Compression of Integral Three-Dimensional Television Pictures

Matthew Charles Forman

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

De Montfort University
September 1999

ABSTRACT

Three dimensional image and video communication systems offer significant benefits over conventional two dimensional systems. The presence of natural depth information in the scene affords the viewer an overall improved sense of reality and naturalness. A variety of systems attempting to reach this goal have been designed by many independent research groups, though the images displayed by these stereoscopic systems lack a sense of solidity and have been shown to cause viewer fatigue due to a physiological conflict resulting from supplying a different disparate planar image to each eye.

A system developed at De Montfort University, integral three dimensional imaging, overcomes these limitations. It encodes a volume spatial model of an object scene on a plane surface, and is capable of real-time capture and display. The resolution requirements, however, are significantly greater than those of a conventional two dimensional display of a similar size. Hence for transmission via existing broadcast methods a reduction of the bandwidth required is necessary.

The objective of this thesis is the development of compression systems optimised for use with the intensity distributions recorded by the integral 3D imaging system. Such a compression system must be designed such that the transmission bandwidth consumed by the data stream resulting from compression of the captured integral 3D image data is no greater than that available within a conventional television broadcast channel, while ensuring that the reconstructed decoded image contains no objectionable distortion noticeable to the viewer. The thesis also develops procedures for capture, pre-processing, post-processing and display for evaluation of the compression schemes.

ACKNOWLEDGEMENTS

Firstly I wish to thank my supervisors Dr. Amar Aggoun and Prof. Malcolm McCormick for invaluable support and guidance throughout the period of my research.

I would like to thank Mr. Neil Davies and my fellow postgraduate members of the 3D Imaging Group for many helpful discussions and inspiring arguments.

I gratefully acknowledge the financial support of the Engineering and Physical Sciences Research Council.

Thanks to my friends for the providing the encouragement when it was needed.

Last but not least, I am deeply grateful to my father for encouraging and supporting my education to the level of PhD.

STATEMENT OF ORIGINALITY

The work contained within this thesis is purely that of the author unless otherwise stated. This work has not been submitted by any party for any examination, qualification or publication.

Matthew C. Forman
September 1, 2000, Leicester.

CONTENTS

1. Introduction 1

1.1 Overview 1

1.2 Scope of the Thesis 2

1.3 Outline of the Thesis 3

2. Three Dimensional Imaging for Television 5

2.1 Integral (Holoscopic) Imaging 5

2.1.1 Classical Integral Photography 5

2.1.2 Real-time Holoscopic Integral Image Capture 7

2.1.3 Electronic Capture of Integral Images 11

2.1.4 Electronic Display of Integral Images 12

2.1.5 The Lenticular-Integral Image 13

2.2 Alternative Three Dimensional Imaging Methodologies 14

2.2.1 Holography 15

2.2.2 Lenticular Stereoscopic Imaging 15

2.2.3 A Comparison of Autostereoscopic 3D Imaging Systems for Television 16

3. Communication and Compression of Image Data 19

3.1 Digital Video Image Representation 19

3.1.1 Colour encoding 20

3.1.2 Spatial and temporal sampling characteristics 20

3.2 Radio Frequency Broadcast of Digital Images 21

3.2.1 Digital signal modulation 21

3.2.2 Digital television bit rate reduction 22

3.3 Data Compression 22

3.3.1 A generic compression system model 23

3.3.2 Image data compression techniques 24

3.3.3 The JPEG system [1, 2] 28

4. Acquisition, Preparation and Replay of Integral Image Data 33

4.1 Integral Image Acquisition . . . 33
4.2 Intensity Distribution Sampling . . . 34
4.2.1 Use of Block-orientated Processing Techniques . . . 34
4.3 Electronic Scanning of the Intensity Distribution Transparency . . . 35
4.3.1 Scanning parameters . . . 35
4.3.2 Test image scanning . . . 36
4.4 Correction of Geometric Distortions in Scanned Intensity Distribution Data 37
4.4.1 Rotational correction . . . 37
4.4.2 Scale correction by direct examination . . . 39
4.5 Noise Reduction in Scanned Intensity Distribution Data . . . 40
4.5.1 Frequency domain noise reduction filter design . . . 41
4.5.2 Use of one dimensional spatial filtering . . . 46
4.6 Three Dimensional Display of Processed Integral Intensity Distributions . . 49
4.7 Evaluation of Replayed Image Fidelity . . . 50
4.8 An Alternative Scale Correction Method using Moire Analysis . . . 51

5. Evaluation of the Characteristics of Captured Integral Image Data . . . 56
5.1 The Captured Intensity Distribution . . . 56
5.2 Inter-band Correlation . . . 58
5.2.1 Evaluation of Inter-band Correlation . . . 58
5.3 Intra-band Correlation . . . 60
5.3.1 Evaluation of Intra-band Correlation . . . 60

6. An Integral 3D Image Compression Scheme using a Hybrid DPCM/DCT Encoder 64
6.1 Hybrid DPCM/DCT Mapping for 2D Video Encoding . . . 64
6.2 Mapper Design for Integral 3D Image Compression . . . 66
6.2.1 Inter-band decorrelation . . . 67
6.2.2 Intra-band decorrelation . . . 67
6.3 Quantisation of the Mapped Subimage Data . . . 69
6.3.1 Quantisation table selection . . . 69
6.4 Entropy Coding of the Quantised Subimage Coefficients . . . 72
6.5 The Base Hybrid DPCM/DCT Encoder and Decoder . . . 73
6.6 Enhancement to Reduction of Vertical Intra-band Correlation . . . 73
6.7 Implementation of the Hybrid Schemes for Simulation . . . 75
6.8 Global Huffman Code Generation . . . 80
6.9 Simulation Results . . . 83
6.9.1 Coding of luma channel only . . . 83
6.9.2 Full colour coding . . . 90

6.9.3 Reconstructed intensity distributions 95

7. A 3D-DCT-based Integral 3D Image Compression Scheme 110
7.1 The 3D-DCT Mapping Stage 110
7.2 Quantisation of 3D-DCT-decorrelated Subimage Groups 112
7.2.1 Quantisation array selection 112
7.3 Entropy Coding of Quantised Coefficient Groups 113
7.3.1 Per-plane zigzag scanning and coding 113
7.3.2 All-plane, per-row scanning and coding 113
7.4 The Complete 3D-DCT Scheme Encoder and Decoder 117
7.5 Implementation of the 3D-DCT Scheme for Simulation 118
7.6 Global Huffman code generation 122
7.7 Simulation Results 123
7.7.1 Coding of luma channel only 123
7.7.2 Full colour coding 128
7.7.3 Reconstructed intensity distributions 131

8. Efficient Quantisation in the 3D-DCT-based Compression Scheme 145
8.1 Statistical Analysis of 3D-DCT Coefficients 145
8.2 Use of Optimisation Techniques in Quantisation Array Design 146
8.2.1 Simulated annealing 149
8.2.2 Problem formulation: The parameter space 151
8.2.3 Problem formulation: The cost function 153
8.2.4 The optimisation procedure 155
8.2.5 Single-point optimisation example 155
8.2.6 Development of a multi-point quantisation array generation method 158

9. Conclusions 169
9.1 Research Undertaken 169
9.1.1 Acquisition and Pre-processing of Integral 3D Image Data 169
9.1.2 The Nature of the Digital Integral 3D Image Representation 169
9.1.3 Coding Techniques Developed for Use with Integral 3D Image Data 170
9.1.4 Coding Scheme Quantisation Stage Optimisation 170
9.2 Suggestions for Further Work 173
9.2.1 Use of Microlens Captured Full Integral Image Data 173
9.2.2 Addition of Local Image-Adaptive Features 174
9.2.3 Extension to Motion Video Integral Image Data 174
9.2.4 Design Prototyping of the 3D-DCT based Scheme 175

9.2.5 Integration of Application-Orientated Data Structures and Syntax . 175

Appendices 176

A. Test Integral Image Intensity Distributions 177

B. Conventional Test Images for Statistical Comparisons 180

C. Publications 183

References 184

LIST OF TABLES

2.1 Comparison of 3D Imaging Systems for Television Applications 17

3.1 Digital television picture sampling characteristics. 20
3.2 Standard JPEG quantisation matrices. 29

4.1 Physical characteristics of test integral intensity distributions. 34

5.1 Mean normalised cross correlation evaluated between adjacent bands on the luma channel of four lenticular-integral image intensity distributions. 59

6.1 Hybrid DPCM/DCT-based international video compression standards. . . . 65
6.2 Global Huffman code length table for luma channel data in the base hybrid DPCM/DCT scheme. 81
6.3 Summary of performance for luma channel-only hybrid DPCM/DCT based coding of four lenticular-integral intensity distributions. 89
6.4 Summary of performance for full colour hybrid DPCM/DCT based coding of three lenticular-integral intensity distributions. 95

7.1 Summary of performance for luma channel-only 3D-DCT based encoding of four lenticular-integral intensity distributions. 128
7.2 Summary of performance for full colour 3D-DCT based encoding of three lenticular-integral intensity distributions. 132

8.1 Constraints and initial conditions for the quantisation array generating function. 153
8.2 Final parameter values for the quantisation array generating function after an ASA run with $R_0 = 0.250061$ bpp. 157
8.3 Comparison of objective fidelity between use of fixed quantisation and optimal quantisation arrays. 160

LIST OF FIGURES

2.1 Recording of an integral photograph. 6
2.2 Replay of an integral photograph. 6
2.3 Production of a second-stage integral photograph. 6
2.4 Replay and viewing of the orthoscopic image. 7
2.5 True-scale spatial depth inversion using a retroreflective screen. 8
2.6 True-scale microtelescopic spatial depth inversion using an afocal combination of microlens arrays. 9
2.7 Two-tier retroreflective depth-inverting image transfer device. 10
2.8 Two-tier transmissive depth-inverting image transfer device. 10
2.9 A system for real time electronic capture of integral 3D images. 12
2.10 Diagrammatic representation of encoder/decoder arrays. 13
2.11 Categorisation of 3D imaging methodologies. 14

3.1 Generic model of a data compression system. 23
3.2 Transfer characteristic for a uniform symmetric quantiser. 26
3.3 Baseline sequential-DCT JPEG encoder. 28
3.4 The ‘zigzag’ scan of quantised DCT coefficients used in Baseline sequential-DCT JPEG encoding. 30
3.5 Baseline sequential-DCT JPEG decoder. 31

4.1 Rotational error in a scanned intensity distribution. 38
4.2 Determining scale error in a scanned intensity distribution. 40
4.3 Randomly distributed high frequency image information in a scanned intensity distribution. 42
4.4 A scanned intensity distribution without information at high spatial frequencies. 42
4.5 Intensity plot of frequency spectra for *Horseman*. 43
4.6 Intensity distribution *Horseman* processed using a 1D Butterworth noise rejection filter. 45
4.7 Noise reduction filter frequency response. 46
4.8 Operation of a one dimensional median filter. 47

4.9 Intensity distribution *Horseman* processed using a 1D median filter. 48
4.10 Moire fringes observed on a replayed integral intensity distribution. 52
4.11 Formation of Moire fringes due to pitch mismatch in integral three dimensional image replay. 53

5.1 Example integral image intensity distribution captured using a lenticular encoding element. 57
5.2 General structure of the lenticular-integral intensity distribution. 57
5.3 Normalised cross correlation coefficient evaluated between adjacent bands on the luma channel of four lenticular-integral image intensity distributions. 59
5.4 Pixel difference histograms. 61
5.5 Spatial domain correlation in a lenticular-integral intensity distribution. . . 63

6.1 Basic hybrid DPCM/DCT two dimensional video encoder. 66
6.2 First order linear prediction for inter-subimage spatial decorrelation. 68
6.3 Standard deviation of hybrid DPCM/DCT decorrelated coefficients. 71
6.4 The Base hybrid DPCM/DCT still integral 3D image compression scheme. 74
6.5 Modification to the base hybrid DPCM/DCT encoder for improved vertical intra-band correlation reduction. 75
6.6 Performance of the base hybrid DPCM/DCT scheme using fixed quantisation and a predefined code table — luma channel only. 84
6.7 Improvement in performance of the base hybrid DPCM/DCT scheme using fixed quantisation due to use of a local rather than predefined code table — luma channel only. 85
6.8 Difference in objective fidelity of the reconstructed intensity distribution between use of the enhanced hybrid DPCM/DCT scheme and the base hybrid DPCM/DCT scheme using fixed quantisation — luma channel only. 87
6.9 Difference in objective fidelity of the reconstructed intensity distribution between use of LVV and fixed quantisation under the base and enhanced hybrid DPCM/DCT schemes — luma channel only. 89
6.10 Performance of the base hybrid DPCM/DCT scheme using fixed quantisation — full colour source data. 91
6.11 Improvement in performance of the base hybrid DPCM/DCT scheme using fixed quantisation due to use of a local rather than predefined code table — full colour source data. 92
6.12 Difference in objective fidelity of the reconstructed intensity distribution between use of the enhanced hybrid DPCM/DCT scheme and the base hybrid DPCM/DCT scheme using fixed quantisation — full colour source data. . . 93

6.13 Difference in objective fidelity of the reconstructed intensity distribution between use of LVV and fixed quantisation in the base and enhanced hybrid DPCM/DCT schemes — full colour source data. 94
6.14 Magnified 16×16-subimage sections of *Horseman* resulting from coding using hybrid DPCM/DCT and JPEG schemes. 101
6.15 Magnified 16×16-subimage sections of *Toyplane* resulting from coding using hybrid DPCM/DCT and JPEG schemes. 106

7.1 Assembly of a subimage group for 3D-DCT based lenticular-integral image coding. 111
7.2 Use of the 'ESB' and 'ESG' symbols in 3D-DCT scheme coefficient-to-symbol mapping where group length $N = 4$. 114
7.3 Variances of 3D-DCT coefficients along the zigzag scan for each (u,v)-plane in a group of length 4. 115
7.4 Log of variance of 3D-DCT coefficients by (u,v)-position in each plane; group length $N = 4$. 116
7.5 An improved scanning strategy for 3D-DCT based lenticular-integral image coding; in this case $N = 4$. 116
7.6 Variances of 3D-DCT coefficients along the scan for the improved scanning strategy in a group of length 4. 117
7.7 The 3D-DCT based still integral 3D image compression scheme. 118
7.8 Difference in objective fidelity of the reconstructed intensity distribution between use of the 3D-DCT based scheme with fixed quantisation and the per-plane zigzag scanning strategy for several group lengths and the hybrid DPCM/DCT scheme — luma channel only. 124
7.9 Improvement in performance of the 3D-DCT based scheme with fixed quantisation and the per-plane zigzag scanning strategy due to use of a local rather than predefined code table for a group length of 8 — luma channel only. . . 126
7.10 Improvement in objective fidelity due to use of the all-plane scanning strategy rather than the per-plane zigzag strategy in the 3D-DCT scheme with fixed quantisation for a group length of 8 — luma channel only. 127
7.11 Improvement in objective fidelity due to use of the 3D-DCT based scheme with $N = 8$ and fixed quantisation, and the per-plane zigzag scanning strategy rather than the hybrid DPCM/DCT scheme — full colour source data. . . . 129
7.12 Improvement in performance of the 3D-DCT based scheme with fixed quantisation and the per-plane zigzag scanning strategy due to use of a local rather than predefined code table for a group length of 8 — full colour source data. 130

7.13 Improvement in objective fidelity due to use of the all-plane scanning strategy rather than the per-plane zigzag strategy in the 3D-DCT scheme with fixed quantisation for a group length of 8 — full colour source data. 131
7.14 Magnified 16×16-subimage sections of *Horseman* resulting from coding using the 3D-DCT and base hybrid DPCM/DCT schemes. 137
7.15 Magnified 16×16-subimage sections of *Toyplane* resulting from coding using the 3D-DCT and base hybrid DPCM/DCT schemes. 141

8.1 Standard deviation of 3D-DCT decorrelated subimage group coefficients for a group length of 4. 147
8.2 Derivation of rate-distortion optimal quantisation arrays. 148
8.3 Global and local minima in a hypothetical cost function. 150
8.4 Environment for derivation of optimal quantisation parameters for the 3D-DCT based coding scheme. 156
8.5 Intensity distribution section used in derivation of optimal quantisation parameters. 156
8.6 First (u,v)-plane of optimal quantisation array obtained for a bit rate $R_0 = 0.250061$ bpp. 158
8.7 Plot of final ten-point generating function parameters against output bit rate. 162
8.8 Improvement in objective fidelity due to use of optimal model-derived quantisation arrays rather than fixed quantisation arrays; all-plane scanning strategy and code table generated for 0.5 bpp employed — luma channel only. . 163
8.9 Magnified 16×16-subimage sections of luma channels of *Horseman* and *Toyplane* resulting from coding using the 3D-DCT scheme with fixed and optimal model-derived quantisation. 167

9.1 Staged performance data for six coding schemes applied to lenticular-integral intensity distribution data (luma channel only). 171
9.2 Staged performance data for five coding schemes applied to lenticular-integral intensity distribution data (full colour coding). 172
9.3 Assembly of a subimage group for 3D-DCT based full integral 3D image coding. 174

B.1 Luma channel of conventional test image *Lena*. 180
B.2 Luma channel of conventional test image *Airplane*. 181
B.3 Luma channel of conventional test image *Peppers*. 181
B.4 Luma channel of conventional test image *Sailboat*. 182

LIST OF PROCEDURES

6.1 Top level base/enhanced hybrid DPCM/DCT simulation system. 76
6.2 Main loop for hybrid DPCM/DCT simulation. 76
6.3 Symbol frequency table generation. 77
6.4 DC difference symbol frequency table generation for enhanced mode. . . . 78
6.5 Calculate bit count for main Huffman coder output. 79
6.6 Calculate bit count for DC Huffman coder output in enhanced mode. 79

7.1 Top level 3D-DCT scheme simulation system. 119
7.2 Main loop for 3D-DCT scheme simulation. 120
7.3 3D-DCT scheme symbol frequency table generation (per-plane scanning). . 120
7.4 3D-DCT scheme symbol frequency table generation (all-plane scanning). . 121
7.5 Calculate bit count for coded intensity distribution (per-plane scanning). . . 121
7.6 Calculate bit count for coded intensity distribution (all-plane scanning). . . 122

LIST OF INTEGRAL INTENSITY DISTRIBUTIONS

The following list references intensity distributions which have been scaled suitably for replay as three dimensional integral images and output using a dye sublimation printer. The lenticular sheet material provided in the inside front cover of this thesis may be used to view them. Two types of lenticular sheet have been provided — Screens 'A' and 'B' — for decoding images recorded using lenticular pitches of 1270 μm and 600 μm respectively. The correct decoding screen for viewing each image is specified in its caption using the keys '[SCREEN A]' and '[SCREEN B]'.

6.1 *Horseman* compressed using the enhanced hybrid DPCM/DCT scheme with fixed quantisation; output bit rate is 1.71 bpp. 97
6.2 *Horseman* compressed using the enhanced hybrid DPCM/DCT scheme with fixed quantisation; output bit rate is 0.76 bpp. 98
6.3 *Horseman* compressed using the base hybrid DPCM/DCT scheme with LVV quantisation; output bit rate is 0.37 bpp. 99
6.4 *Horseman* compressed using the baseline JPEG scheme with JPEG standard quantisation; output bit rate is 0.37 bpp. 100
6.5 *Toyplane* compressed using the enhanced hybrid DPCM/DCT scheme with fixed quantisation; output bit rate is 1.75 bpp. 102
6.6 *Toyplane* compressed using the enhanced hybrid DPCM/DCT scheme with fixed quantisation; output bit rate is 0.76 bpp. 103
6.7 *Toyplane* compressed using the base hybrid DPCM/DCT scheme with LVV quantisation; output bit rate is 0.41 bpp. 104
6.8 *Toyplane* compressed using the baseline JPEG scheme with JPEG standard quantisation; output bit rate is 0.41 bpp. 105

7.1 *Horseman* compressed using the 3D-DCT scheme with fixed quantisation; output bit rate is 1.65 bpp using the all-plane scanning strategy. 134
7.2 *Horseman* compressed using the 3D-DCT scheme with fixed quantisation; output bit rate is 0.75 bpp using the all-plane scanning strategy. 135
7.3 *Horseman* compressed using the 3D-DCT scheme with fixed quantisation; output bit rate is 0.35 bpp using the all-plane scanning strategy. 136

7.4 *Toyplane* compressed using the 3D-DCT scheme with fixed quantisation; output bit rate is 1.74 bpp using the all-plane scanning strategy. 138
7.5 *Toyplane* compressed using the 3D-DCT scheme with fixed quantisation; output bit rate is 0.76 bpp using the all-plane scanning strategy. 139
7.6 *Toyplane* compressed using the 3D-DCT scheme with fixed quantisation; output bit rate is 0.39 bpp using the all-plane scanning strategy. 140

8.1 Luma channel of *Horseman* compressed using the 3D-DCT scheme with optimal model-derived quantisation; output bit rate is 0.14 bpp. 165
8.2 Luma channel of *Toyplane* compressed using the 3D-DCT scheme with optimal model-derived quantisation; output bit rate is 0.16 bpp. 166

A.1 *Toyplane*, original. 177
A.2 *Microscope*, original. 178
A.3 *Tank*, original. 179
A.4 *Horseman*, original. 179

1. INTRODUCTION

1.1 Overview

Many potential applications exist for three dimensional (3D) image and video communication systems. One much discussed application is three dimensional television (3D-TV), which possesses several advantages over conventional standard and high definition colour television systems. It has been reported that the greatly improved sensations of depth and naturalness provided by a three dimensional display cause viewers to perceive an increase in overall picture quality, leading them to prefer 3D presentation [3].

A large variety of three dimensional video recording and display systems have been proposed by a number of independent research groups [3–5]. Much work on the development of suitable autostereoscopic (free viewing) systems has centred around the use of multi-camera capture and multi-view display using lenticular sheets to afford the viewer a certain degree of look-around capability. Such systems do not produce truly spatial 3D displays since they lack vertical parallax and rely on the brain's fusing of two disparate planar images.

An innovative and unique optical system for transferring full parallax three dimensional images has been developed by the 3D Imaging Group at De Montfort University [6, 7]. The technique is based around Lippmann's method of Integral Photography [8], in which three dimensional volume spatial information from a scene is represented as related series of intensity distributions recorded on a plane surface. The recording and replay processes take place in constant time, hence with the use of flat panel format electronic image capture and display devices with adequate spatial and temporal resolution, real time digital 3D video sequence capture and display systems using this technique are practicable. Past estimates of the bandwidth required for communication of integral image data have precluded the use of the technique, however it has been demonstrated by the Group that spatial resolution, and hence bandwidth required is far lower than previously predicted. Despite this, to enable transmission and storage of integral 3D video sequence data in formats compatible with

conventional digital video studio and broadcast systems and with adequate scene size, information reduction is necessary. Compression techniques to carry out this function therefore form an essential component of a three dimensional image communication system using the integral 3D imaging method.

In order to achieve the data reduction necessary, a lossy compression scheme is required, which removes certain components of the source data which are not significant in reconstructing an image with high perceived fidelity. In design of such a system, there is a fundamental compromise between the data reduction possible and the distortion present in the reconstructed, decoded image. Where integral 3D images are concerned, a suitable coding system must provide data reduction which is sufficient for transmission via conventional broadcast methods whilst being capable of high quality reconstruction of the source object scene such that no objectionable distortions are noticeable to the viewer.

1.2 Scope of the Thesis

Many lossy compression schemes exist, which have been developed for use with conventional two dimensional still images and video data. Since the captured representation of an integral 3D image is a planar intensity distribution, such data can be stored using conventional image formats. Therefore, the captured integral 3D image would seem to be compatible with two dimensional image compression schemes. However, due to the optical characteristics of the capture system the resulting intensity distribution possesses a structure with unique statistical characteristics.

In this thesis the development of compression schemes specifically for application to integral 3D image data is presented, which exploit the unique structure of these data. Two schemes, based on hybrid DPCM/DCT and 3D-DCT based encoders are described, and simulation results are presented which show that greatly improved rate-distortion performance can be obtained compared with use of conventional two dimensional image compression schemes when the characteristics of captured integral 3D images are exploited. An optimisation of the quantisation stage, critical to rate-distortion performance, is also presented, resulting in further improved reconstructed image quality at a given level of data reduction.

This thesis also describes front-end capture and processing techniques developed to ensure that the intensity distributions presented to the coding schemes are free of distortion and represented in a common format, post-processing of decoded intensity distributions for display,

and the replay of the resulting data as three dimensional images.

1.3 Outline of the Thesis

The thesis is organised into nine chapters. This chapter has introduced the research undertaken by providing an overview of integral three dimensional imaging, the need for information reduction for transmission of integral images and the development of compression schemes specifically for application to this type of image.

Chapters 2 and 3 summarise the two main concepts brought together by the work: three dimensional imaging and data compression, providing background for the remainder of the thesis. Specifically, chapter 2 gives a short treatment of the optical components of the integral imaging system and introduces the electronic capture and display processes. Brief descriptions of alternative autostereoscopic imaging systems are given and comparisons made with the integral imaging system. An overview of digital image and video communication methods is given in chapter 3, describing the purpose and main components of data compression systems and also summarising the operation of the JPEG lossy compression system for conventional natural still images.

In chapter 4, procedures developed for capture of integral image data are described. Additionally, this chapter describes processing techniques designed to correct distortions introduced by the scanning process.

A study of the statistical properties of the intensity distribution which represents captured integral 3D images is given in chapter 5, such that the correlation redundancy present within the data can be exploited effectively by a compression scheme.

Chapters 6 and 7 describe two coding schemes developed for application to integral 3D image data. The first makes use of a hybrid DPCM/DCT encoder structure and the second a three dimensional DCT stage providing more effective decorrelation of the source data. It is shown through simulation results that these schemes possess significantly greater rate-distortion performance than a conventional two dimensional image coding scheme, JPEG, when applied to integral 3D image data.

Improvements to the quantisation stage made through the use of numerical optimisation and model fitting techniques are presented in chapter 8. It is demonstrated that further improved

performance is possible if the quantisation parameters of the 3D-DCT scheme are matched to the characteristics of the transformed intensity distribution.

Finally, chapter 9 concludes the thesis and suggests possible future directions for extension of the research.

2. THREE DIMENSIONAL IMAGING FOR TELEVISION

Introduction

As stated in the introduction to this thesis, it has been found that viewers generally prefer 3D display systems, in which the depth dimension of a scene is naturally presented, to conventional 2D systems. Several classes of 3D imaging system have been proposed and constructed by many independent research groups either as general purpose display units or targetting a specific application. This chapter gives an outline of the construction and operation of the integral (*holoscopic*) 3D imaging system developed by the 3D Imaging Group at De Montfort University. The system makes use of unique optical capture apparatus to encode a true volume spatial optical model of the object scene in the form of a planar intensity distribution which may then be stored or transmitted as a conventional two dimensional pixel array. Following is a brief review of alternative 3D display methodologies, concentrating on free-viewing (autostereoscopic) systems. Such systems are compared with the integral 3D imaging system for their suitability for use in a consumer television application.

2.1 Integral (Holoscopic) Imaging

2.1.1 Classical Integral Photography

The recording and display of true 3D images was pioneered by Lippmann [8]. His technique, termed *integral photography (IP)* utilises a two dimensional microlens array to capture a related series of intensity distributions which encodes the locations of points in the original scene on photographic film (see Figure 2.1). In order to replay a 3D image of the original scene, the processed photographic transparency is overlaid with a microlens array identical to that used in the capture process and back-illuminated with incoherent light (see Figure 2.2). When the scene is replayed in this way the viewer observes a pseudoscopic (spatially depth

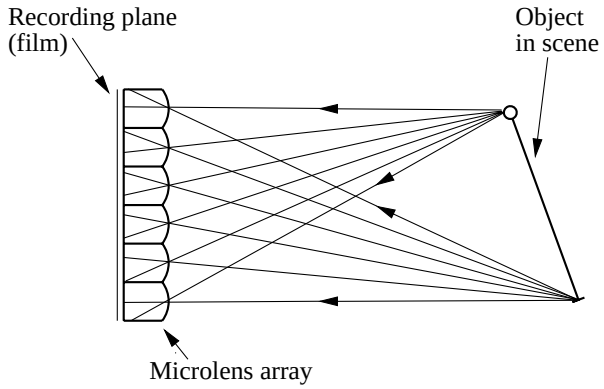


Figure 2.1: Recording of an integral photograph.

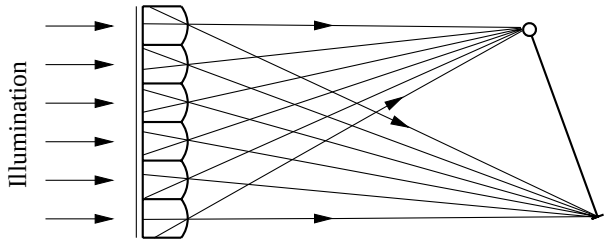


Figure 2.2: Replay of an integral photograph.

inverted) image. To overcome this problem a second recording stage is introduced, in which a second integral photograph is taken of the first in replay, (see Figure 2.3) as proposed by Ives [9]. When the second-stage photograph is replayed as shown in Figure 2.4 an orthoscopic (spatially depth corrected) 3D image results.

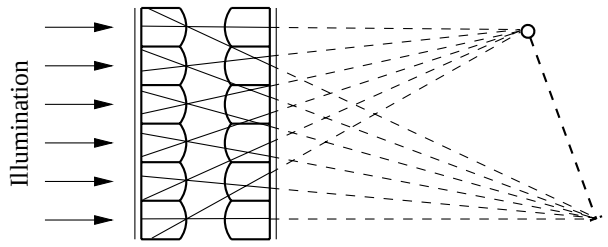


Figure 2.3: Production of a second-stage integral photograph.

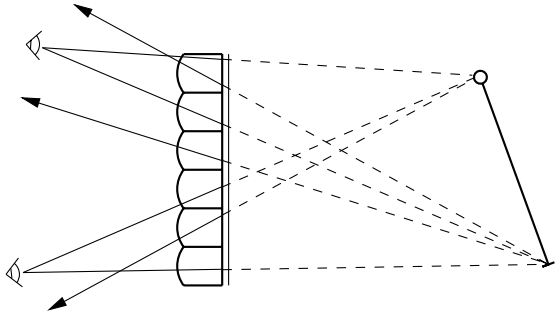


Figure 2.4: Replay and viewing of the orthoscopic image.

2.1.2 Real-time Holoscopic Integral Image Capture

The two-stage recording process outlined above possesses several disadvantages as a practical imaging system:

1. The process does not take place in constant time and is therefore unsuitable for use within a real time image capture and display environment;
2. Only remote imaging is possible since it is inherently impossible to place the capture plane at arbitrary locations in a solid scene;
3. The final image is substantially degraded due to distortions introduced in the microlenses and film emulsion, and possible stray light effects.

In order to address the first concern, the capture of the orthoscopic image must take place in a single operation, that is, the camera unit must perform the necessary spatial depth inversion optically. Techniques which have been used by the 3D Imaging Group to accomplish this will now be summarised.

Single-tier spatial inversion methods

Depth inversion may be accomplished through the use of a retroreflective screen, which always reflects an incident ray of light back in its original direction. This device is used in combination with a half-mirror to produce the required pseudoscopic aerial image. Such a scheme is shown in Figure 2.5. Alternatively a more optically efficient scheme using an

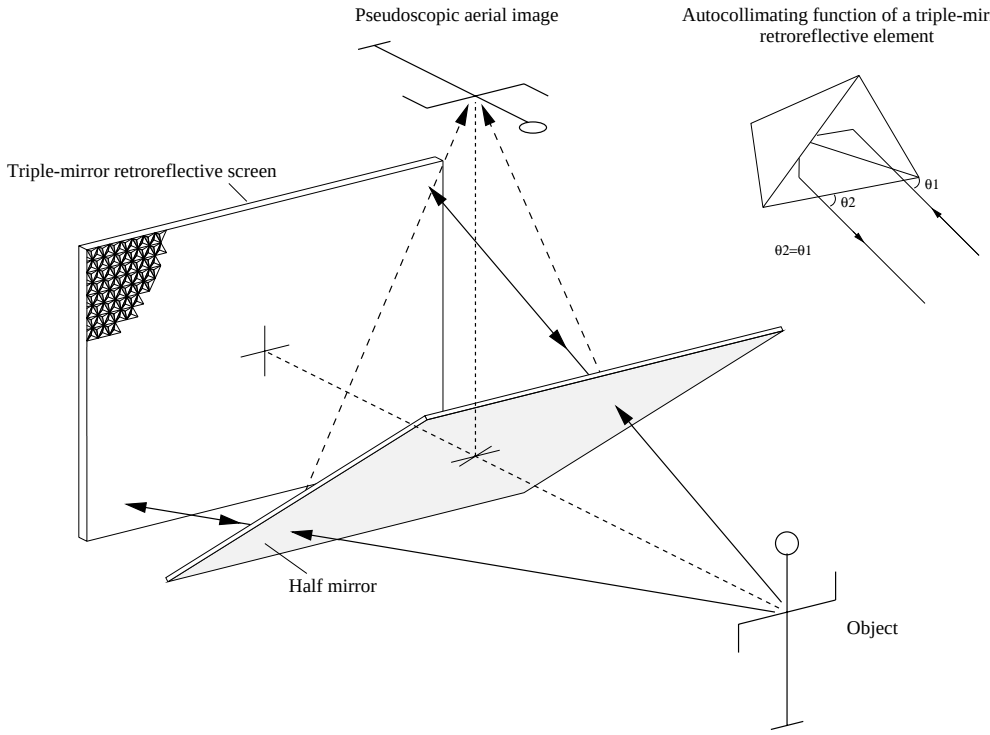


Figure 2.5: True-scale spatial depth inversion using a retroreflective screen.

afocal combination of two microlens arrays allows direct transmission of the optical model on the axis of the system (Figure 2.6).

Two-tier spatial inversion methods

Unfortunately the simple systems described above are incapable of providing sufficient lateral and angular resolution for use in a high quality imaging system, that is sharp focus cannot be maintained in the near and far fields [10, 11]. These systems are therefore unsuitable for use in a practical high quality imaging system. The resolution limitation can be overcome if it is arranged that the object scene appears entirely in the near field of the retrodirective screen [7] or afocal microlens combination [6]. For this purpose an additional lens system is included to image the scene around the central plane of the microlens array combination or front face of the retroreflective screen. The resulting two-tier system is illustrated in Figure 2.7 for the retroreflective case and in Figure 2.8 for the transmissive case. In order to provide a large aperture the relay lens system is composed of a two dimensional ‘macrolens’ array comprised of multiple identical simple (plano-convex) lenses.

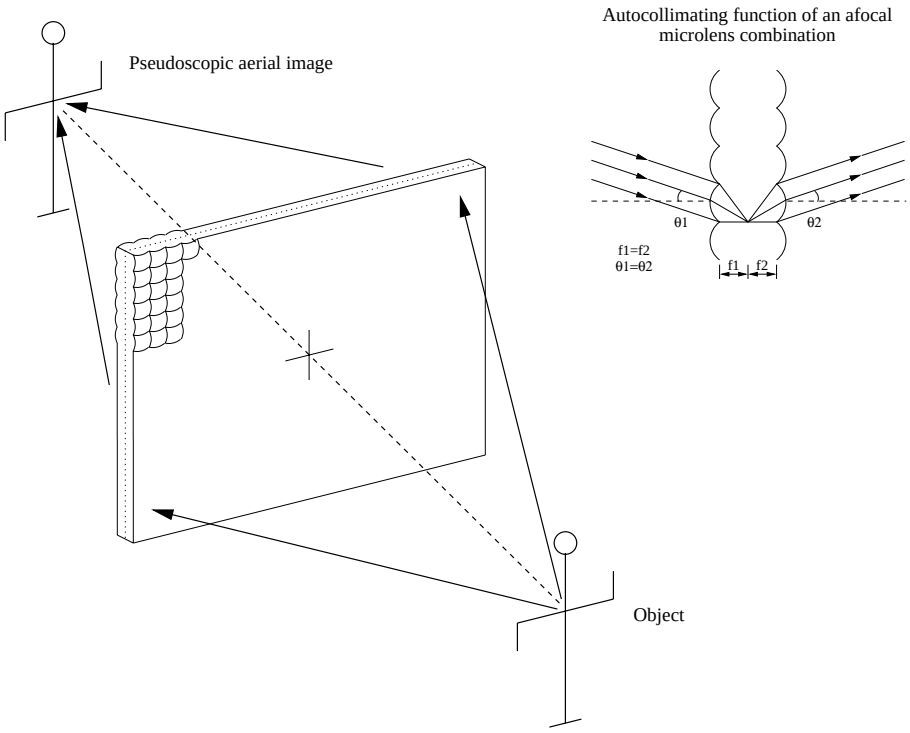


Figure 2.6: True-scale microtelescopic spatial depth inversion using an afocal combination of microlens arrays.

