$

Should the second test of Inequality (3.36) fail to identify the presence of data it is concluded that no data was transmitted and the system proceeds to process the second block of five pels. This data extraction procedure continues until every block in the image has been processed.

3.2.2.2 Post Filtering For System 2

If a block contains a data bit of logical 1 by means of a scrambled fourth pel, and the subsequent block contains a data bit of logical 0 with its second pel scrambled, then the pel at the end of the first block may be received with a significant error. This error derives

from the dispersive properties of the transmission channel. Consequently, when the data in successive blocks is a logical 1 followed by a logical 0 the luminance value of the pel at the end of the first block is replaced by the luminance formulated by averaging the luminance values of the two pels at the end of the first block and the first pel in the subsequent block. As an example, if the first block and second blocks in the frame contain a logical 1 and logical 0, respectively, x'_5 is modified to

$$x_5 = \frac{\hat{x}_4 + x'_5 + x'_6}{3} \quad (3.39)$$

3.2.3. System 3

This is a constant transmission data rate system, unlike Systems 1 and 2, and the block size is $N=3$. Hence, from eq.(3.10) the transmission bit rate of the data embedded in a television signal is

$$TBR = f_F M_1 M_2 / 3 \quad (3.40)$$

Prior to embedding data, the centre pels in the blocks are replaced by the average luminance of the block. Thus for the first block, the luminance x_2 of the second pel is changed to $\hat{x}_2$ in accordance with eq.(3.3). The output sequence applied to the interpolating filter becomes

$$S_T = \begin{cases} x_1, \hat{x}_2, x_3 & ; \text{data of logical 0} \\ x_1, \tilde{x}_2, x_3 & ; \text{data of logical 1} \end{cases} \quad (3.41)$$

where

$$\tilde{x}_2 = \hat{x}_2 \oplus 128 \text{ mod } 256 \quad (3.42)$$

3.2.3.1. Receiver For System 3

Let us commence by considering the sequence of pels in the first block of the received frame. The luminance levels of the three pels are x'_1, x'_2 and x'_3 , and will again form λ_2 , see eq.(3.12). If

$$|x'_2 - \lambda_2| > \Omega ; \text{ data of logical 1} \quad (3.43)$$

the data is considered to be a logical 1 as x'_2 must have been scrambled by modulo addition at the transmitter for Inequality (3.43) to apply. Should Inequality (3.43) be invalid it is known that the luminance level of the second pel was smoothed prior to transmission. Hence a logical 0 is deemed present. The regenerated luminance level of x'_2 is

$$x_2 = \begin{cases} (x'_1 + x'_2 \oplus 128 \text{ mod } 256 + x'_3)/3; & \text{data of logical 1} \\ x'_2 & ; \text{data of logical 0} \end{cases} \quad (3.44)$$

3.2.3.2. Post Filtering For System 3

The procedure for this System is the same as that for System 1, see section 3.2.1.2, except that the smoothing is applied to the last pel in all the blocks in the image.

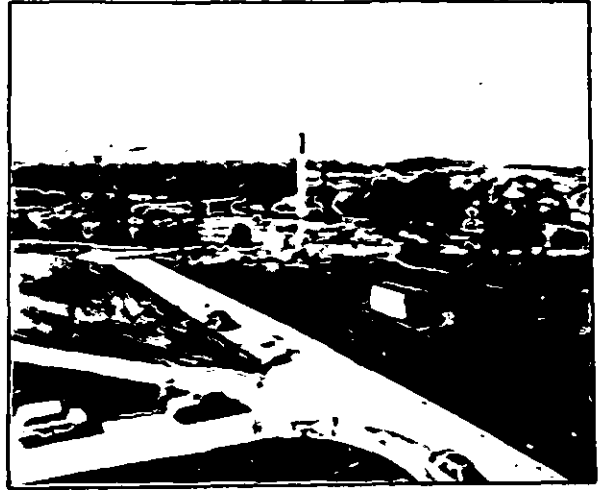
3.3. Experimental Procedure

The experiments were performed by means of computer simulation on six monochrome pictures, each bandlimited to 2.5 MHz, sampled at 5.5 MHz, and composed of 256 lines with 256 pels/line. The picture elements were linearly quantized to one of 256 luminance levels between 0 and 255. Figure 3.1 displays the six pictures, where (a) is the well-known Girl image, and two further head and shoulder images are presented in (e) and (f). A portion of a test card, an apartment building scene, and a semi-rural setting are shown in (b), (c) and (d), respectively. Figure 3.2 shows the luminance level histograms of the six pictures displayed in figure 3.1. Each head and shoulder picture exhibits a substantial peak, but the largest peak in the histograms occurs for the test-card image. The apartment building picture has a relatively flat histogram, except at the higher luminance levels, while the semi-rural scene has a double hump characteristic. The six pictures have luminance values which span the entire 256 quantized levels. Thus the images in figure 3.1 have different statistical properties, and for Systems 1 and 2 it was found that they support different amounts of embedded data.

The systems proposed here are suitable for industrial applications, and not broadcast quality television. Selecting a channel for the experiments was therefore difficult as the variety of industrial applications with their attendant channels are legion. Picture signals tend to be conveyed over coaxial quality channels and therefore it was decided to opt for a mildly dispersive channel in



(a)



(d)



(b)



(e)

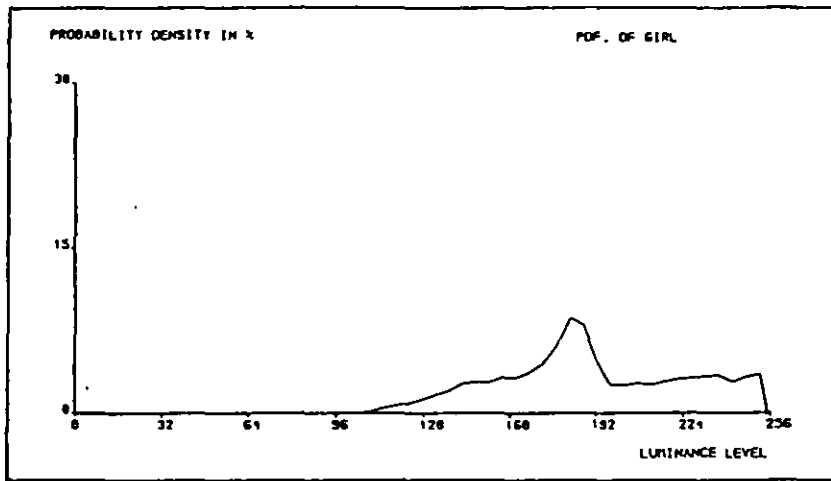


(c)

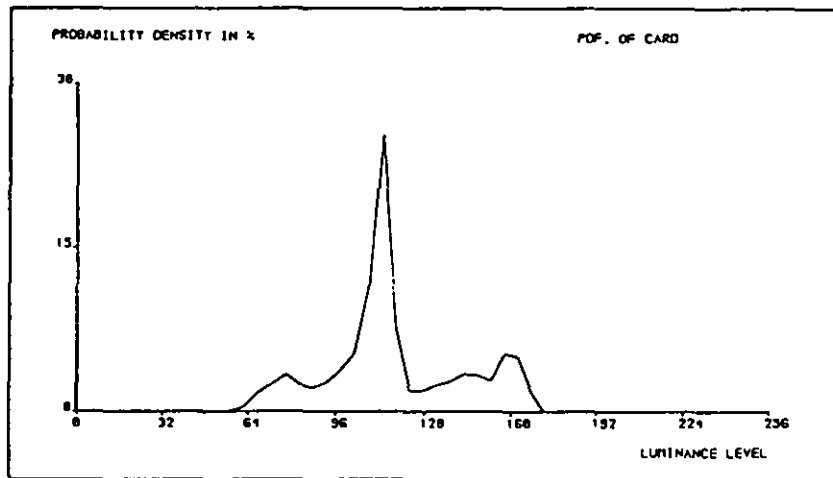


(f)

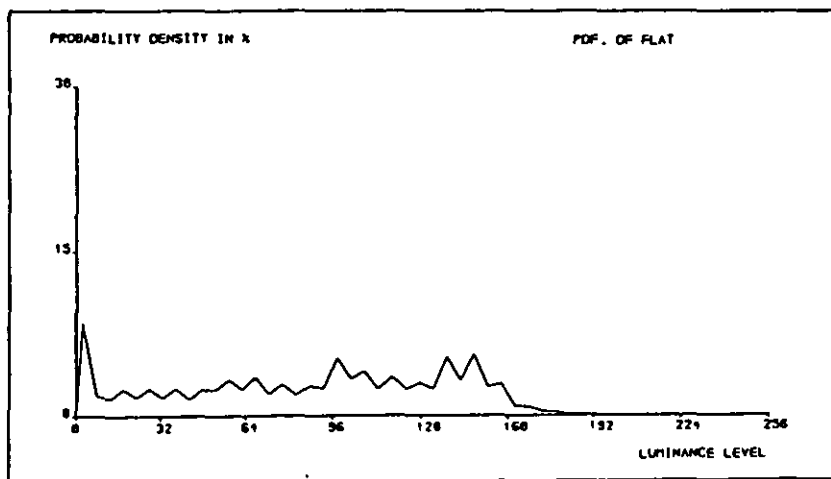
Fig.3.1. Images used to convey data (a) GIRL (b) CARD
(c) FLAT (d) SCENE (e) GIRL 2 (f) GIRL 3



(a)

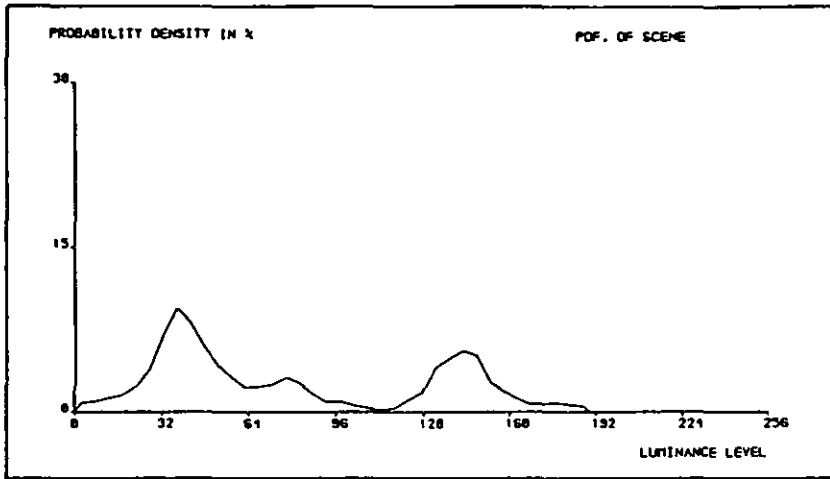


(b)

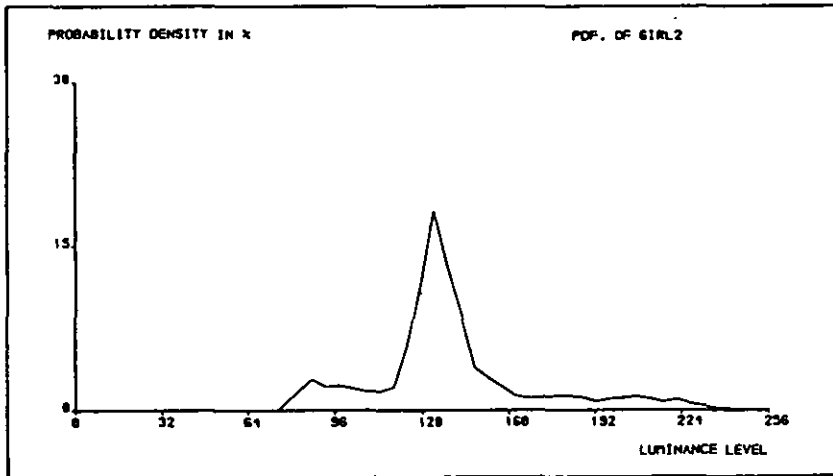


(c)

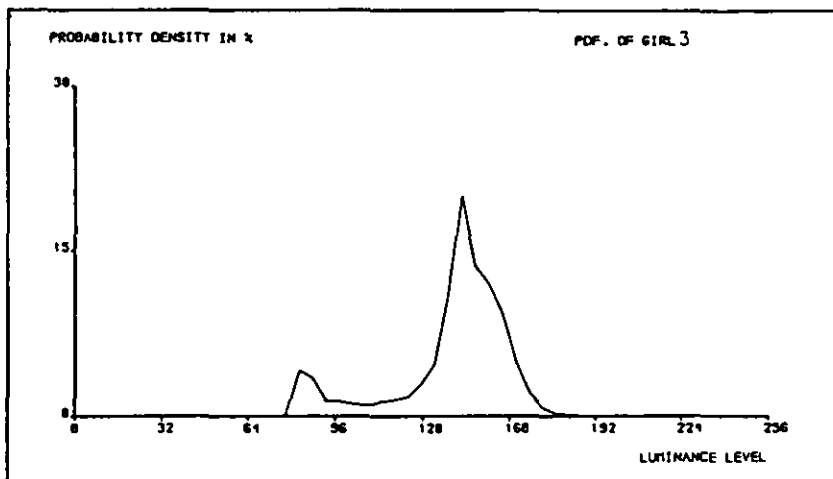
Fig.3.2. Histograms for the images in Fig.3.1
(a)-(c) are the PDF's for the images
in Fig.3.1 (a)-(c)



(d)



(e)



(f)

Fig.3.2. Histograms for the images in Fig.3.1
(d)-(f) are the PDF's for the images
in Fig.3.1 (d)-(f)

the form of a fourth order lowpass Butterworth filter having a cut-off frequency of 2.5 MHz, to which Gaussian white noise was added. The attenuation and phase characteristics of the channel filter are displayed in figure 3.3. The combined data and video sequence S_T at the transmitter was conveyed directly to the Butterworth filter there being no point in using an interpolating filter that offered little amplitude and phase distortion to S_T if it was to be followed by this Butterworth filter with its significantly worse frequency response. Alternatively, the filter of figure 3.3 may be viewed as the combined interpolating-channel filter. The channel SNR (SNR_C) was found by replacing the dispersive channel by an ideal channel, and then measuring the ratio of the mean square value of the embedded data picture signal to that of the noise signal. Although experiments for SNR_C of 30 dB were performed the picture quality was unacceptable, and an SNR_C of 40 dB was chosen, a value significantly below that encountered in broadcast quality channels. Thus, unless otherwise stated, SNR_C is assumed to be 40 dB.

The synchronization of the transmitter and receiver clocks may be performed using the line synchronization pulses. However, should the accuracy of the clock synchronization be inadequate, timing pulses may be introduced during the line synchronization periods that are precisely locked to the generation of pels at the transmitter. This procedure will enable the receiver to sample to the combined video and data signal to an acceptable accuracy. In the simulations the sampling

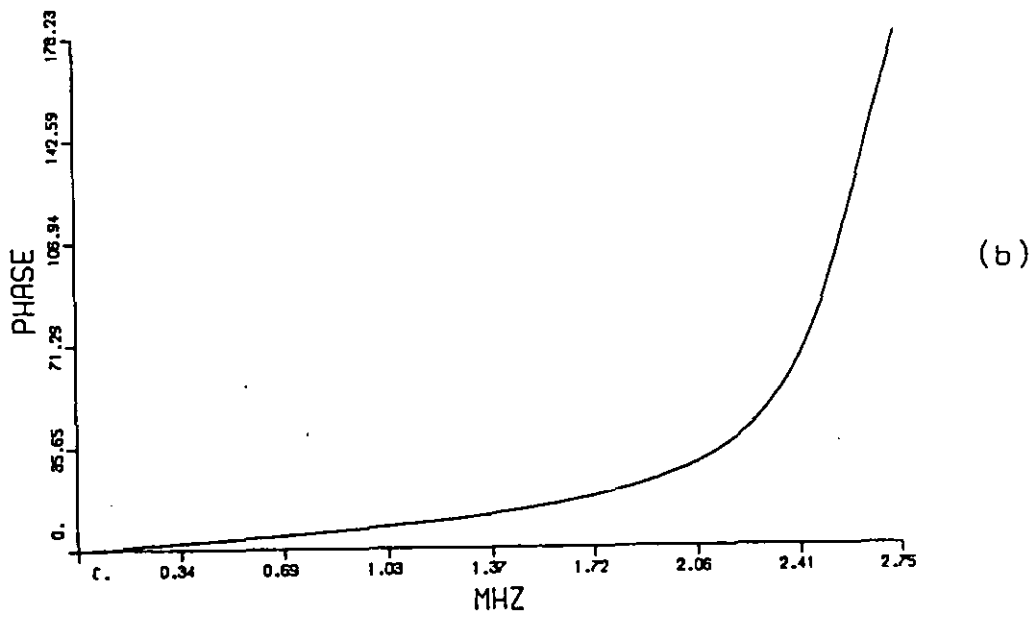
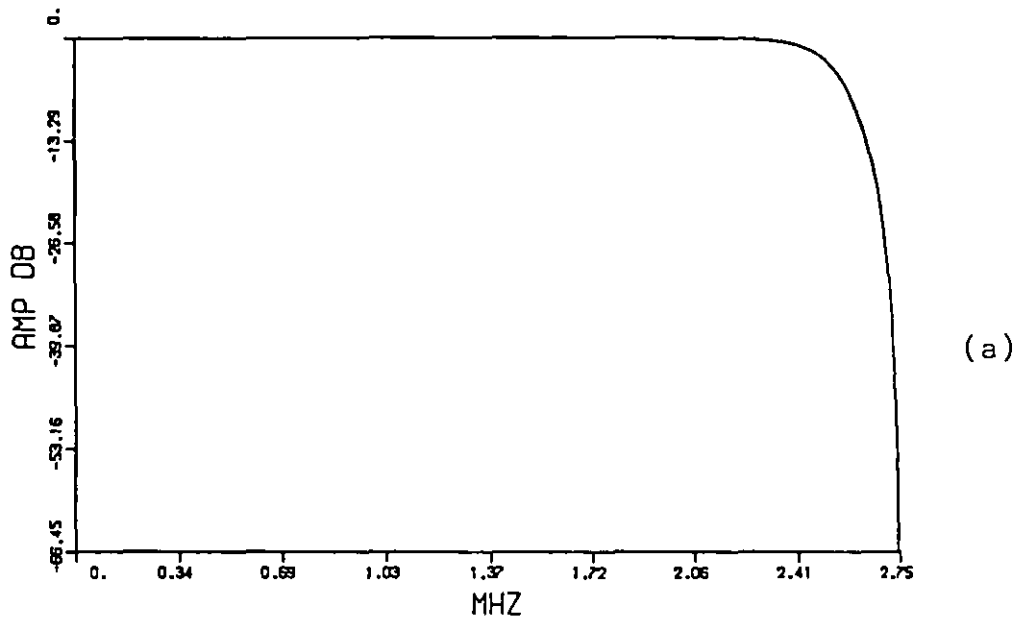


Fig.3.3. Characteristic of channel filter
(a) attenuation (b) phase

instants at the receiver has an exact relative correspondance to those at the transmitter. Although it was not implemented in the data embedding systems, it must be emphasised that the simulation procedures were those which can be conveniently executed by a microprocessor.

3.4. Results

3.4.1. System 1

This system has three system parameters; an activity threshold $T_{a,1}$, and two distance thresholds $T_{d,1}$ and $T_{d,2}$. These parameters are displayed in Table 3.1. In selecting the distance thresholds the histograms were computed for the distance measure D for the case when no data was transmitted, when data was a logical 0 and when the data was a logical 1. The shape and position of these PDFs was found to be dependent on the choice of the activity threshold $T_{a,1}$, and on the channel. For the dispersive channel with its additive noise experiments were conducted with $T_{a,1}$ to ensure that these three histograms did not overlap. By separating the histograms data errors could not occur. The activity threshold $T_{a,1}$ also determined the amount of data that could be embedded in the picture, lower $T_{a,1}$ and the data rate decreased, but if $T_{a,1}$ became too high data errors ensued.

Figure 3.4 shows the histograms measured at the receiver for the apartment building picture of figure 3.1(c). The histograms for this picture are presented because it was the most demanding of the six images shown in figure 3.1, producing the smallest separations between the histograms for no data transmitted, data logical 0, and data logical 1.

System	Transmitter	Receiver	
1	$T_{a,1} = 50$	$T_{d,1} = 43$	$T_{d,2} = 99$
2	$T_{a,2} = 50$	$\Gamma_1 = 100$	$\Gamma_2 = 100$
3		$\Omega = 57$	

Table 3.1 System Parameters

Image	Bits Per Frame		
	System 1	System 2	System 3
GIRL	17025	9862	21760
CARD	14745	8260	21760
FLAT	14435	7780	21760
SCENE	18948	11121	21760
GIRL 2	19564	11543	21760
GIRL 3	19866	11703	21760

Table 3.2 Data Rates Comparisons for the Three Systems

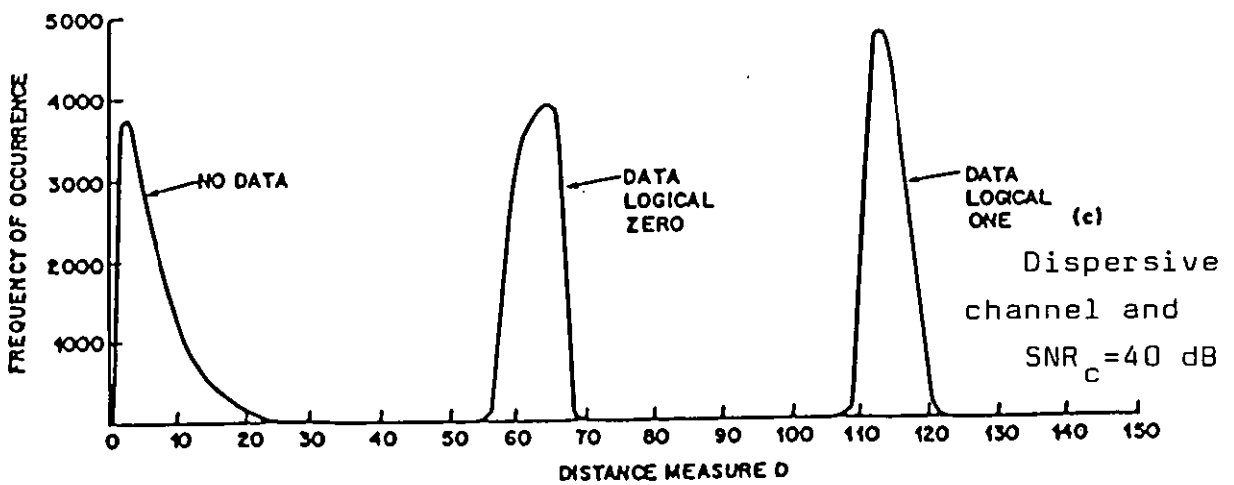
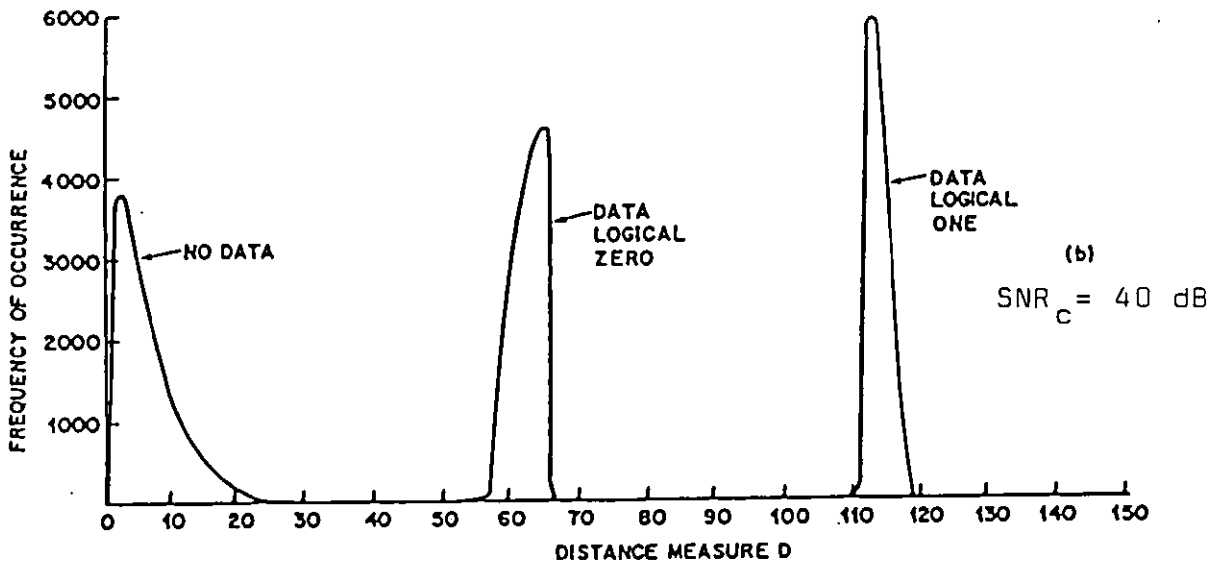
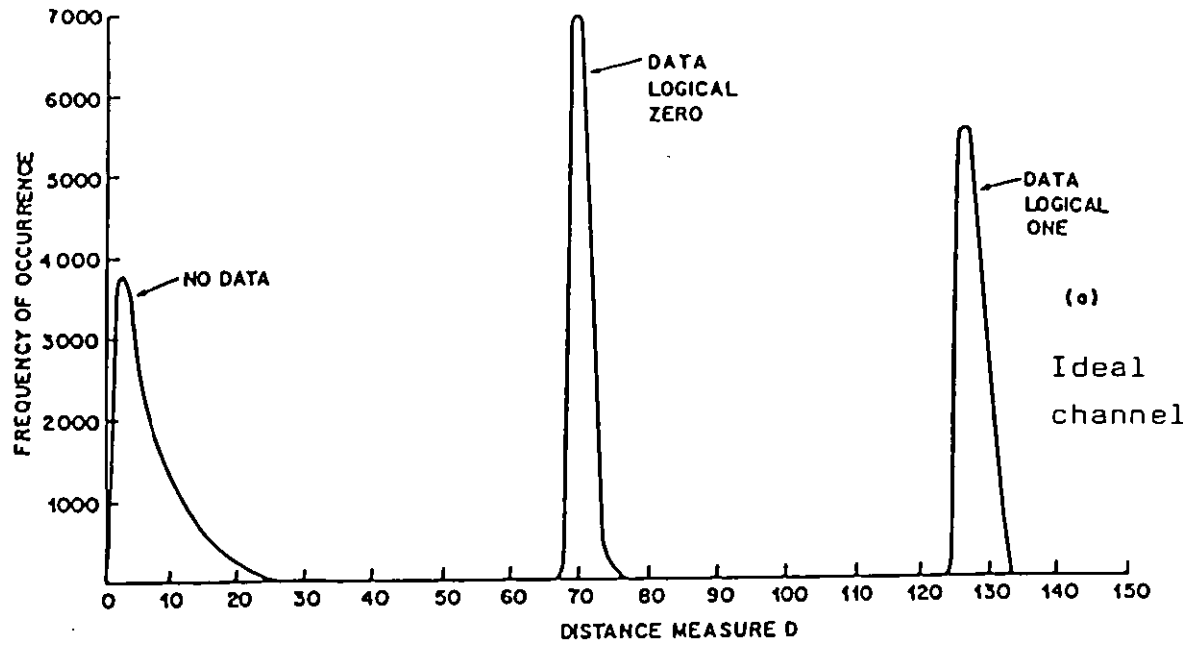


Fig.3.4. Histograms of the luminance distance measure for the picture in Fig.3.1c

In figure 3.4(a) ideal channel conditions prevailed. When the channel had a SNR of 40 dB the histograms for the data moved to lower values of D as shown in figure 3.4(b), and when the channel was made dispersive the histograms broaden, figure 3.4(c). The distance thresholds were selected using figure 3.4(c) and similar histograms relating to the other five pictures. The values of the thresholds are shown in Table 3.1. No additional histograms for other values of $T_{a,1}$ were shown; rather it is stated that this value of $T_{a,1}$, although sub-optimum, is a satisfactory value for the six pictures used in the experiments. By this choice of $T_{a,1}$ it was ensured that the bit-error-rate (BER) would be zero for the six pictures of figure 3.1. With different pictures a non-zero BER may be expected.

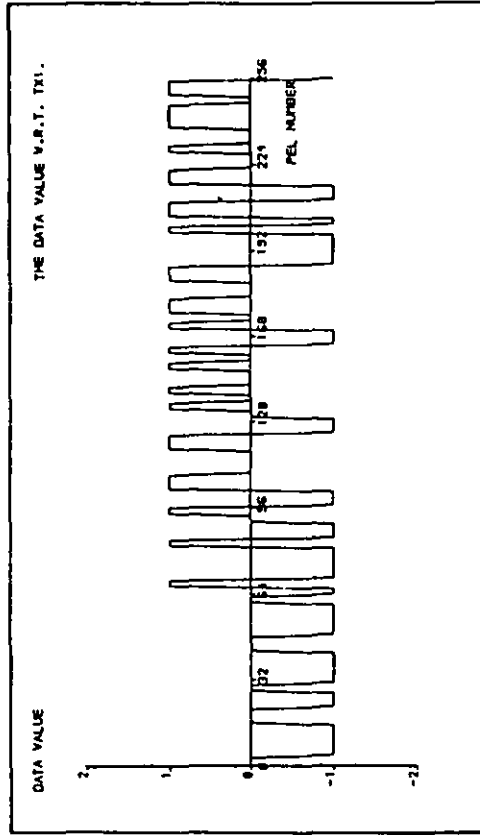
Performance Of System 1

The maximum number of bits that could be embedded in a scan-line for a block size of $N=3$ was 85. As each image has 256 scan-lines the maximum data that an image could support was 21760 bits. The amounts of data embedded in each picture, that were conveyed over the dispersive channel having an SNR_c of 40 dB, and extracted without a single bit error, are displayed in Table 3.2. From these results it was observed that the bit rate efficiency

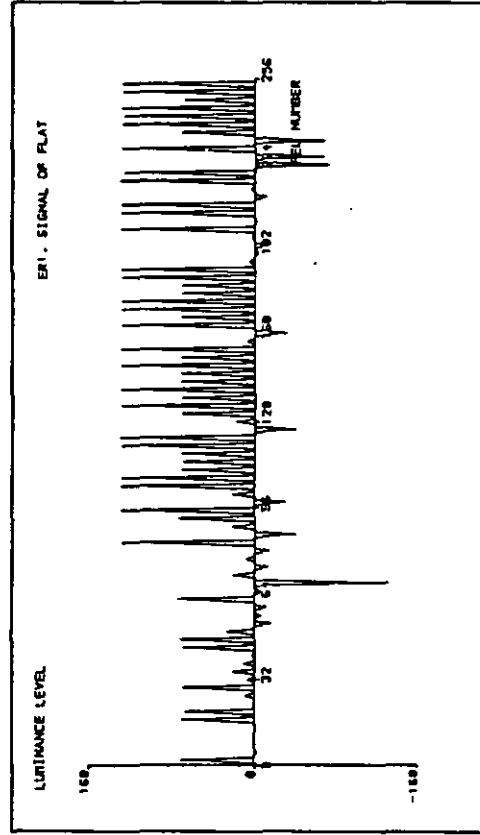
$$\eta = \frac{\text{actual data/frame}}{\text{possible data/frame}} \quad (3.45)$$

was as high as 91.3% for the head and shoulder 2 picture, and the lowest η was 66.3% for the apartment building scene.