With the use of either two-tier image transfer device the integral image capture system can be operated in a ‘close imaging’ mode, since it is possible to select any plane within or around the pseudoscopic aerial image as the reference (display screen) plane. Thus the second limitation of two-stage integral photography as the basis of a practical imaging system has been addressed. The capture process now takes place as a single-stage process, and therefore the imaging system can be operated in real time.

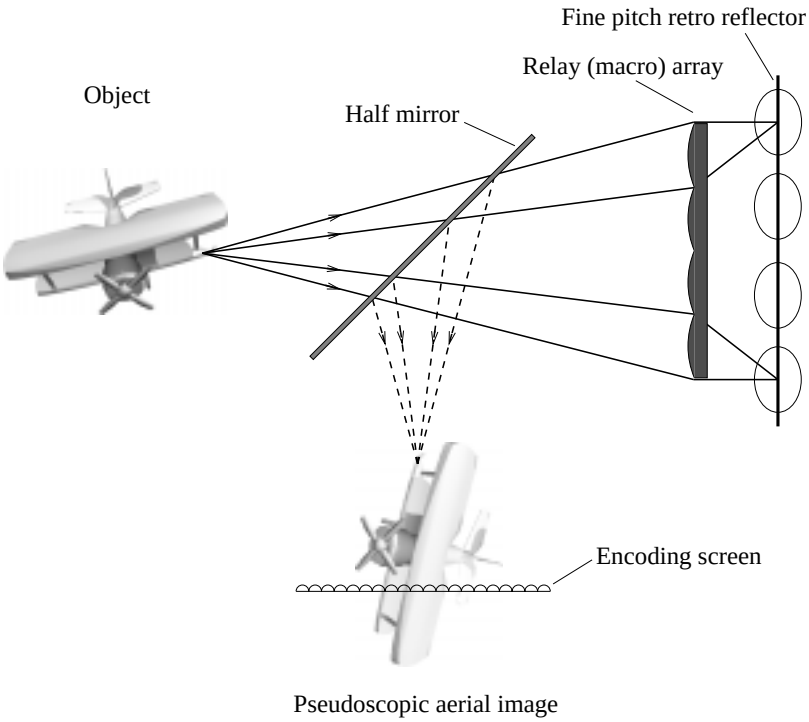


Figure 2.7: Two-tier retroreflective depth-inverting image transfer device.

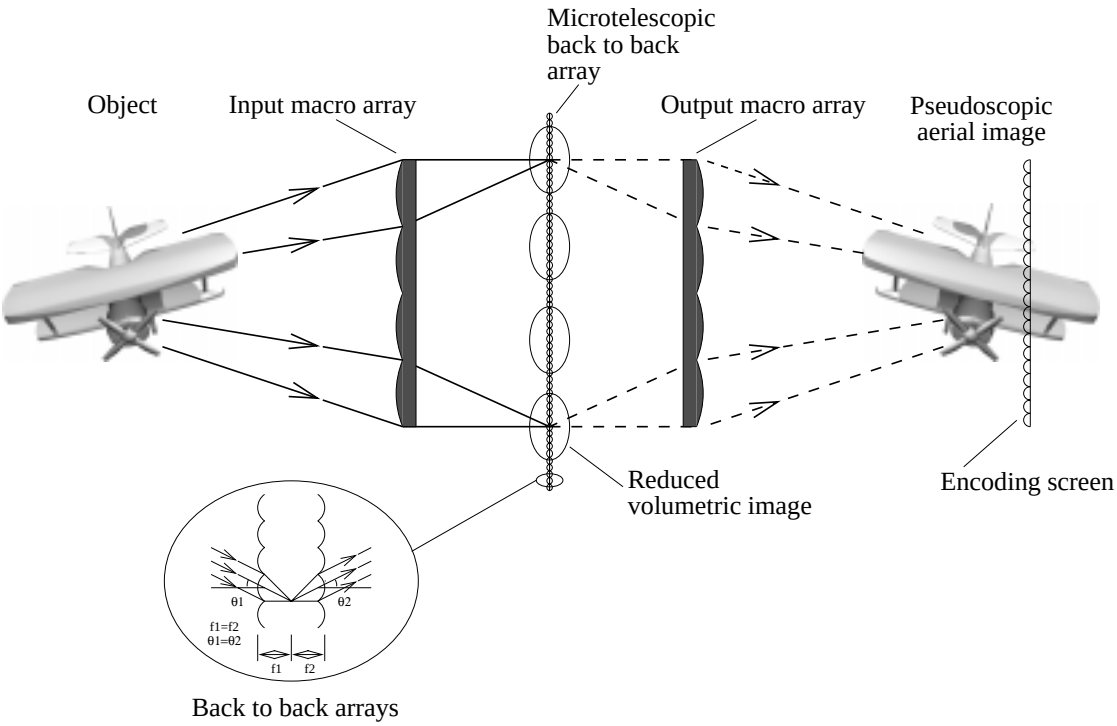


Figure 2.8: Two-tier transmissive depth-inverting image transfer device.

2.1.3 Electronic Capture of Integral Images

Any conventional image may be captured electronically via one of two routes, to which the author refers as *direct* and *indirect* capture. Since a captured integral 3D image is represented as a planar intensity distribution and may be recorded in a single optical operation, the same applies.

Direct Capture

This refers to the use of a real-time imaging unit capable of generating either discrete images or a continuous video signal. Where conventional television pictures are concerned, a television camera performs this function. For direct capture of integral three-dimensional image frames, the requirements of the optical image transfer system preclude the use of a conventional camera. This is the case since it is imperative that the encoded planar intensity distribution is recorded exactly parallel to its encoding plane (the rear surface of the microlens or lenticular encoding array), requiring a mechanically stable self-contained capture unit containing both the image transfer optical elements and the target electronic imaging array. Work is under way within the 3D Imaging Group to develop such a system utilising charge coupled imaging devices (CCDs) in the intensity distribution recording plane.

Figure 2.9 shows the principal elements of a simple camera unit. A pseudoscopic optical model of the object scene is produced by the two tier transmissive image transfer unit and the encoding microlens array produces at its back face an intensity distribution representing the object scene. In the case of the still camera developed by the Group, photographic film would be placed at this point. In this real time capture device, however, a copy lens arrangement images the current planar encoding of the object scene on to a high resolution CCD imaging device, which then produces an electronic representation of the captured intensity distribution.

Indirect Capture

Indirect electronic capture refers to the use of photographic film as an intermediate image recording medium. The processed photograph is digitised at a later time using an electronic scanning technique. Real-time (video) capture is evidently not possible using this technique.

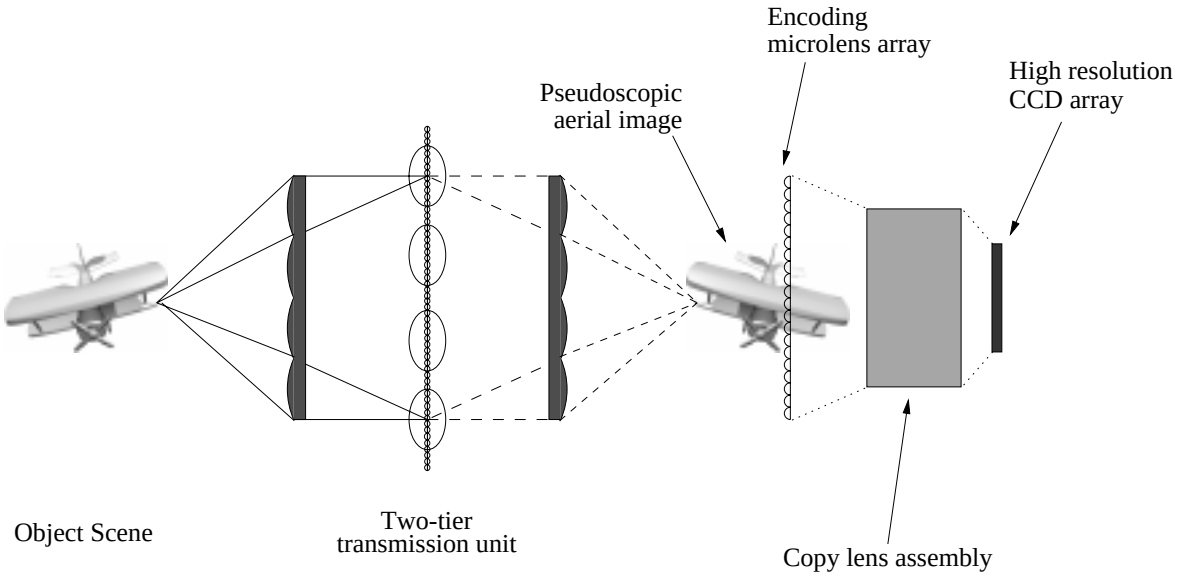


Figure 2.9: A system for real time electronic capture of integral 3D images.

For the purpose of the work described in this thesis involving the use of still integral 3D images, intensity data were obtained by electronically scanning photographic transparencies which had been produced by the still version of the integral image capture device.

2.1.4 Electronic Display of Integral Images

In order to replay the original scene in three dimensions with continuous parallax, it is necessary to overlay the intensity distribution captured by the CCD imager with a microlens array, with lenslets aligned with the displayed data as they were in capture. To achieve this a planar, high contrast, colour display medium is required, such as an active-matrix liquid crystal display. The display screen is overlaid directly with the decoding microlens array.

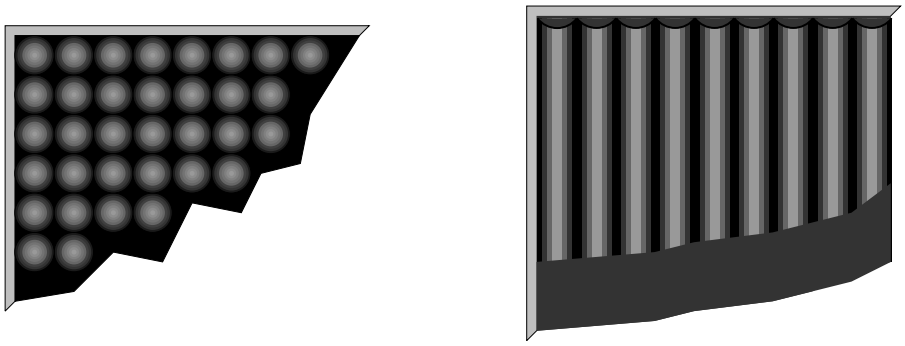
For the purpose of work described in this thesis which involves only still integral images, electronic display was simulated using a 300 dpi dye sublimation colour printer. This device produces continuous colour tones as square pixels; no dithering is required as is generally the case in the process of most other types of desktop colour hardcopy device. Such techniques give rise to unwanted interference effects and lower the effective colour output resolution [12].

2.1.5 The Lenticular-Integral Image

The capture and display processes outlined above make use of microlens arrays in the encoding and decoding of the planar intensity distribution. The arrays used for this purpose are typically of square base, and are capable of encoding the object scene with continuous parallax in all directions parallel to the capture plane. A section of such a lens array is illustrated in Figure 2.10a. In the figure, intensity represents depth.

It is also possible to record and replay integral 3D images using lenticular sheets, which comprise many thin cylindrical lenses (see Figure 2.10b). Again, in the figure intensity represents depth. Integral images recorded in this way, with the lenticular encoder/decoder sheets oriented such that the lenticules run vertically, possess parallax only horizontally though still continuously. It is important to note that images recorded in this way do *not* replay as lenticular stereoscopic images, where the lenticular sheet is used merely for spatial demultiplexing of multiple separate views of an object scene—see section 2.2 [10]. To emphasise this distinction the author uses the term 'lenticular-integral image' to refer to an integral 3D image recorded using lenticular sheet.

All work carried out on compression of integral 3D image data, which is the subject of this thesis, made use of lenticular-integral 3D image data, as these were more readily available from the camera system under development by the Group.



a)
Section of square based microlens array

b)
Section of lenticular sheet

Figure 2.10: Diagrammatic representation of encoder/decoder arrays.

2.2 Alternative Three Dimensional Imaging Methodologies

A number of novel 3D imaging systems, capable of reproducing scene depth and parallax, have been proposed. Figure 2.11 presents a categorisation of techniques based on whether or not viewing aids are required (non-autostereoscopic or autostereoscopic). Of all of these methods, only free-viewing (autostereoscopic) methods are appropriate for television applications as a requirement for viewing aids detracts from the natural viewing experience expected by the user. The furthest developed of the autostereoscopic techniques are holography and lenticular sheet based stereoscopic systems. These methods will be briefly described and compared with holoscopy (integral imaging), the technique developed by the 3D Imaging Group at De Montfort University.

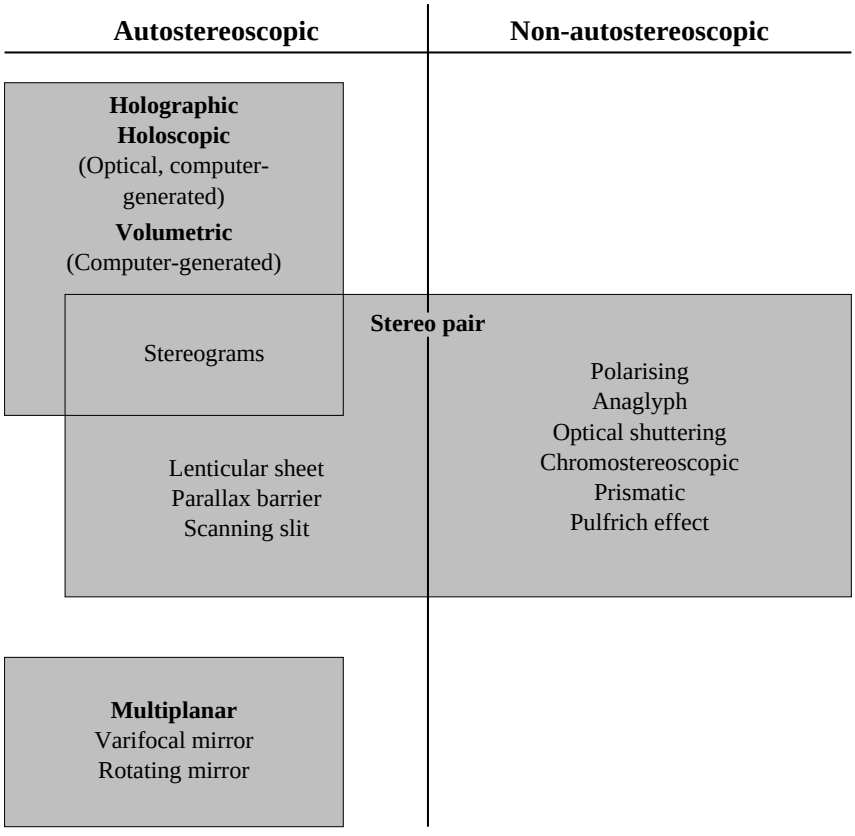


Figure 2.11: Categorisation of 3D imaging methodologies.

2.2.1 Holography

The method of holography, originally proposed by Gabor in 1948, allows the capture of a volume spatial image which may be replayed with full parallax within a wide viewing zone. Rather than a separate fine lens sheet as used in integral (holoscopic) imaging for direction-selective recording, holographic encoding makes use of a large number of diffraction gratings which are intrinsic to the image. A coherent light source is required in the capture process in order to produce interference fringes on the recording medium encoding a spatial image of the object scene. Dependent on the precise materials and technique used in the capture process, reconstruction may take place under coherent or incoherent illumination [10]. Many variations on and enhancements to the basic process have been proposed and demonstrated, and much research effort is currently concentrated on the production of natural colour holograms.

Practical holographic systems have concentrated on the production of still images; these are of very high quality and give an extremely realistic reproduction of the object scene. The transfer of holographic techniques to the display of moving spatial images of equivalent quality presents several problems. These are centred around the extremely high information content of a hologram [10] and the related difficulty in producing a display medium which can be addressed and refreshed with sufficient frequency. Making use of information reduction techniques, holographic display devices have been constructed, such as the 'holovideo' computer-generated hologram display system developed by S. A. Benton and colleagues [13]. Such systems are highly complex, however, and require very high speed processing systems to achieve even relatively low frame rates. An additional problem in the capture of natural scenes stems from the requirements for coherent light sources, dark room conditions and high mechanical stability during recording. These are incompatible with the conditions under which television programme production must in general be able to take place. These considerations reduce the practical utility of the technique for general spatial video image recording applications.

2.2.2 Lenticular Stereoscopic Imaging

Extensive research has been carried out on the use of this form of stereoscopic display for television. The basic principle behind all stereoscopic displays is the presentation of two separate images of the object scene, one to each eye. These are then fused by the brain, producing a convincing three dimensional image. The sensation of depth experienced by the

viewer is enhanced compared with viewing of planar images since the use of two separate views allows binocular physiological depth cues—important for natural depth perception—to be used [14]. The many stereoscopic display techniques developed differ only in the method and apparatus used to present these disparate planar images to the viewer.

Lenticular stereoscopic displays exploit the image channelling property of lenticular sheet material, directing the left-eye and right-eye views of the stereoscopic image to the correct eyes. The simplest form of such a display is comprised of alternating vertical strips of the right-eye/left-eye images, with each stereo pair of view strips located precisely behind a lenticule. This delivers a stereo image observed from a single viewpoint to the viewer. To integrate a 'look-around' capability, enhanced systems provide multiple stereo viewpoints by locating bands derived from many views behind each lenticule. Typically four to eight views are provided [5].

Provided that the image data from all appropriate views can be interlaced into the multi-view banded arrangement in real time, live stereoscopic video can be produced using a lenticular stereoscopic display. The main disadvantage, however, of any stereoscopic imaging system is the requirement for multiple cameras at the image capture stage if live capture is to be possible. In general the number of cameras required is equal to the number of views to be presented at the display; although interpolation schemes have been used to reduce the number of cameras needed [15], the requirement for more than one results in increased costs, bulky equipment and the need for careful alignment of cameras before recording can take place. Additionally, the optical parameters of the camera lenses must be carefully matched.

2.2.3 A Comparison of Autostereoscopic 3D Imaging Systems for Television

Table 2.1 presents a comparison of the autostereoscopic three dimensional imaging systems discussed. The comparison is based on certain properties which are important in a practical consumer 3D television system with live capture, transmission and display capabilities.

Property	*Integral Imaging*	*Holography*	*Lenticular Stereoscopic/ Multiview*
Image ‘solidity’	Good	Very good	Reasonable: • Can suffer from visible discrete depth planes (‘cardboarding’)
Look-around capability	Good: • continuous parallax; • wide viewing zone	Very good: • continuous parallax; • very wide viewing zone	Reasonable: • Depends on number of views; • discrete parallax; • can ‘flip’ between views
Colour realism	Very good	Recent research has produced reasonable–good results	Very good
Live capture practicality	Very good: • single ‘camera’ unit	Poor	Reasonable: • multi-camera so expensive and unwieldy
Live display practicality	Very good	Poor	Very good

Table 2.1: Comparison of 3D Imaging Systems for Television Applications

Summary

In this chapter the operation of the integral three-dimensional imaging system has been outlined, from its basis in integral photography to a unique optical system capable of real-time image capture. The use of electronic imaging devices for recording of the planar intensity distribution produced by the camera system has also been discussed. Brief descriptions of alternative autostereoscopic imaging systems have been given and comparisons made with the integral imaging system.

3. COMMUNICATION AND COMPRESSION OF IMAGE DATA

Introduction

Communication of images, both still and in time domain sequences, has many and varied applications. It may be necessary to transmit image data over significant distances, often in real time, or alternatively offline storage may be required for archiving purposes. In either case it is desired to make use of transmission bandwidth or storage capacity as efficiently as possible so that resource requirements and hence cost can be minimised or a greater number of images can share finite bandwidth or storage resources. Data compression techniques have become extremely important to fulfill this need, and many algorithms have been internationally standardised for various applications.

This chapter provides an overview of the subject of digital image formats and transmission for digital video broadcasting applications, and outlines general data compression concepts and specific image data compression techniques. The ISO/IEC JPEG standard for still continuous-tone image compression is used to illustrate the concepts developed.

3.1 Digital Video Image Representation

In the general case a television picture is a continuous colour image of a scene captured by a camera system. For digital processing of such an image to take place, certain parameters must be set. The most important of these are the dimensions of the digital image and the method of encoding of its colour information. Digital television formats have been standardised within the telecommunications industry, both for standard definition and high definition television (HDTV) systems [16–18].

3.1.1 Colour encoding

Digital television picture standards specify the use of the $Y'C_bC_r$ colour coordinate system. This is based on the $Y'UV$ system employed by the NTSC, PAL and SECAM composite television standards, modified for use with digital signals. Y' (luma) represents the brightness of pixels in the colour image and C_b and C_r represent colour differences. Digital colour image data obtained from a video camera or scanner is stored using the $R'G'B'$ coordinate system. The transformation to $Y'C_bC_r$ for communication and compression is [16]:

$$\begin{bmatrix} Y' \\ C_b \\ C_r \end{bmatrix} = \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix}. \tag{3.1}$$

All signals are represented using integers defined on $[0, 255]$.

3.1.2 Spatial and temporal sampling characteristics

The size of the image frame is determined by the sampling frequency required. The colour difference signals C_b and C_r are sampled at one half the rate of the luma signal Y', and thus there are precisely half the number of C_b and C_r pixels per line as Y' pixels per line. The maximum field rate is 60 s^{-1}. Table 3.1 lists the parameter values defined for both standard and high definition television. The total image size is that which must be processed and transmitted to allow for overscan on television receivers. The active image size for high definition television relates to the 'Common Image Format' for international use [17].

Display definition	*Maximum total image size*	*Active (displayed) image size*	*Maximum frame rate*
Standard [16]	864×625 pixels	720×576 pixels	30 s^{-1}
High (Europe) [17]	2304×1250 pixels	1920×1080 pixels	25 s^{-1}

Table 3.1: Digital television picture sampling characteristics.

3.2 Radio Frequency Broadcast of Digital Images

Terrestrial and cable transmission networks used for television picture distribution provide many channels each with a fixed bandwidth. The channel bandwidth, BW_c, is typically between 6 and 8 MHz depending upon the television standard used in a particular geographic region [19].

3.2.1 Digital signal modulation

In order to allow broadcast of digital television signals, a digital modulation system must be used. Many modulation techniques have been developed. Quadrature Amplitude Modulation (QAM) is commonly used [20, 21], for example by the European Telecommunications Standards Institute (ETSI) Digital Video Broadcasting–Cable (DVB-C) standard [22]. Many planned digital cable television systems employ this standard. The example of digital cable television transmission is used in the channel capacity calculations below.

QAM belongs to a class of digital modulation schemes which provides multisymbol signalling, that is successive digits in the input bit stream are combined to produce one of several possible states in the modulated carrier. This approach reduces the bandwidth required to transmit a sequence of a given length, but at the expense of noise immunity. A fundamental parameter of any such system is its spectral efficiency, which provides a measure for the number of binary digits in the baseband modulating signal which may be transmitted within a given channel bandwidth. The spectral efficiency of QAM with M states (denoted M-ary QAM),

$$\eta = \frac{\log_2 M}{1+\alpha} \quad \text{bit.s}^{-1}\text{Hz}^{-1} \tag{3.2}$$

where α is the roll-off factor of the modulation shaping filter.

By way of example, in digital video broadcast applications for cable systems 64-ary QAM is extensively used. The roll-off factor used by the baseband shaping filter in the DVB transmission system is 0.15. In this case the spectral efficiency is approximately $5.22\ \text{bit.s}^{-1}\text{Hz}^{-1}$ and thus for a channel bandwidth of 8 MHz, used throughout Europe, the basic bit rate available for the baseband signal without error correction coding,

$$BR_c = \eta BW_c \approx 41.74\ \text{Mbit.s}^{-1}.$$

The basic bit rate required for a digital video signal,

$$BR = \left(1+\tfrac{1}{2}+\tfrac{1}{2}\right)\big[(\text{Bits/sample}) \times (\text{Samples/line}) \times \\ \times\ (\text{Lines/frame}) \times (\text{Frames/s})\big] \quad \text{bit.s}^{-1} \tag{3.3}$$

where the first term represents the contributions of the luma and two colour difference channels. For signals with characteristics as described in section 3.1, the raw bit rate for standard definition pictures at a frame rate of 25 s^{-1},

$$BR_{\text{SD}} = (2)(8)(864 \times 625)(25) = 216 \text{ Mbit.s}^{-1},$$

and for high definition at 25 frames per second,

$$BR_{\text{HD}} = (2)(8)(2304 \times 1250)(25) = 1.152 \text{ Gbit.s}^{-1}.$$

3.2.2 Digital television bit rate reduction

Evidently in order to transmit such signals using the modulation scheme described requires bit rate reduction. The necessary minimum data compression ratios are $5.175:1$ and $27.6:1$ for standard and high definition sequences respectively. The MPEG-2 scheme [23, 24], whose hybrid DPCM/DCT based encoder is outlined in chapter 6, is used for video compression and coding as part of the DVB system. MPEG-2 has been designed to provide sufficient reduction of transmission bit rate for modulation while decoding video with no distortions noticeable to the viewer. It also allows for simultaneous encoding of accompanying audio.

3.3 Data Compression

It is often desirable to reduce the amount of storage or bandwidth required to improve efficiency in communication of information. For transmission of digital television sequences bit rate reduction is essential to enable broadcast to take place. Fundamentally data compression schemes operate by exploiting redundancies which are inherently present in many types of data stream due to correlation between data elements, and the fact that the occurrences of possible elements are rarely equiprobable [25].

3.3.1 A generic compression system model

A generic compression system may be modelled as a three stage process [26], as illustrated in Figure 3.1. The first stage, the mapping operation, does not contribute to quantitative data reduction. It converts the source data stream to an alternative representation, a set of coefficients, in order that correlations present may be more effectively exploited. It is often known as the *decorrelation* stage.

In many applications where the decoded data are destined for human viewing, for example compression of natural images[1] or video sequences, it is not necessary for the decompressed data to have exactly the same set of values as the original. In this case a quantisation stage is used to enable extremely high information reduction. The quantiser carries out a many-to-one mapping of the decorrelated data elements to a smaller set of *quantisation levels*. This results in the final stage providing greater data reduction than otherwise possible since fewer discrete data objects must be represented. In other applications such as compression of executable computer programs or data resulting from medical scanning equipment, it is essential that the decoded data are identical with that originally presented to the compression scheme. Compression schemes are classed as *lossy* or *lossless* if they contain or do not contain a quantisation stage respectively. The quantiser must be chosen to be compatible with the preceding mapping stage.

The final stage, the coder, is responsible for quantitative data reduction. It takes advantage of the fact that the probability of occurrence of different data objects, whether previously quantised or not, is very rarely equal. The coder produces a *codeword* comprised of a unique sequence of binary digits, for each possible input symbol. The objective is to compact the data resulting from decorrelation (in a lossless scheme) or quantisation (in a lossy scheme) such as to approach its entropy[2] as closely as possible. Hence this process is often known as *entropy coding*.

[1] A *natural image* is any image which has not been computer-generated.

[2] Entropy is defined as the average information content of a source and is hence a measure of uncertainty in an information source [27].

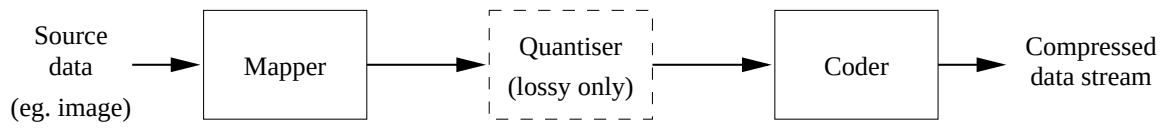


Figure 3.1: Generic model of a data compression system.

3.3.2 Image data compression techniques

This subsection outlines mapping, quantising and coding methods which are commonly used where the data to be compressed represent still natural images.

The mapper

Among common techniques for performing the mapping from the highly correlated image space to a decorrelated domain are *linear prediction* and the use of transforms such as the Fourier, Hadamard, cosine and wavelet transforms [26]. Other methods, not described here, make use of filter banks to perform frequency domain subband decomposition [28]. Such *subband coding* systems are often used in applications where it is appropriate to apply perceptual weighting in the quantisation stage directly in the frequency domain to effect lossy compression without distortion being noticeable to the consumer. *Object-orientated* and *model-based* systems make use of source models to describe elements of an image or moving sequence in terms of shape, colour, etc., and transmitting the model information to effect data rate reduction. Strictly, such methods may not be considered to be mapping stages separate from subsequent quantisation and entropy coding stages, however; their operation is incompatible with the generic three-stage compression model. The use of linear prediction and the discrete cosine transform will now be discussed further.

Linear prediction [28, 29]: This method, also referred to as *Differential Pulse Code Modulation (DPCM)* decorrelates image data by forming an estimate of a given pixel value as a weighted sum of associated pixel values (the neighbours of the pixel). This prediction is then subtracted from the current pixel value and the result output to the decorrelated data stream. A first order linear predictor produces a pixel estimate based only on the value of only a single neighbouring pixel, usually the left neighbour. First order predictors are used especially in situations where no other associated pixel values exist, for example in temporal decorrelation of sequentially-transmitted video. Higher order linear predictors, where two or more neighbouring pixels are used in forming the prediction, require more complex implementations but provide more effective decorrelation.

The discrete cosine transform [28]: Orthogonal transforms are often used as the mapping stage [26]. The *Discrete Cosine Transform (DCT)* [30], a relative of the Fourier transform,

is popular for several reasons:

- Its decorrelation properties are near those of the only transform known to be optimal, the Karhunen-Loeve transform (KLT), which cannot be evaluated for arbitrary data and has no fast implementation;
- It produces purely real coefficients and therefore requires only real arithmetic processes to be implemented;
- It is orthogonal and separable, and therefore can readily be extended to multiple dimensions;
- Fast computation algorithms exist and hence implementation is practical for real time applications.

The forward DCT (FDCT) is generally applied on a two dimensional array of fixed size to blocks of image data in turn. The inverse DCT (IDCT) is used at the decoder to transform entropy decoded, dequantised data back to the image domain.

Run-length encoding: *Run-length encoding* is often used where significant runs of identical values are expected. A run of pixels is mapped to a pair of parameters (g_i, l_i) where g_i is the pixel value and l_i is the length of the run. This approach is particularly efficient in coding of bilevel images and has been used as part of a facsimile coding system [31].

The quantiser [26, 28]

The function of the quantiser is to discard image data selectively to improve overall compression efficiency. Quantisation is inherently a lossy process and therefore introduces noise. The quantiser is designed to maximise rate-distortion performance where the benefit to compression efficiency outweighs objectionable distortion due to the inevitable misrepresentation of its input data.

Since it operates in the decorrelated image domain, the quantiser must be matched to the mapping stage. Generally quantiser designs fall into two classes. *Scalar quantisation* methods operate independently on each decorrelated sample, while *vector quantisers* treat arrays of samples as independent quantisation 'units', substituting a code representing the best

match to the particular array of sample values. If the matching templates are chosen appropriately, the set of possible codes is significantly smaller than the set of all possible sample combinations in the array, resulting in data compression. The performance of an encoder using a vector quantisation scheme is dependent on the number and choice of templates available, with the goal being to produce the best possible match to each array, given the statistical characteristics of the data being quantised, while minimising the number of templates. It has been shown theoretically that vector quantisation always provides superior performance to scalar quantisation based on a rate-distortion criterion. However it has been found that for this to be the case in practical situations, a vector-orientated mapping stage must be designed to match the vector quantiser [32]. The use of vector mapping and quantisation methods in integral three dimensional image coding is outside the scope of this thesis, hence this outline is limited to a discussion of scalar quantisation techniques.

The transfer relationship of a general quantiser is shown in Figure 3.2. Here $\pm d_n$ are the *decision levels* of the quantiser and $\pm r_n$ are the *representation levels*. If the input lies in the interval $[d_1, d_2)$ it is quantised to an output level of r_1. The lossy nature of the process can be seen in that any input in this interval leads to the same output level. The figure represents

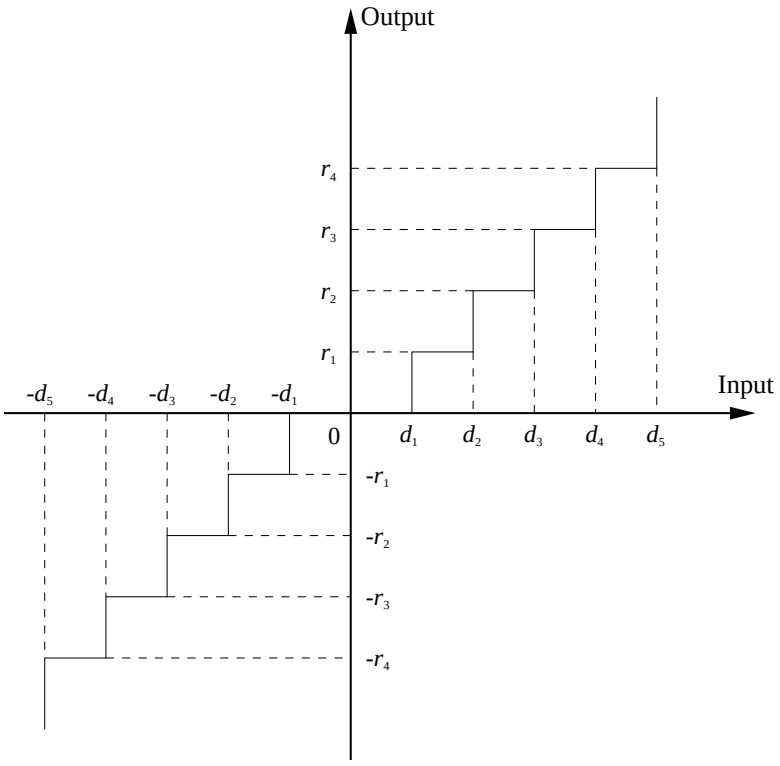


Figure 3.2: Transfer characteristic for a uniform symmetric quantiser.

a *uniform* quantiser, where the decision and representation levels are spaced equally, that is it has a constant step size. It is also *symmetric*, since negative values are represented by the same levels as positive values but with opposite sign.

The coder

The function of the coder is to reduce the statistical redundancy in the quantised data elements and produce a stream of uniquely decodable binary codewords at the bit rate required to give the desired compression ratio. A code may be described as *equal length* or *variable length*. An equal length code is a straightforward mapping from each possible quantised data symbol to a unique sequence of binary digits of fixed length. Examples are the natural binary code and the Gray code. A variable length code has no specified code length for any input symbol, and therefore the potential exists to produce far more efficient codes if the probability of occurrence of each symbol is taken into account when designing the code.

Huffman coding [27, 33]: A method for producing a highly efficient variable length code for a set of input symbols with defined probabilities due to Huffman [34] is often used. A minimum-length code results, that is one whose average output symbol length is less than or equal to that of all other possible uniquely-decodable codes. Unfortunately in a practical image compression system which must operate efficiently on all possible images with which it is presented, the coder input symbol probabilities are unknown. Two approaches for construction of a Huffman code for data compression are:

1. Evaluate symbol probabilities individually for every set of quantised image data presented to the coder by computing the frequencies of those symbols;
2. Produce a set of symbol probabilities based on the frequencies of a number of quantised image data sets produced for a number of images believed to be representative of the type of image data with which the compression scheme will be presented in the future.

Method (1), which gives rise to a *custom* or *local* code table, will always produce a minimum-length code but at the expense of extra computation required to generate the code in every compression cycle. It also has the disadvantage that since the code is unique to that set of input data it must be communicated to the decoder leading to a bit rate overhead, though this

is virtually insignificant in length alongside the coded data themselves. Method (2) (use of a *fixed* or *global* table) will never produce the optimum code but does not introduce extra computational or communication overhead. In practice both techniques are used dependent upon whether speed or compression performance is to be optimised [1].

3.3.3 The JPEG system [1, 2]

From 1985 an ISO/CCITT committee group known as the Joint Photographic Experts Group (*JPEG*) worked to produce a compression scheme for continuous-tone still images. As a result an international standard has been established, informally known as JPEG itself, which provides several encoding modes for various image compression applications. JPEG is a practical, popular scheme and contains many of the elements outlined in subsection 3.3.2. Hence it will be used to illustrate the use of those methods and to introduce certain other concepts utilised in the development of integral three dimensional image compression schemes. This summary description will cover only the most fundamental but also widely-used encoding mode, known as *baseline sequential-DCT JPEG*.

The encoder

Figure 3.3 represents the encoding process used by sequential-DCT JPEG in the format used for the generic compression system in subsection 3.3.1.

The subject image is presented to the encoder having initially been subdivided into blocks of 8×8 pixels each.

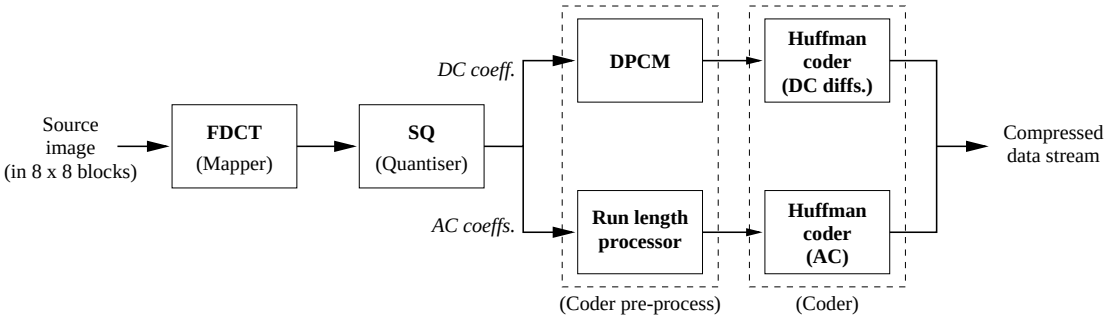


Figure 3.3: Baseline sequential-DCT JPEG encoder.

The mapping stage is comprised of a forward DCT ('FDCT'), which decomposes the source block into cosine weighting coefficients. Most of the energy in the source image block is concentrated in the top left coefficient (the *DC coefficient*) representing the mean intensity of pixels comprising the block. The remaining *AC coefficients* represent horizontal and vertical spatial frequencies increasing with distance from the DC coefficient.

The scalar quantisation stage ('SQ') used by the JPEG system effectively processes each DCT coefficient using a different uniform quantiser, allowing coarseness of quantisation to be set independently for each transform coefficient position. This enables the coefficients to be quantised based on the visibility of image data at specific coefficient positions to ensure that image features to which the viewer is particularly sensitive are retained while those which are not visible are discarded. Generally, psychovisually weighted quantisation stages such as these have been developed through measurements of the visibility of the 64 DCT basis functions [35, 36] and using experimental results on the spatial resolving capabilities of the human visual system [37]. The quantisation parameters which result are in the form of an 8×8 matrix of step sizes which correspond to transform coefficient positions. Table 3.2 shows the standard psychovisually weighted quantisation matrices used by JPEG for the luma (Y') and chrominance (C_b, C_r) signals derived from $R'G'B'$ colour image data. If S are the DCT coefficients and Q represents the quantisation parameter array, the quantised

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

a)

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

b)

Table 3.2: Standard JPEG quantisation matrices. a) Luma. b) Chrominance.

coefficients,

$$S_q(u,v) = \text{round}\left(\frac{S(u,v)}{Q(u,v)}\right) \qquad \text{for } u,v = 0,1,\ldots,7,$$
$$\text{with round}(a) = \begin{cases} \lfloor 1/2 + a \rfloor & \text{if } a \geq 0, \\ -\lfloor 1/2 - a \rfloor & \text{otherwise.} \end{cases} \tag{3.4}$$

Due to its simplicity this quantisation method lends itself to straightforward, fast implementation. The quantised data are likely to contain significant numbers of zero valued positions, particularly those corresponding to high spatial frequencies.

The DC coefficient resulting from quantisation generally has a significantly larger value and variance than the 63 AC coefficients. For this reason and the fact that the human visual system is most sensitive to the 'blocking' errors which result from inaccurately encoded DC coefficients, it is treated separately from the AC coefficients. The DC coefficient is differentially coded with respect to the previously-coded (8×8) block. The result is passed to a Huffman coder which may use either a fixed table of symbol probabilities or one generated locally for the particular image being compressed.

The AC coefficients are formatted into a one dimensional sequence by performing a 'zigzag' scan through the quantised block, as shown in Figure 3.4. This scan places the AC coeffi-

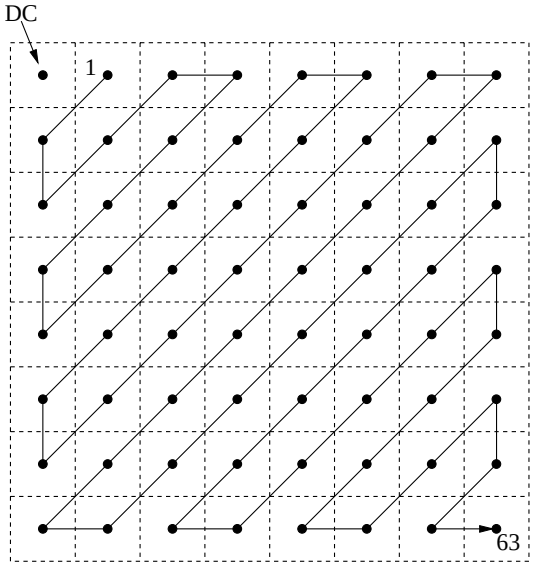


Figure 3.4: The 'zigzag' scan of quantised DCT coefficients used in Baseline sequential-DCT JPEG encoding.

cients in order of increasing spatial frequency and in general decreasing variance. The 63-coefficient sequence produced is highly likely to contain long runs of zero and possibly nonzero values. To increase the efficiency of the AC Huffman coder, the sequence is processed to represent its members by single codes combining their values and run lengths. The resulting codes are input to a second Huffman coder, again which may use either fixed or local symbol probabilities. Finally, the codewords produced by the DC and AC Huffman coders are combined and application-orientated control information is added to conform to the data syntax specified by the standard.

The decoder

The analogous decoding process for baseline sequential-DCT JPEG is illustrated in Figure 3.5. The incoming compressed stream is split into segments representing coded (8×8) blocks and each segment further divided into the coded DC coefficient difference and the coded AC coefficients. Each set of codewords is input to the appropriate Huffman decoder, which must have a copy of the *codebook* (list of symbols and associated codewords) used by the encoder. Having recovered the original symbols, the previously quantised DC coefficient is recovered by recombination with that decoded for the previous block and the 63 AC coefficients are recovered by the reverse of the run length processing used at the encoder. The quantised coefficient block is reconstructed from the DC coefficient and the AC coefficients by the reverse of the scanning process described previously.

The scalar dequantisation stage ('SDQ' in Figure 3.5) now normalises the quantised coefficient block using the quantisation table employed at the encoder such that it is suitable for

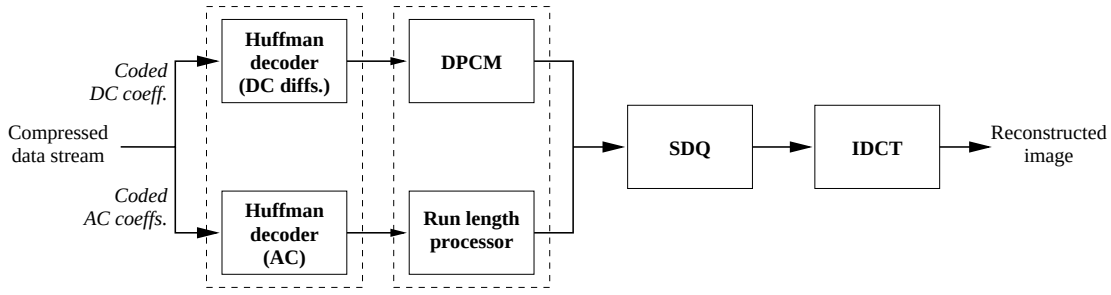


Figure 3.5: Baseline sequential-DCT JPEG decoder.

inverse transformation. The DCT coefficients to be input to the inverse transform,

$$R(u,v) = S_q(u,v)Q(u,v) \qquad \text{for } u,v = 0,1,\ldots,7. \tag{3.5}$$

Finally the inverse discrete cosine transform is applied ('IDCT') to produce the 8×8-pixel image block corresponding to the encoded input stream segment.

The compromise between data reduction and the fidelity of the reconstructed image, which is application-specific, may be varied by scaling the entire quantisation table or a selected area of it to adjust overall coarseness of quantisation. For compression of full colour images the source data are transformed into the $Y'C_bC_r$ colour space. Each channel is treated individually as described above, with a separate quantisation table (the chrominance table) used to quantise DCT coefficient blocks representing the C_b and C_r channels.

Summary

This chapter has outlined representation and broadcast of digital video sources and provided a summary treatment of techniques in image data compression. The ISO/IEC JPEG standard was used to illustrate the general compression concepts described for the fundamental application of still continuous-tone image compression.

Chapter 4 will describe methods developed for acquisition, preparation for input to a coding scheme and eventual display of integral three dimensional images.

4. ACQUISITION, PREPARATION AND REPLAY OF INTEGRAL IMAGE DATA

Introduction

As was stated in chapter 2, a captured integral image is represented entirely by a planar intensity distribution, formed at the focal plane of the encoding microlens or lenticular array. This may be recorded on to photographic film for later electronic scanning and processing, or directly on to an electronic imaging device such as a CCD array. The data resulting in either case are in the form of an array of pixels with values representing either white intensity (*grey level*) or red, green and blue intensities in the case of full colour capture.

To facilitate development of a compression scheme for integral three dimensional image data, suitable image data must be acquired, processed and displayed for evaluation. Since the integral imaging system is a unique technology there is no established framework for these procedures. This chapter describes methods developed to address this requirement.

4.1 Integral Image Acquisition

Four still test images were available for use in compression scheme development. These are shown, post processing, in appendix A. The physical characteristics of these images are listed in Table 4.1.

An indirect capture method is used, as summarised in chapter 2. The initial stage of the process consists of capturing the object scene using the still integral imaging camera. The resulting transparency, which carries the intensity distribution representing the volume of the scene, is submitted for photographic processing and is then ready for use.

Image	*Dimensions*	*Lenticular pitch / μm*	*Medium*	*Colour?*
Microscope	8” × 10”	1270	Paper print	No
Tank	5” × 4”	600	Transparency	Yes
Horseman	4” × 5”	600	Transparency	Yes
Toyplane	4” × 5”	1270	Transparency	Yes

Table 4.1: Physical characteristics of test integral intensity distributions.

4.2 Intensity Distribution Sampling

In order that digital processing of an integral intensity distribution for compression and transmission can take place, a sampling process is inevitably carried out by the electronic imaging element at the time of capture. The overriding consideration in choice of spatial sampling frequency is that a high fidelity three dimensional display is produced when the decoded intensity distribution is replayed. The sampling frequency chosen is dependent upon several factors of the image display system, the most important of which are the viewing distance, the pitch of the particular decoding lenticular array and the viewing cone angle produced. For a practical system the lower bound has been determined by the 3D Imaging Group to be 6 pixels horizontally per lenticule.

4.2.1 Use of Block-orientated Processing Techniques

In order to carry out any form of block-orientated digital processing as may be required by a compression scheme, it is necessary to determine the size of image data block for processing. Several important considerations influence this choice.

For any such processing algorithm the horizontal block dimension should map to the width of a lenticular band to allow for flexibility in the method used to take account of the inter-band redundancy identified. A lower bound on horizontal sampling frequency of 6 pixels per lenticular band has been determined to facilitate high quality integral image capture and display (see above). For practical reasons concerned with the feasibility of implementation of processing algorithms, it is desirable to set a block with sides which are powers of two. A processing block width of 8 pixels has been chosen to satisfy both of these considerations.

All electronic capture technologies for potential use with the imaging system employ a uni-

form horizontal and vertical sampling rate, that is, square pixels are produced. Therefore in deciding upon the processing block height, vertical sampling frequency cannot be considered. The principal consideration is the practicality of implementing block processing algorithms. To ensure flexibility in implementation, a square processing block is used.

The 8×8-pixel processing block to be used in integral three dimensional image compression will be termed an integral *subimage*.

4.3 Electronic Scanning of the Intensity Distribution Transparency

To produce data in a form suitable for compression development, the intensity distribution information recorded on the photographic transparency produced must first be digitised. Experiments have been conducted by the 3D Imaging Group on resolution requirements of captured integral image data [38]. These employed a high resolution drum scanner of a type used in pre-press stages in the commercial print industry. To avoid introducing distortion into a scanned image due to aliasing in the sampling process, it was understood that such high quality output scanners would implement band-limiting filtering techniques. This was confirmed by the observation that no such distortion was detectable in any of the integral intensity distribution data resulting from the scanning process.

4.3.1 Scanning parameters

The resolution at which scanning takes place must be chosen. This parameter is decided chiefly by the lateral resolution which is required in the resulting image data. Lenticular band width varies with the choice of lenticular sheet used in encoding of the intensity distribution during photographic capture, therefore scanning resolution for a particular instance is dependent upon the pitch of lenticular encoding sheet used during its capture. The minimum acceptable resolution $R_{s_{\min}}$ is the resolution $R_{s_{\text{proc}}}$ required to produce data with the correct scale for compression processing. If p is the lenticular pitch expressed in microns,

$$\begin{aligned} R_{s_{\min}} = R_{s_{\text{proc}}} &= \frac{8}{p/25400} \\ &= \frac{203200}{p} \quad \text{dots per inch (dpi).} \end{aligned} \tag{4.1}$$

A value for the *minimum* resolution required is produced, based on the fact that a scan performed with any greater resolution can be resized to the required target resolution of 8 pixels per lenticular band without detrimental effect on eventual quality of replay of the intensity distribution data.

It must be borne in mind that any single scanning device is capable of only a limited discrete number or range of scanning resolutions, therefore it is unlikely that a scanner can operate precisely at the target resolution $R_{s_{\min}}$.

4.3.2 Test image scanning

Images 'Tank' and 'Horseman'

A Linotype-Hell drum scanner, as employed in experiments on resolution requirements noted above, was used to digitise these intensity distributions.

Using information given in Table 4.1 and Equation 4.1, the minimum scan resolution for both images, $R_{s_{\min}} = 338.67$ dpi. The resolutions available using this scanner were found to be 12, 24 and 36 dots per millimetre (dpmm), the imperial equivalents being 304.8, 609.6 and 914.4 dpi respectively. A resolution of 24 dpmm was selected for scanning both transparencies.

Image 'Toyplane'

A Crosfield drum scanner was used to digitise the *Toyplane* image intensity distribution. The minimum scan resolution, $R_{s_{\min}} = 160$ dpi. A scan resolution of 16 dpmm (406.4 dpi) was decided to be the most suitable of those available, and was selected for use.

Image 'Microscope'

Since the image *Microscope* exists only as a paper print, it cannot be digitised using a drum scanner. Additionally this image was photographically captured in grey-scale and has a relatively coarse lenticular pitch. For these reasons it was expected that a high quality desktop

flat-bed scanner could be used in this case, the limitations of which would not be expected to lead to any significant degradations in the resulting image.

From Table 4.1 and Equation 4.1, the minimum scan resolution for this intensity distribution print, $R_{s_{\min}} = 160$ dpi. A Hewlett-Packard 'ScanJet' flat-bed scanner was used, which is capable of operating only at certain fixed resolutions. A resolution of 300 dpi was selected as this was greater than the minimum acceptable resolution and was a commonly used resolution which was known to produce satisfactory results when scanning conventional photographs.

4.4 Correction of Geometric Distortions in Scanned Intensity Distribution Data

For processing of the scanned intensity distribution data to take place in the development of compression schemes, the resolution must be such that each band is comprised of exactly 8 pixels horizontally, aspect ratio must be preserved, and lenticular bands must be precisely vertical.

It is highly likely that geometric distortions exist in the data resulting from the scanning process since:

1. The band structure on the original transparency or print is unlikely to have been aligned precisely with the scanner's vertical axis — this will introduce a rotational error;
2. The exact resolution at which the scanner operates is likely to deviate from the nominal value provided — this will add a scale error.

Methods employed for correction of these errors will now be described.

4.4.1 Rotational correction

This procedure is comprised of two stages. Firstly the degree of rotational error in the scanned data is ascertained, and secondly a correction is applied in such a way that the structure of the intensity distribution is preserved.

Figure 4.1 illustrates scanned data with an exaggerated clockwise rotational error. The

scanned intensity distribution data are viewed using the program 'display', part of the 'ImageMagick' image manipulation package [39]. This is an image display program capable of identifying individual pixel indices and values using a cursor. Points 'A' and 'B' are chosen as opposite extents of the centre of the low intensity area (black band) between two lenticular bands. The dimensions h and w are then taken and the error angle,

$$\theta_e = \tan^{-1} \frac{w}{h}. \tag{4.2}$$

If the data contain an anticlockwise rotational error, a similar procedure is followed with all measurements reflected around the vertical axis.

To effect correction the image must be rotated, in the opposite sense to the error, by an angle θ_e. To achieve this a routine is used which is capable of rotating an image by an arbitrary angle while preserving the original orthogonal sampling grid. A program 'pnmrotate' from the 'NetPBM' image processing package was used to carry out rotational corrections [40, 41].

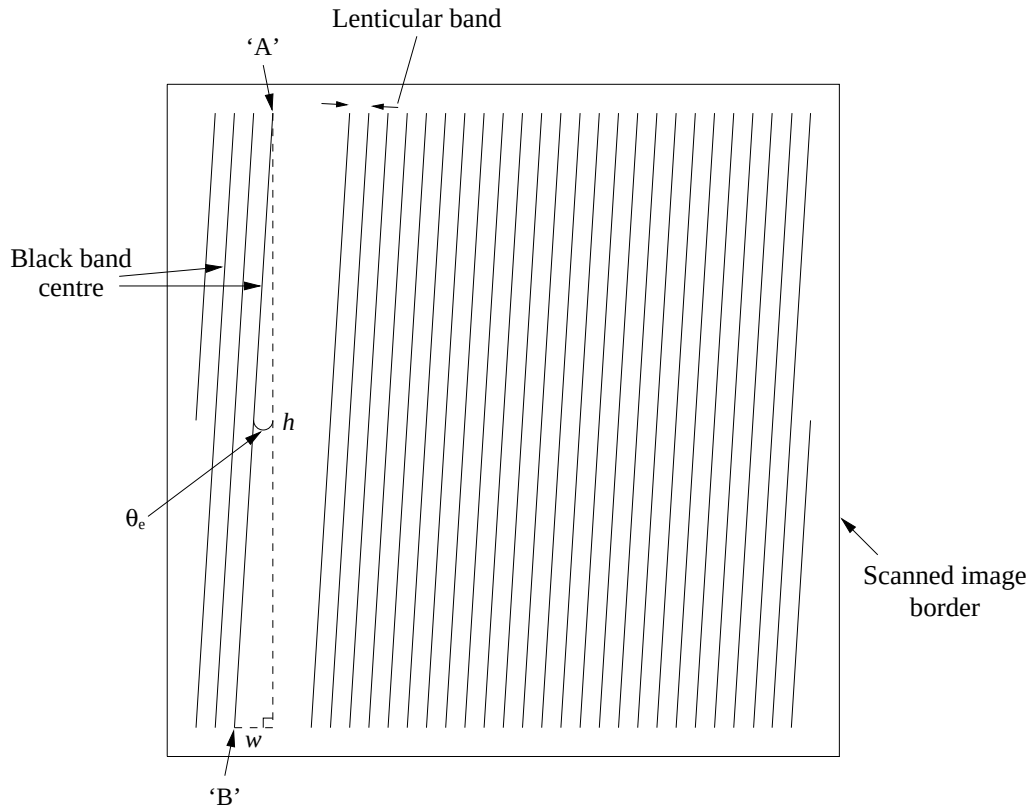


Figure 4.1: Rotational error in a scanned intensity distribution.

It was found that error angles and thus corrections required were in the range $0.5-2.0^{\circ}$.

4.4.2 Scale correction by direct examination

The purpose of this process is to correct the size of scanned data where the rotational correction procedure has been carried out, such that it is suitable for compression scheme development, with each lenticular band being composed of 8 pixels. Such a size correction is necessary since the true scan resolution often deviates from the nominal value given.

By the nature of the problem the precise scan resolution is unknown, therefore a value for this quantity must first be obtained from the data. A scale factor is then derived and a routine used to resize the data by this factor. The methods used will now be described.

Firstly the number of complete lenticular bands between two band edges is ascertained. Secondly the length of the line between the two chosen band edges is taken. The measurement points are chosen such that they are on the same vertical axis and are as close as possible to the horizontal extremes of the image data to minimise the significance of any experimental error in lenticular band measurement. The vertical axis used is that which contains the largest number of clearly defined bands near its horizontal extremes. Figure 4.2 illustrates the process. Note that in this figure all lenticular bands except those at the extremes of the data have been omitted for clarity.

If p is the pitch of the lenticular sheet used during capture (in microns) and n and l are the lenticular band count and pixel span of that count respectively then the true scan resolution,

$$R_{s_{\text{true}}} = \frac{25400 \times l}{np} \quad \text{dpi.} \tag{4.3}$$

Substituting Equation 4.3 into Equation 4.1, the scale correction ratio required,

$$\begin{aligned} S_i &= \frac{R_{s_{\text{proc}}}}{R_{s_{\text{true}}}} \\ &= \frac{8n}{l}. \end{aligned} \tag{4.4}$$

Since it is essential that the coherent structure of the intensity distribution is retained, the process used to carry out scaling must be chosen carefully. The ‘resize’ or ‘rescale’ functions available in some image processing packages were found to introduce periodic discontinuities into an image, resulting in a nonlinear distortion to the original regular sampling

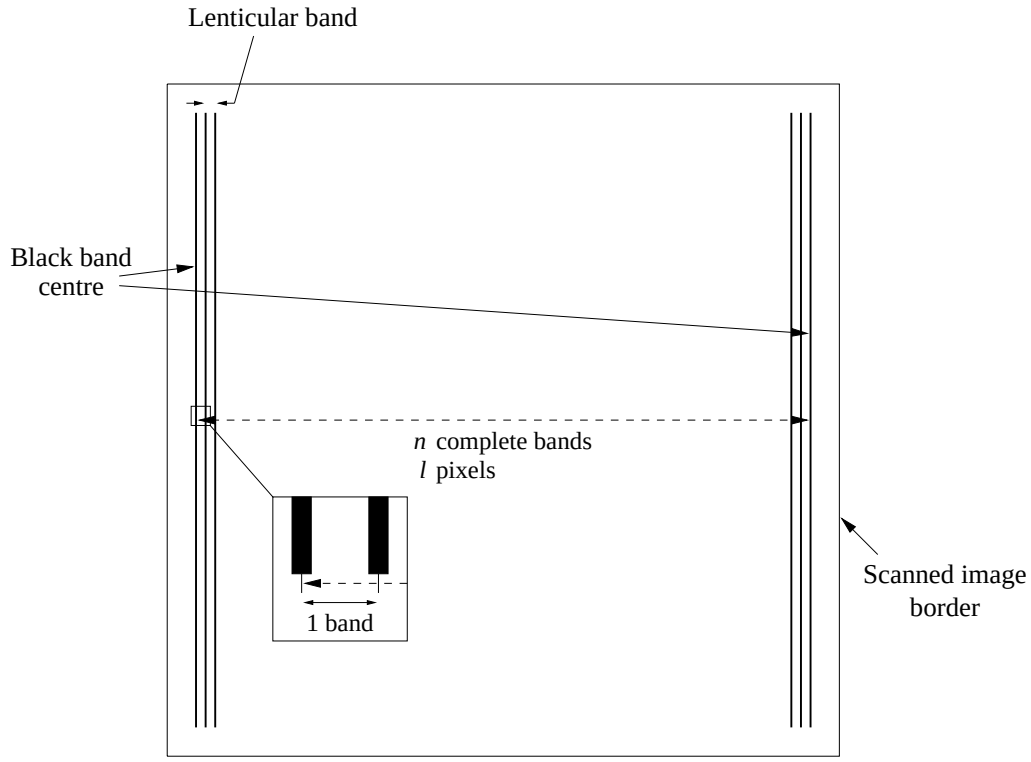


Figure 4.2: Determining scale error in a scanned intensity distribution.

structure. This renders the resulting image unusable.

A general filtered image rescaling program, 'pnmscale', from the 'NetPBM' package, is used to carry out the necessary resizing by the factor S_i. This program is capable of scaling an image by an arbitrary factor, using a technique which calculates the contributing source pixels to a given destination pixel and applies a filtering/resampling process to determine the value of each destination pixel without introducing aliasing [42]. The regular sampling structure in the original scanned data is preserved.

4.5 Noise Reduction in Scanned Intensity Distribution Data

During observation of scanned integral intensity distribution data, it was found that in certain cases noticeable information was present at high spatial frequencies, appearing to be evenly distributed across the data. This phenomenon can be seen in the intensity distribution at Figure 4.3.

This effect has been observed only in data originating in transparency form and electronically captured using a high resolution drum scanner. It is postulated that it originates as the result of an interaction between the drum scanning process and the fine grain structure of the transparency film used to capture the scene. Figure 4.4 shows an intensity distribution originating from a print on photographic paper in which the grain structure is significantly coarser. These data, which were imaged using a medium resolution desktop scanner, do not suffer from the same effect.

This effect is therefore considered as noise for the purpose of integral three-dimensional image display and should be reduced as far as possible in order not to bias the performance of an integral image compression scheme.

4.5.1 Frequency domain noise reduction filter design

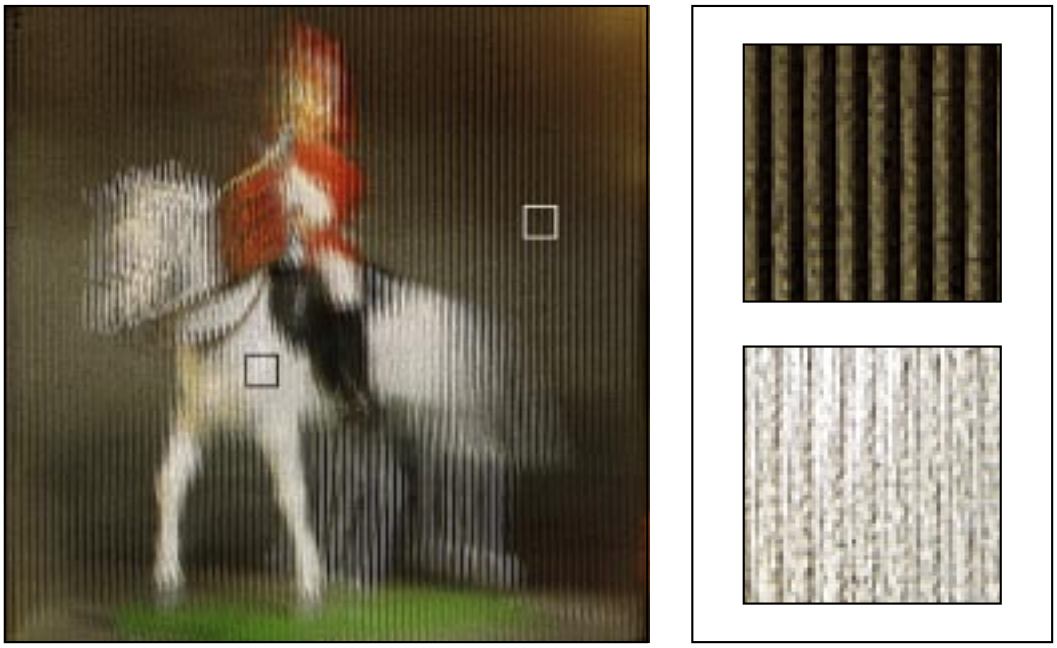
Spatial domain methods such as neighbourhood averaging and median filtering are often used in high frequency noise reduction applications for conventional two dimensional image data [26]. Spatial domain filtering techniques however are considered unsuitable for use with integral intensity distribution data. This is since unlike conventional images, integral intensity distributions inherently contain a structure which is fundamental to successful three dimensional display. The matrix operations employed by spatial domain methods are likely to disturb this structure by operating across lenticular band boundaries.

In the frequency domain filters can be designed with great flexibility for a specific application, such that they do not adversely affect the band structure in the integral intensity distribution. Frequency domain filtering methods operate by attenuating unwanted components of the Fourier spectrum of an image. If $f(x,y)$ is the subject image data, then the filtered output image is

$$g(x,y) = \mathscr{F}^{-1}\{H(u,v).\mathscr{F}\{f(x,y)\}\}, \tag{4.5}$$

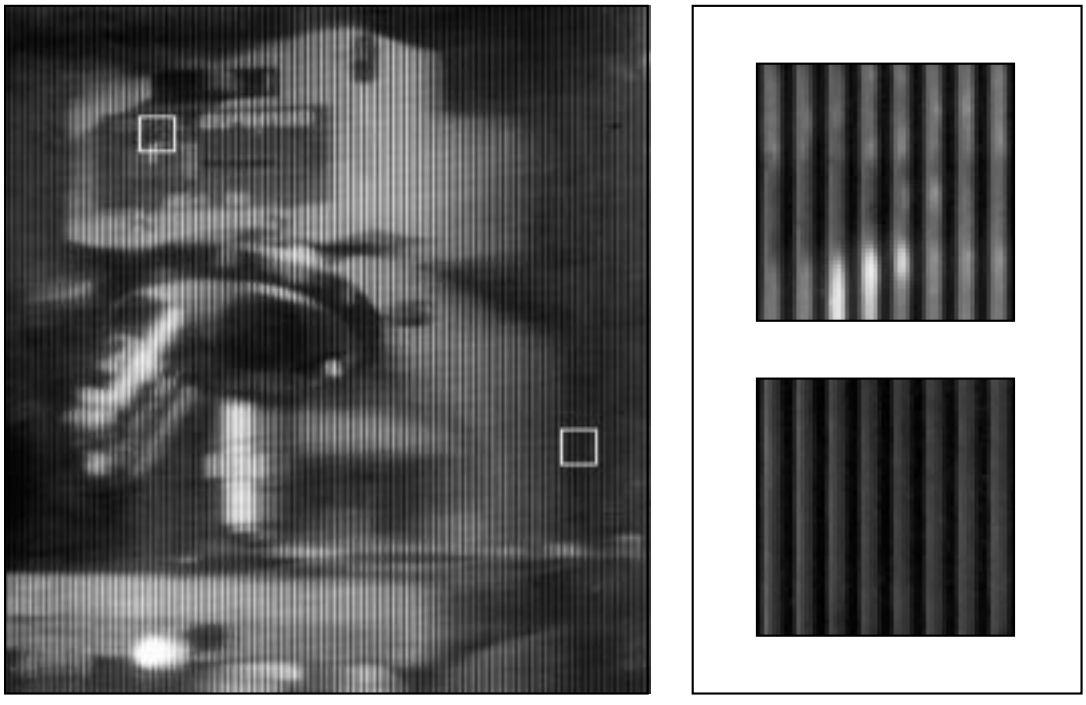
where $H(u,v)$ is the filter transfer function and $\mathscr{F}\{\}$ and $\mathscr{F}^{-1}\{\}$ represent the forward and inverse Fourier transform operations respectively [26].

In order to select a suitable transfer function for this application the general form of frequency spectra of the scanned and geometrically corrected intensity distribution data must be examined. Figure 4.5 is a display of $\log(1+|\mathscr{F}\{f(x,y)\}|)$ for the luma component (Y') of the scanned and geometrically corrected intensity distribution *Horseman*. The spatial fre-



a b

Figure 4.3: Randomly distributed high frequency image information in a scanned intensity distribution. a) Full image. b) Magnified 64×64-pixel areas.



a b

Figure 4.4: A scanned intensity distribution without information at high spatial frequencies. a) Full image. b) Magnified 64×64-pixel areas.

quency origin has been shifted to the centre of the transform image. Also to the right and top of the display are shown plots of the maximum value of the transform along each row and column respectively.

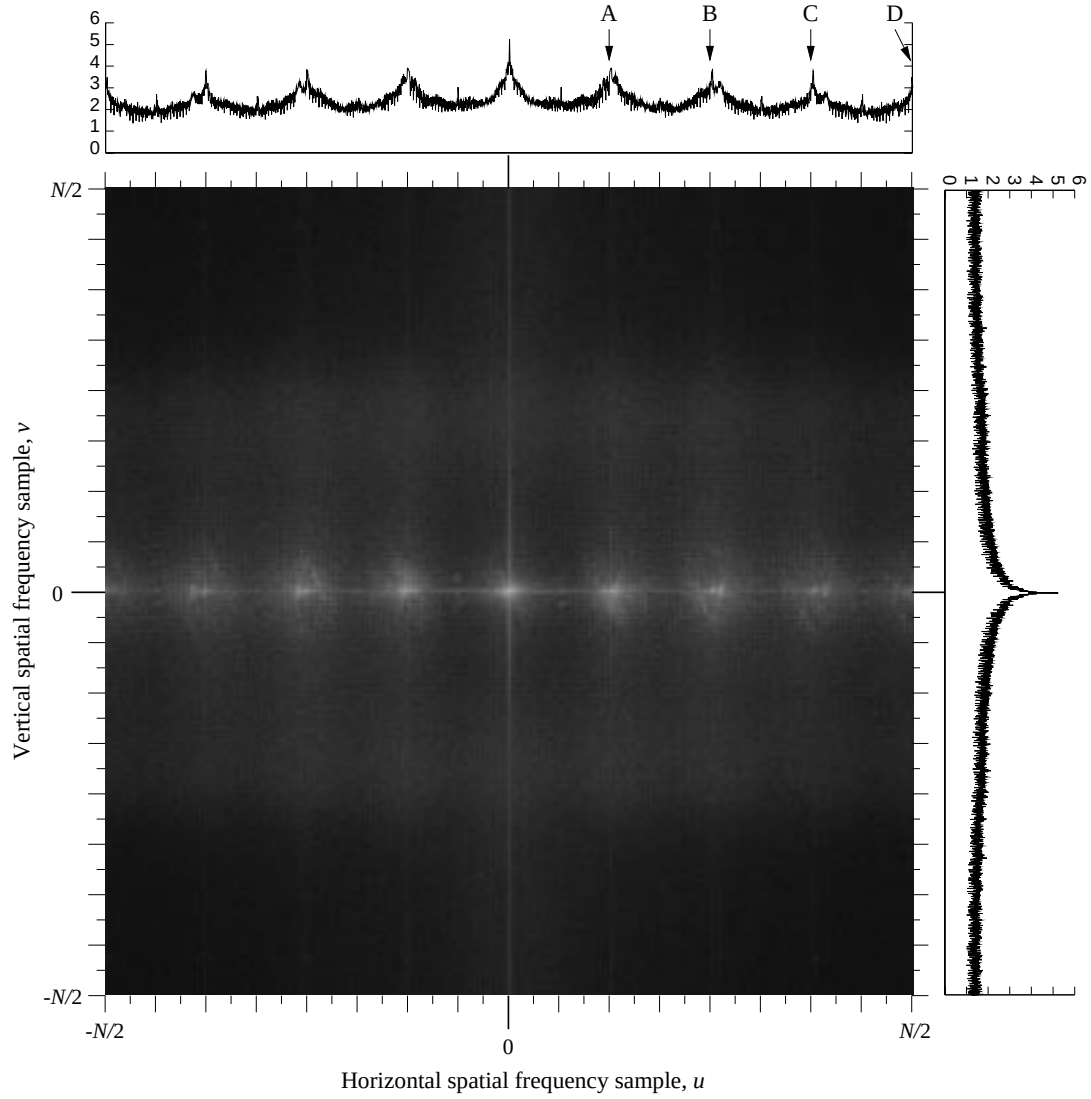


Figure 4.5: Intensity plot of frequency spectra for *Horseman*.

Peaks can be seen at certain values of u for all v at points 'A'–'D', especially significant around the vertical origin. These are attributed to the nature of the lenticular band structure. Since the great majority of image information which contributes to integral three dimensional replay is concentrated within this area, a filter operating in the vertical dimension only will be used to reject unwanted components. The transform magnitude was examined for the luma component of a second intensity distribution captured using the same technique. Its

general form was found to be identical to that of *Horseman*, described above.

The transfer function of the general (two dimensional) Butterworth filter is defined by

$$H(u,v) = \frac{1}{1+[D(u,v)/D_0]^{2n}}, \tag{4.6}$$

where n is the order of the filter,

$$D(u,v) = \sqrt{u^2+v^2}$$

is the distance from the point (u,v) to the origin of the frequency domain and D_0 represents the distance of the cut-off frequency locus from the origin [26]. For the case of the filter to be designed for integral intensity distribution noise rejection, where no filtering is to be performed in the horizontal spatial frequency dimension, $D(u,v) = |v|$ and the transfer function,

$$H(u,v) = \frac{1}{1+[v/v_0]^{2n}} \tag{4.7}$$

where v_0 is the vertical cut-off frequency.

Experiments were conducted to determine the optimum cut-off frequency using values for v_0 of 64, 128, 192 and 256. Magnified sections of the images resulting from these experiments are shown in Figure 4.6. The areas shown are taken at the same positions as those in Figure 4.3. The optimum cut-off frequency value, which introduced no noticeable distortion to the replayed three dimensional image while having sufficient noise rejection performance, was found to be 128. The results where a filter of order 2 was used were deemed to be satisfactory, therefore it was not felt necessary to implement the filter with a higher order. The response of the filter designed is shown in Figure 4.7 superimposed on to a plot of the maximum transform value along each row for two intensity distributions. The response is plotted only over one half of the transform, sufficient since $f(x,y)$ is purely real and the transform is then symmetrical about its origin.