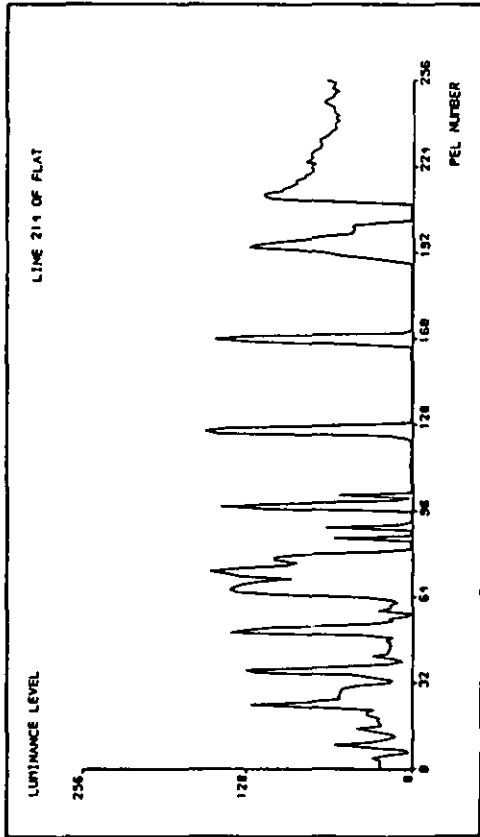
The luminance waveform of the 214th scan-line in the 'FLAT' image is displayed in figure 3.5(a). This scan-line was selected because the rapidly changing luminance levels of the FLAT picture provided a highly active region. The waveform of the data embedded into the luminance signal is shown in figure 3.5(b), where +1, 0 and -1 levels signify when data of logical 1 or when data of logical 0 were embedded, and when no data were embedded, respectively. The effect of embedding the data into the relatively inactive regions of the line waveform was to cause them to exhibit rapid luminance changes. The data embedded signal is displayed in figure 3.5(c). Highly active blocks of pels that were unable to support data were rendered less active by smoothing the pels in the centre of the blocks in accordance with eq.(3.3). Figure 3.5(d) shows the difference between the input scan-line and the transmitted signal, where the effects of modulo addition and luminance smoothing are evident. After transmission via the dispersive noisy channel, data extraction (without error) and post filtering at the receiver, the recovered luminance waveform appeared as displayed in figure 3.5(e). Because the error in the recovered luminance signal was relatively small shown in figures 3.5(f) and (g) are the overall luminance error signal with and without post filtering, respectively. It was found that post filtering made a significant contribution to the reduction of luminance errors. These errors are the penalty incurred for using the video signal as a data carrier over an imperfect channel.



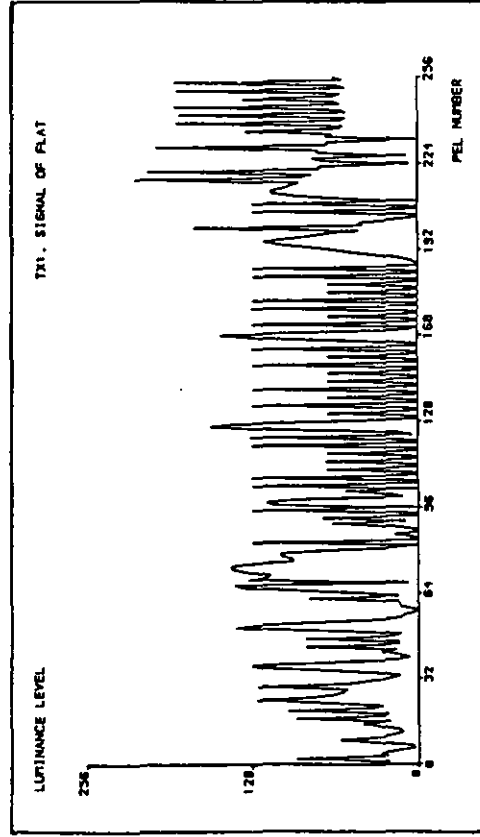
(b)



(d)

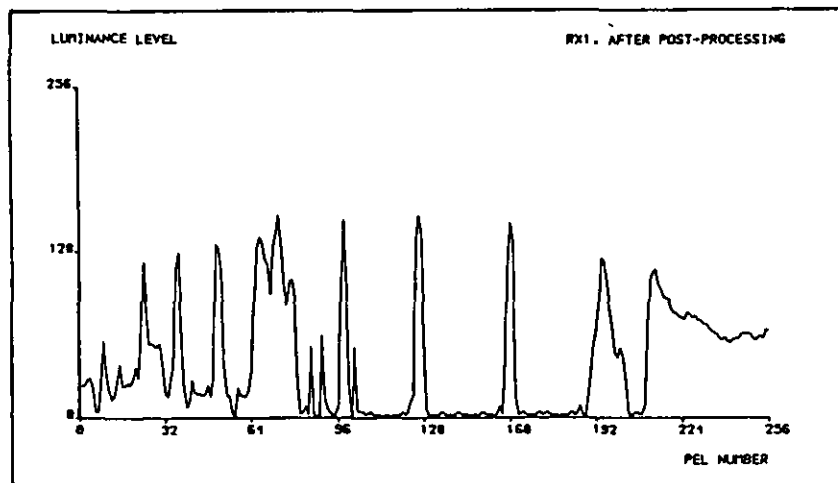


(a)

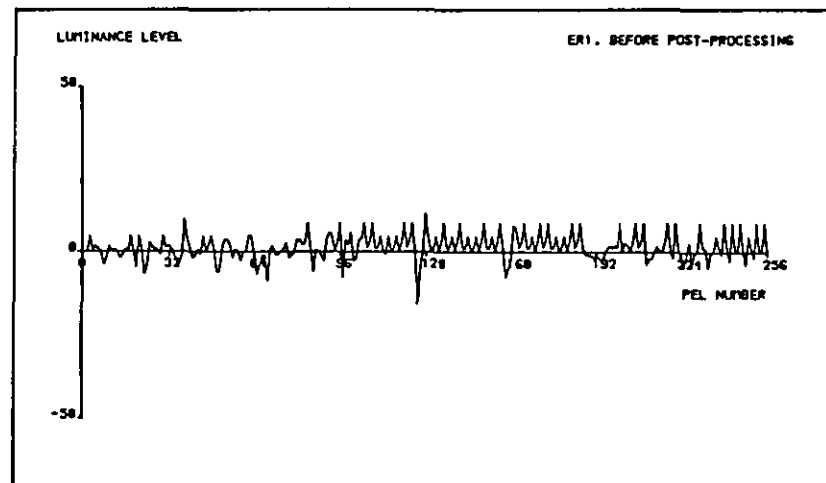


(c)

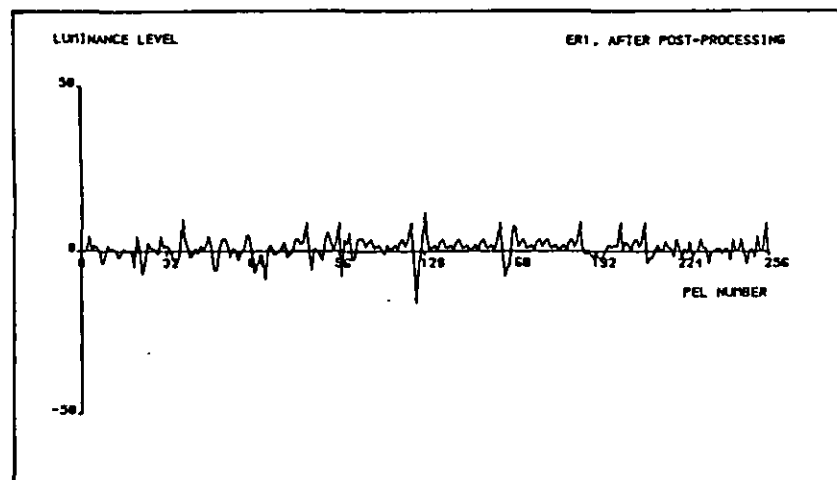
Fig.3.5. Captions overleaf



(e)



(f)



(g)

Fig.3.5. System 1: waveforms of a scan line

- (a) Original luminance waveform
- (b) Data signal
- (c) Combined luminance and data signal
- (d) Difference between (a) and (c)
- (e) Luminance waveform after data extraction and post filtering
- (f) Error in the recovered luminance signal without post filtering
- (g) Error in the recovered luminance signal when post filtering is used

Photographs of only 3 recovered pictures, when the data embedding scheme of System 1 was employed, are displayed fig. 3.6, because all the recovered pictures show no discernable differences from photographs of the original pictures that have been conveyed directly over the channel. However, small degradations in the recovered images due to the data embedding process were observed on the video monitor.

3.4.2. System 2

It will be recalled that System 2 uses five pels to support one bit of data, embedding the data by modulo 256 addition of 128 with the luminance level of the second or fourth pel in the block according to whether the data is a logical 0 or logical 1, respectively. However, if the activity of the five pels in the block exceeds threshold $T_{a,2}$ no data is inserted into the pels. At the receiver data is detected by comparing distance measures against system thresholds Γ_1 and Γ_2 . Inequalities (3.34) and (3.36) are used in the test procedure for the first block of pels in the image.

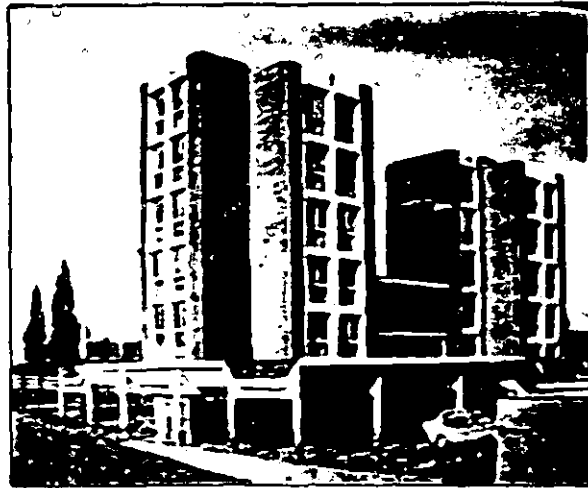
In figure 3.7 histograms of the luminance distance measure are shown, i.e., the difference between the luminance value of the second or fourth pel compared to the luminance average between the first and third pels, or the third and fifth pels, respectively, in blocks containing $n=5$ pels. These distance measures will be referred to as D_2 and D_4 , the sub-script relating to the pel number in a block. Thus in figure 3.7(a) and (b) histograms are displayed D_2 and D_4 , respectively, measured for all non-supporting



(a)



(b)



(c)

Fig.3.6. Recovered images from System 1 after being conveyed through the dispersive channel and $\text{SNR}_c = 40 \text{ dB}$ (a) GIRL (b) CARD (c) FLAT

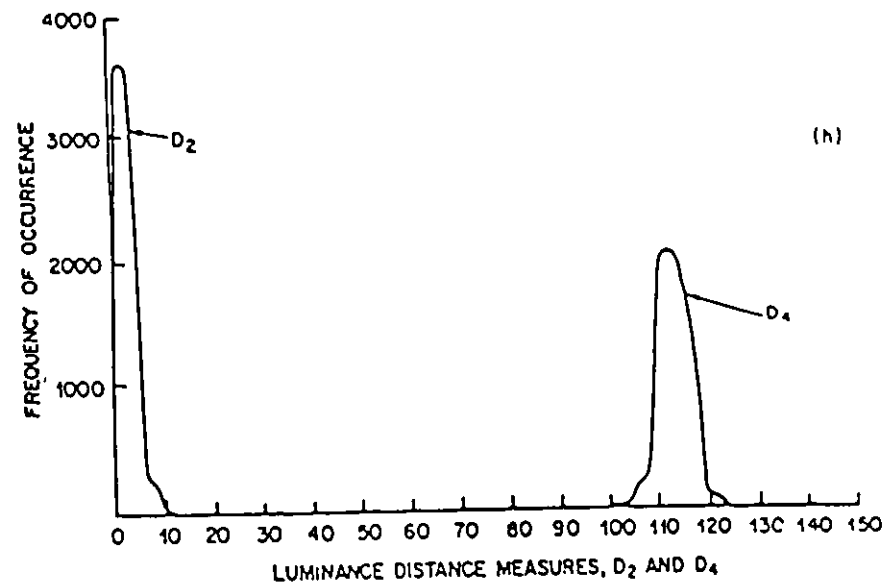
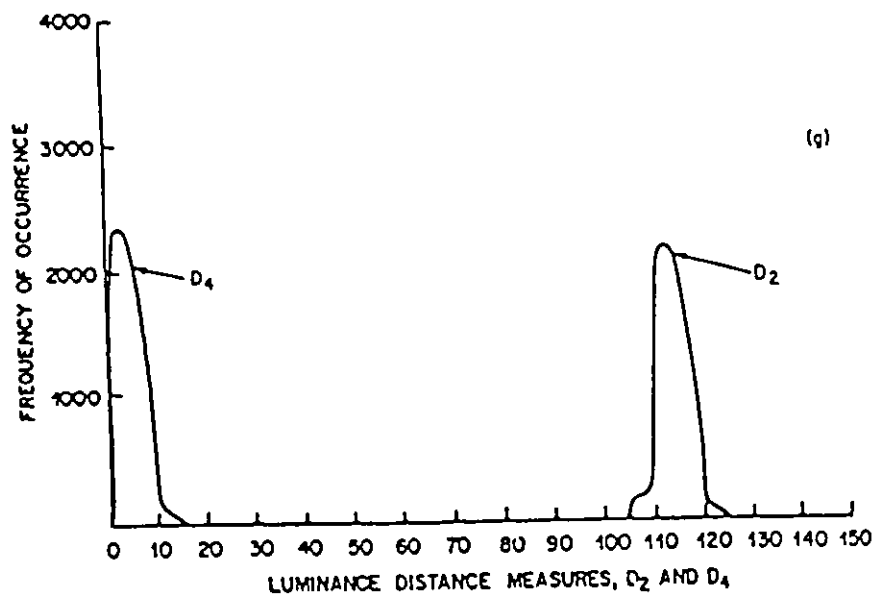
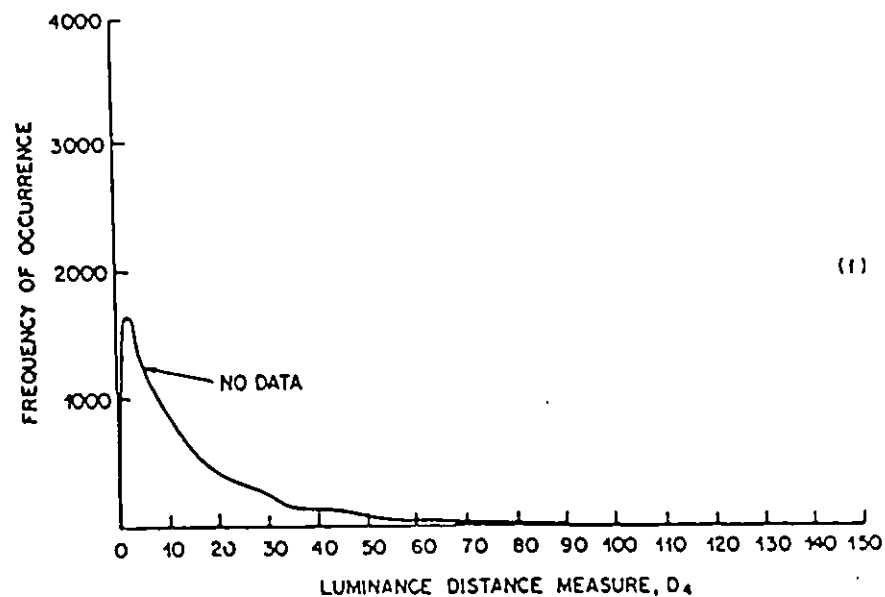
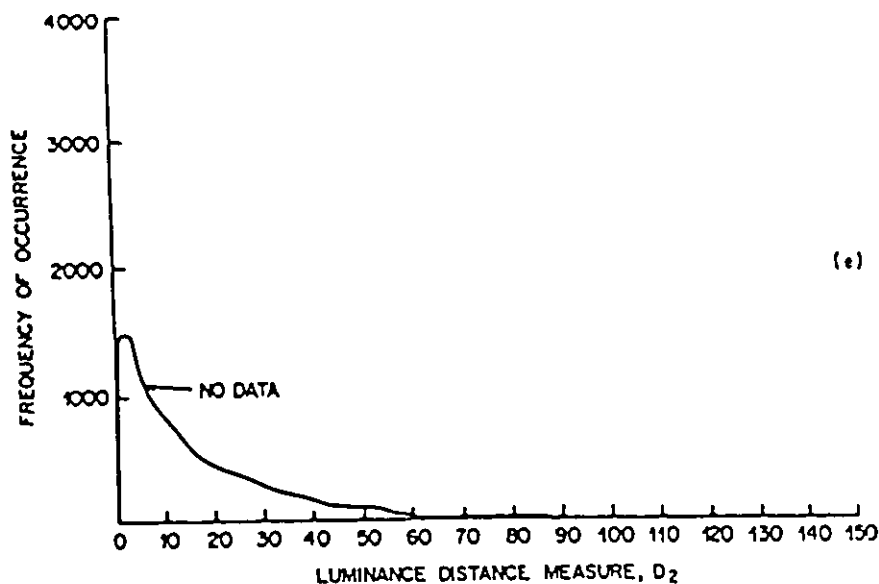


Fig.3.7. Histograms of the luminance distance measure for the picture in Fig.3.1c

(e)-(h) Imperfect channel

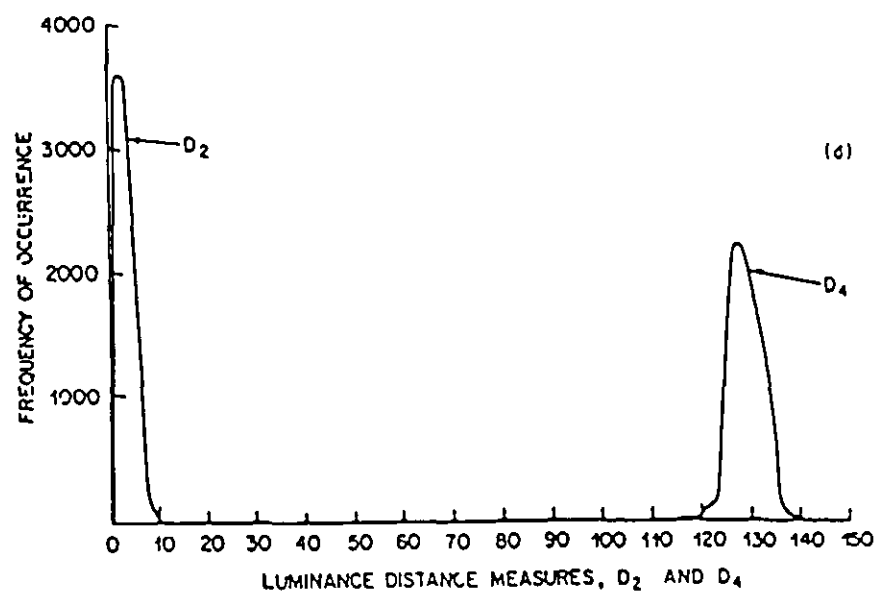
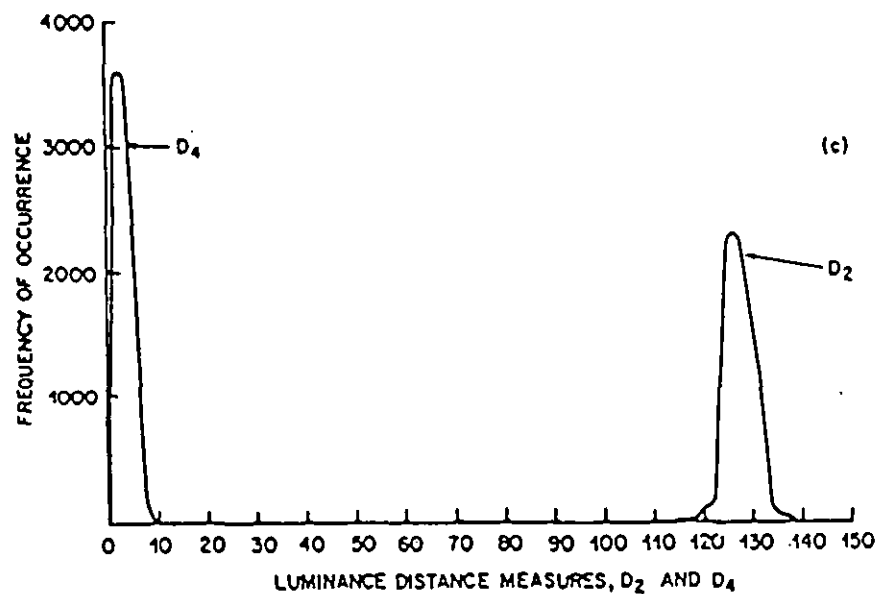
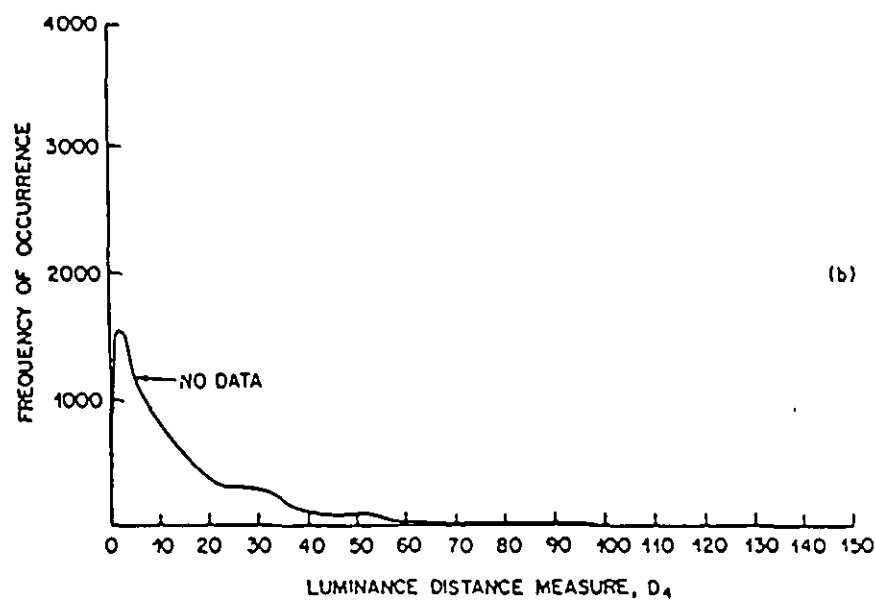
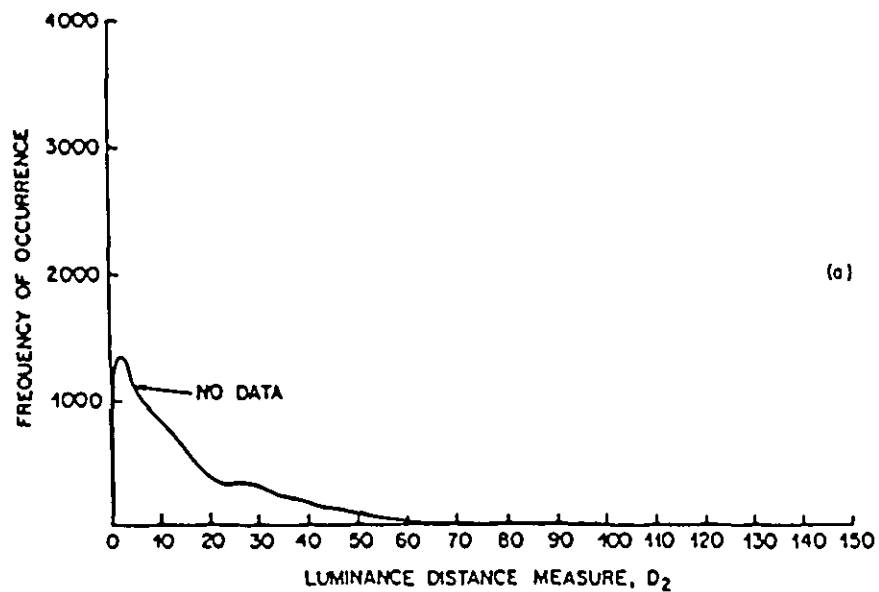


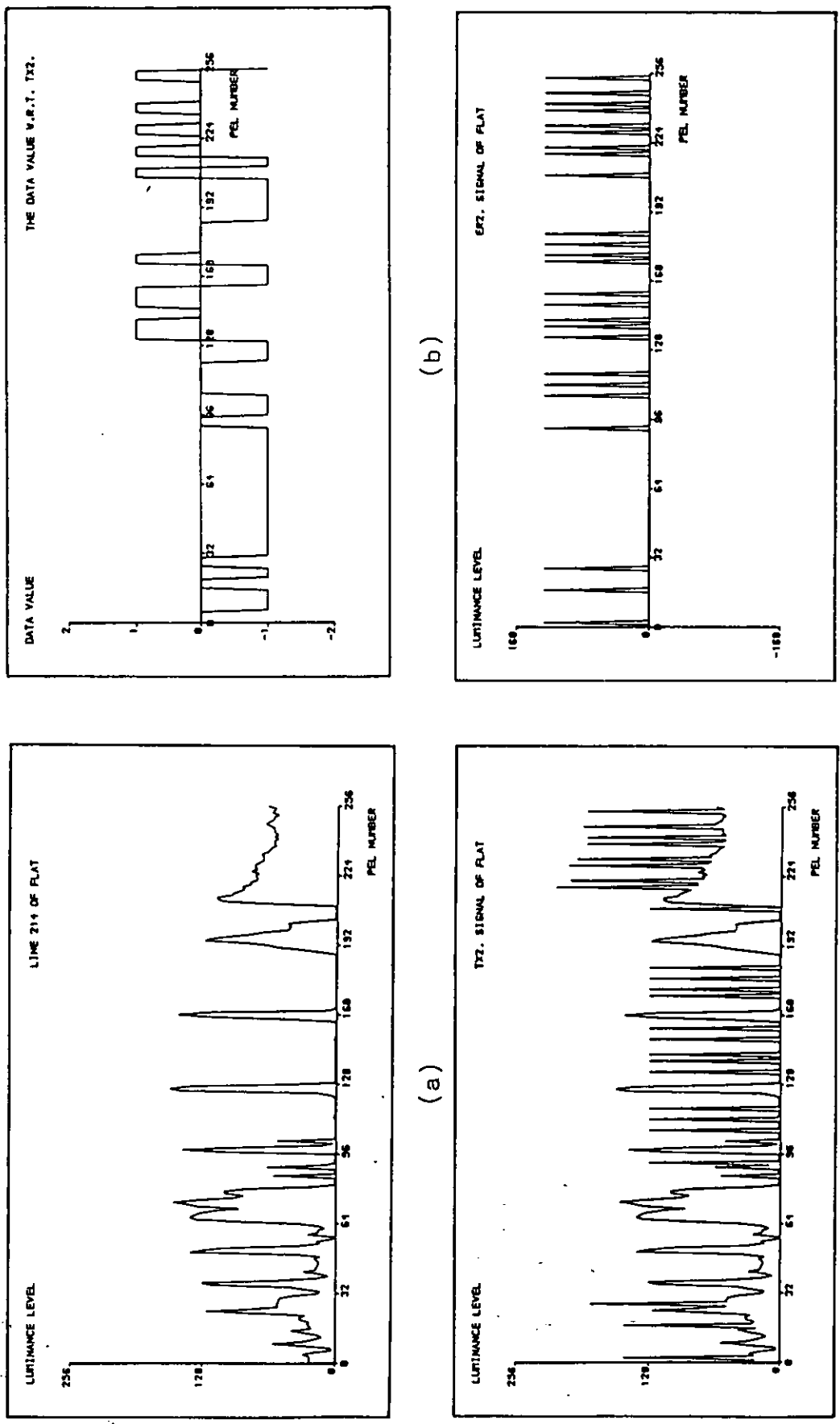
Fig.3.7. Histograms of the luminance distance measure for the picture in Fig.3.1c

(a)-(d) Ideal channel (e)-(h) overleaf

data blocks of pels. Figure 3.7(c) shows the histograms for D_2 and D_4 for those blocks supporting a logical 0. As expected the histogram for D_2 was moved to high values of D_2 due to the substantial increases in the luminance values resulting from modulo addition. The histogram for D_4 occurred for low values of D_4 as no logical 1 values were embedded in the picture. The histograms for D_2 and D_4 in figure 3.7(d) are for blocks supporting data of logical 1. Observe that the histograms in figure 3.7(c) and (d) relating to the case of data of logical 0 and logical 1, respectively have very similar shapes and positions, although the labelling of the histograms is reversed. This reversal occurred because the picture was capable of supporting as many bits of logical 0 as logical 1, the data embedded in the image being of a random nature. Figure 3.7(a)-(d) apply for an ideal channel, while 3.7(e)-(h) are the corresponding histogram for the dispersive, noisy channel. The presence of the imperfect channel is to modify the shape of the histograms, but more importantly, to reduce the spacing between the histograms of D_2 and D_4 . Changing threshold $T_{a,2}$, or processing a different picture, modifies the histograms of D_2 and D_4 . The histograms displayed in figure 3.7 are for the apartment building scene, the most difficult picture to accommodate. The distance measure threshold Γ_1 and Γ_2 were selected as follows. The histograms of figure 3.7(e) and (g) were over-laid, and a threshold Γ_1 was sought for data of logical 0 condition such that the histograms for D_2 never over-lapped. Then the

histograms of figure 3.7(f) and (h) were over-laid, and Γ_2 placed in the space between the histogram for data of logical 1 and the histogram for no data embedded, the distance measure being D_4 . This procedure was repeated for all six pictures. The values of Γ_1 and Γ_2 that prevented any bit errors occurring in the six images are listed in Table 3.1.