The filter was implemented in the frequency domain using the 'Cantata' visual programming interface, part of the 'Advanced Khoros' software system [43].

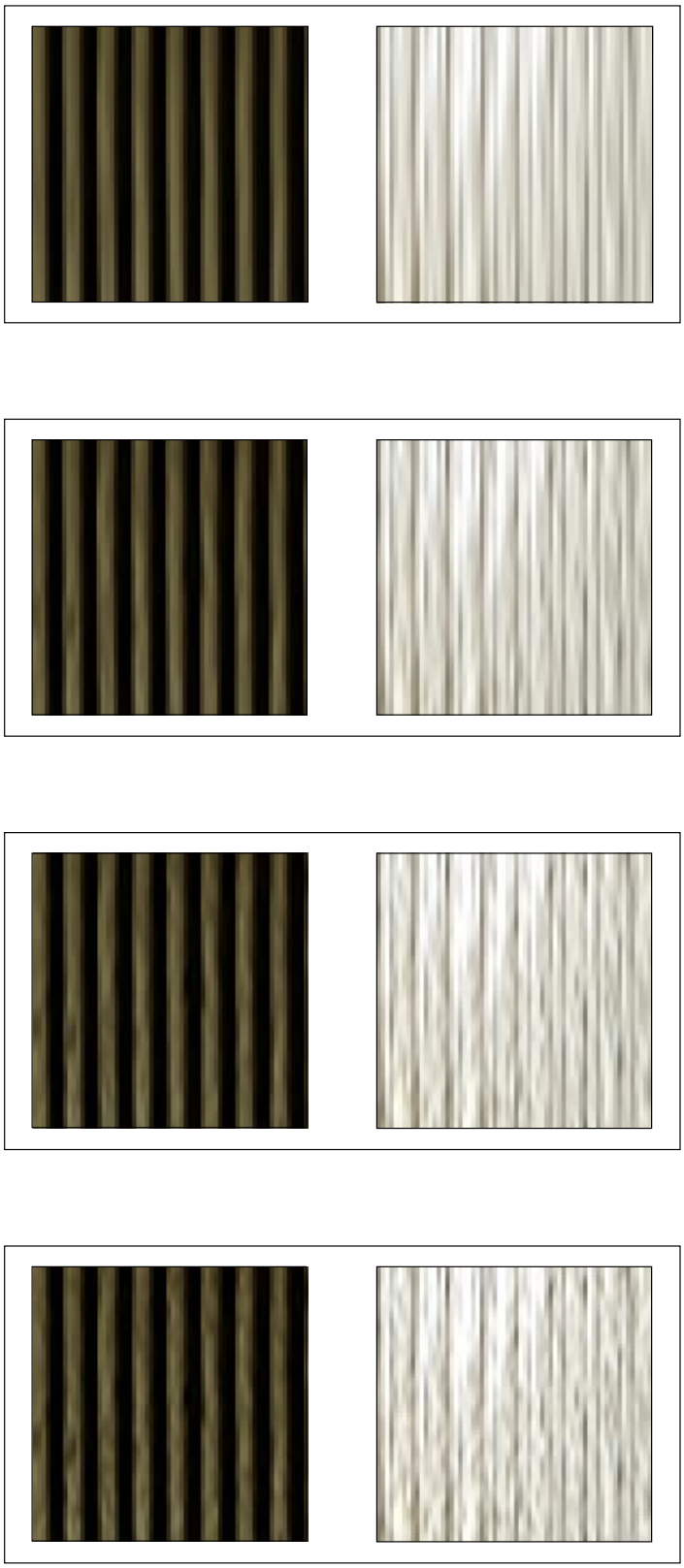


Figure 4.6: Intensity distribution *Horseman* processed using a 1D Butterworth noise rejection filter applied to each column. a) $v_0 = 64$. b) $v_0 = 128$. c) $v_0 = 192$. d) $v_0 = 256$.

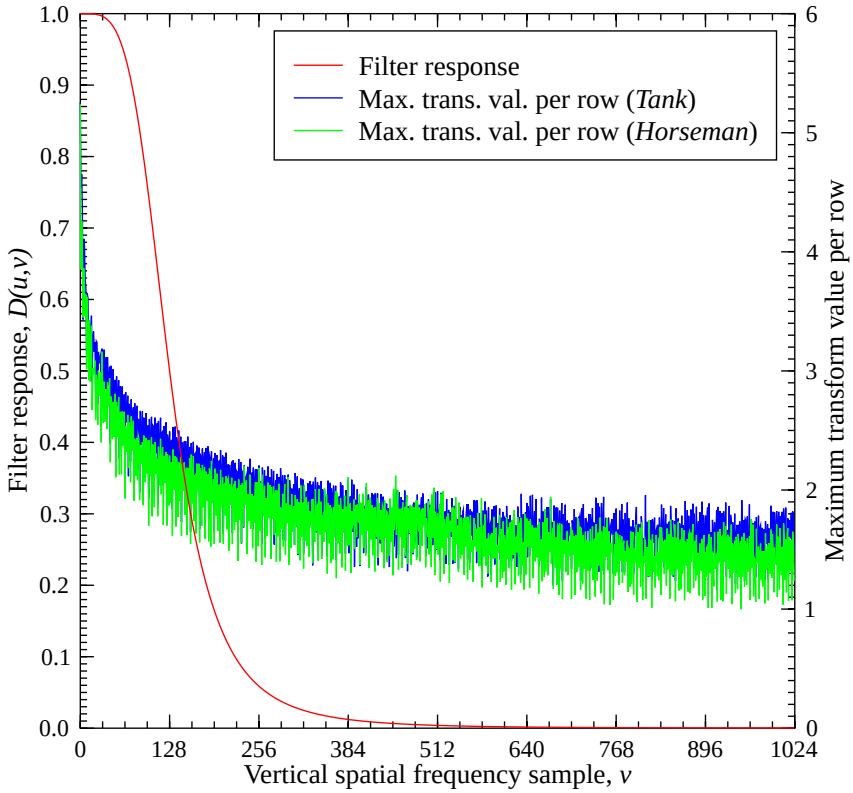


Figure 4.7: Noise reduction filter frequency response.

4.5.2 Use of one dimensional spatial filtering

The noise rejection filter described was implemented in the frequency domain primarily because the design process involved examination of the frequency domain representation of the integral intensity distribution. However following the decision that the filter should operate only in the vertical transform dimension and thus image dimension, a spatial domain filter can be used which operates only in the vertical dimension. This allows the use of a larger set of filtering techniques, including nonlinear methods which may give results superior to the frequency domain method previously employed.

Median filtering is a nonlinear technique which is often used to remove unwanted spurious features from images with a reasonably high level of detail [26]. It has an advantage over linear filtering methods such as Butterworth filtering since it tends not to introduce a 'blurring' effect — it effectively causes points with distinct intensities to take on intensity values closer to those of their neighbours.

A median filter is generally implemented using a neighbourhood of pixels forming a square matrix. It is the likely detrimental effect of use of such a matrix that was decided to mitigate against the use of such spatial filters across integral intensity distributions. However a one dimensional median filter, operating only in the vertical dimension is more appropriate. Figure 4.8 represents a section of an integral intensity distribution, $f(x,y)$. In the figure a one dimensional median filter with a window length n of 5 is currently operating on the pixel value $f(1,4)$. The intensities of pixels $f(1,2\ldots6)$ are sorted and the median value assigned to $f(1,4)$. This process is repeated for each pixel in the intensity distribution.

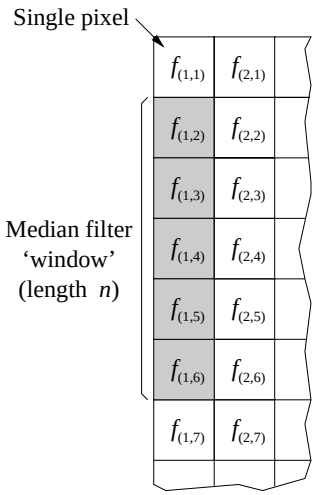


Figure 4.8: Operation of a one dimensional median filter.

A filter of this type was implemented, as before using Khoros' 'Cantata' system. The implementation was used to carry out one dimensional median filtering on the *Horseman* intensity distribution, using window lengths of $3,5,7\ldots25$. Where the filter window overlaps the vertical boundary of the data, any values outside the boundary are taken to be zero. Magnified sections of the resulting intensity distribution data are shown in Figure 4.9. The sections shown are again taken at the same positions as those in Figure 4.3. From examination of the results, magnified sections of the intensity distribution for $n<7$ still exhibited significant noise. When replayed as three dimensional images, it was found that for $n>17$ loss of edge definition occurred. It was decided that on the basis of these subjective evaluations the optimum window length was 11 pixels. This vertical one dimensional median filter was used to process the *Horseman* and *Tank* intensity distributions prior to their use in developing the coding schemes described in the following chapters.

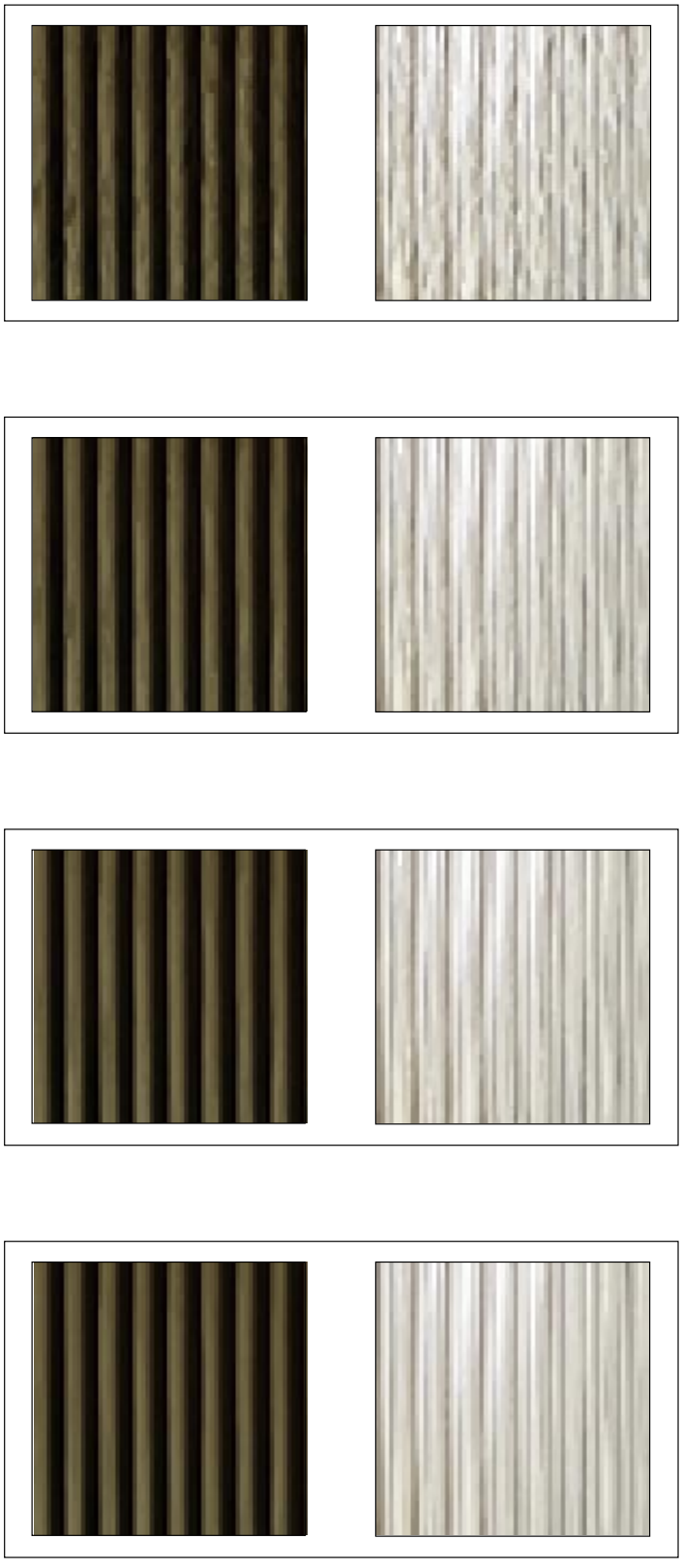


Figure 4.9: Intensity distribution *Horseman* processed using a 1D median filter applied to each column. a) $n = 5$. b) $n = 11$. c) $n = 17$. d) $n = 25$.

4.6 Three Dimensional Display of Processed Integral Intensity Distributions

In order to assess the fidelity of image data resulting from lossy compression and decompression of integral intensity distributions they must be replayed in their final form, as integral three dimensional images. As described in chapter 2 fundamentally this involves the display of the intensity distribution on a plane surface and reintegration of the object scene through the use of an overlaid lenticular decoding element. For the purpose of this thesis, prints on paper produced by a colour hardcopy device are used as the display medium.

A suitable hardcopy device must be mechanically stable and have an accurately specified printing resolution in order that the displayed intensity distribution is dimensionally consistent with the original photographic capture medium. Dye sublimation printers were found to comply with these requirements and also produce a close approximation to block pixelation due to the fact that they produce continuous-tone colour without use of halftone techniques. A Tektronix 'Phaser 480X' dye sublimation printer was used to display processed intensity distributions. It operates at a resolution (R_o) of 300 dpi.

To prepare processed intensity distribution data for output, they must first be scaled in such a way that the dimensions of the original scanned photographic print or transparency are preserved in the display. Making use of Equation 4.1, the output scale ratio required for an output resolution of 300 dpi is:

$$\begin{aligned} S_o &= \frac{R_o}{R_{s_{\text{proc}}}} \\ &= \frac{3p}{2032} \end{aligned} \tag{4.8}$$

where p is again the pitch (in microns) of the lenticular encoding sheet used during capture. The data to be displayed are resized by a factor S_o using the 'pnmscale' program (refer to subsection 4.4.2).

4.7 Evaluation of Replayed Image Fidelity

The decoded image output of a coding scheme employing a quantisation stage (a lossy scheme) is never identical to the source image. Since the losses inevitably affect the quality of the decoded image, it is important to be able to measure decoded fidelity as part of the evaluation process for a lossy compression scheme. Unfortunately it is difficult to provide meaningful measures of image quality. Many researchers have investigated this problem for the case of conventional two dimensional images [44, 45], which arises from the difficulty of providing a metric which is both explicit and relevant in terms of quality as perceived by a human observer. No perceptually relevant quality criteria have been found for the case of viewing of volume spatial optical models as produced by the integral three dimensional imaging system, therefore an objective fidelity metric, applied to the decoded planar intensity distribution, is used to evaluate decoded integral image quality in this thesis. Objective quality metrics, despite not taking into account the characteristics of the human visual system, produce explicit results which are generally relevant for relative comparisons, and can be produced directly using a single calculation for any image.

The peak signal to noise ratio is widely used in the field and readily recognised when presented in evaluation of lossy coding schemes. For a source image array $f(x,y)$ with dimensions $X \times Y$, the peak signal to noise ratio PSNR of the decoded image array $g(x,y)$ where each pixel is represented by n bits [33],

$$\mathrm{PSNR} = 20 \log_{10} \left[\frac{2^n - 1}{\mathrm{RMSE}} \right] \quad \mathrm{dB}, \tag{4.9}$$

where RMSE is the root mean square error in the decoded image array:

$$\mathrm{RMSE} = \sqrt{\frac{1}{XY} \sum_{y=0}^{Y-1} \sum_{x=0}^{X-1} \left[g(x,y) - f(x,y) \right]^2}. \tag{4.10}$$

The PSNR is employed in this thesis as a measure of distortion in decoded integral image data. It is applied directly to the source and decoded planar intensity distributions, its value decreasing with increasing distortion.

For the case of greyscale images, which contain only a single intensity channel, Equation 4.9 is used to compute the PSNR. Where colour images are to be evaluated a mean value, weighted by factors determined by human visual sensitivity to errors in each of the three

channels R' (red), G' (green) and B' (blue) [46, 47], is produced:

$$\mathrm{PSNR}_{\mathrm{rgb}} = 20 \log_{10} \left[\frac{2^n - 1}{0.299\,\mathrm{RMSE}_{R'} + 0.587\,\mathrm{RMSE}_{G'} + 0.114\,\mathrm{RMSE}_{B'}} \right] \quad \mathrm{dB}, \quad (4.11)$$

where $\mathrm{RMSE}_{R'}$, $\mathrm{RMSE}_{G'}$ and $\mathrm{RMSE}_{B'}$ are given by applying Equation 4.10 to the R', G' and B' channels respectively of the source and decoded intensity distributions. In the intensity distributions under consideration in this thesis each channel is represented using 8 bits ($n = 8$).

4.8 An Alternative Scale Correction Method using Moire Analysis

Subsection 4.4.2 described a method for correction of the size of scanned intensity distribution data which may be necessary due to uncertainty about the precise resolution at which scanning took place. Since the method relies upon counting of small features in the image data, it is inherently error prone and several passes are typically made to ensure that the count is accurate. This section develops an alternative technique which does not have this disadvantage and hence is expected to produce more reliable results.

If two gratings of parallel lines, each of different pitch are placed together with their lines running in the same direction, interference is produced in the form of Moire fringes [48]. Moire fringes are often an unwanted feature of optical systems, particularly in imaging and printing systems, and significant design effort is concerned with avoiding them. However, the phenomenon has been used successfully in metrology applications such as mechanical strain analysis since by measurement of the fringes the difference in pitch between the two gratings can be determined.

When an attempt is made to replay the three dimensional image encoded in an integral intensity distribution using a lenticular sheet whose pitch is mismatched from that of the banded structure in the intensity distribution, the same effect is produced. Figure 4.10 is a photograph of an image in replay which exhibits Moire fringes for this reason. The intensity distribution was scanned at a nominal resolution of 300 dpi. By considering the displayed intensity distribution and the decoding lenticular sheet as coarse gratings, the theory of Moire fringes can be used to quantify the pitch mismatch and hence determine a scale correction for the intensity distribution to give correct three dimensional replay. The pitch of the lenticular decoding sheet, p, is known, therefore the problem is to determine the pitch of the image bands in the replayed intensity distribution with which the decoding sheet lenticules must

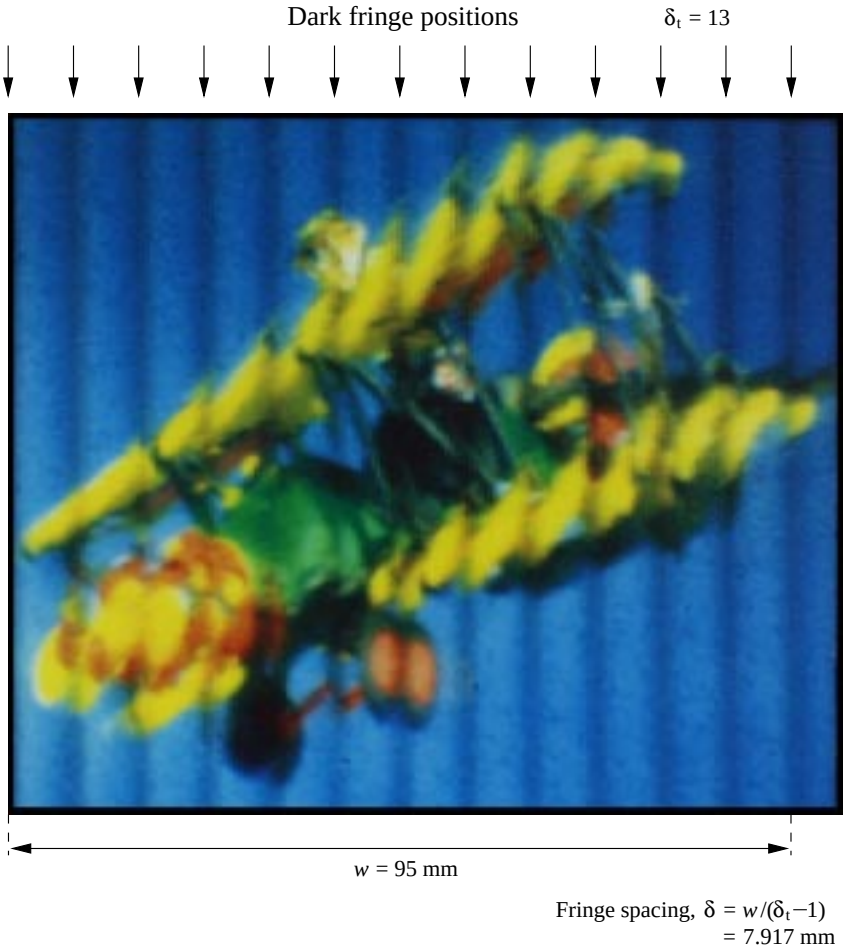


Figure 4.10: Moire fringes observed on a replayed integral intensity distribution (not to scale).

align.

The scanned intensity distribution is first scaled for display at the resolution of the hardcopy device (300 dpi) directly from the nominal scan resolution and then printed.

Figure 4.11 illustrates the formation of Moire fringes caused by a pitch mismatch during replay. In the figure the pitch of the bands on the displayed intensity distribution (p_d) is greater than that of the lenticular decoding element (p). Where the centre of a lenticule of the decoding sheet coincides with the centre of an image band, a light Moire fringe is produced in the form of an area of correctly displayed image. Conversely at points at which a lenticule is aligned with the black band border between two image bands, a dark fringe arises. If m represents the number of lenticules between two dark fringes and δ is the fringe

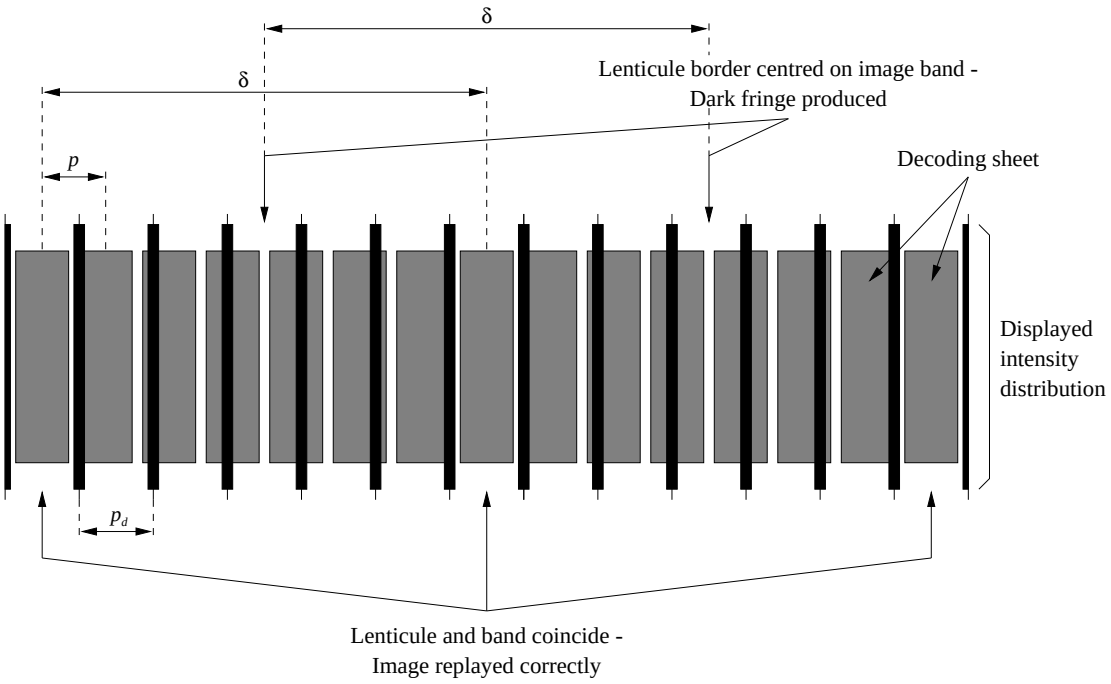


Figure 4.11: Formation of Moire fringes due to pitch mismatch in integral three dimensional image replay.

spacing then

$$\delta = mp = (m \pm 1)p_d. \tag{4.12}$$

It is only possible using this method to calculate the *magnitude* of the pitch mismatch — it cannot be determined whether the displayed pitch is greater or smaller than that of the lenticular decoder, hence the use of the '$\pm$' sign in Equation 4.12.

Using Equation 4.12 an expression for the pitch of the displayed intensity distribution bands can be derived:

$$p_d = \frac{mp}{m \pm 1} = \frac{\delta}{\delta/p \pm 1}. \tag{4.13}$$

Knowing p_d, the disparity between displayed and correct pitch can be determined and hence

the scale correction required to be applied to the scanned data may be calculated directly:

$$\begin{aligned} S_m &= \frac{p}{p_d} \\ &= \frac{p}{\delta/(\delta/p \pm 1)} \\ &= 1 \pm \frac{p}{\delta}. \end{aligned} \tag{4.14}$$

Equation 4.14 gives two possible values for S_m, one corresponding to an increase and the other to a decrease in the scale of the scanned intensity distribution data.

A value for δ is obtained from the intensity distribution previously printed, as the distance between two neighbouring dark fringes. To minimise the effect of experimental error in measurement, this is carried out by counting the number of dark fringes observed (δ_t) inclusively between two points each in the centre of a dark fringe on the displayed image; the distance between these points is measured as w. Then the fringe spacing $\delta = w/(\delta_t - 1)$. This method is illustrated on the photograph of Figure 4.10. Since the direction of scale correction required is unknown, the data are then both enlarged:

$$S_i = S_{m_+} = 1 + \frac{p}{\delta}, \tag{4.15a}$$

and reduced:

$$S_i = S_{m_-} = 1 - \frac{p}{\delta}. \tag{4.15b}$$

In previous use of p to represent lenticular pitch, the quantity has been expressed in microns. This must be taken into account in the use of Equations 4.15a and 4.15b and either δ specified in microns or p given in millimetres such that both have the same unit.

The 'pnmscale' program (refer to subsection 4.4.2) is used to carry out resizing and finally both sets of intensity distribution data are scaled to hardcopy resolution (300 dpi) and output to the printer. The lenticular decoding element is overlaid and the correct scale factor (S_{m_+} or S_{m_-}) can be determined as the intensity distribution corresponding to it will replay as a three dimensional integral image without Moire fringes. In the example of Figure 4.10, $S_{m_+} \approx 1.076$ and $S_{m_-} \approx 0.924$. In this case reduction was the correct scaling, that is $S_i = S_{m_-}$.

Summary

This chapter has covered the acquisition, translation into a common format and display of integral three dimensional images as required for the development of integral image compression schemes. The following processes are involved:

1. The original photographic medium on to which the subject scene has been encoded is scanned electronically;
2. Geometric distortions in the scanned data are removed: rotational errors may be corrected, and scale may be corrected either via direct inspection of the intensity distribution or by analysis of Moire fringes in replay;
3. A filtering process may be applied if necessary to reject high frequency noise which may have been introduced at the scanning stage;

   At this point the data are in a form suitable for digital processing by an integral image compression scheme;
4. The resulting intensity distributions are resized to the display resolution and output for evaluation.

Chapter 5 will discuss the structure of captured lenticular-integral image data and address their statistical characteristics. Conclusions are drawn concerning compression scheme design based on the observations made.

5. EVALUATION OF THE CHARACTERISTICS OF CAPTURED INTEGRAL IMAGE DATA

Introduction

To facilitate the design of an efficient class of compression schemes for use with integral three-dimensional television pictures, the nature of the recorded intensity distribution data with respect to redundancies present must be understood. To this end, this chapter investigates the structure of the data representing the imaged object scene.

5.1 The Captured Intensity Distribution

Figure 5.1a shows an intensity distribution produced by the electronic integral image capture process where a lenticular sheet has been used as the encoding element. Figure 5.1b magnifies sections of this intensity distribution, showing the effect of varying levels of detail and also of scratch defects on the encoding array. The regular structure of the cylindrical lenses on the encoding array gives rise to a regular banded structure in the intensity distribution.

Observations of the captured intensity distribution indicate that in general a band is comprised of two distinct components (see Figure 5.2):

1. A central section containing intensity information significant in reconstruction of the captured scene;
2. Two low intensity borders to the left and right of the central component.

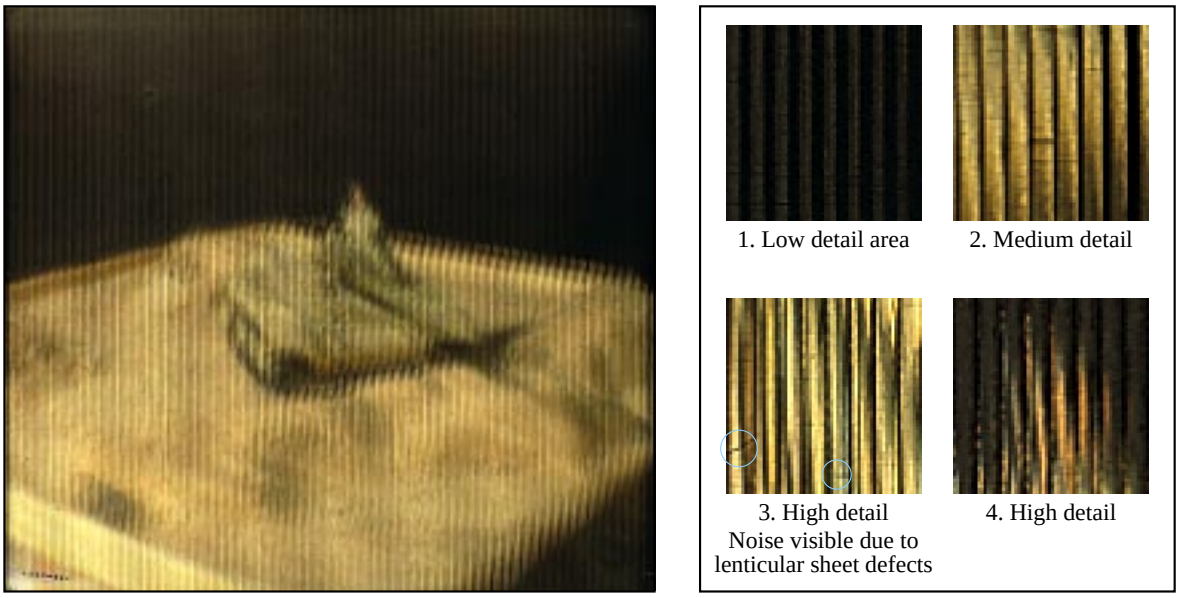


Figure 5.1: Example integral image intensity distribution captured using a lenticular encoding element. a) Full image. b) Magnified sections.

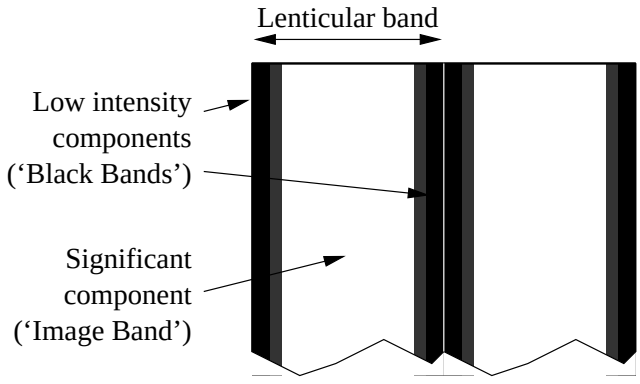


Figure 5.2: General structure of the lenticular-integral intensity distribution.

5.2 Inter-band Correlation

An inter-band correlation exists in the recorded intensity distribution which is unique to an integral image, since all bands on the intensity distribution have the same general structure as described above. This is borne out by observations of recorded intensity distributions (Figure 5.1).

The data reduction potential of a compression system is governed by its ability to exploit redundancies present in the source data stream due to correlation between data elements. For this reason, in order to maximise the efficiency of a compression scheme for use with the integral image intensity distribution, inter-band correlation should be evaluated and exploited.

5.2.1 Evaluation of Inter-band Correlation

To provide a quantitative analysis of inter-band correlation, the normalised cross correlation coefficient was evaluated between pairs of neighbouring bands in lenticular-integral image intensity distributions [26]. If f_i and f_{i+1} are a pair of adjacent bands each with horizontal and vertical sides of L and M respectively, the normalised cross correlation coefficient evaluated between bands i and $i+1$,

$$r(i) = \frac{\sum_{l=0}^{L-1}\sum_{m=0}^{M-1} f_i(l,m)\, f_{i+1}(l,m)}{\left[\sum_{l=0}^{L-1}\sum_{m=0}^{M-1} f_i^2(l,m) \times \sum_{l=0}^{L-1}\sum_{m=0}^{M-1} f_{i+1}^2(l,m)\right]^{\frac{1}{2}}}. \tag{5.1}$$

The luma channels of the four intensity distributions *Microscope*, *Tank*, *Horseman* and *Toyplane* (see appendix A) were used in this experiment. Values of $r(i)$ are plotted for each in Figure 5.3, and the mean values of normalised cross correlation coefficient for each intensity distribution, evaluated between all contiguous band pairs on each, $\bar{r}$, are listed in Table 5.1. The low values of $r(i)$ at the extreme left and right of the data sets for *Tank* and *Horseman* visible in Figure 5.3 arise from discontinuities at the edges of the encoding lenticular sheets. These result from small areas in which captured data were not fully covered by lenslets, and hence do not relate to valid integral image data. The cross-correlation values are, in general, greater than would be expected for a conventional image. Considering the unpredictable nature of natural image pixel values, both conventional two-dimensional and

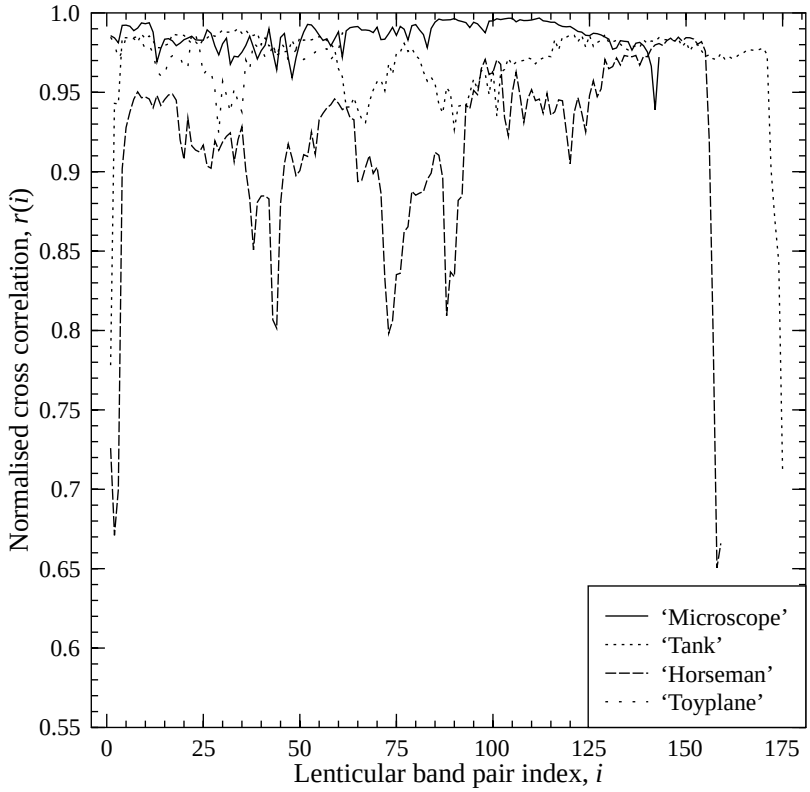


Figure 5.3: Normalised cross correlation coefficient evaluated between adjacent bands on the luma channel of four lenticular-integral image intensity distributions.

Intensity distribution	$\bar{r}$
Microscope	0.9858
Tank	0.9688
Horseman	0.9176
Toyplane	0.9713

Table 5.1: Mean normalised cross correlation evaluated between adjacent bands on the luma channel of four lenticular-integral image intensity distributions.

three-dimensional, the observed behaviour of the cross correlation coefficient value confirms that considerable significance should be attached to the inter-band correlation inherent to integral three-dimensional intensity distributions.

5.3 Intra-band Correlation

In compression of two-dimensional natural images, data reduction is achievable chiefly due to the fact that the spatial frequencies encountered are generally low. The technique of linear prediction, relies solely on this significant 'inter-pixel correlation'. Natural three-dimensional integral intensity distributions have been observed to share this property in a local sense within the bounds of individual lenticular bands. This will be termed 'intra-band correlation'.