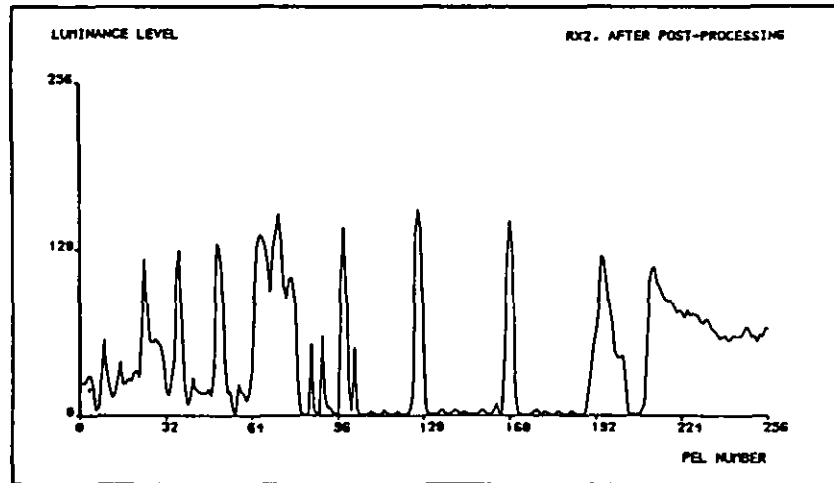
The waveforms for the same scan-line in figure 3.5(a), repeated for convenience in figure 3.8(a), are displayed in figure 3.8(b)-(e) for System 2. The waveform in figures 3.5(c) and 3.8(c) are very different because of the amount of data inserted in scheme 2 is less than in scheme 1, and the manner in which the data is inserted is dissimilar. The difference between the combined data and video signal and the original signal shown in figure 3.8(d) only reflects the effect of modulo addition, as scheme 2 does not require those blocks that are not carrying data to have their activity lowered. The recovered luminance signal is shown in figure 3.8(e), and the luminance error signals before and after post filtering are displayed in figure 3.8(f) and 3.8(g), respectively. The effect of post filtering is to reduce the error signal by an amount that is perceptually significant. Observe that the error waveform in figure 3.8(g) is marginally of lower power than the error waveform of System 1, see figure 3.5(g). The spatial frequency components in figure 3.5(g) are higher than those in figure 3.8(g) due to System 1 having the greater data rate. The recovered images are of slightly superior quality to those obtained using System 1, therefore only 3 recovered



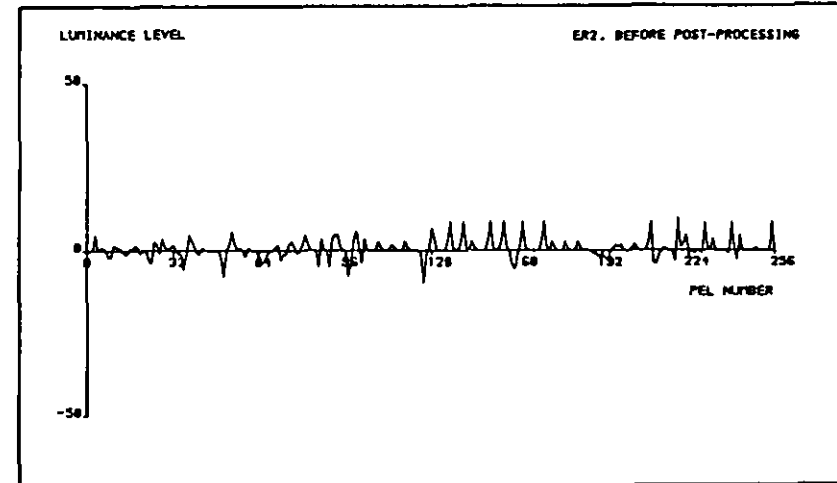
(c)

(d)

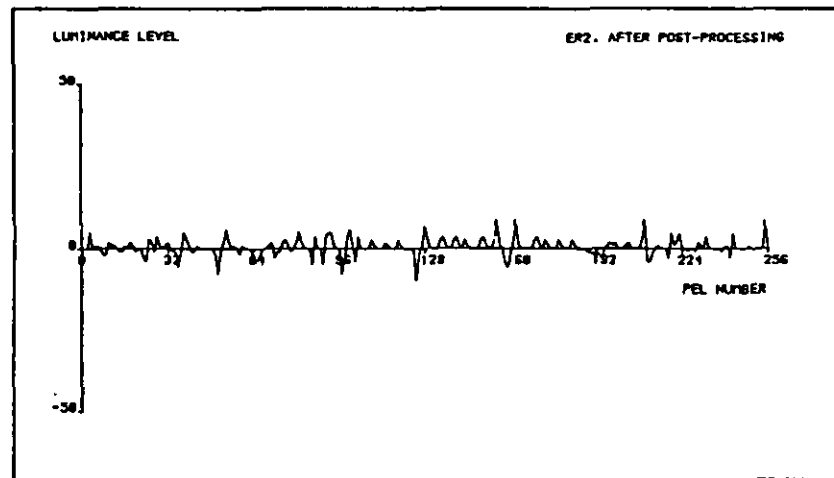
Fig.3.8 Captions overleaf



(e)



(f)



(g)

Fig.3.8. System 2: waveforms of a scan line

- (a) Original luminance waveform
- (b) Data signal
- (c) Combined luminance and data signal
- (d) Difference between (a) and (c)
- (e) Luminance waveform after data extraction and post filtering
- (f) Error in the recovered luminance signal without post filtering
- (g) Error in the recovered luminance signal when post filtering is used

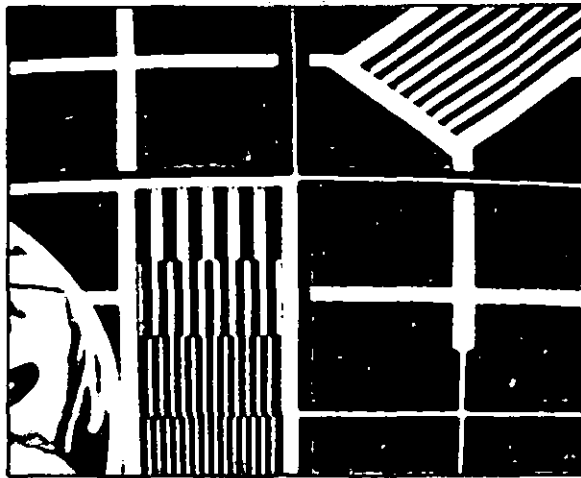
pictures are displayed, fig.3.9, because the differences are indiscernable when photography is employed. Thus, System 2 handles less data per image than System 1, as shown in Table 3.2 and recovers pictures with a marginally higher quality.

3.4.3. System 3

Irrespective of picture activity, System 3 embedded one bit in every three pels to produce a constant data rate transmission. It also provided the highest data rate of the three systems with 21760 bits per picture, see Table 3.2. System 3 is the simplest of the three systems, having no activity factor and only one distance threshold parameter Ω . Figure 3.10 shows the histograms for the distance measure D_3 , where D_3 is the magnitude of the difference between the luminance level of the pel at the centre of the block and the average luminance of the two pels at the ends of the block. The luminance of the pel at the centre of each block was replaced by the average luminance of the block, and when a logical 0 was embedded no modification to the centre pel occurred. However, if a logical 1 was transmitted the modified centre pel experienced modulo 128 addition. Hence the histograms for data of logical values 0 and 1 are spaced by a significant distance in figure 3.10, even for the worst channel condition shown in (c). Histograms were produced of the type in figure 3.10(c) for all the pictures, and found the distance mid-way between the histograms for each picture. The average of these distance measures was designated the system threshold Ω , and its value was 57.



(a)



(b)



(c)

Fig.3.9. Recovered images from System 2 after being conveyed through the dispersive channel and $\text{SNR}_c = 40 \text{ dB}$ (a) GIRL (b) CARD (c) FLAT

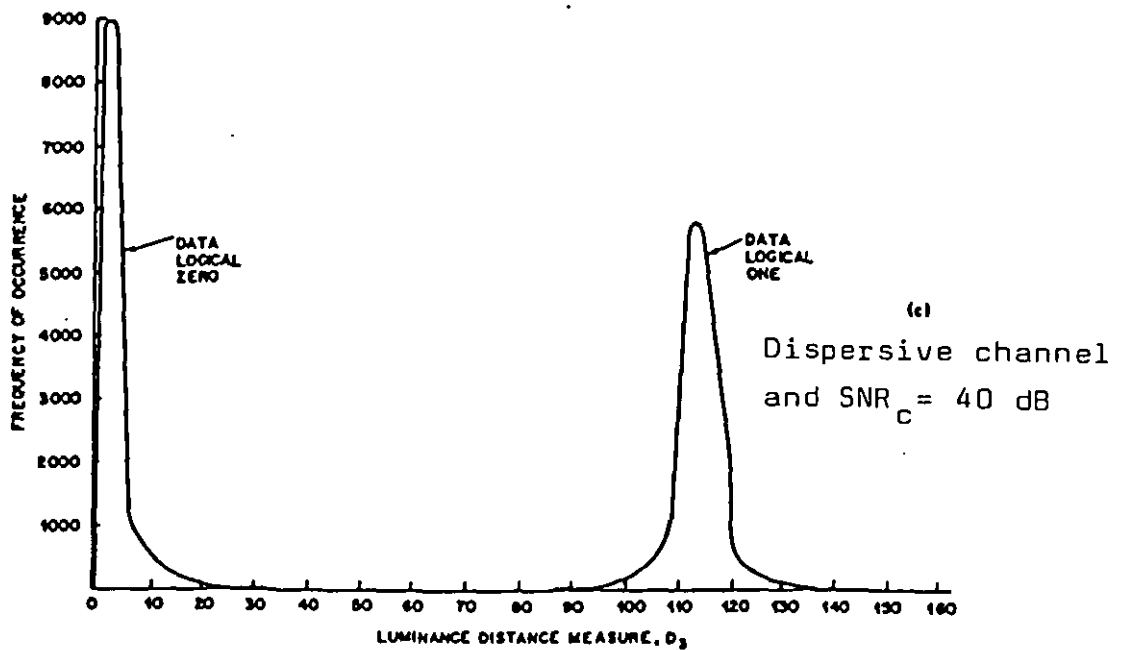
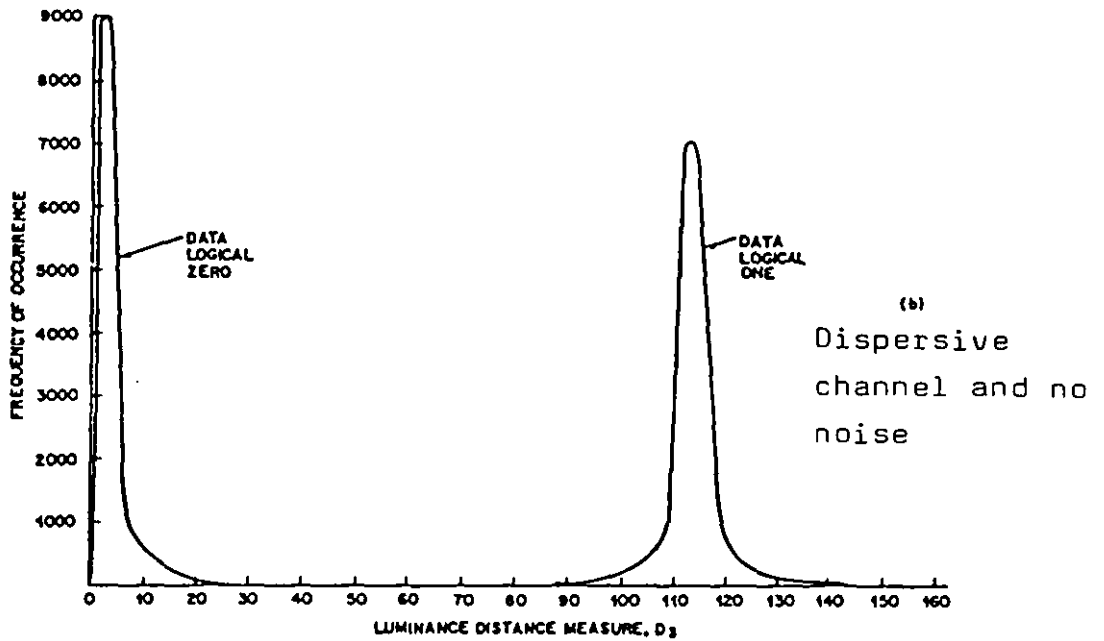
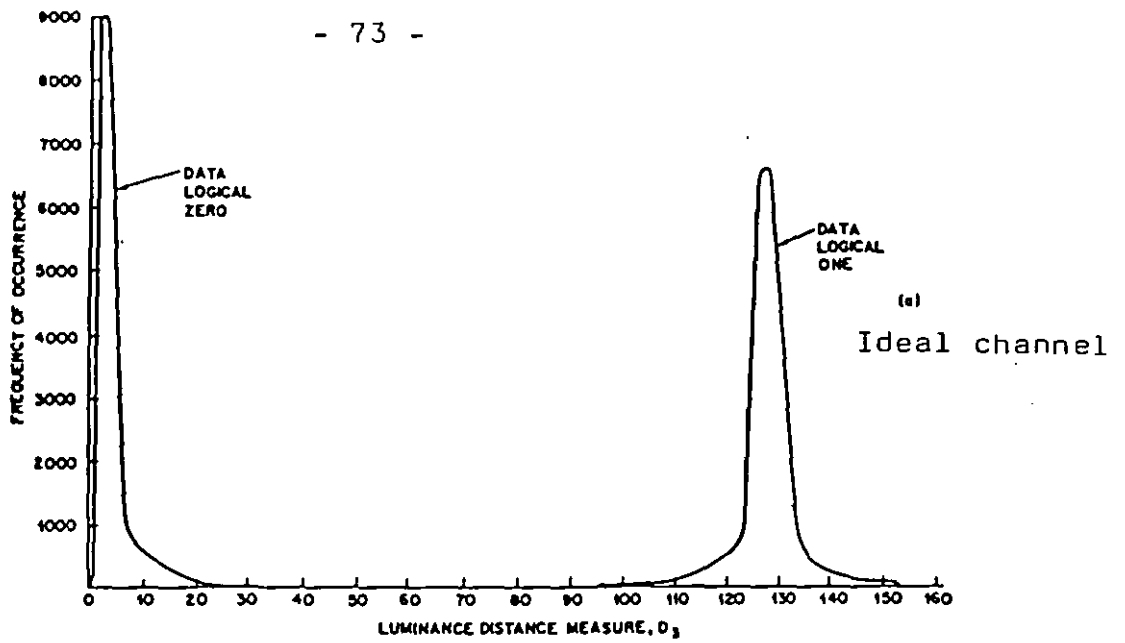


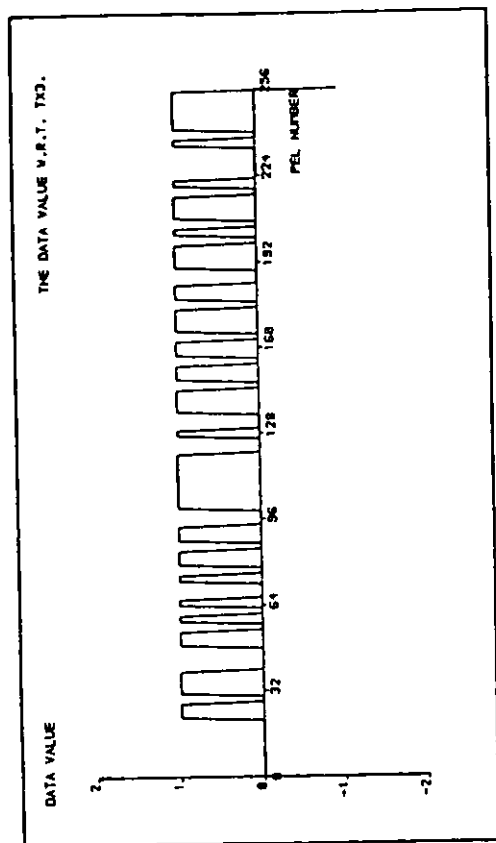
Fig.3.10. Histograms of the luminance distance measure for the image of Fig.3.1.c

The waveforms associated with System 3 for embedding data into the same scan-line as used in figure 3.5(a) are presented in figure 3.11. Observe that because data was inserted into every block, there was considerable distortion in the vicinity of the small vertical bars, as seen in figure 3.11, or by comparing figures 3.11(a) with 3.11(e). These errors are sufficiently large to cause a clearly discernable degradation as can be seen by comparing the first three images of figure 3.1 with the recovered pictures shown in figure 3.12 which have conveyed 21760 bits over the channel. It must be emphasized that of the six pictures shown in figure 3.1 and processed by System 3, only the test card image had sufficient activity for the distortion due to System 3 to be clearly visible when photographed. Inspection of the monitor revealed that the quality of all the recovered pictures was worse for System 3 than for the other two systems.

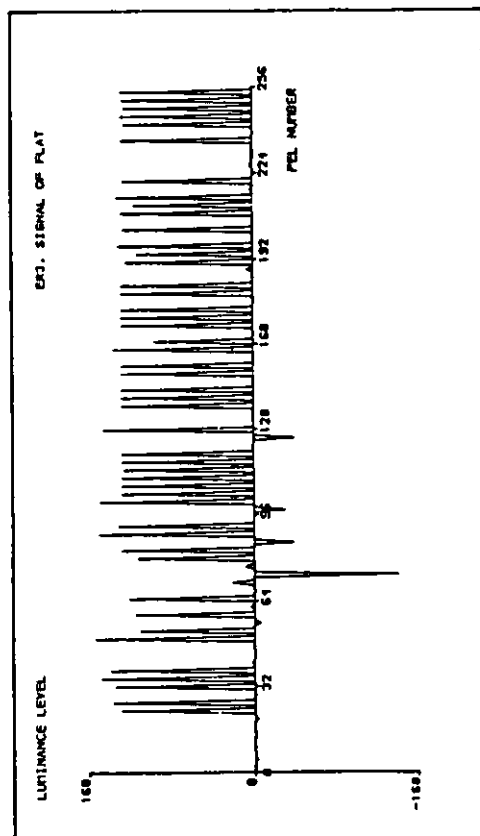
The effect of privacy provided by the scrambling algorithm of the three systems is shown in figure 3.13, where the data manifests as a noise mask of black and white dots on the original image.

3.4.4. Average Picture Signal-To-Noise Ratios

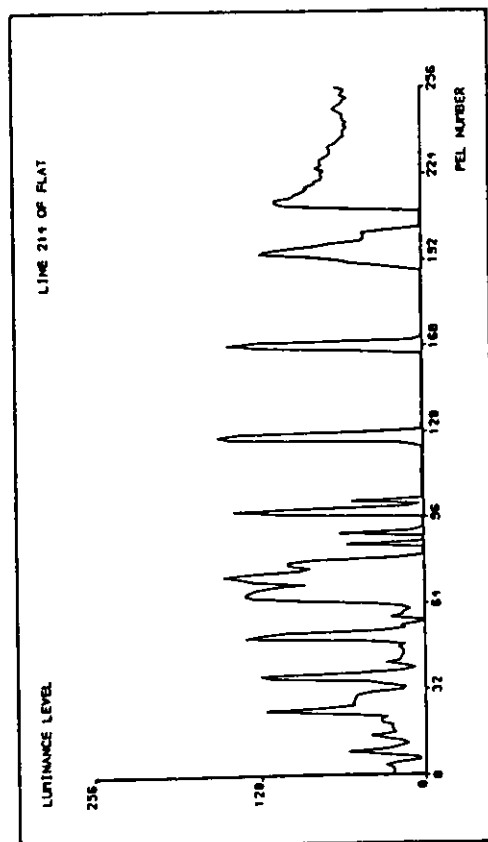
The inband noise power in the recovered pictures was computed, and thence the picture signal power to noise power ratio (SNR) in dB was found. The SNR values in dB for each picture was then averaged to give $\overline{\text{SNR}}$. When the channel was ideal, the data embedding schemes of System 1, 2 and 3 yielded $\overline{\text{SNR}}$ values of 41, 44 and 30 dB, respectively. The effect of including the dispersive channel, the second-



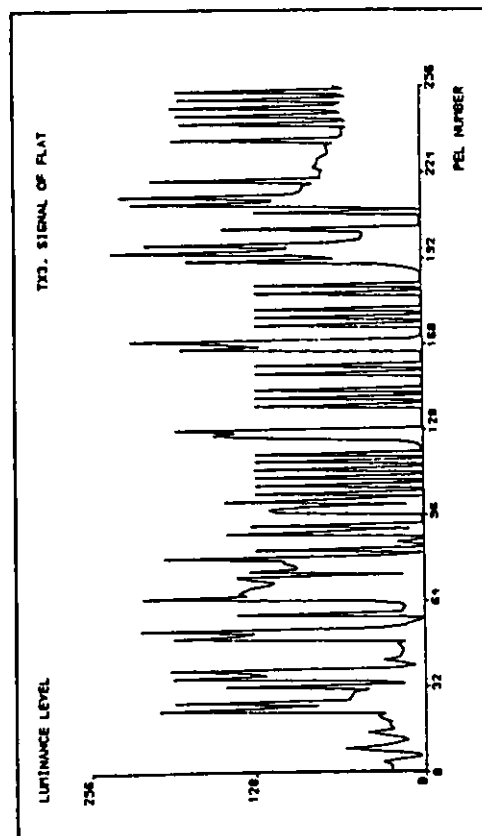
(b)



(d)

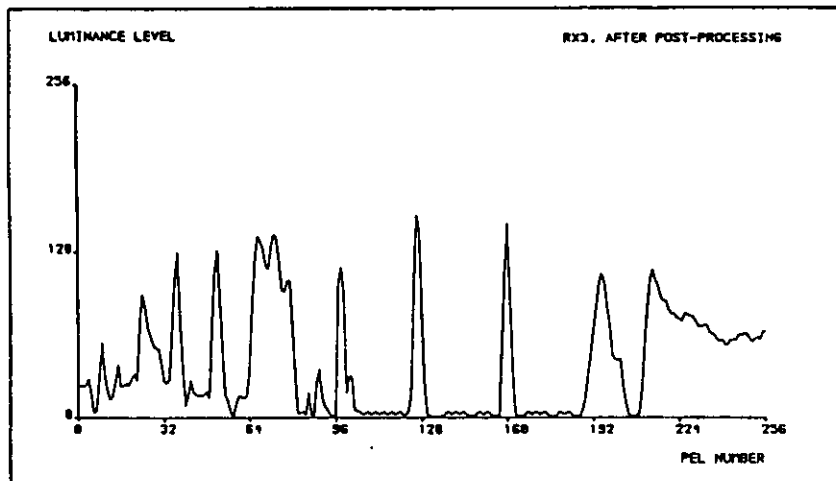


(a)

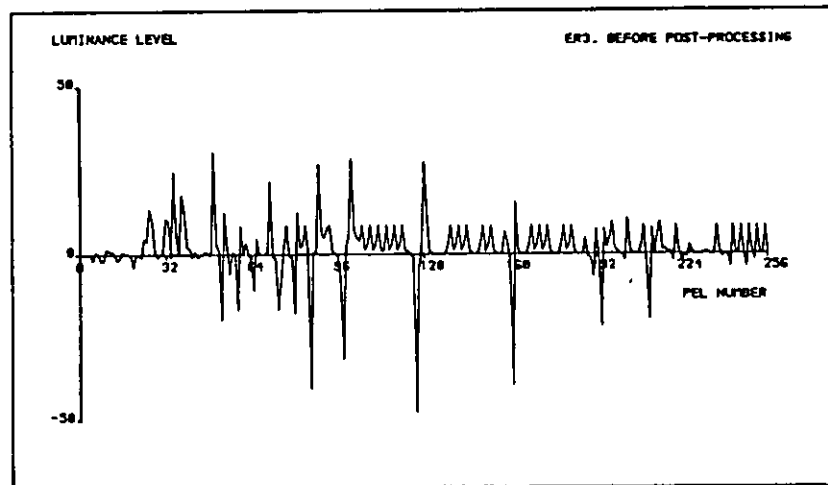


(c)

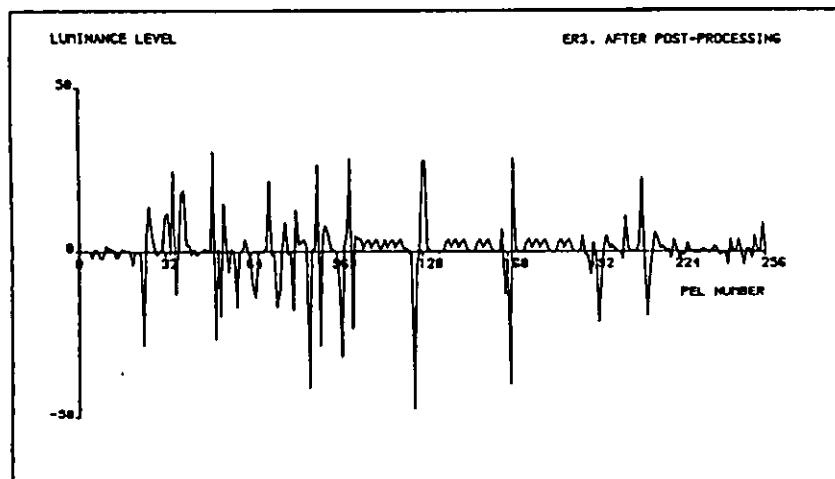
Fig.3.11. Captions overleaf



(e)



(f)



(g)

Fig.3.11. System 3: waveforms of a scan line

- (a) Original luminance waveform
- (b) Data signal
- (c) Combined luminance and data signal
- (d) Difference between (a) and (c)
- (e) Luminance waveform after data extraction and post filtering
- (f) Error in the recovered luminance signal without post filtering
- (g) Error in the recovered luminance signal when post filtering is used



(a)

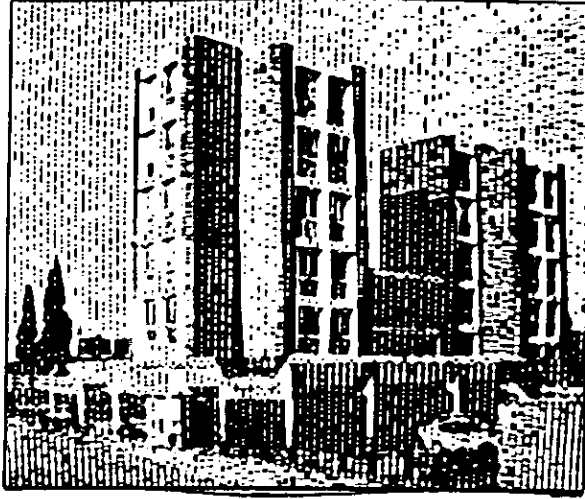


(b)

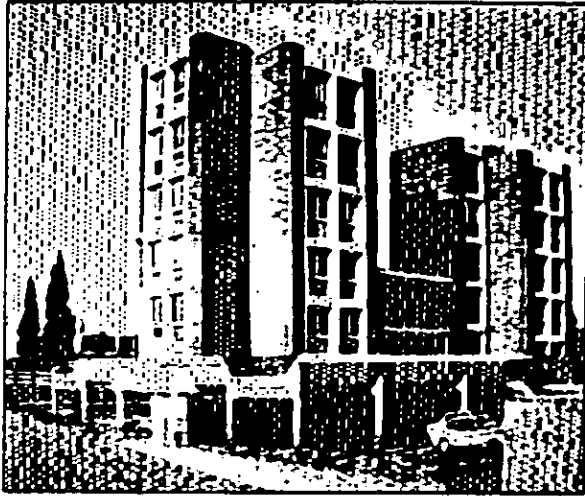


(c)

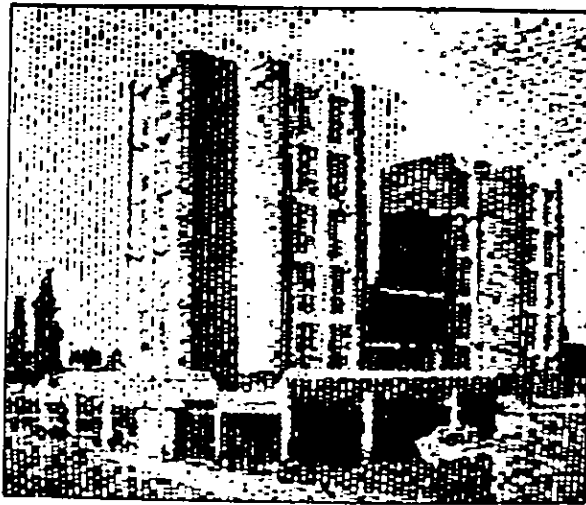
Fig.3.12. Recovered images from System 3 after being conveyed through the dispersive channel and $\text{SNR}_c = 40 \text{ dB}$ (a) GIRL (b) CARD (c) FLAT



(a)



(b)



(c)

Fig.3.13..The transmitted signal from (a) System:1
(b) System 2 (c) System 3

order Butterworth filter, was to lower the $\overline{\text{SNR}}$ numbers to 34, 36 and 29 dB for Systems 1, 2 and 3, respectively. The additive channel noise that yielded a channel SNR of 40 dB, is therefore a second order effect compared to the noise originating from the dispersion in the channel. However, inspection of the recovered pictures on the monitor leaves no doubt that even in the presence of the dispersive channel with its additive noise, System 2 is marginally superior to System 1, and the picture quality obtained with System 3 is significantly lower than that of the other two systems.

3.5. Discussion

Three systems for embedding data into industrial quality monochrome pictures have been proposed. The systems use the picture signal as a data carrier, and the combined picture-data signal does not require a greater bandwidth than that of the original picture signal. In addition the transmitted images have some degree of privacy because the effect of embedding data is to speckle the image with unexpected bright and dark dots, thereby making viewing fatiguing for an intruder.

Using six images having 256 scan-lines with 256 pels per line, Systems 1, 2 and 3 conveyed an average of 17430, 9713 and 21760 bits per image, an average bit rate efficiency η (see eq. (3.43)) of 80, 74 and 100 percent. The values of η show that a large majority of the blocks in Systems 1 and 2 were able to support data, and all the blocks in System 3 were embedded with data. The average number of pels required for the transmission of one bit

of data was 3.76, 6.75 and 3 for Systems 1, 2 and 3, respectively.

It must be emphasized that no bit errors were obtained for the six images that comprised the source signals. Other images will probably cause bit errors. The BER can be reduced to an acceptable value by introducing sufficient channel protection coding bits at the expense of the transmitted bit rate.