5.3.1 Evaluation of Intra-band Correlation

A quantitative analysis of this property has been conducted. In this case it is relevant to compare results for integral intensity distributions with those determined for a representative set of natural two-dimensional images. Differential images were calculated for the luma channels of each of the lenticular-integral intensity distributions *Microscope*, *Tank*, *Horseman* and *Toyplane*[1] and also for each of the conventional two-dimensional images *Lena*, *Airplane*, *Peppers* and *Sailboat* (these are reproduced in appendix B). Both horizontal and vertical pixel differences were calculated for each. Histograms of these differential images are plotted in Figure 5.4.

In general, adjacent pixel differences for natural two-dimensional images follow a Laplacian (double sided exponential) distribution [33] as evidenced by the histograms in Figures 5.4a and 5.4b. No specific significance is attached to either the horizontal or vertical pixel directions, since theoretically a conventional two-dimensional natural image has no inherent structure, though the scanning process employed in digitising video or stills often results in a slight imbalance between peak height for horizontal and vertical difference histograms.

The histograms shown for the case of the lenticular-integral intensity distributions, in Figures 5.4c and 5.4c, demonstrate that although a significant difference exists between the

[1] Only the central six pixels from each eight pixel band were included in these horizontal-dimension calculations, to avoid invalid results due to consideration of pixel pairs straddling black-band/image-band boundaries.

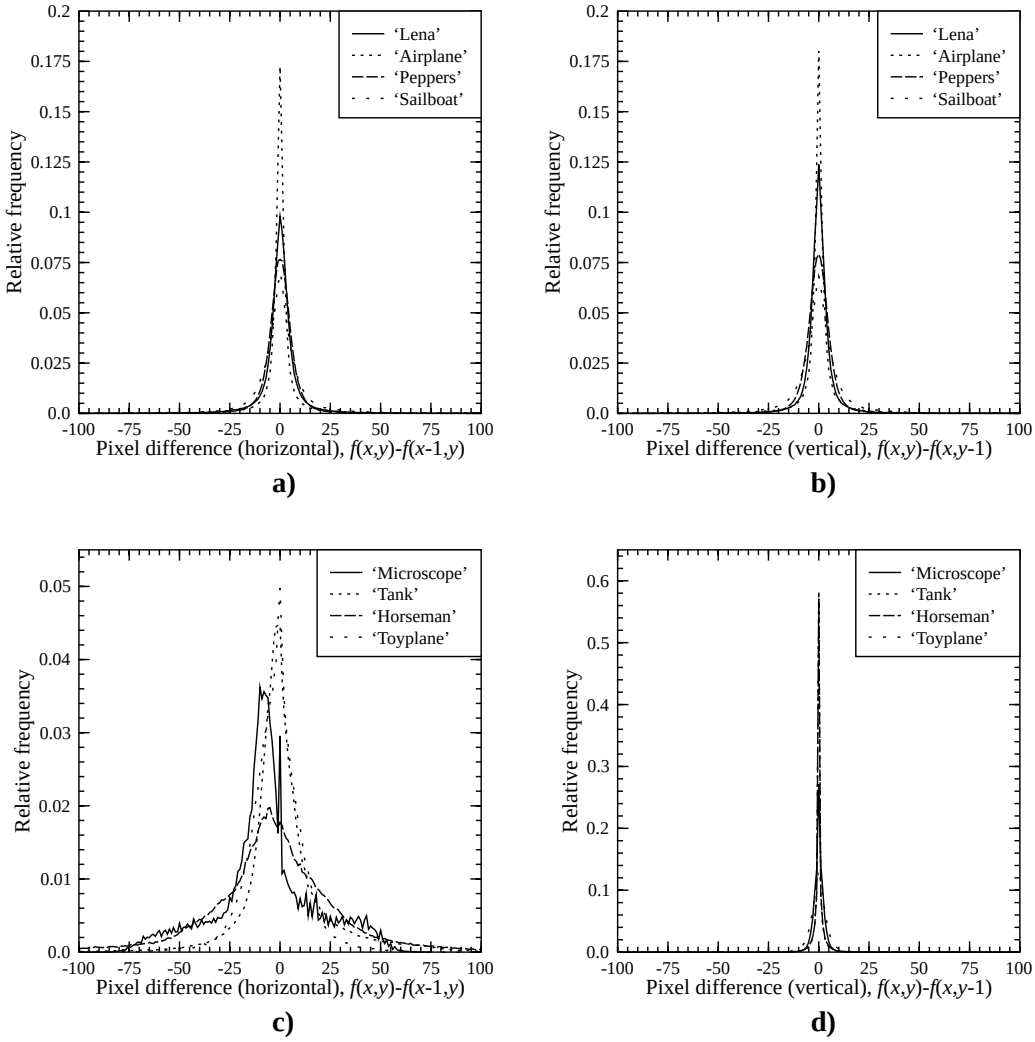


Figure 5.4: Pixel difference histograms. a) Conventional images, horizontal. b) Conventional images, vertical. c) Lenticular-integral, horizontal. d) Lenticular-integral, vertical.

nature of distributions for horizontal and vertical differences, both have similar envelopes. The distributions for horizontal differences in the integral intensity distributions each possess a low central peak in comparison with all other distributions shown. The apparent bimodality and rough nature of some of these histograms are not as significant as might be assumed at first, since all frequency values are so much smaller than are the case in all other plots in the figure.

Overall it is observed that in a lenticular-integral intensity distribution highly significant image information is distributed over a greater number of contiguous pixels horizontally than it is vertically or in either direction in two-dimensional natural images. This is due to the

fact that lenticular band boundaries, included in the difference calculation, run horizontally.

The conclusions drawn are twofold. Firstly, lenticular-integral intensity distributions possess inter-pixel correlation in common with conventional images. Secondly, care must be taken in any processing technique which introduces losses to ensure that any lossy treatment of data in horizontal and vertical directions can be controlled independently to avoid destroying important information in a replayed integral three-dimensional scene.

Summary

This chapter has described work carried out to discover the nature of the information plane recorded by the integral imaging system. The following observations and conclusions have been made:

1. The captured intensity distribution has an inherent banded structure;
2. A significant cross-correlation exists between bands in the intensity distribution, a feature of which a compression scheme should take account;
3. Inter-pixel correlation redundancy is present, though to a greater degree in vertical pixel directions than in horizontal directions. This disparity must be observed when designing a compression scheme for use with integral three dimensional image data.

Figure 5.5 illustrates the spatial correlation dimensions encountered in lenticular-integral image data.

In chapter 6 a coding scheme developed for use with integral three dimensional image data, based on a hybrid DPCM/DCT mapping stage, is described. Simulations demonstrate its superior rate-distortion performance compared with that of a conventional lossy image coding system.

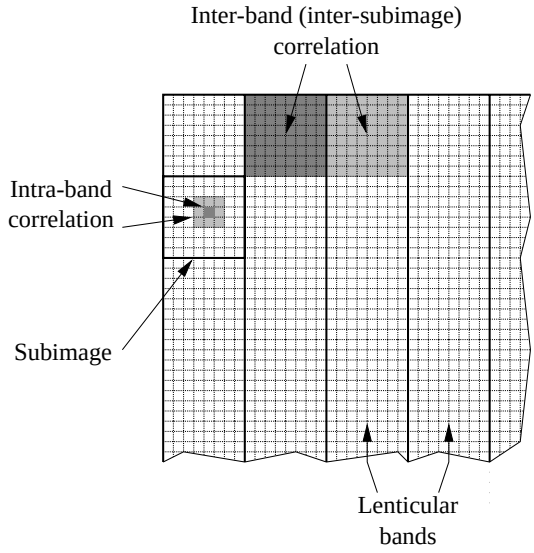


Figure 5.5: Spatial domain correlation in a lenticular-integral intensity distribution.

6. AN INTEGRAL 3D IMAGE COMPRESSION SCHEME USING A HYBRID DPCM/DCT ENCODER

Introduction

The combination of first-order linear prediction (DPCM) and the discrete cosine transform (DCT) has been researched extensively for use as the mapping stage in schemes for compression of conventional two dimensional video sequences. The difference between successive frames is computed to effect inter-frame (temporal domain) decorrelation, after which the DCT is applied to small blocks of the resulting frame difference data to reduce spatial domain correlation.

This chapter describes a method for compressing still integral three dimensional images by the application of a hybrid DPCM/DCT algorithm. Inter-band spatial decorrelation is performed by linear prediction of entire subimages of neighbouring lenticular bands. The DCT is computed for each resulting difference subimage to exploit intra-band spatial redundancies. Scalar quantisation of the resulting coefficients is followed by a Huffman-based entropy coding stage.

Simulations have been carried out for two variants of the scheme using two forms of quantisation parameter array with both luma channel only and full colour lenticular-integral intensity distributions as source data, the results of which are presented and discussed.

6.1 Hybrid DPCM/DCT Mapping for 2D Video Encoding

For the compression of certain types of data, multidimensional decorrelation, where the encoder's mapping stage is comprised of two or more 'phases', is desirable to take full advan-

tage of the redundancy in and hence size or rate reduction potential of the data. To identify decorrelation phases and therefore suitable mapping methods for effective quantisation and entropy coding, the structure and meaning of the data stream must be understood. The data must be studied extensively to ensure that elements of the data which appear to contain redundancies but are in fact dynamic variations of the data are not misinterpreted as suitable candidates for decorrelation.

An example of such a stream is conventional two dimensional video data. In this case two correlation dimensions exist: spatial redundancy between pixels in a frame as is present in any natural still image and temporal redundancy which occurs as a result of the typical temporal frequencies in sequences caused by movement and distortion of objects in the scene being significantly lower than the frame sampling frequency.

A combination of linear prediction and DCT decorrelation is commonly used for conventional video sequence compression, giving rise to a system known as *hybrid DPCM/DCT encoding*. This method forms the core of several international standards developed for various video compression applications, some examples of which are given in Table 6.1. These systems also possess motion compensation stages to provide further rate reduction by minimising the inter-frame differences to be coded; planar transformations of image sections between frames are coded as displacement vectors.

The basic hybrid DPCM/DCT video sequence encoder, without motion compensation, is shown in Figure 6.1. Each source video frame is split into 8×8-pixel blocks, each of which is encoded separately. Time domain decorrelation is now performed by subtracting the corresponding block from the previous frame (this has been stored for one frame period at '6'). The differential block is passed to a forward DCT stage ('1') which performs spatial domain decorrelation within the input block. The resulting coefficients are quantised ('2') and entropy coded ('3') using methods similar to those described for baseline sequential-DCT JPEG in chapter 3. The quantised coefficients must also be passed through the core of the

Standard	*Application*	*Target bit rates*
ITU-T H.261 [49, 50]	Video telephony/conferencing	64 kbit.s^{-1} – 2 Mbit.s^{-1}
ISO/IEC MPEG-1 [23, 51, 52]	Multimedia (CD-ROM storage)	$\leq$ 1.5 Mbit.s^{-1}
ISO/IEC MPEG-2 [23, 24]	Many — Primarily consumer TV	$\leq$ 15 Mbit.s^{-1} for DVB

Table 6.1: Hybrid DPCM/DCT-based international video compression standards.

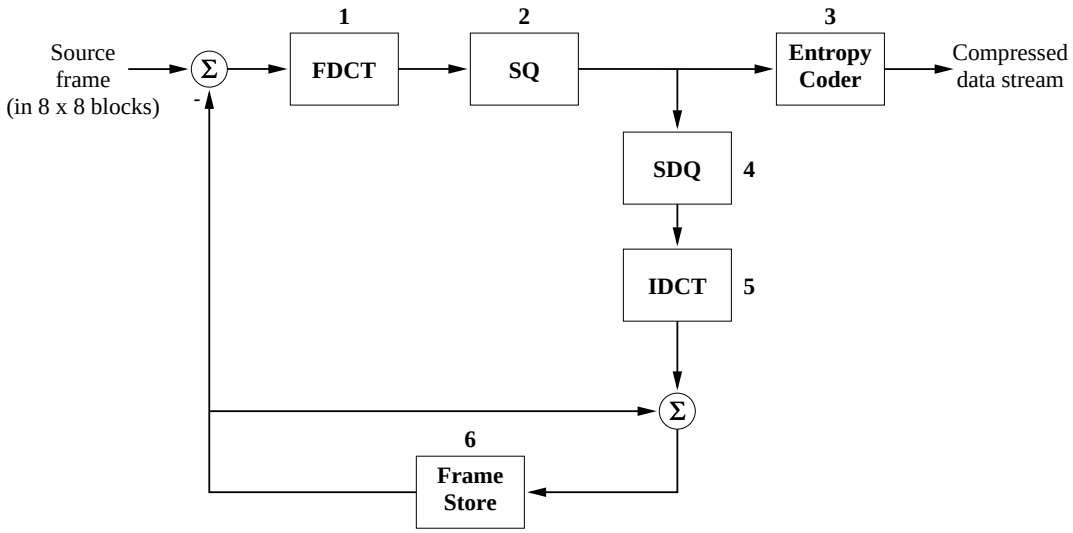


Figure 6.1: Basic hybrid DPCM/DCT two dimensional video encoder.

decoding process. The coefficients are dequantised (‘4’) and inverse transformed (‘5’) and the resulting frame difference blocks added to the previously coded image blocks, which have been stored at ‘6’ for one frame period. This process yields a reconstruction of the original (but quantised) image blocks as they would be at the decoder, ready for encoding of the next frame.

6.2 Mapper Design for Integral 3D Image Compression

It has been established that two spatial correlation dimensions exist in integral three dimensional image data (see chapter 5). These are *inter-band* correlation between corresponding pixels in neighbouring lenticular bands, and *intra-band* correlation between neighbouring pixels within each band. In order to develop efficient compression schemes for use with integral image data, the mapping stage of the encoder must be designed to take account of both inter-band and intra-band redundancies.

A combination of predictive and DCT techniques has been found to perform this function effectively. The operation of the hybrid DPCM/DCT mapper in integral image encoding is now described.

6.2.1 Inter-band decorrelation

In the hybrid DPCM/DCT integral image coding scheme, inter-band decorrelation is carried out using linear prediction (DPCM). Generally, a linear predictor outputs, for each pixel of the source sequence x, the difference between its value and a prediction of its value derived from a linear combination of previously reconstructed pixel values $\hat{x}$. For a predictor using k previous values (a kth order predictor), the mapper output for pixel n is the prediction error,

$$\begin{aligned} e(n) &= x(n) - x_p(n) \\ &= x(n) - \sum_{l=1}^{k} a_l \hat{x}(n-l), \end{aligned} \tag{6.1}$$

where $x_p(n)$ is the predicted value of pixel n and a_l are the predictor weights; $\sum_{l=1}^{k} a_l = 1$. For the case where only the previous reconstructed pixel value is used in the prediction, $k = 1$ and the mapper output,

$$e(n) = x(n) - \hat{x}(n-1). \tag{6.2}$$

Figure 6.2 illustrates this technique applied to inter-subimage decorrelation in an integral three dimensional image. An 8×8-pixel prediction error array $E_{i,j}$ is produced as the output of the inter-subimage spatial decorrelator. If $X_{i,j}$ is the current source subimage and $\hat{X}_{i-1,j}$ is the reconstructed previous subimage, the error for band i,

$$E_{i,j}(u,v) = X_{i,j}(u,v) - \hat{X}_{i-1,j}(u,v) \qquad \text{for } u,v = 0,1,\ldots,7. \tag{6.3}$$

At the first subimage on each line, $\hat{X}_{i-1,j}$ does not exist, ie. $E_{1,j}(u,v) = X_{1,j}(u,v)$.

6.2.2 Intra-band decorrelation

The discrete cosine transform (DCT) is used to reduce intra-subimage correlation, in a manner similar to its use in many block-based 2D image and video compression schemes. The most common form of the DCT for data compression applications, often known as *DCT-II* [28], is used. For a one dimensional sequence $f(l)$ of length L, the DCT is defined

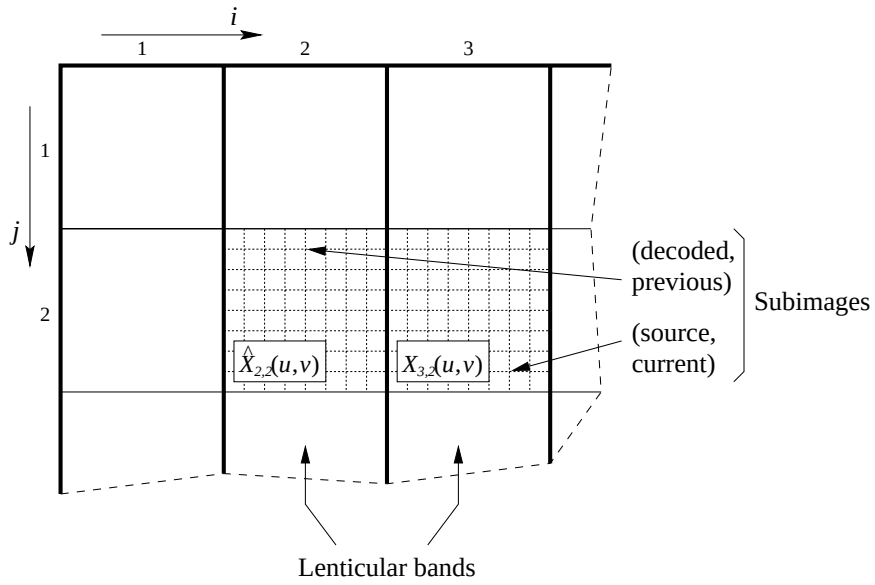


Figure 6.2: First order linear prediction for inter-subimage spatial decorrelation.

thus:

$$F(u) = \sqrt{\frac{2}{L}} K_u \sum_{l=0}^{L-1} f(l) \cos\left[\frac{u(2l+1)\pi}{2L}\right] \qquad \text{for } u = 0,1,\ldots,L-1. \tag{6.4}$$

Using the separability property of the DCT, the two dimensional DCT of an 8×8 array $f(l,m)$ is then:

$$F(u,v) = \frac{1}{4} K_u K_v \sum_{m=0}^{7} \sum_{l=0}^{7} f(l,m) \cos\left[\frac{u(2l+1)\pi}{16}\right] \cos\left[\frac{v(2m+1)\pi}{16}\right] \qquad \text{for } u,v = 0,1,\ldots,7. \tag{6.5}$$

In Equations 6.4 and 6.5 the scaling factors,

$$K_s = \begin{cases} 1/\sqrt{2} & \text{if } s = 0, \\ 1 & \text{otherwise.} \end{cases}$$

The DCT produces an array of highly uncorrelated coefficients with the same dimensions as the array presented to it. For intra-subimage decorrelation in an integral image, the data to be transformed are the difference subimages produced by the inter-subimage linear predictor described in the previous subsection and the transform coefficients then represent the source

subimage fully decorrelated, that is the final output from the mapping stage of the encoder. The overall mapper function for an intensity distribution comprising $I \times J$ subimages is:

$$Y_{i,j} = \begin{cases} \mathscr{C}\left\{X_{i,j}\right\} & \text{if } i = 1, \\ \mathscr{C}\left\{X_{i,j} - \hat{X}_{i-1,j}\right\} & \text{otherwise,} \end{cases} \qquad \text{for } i = 1,2,\ldots,I \text{ and } j = 1,2,\ldots,J. \tag{6.6}$$

where $\mathscr{C}\{\}$ denotes the (8×8) two dimensional DCT of its operand, defined by Equation 6.5.

6.3 Quantisation of the Mapped Subimage Data

Each decorrelated subimage $Y_{i,j}$ is now input to a scalar quantisation stage in which each of the 64 decorrelated coefficients is processed by a different uniform quantiser, allowing individual coefficient positions to be quantised using a different step size. This is achieved by dividing coefficients at each position in the decorrelated array by step sizes at corresponding positions in a quantisation array. If Q is the (8×8) array of quantiser step sizes, the quantised coefficient array at subimage position (i, j),

$$S_{i,j}(u,v) = \text{Nearest integer}\left(\frac{Y_{i,j}(u,v)}{Q(u,v)}\right) \quad \text{for } u,v = 0,1,\ldots,7. \tag{6.7}$$

The 64 values comprising Q are parameters of the encoder and decoder and are chosen by considering the significance of each DCT coefficient to maintaining high fidelity replay of the decoded intensity distribution data when replayed as a three dimensional integral image.

6.3.1 Quantisation table selection

The quantisation method described has previously been applied to compression of conventional two dimensional images intended for human viewing. A technique employed to choose quantisation parameters in this case is to use information on human visual system sensitivity to remove or de-emphasise perceptually insignificant frequency components. Such an approach is not appropriate for application to integral three dimensional image compression. This is the case because the data being quantised, reconstructed and presented are not viewed directly, rather the decoded output of the lenticular decoding sheet is. Due to the complex re-integrating function of the lenticular decoder, the DCT coefficient array presented to the

quantisation process is not compatible with perceptually derived quantisation table design methods.

From this it follows that the standard quantisation parameter arrays developed for use with two dimensional image and video compression standards as being perceptually optimal are not when applied to coefficients derived from integral intensity distributions. For this reason psychovisually weighted quantisation tables have not been utilised during development of the integral image compression systems described in this thesis. Initially the most fundamental form of quantisation table possible, in which each coefficient is quantised using a common step size, is defined. This is the *fixed quantisation table.*

Irrespective of the specific set of quantisation parameters in use it must be possible to vary all 64 step sizes in the quantisation table in unison in order to control the overall coarseness of quantisation. This is necessary to obtain the required balance between the data reduction obtained and distortion in the reconstructed image for a given application. This is accomplished by varying a parameter q, the *quantiser scale,* increasing values of which give rise to increasing data reduction and decreasing fidelity of the decoded image. For fixed quantisation the quantisation table,

$$Q(u,v) = q \quad \text{for } u,v = 0,1,\ldots,7 \text{ and } q \in \mathbb{N}^{+}. \tag{6.8}$$

Variability of decorrelated subimage data

Due to the nature of the hybrid DPCM/DCT decorrelation method used, and the structure of the lenticular-integral intensity distribution, the DCT coefficient block resulting from decorrelation of data in a subimage is expected to possess certain unique features. In particular it is reasonable to suggest that an imbalance exists between the significance of spatial frequency components occurring vertically and horizontally within each subimage due to the banded structure of the captured intensity distribution. To corroborate this the variability of decorrelated coefficients has been examined by computing the standard deviation from the mean of each for each subimage from several example intensity distributions.

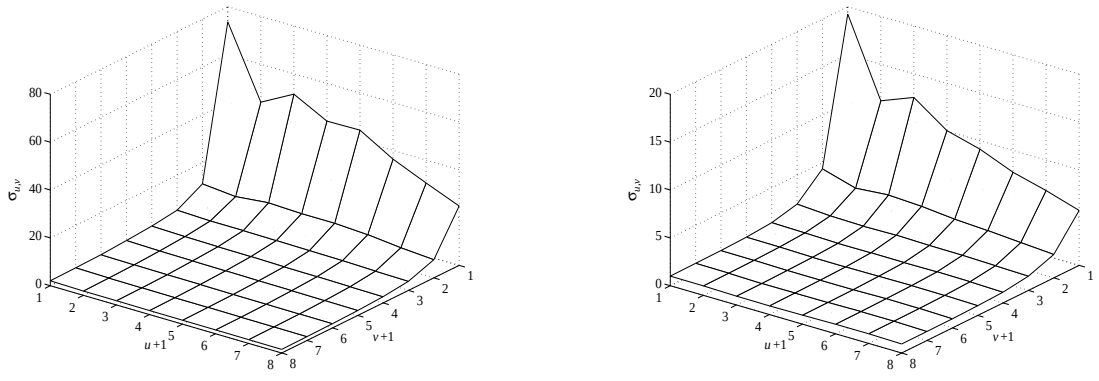
In this experiment the decorrelated data were obtained directly from the hybrid DPCM/DCT decorrelation stage. A quantisation stage was not included since this would distort the resulting statistics. For the decorrelated coefficient position (u,v), the standard deviation over

all $I \times J$ subimages from K intensity distributions [53],

$$\sigma_{u,v} = \sqrt{\frac{1}{IJK-1}\sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{i=1}^{I}\left[Y_{i,j,k}(u,v) - \overline{Y}(u,v)\right]^2} \quad \text{for } u,v = 0,1,\ldots,7, \tag{6.9}$$

where $Y_{i,j,k} = Y_{i,j}$ taken from intensity distribution k and $\overline{Y}(u,v)$ is the sample mean for coefficient position (u,v) evaluated over all $I \times J$ decorrelated subimages from all K intensity distributions.

Equation 6.9 was evaluated for the coefficient blocks resulting from decorrelation of the luma channels of the four intensity distributions *Microscope*, *Tank*, *Horseman* and *Toyplane* and the chrominance-blue/chrominance-red channels of *Tank*, *Horseman* and *Toyplane*. Figures 6.3a and 6.3b are the resulting plots of $\sigma_{u,v}$. It is observed that the variability of coefficients representing moderate to high horizontal spatial frequencies is substantially greater than that of those representing corresponding frequencies in the vertical dimension in both the luma and combined chrominance data. The form of the plots for luma and combined chrominance are very similar, though in the chrominance case standard deviation values are between 1.5 and 5 times lower than for the luminance case. The conclusion is drawn that within each subimage vertically-running features are less significant than horizontally-running features, and can therefore be quantised more coarsely leading to improved data reduction. The quantisation array produced to effect this has monotonically increasing step



a) **b)**

Figure 6.3: Standard deviation from the mean of hybrid DPCM/DCT decorrelated subimage coefficients. a) Luma channel. b) Combined chrominance channels.

sizes vertically, while constant horizontally:

$$Q(u,v) = q(1+v) \quad \text{for } u,v = 0,1,\ldots,7 \text{ and } q \in \mathbb{N}^+. \tag{6.10}$$

This will be called the *linear variable vertical* (*LVV*) quantisation table.

6.4 Entropy Coding of the Quantised Subimage Coefficients

An entropy coder is used to reduce the high statistical redundancy present in the quantised coefficient blocks. The original subimage data will be represented by a sequence of binary codewords as the final output of the encoder. The entropy coder is a Huffman-based coder as used by the JPEG system for AC coefficient coding. In the base hybrid DPCM/DCT integral image compression scheme the DC coefficient in a block is not coded differentially and separately from the AC coefficients as it is in the JPEG scheme. All coefficients are treated in the same way.

The quantised coefficients are first ordered into a sequence using the zigzag scan described in chapter 3 for the JPEG encoder, but in this case the scan includes the DC coefficient as its first element, resulting in a sequence of length 64. This ordering results in a sequence of coefficients with increasing spatial frequencies. Due to the nature of the quantised coefficient block the higher order coefficients have no spatial significance and as the zigzag scan index increases the likelihood of zero-valued coefficients increases.

The mapping of coefficients from the zigzag scan to the set of symbols which are input to the Huffman coder (the coder alphabet), each of which is one byte long, now takes place. A symbol representing a nonzero value from the zigzag scan is produced by combining the number of zero valued coefficients up to run length of 15, R, with the *magnitude category* into which the coefficient falls. The magnitude category for a nonzero coefficient a,

$$S = \left\lfloor 1 + \log_2(|a|) \right\rfloor. \tag{6.11}$$

Due to the nature of scaling in the DCT operation, the maximum coefficient value is ± 2047, leading to a maximum value for S of 11. The final symbol is constructed with the length of the run of zeros preceding the nonzero coefficient under consideration as the upper nibble and the magnitude category as the lower nibble of the symbol, that is $16R+S$.

A run of zero valued coefficients of length greater than 15 which is followed by a nonzero coefficient is encoded using one or more special 'ZRL' (zero run length) symbols. A ZRL symbol represents a run of 16 zeros and any further zero run of less than 16 up to the next nonzero coefficient is encoded as part of the symbol for the nonzero coefficient. A further special symbol, 'ESB' (end of subimage block) is placed after the final nonzero coefficient to state that any further coefficients in the block all have values of zero.

Following construction of the alphabet for the Huffman coder, the total individual symbol frequencies are calculated for the zigzag-scanned sequence resulting from the decorrelation and quantisation of every subimage in the source intensity distribution. The frequency table produced is used to generate the table of codewords using the Huffman algorithm [26], giving a minimum length code for the quantised coefficients representing that particular integral intensity distribution.

The symbols representing nonzero coefficients do not contain information on the magnitudes of those coefficients. Hence to fully specify such a coefficient and its associated preceding run of zero valued coefficients, each codeword produced at the final output of the encoder is comprised of the Huffman coded symbol followed by S bits to specify the sign and magnitude of the coefficient.

6.5 The Base Hybrid DPCM/DCT Encoder and Decoder

The mapping, quantisation and coding stages in the hybrid DPCM/DCT integral image compression scheme have been described. The complete encoding and complementary decoding processes are shown in Figures 6.4a and 6.4b respectively.

6.6 Enhancement to Reduction of Vertical Intra-band Correlation

A variation of the base encoder described above has been developed to provide enhanced compression performance. The non-parallax direction in an integral intensity distribution recorded using a lenticular sheet is captured as a continuous stripe, hence the division of subimages in that dimension is arbitrary; vertical division points are not dictated by the structure of the intensity distribution. Intra-band data which is split vertically across two subimages are therefore not decorrelated. In order to take account of this redundancy a

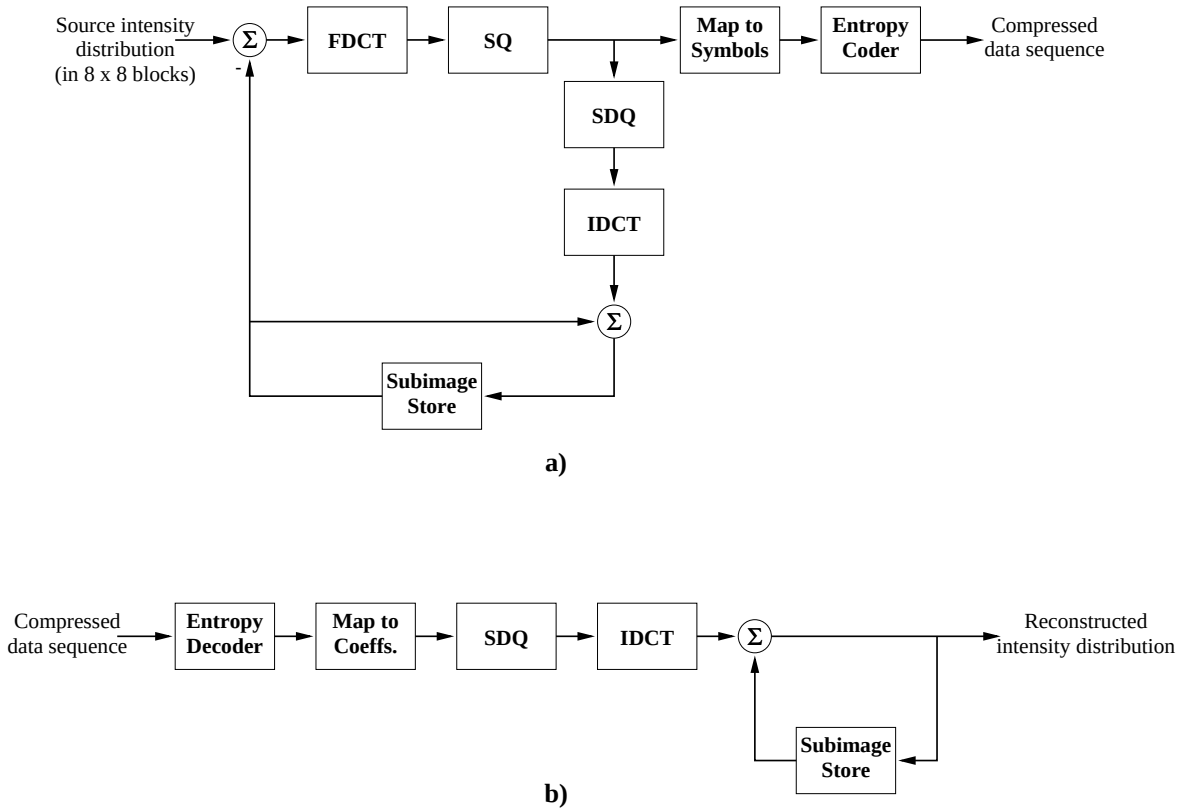


Figure 6.4: The Base hybrid DPCM/DCT still integral 3D image compression scheme. a) Encoder. b) Decoder.

DPCM loop is added to the entropy coder pre-processing stage such that the highly significant quantised DC coefficients are differentially encoded vertically in a manner similar to the use of DPCM between block DC coefficients horizontally in the JPEG system. A separate Huffman stage is used to code these vertical DC differences. The method of construction of symbols from DC differences for this Huffman coder is similar to that used for AC coefficients (explained in section 6.4). Run lengths of differences of zero are not encoded, so for a vertical DC coefficient difference Δd the symbol,

$$S_{\mathrm{DC}} = \begin{cases} 0 & \text{if } \Delta d = 0, \\ \lfloor 1 + \log_2(|\Delta d|) \rfloor & \text{otherwise.} \end{cases} \tag{6.12}$$

Due to the scaling in the DCT, the maximum difference between two DC coefficients from neighbouring subimages is 2047. Hence the maximum possible value for S_{DC} is 11. The final output codeword is constructed by appending S_{DC} bits to specify the sign and magnitude within the magnitude category, to the Huffman codeword corresponding to symbol S_{DC}.

Since this *enhanced hybrid DPCM/DCT* scheme treats the DC coefficient differently from the AC coefficients, the zigzag coefficient ordering process is modified to scan only the 63 AC coefficients. The encoder modification, placed in parallel with the AC coefficient-to-symbol mapping and entropy coding stages in Figure 6.4, is shown in Figure 6.5.

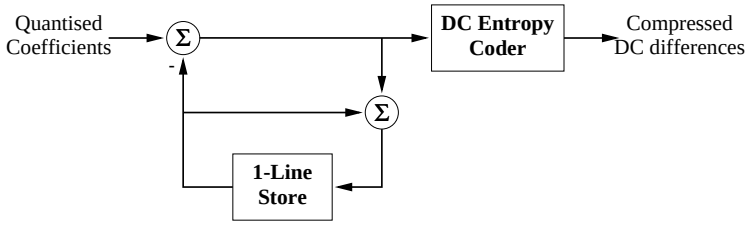


Figure 6.5: Modification to the base hybrid DPCM/DCT encoder for improved vertical intra-band correlation reduction.

6.7 Implementation of the Hybrid Schemes for Simulation

For the purpose of evaluation of the base and enhanced hybrid DPCM/DCT schemes the encoder and certain portions of the decoder have been implemented in software, instrumented to allow the following data relevant to the performance of the algorithms to be obtained:

1. The decoded intensity distribution;
2. The data reduction and hence output bit rate obtained;
3. An objective measure of the fidelity of the decoded intensity distribution.

The pseudo-code listed as Procedure 6.1 represents the complete simulation system for evaluation of the base and enhanced hybrid DPCM/DCT schemes. The system has been implemented using the C language under the UNIX operating system.

The procedure ‘Encoder_main_loop’ carries out the mapping and quantisation functions, producing an array of quantised coefficients for all decorrelated subimages and the fully reconstructed intensity distribution as would be output by the decoder. It is listed as Procedure 6.2 for both the base and enhanced schemes. The forward and inverse 2D-DCT in this procedure are carried out using separate row/column application of a forward or inverse 1D-DCT

Procedure 6.1 Top level base/enhanced hybrid DPCM/DCT simulation system.

Arguments:
INIMG {*source intensity distribution*}
QMTX {*quantisation array*}
[HUFF [, DCHUFF]] {*Huffman code lengths (if using predefined code tables)*}
Results:
OUTIMG {*reconstructed decoded intensity distribution*}
OBPP {*output bit rate*}
QUAL {*objective fidelity of reconstructed intensity distribution*}

(QUANT_COEFFS, OUTIMG) ⇐ Encoder_main_loop(INIMG, QMTX)
if not using predefined code tables **then**
  FREQ ⇐ Gen_sym_table(QUANT_COEFFS) {*Generate table of coder symbol frequencies*}
  HUFF ⇐ Gen_code_lengths(FREQ) {*Compute lengths of Huffman codes for each symbol*}
  if using enhanced scheme **then**
    DCFREQ ⇐ Gen_DC_sym_table(QUANT_COEFFS) {*Generate table for DC differences*}
    DCHUFF ⇐ Gen_code_lengths(DCFREQ) {*Compute DC Huffman code lengths*}
  end if
end if
OBITS ⇐ Calc_bit_count(QUANT_COEFFS, HUFF) {*Calculate total bit count for coded data*}
if using enhanced scheme **then**
  OBITS ⇐ OBITS + Calc_DC_bit_count(QUANT_COEFFS, DCHUFF) {*Add DC bit count*}
end if
OBPP ⇐ OBITS / Number_of_pixels(INIMG) {*Compute output bit rate (bits per pixel)*}
QUAL ⇐ Compute_PSNR(OUTIMG, IMG) {*Compute objective fidelity for decoded data*}

Procedure 6.2 Main loop for hybrid DPCM/DCT simulation.