Systems 1 and 2 operate with a variable data rate, and because highly active blocks are excused from conveying data, the recovered picture quality is greater than that obtained with System 3, which has a constant data rate of 21760 bits per image. Although System 2 provides a recovered image having a marginally higher quality than that obtained with System 1, it does have a significantly lower data rate. System 3, however, has the virtue of being significantly more robust to bit errors compared to Systems 1 and 2, see figures 3.4, 3.7 and 3.10. Thus it can be observed that each system has its strengths and weaknesses, and the system to be deployed depends on the application.

The high data rates per frame, see Table 2, would if maintained for different images, yield high data rates for television signals with their 25 or 30 frames per second. Thus one might envisage a video-telephone service conveying data of the order of half a mega-bit per second, or the equivalent of approximately eight digitized voiced signals, in addition to the picture signal.

Note On Publications

Two papers, based on the work described in this chapter, were produced both entitled "Embedding data into pictures by modulo masking". The first paper, which deals with systems 1 and 3, was presented at the International Conference on Communications, 19-22 June 1983, Boston, M.A. and is published in the conference proceedings. The second paper, which deals with all three systems of this chapter, has been accepted for publication in the IEEE Trans. on Commun., as a full paper, to be published this year. All the above papers are in joint authorship with Dr. C. Xydeas and Dr. R. Steele.

CHAPTER IV

SWITCHED QUANTIZATION - FIXED BIT RATE

4.1. Introduction

An important form of predictive coding is differential pulse code modulation (DPCM). In this predictive coding system 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 then quantized. The quantized output may then be binary coded to either a fixed or variable word length. Hence a DPCM encoder consists of three basic components, predictor, quantizer, and code assigner.

In this chapter a DPCM coder, coupled with a scrambling strategy yielding a novel switched quantization scheme, will be described. The scheme is applied with 1-D and 2-D fixed predictors and with fixed word length assignment. However, the focus of attention is first centred on some relevant work reported on video differential coders.

4.1.1. Linear Predictors in Video Differential Coders

Predictors are usually optimised in a mean square sense, using the short or long term correlation function of the input video signal. This approach produces a set of linear equations the solution of which provides the 'optimum' coefficients of the linear predictor. Prediction coefficients optimised under the mean square criterion have been computed by Habibi⁷⁰ for predictors which use

different numbers of picture elements. His results show that if the predictor coefficients are matched to the statistics of a picture the m.s.e. decreases significantly by using up to three picture elements. However, if the coefficients are not exactly matched, the decrease in m.s.e. is not significant by using three previous elements as compared to one element. In nonadaptive DPCM systems predictors are usually designed employing the 'average' (long term) statistics of the input signal and not according to the short term statistics, for example every picture frame. The picture signal is however, highly nonstationary it is therefore advantageous to adapt the predictor to the local property of the video signal. This adaptive prediction approach has been tried by Zschunke⁷¹, and Dukhovich and O'Neal⁷². The technique used was to select a set of linear predictors based on some local property of the previously transmitted neighbouring picture elements. However, the results have not been very encouraging in terms of m.s.e. or the entropy of the prediction error. The development of a more complex prediction algorithm has been difficult due to the processing time restrictions imposed by the high sampling rate of the video signal. Nevertheless Girod⁷³ developed a gradient type algorithm such that a high speed real time implementation would be possible. His results show that additional gain by the use of adaptive predictor control, compared to optimum linear fixed predictors reported by Pirsch⁷⁴, in terms of prediction error variance or entropy is not significant. The SNR increase of 1.3 dB over optimum linear fixed

predictors obtained, could hardly justify the increase in complexity due to the introduction of the gradient algorithm. Using this adaptive prediction scheme the recovered image was indistinguishable from the original by a human observer under normal viewing conditions, and there was about 2.9 bits/pel used in conveying the picture information.

The video delta codec used in reference 75 adapts the predictor coefficients with a forward estimation scheme. The predictor uses either the horizontal or vertical samples for the prediction of the incoming signal depending on which one produced a smaller difference signal. The side information for the control of such a scheme is as large as 1 bit/pel and can be justified since it was found that the picture quality was better compared to a fixed delta-codec operating at twice the original sampling rate. The recovered picture produced from this scheme improves upon the edge busyness and most contours were well reconstructed.

Frei, Schindler and Vettiger⁷⁶ describe a dual mode predictive coding system using a DPCM and delta modulator. The delta modulator is used to code the video waveform in quiet and relatively smooth areas of the picture while the DPCM is used in regions of sharp transients. The switching between the two modes of operation is done by a sensor that responds to the input signal variation. The delta coder produces a bit-rate of $3f_s$ (f_s is the Nyquist sampling rate) which are later grouped and compressed according to a majority decision rule, in order to achieve a compression ratio 3:1. The receiver is able to reconstruct the compressed delta bits from a coding table. Upon

producing three equal bits within one sampling interval, from the delta coder, a sensor detects this as a deviation from a smooth region in the video waveform and switches the system into the DPCM mode. The system remains in this mode until the idling of the DPCM is detected. The idling condition of the DPCM is defined as the occurrence of the smallest DPCM word followed by the same one of opposite sign. When this idling is sensed the transmitter and receiver will each switch to the delta mode of operation in turn. For the system to operate successfully i.e. for the switching of modes to be the same at the transmitter and receiver, it was necessary to insert redundant marker bits. Nevertheless an average bit rate of 1.5 bits/pel was reported with good subjective quality pictures for single frame monochrome images.

4.1.2. Quantizers in Video Differential Coders

Quantization is an approximation process which maps continuous in amplitude levels (X) into discrete in amplitude output levels (Y). Hence the quantizer introduces distortion and since the quantization process is irreversible any good quantizer design must attempt to minimize this distortion. The basic quantizer problem is to determine the optimum decision and reconstruction levels, given the signal probability density function $p(X)$ and an optimization criterion.

In DPCM systems, optimization of the quantizer is carried out based mainly on two criteria : mean square quantization error and the subjective visibility threshold of the quantization error. Since the mean square error

criterion does not adequately take into account the human observer and the special properties of visual perception, its usefulness is restricted. For better picture quality, quantizers should be designed on the basis of the psychovisual criteria, however the debate continues on what is a good criteria to use^{77,78}. Optimum quantizers based on a minimum mean-square error criterion have been derived using the work of Max⁴³.

Max showed that the mean square error can be represented by :

$$MSE = \sum_{i=1}^L \int_{d_i}^{d_{i+1}} (X - Y_i)^2 p(X) dX \quad (4.1)$$

where

X is the input signal to be quantized

Y_i is the quantizer reconstruction levels

$p(X)$ is the probability density function of the input signal

d_i is the quantizer decision levels

The mean square error is minimised if the following two criteria are met :

$$d_i = \frac{(Y_i + Y_{i+1})}{2} \quad (4.2)$$

$$\int_{d_i}^{d_{i+1}} p(X)(Y_i - X) dX = 0 \quad (4.3)$$

Equations (4.2) and (4.3) describe the overall relationship of the optimum quantizer. Equation (4.2) states that the decision level d_i should lie half-way between Y_i and Y_{i+1} . Equation (4.3) shows Y_{i+1} to be the centroid of the area of $p(X)$ between d_i and d_{i+1} .

One method of solving these equations is to apply a search procedure. For example; for a given number of levels (L) with respect to $d_1 = -\infty$ and arbitrary selection of d_2 , the most negative quantizer output Y_1 can be obtained. If Y_L is the centroid of the area of $p(X)$ between d_L and ∞ , Y_1 was chosen correctly. Otherwise a new value must be assigned to d_1 and the procedure repeats.

The performance of DPCM is dependent upon the method used to set the quantizer's decision and reconstruction levels. The quantization error, i.e. the difference between the input and output of the quantizer, can produce a characteristic pictorial error called granularity error that forms a texture pattern in regions of constant brightness. Quantization step size may also limit the slew rate of the coder and lead to an error called slope overload, producing an image appearance similar to low pass filtering the image. There is one other degradation associated with improper quantizer design called 'edge busyness'. This occurs at slow varying contrast edges when the quantizer output oscillates around the signal value giving the appearance of a 'busy edge'.

Due to the problems outlined above, with respect to using nonadaptive quantizers, it is therefore advantageous to adapt the DPCM quantizer. Adaptation of the quantizer to signal statistics or visual perception is accomplished using various approaches^{79,80,81}. Some of these techniques are discussed in the following chapter. It is felt sufficient here to quote Netravali and Limb⁹; "In summary, for intraframe coding, although the optimum rules for adaptation are not known, rules based on intuition, and trail and error have shown significant reduction in bit rate over nonadaptive systems".

In what follows a novel feed forward switched quantizer system will be described while operating in a DPCM coder. Even though the scheme uses forward estimation it does not require any side information. Since the schemes described in chapter V rely on the knowledge acquired from the schemes in this chapter it is recommended that the schemes in this chapter be fully understood.

4.2. Fixed Bit Rate Switched Quantization (FSQ) Scheme (1-D)

A novel switched 3 bit/pel quantization scheme will be considered, as applied to a DPCM codec with a first order predictor, without the necessity of transmitting any side information. The basic idea originates from the analogue system described in the previous chapter. The analogue system exploited the statistical dependence that resides in video signals enabling them to be data carriers at the same time. However, a DPCM codec already uses this statistical dependence to reduce the channel

capacity necessary for conveying the image information. Since nonadaptive DPCM encoders are usually designed on a long term average basis they are not optimum. This small but inherent presence of redundancy will aid the design of the proposed quantization scheme. By rearranging, or not rearranging i.e. scrambling or not scrambling, the quantizer output levels in a predetermined manner and assuming the decoder is able to distinguish between the two sequences a switching code can be transmitted. This switching code will be used to switch between two quantizers present in the system. Hence the receiver must determine whether the scrambling key was used or not in order to establish which quantizer (of two quantizer) was used so that the correct video signal may be recovered.

4.2.1. The Transmitter

The encoder section of the switched quantization scheme is shown in figure 4.1. It basically consists of two conventional encoders, decoders and a processor section. The two encoders have the same first order predictor P but use quantizers ($Q1$ and $Q2$) with a different step size. The quantizers are of the type described by Max⁴³ with a Laplacian distribution. The two decoders are labelled $Q1$ and $Q2$ respectively. These signify decoding with respect to $Q1$ and $Q2$ step sizes. It is assumed that the step size of quantizer $Q1$ is smaller than the step size of quantizer $Q2$. The boxes labelled SC indicate a scrambling operation. As stated earlier the scrambling strategy is applied to the quantizer output levels which are scrambled by inversion about the midpoint level of

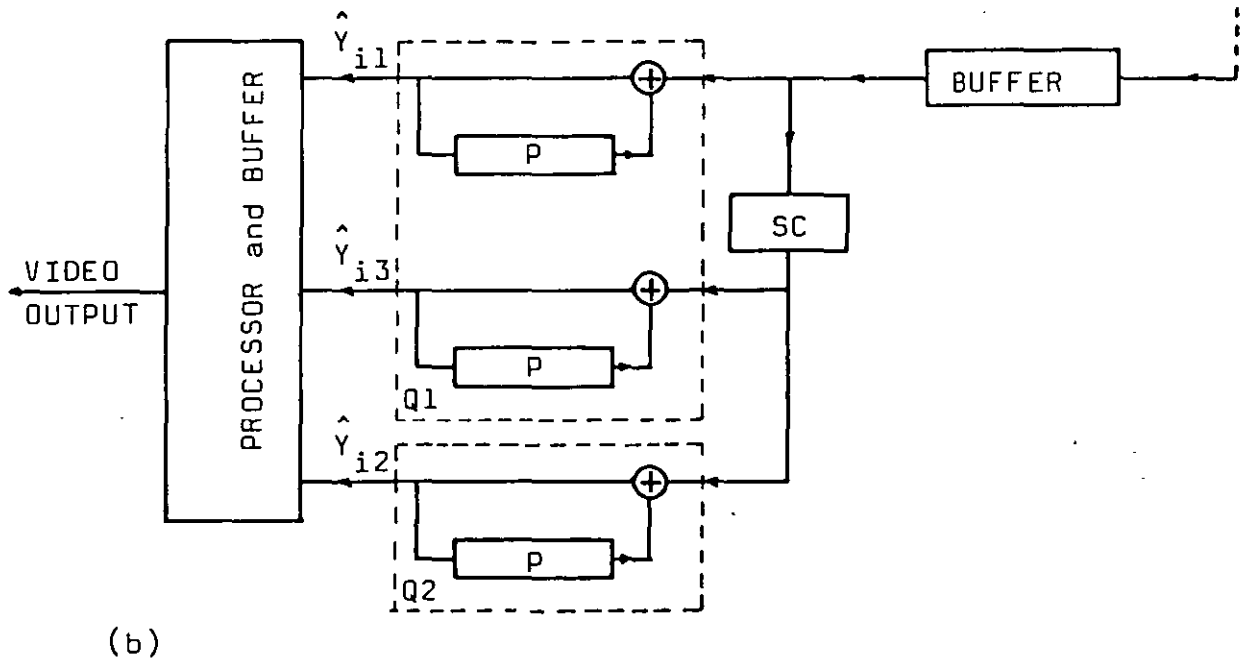
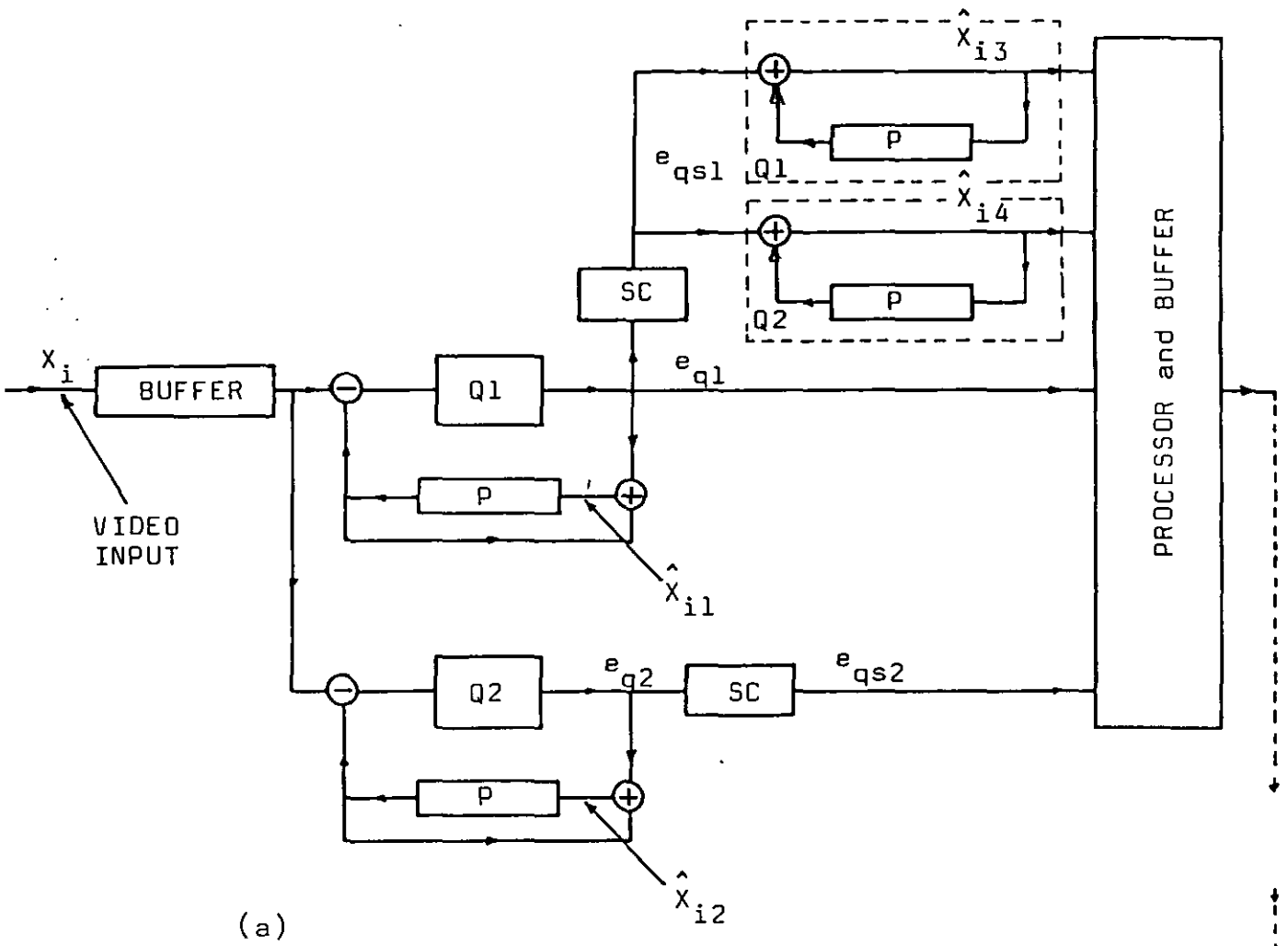


Fig.4.1. Block diagram of the FSQ 1-D scheme
(a) transmitter (b) receiver

the positive or negative half of the quantizer's range, depending on whether the quantizer's output was positive or negative respectively. The aim of this type of scrambling strategy is to push the inner quantization levels to their outer levels and hence obtain a signal that has a higher activity (defined later in the text) than its unscrambled version. The processor which is also a decision maker is given the information about the two locally decoded sample values $\hat{X}_{i1}$, $\hat{X}_{i2}$, and $\hat{X}_i$. As a criterion for system optimization it was decided to minimize the quantization noise power. The system is designed to operate on blocks of eight samples at any given time.

To be able to follow the description the reader is advised to keep in mind that a scrambled sequence produces a higher activity (defined later) signal than an unscrambled sequence. Although this is not always true it will help in the better understanding of the scheme. Furthermore whenever an unscrambled sequence e_{q1} is transmitted it is always associated with the first encoder i.e. using $Q1$ to encode the input signal. Similarly whenever a scrambled sequence e_{q2} is transmitted it is always associated with the second encoder i.e. using $Q2$ to encode the input signal. As a further aid the transmitter operation is presented in the flow chart of figure 4.2.

Consider that the first block of 8 input samples is processed by the system. Upon receiving the 8 sample values for $\hat{X}_i$, $\hat{X}_{i1}$, $\hat{X}_{i2}$, $\hat{X}_{i3}$, $\hat{X}_{i4}$ the processor computes the noise power from :

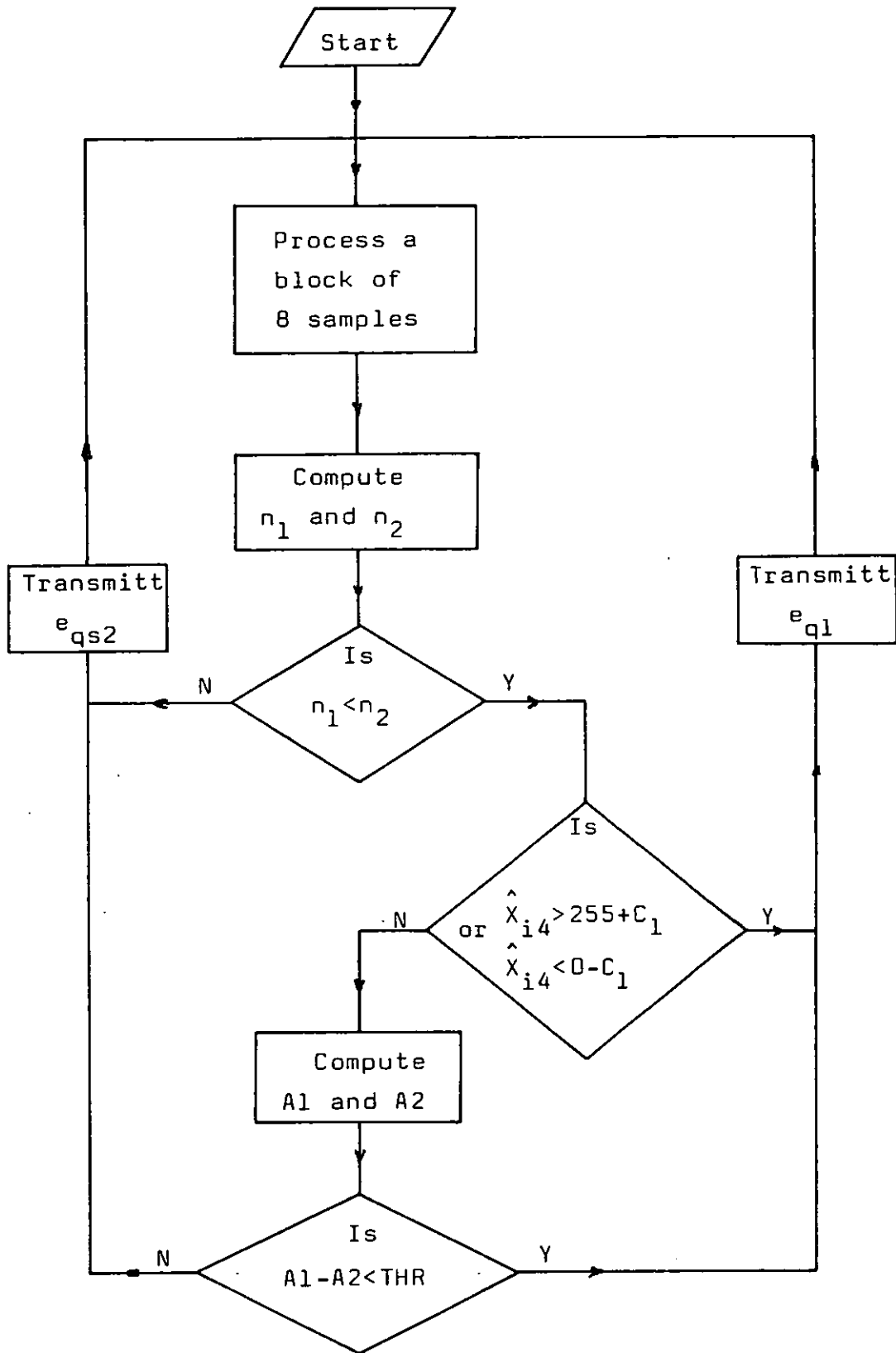


Fig.4.2. Flow chart for the transmitter section of the FSQ 1-D scheme

$$n_1 = \sum_{i=1}^8 (x_i - \hat{x}_{i1})^2 \quad (4.4)$$

$$n_2 = \sum_{i=1}^8 (x_i - \hat{x}_{i2})^2 \quad (4.5)$$

If $n_1 \leq n_2$, it means that the quantization noise produced by the first encoder (associated with Q1) was less than that of the second encoder. Therefore it would be desirable to transmit the output from the first encoder; and vice versa if $n_2 > n_1$.

Assuming that $n_1 < n_2$ a decision has to be made on whether it is possible to transmit the output from the first encoder such that the receiver processor will not make an error in deciding which encoder was used, without sending any side information. All further examination is therefore performed on the sample values $\hat{x}_{i1}$, $\hat{x}_{i3}$ and $\hat{x}_{i4}$ since the receiver, figure 4.1, is able to produce these samples as well. Now, the source signal was obtained using an 8-bit uniform quantizer, consisting of discrete levels ranging from 0 to 255 and any large deviation from that range may be exhibited by $\hat{x}_{i4}$ and $\hat{x}_{i3}$. This is because the scrambling process might drive the predictors unstable. Clearly the examination of $\hat{x}_{i4}$ against the stated deviation range (with some confidence limits) is sufficient to observe whether the scrambled quantized output from Q1 produces this deviation.

Thus the processor first examines whether there

are any samples in $\hat{X}_{i4}$ such that :

$$\hat{X}_{i4} > 255 + C_1 \quad (4.6)$$

or

$$\hat{X}_{i4} < 0 - C_2 \quad (4.7)$$

where C_1 and C_2 are experimentally obtained constants
or confidence limits

If either eq.(4.6) or (4.7) are satisfied the processor decides to transmit the set of binary values associated with the first encoder i.e. e_{q1} . This is possible since the receiver is able to produce equivalent samples of $\hat{X}_{i1}$, $\hat{X}_{i3}$ and $\hat{X}_{i4}$ and following the same procedure arrive at the same conclusion i.e. an unscrambled sequence was sent using $Q1$. If eqs.(4.6) or (4.7) are not satisfied the processor examines the activity of $\hat{X}_{i1}$ and $\hat{X}_{i3}$ defined as :

$$A1 = \frac{1}{(N-1)} \sum_{i=1}^{N-1} (\hat{X}_{i1} - \hat{X}_{i1+1})^2 \quad (4.8)$$

$$A2 = \frac{1}{(N-1)} \sum_{i=1}^{N-1} (\hat{X}_{i3} - \hat{X}_{i3+1})^2 \quad (4.9)$$

$$N = 8$$

As stated earlier it is expected that $A1$ will be less than $A2$. To make certain that this statement always holds,

when transmitting e_{q1} , it was necessary to introduce a threshold measurement parameter THR. Upon computing values for A1 and A2 the processor examines if:

$$A1 - A2 < THR \quad (4.10)$$

where THR is a system threshold constant

This is an experimentally obtained constant and will be discussed later in the experimental procedure (section 4.2.3.). It is sufficient to know that such a threshold exists.

If condition (4.10) is satisfied the processor permits the values, corresponding to the first encoder, to be transmitted i.e. e_{q1} . Since A1 will have a value less than A2 all the receiver has to select, after satisfying condition (4.10), is the set of samples with the lower activity in order to recover the correct video signal.

If however condition (4.10) is not satisfied the processor permits the scrambled values e_{q2} corresponding to the second encoder to be transmitted even though $n_1 < n_2$. This is possible since it was experimentally found that the activity difference between the scrambled sequence and the unscrambled sequence always produces a value that falls above the system measurement threshold THR. The reason why a similar examination such as eq.(4.10) is not performed at the transmitter is that it was experimentally found to hold for all blocks of the

scrambled sequence e_{qs2} . It is sufficient at present to know that such a threshold exists. Its derivation (of THR) will be explained in the experimental procedure (section 4.2.3.).

Assuming that $n_1 \geq n_2$ a scrambled sequence e_{qs2} would be transmitted without any further examination. This is possible since it was experimentally found that the activity difference between a scrambled sequence and an unscrambled sequence always produces a value that falls above the system threshold parameter THR. It is sufficient at present to know that such a threshold exists. Its derivation (of THR) will be explained in the experimental procedure (section 4.2.3.).

After encoding the last pel in each block of samples the last decoded samples ($\hat{x}_{81}, \hat{x}_{82}, \hat{x}_{83}$ and $\hat{x}_{84}$) of the system are set equal to the decoded sample, either $\hat{x}_{81}$, or $\hat{x}_{82}$, of the encoder chosen for processing the video signal. This ensures the exact starting condition of all the predictors.

To summarize in brief, a scrambled sequence of binary levels is always transmitted when it is decided to select the second encoder and an unscrambled sequence of binary levels is always transmitted when the first encoder is selected and used for the transmission of the video signal.

4.2.2. The Receiver

The receiver section of the switched quantizer scheme is similar to the transmitter and is shown in fig.4.1. It basically consists of three decoders and a

processor section. Two decoders decode the incoming signal with respect to the first quantizer (Q1) and one with respect to the second quantizer (Q2). The receiver also operates on blocks of eight samples and follows a similar examination of the incoming signal to that of the transmitter.

The processor first examines the samples $\hat{Y}_{i2}$ to find out if :

$$\hat{Y}_{i2} > 255 + C_1 \quad (4.11)$$

or

$$\hat{Y}_{i2} < 0 - C_2 \quad (4.12)$$

where C_1 and C_2 are system constants

If either eq.(4.11) or (4.12) are satisfied the video output signal contains values from $\hat{Y}_{i1}$. For all other cases the activity of sample values $\hat{Y}_{i1}$ and $\hat{Y}_{i3}$ are computed from :

$$A3 = \frac{1}{(N-1)} \sum_{i=1}^{N-1} (\hat{Y}_{i1} - \hat{Y}_{i1+1})^2 \quad (4.13)$$

$$A4 = \frac{1}{(N-1)} \sum_{i=1}^{N-1} (\hat{Y}_{i3} - \hat{Y}_{i3+1})^2 \quad (4.14)$$

and if

$$A3 - A4 < \text{THR} \quad (4.15)$$

eq.(4.15) is satisfied the video output will contain

sample values from $\hat{Y}_{i1}$ otherwise it will be from $\hat{Y}_{i2}$. Note that by experimental definition of the system measurement parameter THR, whenever the activity difference (A3-A4) resides below THR, an unscrambled sequence was transmitted using Q1. Similarly when the activity difference (A3-A4) falls above/or equal to THR a scrambled sequence was transmitted using Q2.

Final note : If the final video output consists of sample values from $\hat{Y}_{i1}$ the last decoded samples of the other decoders are set equal to the last decoded sample of $\hat{Y}_{i1}$. This only occurs at the end of each decoded block of pels after the processor has decided what the final video output signal should be and ensures the exact starting condition of all the predictors.

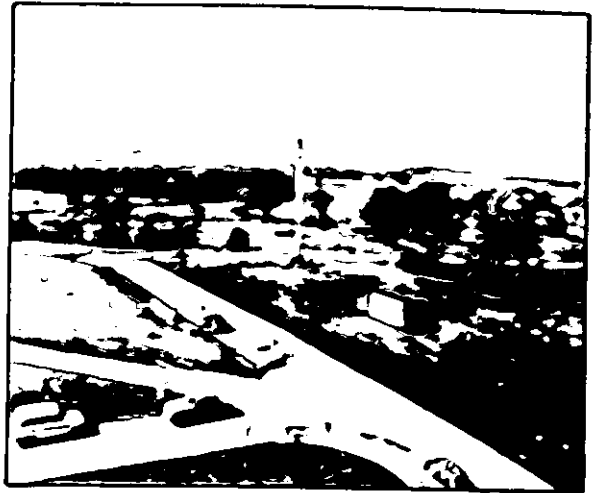
4.2.3. Experimental Procedure

The system parameters have been obtained using the six pictures shown in fig.4.3 and stored on a PDP11/34 minicomputer. Each source picture was quantized using an 8-bit uniform quantizer and consists of discrete levels ranging from 0 to 255. The predictor tap gain was obtained by computing the horizontal autocorrelation of each image. The six values so obtained were then averaged to produce one number, this being 0.99.