Arguments:
INIMG {*source intensity distribution*}
QMTX {*quantisation array*}
Results:
QUANT_COEFFS {*quantised transform coefficients*}
OUTIMG {*reconstructed decoded intensity distribution*}

I ⇐ number of subimage columns in INIMG
J ⇐ number of subimage rows in INIMG
for $j = 1$ to J **do**
  PREV ⇐ clear to all zero
  for $i = 1$ to I **do**
    DIFF ⇐ $\text{INIMG}_{i,j}$ − PREV {*Inter-subimage decorrelation*}
    DCT ⇐ Forward_2D_DCT(DIFF) {*Intra-subimage decorrelation*}
    $\text{QUANT_COEFFS}_{i,j}$ ⇐ Nearest_integer(DCT / QMTX) {*Quantisation*}
    DQUANT ⇐ $\text{QUANT_COEFFS}_{i,j}$ × QMTX {*Dequantisation*}
    IDCT ⇐ Inverse_2D_DCT(DQUANT)
    $\text{OUTIMG}_{i,j}$ ⇐ IDCT + PREV {*Subimage reconstruction*}
    PREV ⇐ $\text{OUTIMG}_{i,j}$ {*Store reconstructed subimage for next pass*}
  end for
end for

respectively. These processes were implemented using routines from the 'FFTPACK' library [54].

The simulation system accommodates two different sources for the Huffman codes required by the entropy coding stages. Since the particular bit sequences output are not relevant to the analysis, only the *lengths* in bits of the codes representing each possible symbol are required. The necessary table of code lengths may either be predefined and passed to the simulation system or it may be generated for a particular encoding instance from the statistics of the transformed and quantised coefficient data. When the hybrid DPCM/DCT scheme is operating in enhanced mode, a second code length table is required to represent vertical DC coefficient differences.

If predefined Huffman code tables are to be used, they are presented to the top level procedure as HUFF and additionally DCHUFF if the enhanced scheme mode is in operation. If, instead, local Huffman tables are to be used, Procedure 6.3 ('Gen_sym_table') is invoked to generate a table of symbol frequencies for the quantised coefficients produced by the main encoder loop. If the enhanced scheme is being used the zigzag scan does not include the DC coefficient

Procedure 6.3 Symbol frequency table generation.

Arguments:
  QUANT_COEFFS {*quantised transform coefficients*}
Results:
  FREQ {*table of coder symbol frequencies*}

```
I ⇐ number of subimage columns in QUANT_COEFFS
J ⇐ number of subimage rows in QUANT_COEFFS
FREQ ⇐ clear to all zero
if using enhanced scheme then
  ZZSTART ⇐ 2                                   {Don't scan DC coefficient}
else                                            {using base scheme}
  ZZSTART ⇐ 1                                   {Do scan DC coefficient}
end if
for j = 1 to J do
  for i = 1 to I do
    ZZ ⇐ Zigzag_scan(QUANT_COEFFS_{i,j}, ZZSTART)   {Scan starts at coeff. ZZSTART}
    SYM ⇐ Map_to_coder_symbols(ZZ)
    for all symbols S in SYM do
      FREQ(S) ⇐ FREQ(S) + 1
    end for
  end for
end for
```

of each subimage and frequencies are only produced for those symbols mapped from AC coefficients. The Huffman code length table is produced from this frequency table by the procedure 'Gen_code_lengths'. This routine is described in Annex K of the JPEG Draft International Standard [1].

'Gen_DC_sym_table' (Procedure 6.4) is executed only during simulation of the enhanced hybrid scheme described in section 6.6, and only if predefined Huffman code tables are not being used. It produces a table of frequencies of the symbols representing differences between DC coefficients from neighbouring decorrelated subimages running vertically. The code lengths for the DC Huffman coder are derived from the DC difference symbol frequency table generated by 'Gen_DC_sym_table' using the same procedure employed to compute code lengths for the main Huffman coder.

Next the total bit count for the coded intensity distribution is calculated. This is carried out using the Huffman code tables which were either predefined or generated from the quantised coefficients as described above, by summing code lengths for all generated symbols for every quantised subimage block. For all coefficients where the base scheme is being simulated, and for the AC coefficients in the case of the enhanced scheme, Procedure 6.5 ('Calc_bit_count') is executed. For the enhanced scheme the code lengths for symbols representing vertical DC coefficient differences are then added by 'Calc_DC_bit_count', listed as Procedure 6.6.

Procedure 6.4 DC difference symbol frequency table generation for enhanced mode.

Arguments:
QUANT_COEFFS {*quantised transform coefficients*}
Results:
DCFREQ {*table of DC vertical difference coder symbol frequencies*}

$I \Leftarrow$ number of subimage columns in QUANT_COEFFS
$J \Leftarrow$ number of subimage rows in QUANT_COEFFS
DCFREQ $\Leftarrow$ clear to all zero
for $i = 1$ to I **do**
  PREVDC $\Leftarrow 0$
  for $j = 1$ to J **do**
    DCDIFF $\Leftarrow$ QUANT_COEFFS$_{i,j}$ $-$ PREVDC
    $S_{DC} \Leftarrow$ Map_to_DC_coder_symbol(DCDIFF) {*Described by equation 6.12*}
    DCFREQ(S_{DC}) $\Leftarrow$ DCFREQ(S_{DC}) $+ 1$
    PREVDC $\Leftarrow$ QUANT_COEFFS$_{i,j}$
  end for
end for

Procedure 6.5 Calculate bit count for main Huffman coder output.

Arguments:
QUANT_COEFFS {*quantised transform coefficients*}
HUFF {*Huffman code size per symbol*}
Results:
BITS {*coded bit count for output of main Huffman coder*}

$I \Leftarrow$ number of subimage columns in QUANT_COEFFS
$J \Leftarrow$ number of subimage rows in QUANT_COEFFS
BITS $\Leftarrow 0$
if using enhanced scheme **then**
ZZSTART $\Leftarrow 2$ {*Don't scan DC coefficient*}
else {*using base scheme*}
ZZSTART $\Leftarrow 1$ {*Do scan DC coefficient*}
end if
for $j = 1$ to J **do**
for $i = 1$ to I **do**
ZZ $\Leftarrow$ Zigzag_scan(QUANT_COEFFS$_{i,j}$, ZZSTART) {*Scan starts at coeff.* ZZSTART}
SYM $\Leftarrow$ Map_to_coder_symbols(ZZ)
for all symbols S in SYM **do**
BITS $\Leftarrow$ BITS + (HUFF(S) + Magnitude_category(COEFF_GENERATING_S))
end for
end for
end for

Procedure 6.6 Calculate bit count for DC Huffman coder output in enhanced mode.

Arguments:
QUANT_COEFFS {*quantised transform coefficients*}
DCHUFF {*DC difference Huffman code size per symbol*}
Results:
DCBITS {*coded bit count for output of DC Huffman coder*}

$I \Leftarrow$ number of subimage columns in QUANT_COEFFS
$J \Leftarrow$ number of subimage rows in QUANT_COEFFS
DCBITS $\Leftarrow 0$
for $i = 1$ to I **do**
PREVDC $\Leftarrow 0$
for $j = 1$ to J **do**
DCDIFF $\Leftarrow$ QUANT_COEFFS$_{i,j}$ − PREVDC
$S_{\mathrm{DC}} \Leftarrow$ Map_to_DC_coder_symbol(DCDIFF)
DCBITS $\Leftarrow$ DCBITS + (DCHUFF(S_{DC}) + Magnitude_category(DCDIFF))
PREVDC $\Leftarrow$ QUANT_COEFFS$_{i,j}$
end for
end for

The output bit rate for the fully coded intensity distribution is then computed, expressed as the number of coded bits per pixel (bpp) of the source intensity distribution. Finally the peak signal to noise ratio (defined by Equations 4.9 and 4.11 for greyscale and colour images respectively) is calculated for the reconstructed decoded intensity distribution to provide a measure of the distortion introduced as a result of the quantisation process.

6.8 Global Huffman Code Generation

As described in the previous section, if it is not desired to generate Huffman code length tables directly from the statistics of the quantised coefficient data, predefined tables may be used. The per-image code generation technique generally leads to significantly improved performance due to the generated code being statistically matched to the transformed and quantised representation of the image. However in a practical implementation this requires the transmission of the Huffman code table for every image leading to a higher encoder overhead. The coding gain achieved using the per-image method and the inherent overhead must be compared for a given application of the hybrid DPCM/DCT coding scheme. As a worst-case scenario, the results of simulations given later in this chapter have been obtained from use of predefined Huffman code tables, though results where the per-image local code table approach has been used are given for comparison. This section describes the preparation of the *global* (predefined) Huffman code tables.

In constructing such general purpose code tables, it is necessary to make use of several source intensity distributions which should be as far as possible representative of the entire set of intensity distributions which the scheme may be expected to encode. This is the case since the statistical distribution of symbols from a given necessarily finite set of intensity distributions is used to infer symbol probabilities which will construct a code for use with all further source intensity distributions. For this reason, to provide a reasonable guarantee of satisfactory future performance of the coding scheme, the number and nature of sample intensity distributions used in generating the code must be appropriately chosen. The process employed to produce the Huffman code tables is now described.

Initially, the Huffman coder input statistics are collected in terms of the frequency of each symbol, for several source intensity distributions. This is carried out for the base scheme mode using the routine ‘Gen_sym_table’ (Procedure 6.3) and for the enhanced scheme mode using ‘Gen_sym_table’ and ‘Gen_DC_sym_table’ (Procedure 6.4). The corresponding symbol frequencies produced from the source intensity distributions are then summed and presented

to the routine 'Gen_code_lengths' to construct tables of Huffman word lengths.

This process was carried out using the luma channels of the four intensity distributions *Horseman*, *Microscope*, *Tank* and *Toyplane* and separately for the six chrominance channels from *Horseman*, *Tank* and *Toyplane*. The quantisation array used in this process was fixed; the quantiser scaling parameter q was set such that for the luma channel of each source intensity distribution, the output bit rate produced using local Huffman codes was approximately equal to a specific value (the *target* bit rate). Initially the target bit rate was set to 0.5 bits per pixel. This value was selected as it is a moderate bit rate around which the schemes are expected to be of most practical utility. For the data resulting from the chrominance components of the source intensity distributions, q was set to the same value as was used to generate the luma coefficient frequency tables. To ensure that a Huffman code was generated in all cases, even for a symbol which did not occur for the particular intensity distribution samples used, unity was added to each total symbol frequency value. By way of example, Table 6.2 gives the generated code lengths for the luma channel data in the base hybrid DPCM/DCT scheme. The symbol indexes are listed in hexadecimal. The special symbols 'ESB' and 'ZRL' correspond to symbol indexes 00 and F0 respectively.

Table 6.2: Global Huffman code length table for luma channel data in the base hybrid DPCM/DCT scheme.

$16R+S$	*Length*	$16R+S$	*Length*	$16R+S$	*Length*	$16R+S$	*Length*
00	3	40	19	80	19	C0	19
01	2	41	5	81	6	C1	5
02	3	42	8	82	8	C2	9
03	5	43	11	83	11	C3	13
04	8	44	13	84	15	C4	18
05	10	45	19	85	19	C5	19
06	12	46	19	86	19	C6	19
07	18	47	19	87	19	C7	19
08	18	48	19	88	19	C8	19
09	18	49	19	89	19	C9	19
0A	18	4A	19	8A	19	CA	19
0B	18	4B	19	8B	19	CB	19
0C	19	4C	19	8C	19	CC	19
0D	19	4D	19	8D	19	CD	19
0E	19	4E	19	8E	19	CE	19
0F	19	4F	19	8F	19	CF	19
10	19	50	19	90	19	D0	19
11	4	51	5	91	6	D1	6
12	6	52	9	92	10	D2	12
13	9	53	13	93	14	D3	15
14	13	54	16	94	19	D4	19
15	15	55	19	95	19	D5	19

Table 6.2: (continued)

$16R+S$	*Length*	$16R+S$	*Length*	$16R+S$	*Length*	$16R+S$	*Length*
16	19	56	19	96	19	D6	19
17	19	57	19	97	19	D7	19
18	19	58	19	98	19	D8	19
19	19	59	19	99	19	D9	19
1A	19	5A	19	9A	19	DA	19
1B	19	5B	19	9B	19	DB	19
1C	19	5C	19	9C	19	DC	19
1D	19	5D	19	9D	19	DD	19
1E	19	5E	19	9E	19	DE	19
1F	19	5F	19	9F	19	DF	19
20	19	60	19	A0	19	E0	19
21	5	61	6	A1	6	E1	7
22	8	62	8	A2	9	E2	13
23	9	63	10	A3	12	E3	18
24	12	64	14	A4	17	E4	19
25	15	65	19	A5	19	E5	19
26	19	66	19	A6	19	E6	19
27	19	67	19	A7	19	E7	19
28	19	68	19	A8	19	E8	19
29	19	69	19	A9	19	E9	19
2A	19	6A	19	AA	19	EA	19
2B	19	6B	19	AB	19	EB	19
2C	19	6C	19	AC	19	EC	19
2D	19	6D	19	AD	19	ED	19
2E	19	6E	19	AE	19	EE	19
2F	19	6F	19	AF	19	EF	19
30	19	70	19	B0	19	F0	6
31	5	71	5	B1	5	F1	8
32	7	72	7	B2	7	F2	17
33	8	73	9	B3	10	F3	19
34	11	74	13	B4	16	F4	19
35	14	75	19	B5	19	F5	19
36	19	76	19	B6	19	F6	19
37	19	77	19	B7	19	F7	19
38	19	78	19	B8	19	F8	19
39	19	79	19	B9	19	F9	19
3A	19	7A	19	BA	19	FA	19
3B	19	7B	19	BB	19	FB	19
3C	19	7C	19	BC	19	FC	19
3D	19	7D	19	BD	19	FD	19
3E	19	7E	19	BE	19	FE	19
3F	19	7F	19	BF	19	FF	19

The tables generated by the procedure described have been used to produce the simulation results presented in the next section for the cases where predefined code words were used. It should be noted that where only four sample source data sets have been used, as in this case, the codes produced can not be expected to apply efficiently to future possible intensity distributions as such a small sample is highly unlikely to be representative. When a greater number of sample source data sets is available, the same method can, however, be applied to produce more statistically accurate code tables.

6.9 Simulation Results

This section presents the results of simulations performed of the base and enhanced hybrid DPCM/DCT schemes. To provide a reference for comparison of these schemes the JPEG system was used to generate results in parallel with the schemes under test. The JPEG simulations made use of the default JPEG luma and chrominance quantisation arrays and the default JPEG Huffman code tables. The bit rate statistics generated by the simulation models represent only the raw coded intensity distribution data, that is no control or synchronisation data are included in the output. To ensure a fair comparison with the JPEG scheme its baseline sequential-DCT coding mode has been re-implemented for this purpose to remove the control codes which would normally be present in its fully encoded output.

The results are presented in the form of plots of peak signal to noise ratio (PSNR) as an objective quality metric against output bit rate. In each plot shown a given scheme has been simulated using the intensity distributions in appendix A as source data and mean curves produced. Values of output bit rate (the independent variable) do not correspond for each image, therefore in order to calculate mean values of PSNR, the data for each intensity distribution were first remapped to a common independent variable. This was carried out using linear interpolation with an x-axis increment of 0.01 bpp.

6.9.1 Coding of luma channel only

Initially simulation results are presented where the source data are the luma (Y') channel only of the source intensity distributions. In this case the source data represent only the white intensity of the pixels comprising each input distribution, the source bit rate being 8 bits per pixel.

Absolute comparison with baseline JPEG

The plot in Figure 6.6 shows performance curves for the base hybrid DPCM/DCT scheme and baseline sequential-DCT JPEG. The results for the base hybrid scheme relate to use of a fixed quantisation array as per Equation 6.8, and the predefined luma Huffman code table derived in section 6.8. It is observed that the base hybrid DPCM/DCT scheme gives a substantial increase in rate-distortion performance over the baseline JPEG system for the integral intensity distribution luma data used. An improvement is expected since baseline JPEG cannot take account of the inter-band correlation inherent in the structure of integral image data.

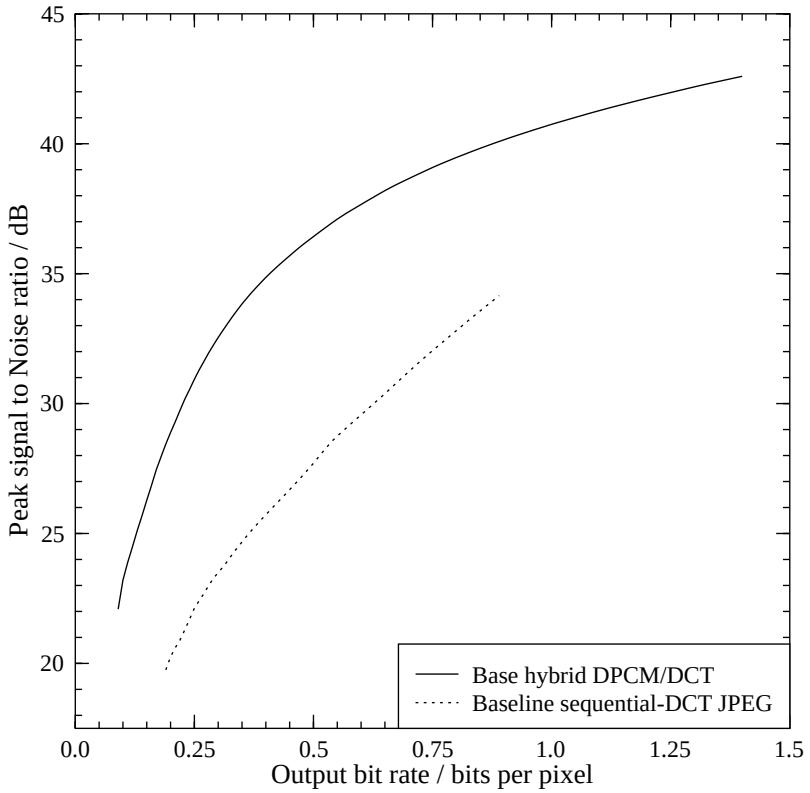


Figure 6.6: Performance of the base hybrid DPCM/DCT scheme using fixed quantisation and a predefined code table on the luma channel only of the source data; baseline sequential-DCT JPEG performance is shown for reference.

Use of local Huffman code tables

Figure 6.7 gives a plot of the difference in reconstructed fidelity between use of two predefined luma code tables, constructed for two bit rates, and code tables generated locally as part of the encoding process to apply to the specific intensity distribution being encoded. The overhead due to the necessary transmission of the code table has not been included in the output bit rate calculation for the local case, since the method by which the table is encoded is seen as an implementation detail and directly influences the cost of transmitting it. It is, however, a constant overhead unaffected by quantisation parameters, and is expected to be negligible compared with the cost of transmitting the coded intensity distribution at a bit rate capable of giving high quality image reproduction. It is observed that across the range of output bit rates for which the scheme was simulated, where a local Huffman code table is used the scheme outperforms the global code table case. This is expected since the global

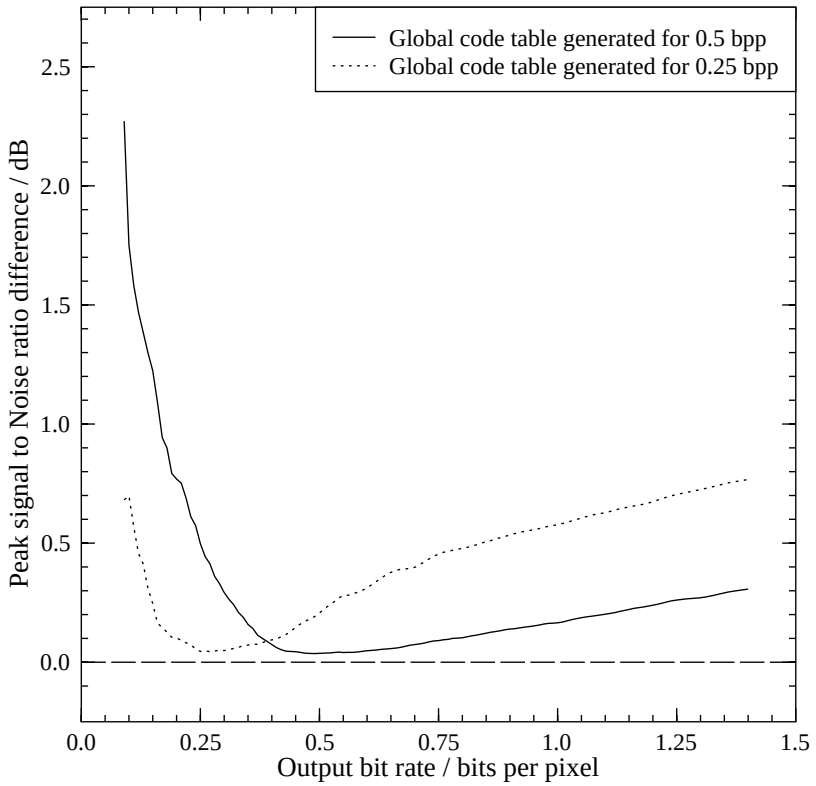


Figure 6.7: Improvement in performance of the base hybrid DPCM/DCT scheme using fixed quantisation due to use of a local rather than predefined code table on the luma channel only of the source data; results for predefined code tables generated for bit rates of 0.25 bpp and 0.5 bpp are shown.

code table is not statistically matched as accurately to a given source intensity distribution as is a local code table. The difference in objective fidelity can be seen to reach its minimum value at approximately the bit rate for which the global code table was constructed. The differences at this point are barely significant, being 0.036 and 0.045 dB for the 0.5 and 0.25 bpp target values respectively. However, if the scheme is used to produce output bit rates far from that for which the global code table was constructed, a further decrease in performance results whose value and gradient are particularly great when it is scaled to produce a lower output bit rate. All further results presented in this section result from the use of predefined Huffman code tables.

The enhanced hybrid DPCM/DCT scheme

The enhanced hybrid DPCM/DCT scheme was simulated under the same conditions as the base hybrid DPCM/DCT scheme, that is fixed quantisation was used. The plot of Figure 6.8 represents the difference in objective fidelity between use of the enhanced hybrid DPCM/DCT scheme and the base scheme, again where the predefined code tables used were generated for two target bit rates (0.25 and 0.5 bpp). Positive values represent an improvement in performance due to use of the enhanced scheme. It may be seen that the enhanced scheme provides an improvement in performance over the base scheme when used at moderate bit rates between approximately 0.25 and 0.75 bpp. The gain in objective fidelity for both curves reaches its maximum value at approximately the bit rate used as a target when producing the global code tables. At this point an increase of between approximately 0.15 and 0.25 dB has been observed. If the scheme is used to produce output bit rates far below that for which the global code tables were generated, however, a comparatively dramatic decrease in performance results. This occurs because at such low bit rates, the large quantisation table step sizes used result in many blocks which are entirely zero valued beyond the DC coefficient. At this point the vertical differential encoding of the DC coefficient used by the enhanced scheme becomes a disadvantage, as the (zero valued) AC coefficients must still be encoded separately. Such a block may be encoded more efficiently where the DC coefficient is included, as is the case in operation of the base scheme. It is also observed that for a global table target bit rate of 0.5 bpp, the decrease in objective fidelity upwards from the target bit rate has a shallow gradient and at every point constitutes an improvement in performance over the base hybrid DPCM/DCT scheme. In this region its performance is significantly better than for the 0.25 bpp target bit rate case. However, these differences appear to be balanced by those occurring below the target bit rate, where the fidelity loss associated with the 0.25 bpp target case is far less significant in value and gradient than in the 0.5 bpp

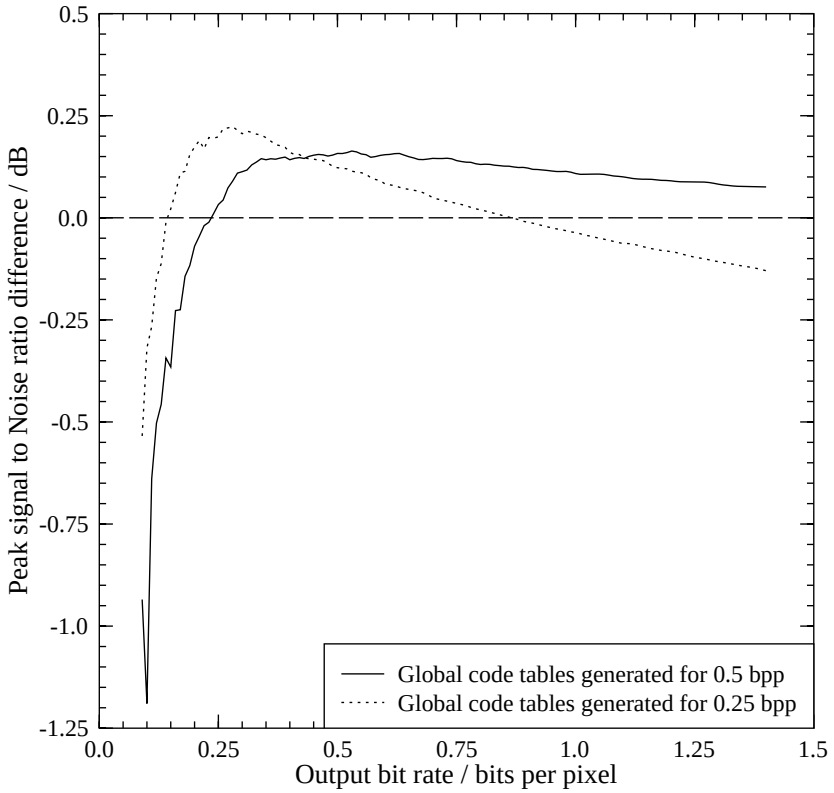


Figure 6.8: Difference in objective fidelity of the reconstructed intensity distribution between use of the enhanced hybrid DPCM/DCT scheme and the base hybrid DPCM/DCT scheme using fixed quantisation on the luma channel only of the source data; results for predefined code tables generated for bit rates of 0.25 bpp and 0.5 bpp are shown. Positive ordinate values constitute an improvement with use of the enhanced scheme.

target case.

The LVV quantisation array

The base and enhanced hybrid DPCM/DCT schemes were simulated where the LVV quantisation array, defined by Equation 6.10, was substituted for the fixed quantisation array. Figure 6.9 plots the difference in objective fidelity between both the base and enhanced schemes where LVV quantisation was used, and the enhanced scheme using the fixed array. When using LVV quantisation, the hybrid DPCM/DCT schemes have been found to give improved performance over their use with a fixed quantisation only at low bit rates, below approximately 0.4 bpp. At higher bit rates, use of the fixed quantisation table gives better rate-distortion performance. At low bit rates, where coarse quantisation has taken place, the most significant spatial frequency components, that is those with low vertical frequency values have been better represented by LVV quantisation, which has discarded unimportant high vertical frequency components. At such low bit rates, the use of fixed quantisation results in more of the low significance frequency components being preserved, leading to occasional nonzero values at these frequency positions. These must then be scanned and represented by the entropy coder resulting in a higher bit cost with a barely significant variation in reconstructed image fidelity between the two cases. It is evident, therefore, that the profile of the LVV quantisation array is unsuitable for use with simple linear scaling across the full output bit rate range. A more appropriate quantisation scheme should include dynamic variations with bit rate to ensure it remains efficient across the entire range. It is noted that the improvement at low bit rates due to use of LVV quantisation is smaller for either the base or enhanced scheme than the performance difference at moderate to high bit rates between the base and enhanced schemes using the same quantisation table. The effect of different target bit rates for the predefined Huffman code tables is barely significant, with any difference in objective fidelity accounted for by the general observed difference between fidelity when using the base and enhanced schemes at different values of target bit rate (Figure 6.8).

Table 6.3 gives a summary of the performance of the two scheme variations and two quantisation methods at three bit rates when used to encode the luma channels of four lenticular-integral intensity distributions. Where global (predefined) Huffman code tables have been used, these have been generated for a target bit rate of 0.25 bpp.

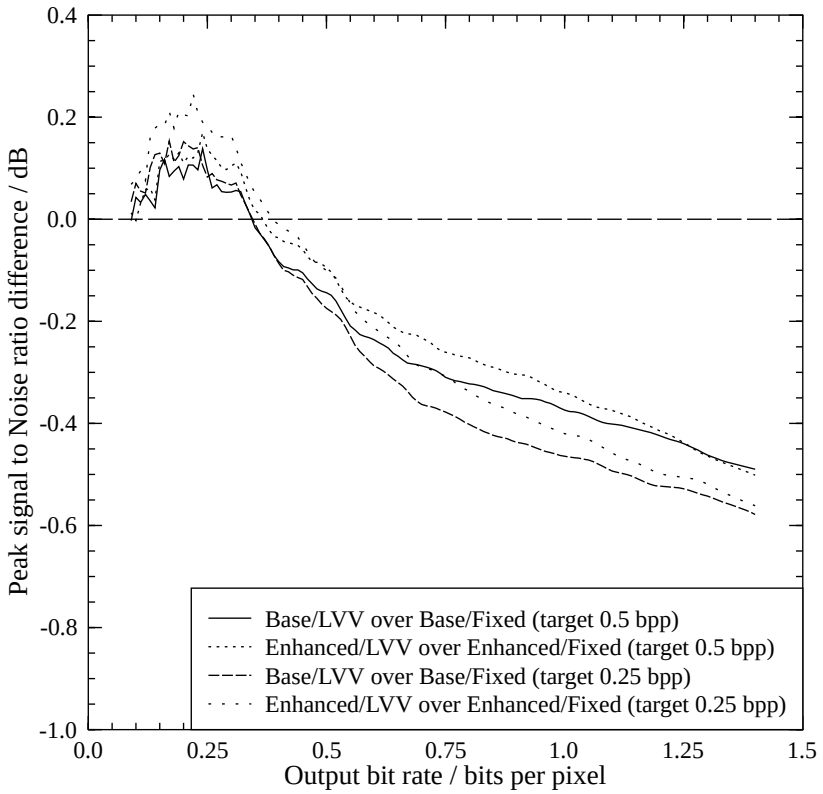


Figure 6.9: Difference in objective fidelity of the reconstructed intensity distribution between use of LVV and fixed quantisation under the base and enhanced hybrid DPCM/DCT schemes on the luma channel only of the source data; results for predefined code tables generated for bit rates of 0.25 bpp and 0.5 bpp are shown. Positive ordinate values constitute an improvement with use of LVV quantisation.

	Peak signal to Noise ratio / dB		
Encoder	@ 1 bpp	@ 0.5 bpp	@ 0.25 bpp
Baseline sequential-DCT JPEG	—	27.70	22.12
Base hybrid DPCM/DCT; fixed quant.; global codes	40.33	36.26	31.39
Base hybrid DPCM/DCT; fixed quant.; local codes	40.91	36.46	31.43
Enhanced hybrid DPCM/DCT; fixed quant.; global codes	40.30	36.38	31.59
Base hybrid DPCM/DCT; LVV quant.; global codes	39.87	36.08	31.47
Enhanced hybrid DPCM/DCT; LVV quant.; global codes	39.88	36.28	31.77

Table 6.3: Summary of performance for luma channel-only hybrid DPCM/DCT based coding of four lenticular-integral intensity distributions.

6.9.2 Full colour coding

Results are now presented of simulations of the hybrid DPCM/DCT schemes applied to full colour lenticular-integral intensity distribution data. The intensity distribution *Microscope*, being a greyscale example, has not been included in these tests. The source data are stored using the $R'G'B'$ colour model where each colour component is represented using an 8 bit integer; the source bit rate is therefore 24 bits per pixel. The source is converted to the $Y'C_bC_r$ colour space using $4:4:4$ sampling (that is no subsampling of the chrominance channels takes place) and full colour coding is implemented using three coding loops operating on the Y', C_b and C_r channels of the source separately. The reconstructed intensity distribution emerging from the decoder is remapped to the $R'G'B'$ colour space for display and evaluation. Quantisation is performed using two separate tables, one applied to luma and the other to chrominance-red/chrominance-blue data. The luma and chrominance tables are identical. Where predefined Huffman code tables are used they have been constructed as described in section 6.8. Separate tables were generated for use with luma and chrominance-blue/chrominance-red data. In all cases the target bit rate for the luma code tables was 0.25 bpp, as this value was found to lead to a lower overall performance deviation across all bit rates tested in the luma channel-only simulations.

Absolute comparison with baseline JPEG

Figure 6.10 shows performance curves for the base hybrid DPCM/DCT scheme using fixed quantisation and baseline sequential-DCT JPEG. Predefined Huffman code tables were used in the base hybrid scheme simulations, as described above. The hybrid DPCM/DCT schemes give a significant increase in performance over baseline JPEG for full colour lenticular-integral image data using identical luma and chrominance quantisation tables, as was the case in coding of the luma channel only.

Use of local Huffman code tables

Simulations were carried out using the base hybrid DPCM/DCT scheme with local Huffman code tables, generated for a particular encoding instance. The improvement in rate-distortion performance due to use of local code tables is illustrated by the difference in objective fidelity of the reconstructed intensity distribution for the local and global table cases, plotted in Fig-

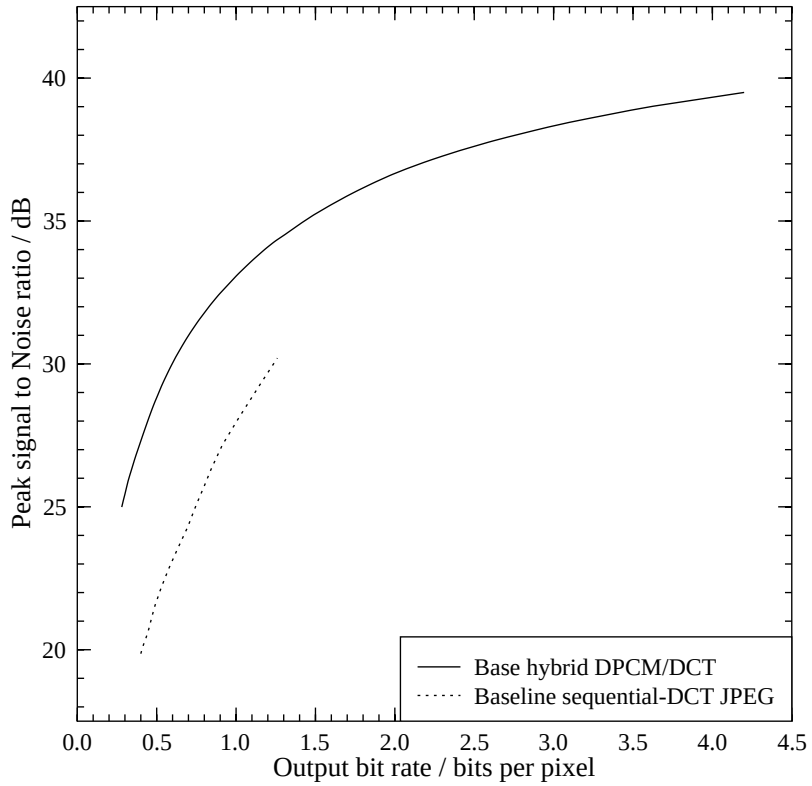


Figure 6.10: Performance of the base hybrid DPCM/DCT scheme using fixed quantisation on full colour source data; baseline sequential-DCT JPEG performance is shown for reference.

ure 6.11. As was the case in the luma channel only simulations, code table transmission overhead has not been included in output bit rate calculations in the local table case. The resulting difference can be seen to follow the same trend as observed in the same comparison for coding of only luma channel data, though it is smaller in both magnitude and range across the full bit rate range. The minimum difference in distortion occurs at a bit rate approximately equivalent to the global Huffman code table target value at which the minimum was observed in luma channel coding. When encoding three-channel intensity distribution data, the minimum would be expected to occur at a bit rate of approximately 0.75 bpp. The reason for its departure from this point is that the contributions made by the chrominance channels to the aggregate bit rate value produced in full colour coding is far smaller than that made by the luma channel. This is due to far more redundancy in chrominance data caused by extremely high intra-band and inter-band correlation, in turn leading to a far greater rate reduction than is achieved for luma channel data. At the minimum the difference of 0.070 dB is unlikely to render use of local code tables beneficial, considering the coding overhead re-

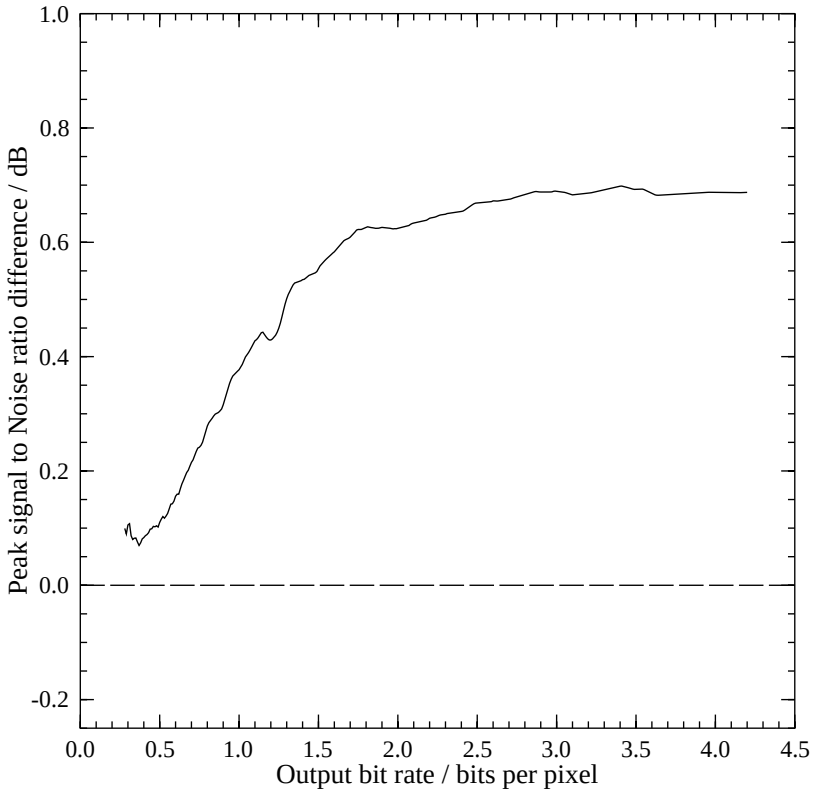


Figure 6.11: Improvement in performance of the base hybrid DPCM/DCT scheme using fixed quantisation due to use of a local rather than predefined code table on full colour source data.

quired due to communication of the tables to the decoder.