Initial experiments were carried out on the GIRL and the FLAT pictures only, to see the effect of the threshold THR with respect to the step size used for Q1 and Q2 under 'ideal conditions' i.e. assuming that the choice of encoders is only governed by the minimisation of the noise power and is known by the receiver. The



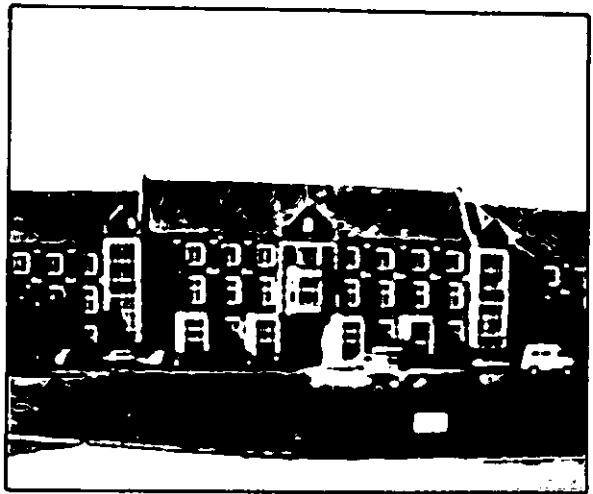
(a)



(d)



(b)



(e)



(c)



(f)

Fig.4.3. The source images used in the experiments

(a) GIRL (b) CARD (c) FLAT (d) SCENE

(e) HALL (f) GIRL 2

term 'non ideal conditions' being used for the case when no additional information is given to the receiver other than the system constants. The p.d.f. of the activity difference when Q1 or Q2 is used, under 'ideal conditions', was plotted and is shown in fig.4.4 for the GIRL, CARD and FLAT images (the p.d.f. plots are not shown for the other images since they were of similar characteristics to the ones in fig.4.4). As the step size of Q1 approaches the step size of Q2 the area of graph F1 would move to the right and reduces in size. The overlap region between F1 and F2 must be avoided so that the receiver never makes an error in deciding which quantizer was used. This is done by placing a threshold level THR with respect to the area of F2 e.g. in fig.4.4, and with respect to the p.d.f. plots for all six pictures so forcing the area of F1 to the right of THR to be mapped into the area of F2. This will have the effect of forcing the use of quantizer Q2 in the area where Q1 gives a lower noise power value. Therefore the choice of step size for Q1 and Q2 must be such that it does not have very noticeable visual impairments on the recovered video signal due to the overlap region of F1 and F2. The step size of Q1 and Q2 was changed by having two different normalization constants for the error signal being fed to the quantizer. The system was first run under 'ideal conditions'. For a given variation of normalization constants (for Q1 and Q2) it was necessary to establish the system constants THR, C_1 and C_2 , shown in Table 4.1, with respect to the six pictures. It was then necessary to decide which set of normalization constants

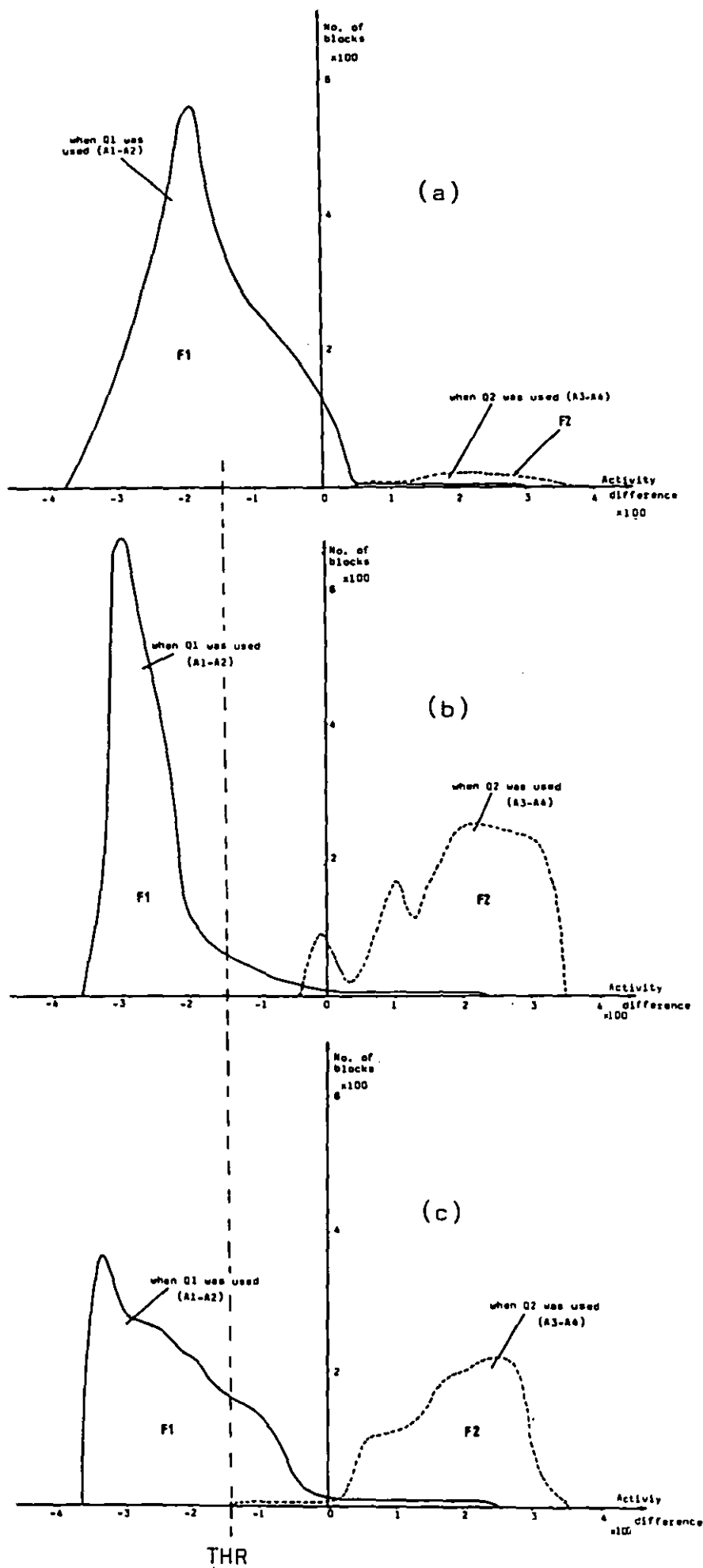


Fig.4.4. P.D.F. plots from the FSQ 1-D scheme, 'ideal conditions'
(a) GIRL (b) CARD (c) FLAT

		NORMALIZATION CONSTANT FOR Q1		
		4	5	6
NORMALIZATION CONSTANT FOR Q2	20	THR= -126.63 C ₁ = 8.05 C ₂ = 13.92	THR= -197.79 C ₁ = 10.88 C ₂ = 14.06	THR= -284.75 C ₁ = 11.19 C ₂ = 14.10
	25	THR= -109.21 C ₁ = 16.48 C ₂ = 17.56	THR= -170.62 C ₁ = 15.47 C ₂ = 17.67	THR= -244.11 C ₁ = 16.99 C ₂ = 17.63
	30	THR= -92.73 C ₁ = 12.12 C ₂ = 21.01	THR= -170.84 C ₁ = 12.58 C ₂ = 21.11	THR= -208.43 C ₁ = 14.34 C ₂ = 21.11

Table 4.1. System parameters for the FSQ 1-D scheme under 'ideal conditions'

(for Q1 and Q2) gave the higher SNR and/or the more pleasing subjective quality. The SNR values were computed from the root mean square error and the peak-to-peak signal amplitude. This was only done for the GIRL and FLAT picture as their picture statistics are very different from one another. Presented in Table 4.2 and Table 4.3 are the results when the decoder processor calculated correctly the decision made by the encoder processor, i.e. 'non ideal conditions', for the above mentioned two pictures.

Following our experimental procedure it was decided to take 6.0 and 30.0 as the two normalization constants for Q1 and Q2 respectively. For these values of scaling constants the corresponding system parameters are (from Table 4.1):

$$C_1 = 14.34 \quad C_2 = 21.11 \quad \text{THR} = -208.43$$

4.2.4. Results

With all the necessary parameters included in the scheme the system was run under 'non ideal conditions' and the SNR values produced by processing the six pictures are presented in Table 4.4. The results compare quite favourably to the case when one Laplacian quantizer was used and optimised for each picture separately with the method described by Max⁴³, as seen in Table 4.5. Given in Table 4.6 are SNR values for the just mentioned scheme when run under 'ideal conditions'. The results of Table 4.6 are a little surprising since they do not show any large deviations from the case when the system was run under

		NORMALIZATION CONSTANT FOR Q1		
		4	5	6
NORMALIZATION CONSTANT FOR Q2	20	36.67 dB Q2 = 3624 FQ2 = 2405	37.05 dB Q2 = 3039 FQ2 = 2246	37.21 dB Q2 = 2697 FQ2 = 2150
	25	35.82 dB Q2 = 3063 FQ2 = 1995	36.35 dB Q2 = 2487 FQ2 = 1773	36.77 dB Q2 = 2048 FQ2 = 1599
	30	35.34 dB Q2 = 2398 FQ2 = 1434	35.99 dB Q2 = 1908 FQ2 = 1285	36.68 dB Q2 = 1407 FQ2 = 998

Note : Q2=total number of blocks used by
the quantizer Q2 to encode the
input signal

FQ2=total number of forced blocks used
by the quantizer Q2 to encode the
input signal

Table 4.2. SNR results for the FSQ 1-D scheme,
for the GIRL image, under 'non ideal
conditions'

		NORMALIZATION CONSTANT FOR Q1		
		4	5	6
NORMALIZATION CONSTANT FOR Q2	20	30.98 dB Q2 = 5390 FQ2 = 2017	31.02 dB Q2 = 4954 FQ2 = 1737	31.03 dB Q2 = 4685 FQ2 = 1550
	25	32.63 dB Q2 = 4725 FQ2 = 1413	32.74 dB Q2 = 4318 FQ2 = 1133	32.80 dB Q2 = 4006 FQ2 = 959
	30	32.97 dB Q2 = 4143 FQ2 = 883	33.02 dB Q2 = 3915 FQ2 = 773	33.16 dB Q2 = 3564 FQ2 = 534

Note : Q2=total number of blocks used by
the quantizer Q2 to encode the
input signal

FQ2=total number of forced blocks used
by the quantizer Q2 to encode the
input signal

Table 4.3. SNR results for the FSQ 1-D scheme,
for the FLAT image, under 'non ideal
conditions'

IMAGE	SNR
GIRL	36.68 dB
CARD	31.44 dB
FLAT	33.16 dB
SCENE	38.41 dB
HALL	34.67 dB
GIRL 2	40.40 dB

Table 4.4. SNR results for the FSQ 1-D scheme
under 'non ideal conditions'

IMAGE	SNR	NORMALIZATION CONSTANT
GIRL	36.25 dB	
CARD	32.39 dB	
FLAT	30.95 dB	
SCENE	31.12 dB	
HALL	30.24 dB	
GIRL 2	36.43 dB	4.9

Table 4.5. SNR results for the 1-D DPCM when a Laplacian quantizer was optimised per frame (3 bits/pel)

IMAGE	SNR
GIRL	38.77 dB
CARD	31.62 dB
FLAT	33.39 dB
SCENE	39.65 dB
HALL	35.02 dB
GIRL 2	42.48 dB

Table 4.6. SNR results for the FSQ 1-D scheme
under 'ideal conditions'

'non ideal conditions' shown in Table 4.4. Upon further investigation it was found that under 'non ideal conditions' the amount of extra added noise (due to the use of Q2 in cases where Q1 should have been used) was a relatively small percentage of the total noise power. Presented in fig.4.5 are the recovered pictures of our new scheme corresponding to the results obtained in Table 4.4. Also shown in fig.4.6 are the recovered pictures when one Laplacian quantizer was optimised for each picture individually, using the mean square error criterion, corresponding to the results obtained in Table 4.5.

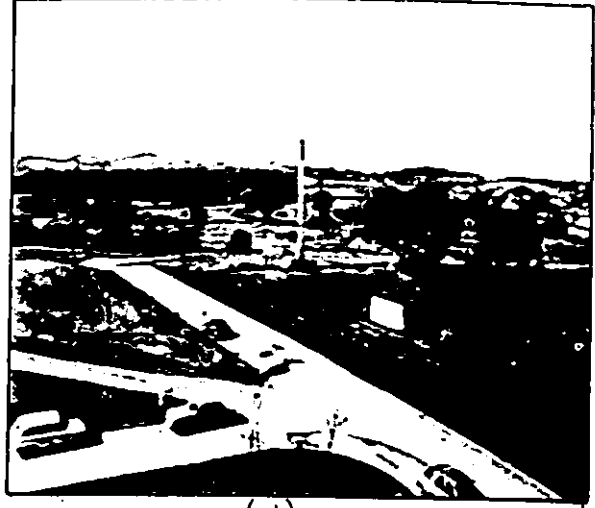
Subjectively the recovered video images from the new switched quantizer scheme were of equal quality to the case when one Laplacian quantizer was optimised for each picture individually using the minimum mean square criterion described in reference 43. Note however that the latter system requires the transmission of the step size information at a rate equal to the frame rate of the video input.

4.3. Fixed Bit Rate Switched Quantization (FSQ) Scheme (2-D)

The switched quantization scheme used for the 1-D case will be applied to a 2-D predictive differential system, with some modifications. It was decided to use a third order predictor as higher order predictors would not produce any noticeable improvement in picture quality⁷⁰. Due to the use of 2-D predictors it was more challenging to reduce the number of bits/pel to 2 bits/pel. The scrambling strategy used for the 1-D scheme will be



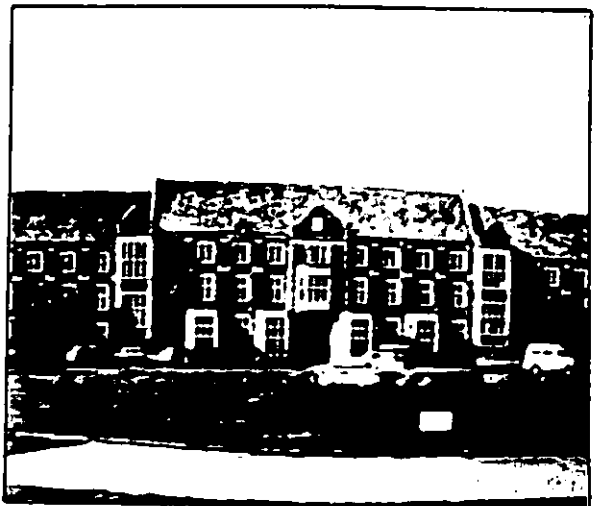
(a)



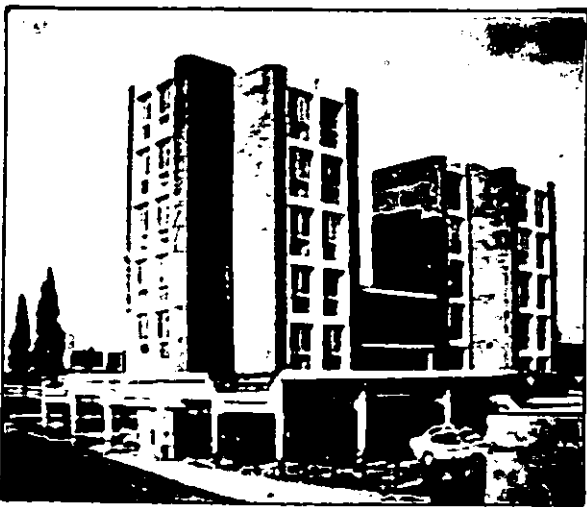
(d)



(b)



(e)



(c)

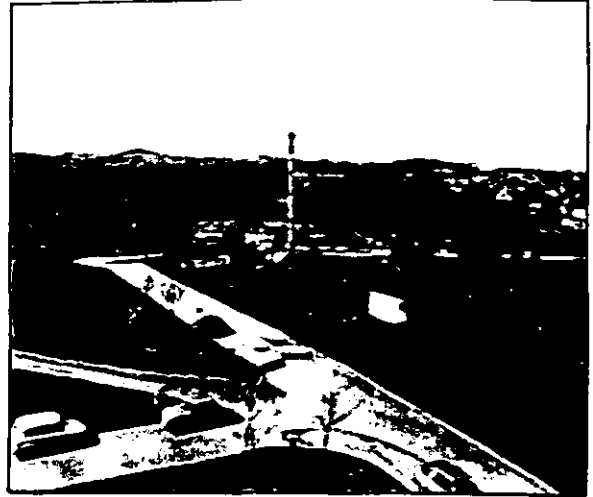


(f)

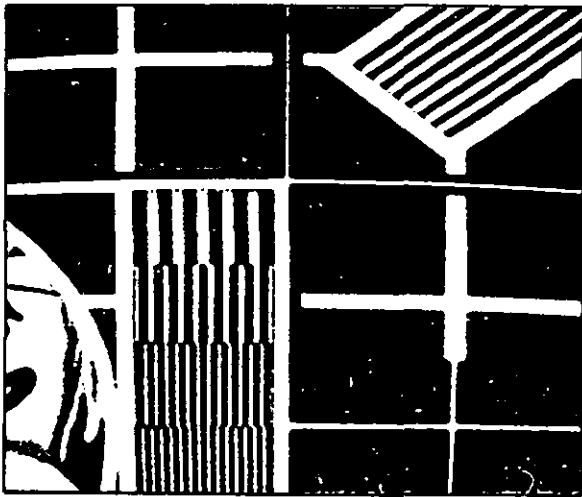
Fig.4.5. Recovered images from the FSQ 1-D scheme under
'non ideal conditions' (a) GIRL (b) CARD
(c) FLAT (d) SCENE (e) HALL (f) GIRL 2



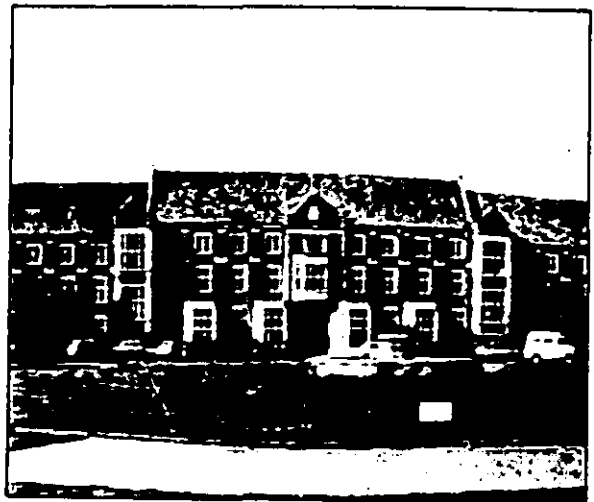
(a)



(d)



(b)



(e)



(c)



(f)

Fig.4.6. Recovered images using a Laplacian quantizer
optimized per frame (a) GIRL (b) CARD (c) FLAT
(d) SCENE (e) HALL (f) GIRL 2

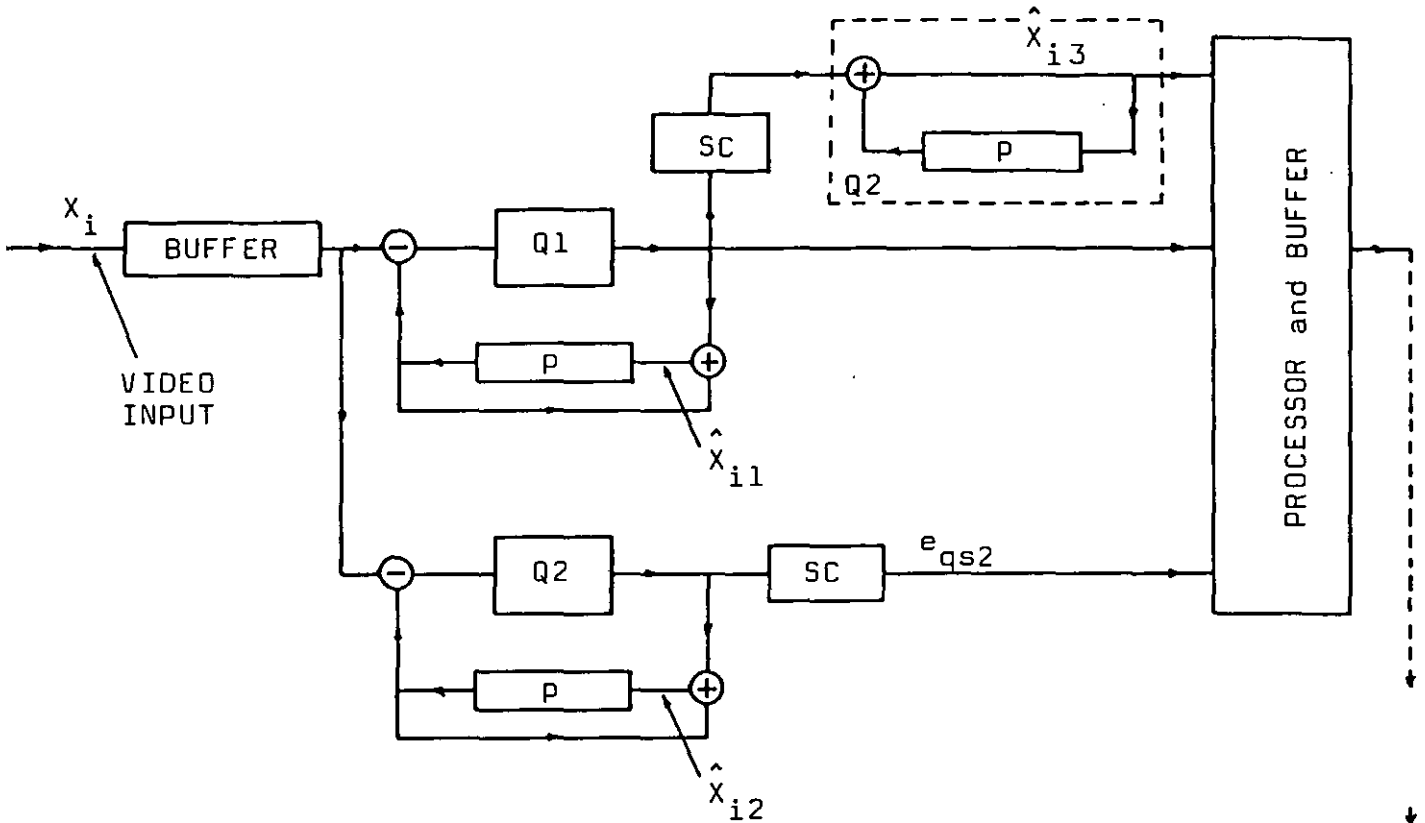
used here in exactly the same way. However, since the quantizers (Q1 and Q2) were allocating only 2 bits/pel this rather limits the effectiveness of the scrambling strategy to produce a scrambled signal within 8 pels such that it is distinguishable from its unscrambled version. Both quantizers were of the type described by Max⁴³ with a Laplacian distribution.

4.3.1. The Transmitter

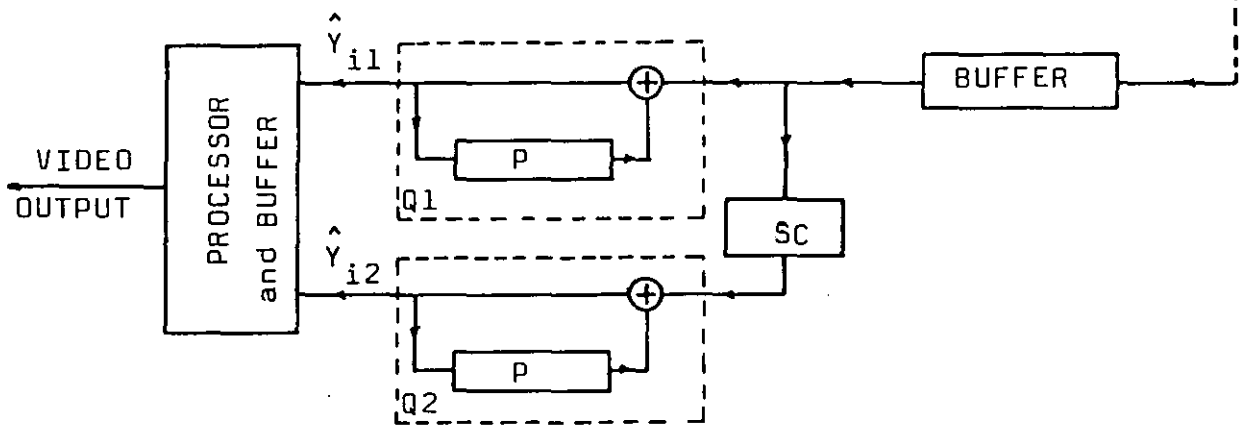
Shown in fig.4.7 is the block diagram of the transmitter and receiver section of 2-D scheme. The transmitter section basically consists of two encoders, one decoder and a processor unit. All the predictors P are 2-D third order predictors. The two encoders use two quantizers of the same distribution but with a different step size i.e. Q1 has a smaller step size than Q2. The remaining decoder performs decoding of the scrambled quantized output from Q1 with respect to the step size of Q2. The boxes labelled SC indicate a scrambling operation. As an optimization criterion, the minimization of the quantizers error signal is used subject to the constraint that other system conditions are satisfied as well.

The input video signal is first segmented into 8 samples per block. Hence commencing with the first block of pels the noise power is computed from :

$$n_1 = \sum_{i=1}^N (x_i - \hat{x}_{i1})^2 \quad (4.16)$$



(a)



(b)

Fig.4.7. Block diagram of the FSQ 2-D scheme
(a) transmitter (b) receiver

$$n_2 = \sum_{i=1}^N (x_i - \hat{x}_{i2})^2 \quad (4.17)$$

If $n_1 < n_2$ it means that it would be desirable to transmit the error signal associated with the first encoder. However, this is subject to one of the following constraints being satisfied. Therefore the examination of all the samples in $\hat{x}_{i3}$ by :

$$\hat{x}_{i3} > 255 + B_1 \quad (4.18)$$

or

$$\hat{x}_{i3} < 0 - B_2 \quad (4.19)$$

where B_1 and B_2 are system constants or
or confidence limits.

Since the source signal was obtained using an 8-bit uniform quantizer, consisting of discrete levels ranging from 0 to 255 any large deviation from that range will be exhibited by $\hat{x}_{i3}$. This is because the scrambling process might drive the predictors unstable. Clearly the examination of $\hat{x}_{i3}$ against the stated deviation range (with some confidence limits) is sufficient to observe whether the scrambled quantized output from Q1 produces this deviation.

If any of the samples $\hat{x}_{i3}$ exceed the inequalities imposed by eqs.(4.18) or (4.19) an unscrambled sequence e_{q1} , associated with Q1, is transmitted. Since the receiver

is able to produce equivalent samples of $\hat{X}_{i3}$, after performing an examination similar to eqs.(4.18) and (4.19) it would arrive at the same result i.e. an unscrambled sequence was transmitted.

In the event that inequalities of eqs.(4.18) and (4.19) are not satisfied the error signal e_{qs2} , associated with the second encoder, is transmitted even though this encoder produced a higher noise value.

If $n_1 \geq n_2$ it means that the second encoder produced a lower or equal noise value to that of the first encoder. Due to the design of the system the receiver will not be able to detect any adverse instability produced by the sequence e_{qs2} and will opt for the condition (defined by the system) that a scrambled sequence was transmitted.

After encoding the last pel in each block of samples the contents of the predictor delays of the system are set equal to the contents of the predictor delays of the encoder chosen for processing the video signal. This ensures the exact starting condition of all the predictors.

4.3.2. The Receiver

Due to the simplified structure of the transmitter the receiver, in fig.4.7, consists of two decoders and a processor unit. One decoder uses Q1 to decode the incoming digits and the other used Q2.

The examination of the recovered signals is performed only on $\hat{Y}_{i2}$ since this corresponds to $\hat{X}_{i3}$ generated by the transmitter section. Hence the examination :

$$\hat{Y}_{i2} > 255 + B_1 \quad (4.20)$$

and

$$\hat{Y}_{i2} < 0 - B_2 \quad (4.21)$$

Upon any of the samples exceeding the inequalities of eqs.(4.20) or (4.21) it is concluded that an unscrambled sequence was transmitted and the final video output is taken as $\hat{Y}_{i1}$. Otherwise the final video output is from $\hat{Y}_{i2}$. The receiver processor follows the same steps as that of the transmitter processor. Having the constants B_1 or B_2 properly set, and assuming there are no channel error, the transmitter and receiver will always make the same decision.

Final note : If the final video output consists of sample values from $\hat{Y}_{i1}$ the contents of the predictor delays of the second decoder are set equal to the contents of the predictor delays of the first decoder and vice versa. This only occurs at the end of each decoded block of pels, after the processor has decided what the final output should be, to ensure the same starting position of all predictors.

4.3.3. Experimental Procedure and Results

The multipliers used in the predictor are the same as those reported by Pirsch and Stenger⁷⁴ :

$$a_h = \frac{7}{8}$$

$$a_v = \frac{3}{4}$$

$$a_d = \frac{-5}{8}$$

Initial testing was carried out on the GIRL and FLAT pictures only to establish the SNR variation with respect to the normalization constants used for quantizers Q1 and Q2. These results are shown in Table 4.7 and Table 4.8. It was decided that 4.0 and 15.0 would be suitable normalizing constants for Q1 and Q2. The system parameters B_1 and B_2 were obtained by processing the six images through a 'dummy run'. This yielded values for B_1 and B_2 of 24.56 and 13.91 respectively.

With the constants B_1 and B_2 fixed the six images were processed again under 'non ideal conditions' to produce SNR values as shown in Table 4.9. These SNR values compare quite favourably to the case when considering system performance, under 'ideal conditions', shown in Table 4.10. When the pictures were evaluated subjectively the CARD picture had a noticeable increase in granular noise. This was remedied by reducing the normalization constant of Q2. Having chosen a lower normalization constant for Q2 (10.0) new values for B_1 and B_2 had to be found. The same procedure was adopted as in the previous case giving values for B_1 and B_2 of 22.0 and 12.8.

Consequently new pictures were obtained for the case when the normalizing constants of Q1 and Q2 were 4.0 and 10.0. Given in Table 4.11 are the results regarding the SNR values under the 'non ideal conditions'.

		NORMALIZATION CONSTANT FOR Q1		
		4	5	6
NORMALIZATION CONSTANT FOR Q2	15	36.79 dB Q2 = 627	36.87 dB Q2 = 472	36.58 dB Q2 = 438
	20	35.89 dB Q2 = 531	36.27 dB Q2 = 331	36.21 dB Q2 = 247
	25	35.42 dB Q2 = 456	35.71 dB Q2 = 273	35.88 dB Q2 = 179

Note: Q2=total number of blocks used by
the quantizer Q2 to encode the
input signal

Table 4.7. SNR results for the FSQ 2-D scheme,
for the GIRL image under 'ideal
conditions'

		NORMALIZATION CONSTANT FOR Q1		
		4	5	6
NORMALIZATION CONSTANT FOR Q2	15	31.58 dB Q2 = 2751	31.74 dB Q2 = 2449	31.71 dB Q2 = 2208
	20	32.18 dB Q2 = 2442	32.36 dB Q2 = 2001	32.68 dB Q2 = 1695
	25	31.61 dB Q2 = 2286	31.93 dB Q2 = 1831	32.29 dB Q2 = 1416

Note: Q2=total number of blocks used by
the quantizer Q2 to encode the
input signal

Table 4.8. SNR results for the FSQ 2-D scheme
for the FLAT image under 'ideal
conditions'

IMAGE	SNR
GIRL	32.44 dB
CARD	30.16 dB
FLAT	30.61 dB
SCENE	32.12 dB
HALL	30.59 dB
GIRL 2	33.83 dB

Table 4.9. SNR results for the FSQ 2-D scheme
under 'non ideal conditions'
(normalization constants for Q1 and
Q2 were 4 and 15)

IMAGE	SNR
GIRL	36.79 dB
CARD	33.19 dB
FLAT	31.58 dB
SCENE	34.55 dB
HALL	31.81 dB
GIRL 2	41.02 dB

Table 4.10. SNR results for the FSQ 2-D scheme
under 'ideal conditions'
(normalization constants for Q1 and
Q2 were 4 and 15)

IMAGE	SNR
GIRL	34.46 dB
CARD	31.38 dB
FLAT	28.42 dB
SCENE	31.74 dB
HALL	28.60 dB
GIRL 2	36.78 dB

Table.4.11. SNR results for the FSQ 2-0 scheme
under 'non ideal conditions'
(normalization constants for Q1 and
Q2 were 4 and 10)

The recovered pictures from this scheme are shown in fig.4.8. The SNR values of Table 4.11 compare very well to the values of Table 4.12 when the scheme was run under 'ideal conditions'. With the new normalizing constant for Q2 there was much less visible granular noise on the CARD picture. The other pictures did not show any visible degradation in picture quality as compared to those obtained when the normalizing constant of Q2 was 15.0.

A scheme was simulated using the same predictors as before, third order 2-D, however one Laplacian quantizer was used and optimized for each picture separately using the method described by Max⁴³. The SNR values so obtained are shown in Table 4.13 and the corresponding recovered pictures in fig.4.9. Although, overall, the scheme produced pictures with higher SNR it was not high enough to show subjective quality improvement.

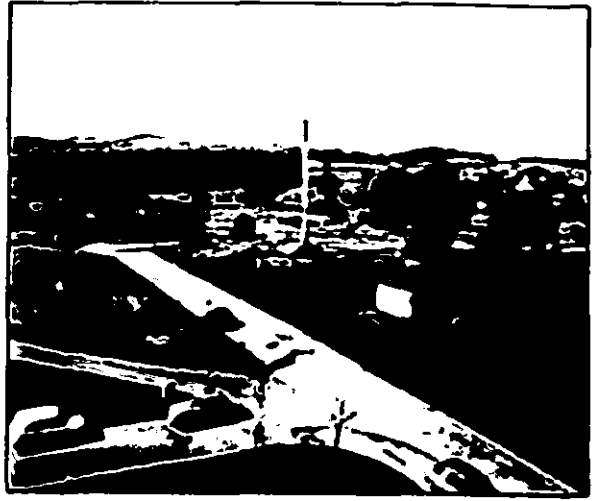
4.4. Discussion

The two schemes described in this chapter started from the principle that a switching type operation could be transmitted by using, or not using a scrambling key, and casting the receiver in the role of a code breaker. The systems use the inherent redundancy present in a differential coder when designed on a long term average basis using a fixed bit rate code assigner.

The results for the 1-D scheme show that a SNR improvement of up to 7 dB, compared to the scheme when the quantizer was optimized per frame as described by Max⁴³, is possibly operating at the same bit rate and using the same predictor. The scrambling strategy used,



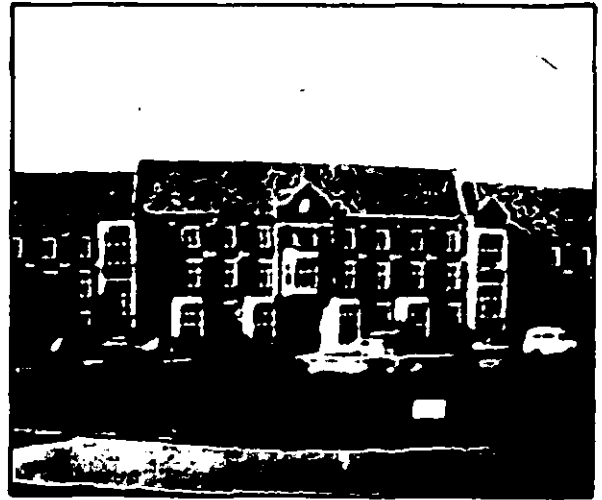
(a)



(d)



(b)



(e)



(c)



(f)

Fig.4.8. Recovered images from the FSQ 2-D scheme

(a) GIRL (b) CARD (c) FLAT (d) SCENE

(e) HALL (f) GIRL 2

IMAGE	SNR
GIRL	37.15 dB
CARD	32.91 dB
FLAT	28.68 dB
SCENE	32.70 dB
HALL	28.95 dB
GIRL 2	41.04 dB

Table 4.12. SNR results for the FSQ 2-D scheme
under 'ideal conditions'
(normalization constant for Q1 and
Q2 were 4 and 10)

IMAGE	SNR	NORMALIZATION CONSTANT
GIRL	33.52 dB	
CARD	23.76 dB	
FLAT	28.17 dB	
SCENE	29.92 dB	
HALL	28.07 dB	
GIRL 2	37.76 dB	3.3

Table 4.13. SNR results for the 2-D DPCM when a Laplacian quantizer was optimised per frame (2 bits/pel)



(a)



(d)



(b)



(e)



(c)



(f)

Fig.4.9. Recovered images when Laplacian optimized
per frame (a) GIRL (b) CARD (c) FLAT
(d) SCENE (e) HALL (f) GIRL 2

performed very well, this being due to the relatively high bit allocation per pel i.e. 3 bits/pel. It was therefore disheartening to find that any subjective quality improvement was minimal.

The results for the 2-D scheme show that an average SNR improvement of 1.8 dB., compared to the scheme when the quantizer was optimised, per frame, as described in reference 43, is possibly operating at the same bit rate and using the same predictor. The scrambling strategy used, did not perform well, this being due to the relatively low bit allocation per pel i.e. 2 bits/pel. However, the subjective quality of pictures obtained was improved upon the 1-D scheme mainly due to the 2-D predictor.

The difficulty these schemes presented was that the quantizer switching information was carried within the block that was to be decoded. In other words the received incoming block of pels could have been encoded with either one of the two quantizers. Hence it was necessary to devise a measurement technique to cover both possibilities i.e. if the block was encoded with Q1 or Q2.

In chapter V a variable bit rate system is used to enhance the SNR results obtained here and improve upon the subjective quality of the six pictures.

It must be emphasized that both systems were designed such that no errors occurred in the switching process at the transmitter and receiver, although this is not guaranteed for other input images.

CHAPTER V

SWITCHED QUANTIZATION - VARIABLE BIT RATE

5.1. Introduction

Video signals generally contain two types of redundancy which can be exploited in encoding. First, there is the redundancy associated with the correlation between signal samples which can be removed in differential coders. Second, there is redundancy due to the non uniform probability density of the discrete levels produced by the encoding algorithm. This second type of redundancy is removed by entropy coding using a variable length code word. Entropy coding is aimed at reducing the transmission bit rate but it will in return produce a variable data rate, hence requiring a buffer memory and overflow/underflow control strategies.

In chapter IV the use of an adaptive quantizer, in a feed forward estimation scheme, with an intraframe DPCM system was examined. Here the FSQ schemes, of chapter IV, are altered to accommodate quantizers with a different number of levels. This will in turn result in a more complex scheme producing a variable bit stream and requiring a buffer memory control (not discussed in this thesis). Whereas previously, in chapter IV, the encoded block of pels carried its own information as to which quantizer was used to encode the block of samples, in the schemes presented in this chapter the encoded block of pels will carry the switching information for the next block of samples. In other words the switching information

is carried one block of samples in advance. Without claim to completeness the reader's attention is first focused on some relevant work as regards adaptive quantization.

5.1.1. Video Adaptive Quantizers

It appears that adaptive quantization has been studied more intensely for delta modulation than for differential pulse code modulation in digital encoding of pictures. Among few papers published on adaptive quantizers, for DPCM encoders, the paper by Jayant⁸¹ treats the design of an adaptive quantizer scheme with a one-word memory. This simple algorithm uses a backward estimation scheme in recursively updating the scaling of the quantization levels by a factor g_n for each new sample :

$$g_n = M(|q_{n-1}|)g_{n-1} \quad (5.1)$$

where q_{n-1} stands for the quantizer index at time index $n-1$

$M(|q_{n-1}|)$ is a function of the normalized quantizer output

In all cases $M(q)$ is symmetrical around the zero level hence the absolute sign $M(|q|)$. This design can be applied to both speech and picture signals.

The paper by Zetterberg, Ericsson and Brusewitz⁸² presents a comprehensive comparison between various adaptive quantization schemes. Their study is centred on a DPCM type system with a fixed predictor using a

three-level quantizer. It was shown that Jayant's adaptive quantization strategy produced a 1 dB SNR increase over a fixed quantizer with modest subjective quality improvement. Among other adaptive quantizers used their results point to the fact that a feed forward estimation scheme gave the best performance in terms of SNR and subjective quality improvement. The quantizer was made adaptive by selecting three different gain parameters (step sizes) for each horizontal block of 8 pels. This means that it was necessary to transmit about 0.2 bits/pel as side information. Furthermore the simulations showed that more than three quantization levels are needed for acceptable picture quality, especially with scenes of a more complex nature.

In a more recent report by Zetterberg, Ericsson and Couturier⁸³ the performance of various adaptive quantization schemes was examined using a five level quantizer. They again use the so called Jayant quantizer⁸¹ and extend it to make the algorithm two-dimensional. The rule is applied to the horizontal and vertical direction from which a resultant is obtained by arithmetic averaging :

$$g_{n,m} = \frac{1}{2} \left[g_{n,m-1}^M(|g_{n-1,m}|) + g_{n-1,m}^M(|g_{n,m-1}|) \right] \quad (5.2)$$

The results obtained using a two-dimensional quantizer control was 2 dB better from the one-dimensional case and produced an improved subjective quality impression.

In addition to this, results from an adaptive quantizer with forward estimation using four gain constants was compared. The forward estimation scheme, although requiring side information, gave the highest SNR and the best picture quality with an average transmission rate of 1.8 bits/pel (in addition to 0.125 bits/pel side information).

From reference 82 and reference 83 it seems that there is a lot to be gained by using an adaptive quantizer with a forward estimation scheme. The disadvantage of such a scheme is that the necessary side information contributes a significant proportion to the average transmitted bit rate. In the following section two adaptive quantization schemes are described using feed forward estimation without the necessity of transmitting any side information. The difference between these schemes and those of chapter IV is that here the quantizers use a different number of levels and the switching information is carried one block of pels in advance.

5.2. Variable Bit-Rate Switched Quantization (VSQ)

Scheme (1-D)

The FSQ 1-D scheme (section 4.2) transmitted the switching information by allowing the processed block of pels to carry its own information about the choice of quantizer it was encoded with. Furthermore the results obtained showed clearly that a 3 bit/pel system using first order prediction was not able to cope well with the sharp transitions exhibited by some of our pictures. Since major parts of an image do not contain sharp

transitions between pels they could be coded with fewer quantization levels for an acceptable visual fidelity. However, at sharp transients it would be necessary to allocate more quantization levels. The result of such a bit allocation strategy would probably be an overall reduction in the total number of bits required for transmission with an improved image fidelity at the expense of a more complex scheme. Hence it was decided to use a similar switching strategy, as described in section 4.2, but to allocate a different number of levels to the two quantizers. This means that each block of pels would be encoded and transmitted with respect to one quantizer (of two quantizers) only. Therefore at each received block of digits decoding would be performed with respect to one of the two quantizers only. In such a situation it is possible to have a different set of examination processes depending on which quantizer was used to decode the received digits. Furthermore the receiver would need to know, in advance, which quantizer was used to encode the block of pels in order to perform the decoding process. This means that the switching information will have to be carried one block of pels in advance.

5.2.1. The Transmitter

For ease of explanation the transmitter and receiver section of the new scheme is shown in fig.5.1. To aid the better understanding of the transmitter operation a flow chart is provided in fig.5.2. The transmitter mainly consists of two encoders, two decoders and a

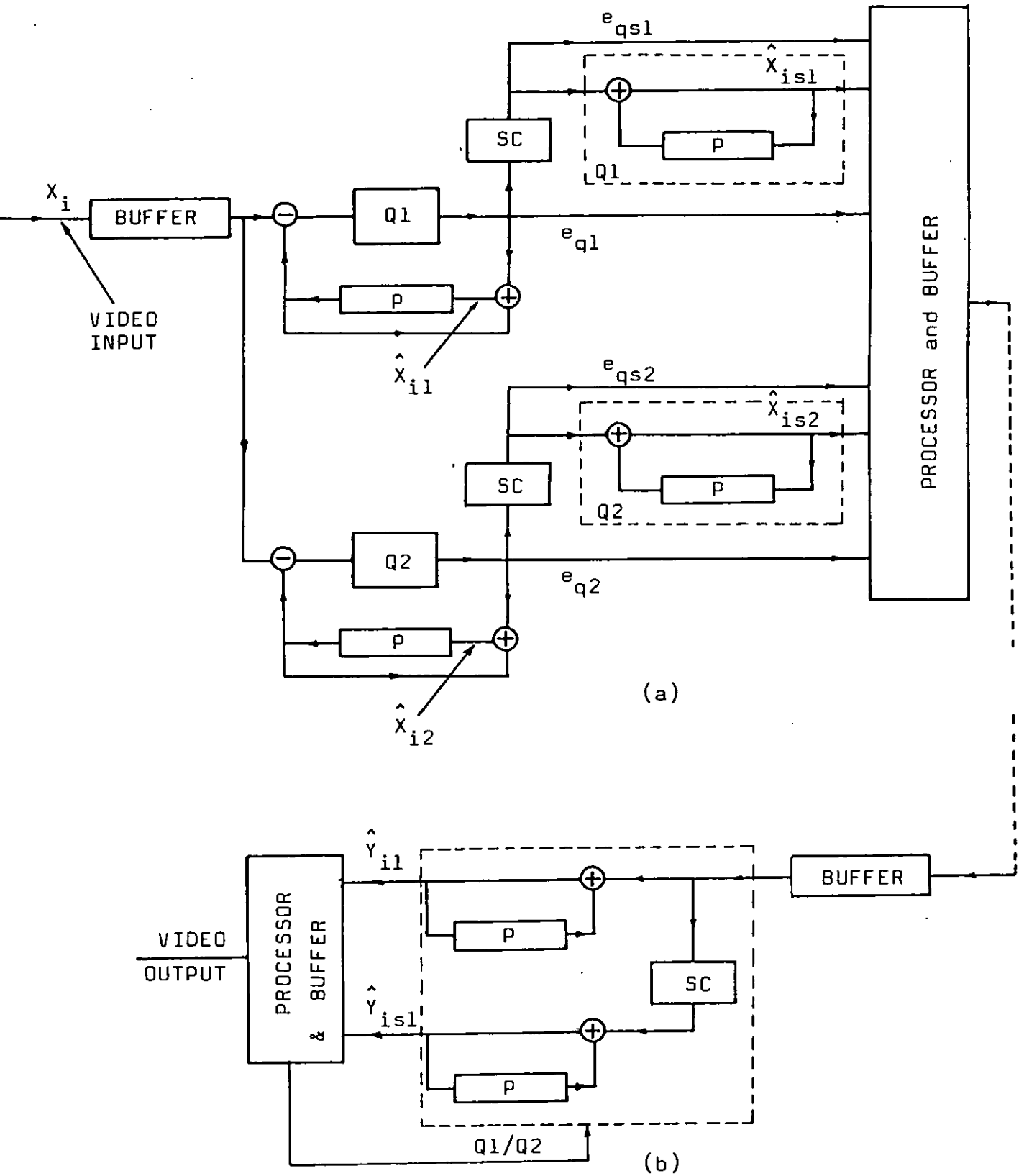
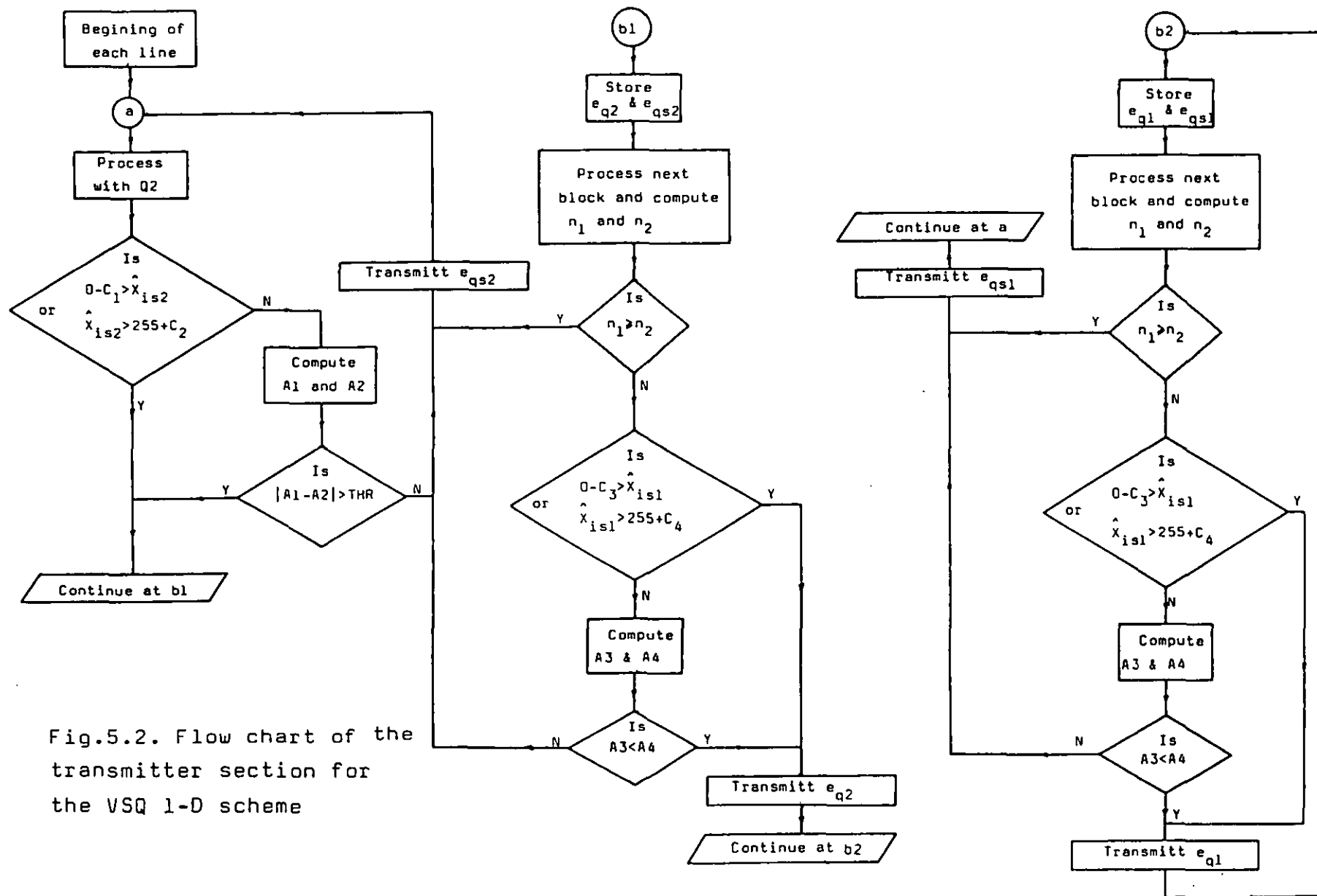


Fig.5.1. Block diagram of the VSQ scheme
(a) transmitter (b) receiver



processing unit. All the predictors P are first order predictors. The two encoders use quantizers ($Q1$ and $Q2$) having a Laplacian distribution designed on minimum mean square criterion as reported by Max⁴³. The quantizers use a different number of levels i.e. $Q1$ is a 2-bit quantizer and $Q2$ is a 4-bit quantizer. Labels SC indicate a scrambling operation of the quantizers output levels. This is achieved by inversion about the midpoint level of the quantizer's positive and negative half depending on whether the quantizer's output was positive or negative respectively. Such a scrambling process forces the inner quantization levels (these are more frequent) to their outer levels so that the recovered signal produces a higher activity (defined later) than its recovered unscrambled signal. At the two decoders there are labels $Q1$ and $Q2$ indicating decoding with respect to the two quantizers. It is assumed that the processor which is also a decision maker is provided with values of X_i , $\hat{X}_{i1}$, $\hat{X}_{is1}$, $\hat{X}_{i2}$, and $\hat{X}_{is2}$. The input signal is segmented into blocks of 8 pels before encoding begins.

To start processing the input video correctly at the transmitter and receiver it was decided to encode the first 8 pels of each video line using the second encoder i.e. $Q2$, 4-bit quantizer. Hence commencing with the first block of pels at the beginning of a video line, fig.5.2 :

- (a) The processor examines the sequence $\hat{X}_{is2}$ if :

$$0 - C_1 > \hat{X}_{is2} > 255 + C_2 \quad (5.3)$$

where C_1 and C_2 are system constants

Since the source signal was digitized using an 8-bit uniform quantizer, having discrete levels ranging from 0 to 255, any large deviation from that range would be exhibited by $\hat{X}_{is2}$ and not by $\hat{X}_{i2}$. This is because the scrambling process could drive the predictors unstable. Hence an examination such as inequality of eq.(5.3) would enable the selection of an unscrambled set of error samples, in some instances.

If any of the $\hat{X}_{is2}$ exceeds the limitations imposed by the inequalities of eq.(5.3), the transmitted sequence can be either from e_{q2} or e_{qs2} depending on the next block of pels. The next block of samples to be encoded continues to be examined from section (b), described later on in the text, or (b1) in fig.5.2.

However, if $\hat{X}_{is2}$ are confined to the inequalities imposed by eq.(5.3) an activity measurement is performed on the sequence $\hat{X}_{i2}$ and $\hat{X}_{is2}$ by :

$$A1 = \frac{1}{(N-1)} \sum_{i=1}^{N-1} (\hat{X}_{i2} - \hat{X}_{i2+1})^2 \quad (5.4)$$

$$A2 = \frac{1}{(N-1)} \sum_{i=1}^{N-1} (\hat{X}_{is2} - \hat{X}_{is2+1})^2 \quad (5.5)$$

and if

$$| A1 - A2 | > THR \quad (5.6)$$

where THR is a system constant

the transmitted sequence can be either from e_{q2} or e_{qs2} depending on the next block of 8 pels. Since an unscrambled sequence nearly always produces a lower activity signal than a scrambled sequence it is necessary to incorporate THR as an additional condition to accommodate the cases when this may not be true and there is a possibility of an error being made by the receiver. This is an experimentally derived threshold parameter and it was found that above this value the scrambled sequence always produces a signal of higher activity than the unscrambled sequence. For this particular situation i.e. when $|A1-A2| > THR$ the receiver is able to drive at the unscrambled sequence, regardless of which sequence (scrambled or unscrambled) was transmitted, by following the sequence with a lower activity value. The exact derivation of the parameter THR will be discussed in section 5.2.3. It is sufficient at present to know that such a threshold exists.

For such cases when the absolute activity difference falls below THR the transmitted sequence would be from e_{qs2} , by definition of the system, and the next block of pels would be encoded using Q2 starting from section (a).

(b) Consider the next block of samples to be encoded. If this is the second block of pels to be encoded then the processor stores the previous sequences e_{q2} and e_{qs2} shown in the flow chart fig.5.2 from position (b1). The noise power of the two encoders is measured from :

$$n_1 = \sum_{i=1}^N (\hat{x}_{i1} - x_i)^2 \quad (5.7)$$

$$n_2 = \sum_{i=1}^N (\hat{x}_{i2} - x_i)^2 \quad (5.8)$$

If $n_1 \geq n_2$ it means that the second encoder (4-bit quantizer) produced the same or lower noise value than the first encoder (2-bit quantizer). Assuming that the present block is the second block of pels to be encoded the processor transmits the previous block as e_{qs2} . Further examination of the signal to be encoded continues from (a).

For all other cases i.e. $n_2 > n_1$ the sequence $\hat{x}_{is1}$ is further examined by the inequalities :

$$0 - C_3 > \hat{x}_{is1} > 255 + C_4 \quad (5.9)$$

where C_3 and C_4 are system constants

If any $\hat{x}_{is1}$ lie outside the inequalities defined in (5.9) it means that the signal will be encoded using the first encoder. Assuming that the present block is the second block of pels to be encoded the previous block is transmitted as e_{q2} . The examination of the following block, or third block continues from section (b) with the processor stored and possible transmitted values of e_{q2} and e_{qs2} substituted by e_{q1} and e_{qs1} . This is

shown in the flow chart of fig.5.2 at position (b2).

Note, from fig.5.2, that the examination of the encoded samples through paths (b1) and (b2) are the same.

If all $\hat{X}_{isl}$ lie within the bounds of the inequalities defined in eq.(5.9) a further inspection is performed on the two sequences $\hat{X}_{il}$ and $\hat{X}_{isl}$ by forming an activity measure :

$$A3 = \frac{1}{(N-1)} \sum_{i=1}^{N-1} (\hat{X}_{il} - \hat{X}_{il+1})^2 \quad (5.10)$$

$$A4 = \frac{1}{(N-1)} \sum_{i=1}^{N-1} (\hat{X}_{isl} - \hat{X}_{isl+1})^2 \quad (5.11)$$

If $A3 < A4$ the signal will be encoded using the first encoder (2-bit quantizer). Assuming that the present block is the second block of pels to be encoded the previous block is transmitted as e_{q2} . The examination of the following block, or third block, continues from section (b) with the processor stored and possible transmitted values of e_{q2} and e_{qs2} substituted by e_{q1} and e_{qs1} . This is shown in the flow chart of fig.5.2 at position (b2). Note, from fig.5.2, that the examination of the encoded samples path (b1) and (b2) are the same.

For any other case i.e. $A3 \geq A4$, the signal will be encoded using the second encoder (4-bit quantizer) continuing its examination from section (a). Assuming that the present block is the second block of pels to be encoded the previous block is transmitted as e_{q2} .

Final note for the transmitter :

It must be pointed out that upon deciding which encoder is to be used, in encoding a block of pels, the last decoded pels of all other predictors are made equal to the decoded pel of the predictor chosen for encoding. This ensures the same starting position for all predictors. From the description of the transmitter operation, and flow chart fig.5.2, it can be concluded that :

- (1) A scrambled sequence is transmitted if the following block of pels is to be encoded with the second encoder (Q2), satisfying the necessary system conditions.
- (2) An unscrambled sequence is transmitted if the following block of pels is to be encoded with the first order (Q1), satisfying the necessary system conditions.

5.2.2. The Receiver

The block diagram of the receiver section is shown in fig.5.1. It is composed of mainly two encoders and a processing unit. Since at any one block the incoming signal could be encoded using either Q1 or Q2 the receiver processor also switches the decoding process with respect to Q1 or Q2. As stated earlier in section 5.2.1. the first block of each video line is encoded using the 4-bit quantizer. It is justified therefore to use this initial starting condition at the receiver. It shall also be assumed that the receiver has knowledge about all the system constants i.e. C_1 , C_2 , C_3 , C_4 and THR.

(a) Starting with the first block of pels in a video line the decoding is performed with respect to Q2 (4-bit quantizer). The two recovered sequences $\hat{Y}_{i1}$

and $\hat{Y}_{isl}$ are examined by the inequalities defined :

$$0 - C_1 > \hat{Y}_{il} > 255 + C_2 \quad (5.12)$$

$$0 - C_1 > \hat{Y}_{isl} > 255 + C_2 \quad (5.13)$$

where C_1 and C_2 are system constants.

If any of the $\hat{Y}_{il}$ was found to lie outside the limitations imposed by eq.(5.12) the recovered signal would be chosen as $\hat{Y}_{isl}$ and the next block would be decoded with respect to Q2. In case any $\hat{Y}_{isl}$ was found to lie outside the limitations of eq.(5.13) the recovered signal would be chosen as $\hat{Y}_{il}$ and the next block would be decoded with respect to Q1 (2-bit quantizer). Failing this the activity of $\hat{Y}_{il}$ and $\hat{Y}_{isl}$ is obtained using :

$$A1 = \frac{1}{(N-1)} \sum_{i=1}^{N-1} (\hat{Y}_{il} - \hat{Y}_{il+1})^2 \quad (5.14)$$

$$A2 = \frac{1}{(N-1)} \sum_{i=1}^{N-1} (\hat{Y}_{isl} - \hat{Y}_{isl+1})^2 \quad (5.15)$$

and if

$$| A1 - A2 | > THR \quad (5.16)$$

where THR is a system constant

the final video output is chosen as the one associated with the lower activity value. Assuming the sequence

$\hat{Y}_{i1}$ produced a lower activity measure, it would be taken as the final video output and the decoding of the next block would be with respect to Q1. Similarly if $\hat{Y}_{is1}$ produced a lower activity measure, it would be taken as the final video output and the decoding of the next block would be with respect to Q2.

If the absolute activity difference (with respect to $\hat{Y}_{i1}$ and $\hat{Y}_{is1}$ is less than or equal to the threshold THR, the final video output is taken as $\hat{Y}_{is1}$ and decoding of the next block is done with respect to Q2.

(b) For the sake of completeness it is now necessary to consider the situation when the previous video output was from the sequence $\hat{Y}_{i1}$. This situation can occur if the transmitter encoded and sent a block either as e_{q1} or e_{q2} . Detecting this state the receiver decodes the next incoming block with respect to Q1. The so produced samples $\hat{Y}_{i1}$ and $\hat{Y}_{is1}$ are examined by the inequalities defined as :

$$0 - C_3 > \hat{Y}_{i1} > 255 + C_4 \quad (5.17)$$

$$0 - C_3 > \hat{Y}_{is1} > 255 + C_4 \quad (5.18)$$

where C_3 and C_4 are system constants

If any $\hat{Y}_{i1}$ are found to lie outside the limitations imposed by eq.(5.17) the recovered video signal would be chosen as $\hat{Y}_{is1}$ and the next block would be decoded with

respect to Q2. In case any $\hat{Y}_{isl}$ are found to lie outside the limitations of eq.(5.18) the recovered signal would be chosen as $\hat{Y}_{il}$ and the next block would be decoded with respect to Q1 (2-bit quantizer). Failing this the activity of $\hat{Y}_{il}$ and $\hat{Y}_{isl}$ is obtained using :

$$A3 = \frac{1}{(N-1)} \sum_{i=1}^{N-1} (\hat{Y}_{il} - \hat{Y}_{il+1})^2 \quad (5.19)$$

$$A4 = \frac{1}{(N-1)} \sum_{i=1}^{N-1} (\hat{Y}_{isl} - \hat{Y}_{ils+1})^2 \quad (5.20)$$

In this situation the processor is only looking for the set of samples producing a lower measure of activity since that will be the one chosen as the final video output.