The enhanced hybrid DPCM/DCT scheme

Figure 6.12 contains a plot of the difference in objective fidelity obtained across a range of values of output bit rate between simulations of the enhanced and base hybrid DPCM/DCT schemes with fixed quantisation arrays. Positive values represent an improvement in performance due to use of the enhanced scheme. At bit rates between approximately 0.6 and 1.8 bpp an improvement is observed, with a maximum value of 0.053 dB at a bit rate of 0.98 bpp. The curve generally follows the same trend as in the case of encoding the luma channel only of the source intensity distributions, with a dramatic performance decrease at lower bit rates for the same reason as in the luma only case, and a performance loss with a comparatively small gradient and value at higher bit rates.

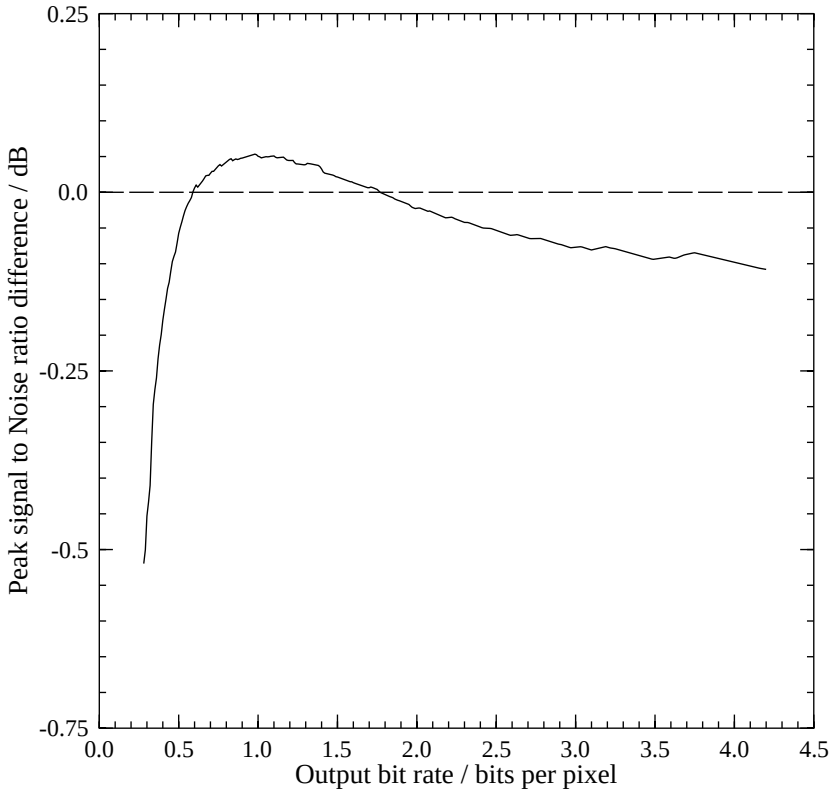


Figure 6.12: Difference in objective fidelity of the reconstructed intensity distribution between use of the enhanced hybrid DPCM/DCT scheme and the base hybrid DPCM/DCT scheme using fixed quantisation on full colour source data. Positive ordinate values constitute an improvement with use of the enhanced scheme.

The LVV quantisation array

Both variations of the hybrid DPCM/DCT coding scheme were simulated where the LVV quantisation array was used in place of the fixed array for both the luma and chrominance channels of full colour source data. The differential performance curves of Figure 6.13 show the gain and loss in objective fidelity obtained for the base and hybrid DPCM/DCT scheme with LVV quantisation relative to use of fixed quantisation arrays; positive values correspond to a performance improvement when LVV quantisation tables are used. Use of the LVV quantisation array can be seen to provide a performance benefit over use of a fixed quantisation table only at low bit rates. When encoding to bit rates above approximately 0.5 bpp takes place using LVV quantisation, however, a significant performance loss occurs. As discussed in the case of luma channel only coding, this is the case since the structure of the

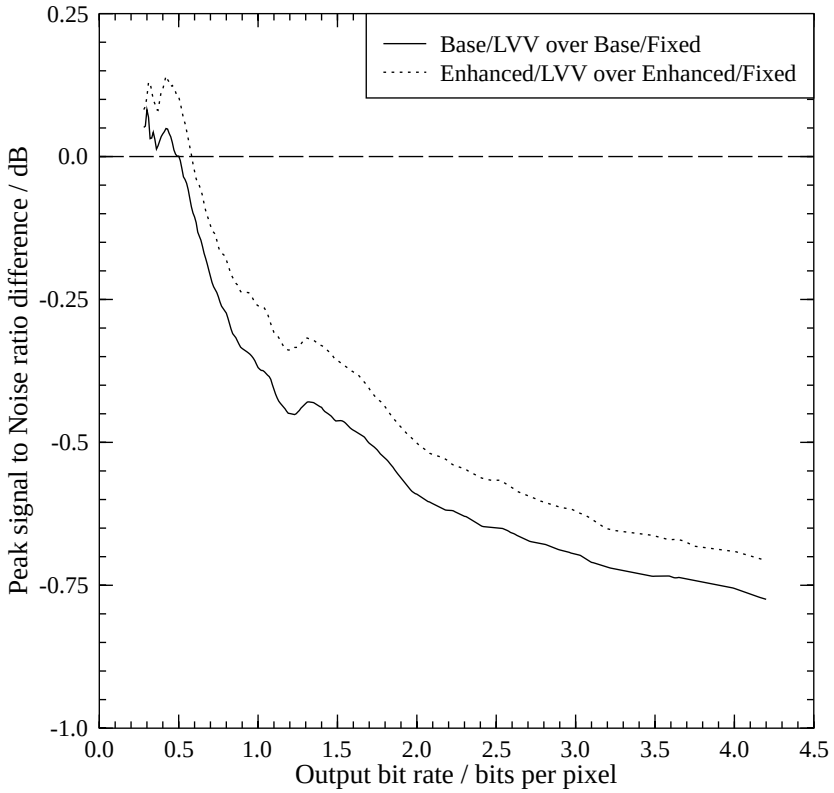


Figure 6.13: Difference in objective fidelity of the reconstructed intensity distribution between use of LVV and fixed quantisation in the base and enhanced hybrid DPCM/DCT schemes on full colour source data. Positive ordinate values constitute an improvement with use of LVV quantisation.

linearly scaled structure of the LVV quantisation array is only appropriate to the statistical characteristics of the integral intensity distribution when coarsely scaled. A quantisation array structure suitable for use when coding across the full bit rate range must therefore include dynamic variation with output bit rate. The general form of the difference in performance between use of the two quantisation schemes is similar when examining the case for both the base and enhanced hybrid DPCM/DCT schemes, with a greater increase in objective fidelity across the range of bit rates tested for the enhanced scheme. Unfortunately the performance gain noted for the enhanced scheme at low bit rates does not compensate for the poor performance of the enhanced scheme in general in this range (refer to Figure 6.12), and therefore no net gain is seen for the enhanced scheme with LVV quantisation over any other scheme/quantisation table combination tested. In this low bit rate range, the base scheme with LVV quantisation provides superior rate-distortion performance to all other schemes and quantisation tables simulated.

	Peak signal to Noise ratio / dB		
Encoder	@ 2 bpp	@ 1 bpp	@ 0.4 bpp
Baseline sequential-DCT JPEG	—	27.96	19.86
Base hybrid DPCM/DCT; fixed quant.; global codes	36.66	33.07	27.30
Base hybrid DPCM/DCT; fixed quant.; local codes	37.29	33.45	27.39
Enhanced hybrid DPCM/DCT; fixed quant.; global codes	36.64	33.12	27.12
Base hybrid DPCM/DCT; LVV quant.; global codes	36.07	32.70	27.34
Enhanced hybrid DPCM/DCT; LVV quant.; global codes	36.14	32.86	27.24

Table 6.4: Summary of performance for full colour hybrid DPCM/DCT based coding of three lenticular-integral intensity distributions.

Table 6.4 gives a summary of the performance of the two scheme variations and two quantisation methods in terms of objective fidelity at three bit rates, when used to encode three full colour lenticular-integral intensity distributions.

6.9.3 Reconstructed intensity distributions

Reconstructed decoded intensity distributions resulting from the simulations described above are reproduced in this subsection. The examples shown are from full colour application of both baseline sequential-DCT JPEG using standard JPEG quantisation, enhanced hybrid DPCM/DCT using fixed quantisation and base hybrid DPCM/DCT using LVV quantisation. Global Huffman code tables generated for a luma channel target bit rate of 0.25 bpp, previously used in simulation performance plots above, were employed. For each case the appropriate source intensity distribution, marked '[Original]', is also given for comparison.

Images 6.1 and 6.2 show the reconstructed image *Horseman*, having been coded to 1.71 bpp and 0.76 bpp respectively using the enhanced hybrid DPCM/DCT scheme with fixed quantisation. No differences can be seen between either of these images and the source image either in intensity distribution form or when replayed as integral 3D images. Image 6.3 shows *Horseman* having been reconstructed after coding to 0.37 bpp using the base hybrid DPCM/DCT scheme with LVV quantisation. In this case over-quantisation has led to misrepresentation of the contents of each subimage and though the intensity distribution does replay correctly as an integral 3D image, some loss of detail is noticeable. The reconstructed intensity distribution encoded using the baseline JPEG scheme to 0.37 bpp (Image 6.4) con-

tains similar unwanted artifacts, though here blocking artifacts and colour misrepresentation effects have become highly objectionable. Additionally in this case the delineation between lenticular bands important to maintain the structure of the intensity distribution and hence for high quality three dimensional display, has not been reproduced clearly. In three dimensional replay the main noticeable distortions are a general loss of detail and incorrect colour representation. To enable the distortion in the reconstructed intensity distributions to be more clearly seen, Figure 6.14 shows magnified sections of the source intensity distribution and reconstructed distributions of Images 6.2, 6.3 and 6.4.

Reconstructed intensity distributions are also shown for compression of *Toyplane* to similar bit rates to those used in preparation of the *Horseman* examples and for the same coding schemes (Images 6.5 to 6.8). Magnified sections of the decoded intensity distributions are again shown, in Figure 6.15.



[Original]



Image 6.1: *Horseman* compressed using the enhanced hybrid DPCM/DCT scheme with fixed quantisation; output bit rate is 1.71 bpp. [SCREEN B]



[Original]



Image 6.2: *Horseman* compressed using the enhanced hybrid DPCM/DCT scheme with fixed quantisation; output bit rate is 0.76 bpp. [SCREEN B]



[Original]



Image 6.3: *Horseman* compressed using the base hybrid DPCM/DCT scheme with LVV quantisation; output bit rate is 0.37 bpp. [SCREEN B]



[Original]



Image 6.4: *Horseman* compressed using the baseline sequential-DCT JPEG scheme with JPEG standard quantisation; output bit rate is 0.37 bpp. [SCREEN B]

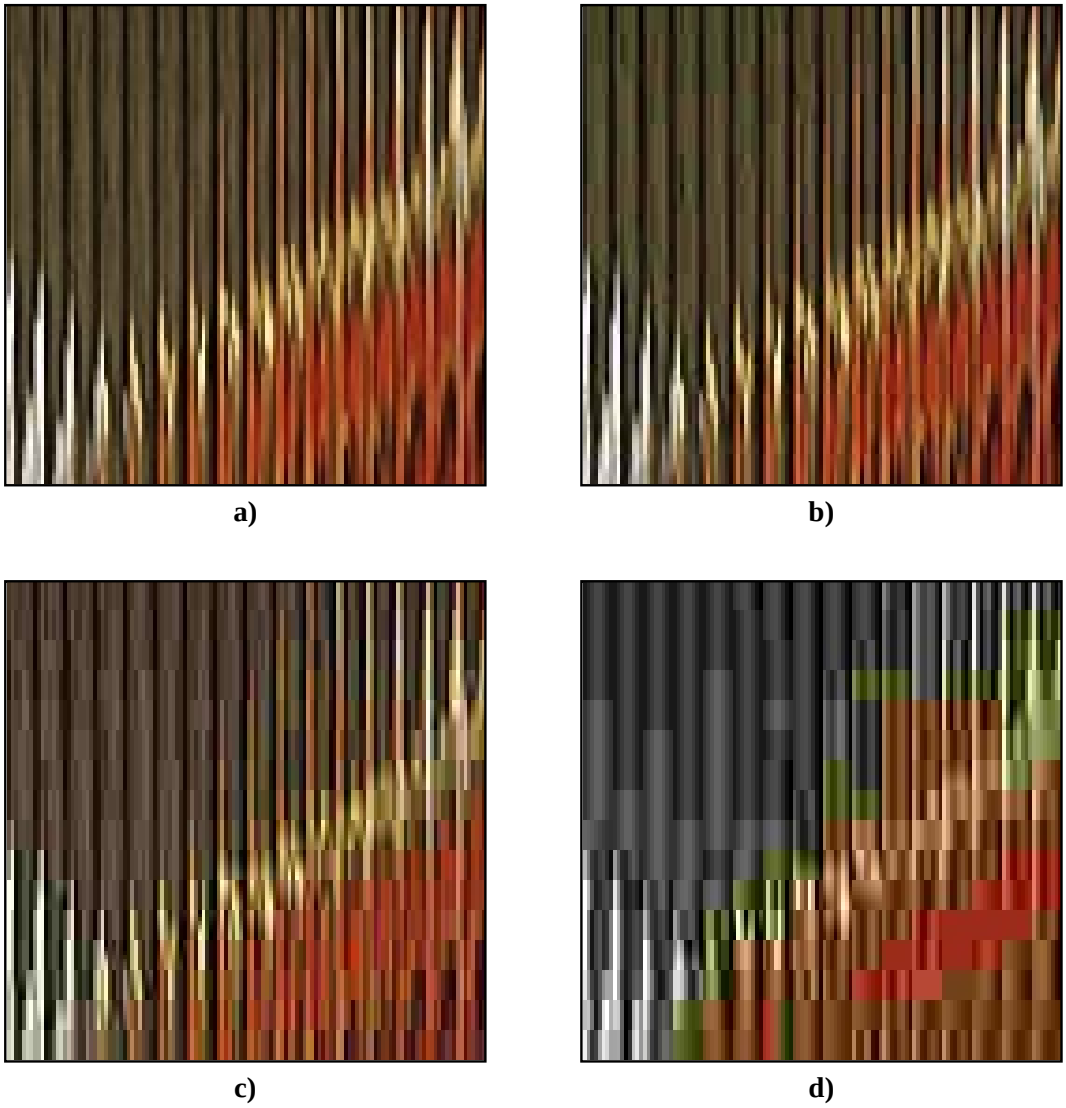


Figure 6.14: Magnified 16×16-subimage sections of *Horseman* resulting from coding using hybrid DPCM/DCT and JPEG schemes.
a) Source.
b) Reconstructed, enhanced hybrid with fixed quantisation at 0.76 bpp.
c) Reconstructed, base hybrid with LVV quantisation at 0.37 bpp.
d) Reconstructed, baseline JPEG with JPEG quantisation at 0.37 bpp.



[Original]



Image 6.5: *Toyplane* compressed using the enhanced hybrid DPCM/DCT scheme with fixed quantisation; output bit rate is 1.75 bpp. [SCREEN A]



[Original]



Image 6.6: *Toyplane* compressed using the enhanced hybrid DPCM/DCT scheme with fixed quantisation; output bit rate is 0.76 bpp. [SCREEN A]



[Original]



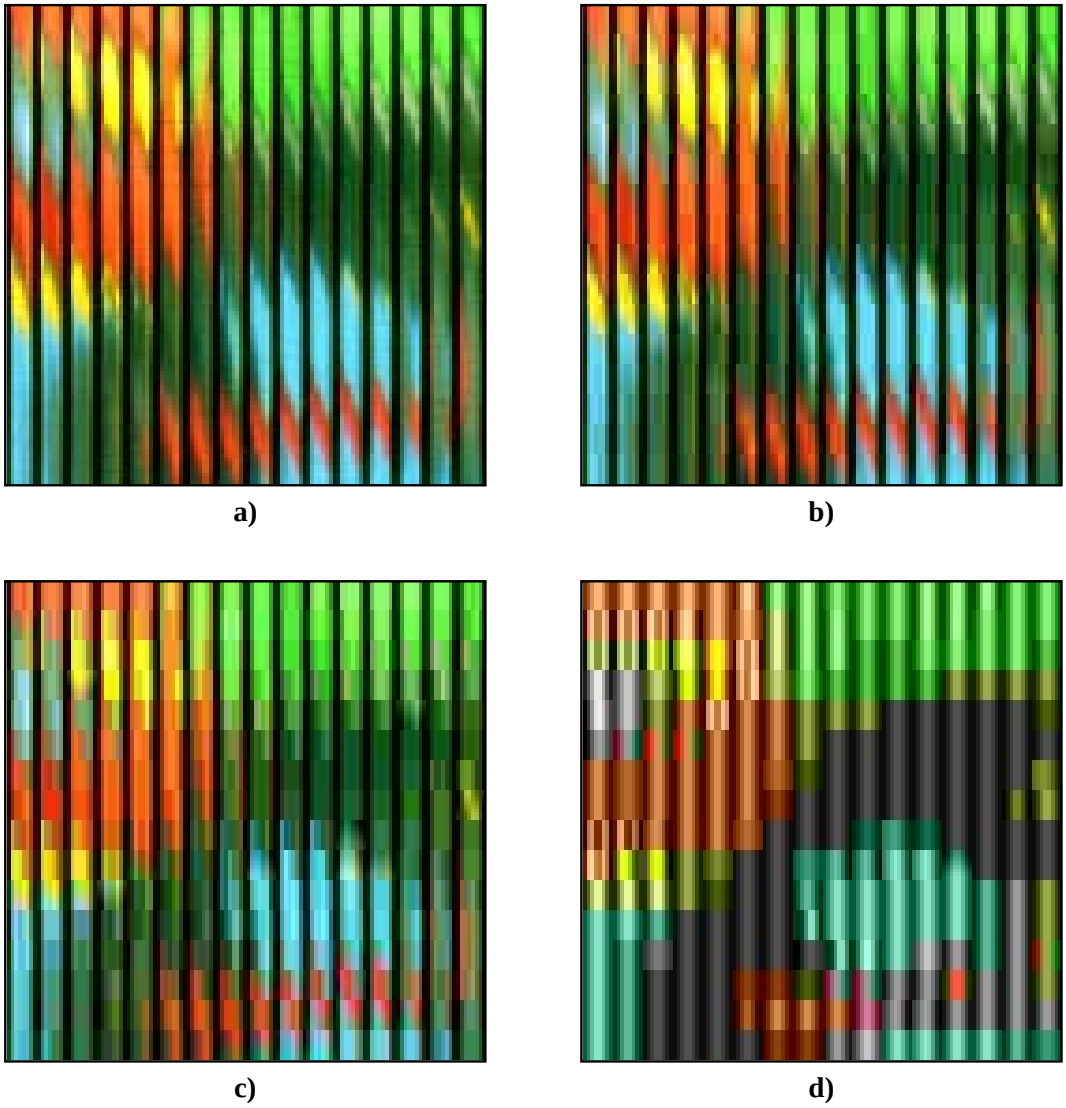
Image 6.7: *Toyplane* compressed using the base hybrid DPCM/DCT scheme with LVV quantisation; output bit rate is 0.41 bpp. [SCREEN A]



[Original]



Image 6.8: *Toyplane* compressed using the baseline sequential-DCT JPEG scheme with JPEG standard quantisation; output bit rate is 0.41 bpp. [SCREEN A]



a) **b)**

c) **d)**

Figure 6.15: Magnified 16×16-subimage sections of *Toyplane* resulting from coding using hybrid DPCM/DCT and JPEG schemes.
a) Source.
b) Reconstructed, enhanced hybrid with fixed quantisation at 0.76 bpp.
c) Reconstructed, base hybrid with LVV quantisation at 0.41 bpp.
d) Reconstructed, baseline JPEG with JPEG quantisation at 0.41 bpp.

Summary

This chapter has described a compression scheme developed for use specifically with intensity distribution data which represent integral three dimensional images recorded using lenticular arrays. Decorrelation of the data both between subimages from neighbouring bands and within subimages is effected using a hybrid DPCM/DCT mapping stage. A scalar quantisation stage is used to reduce the entropy of the symbols presented to the entropy coder, a Huffman-based unit similar to that used by the baseline JPEG conventional continuous tone image compression system.

Two variations of the scheme using slightly different entropy coding stages have been implemented for simulation and presented with intensity distribution data from capture of three dimensional images using the De Montfort University integral imaging camera system. Simulations have been carried out initially with the luma channel only of four intensity distributions as source data, and also in full colour mode on three intensity distributions. Two quantisation arrays have been used: the fixed array in which a constant step size is used in each position and the LVV array which takes account of the expected variability of each coefficient. Both local Huffman code tables, generated by the encoder on a per-image basis, and global code tables, predefined based on a given set of intensity distributions and quantisation parameters, have been used by the coding schemes. In all cases decoded image fidelity was assessed objectively using the peak signal to noise ratio metric and subjectively by examining unwanted coding artifacts and by assessing the quality of their replay as integral 3D images.

For full colour lenticular-integral intensity distribution coding, the peak signal to noise ratio of reconstructed data encoded by the hybrid DPCM/DCT scheme in its fundamental form, the base hybrid DPCM/DCT scheme using predefined Huffman code tables, has been found to be between 5 and 7.4 dB greater than that of the same data from the baseline JPEG scheme at the same bit rate. The improvements in rate-distortion performance between variants of the hybrid DPCM/DCT scheme using two forms of quantisation array have been found to be dependent upon the bit rate required. In all cases where global (predefined) Huffman code tables were used, the rate-distortion performance of the hybrid DPCM/DCT schemes determined across a range of output bit rate values was found to be at a maximum within a small distance from the bit rate for which the Huffman tables were generated. Hence it is important that for a particular application, to achieve a maximum performance level the global Huffman code tables used are constructed for a target bit rate value appropriate to the output bit rate required of the application. Where local Huffman code tables are

used, generated for the specific intensity distribution at the encoder, an improvement in rate-distortion performance is achieved for all output bit rate values tested. Additionally use of local tables removes the variability of performance with bit rate experienced where global code tables are used. For a particular application, these benefits must be compared with the additional encoder overhead due to the necessity to transmit Huffman code tables to the decoder. It is likely that in a motion video coding environment a hybrid approach will be most suitable, with Huffman code tables generated from groups of consecutive frames and used in encoding a finite number of the following frames.

Three dimensional replay of the intensity distributions tested was found to be highly resilient to distortion due to vertical blocking of reconstructed subimages and ringing. Such artifacts occur at very low bit rates due to excessively coarse quantisation. An overall loss of detail and colour definition was found to occur in these cases. In general the fidelity of replayed intensity distributions reconstructed after coding using the hybrid DPCM/DCT schemes was found to be very high. At bit rates above approximately 1 bit per pixel (a compression ratio of $24:1$) the replayed three dimensional image was found by the author to be indistinguishable from the original, uncompressed image.

The compression performance obtained is now compared with that required for transmission of digital video data. It should be noted that the hybrid DPCM/DCT schemes described in this chapter were designed for use with still image data. Therefore, since they do not take account of the significant inter-frame correlation expected to occur in integral 3D video data, the performance achieved is not optimal for such data. Chapter 3 derived compression ratios required for coding of digital video data for cable system broadcast at standard and high definition television resolutions. Those values cannot be compared directly with those produced for the hybrid DPCM/DCT schemes in this chapter since the simulations in this chapter have used data with $4:4:4$ $Y'C_bC_r$ sampling rather than $4:2:2$ sampling and hence 16 bpp source bit rate. When 24 bpp digital video data are to be transmitted, the compression ratios required are $6.9:1$ and $36.8:1$ for the standard and high definition cases respectively.

At standard definition resolution ($6.9:1$ compression ratio), the hybrid DPCM/DCT schemes have reconstructed lenticular-integral intensity distributions with no noticeable distortion. As mentioned above, distortion-free integral images have been found to result from coding at up to approximately $24:1$, and therefore neglecting error correction code overhead, three high quality 3D television pictures could be transmitted at standard definition resolution. Such high quality reconstruction is however not possible in the high definition case, since the fidelity obtained at the required compression ratio of $36.8:1$ is not sufficiently high.

Chapter 7 describes the initial development of a second integral image data compression scheme, based on the use of a three dimensional discrete cosine transform in the mapping stage, with improved rate-distortion performance.

7. A 3D-DCT-BASED INTEGRAL 3D IMAGE COMPRESSION SCHEME

Introduction

An integral 3D image compression scheme employing hybrid DPCM/DCT decorrelation has been developed in the previous chapter. Here a scheme is described which uses a three dimensional DCT to exploit both intra-subimage and inter-subimage correlation simultaneously, resulting in more efficient decorrelation of the source intensity distribution. A three dimensional scalar quantisation stage is applied to the DCT coefficients which are then entropy encoded by a Huffman-based coder.

The 3D-DCT based scheme has been implemented for simulation and applied to lenticular-integral image data. Results of these simulations are presented and discussed, and compared with those obtained from simulation of the previous hybrid DPCM/DCT schemes.

7.1 The 3D-DCT Mapping Stage

The three dimensional DCT has previously been used in image compression applications, for example low bit rate coding for videoconferencing systems [55]. In the 3D-DCT based integral image compression scheme described in this chapter, both inter-subimage and intra-subimage decorrelation of the source intensity distribution data are performed simultaneously by a three dimensional DCT. A *group* of neighbouring subimages from the source intensity distribution, between which inter-band correlation is high, is taken to form a volume of input data for the DCT. The group of length N, beginning with subimage (i, j),

$$G_{i,j}(l,m,n) = X_{i+n,j}(l,m) \qquad \text{for } l,m = 0,1,\ldots,7 \text{ and } n = 0,1,\ldots,N-1, \tag{7.1}$$

where $X_{i,j}$ is the subimage at position (i, j) in the source intensity distribution. The process is illustrated by Figure 7.1 for the case $N = 4$. The 3D-DCT is applied to the resulting volume to produce the decorrelated subimage group. By extension of Equation 6.4, the forward DCT in three dimensions of lengths 8, 8 and N respectively is:

$$Y_{i,j}(u,v,w) = \frac{1}{\sqrt{8N}} K_u K_v K_w \sum_{n=0}^{N-1} \sum_{m=0}^{7} \sum_{l=0}^{7} G_{i,j}(l,m,n)\, C_{l,u,8}\, C_{m,v,8}\, C_{n,w,N}$$

$$\text{for } u,v = 0,1,\ldots,7, \quad w = 0,1,\ldots,N-1, \tag{7.2}$$

$$\text{where } K_s = \begin{cases} 1/\sqrt{2} & \text{if } s = 0, \\ 1 & \text{otherwise,} \end{cases}$$

$$\text{and } C_{a,b,c} = \cos\left(\frac{b(2a+1)\pi}{2c}\right).$$

The overall effectiveness of the 3D-DCT mapping stage is expected to be greater than that of the hybrid DPCM/DCT method used previously. This is the case for two reasons. Firstly, the DCT in general provides more effective decorrelation of image data than linear prediction and secondly the DPCM/DCT decorrelator takes account of the correlation between only two

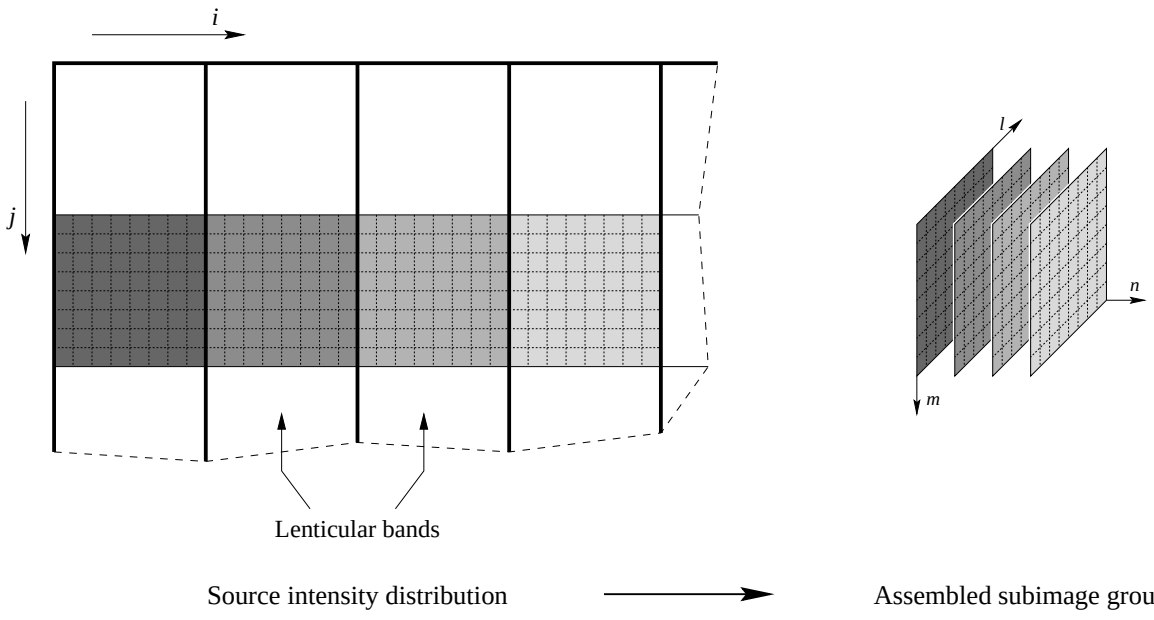


Figure 7.1: Assembly of a subimage group for 3D-DCT based lenticular-integral image coding.

neighbouring subimages while when $N>2$ the 3D-DCT method includes several contiguous subimages in one transform computation. The group length N is a parameter of the system, and due to implementation considerations $N = \{2^a; a \in \mathbb{N}\}$. The group length is chosen to provide maximally efficient decorrelation of all subimages from the full source intensity distribution.

7.2 Quantisation of 3D-DCT-decorrelated Subimage Groups

Decorrelated subimage groups are presented to a scalar quantisation stage operating in a similar manner to that used for hybrid DPCM/DCT scheme coefficient quantisation. Each of the $64N$ coefficient positions in each group has a specific uniform quantiser, parameterised by a three dimensional array of step sizes Q with dimensions $8\times8\times N$. The quantised coefficient group starting at subimage position (i, j),

$$S_{i,j}(u,v,w) = \text{round}\left(\frac{Y_{i,j}(u,v,w)}{Q(u,v,w)}\right) \qquad \begin{array}{l}\text{for } u,v = 0,1,\ldots,7,\\ \quad w = 0,1,\ldots,N-1,\\ \text{with round}(a) = \begin{cases}\lfloor 1/2 + a\rfloor & \text{if } a \geq 0,\\ -\lfloor 1/2 - a\rfloor & \text{otherwise.}\end{cases}\end{array} \tag{7.3}$$

7.2.1 Quantisation array selection

As was the case for quantisation in the hybrid DPCM/DCT schemes, the fundamental *fixed quantisation table* is defined initially since there is no precedent for use of psychovisually-weighted or any other variable step size quantisation array. Where q is the quantiser scale, the fixed quantisation array,

$$Q(u,v,w) = q \qquad \begin{array}{l}\text{for } u,v = 0,1,\ldots,7,\\ \quad w = 0,1,\ldots,N-1,\\ \text{with } q \in \mathbb{N}^+.\end{array} \tag{7.4}$$

In this chapter the fixed quantisation table is used exclusively. Chapter 8 covers the development of improved 3D-DCT scheme quantisation strategies.

7.3 Entropy Coding of Quantised Coefficient Groups

The entropy coding stage of the 3D-DCT based encoder uses a Huffman-based scheme similar to that used by the hybrid DPCM/DCT encoder. Generally this involves scanning the quantised coefficients into a one dimensional vector or vectors with increasing spatial frequency, mapping those vectors to sets of coder symbols and then constructing a table or tables of minimum length codes representing those symbols using the Huffman technique. Two methods are now described which have been developed for this purpose.

7.3.1 Per-plane zigzag scanning and coding

In this strategy, each quantised (u,v)-plane is coded using a separate codebook. A 64-coefficient zigzag scan is first performed on each quantised (u,v)-plane to order the coefficients from each by increasing (u,v) spatial frequency. The resulting N coefficient sequences are mapped to Huffman coder symbols using the strategy described for base hybrid DPCM/DCT entropy coding, though a third special symbol, 'ESG' (end of subimage group) is defined. It is used in place of 'ESB' in two cases: in the final sequence of the group (sequence $w = N-1$) to signify that the remaining coefficients are zero, or where *all* remaining coefficients in the current sequence, where the current sequence $w < N-1$, and any following sequences for the group have the value zero. Figure 7.2 illustrates the use of the 'ESB' and 'ESG' symbols in the 3D-DCT scheme.

The frequency of each symbol in each (u,v)-plane derived sequence across every subimage group from the source intensity distribution is then computed. The resulting N frequency tables are used to generate N Huffman codebooks. To fully specify each nonzero coefficient in the zigzag sequences and its preceding run of zero valued coefficients, each final output codeword is comprised of the Huffman code for the appropriate symbol followed by S bits to represent the sign and magnitude, where S is the coefficient's magnitude category.

7.3.2 All-plane, per-row scanning and coding

Since the DCT is an orthogonal transform, following quantisation the spatial significance of each coefficient to producing a reconstructed image with minimum mean square error depends on its variance [28]. To achieve a maximally efficient coding of the symbols rep-

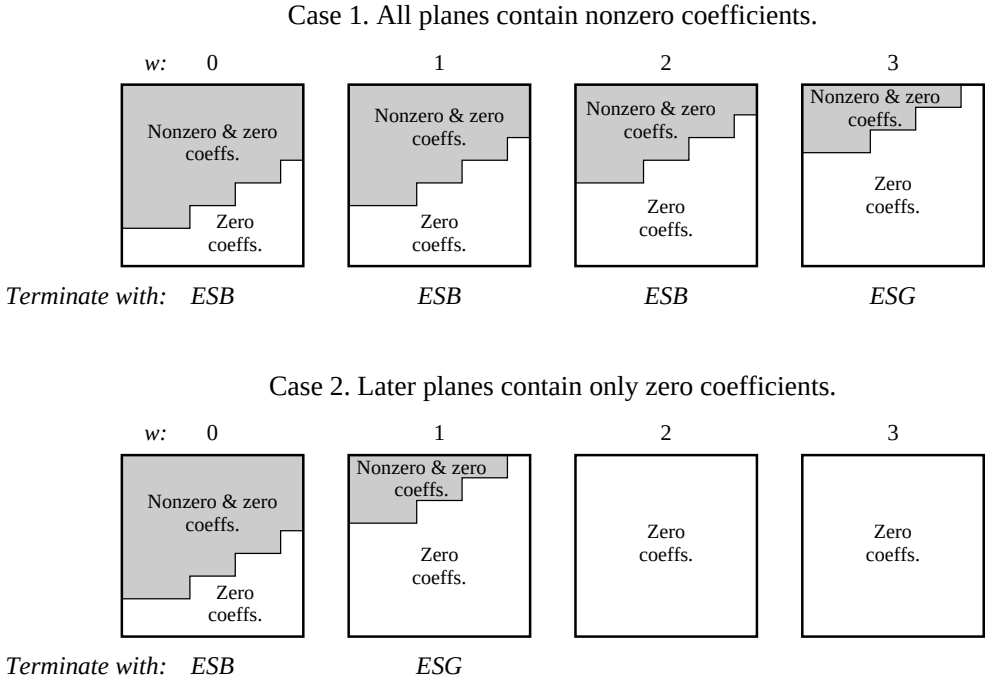


Figure 7.2: Use of the 'ESB' and 'ESG' symbols in 3D-DCT scheme coefficient-to-symbol mapping where group length $N = 4$.

resenting the coefficients from a subimage group, the scanning strategy must order those coefficients for coding approximately in order of steadily decreasing variance. This is the case since such an ordering tends to result in an earlier sequence termination code signalling that the remaining quantised coefficients are zero valued. To judge the efficiency of the per-plane zigzag scanning method described in subsection 7.3.1, the variance of each coefficient in the subimage group has been computed.