Final note for the receiver :

Upon deciding which decoder produced the correct set of video output samples the value of that last decoded pel is transferred in place of the value of the last decoded pel of the other decoder. This ensures the same starting position for all predictors.

5.2.3. Experimental Procedure and Results

The set of source pictures used in chapter IV, fig.4.3. are the same ones used here. Since the chosen predictors were all first order predictors the tap gains were all set to a value of 0.99. The quantizers (Q1 and Q2 in fig.5.1) were of the type described by Max⁴³ with a Laplacian distribution. Therefore it was necessary to

find suitable scaling constants for the error signal being fed to the quantizers Q1 and Q2.

Experiments were initially performed on the GIRL and FLAT picture only to see the effect of the two scaling constants associated with quantizers Q1 and Q2. These experiments were conducted under 'ideal conditions' i.e the minimizations of noise power being the only constraint with the receiver being given all the decisions made by the transmitter, and are shown in Table 5.1 and 5.2. The term 'non ideal conditions' refers to the situation when no additional information is passed on to the receiver other than the system constants. The choice of scaling constants for quantizers Q1 and Q2 were chosen with respect to the average bit rate per pel. The scaling values chosen for the quantizers, Q1 and Q2, were 5.0 and 35.0 respectively.

With the scaling constants chosen, the system was run again under 'ideal conditions' for the rest of the source pictures. Table 5.3 shows the average bit rate and SNR values obtained for the six pictures under the just mentioned conditions. At the same time the p.d.f. of the activity difference was plotted when quantizer Q1 and Q2 was used. These graph plots are shown in fig.5.3 for the GIRL, CARD and FLAT pictures only, because they were the more demanding of the six images considered. It will be recalled that a block of pels can be transmitted as a scrambled or unscrambled set of error samples. With respect to the area F2, fig.5.3, this means that there could theoretically exist another plot of F2 i.e. (A1-A2), mirrored about the vertical axis.

		NORMALIZATION CONSTANT FOR Q2		
		25	30	35
NORMALIZATION CONSTANT FOR Q1	4	39.06 dB Q2 = 4203 TBR = 3.02	38.18 dB Q2 = 3501 TBR = 2.85	37.47 dB Q2 = 3102 TBR = 2.75
	5	38.68 dB Q2 = 4773 TBR = 3.16	37.91 dB Q2 = 3678 TBR = 2.89	37.28 dB Q2 = 2967 TBR = 2.72
	6	38.35 dB Q2 = 5905 TBR = 3.44	37.53 dB Q2 = 4152 TBR = 3.01	38.98 dB Q2 = 3139 TBR = 2.76
	7	38.17 dB Q2 = 6556 TBR = 3.60	37.13 dB Q2 = 5094 TBR = 3.24	36.53 dB Q2 = 3674 TBR = 2.89

Note: Q2=total number of blocks used by
the quantizer Q2 to encode the
input signal
TBR=average bit rate per frame
(in bits/pel)

Table 5.1. SNR results for the VSQ 1-D scheme
for the GIRL image under 'ideal
conditions'

		NORMALIZATION CONSTANT FOR Q2		
		25	30	35
NORMALIZATION CONSTANT FOR Q1	4	37.76 dB Q2 = 4704 TBR = 3.14	37.51 dB Q2 = 4276 TBR = 3.04	37.12 dB Q2 = 4063 TBR = 2.99
	5	37.61 dB Q2 = 4933 TBR = 3.20	37.36 dB Q2 = 4371 TBR = 3.06	37.03 dB Q2 = 4079 TBR = 2.99
	6	37.51 dB Q2 = 5440 TBR = 3.32	37.22 dB Q2 = 4625 TBR = 3.13	36.92 dB Q2 = 4177 TBR = 3.02
	7	37.37 dB Q2 = 5440 TBR = 3.39	37.03 dB Q2 = 4929 TBR = 3.20	36.70 dB Q2 = 4343 TBR = 3.06

Note: Q2=total number of blocks used by
the quantizer Q2 to encode the
input signal
TBR=average bit rate per frame
(in bits/pel)

Table 5.2. SNR results for the VSQ 1-D scheme
for the FLAT image under 'ideal
conditions'

IMAGE	SNR	Q2	TBR
GIRL	37.28 dB	2967	2.72 bits/pel
CARD	34.70 dB	3835	2.93 bits/pel
FLAT	37.03 dB	4079	2.99 bits/pel
SCENE	39.58 dB	1980	2.48 bits/pel
HALL	37.87 dB	2804	2.68 bits/pel
GIRL 2	39.64 dB	1570	2.38 bits/pel

Note: Q2=total number of blocks used by the quantizer
Q2 to encode the input signal

TBR=average bit rate per frame

Table 5.3. Results for the VSQ 1-D. scheme under
'ideal conditions'

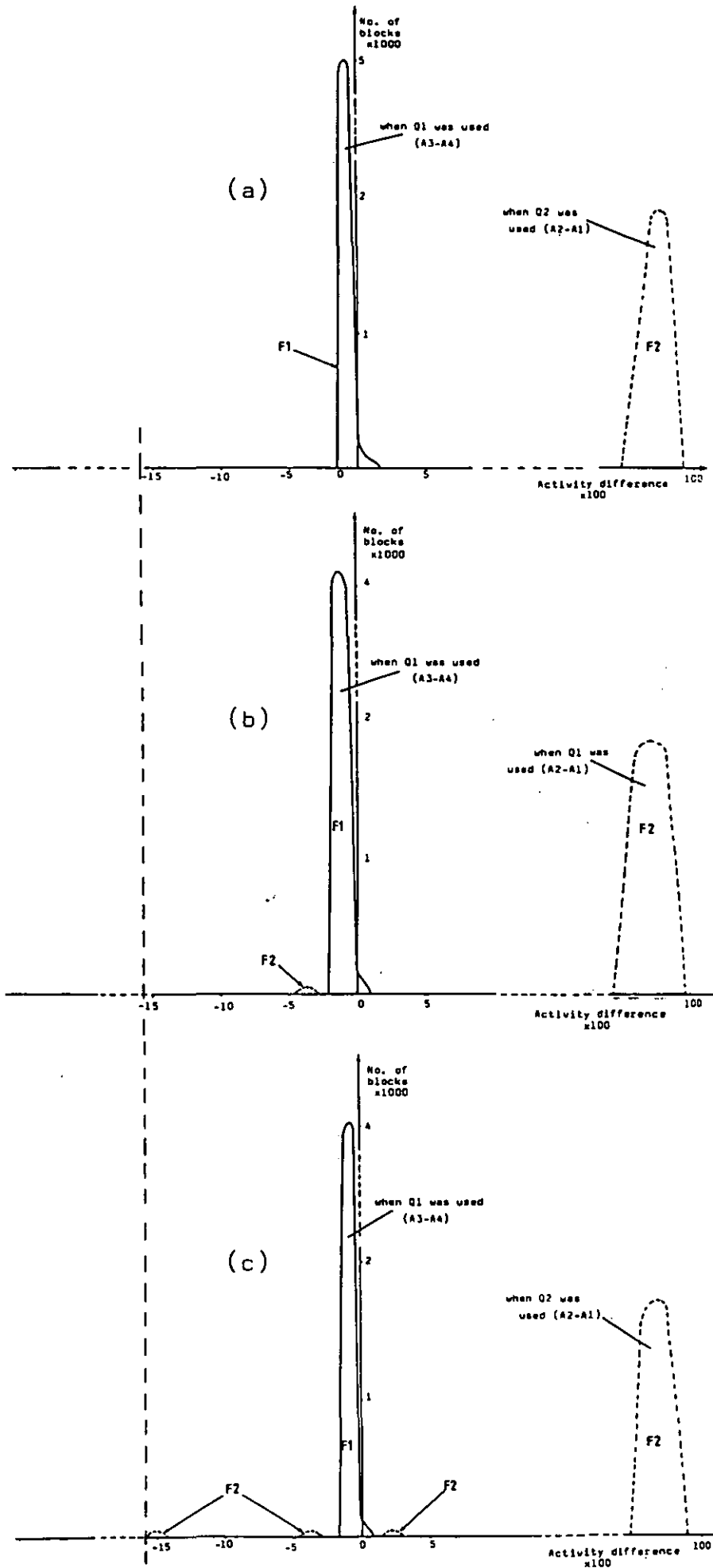


Fig.5.3. P.D.F. plots of the activity difference measure for the VSQ 1-D scheme, 'ideal conditions'

Hence there will exist a region between $\pm \text{THR}$ where it is not possible to make a definite decision as to which set of samples corresponds to the unscrambled signal. Therefore whenever the absolute activity difference falls below the absolute value of THR it was decided to transmit a set of scrambled samples and encode the next block with Q2. For the area of F1, fig.5.3, there could also exist another plot of F1 i.e. (A4-A3), mirrored about the vertical axis. In this case the overlap region is excluded by forcing the use of the second encoder (Q2) whenever $A3 \geq A4$. The threshold level THR was chosen from the lowest value attained by the activity difference of area F2 in fig.5.3, with respect to all six pictures. This value turned out to be a negative number (not very surprising), however, since the proposed system considers the absolute value only, THR was chosen as :

$$\text{THR} = 1641.0$$

During the 'ideal conditions' run it was possible to obtain a good approximation of the other system constants namely C_1 , C_2 , C_3 and C_4

Upon choosing a value for THR the system was run, once under 'ideal conditions' and once under 'non ideal conditions', processing the six pictures in 'dummy' runs so that the remaining system constants C_1 , C_2 , C_3 and C_4 could be finalised. The values produced were :

$$C_1 = 14.5$$

$$C_3 = 4.5$$

$$C_2 = 7.0$$

$$C_4 = 6.0$$

With the system parameters computed in the above mentioned procedure there was no possibility that the receiver would make a different decision other than the one made by the transmitter (for the six pictures that comprise our source data). Other images may cause problems to the system proposed, however this can be overcome by readjusting the system parameters.

Once all the system constants were set the six pictures were processed again under the 'non ideal conditions'. Shown in Table 5.4 are the results obtained under the 'non ideal conditions' of the above mentioned scheme revealing the SNR and average bit-rate for each of the six pictures. Note how close the values of Table 5.4 correspond to that of Table 5.3, in terms of SNR and average bit-rate. This is because the forced use of quantizer Q2, in areas where quantizer Q1 should have been used, was on average about 2% of the total number of blocks processed. Subjective tests conducted using experienced viewers on the received images showed that they were of exceptionally good perceptual quality. These images are shown in fig.5.4. Comparing the FSQ 1-D scheme with the VSQ 1-D scheme it can be seen that the SNR values between the two schemes are not very different, Table 4.4 and Table 5.4. However, the VSQ 1-D scheme produced vastly improved images compared to

IMAGE	SNR	Q2	TBR
GIRL	37.05 dB	3270	2.79 bits/pel
CARD	34.70 dB	3625	2.88 bits/pel
FLAT	36.97 dB	4037	2.98 bits/pel
SCENE	39.45 dB	1994	2.48 bits/pel
HALL	37.83 dB	2809	2.68 bits/pel
GIRL 2	39.49 dB	1666	2.40 bits/pel

Note: Q2=total number of blocks used by the quantizer
Q2 to encode the input signal
TBR=average bit rate per frame

Table 5.4. Results for the VSQ 1-D scheme under
'non ideal conditions'



(a)



(d)



(b)



(e)



(c)



(f)

Fig.5.4. Recovered images from the VSQ 1-D scheme

(a) GIRL (b) CARD (c) FLAT (d) SCENE

(e) HALL (f) GIRL 2

the FSQ 1-D scheme, fig.4.5 and fig.5.4.

5.3. Variable Bit-Rate Switched Quantization (VSQ)

Scheme (2-D)

The results of the 2-D scheme described in section 4.3 showed that four quantization levels were not enough to cope with the sharp transitions exhibited by some of our image. From section 5.2 it is clear that the new strategy provided substantial picture quality improvement at a reduced bit-rate when used with a first order predictor at the expense of a more complex scheme. Hence it was decided to use the same block adaptive strategy in a 2-D system with quantizers Q1 (1-bit) and Q2 (3-bit) designed with a Laplacian distribution. Since the switching information is carried one block of pels in advance, at each received block of digits decoding would be performed with respect to one of the two quantizers only. In such a situation it is possible to have a different set of examination and scrambling processes depending on which quantizer was used to decode the received digits. As Q1 was a 1-bit quantizer the scrambling of these quantized error signals was done by sign inversion.

-The scrambling of the quantized error signal from Q2, was as before, about the mid-point values of the quantizer's positive and negative half. The object of this type of scrambling is to produce a signal whose activity is greater than that of the original.

5.3.1. The Transmitter

The block diagram of the transmitter and receiver, fig.5.1, used in section 5.2 is applicable here with $\hat{X}_{i1}$, $\hat{X}_{is1}$, $\hat{X}_{i2}$ and $\hat{X}_{is2}$ substituted by $\hat{X}_{i1,1}$, $\hat{X}_{is1,1}$, $\hat{X}_{i2,1}$ and $\hat{X}_{is2,1}$ respectively. Also provided is the flow chart of the transmitter section, for this 2-D scheme, in fig.5.5. Note the similarities between fig.5.5 and fig.5.2. However, scrambling of e_{q1} is performed by inversion about the mid-point value of the quantizer's positive and negative half i.e. zero. All the predictors P are third order 2-D predictors of the type used in the FSQ 2-D scheme, section 4.3.1. Each video line (256 pels) is encoded in blocks of 8 pels. The main difference between the VSQ 1-D scheme, described earlier, and this VSQ 2-D scheme, is in part (a) of the transmitter and the corresponding receiver section.

To start processing the input video correctly at the transmitter and receiver it was decided to encode the first 8 pels of each video line using the second encoder i.e. Q_2 , 3-bit quantizer. Commencing with the first block of pels at the beginning of a video line :

(a) The processor considers the sequence $\hat{X}_{is2,1}$:

$$0 - B_1 > \hat{X}_{is2,1} > 255 + B_2 \quad (5.21)$$

where B_1 and B_2 are system constants

$\hat{X}_{is2,1}$ is the present locally decoded sequence

If any of the $\hat{X}_{is2,1}$ samples exceeds the limitations

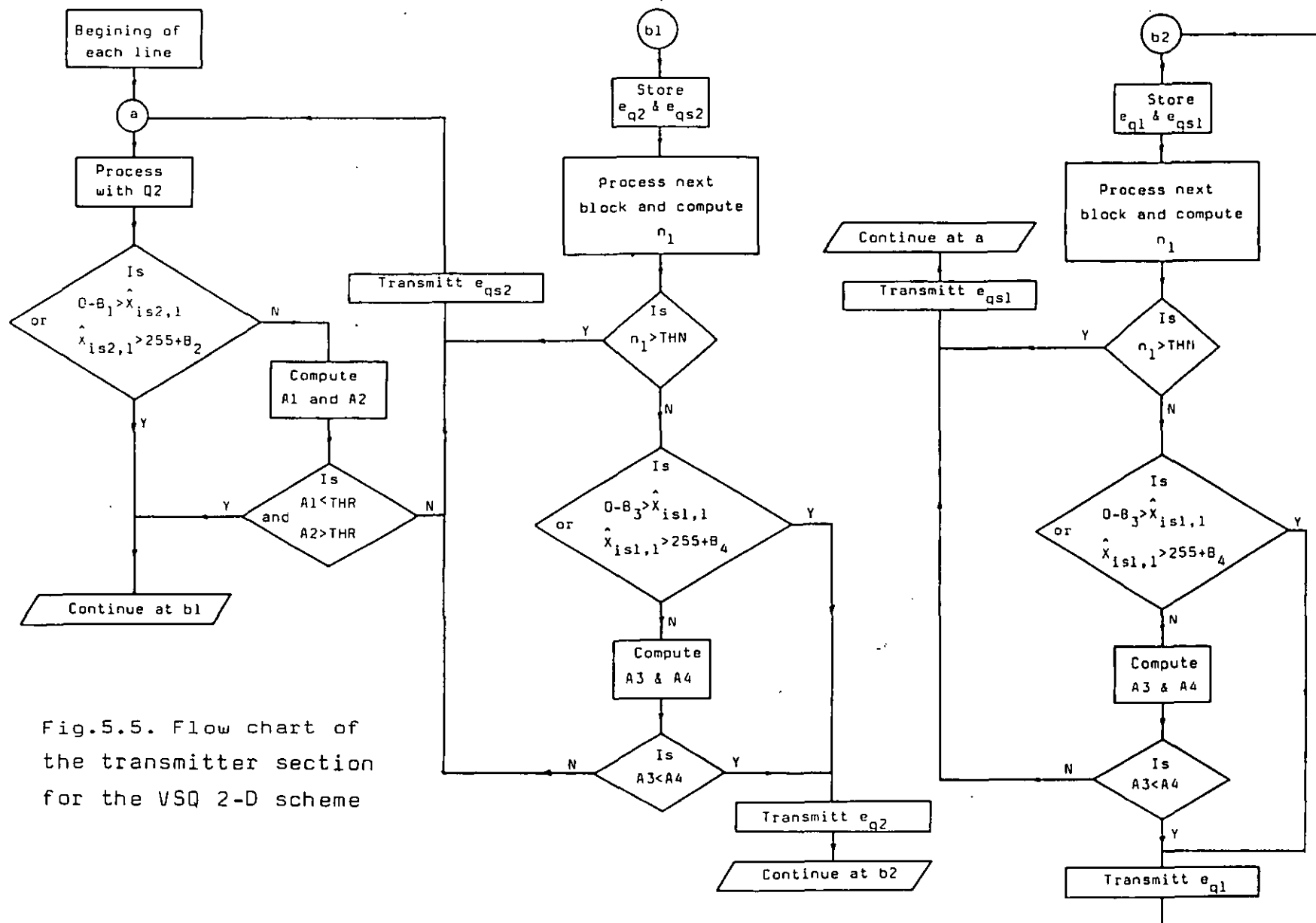


Fig.5.5. Flow chart of the transmitter section for the VSQ 2-D scheme

imposed by the inequalities of eq.(5.21), the transmitted sequence can be either from e_{q2} or e_{qs2} depending on the outcome of the next block of encoded pels. The next block of samples to be encoded continues to be examined from section (b), described later on in the text, or (b1) in fig.5.5.

However, if the sequence $\hat{X}_{is2,1}$ is confined to the inequalities of eq.(5.21) an activity measurement is performed on the samples $\hat{X}_{i2,1}$ and $\hat{X}_{is2,1}$ by :

$$A1 = \sum_{i=1}^N (\hat{X}_{i2,1} - \hat{X}_{i,2})^2 \quad (5.22)$$

$$A2 = \sum_{i=1}^N (\hat{X}_{is2,1} - \hat{X}_{i,2})^2 \quad (5.23)$$

where $\hat{X}_{i,2}$ represents the previous line of decoded pels

and if

$$A1 < THR$$

and (5.24)

$$A2 > THR$$

> the transmitted sequence can be either from e_{q2} or e_{qs2} depending on the next block of encoded pels. It was experimentally found that when eq.(5.24) is satisfied one can definitely differentiate between a scrambled and an unscrambled set of error samples by selecting the one with a lower activity. The way in which this was achieved will be described in the experimental procedure, section 5.3.3.

For other conditions of inequalities (5.24) the transmitted sequence would be from e_{qs2} and the next block of pels would be encoded using Q2 starting its examination from (a).

(b) Let us consider the next block of pels to be encoded. If this is the second block of pels to be encoded then the processor stores the previous sequences e_{q2} and e_{qs2} shown in the flow chart fig.5.5 from position (b1). The noise power of the first encoder is computed from :

$$n_1 = \sum_{i=1}^N (\hat{x}_{i1,1} - x_i) \quad (5.25)$$

This is compared to a noise threshold level :

$$n_1 > THN \quad (5.26)$$

If n_1 exceeds THN the second encoder is chosen for encoding this set of pels. Assuming that the present block of pels is the second block, the previous block of pels is transmitted as e_{qs2} . Further inspection of the present block of pels continues from section (a).

If however, n_1 lies below or is equal to THN, the sequence $\hat{x}_{i1,1}$ is further examined by the inequalities :

$$0 - B_3 > \hat{x}_{i1,1} > 255 + B_4 \quad (5.27)$$

where B_3 and B_4 are system constants

If any $\hat{X}_{isl,1}$ lie outside the inequalities defined in (5.27) it means that the signal will be encoded using the first encoder. Assuming that the present block is the second block of pels to be encoded the previous block is transmitted as e_{q2} . The examination of the following block, or third block continues from section (b) with the processor stored and possible transmitted values of e_{q2} and e_{qs2} substituted by e_{q1} and e_{qs1} . This is shown in the flow chart of fig.5.5 at position (b2). Note, from fig.5.5, that the examination of the encoded sample through paths (b1) and (b2) are the same.

If all $\hat{X}_{isl,1}$ lie within the bounds of the inequalities defined in eq.(5.18) inspection is performed on the two sequences $\hat{X}_{il,1}$ and $\hat{X}_{isl,1}$ by forming an activity measure :

$$A3 = \frac{1}{(N-1)} \sum_{i=1}^{N-1} (\hat{X}_{il,1} - \hat{X}_{il+1,1})^2 \quad (5.28)$$

$$A4 = \frac{1}{(N-1)} \sum_{i=1}^{N-1} (\hat{X}_{isl,1} - \hat{X}_{isl+1,1})^2 \quad (5.29)$$

If $A3 < A4$ the signal will be encoded using the first encoder (1-bit quantizer). Assuming that the present block is the second block of pels to be encoded the previous block is transmitted as e_{q2} . The examination of the following block, or third block, continues from section (b) with the processor stored and possible transmitted values of e_{q2} and e_{qs2} substituted by e_{q1} and e_{qs1} . This is shown in the flow chart of fig.5.5 at position (b2). Note, from fig.5.5, that the examination

of the encoded samples through path (b1) and (b2) are the same.

For any other case i.e. $A3 \geq A4$ the signal will be encoded using the second encoder (3-bit quantizer) continuing its examination from section (a).

Note :

Upon deciding which encoder is used in processing the video signal the contents of the tap delays, in that encoder, are transferred to the corresponding tap delays of all the other predictors. This ensures the same starting position for all predictors.

5.3.2. The Receiver

A brief description of the receiver is given in this section since it functions relatively similar to the transmitter section. Shown in fig.5.1 is the block diagram of the receiver section. Since the first block of each video line is encoded using the 3-bit quantizer, at the transmitter, it is therefore justified to use this initial starting condition at the receiver. It is also assumed that the receiver has knowledge about the system constants B_1, B_2, B_3, B_4, THR and knows what scrambling to perform when decoding is with respect to Q1 or Q2.

Commencing with the first block of pels in a video line, the two decoded sequences $\hat{Y}_{i1}$ and $\hat{Y}_{is1}$ are examined by the inequalities defined in eq.(5.21) with $\hat{X}_{is2,1}$ being substituted once by $\hat{Y}_{i1}$, and again by $\hat{Y}_{is1}$. If any $\hat{Y}_{i1}$ were found to lie outside the limitations of eq.(5.21) the recovered video signal would be taken

as $\hat{Y}_{isl}$ and the next block of pels is decoded with respect to Q2. In case any $\hat{Y}_{isl}$ were found to lie outside the limitations of eq.(5.21) the recovered video signal would be $\hat{Y}_{il}$, and the next block of pels is decoded with respect to Q1. Failing this the activity of $\hat{Y}_{il}$ and $\hat{Y}_{isl}$ is computed using equations defined by eqs.(5.22) and (5.23) with $\hat{X}_{i2,1}$, $\hat{X}_{i,2}$ and $\hat{X}_{is2,1}$ being replaced by $\hat{Y}_{il}$, $\hat{Y}_{i,2}$ and $\hat{Y}_{isl}$ respectively. If both of these activity measures are found to have a value above THR, the final video signal is taken as $\hat{Y}_{isl}$ and the decoding of the next block is performed with respect to Q2. It was experimentally found that the two just mentioned activity measures never both fall below THR. Otherwise the final video signal output is taken as the one with the lower activity. Assuming the sequence $\hat{Y}_{il}$ produced a lower measure of activity, it would be chosen as the final video output and the decoding of the next block would be with respect to Q1. Making a similar assumption for $\hat{Y}_{isl}$, it would be chosen as the final video output signal and the decoding of the next block is done with respect to Q1.

For the sake of completeness it is now necessary to consider the situation when the previous video output was from the sequence $\hat{Y}_{il}$. This situation can occur if the transmitter encoded and sent a block either as e_{q1} or e_{q2} . Detecting this state the receiver decodes the incoming block with respect to Q1. The so produced samples $\hat{Y}_{il}$, and $\hat{Y}_{isl}$ are examined by the inequalities defined in eq.(5.27) with $\hat{X}_{isl,1}$, being substituted once by $\hat{Y}_{il}$, and again by $\hat{Y}_{isl}$. If any $\hat{Y}_{il}$ are found to lie

outside the limitations imposed by eq.(5.27) the recovered video signal would be chosen as $\hat{Y}_{isl}$, and the next block would be decoded with respect to Q2. In case any $\hat{Y}_{isl}$, were found to lie outside the limitations of eq.(5.27) the recovered video signal would be $\hat{Y}_{il}$, and the next block of pels is decoded with respect to Q1. Failing this the activity of $\hat{Y}_{il}$, and $\hat{Y}_{isl}$, is computed using equations defined by eqs.(5.28) and (5.29) with $\hat{X}_{il,1}$, and $\hat{X}_{isl,1}$ being replaced by $\hat{Y}_{il}$, and $\hat{Y}_{isl}$, respectively. However, in this situation the processor is only looking for the set of samples producing a lower measure of activity since that will be the one chosen as the final video output.

5.3.3. Experimental Procedure and Results

The set of source pictures used in chapter IV, fig.4.3 are the same ones used here. The chosen predictors were all third order fixed predictors whose tap gains contained values :

$$a_h = \frac{7}{8}$$

$$a_v = \frac{3}{4}$$

$$a_d = \frac{-5}{8}$$

The quantizers Q1 and Q2 were of the type described by Max⁴³ with a Laplacian distribution. Previous experience, on selecting suitable scaling constants for the error signals being fed to the two quantizers, was used to

select the required scaling constants. Values of 5.0 and 12.0 were chosen as suitable scaling constants for quantizers Q1 and Q2.

Initially an experiment was carried out only with two encoders at the transmitter and the two decoders at the receiver. The decoder processor was given all the information necessary to recover the original signal i.e. it was operating under 'ideal conditions'. Minimization of the quantization noise power was the only constraint imposed upon such a system set-up. The results of Table 5.5 show the average bit rate achieved and the corresponding SNR values. It is interesting to note that quantizer Q1 was chosen on average about 5.5% of the total number of blocks processed. This resulted in a relatively high average number of bits/pel. Since one of the aims was to try and keep the average bit rate as low as possible, without reducing the perceptual picture quality, it was necessary to incorporate a noise power threshold parameter. This is the reason for having threshold THN (eqs.5.25 and 5.26) in the 2-D system proposed. However, it was first necessary to establish the values of the other system parameters, i.e B_1 , B_2 , B_3 , B_4 , and THR. During the process of the just mentioned experiment it was possible to plot the p.d.f. of the activity measures as defined by eqs.(5.22.) and (5.23). Preliminary values for B_1 , B_2 , B_3 , and B_4 were also obtained. Shown in fig.5.6 are the p.d.f.'s of the activity measures for the GIRL, CARD and FLAT pictures. Although the p.d.f. plots for the other pictures were obtained they are not

IMAGE	SNR	Q2	TBR
GIRL	38.15 dB	7590	2.85 bits/pel
CARD	36.50 dB	7877	2.92 bits/pel
FLAT	35.51 dB	7907	2.93 bits/pel
SCENE	36.93 dB	7592	2.85 bits/pel
HALL	35.68 dB	7809	2.91 bits/pel
GIRL 2	39.88 dB	7647	2.86 bits/pel

Note: Q2=total number of blocks used by the quantizer
Q2 to encode the input signal
TBR=average bit rate per frame

Table 5.5. Results for the VSQ 2-D scheme under
'ideal conditions'

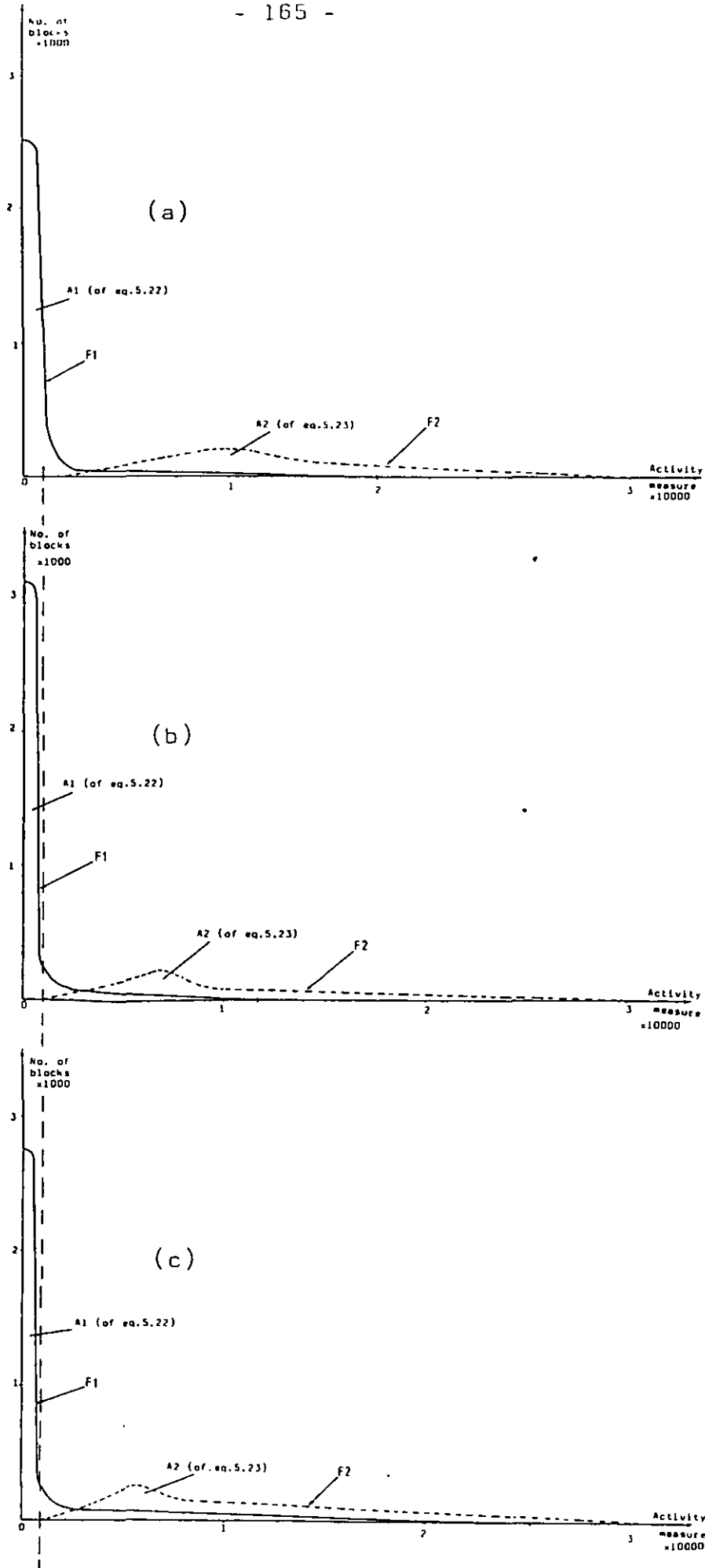


Fig.5.6. P.D.F. plots of the activity measure under 'ideal conditions' with THN = 2000 for the VSQ 2-D scheme (a) GIRL (b) CARD (c) FLAT

shown here since they were of a similar character. to the ones shown in fig.5.6. The value of THR was found by superimposing all the p.d.f. plots and taking the lowest point attained by the curves with respect to F2. This would ensure that one value of THR satisfied all the source images processed here. The chosen threshold yielded :

$$\text{THR} = 354.7$$

The selection of a value for THN was obtained with respect to the recovered picture's subjective quality. The system was run under the so called 'non ideal conditions', with THN included at the transmitter, as the imposing condition instead of the quantization noise power. For each variation of THN, the set of six source pictures were processed, in a 'dummy' run so that the correct values could be given to the rest of the system parameters, B_1 , B_2 , B_3 and B_4 . The six images were then processed again, with the images being recovered and viewed by experienced observers. This was carried on until a just noticeable difference occurred, between the source and the recovered image, on any one of the recovered images. Such an experimental procedure yielded a value for THN of :

$$\text{THN} = 2000.0$$

and the corresponding other system constants :

$$B_1 = 14.0$$

$$B_3 = 4.5$$

$$B_2 = 11.0$$

$$B_4 = 11.0$$

Presented in Table 5.6 are the SNR values attained under 'non ideal conditions' as well as the average bit rates achieved. The corresponding recovered images are shown in fig.5.7. Even though the value of THN was relatively high the final SNR results are also high which is somewhat surprising, since the introduction of THN in the system forces the use of quantizer Q1 in areas where quantizer Q2 should have been used. This curious result can be explained by the fact that the system forced the use of quantizer Q2 (improving the system's performance at an increased average bit rate) in areas where Q1 could have been used. The forced use of quantizer Q2 was on average 14% of the total number of blocks processed. On one hand this means that a further average bit rate reduction of 14% is possible, but on the other hand this reduction would be at the expense of a reduced SNR and a degradation in the recovered image subjective quality.

An experiment was also conducted with 4 pels per block and showed that under the present system set-up it was not possible to send the switching information one block of pels in advance. This was due to a total overlap of the activity measurement p.d.f. plots between the decoded pels and the scrambled decoded pels (the overlap was between the area F1 and F2, similar to the p.d.f.'s shown in fig.5.6). Experiments were conducted to show whether the scheme would function with a larger number of pels per block, e.g. 16 and 32 pels per block. It was anticipated and later verified that the scheme would

IMAGE	SNR	Q2	TBR
GIRL	33.59 dB	3079	1.78 bits/pel
CARD	34.30 dB	3902	1.95 bits/pel
FLAT	33.25 dB	4089	1.99 bits/pel
SCENE	33.65 dB	2113	1.51 bits/pel
HALL	32.88 dB	2851	1.69 bits/pel
GIRL 2	37.40 dB	2214	1.54 bits/pel

Note: Q2=total number of blocks used by the quantizer
Q2 to encode the input signal
TBR=average bit rate per frame

Table 5.6. Results for the VSQ 2-D scheme under
'non ideal conditions'



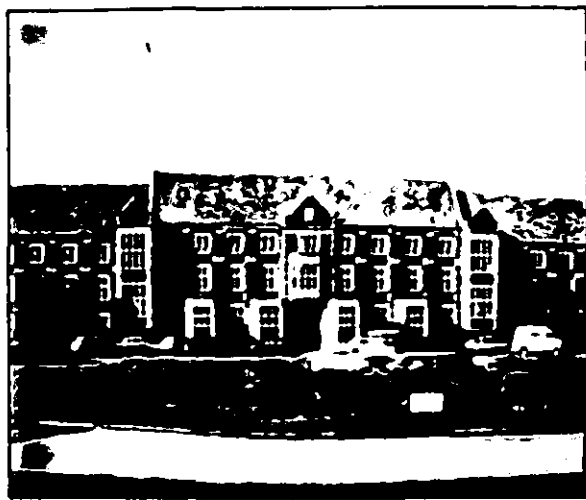
(a)



(d)



(b)



(e)



(c)



(f)

Fig.5.7. Recovered images from the VSQ 2-D scheme

(a) GIRL (b) CARD (c) FLAT (d) SCENE

(e) HALL (f) GIRL 2

function with a larger number of pels per block. However, the average SNR was about the same as in the case of 8 pels per block with an overall increase in the number of bits/pel as the number of pels per block increased. Using, for example 16 or more pels per block in this type of system set-up would mean a saving of 0.0625 bits/pel or less which would hardly justify such an elaborate scheme.

5.4. Discussion

The principle of switched quantization of chapter IV was used here to improve upon the SNR results and subjective picture quality previously obtained. This was achieved by designing two systems that operate in a block adaptive mode with two quantizers using a different number of levels. Inevitably a more complex system design was initiated with the necessity of a buffer memory control (not discussed in this thesis).

The results from the 1-D scheme present a clear indication that the proposed scheme is able to achieve relatively low bit rates, i.e. about 2.7 bits/pel, with extremely good perceptual image quality. In the opinion of some experienced viewers it was not possible to discriminate between the source images and the processed recovered images. Looking at the results of Table 5.3 'ideal conditions', and those of Table 5.4, 'non ideal conditions' it is clear that the proposed 1-D scheme performs exceptionally well. To achieve lower bit rates it would be possible to incorporate a noise threshold THN in a similar manner to that described in the 2-D scheme. However, since there is a small percentage,

about 2%, of the forced use of quantizer Q2 (4-bit), the reduction in bit rate would be at the expense of lower SNR values and lower subjective quality images. It is suggested that any further decrease in the number of bits/pel must be in the area of a 2-D prediction scheme.

The 2-D scheme proposed here presents some encouraging results. The results of Table 5.6., 'non ideal conditions', show that the reduction in the average bit rate was about 1.3 bits/pel, compared to those of Table 5.5, 'ideal conditions', while the decrease in SNR was about 3 dB. From the recovered images shown, in fig.5.7, it is clear that the proposed system produces excellent quality pictures irrespective of the picture content. Furthermore this is achieved at an average bit rate of 1.7 bits/pel. Any additional improvement in the recovered images and SNR values should be by processing the input signal in smaller blocks i.e. less than 8 samples per block. Although an attempt was made to process the input signal with 4 samples per block, using this particular VSQ set-up, it was unsuccessful. This could be remedied by processing the input signal as a two dimensional block of samples(2x2).

CHAPTER VI

CONCLUSION AND SUGGESTION FOR FUTURE WORK

6.1 Introduction

The field of video signal processing is now a rather mature one. The ease with which hardware engineers can currently handle video signals, making comparatively major concession to source signal statistics and perceptual observation, demonstrates that substantial improvements can be made in the field of video signal processing. The limits of image processing and system performance are ultimately set by how well the human visual perception is understood. If perception were thoroughly understood, the quality of a particular video processing system could be ascertained by objective measurement of signal parameters. Our inability to do this, implicit in our ignorance of perception, is a serious impediment both to communication between research scientists and to the video signal processing itself. However, the day when we can, with confidence, objectively evaluate visual impairment without recourse to subjective testing seems very remote.

Through the work presented in this thesis it was shown how further use of scrambling can be achieved in the field of video signal processing. The technique was applied to analogue and digital image transmission. In order to demonstrate the idea an easy signal modification or scrambling process was employed. Using a scrambling technique in video signal transmission, both analogue and digital, inevitably means that a certain amount of

of privacy is achieved against unauthorized receivers. However, this privacy of transmission is accomplished as a by-product of the idea.

6.2. Embedding Data into Pictures by Modulo Masking

Three systems for embedding data into industrial quality analogue monochrome pictures have been proposed. These are referred to as System 1, 2 and 3. The systems use the picture signal as a data carrier, and the combined picture-data signal does not require a greater bandwidth than that of the original signal.

The video signal on each scan-line was sampled, and data bits were inserted into a block of 3 or 5 pels by modulo masking scrambling the luminance level of only one pel in the block. Prior to transmission the combined data and video sequence was converted into a continuous signal.

Experiments were carried out on six images covering a wide range of different video signal statistical properties. Systems 1 and 2 conveyed the data at a variable bit stream while System 3 conveyed the data at a constant bit stream. In System 1 and 2 highly active pel blocks are excused from conveying data and the recovered picture quality was greater than that obtained by System 3. The average number of pels required for transmission of one bit of data was 3.76, 6.75, and 3 for Systems 1, 2 and 3 respectively. Although Systems 1 and 2 were able to produce better quality recovered pictures than System 3, they were less robust to bit errors. Thus each system has its strengths and weaknesses,

and the system to be deployed depends on the application.

It must be emphasized that the systems were designed such that no bit errors were detected when the six images were processed by the Systems 1, 2 and 3. Other images may cause bit errors to be received but this can be remedied by introducing sufficient channel protection coding bits at the expense of the transmitted data bit rate. Thus with the recent advances in VLSI technology one could envisage a video telephone service conveying data of the order of about half a mega-bit per second, or the equivalent of approximately eight digitized voiced signals, in addition to the picture signal.

Furthermore it might be possible to transmit data over analogue video signals at higher rates, maintaining good quality recovered pictures, especially for System 1 and 2 by performing selectively two or more scrambling operations simultaneously on each block of pels. In other words each block of pels would undergo two or more signal modifications where each signal modification would correspond to a different set of data sequence. If this is to be performed selectively on a block of lower activity higher transmission data rates could be achieved with an acceptable recovered image fidelity.

6.3. Switched Quantization - Fixed Bit Rate (FSQ)

The embedding data work reported was basically on system design and development in the field of analogue video transmission. This provided a better familiarization and understanding on the behaviour of video signals, and

helped to develop novel systems in the field of waveform encoding. It was soon realized that, due to the highly nonstationary behaviour of video signals, to obtain good quality pictures a system using feed forward estimation should be used. However, the difficulty with a feed forward estimation system is that at lower transmission rates the necessary additional side information can consume a significant proportion of the overall transmission bit rate.

Two DPCM systems were therefore developed and proposed, employing switched quantization. The two systems operate with a constant bit rate stream and use 1-D and 2-D predictors. The basic idea originated from the analogue systems previously described. By scrambling or not scrambling the quantizer output levels it was possible to transmit a switching type code to the decoder, without the necessity of actually sending any side information, in order to switch between two quantizers. This switching code, used in a feed forward estimation mode, resulted in a more efficient tracking of the input video signal by the codec.

Six pictures were used to test the performance of the FSQ 1-D and FSQ 2-D schemes so that a wide variety of picture content was covered. The quantizer switching information was carried in one dimensional blocks of 8 quantized error signals by scrambling/or not every sample in the sequence. Hence each block of error signals carried its own information as to which quantizer was used in encoding the source input signal. This presented

some difficulty in the decoding and detection process since it was necessary to devise a measurement technique to cover both possibilities i.e. if the block was encoded with one or the other quantizer.

The results from the FSQ 1-D scheme showed that an SNR improvement of up to 8 dB's was possible, compared to the DPCM scheme where a Laplacian quantizer was used, optimized per frame using the mean square error criterion⁴³, and operating at the same bit rate i.e. 3 bits/pel. A similar comparison was made with the FSQ 2-D scheme and showed an average SNR improvement of only 1.8 dB operating at 2 bits/pel. The subjective quality of the recovered images from the FSQ 1-D scheme and the FSQ 2-D scheme were of poor quality. Although the 2-D scheme showed improved picture quality over the 1-D scheme this was mainly due to the use of 2-D predictors.

6.4. Switched Quantization - Variable Bit Rate (VSQ)

Even though the FSQ schemes were not very successful in producing good quality recovered images it did provide the frame-work for the design and developement of two switched quantization schemes with a variable bit stream. The results from the FSQ schemes also imdicated that a fixed bit rate system was not able to cope well with the sharp transitions exhibited by some of the source images. Hence the FSQ schemes were modified to use two quantizers with a different number of levels, both for the 1-D and 2-D predictors.

Since major parts of an image do not contain sharp transients between pels, they were coded with fewer

quantization levels. However, at sharp transients more quantization levels were allocated. This resulted in a lower average number of bits/pel with 1-D and 2-D predictors compared to the FSQ 1-D and FSQ 2-D schemes previously discussed. It was initially anticipated and later verified that this strategy would produce very good quality pictures.

One dimensional blocks of 8 pels were encoded with respect to the appropriate quantizer (of two quantizers) and the necessary decoding information was carried by the previously encoded and transmitted block of pels. In such a situation it was possible to have a different set of detection measurements at the receiver depending on which quantizer was previously used to encode the transmitted block of pels. To avoid sending any side information even at the beginning of the image signal it was decided to encode the very first block of an image signal with a predetermined encoder. This was also repeated at the beginning of each video line to make sure that any error propagation was, hopefully, confined to a video line only.

The results from the VSQ 1-D scheme indicate that the proposed scheme, although more complex than the FSQ 1-D scheme, was able to recover images of exceptionally good perceptual quality. Furthermore this was achieved at an average bit rate of about 2.7 bits/pel. A further reduction in the average bit rate would be possible at the expense of a degraded image quality and therefore this was not pursued.

Due to the efficiency of the 2-D predictors the proposed VSQ 2-D scheme was only a little more complex to impelment than the VSQ 1-D scheme. However the recovered image quality, of the VSQ 2-D scheme, was no less impressive compared to that of the VSQ 1-D scheme. In addition the VSQ 2-D scheme achieved this very good image quality with an average transmitted bit rate of about 1.7 bits/pel.

It must be emphasized that the schemes were designed so that no errors occurred, in the absence of errors generated by the channel, in the quantization switching mode for six images processed. Other images may present difficulties to the schemes mentioned above however this can be overcome by readjusting the system parameters.

Further improvements to the idea of a feed forward estimation system without the necessity of transmitting any side information must clearly proceed from the VSQ 2-D scheme by developing a scheme to process the input video signal in smaller blocks of pels. Although an experiment was conducted with the VSQ 2-D scheme on a one dimensional block of 4 pels it was unsuccessful. It is strongly anticipated that a similar scheme, such as the VSQ 2-D scheme, coupled with a two dimensional (2x2) block processing of the input video signal would yield improved results in terms of SNR and the average transmitted bit rate.

Due to the constraints imposed by the facilities and time available, the research program was carried out on monochrome images only. However the work portrayed could easily be extended to process colour images and specifically moving images. The computer simulated systems and results obtained confirm the notion that significant bit rate reductions are possible by making use of scrambling strategies.

Further work must certainly progress in the direction of three dimensional interframe predictive coding. However, even though video signals exhibit highly non-stationary characteristics it is difficult to foresee any major improvements in the area of adaptive predictors for interframe coders. The currently available three dimensional fixed predictors are quite adequate to cope with scenes of moving images. So far in most adaptive interframe predictive coding work available in literature no systematic investigation of adaptive quantizers has been made. This is primarily due to the lack of understanding of visual perception as well as the perceptual basis for adapting the quantizers. Since the human observer is the end user of an image processing system, a quantizer design based on a subjective criteria would increase the observed fidelity of the processed image. Thus a possible area of research would be to improve the modelling of the human visual system and the design of quantizers based on such models.

REFERENCES

1. P.S.Cart and G.B.Townsend, "*Colour Television*", Vol 1: "*The NTSC System, Principles and Practice*", 1969, Vol 2: "*Pal, Secam and Other Systems*", 1969, London (Illiffe) Newnes-Butterworth.
2. H.V.Sims, "*Principals of Pal Colour Television and Related Systems*", London (Illiffe) Newnes-Butterworth, 1969.
3. D.E.Pearson, "*Transmission and Display of Pictorial Information*", Pentech Press, 1975.
4. D.G.Fink, "*Television Engineering Handbook*", McGraw-Hill, 1975.
5. C.W.B.Reis, "*Technical Reference Book*", IBA Technical Review, No.2, May 1977 (3rd edition).
6. A.Habibi and G.S.Robinson, "*A Survey of Digital Picture Coding*", Computer (IEEE Computer Soc. Magazine), Vol.7, No.5, pp 22-34, May 1974.
7. T.S.Huang and D.J.Tretiak, (Eds) "*Picture Bandwidth Compression*", Proc. IEEE Symposium, MIT, Cambridge, April 1969, Gordon & Breach, 1972.
8. J.O.Limb, C.B.Rubinstein and J.E.Thompson, "*Digital Coding of Colour Video Signals - A Review*", IEEE Trans. on Commun., Vol. COM-25, No.11, pp 1349-1385, November 1977.
9. A.N.Netravali and J.O.Limb, "*Picture Coding - A Review*", IEEE Trans. on Commun., Vol. COM-68, No.3, pp 366-406, March 1980.

10. Several Authors - IEEE Special Issue on, "*Redundancy Reduction*", Proc. IEEE, Vol. 55, No.3, (C.C.Culter), March 1967.
11. Several Authors - IEEE Special Issue on, "*Signal Processing for Digital Communications*", IEEE Trans. on Commun. Techn., Vol. COM-19, No.6, (M.R.Aaron), December 1971.
12. Several Authors - IEEE Special Issue on, "*Digital Picture Processing*", Proc, IEEE Vol. 60, No.7, (H.C.Andrew and L.H.Enloe), July 1972.
13. Several Authors - IEEE Computer Soc. Special Issue on, "*Image Processing*", Computer, Vol. 7, No.5, May 1974.
14. Several Authors - IEE Special on, "*Vision and Sound*", IEE Proc. Vol. 129, No.7, September 1982.
15. C.A.Andrews, J.M.Davis and G.R.Schwartz, "*Adaptive Data Compression*", Proc. IEEE, Vol. 55, No.3, pp 267-277, 1967.
16. N.S.Jayant, "*Waveform Quantization and Coding*", IEEE Press, 1976.
17. L.R.Rabiner and B.Gold, "*Theory and Application of Digital Signal Processing*", Prentice-Hill, 1975.
18. J.C.Simon and A.Rosenfeld, "*Digital Image Processing and Analysis*", NATO Advanced Study Institute Series, 1977-Norhoff Internation Publishing.
19. R.C.Gonzales and P.A.Wintz, "*Digital Image Processing*", Addison-Wesley, 1977.

20. W.K.Pratt, *"Digital Image Processing"*, John Wiley & Son, 1978.
21. H.Hotelling, *"Analysis of a Complex of Statistical Variables into Principal Components"*, J. Educ. Psychol., Vol. 24, pp 417-441, 1933.
22. N.Ahmed, T.Natarjan and K.Rao, *"On Image Processing and a Discrete Cosine Transform"*, IEEE Trans, on Commun., Vol. 23, pp 90-93, January 1974.
23. H.C.Andrews, J.Kane and W.K.Pratt, *"Hadamard Transform Image Coding"*, Proc. IEEE, Vol. 57, pp 58-68, January 1969.
24. W.K.Pratt, W.H.Chen and L.R.Welch, *"Slant Transform Image Coding"*, IEEE Trans. on Commun., Vol. 22, pp 1075-1093, August 1974.
25. A.Haar, *"Zur Theorie der Orthogonalen Functionen-System"*, Inaugural Dissertation, Math. Aunalen, Vol. 5, pp 17-31, 1955.
26. W.H.Chen and C.H.Smith, *"Adaptive Coding of Monochrome and Colour Images"*, IEEE Trans. on Commun., Vol. COM-25, No.11, November 1977.
27. A.H.Reeves, French Pat. 852183 (1938), British Pat. 535860 (1939), U.S.A. Pat. 2270070 (1942)
28. K.W.Cattermole, *"Principles of Pulse Code Modulation"*, London Illiffe, 1973.
29. W.M.Goodal, *"Television Transmission by Pulse Code Modulation"*, BSTJ, Vol. 30, pp 33-49, January 1951.
30. R.L.Cabrey, *"Video Transmission Over Telephone Cable Pairs by Pulse Code Modulation"*, Proc. IRE, Vol. 48, No.9, pp 1546-1561, September 1960.

31. L.H.Harper, *"PCM Picture Transmission"*, IEEE Spectrum, Vol. 3, pp 146, June 1966.
32. V.G.Devereux, *"Pulse Code Modulation of Video Signals : 8-Bit Coder and Decoder"*, BBC Report EL-42, 1970/25.
33. R.E.Fletcher, *"A Video Analogue to Digital Converter"*, IBC 1974, IEE Conf. Publication No.119, pp 47-57.
34. C.C.Cutler, *"Differential Quantization of Communication Signals"*, U.S. Patent No. 2605361, July 1952 (Applied for in June 1950).
35. C.W.Harrison, *"Experiments with Linear Prediction in Television"*, BSTJ Vol. 31, July 1952, pp 764-783.
36. P.Elias, *"Predictive Coding"*, IRE Trans. Inform. Theory IT-1, March 1955, pp 16-33.
37. R.E.Graham, *"Predictive Quantization of TV Signals"*, Wescon Convention Record, Part 4, August 1958, pp 147-156.
38. J.B.O'Neal,Jr., *"Predictive Quantization Systems (Differential Pulse Code Modulation) for the Transmission of Television Signals"*, BSTJ Vol. 45, No.5, May-June 1966, pp 689-721.
39. D.J.Connor, R.F.W.Pease and W.G.Scholes, *"Television Coding Using Two-Dimensional Spatial Prediction"*, BSTJ Vol. 50, March 1971, pp 1049-1061.
40. J.C.Candy and R.H.Bosworth, *"Method for Designing Differential Quantizers Based on Subjective Evaluation of Edge Busyness"*, BSTJ Vol. 51, No.7, 1972, pp 1495-1516.

41. D.K.Sharma and A.N.Netravali, "*Design of Quantizers for DPCM Coding of Picture Signals*", IEEE Trans. Commun., Vol. 25, November 1977, pp 1267-1274.
42. R.F.Panter and W.Dite, "*Quantization Distortion in Pulse Count Modulation with Non-Uniform Spacing of Levels*", Proc. IRE, Vol. 39, Jan. 1951, pp 44-48.
43. J.Max, "*Quantization for Minimum Distortion*", IRE Trans. Inform. Theory, Vol. IT-6, March 1960, pp 7-12.
44. K.Nitadori, "*Statistical Analysis of DPCM*", Electron. Commun., Vol. 48, Feb. 1965.
45. J.B.O'Neal, Jr., "*Predictive Quantization Systems (Differential Pulse Code Modulation) for the Transmission of Television Signals*", BSTJ, Vol. 45, No.5 (May-June 1966), pp 689-721.
46. D.E.Pearson, "*Quantization of Composite Colour Television Signals*", Proc. of Conf. on Digital Process of Signals in Commun., Loughborough, Vol. 49, 1980, pp 413-426.
47. A.N.Netravali and B.Prasada, "*Adaptive Quantization of Picture Signals Using Spatial Masking*", Proc. of IEEE, Vol. 65, No.4, April 1977, pp 536-548.
48. P.Pirsch, "*Design of DPCM Quantizers for Video Signals Using Subjective Tests*", IEEE Trans. Commun., Vol. COM-29, No.7, July 1981, pp 990-1000.
49. A.N.Netravali and C.B.Rubinstein, "*Quantization of Colour Signals*", Proc. IEEE, Vol. 64, August 1977, pp 1177-1187.

50. T.N.Cornsweet, *"Visual Preception"*, Academic Press, New York 1970.
51. G.Wald, *"The Receptors for Human Colour Vision"*, Science, Vol. 45, 3636, September 1964, pp 1007-1017.
52. M.H.Pirenne, *"Vision and the Eye"*, 2nd. ed. Associated Book Publishers, London 1967.
53. P.K.Brown and G.Wald, *"Visual Pigment in Single Rods and Cones of the Human Retina"*, Science, Vol. 44, April 3, 1964, pp 45-52.
54. R.L.Gregory, *"Eye and Brain, the Psychology of Seeing"*, Wiedenfeld and Nicolson, 1966.
55. W.F.Schreiber, *"Picture Coding"*, Proc. IEEE, Vol.55 No.3, pp 320-330, March 1967.
56. A.J.Seyler, *"The Coding of Visual Signals to Reduce Channel-Capacity Requirements"*, Proc. IEEE, Vol. 109, pp 676-684, July 1962.
57. D.E.Pearson, *"Methods for Scaling Television Picture Quality"*, Picture Bandwidth Compression, ed. by T.S.Huang and O.J.Tretiak, Gordon and Breach, New York, 1972, pp 47-95.
58. P.Mertz, D.Fowler and H.N.Christopher, *"Quality Rating of Television Images"*, Proc. IRE, Vol. 38, No.11, 1950, pp 1269-1283.
59. G.L.Fredenhall and W.L.Behrend, *"Picture Quality - Procedure for Evaluating Subjective Effects of Interference"*, Proc. IRE, Vol. 48, No.6, 1960, pp 1030-1040.
60. R.D.Prosser, J.W.Allnatt, and N.W.Lewis, *"Quality Grading of Impaired Television Pictures"*, Proc. IEE, Vol. 111, No.3, 1964, pp 491-502.

61. R.Steele and D.Vitello, "*Simultaneous Transmission of Speech and Data Using Code-Breaking Techniques*", BSTJ, Vol. 60, No.9, November 1981, pp 2081-2105.
62. R.Steel and D.Vitello, "*Embedding Data in Speech Using Scrambling Techniques*", IEEE ICASSP-82, Paris, May3-5, 1982, Session DSP14.
63. W.C.Wong, R.Steele and C.S.Xydeas, "*Transmitting Data on the Phase of Speech Signals*", to be published in BSTJ
64. E.F.Brown and A.N.Netravali, "*Technique for Transmitting Digital Data Together with a Video Signal*", U.S.Patent 4,237,484, December 2, 1980.
65. K.Feher, R.Goulet and S.Morris, "*1,544 Mbit/s Data Above FDM Voice and Data Under FDM Voice Microwave Transmission*", IEEE Trans. on Commun., Vol. COM-23, No.11, November 1975, pp 1321-1327.
66. K.Feher and S.Morris, "*Simultaneous Transmission of Digital Phase-Shift Keying and of Analogue Television Signals*", IEEE Trans. on Commun., Vol. COM-23, No.12, pp 1509-1514.
67. G.B.Lockhart and Y.O.Al-Jalili, "*Method for Superimposing Data on Amplitude-Modulated Signals*", Electronic Letters, Vol. 18, No.9, April 29, 1982, pp 379-380.
68. T.L.Lim and M.S.Mueller, "*Adaptive Equalization and Phase Tracking for Simultaneous Analogue/Digital Data Transmission*", BSTJ, Vol. 60, No.9, November 1981, pp 2039-2063.

69. F.Akashi, Y.Sato and M.Eguchi, *"High-Speed Digital and Analogue Parallel Transmission Technique Over Single Telephone Channel"*, IEEE Commu., Vol.COM-30, No.5, May 1982, pp 1213-1218.
70. A.Habibi, *"Comparsion of n'th-Order DPCM Encoder with Linear Transformations and Block Quantization Techniques"*, IEEE Trans. Commun. Technol., Vol. COM-19, No.6, pp 948-956, December 1971.
71. W.Zschunke, *"DPCM Picture Coding With Adaptive Prediction"*, IEEE Trans. Commun. Technol., Vol. COM-25. No.11, pp 1295-1302, November 1977.
72. I.Dukhovich and J.B.O'Neal,Jr., *"A Three Dimensional Spatial Non-Linear Predictor for Television"*, IEEE Trans. Commun. Technol., Vol. COM-26, No.5, pp 578-583, May 1978.
73. B.Girod, *"Adaptive Prediction for DPCM Coding of TV Signals"*, IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-29, No.6, December 1981.
74. P.Pirsch and L.Stenger, *"Statistical Analysis and Coding of Colour Video Signals"*, Acta Electronica, Vol. 19, No.4, pp 277-287, 1976.
75. T.R.Lei, N.Scheinberg and D.L.Schilling, *"Adaptive Delta Modulation Systems for Video Encoding"*, IEEE Trans. Commin. Technol., Vol. COM-25, No.11, pp 1302-1314, November 1977.
76. A.H.Frei, H.R.Schindler and P.Vettiger, *"An Adaptive Dual-Mode Coder/Decoder for Television Signals"*, IEEE Trans. Commun. Technol., Vol. COM-19, No.6, pp 933-944, December 1971.

77. 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.
78. 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. Technol., Vol. COM-26, No.5, pp 573-578, May 1978.
79. K.Virupaksha and J.B.O'Neal, Jr., *"Entropy-Coded Adaptive Differential Pulse Code Modulation (DPCM) for Speech"*, IEEE Trans. Commun. Technol., Vol. COM-22, No.6, pp 777-787, June 1974.
80. B.Prasada, A.N.Netravali and A.Kobran, *"Adaptive Companding of Picture Signals in a Predictive Coder"*, IEEE Trans. Commun. Technol., Vol. COM-26, No.1, pp 161-164, January 1978.
81. S.Ericsson, *"Interframe Predictive Picture Coding with Delayed Decision"*, TRITA-TTT, November 1980.
82. N.S.Jayant, *"Adaptive Quantization With a One-Word Memory"*, Bell Syst. Tech. J., Vol-52, No.7, pp 1119-1141, September 1973.
83. L.H.Zetterberg, S.Ericsson and H.Brusewitz, *"Interframe DPCM with Adaptive Quantization and Entropy Coding"*, IEEE Trans. Commun. Technol., Vol. COM-30, No.8, pp 1888-1899, August 1982.
84. L.H.Zetterberg, S.Ericsson and C.Couturier, *"DPCM Picture Coding with Two-Dimensional Control of Adaptive Quantization"*, Submitted for publication to IEEE Trans. Commun. Technol., (1982).