For the decorrelated coefficient position (u,v,w), the variance over all $(I \times J)/N$ subimage groups from K intensity distributions,

$$\sigma^2_{u,v,w} = \frac{N}{IJK - N} \sum_{k=1}^{K} \sum_{j=1}^{J} \sum_{\gamma=0}^{\frac{I}{N}-1} \left[Y_{i=(1+\gamma N),j,k}(u,v,w) - \overline{Y}(u,v,w) \right]^2 \tag{7.5}$$

$$\text{for } u,v = 0,1,\ldots,7 \text{ and } w = 0,1,\ldots,N-1.$$

Here $Y_{i,j,k} = Y_{i,j}$ taken from intensity distribution k and $\overline{Y}(u,v,w)$ is the sample mean for coefficient position (u,v,w) evaluated over all subimage groups from all K intensity distributions. Equation 7.5 was evaluated for the decorrelated coefficient groups of the luma channels of the intensity distributions *Microscope*, *Tank*, *Horseman* and *Toyplane*, for the

case $N = 4$. Figure 7.3 shows the variance of each coefficient as a function of its index in the zigzag scan for each (u,v)-plane in the group. The four lines on the plot represent the best fit to the data in a least squares sense. An efficient scanning strategy results in an arrangement of logs of coefficient variances against log of scan index such that the best fit lines are closely surrounded with data points with steadily decreasing values. It is observed that in the plots of Figure 7.3, the data points are spread widely about each line, particularly those points representing the most spatially significant low frequency coefficients. It is also noted that the general form of the distribution of points around its best fit line is the same for each (u,v)-plane but with lower overall variance as w increases. It is concluded that a per-plane zigzag coefficient scanning strategy is not efficient and thus not appropriate for coding of DCT coefficients in the 3D-DCT lenticular-integral image coding scheme.

A more appropriate scanning and entropy coding strategy is now described. Figure 7.4 presents the coefficient variance data as an intensity plot over coefficient positions in each plane. It is observed that there is an imbalance between variances of coefficients at corre-

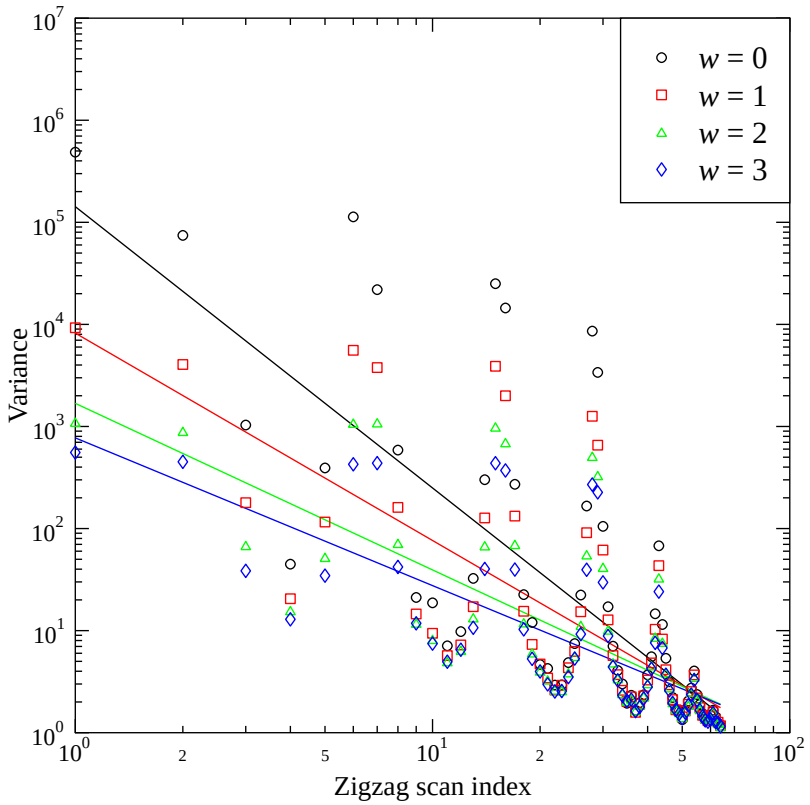


Figure 7.3: Variances of 3D-DCT coefficients along the zigzag scan for each (u,v)-plane in a group of length 4.

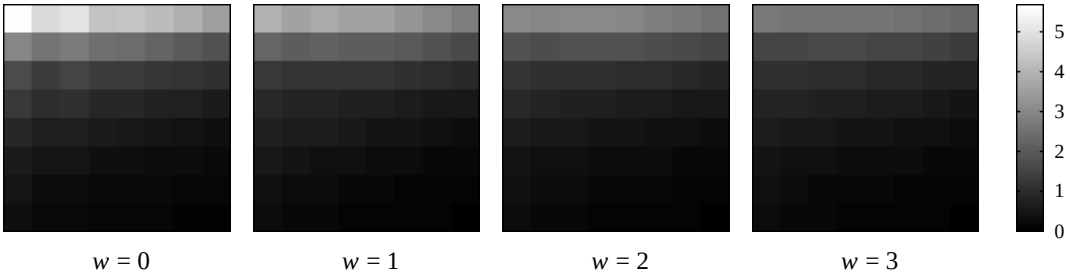


Figure 7.4: Log of variance of 3D-DCT coefficients by (u,v)-position in each plane; group length $N = 4$.

sponding horizontal (u) and vertical (v) spatial frequencies. This property of the transformed subimage group is due to the unique structure of lenticular-integral intensity distribution data, and is in agreement with the differential image histograms presented in chapter 5. It can be seen that an ordering of coefficient positions by steadily decreasing variance, as is required to derive an effective scanning strategy, occurs along a given line of constant v throughout all planes $0 \ldots 3$, rather than on a per plane basis as v is increased. A scanning strategy which follows this ordering is illustrated in Figure 7.5. A single scan is now performed throughout the entire subimage group and the resulting sequence, of length $64N$, mapped to a set of symbols for a single Huffman coder using the technique described for hybrid DPCM/DCT scheme coding. The 'ESG' special symbol is used where all remaining coefficients in the scanned sequence are zero. The improvement in classification of coefficients by reducing variance can be seen in Figure 7.6; the log variance values are clustered far closer around the best fit line than was the case for the per plane zigzag scanning method (Figure 7.3).

The figure shows that the new coefficient ordering results in a scanned sequence with steadily decreasing variance as desired, which is expected to provide a more efficient coding of each transformed subimage group. A further benefit of this scanning and coding method is that

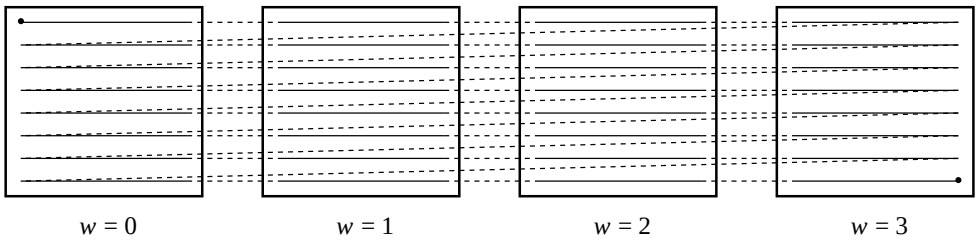


Figure 7.5: An improved scanning strategy for 3D-DCT based lenticular-integral image coding; in this case $N = 4$.

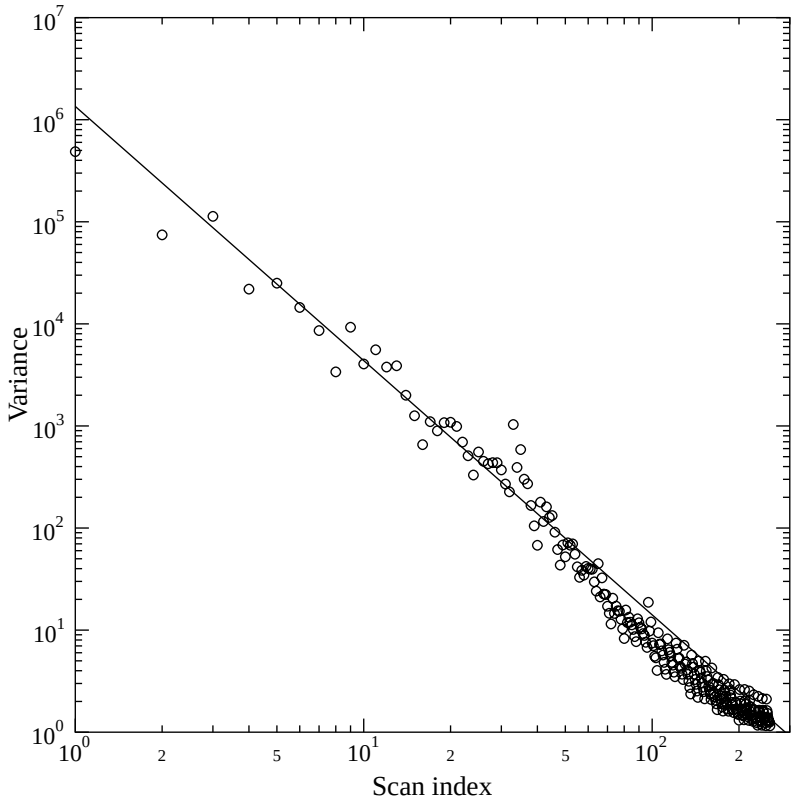


Figure 7.6: Variances of 3D-DCT coefficients along the scan for the improved scanning strategy in a group of length 4.

only one Huffman code table must be produced and therefore transmitted to the decoder, rather than N code tables as is the case for the per plane technique. This leads to lower coding overhead when control information is included in the output bit stream. Reductions in raw output bit rate were observed when using this improved scanning/coding strategy. The comparison between methods is included in section 7.7 which presents results of simulations of the 3D-DCT based scheme.

7.4 The Complete 3D-DCT Scheme Encoder and Decoder

The 3D-DCT based lenticular-integral image encoder and decoder, using the mapping, quantisation and entropy coding stages described above are shown in Figures 7.7a and 7.7b respectively.

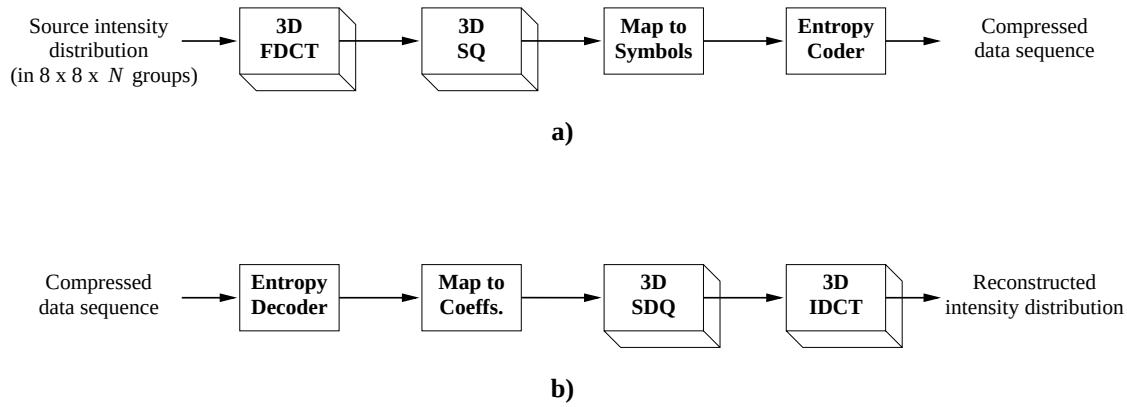


Figure 7.7: The 3D-DCT based still integral 3D image compression scheme. a) Encoder. b) Decoder.

7.5 Implementation of the 3D-DCT Scheme for Simulation

In this section, implementation of the 3D-DCT scheme for performance evaluation is described. The decoded intensity distribution, encoder output bit rate and objective fidelity are obtained for a given source intensity distribution and quantisation array. Procedure 7.1 is a pseudo-code listing for the top level of the simulation system. The system has again been implemented using the C language under the UNIX operating system.

The procedure 'Encoder_main_loop' performs the mapping and quantisation functions and produces as its result an array of quantised coefficient groups for the complete source intensity distribution and the reconstructed intensity distribution as would be output by the decoder (Procedure 7.2). The forward and inverse 3D-DCT are implemented by application of a 1D-DCT in the row, column and depth dimensions. The 1D-DCT was again implemented using routines from the 'FFTPACK' library.

As in the hybrid DPCM/DCT simulation implementation of chapter 6, the Huffman codes used by the entropy coder may either be built on a per-image basis within the encoder or predefined codes generated independently of the encoder may be used. The implementation of the 3D-DCT scheme simulation system also accommodates these two Huffman code sources. If local Huffman code tables are used, the following routines are invoked. If the scanning strategy in use is the per-plane zigzag method, Procedure 7.3 ('Gen_sym_tables_pp') generates, from the quantised coefficient group array, N tables of symbol frequencies, that is one table per (u,v)-plane in the group. When the all-plane improved scanning strategy is in use, Procedure 7.4 ('Gen_sym_table_ap') generates the single table of symbol frequen-

Procedure 7.1 Top level 3D-DCT scheme simulation system.

Arguments:
INIMG {*source intensity distribution*}
QMTX {*quantisation array*}
W {*subimage group length*}
[HUFF] {*Huffman code lengths (if using predefined code tables)*}
Results:
OUTIMG {*reconstructed decoded intensity distribution*}
OBPP {*output bit rate*}
QUAL {*objective fidelity of reconstructed intensity distribution*}

(QUANT_COEFFS, OUTIMG) ⇐ Encoder_main_loop(INIMG, QMTX, *W*)
if using per-plane zigzag scanning strategy **then**
  if not using predefined code tables **then**
    FREQ ⇐ Gen_sym_tables_pp(QUANT_COEFFS, *W*)
      {*Generate tables of coder symbol frequencies for all W planes in group*}
    HUFF ⇐ Gen_code_lengths_pp(FREQ, *W*)
      {*Compute lengths of Huffman codes for each symbol for all W planes*}
  end if
  OBITS ⇐ Calc_bit_counts_pp(QUANT_COEFFS, HUFF, *W*)
    {*Calculate total bit count for all coded groups*}
else {*using all-plane improved scanning strategy*}
  if not using predefined code table **then**
    FREQ ⇐ Gen_sym_table_ap(QUANT_COEFFS, *W*)
      {*Generate table of coder symbol frequencies*}
    HUFF ⇐ Gen_code_lengths_ap(FREQ)
      {*Compute lengths of Huffman codes for each symbol*}
  end if
  OBITS ⇐ Calc_bit_count_ap(QUANT_COEFFS, HUFF, *W*)
    {*Calculate total bit count for all coded groups*}
end if
OBPP ⇐ OBITS / Number_of_pixels(INIMG) {*Compute output bit rate (bits per pixel)*}
QUAL ⇐ Compute_PSNR(OUTIMG, IMG) {*Compute objective fidelity for decoded data*}

cies required. The lengths of Huffman codes representing the symbols generated are then computed using the symbol frequency table(s) generated by the previous stage. The procedures 'Gen_code_lengths_pp' (for per-plane scanning) or 'Gen_code_lengths_ap' (for all-plane scanning) carry out this process using the routine previously described for the hybrid DPCM/DCT coding scheme.

Irrespective of the source of Huffman code tables, the total bit count for the coded intensity distribution is next calculated. If the per-plane scanning method is in operation, the code lengths for all generated symbols representing the coded intensity distribution are calculated for all N planes in the group and summed. Procedure 7.5 ('Calc_bit_counts_pp') performs

Procedure 7.2 Main loop for 3D-DCT scheme simulation.

Arguments:
INIMG {*source intensity distribution*}
QMTX {*quantisation array*}
W {*subimage group length*}
Results:
QUANT_COEFFS {*quantised transform coefficients*}
OUTIMG {*reconstructed decoded intensity distribution*}

$I \Leftarrow$ number of subimage columns in INIMG
$J \Leftarrow$ number of subimage rows in INIMG
for $j = 1$ to J **do**
  for $i = 1$ to I step W **do**
    DCT $\Leftarrow$ Forward_3D_DCT($\text{INIMG}_{i,j}$)
      {*Inter- and intra-subimage decorrelation of group starting at subimage* (i, j)}
    $\text{QUANT_COEFFS}_{i,j} \Leftarrow$ Nearest_integer(DCT / QMTX) {*Quantisation*}
    DQUANT $\Leftarrow \text{QUANT_COEFFS}_{i,j} \times$ QMTX {*Dequantisation*}
    $\text{OUTIMG}_{i,j} \Leftarrow$ Inverse_3D_DCT(DQUANT)
      {*Reconstruction of group starting at subimage* (i, j)}
  end for
end for

Procedure 7.3 3D-DCT scheme symbol frequency table generation (per-plane scanning).

Arguments:
QUANT_COEFFS {*quantised transform coefficients*}
W {*subimage group length*}
Results:
FREQ {*tables of coder symbol frequencies*}

$I \Leftarrow$ number of subimage columns in QUANT_COEFFS
$J \Leftarrow$ number of subimage rows in QUANT_COEFFS
FREQ $\Leftarrow$ clear to all zero
for $j = 1$ to J **do**
  for $i = 1$ to I step W **do**
    for $w = 0$ to $W - 1$ **do**
      $\text{ZZ}_w \Leftarrow$ Zigzag_scan($\text{QUANT_COEFFS}_{i,j}(w)$)
        {*Scan plane w from group starting at* (i, j)}
      $\text{SYM}_w \Leftarrow$ Map_to_coder_symbols(ZZ_w)
      for all symbols S in SYM_w **do**
        $\text{FREQ}_w(S) \Leftarrow \text{FREQ}_w(S) + 1$
      end for
    end for
  end for
end for

Procedure 7.4 3D-DCT scheme symbol frequency table generation (all-plane scanning).

Arguments:
QUANT_COEFFS {*quantised transform coefficients*}
W {*subimage group length*}
Results:
FREQ {*table of coder symbol frequencies*}

$I \Leftarrow$ number of subimage columns in QUANT_COEFFS
$J \Leftarrow$ number of subimage rows in QUANT_COEFFS
FREQ $\Leftarrow$ clear to all zero
for $j = 1$ to J **do**
  for $i = 1$ to I step W **do**
    SCAN $\Leftarrow$ All_plane_scan(QUANT_COEFFS$_{i,j}$)
      {*Scan group starting at* (i, j) *using all-plane strategy*}
    SYM $\Leftarrow$ Map_to_coder_symbols(SCAN)
    for all symbols S in SYM **do**
      FREQ(S) $\Leftarrow$ FREQ(S) + 1
    end for
  end for
end for

Procedure 7.5 Calculate bit count for coded intensity distribution (per-plane scanning).

Arguments:
QUANT_COEFFS {*quantised transform coefficients*}
HUFF {*Huffman code size per symbol per* (u,v)*-plane*}
W {*subimage group length*}
Results:
BITS {*coded bit count for output of main Huffman coder*}

$I \Leftarrow$ number of subimage columns in QUANT_COEFFS
$J \Leftarrow$ number of subimage rows in QUANT_COEFFS
BITS $\Leftarrow 0$
for $j = 1$ to J **do**
  for $i = 1$ to I step W **do**
    for $w = 0$ to $W-1$ **do**
      ZZ$_w \Leftarrow$ Zigzag_scan(QUANT_COEFFS$_{i,j}(w)$)
      SYM$_w \Leftarrow$ Map_to_coder_symbols(ZZ$_w$)
      for all symbols S in SYM$_w$ **do**
        BITS $\Leftarrow$ BITS + (HUFF$_w(S)$ + Magnitude_category(COEFF_GENERATING_S))
      end for
    end for
  end for
end for

this operation. Where the all-plane improved scanning strategy is being used, this function is performed for the single sequence of symbols representing the entire group, by Procedure 7.6 ('Calc_bit_count_ap').

Finally the output bit rate for the coded intensity distribution and peak signal to noise ratio are calculated.

7.6 Global Huffman code generation

The definition of Huffman code tables for general application to future source intensity distributions follows the same procedure for the 3D-DCT scheme as for the hybrid DPCM/DCT schemes, described in chapter 6. When the per-plane zigzag scanning strategy is in use, a separate code table must be supplied for each (u,v)-plane in the transformed, quantised and scanned subimage group. Hence N global tables must be generated. When the all-plane improved scanning strategy is in operation, only one table is generated since the entire subimage group is scanned into a single sequence.

In generating a table, initially the scanned coefficient statistics are collected. The frequency

Procedure 7.6 Calculate bit count for coded intensity distribution (all-plane scanning).

Arguments:

QUANT_COEFFS	{*quantised transform coefficients*}
HUFF	{*Huffman code size per symbol*}
W	{*subimage group length*}

Results:

BITS	{*coded bit count for output of main Huffman coder*}

```
I ⇐ number of subimage columns in QUANT_COEFFS
J ⇐ number of subimage rows in QUANT_COEFFS
BITS ⇐ 0
for j = 1 to J do
  for i = 1 to I step W do
    SCAN ⇐ All_plane_scan(QUANT_COEFFS_{i,j})
    SYM ⇐ Map_to_coder_symbols(SCAN)
    for all symbols S in SYM do
      BITS ⇐ (HUFF(S) + Magnitude_category(COEFF_GENERATING_S))
    end for
  end for
end for
```

of each symbol is computed for several source intensity distributions using the routine appropriate for the scanning strategy in use ('Gen_sym_tables_pp' or 'Gen_sym_table_ap'). The corresponding frequencies produced from the source intensity distributions are then summed, unity added to each to ensure that codes are generated even for highly unlikely symbols, and the result input to the routine 'Gen_code_lengths' to produce the table of Huffman code word lengths. Luma tables were generated using the luma channels of the four intensity distributions *Horseman*, *Microscope*, *Tank* and *Toyplane*. Chrominance tables were produced using the chrominance-red and chrominance-blue channels of *Horseman*, *Tank* and *Toyplane*. The target bit rate was again initially set to 0.5 bits per pixel, as it was in the hybrid DPCM/DCT case.

7.7 Simulation Results

This section presents the results of simulations performed of the 3D-DCT based integral image coding scheme. The enhanced hybrid DPCM/DCT scheme with fixed quantisation is used as a baseline reference. As in previous simulations of hybrid DPCM/DCT schemes, the bit rate values generated represent only the raw coded intensity distribution data; no control data is included, as would be required by a practical application of the coding scheme.

The rate-distortion plots given reflect the performance of the scheme variants simulated when applied to the intensity distributions presented in appendix A. Mean curves are presented in each case, where again the performance data relating to each coded intensity distribution were remapped to a common independent variable using linear interpolation with an x-axis step of 0.01 bpp.

7.7.1 Coding of luma channel only

Initially simulations were carried out of the 3D-DCT schemes applied to the luma (Y') channel only of each source intensity distribution; the source bit rate is therefore 8 bits per pixel.

Comparison with the hybrid DPCM/DCT schemes

Figure 7.8 shows the difference in objective fidelity obtained between application of the 3D-DCT scheme using fixed quantisation and the best mode/quantisation array combination of the hybrid DPCM/DCT scheme across a range of output bit rates. Three curves are given for different values of transform group length N. The best hybrid DPCM/DCT scheme performance data were obtained as the maximum value of objective fidelity at each output bit rate in the simulation range selected from results of applying the base and enhanced hybrid schemes with fixed quantisation and the base scheme with LVV quantisation. The per-plane zigzag scanning method was employed in the 3D-DCT scheme simulations. Global Huffman code tables were used in all cases with a target bit rate of 0.5 bpp. A marked improvement in performance of the 3D-DCT scheme is observed as N is increased. This occurs due to in-

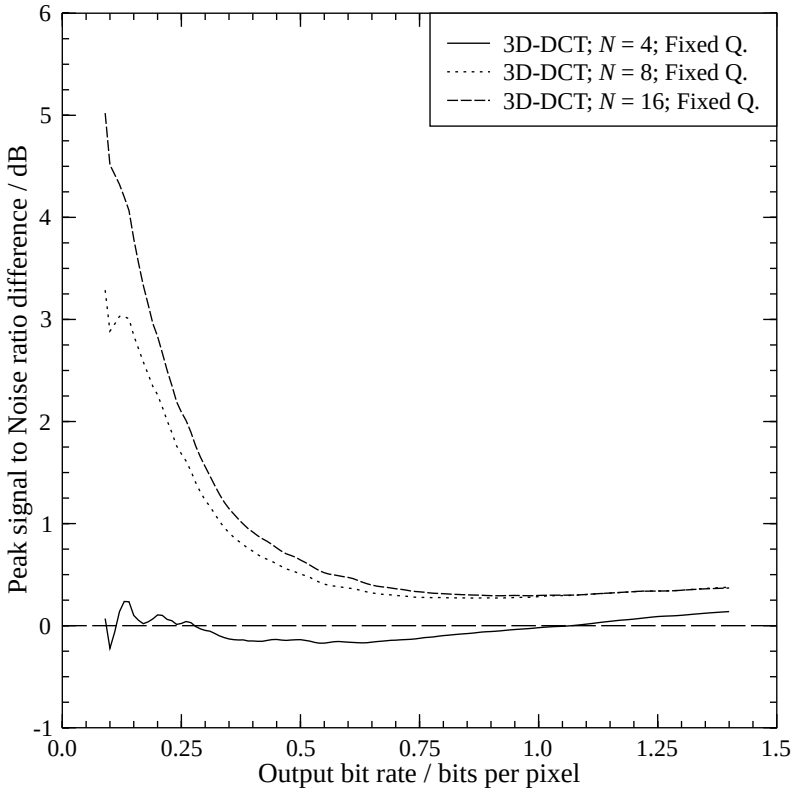


Figure 7.8: Difference in objective fidelity of the reconstructed intensity distribution between use of the 3D-DCT based scheme with fixed quantisation and the per-plane zigzag scanning strategy for several group lengths and the hybrid DPCM/DCT scheme on the luma channel only of the source data. Positive ordinate values constitute an improvement with use of the 3D-DCT scheme.

creasing overall effectiveness of decorrelation of all subimage groups in the source intensity distributions as more subimages are included in a single transform computation. Additionally the performance of the 3D-DCT scheme relative to the hybrid DPCM/DCT scheme can be seen to improve significantly as output bit rate is lowered, thus the chief effect of multiple inter-subimage decorrelation using the 3D-DCT is more even rate-distortion performance across moderate to low bit rates.

The 3D-DCT scheme can be seen to provide a performance improvement over the hybrid DPCM/DCT scheme for $N \geq 8$. For $N = 4$, a performance decrease is observed over most of the tested output bit rate range. This may be explained as follows. Below a certain value of N, in this case 4, the advantage due to the more effective DCT based inter-subimage decorrelation process in the 3D-DCT scheme is outweighed by the negative effect of subimage discretisation inherent in the method due to the processing of individual subimage groups. Despite its inter-subimage decorrelation method being less effective on a local level, the hybrid DPCM/DCT scheme has a global performance advantage in this case since its decorrelation method treats data continuously along a line of subimages. Due to considerations of implementing the real-time transform, it has been determined that N should be set no higher than 16. Simulations could not be performed for group lengths of 32 or greater, since the number of subimage columns in the test intensity distributions (I) are such that

$$I/N \neq \text{mod}(I/N) \qquad \text{for } N \geq 32.$$

Additionally it is predicted from the marginal performance improvement seen between $N = 8$ and $N = 16$ compared with that seen between $N = 4$ and $N = 8$, very little gain is likely for $N \geq 32$. In the remaining 3D-DCT scheme simulations described in this chapter, N is set to 8. If an extra low bit rate performance benefit is required by a specific application of the 3D-DCT based coding scheme, subject to transform implementation being feasible, group length can be set to 16.

Use of local Huffman code tables

Simulations were performed using the 3D-DCT scheme with local Huffman code tables. Figure 7.9 shows the improvement in objective fidelity due to use of local rather than global code tables, where group length $N = 8$. The global code tables used were generated for a target bit rate of 0.5 bpp. In calculating output bit rate in the local table cases, the constant overhead due to transmission of the code table has not been included. The performance

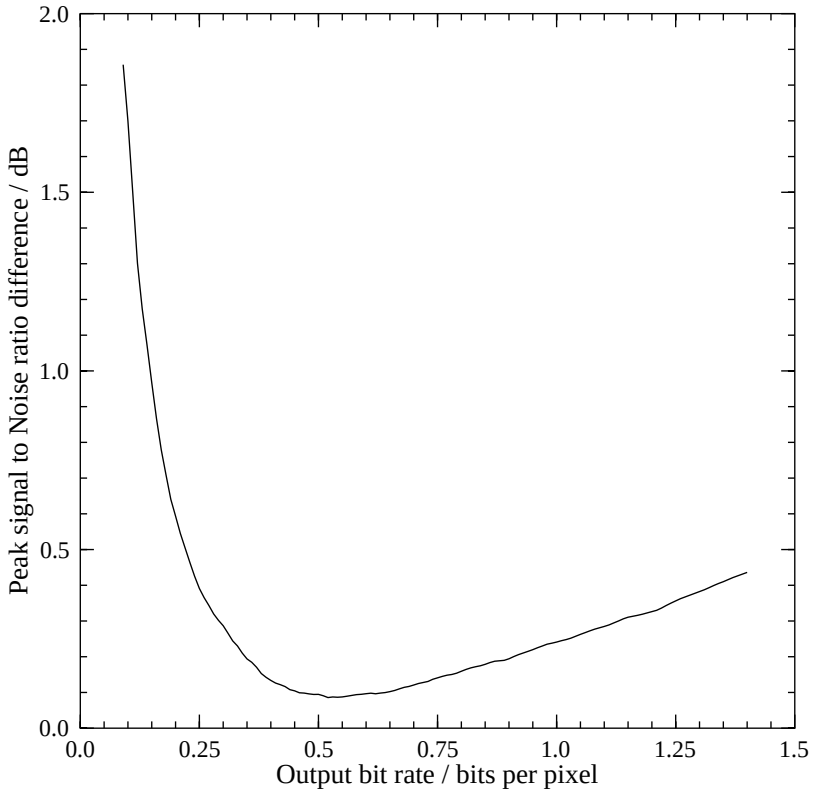


Figure 7.9: Improvement in performance of the 3D-DCT based scheme with fixed quantisation and the per-plane zigzag scanning strategy due to use of a local rather than predefined code table for a group length of 8 on the luma channel only of the source data.

improvement due to use of local code tables is of the same form as observed in the case of the hybrid DPCM/DCT scheme. Close to the code table target bit rate the difference is at a minimum; this is the point at which the encoder using global tables operates most efficiently. As output bit rate is increased above this value the local table improvement increases with a shallow gradient, but at bit rates below the target value a far sharper performance increase occurs.

The all-plane per-row improved scanning strategy

The per-plane zigzag scanning and coding strategy is now replaced with the all-plane strategy described in subsection 7.3.2. Figure 7.10 presents a differential rate-distortion performance curve for the 3D-DCT scheme when this improved scanning and coding strategy is used

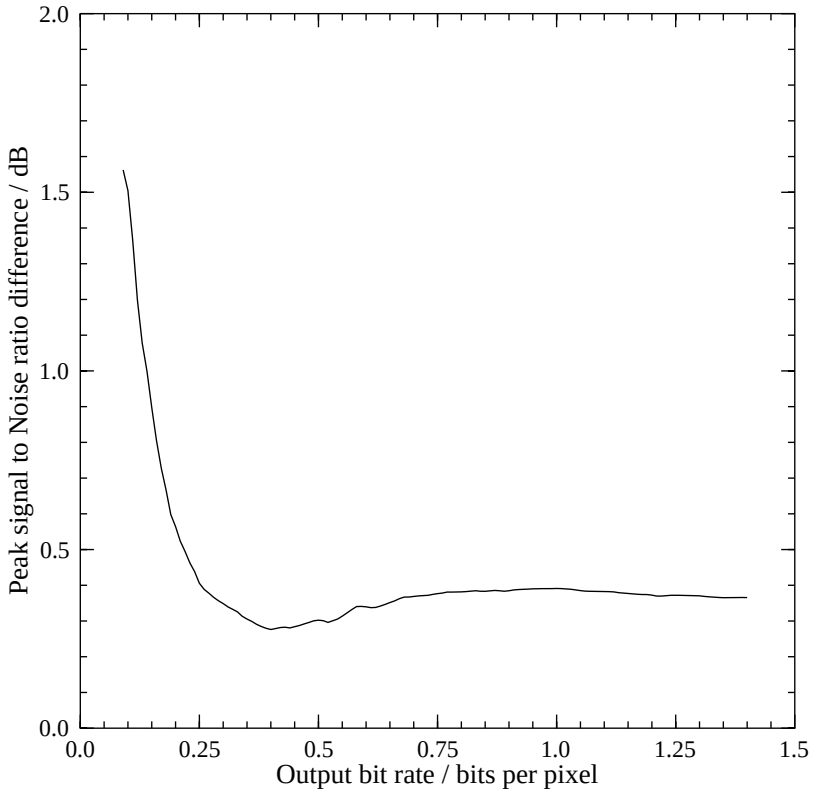


Figure 7.10: Improvement in objective fidelity due to use of the all-plane improved scanning strategy rather than the per-plane zigzag strategy in the 3D-DCT scheme with fixed quantisation for a group length of 8 on the luma channel only of the source data.

rather than the per-plane zigzag method. The transform group length $N = 8$ and the global code tables used were produced for a target bit rate value of 0.5 bpp. An improvement in performance is observed across the range of output bit rates tested. This is due to a decrease in bit rate for coding with a particular reconstructed intensity distribution fidelity, caused by the improved scanning strategy giving rise to a smaller number of Huffman coder input symbols in each coefficient group. The distortion value for a particular coding/decoding instance is unchanged since the quantisation parameters in each case are identical. The gain in reconstructed intensity distribution fidelity peaks at low bit rates but is between 0.27 and 0.4 dB at most output bit rate values (above 0.25 bpp).

A performance summary for the 3D-DCT scheme using fixed quantisation, operating on the luma channels of four lenticular-integral intensity distributions is given by Table 7.1. The table compares the performance of the 3D-DCT scheme with two baseline references simulated previously at three values of output bit rate.

	Peak signal to Noise ratio / dB		
Encoder	@ 1 bpp	@ 0.5 bpp	@ 0.25 bpp
Baseline sequential-DCT JPEG	—	27.70	22.12
Best hybrid DPCM/DCT; fixed quant.; global codes	40.91	36.46	31.77
3D-DCT; $N = 8$; fixed quant.; per-plane scan; global codes	41.14	37.09	32.71
3D-DCT; $N = 8$; fixed quant.; per-plane scan; local codes	41.38	37.18	33.11
3D-DCT; $N = 8$; fixed quant.; all-plane scan; global codes	41.53	37.39	33.12

Table 7.1: Summary of performance for luma channel-only 3D-DCT based encoding of four lenticular-integral intensity distributions.

7.7.2 Full colour coding

In this subsection results are presented from simulations of the 3D-DCT scheme when applied to full colour source data. As in full colour simulations of the hybrid DPCM/DCT schemes, the $R'G'B'$ source data with a bit rate of 24 bpp are converted to the $Y'C_bC_r$ colour model using $4:4:4$ sampling. Coding is performed using three independent 3D-DCT encoders operating on the Y', C_b and C_r channels of the source separately and the reconstructed intensity distribution remapped to $R'G'B'$-space for evaluation. Two separate quantisation arrays are used, applied to luma and chrominance-red/chrominance-blue data respectively. In this case the Y' and C_bC_r quantisation arrays are identical.

Comparison with the hybrid DPCM/DCT schemes

Figure 7.11 shows the improvement in objective fidelity achieved by the 3D-DCT scheme with $N = 8$ over the best case of the hybrid DPCM/DCT scheme for a given bit rate, for full colour coding. Fixed quantisation arrays, the per-plane zigzag scanning strategy and global code tables produced for a luma channel target bit rate of 0.5 bpp were employed in the simulations. A significant improvement is seen throughout the full bit rate range, with the benefit in objective fidelity increasing as output bit rate is decreased.

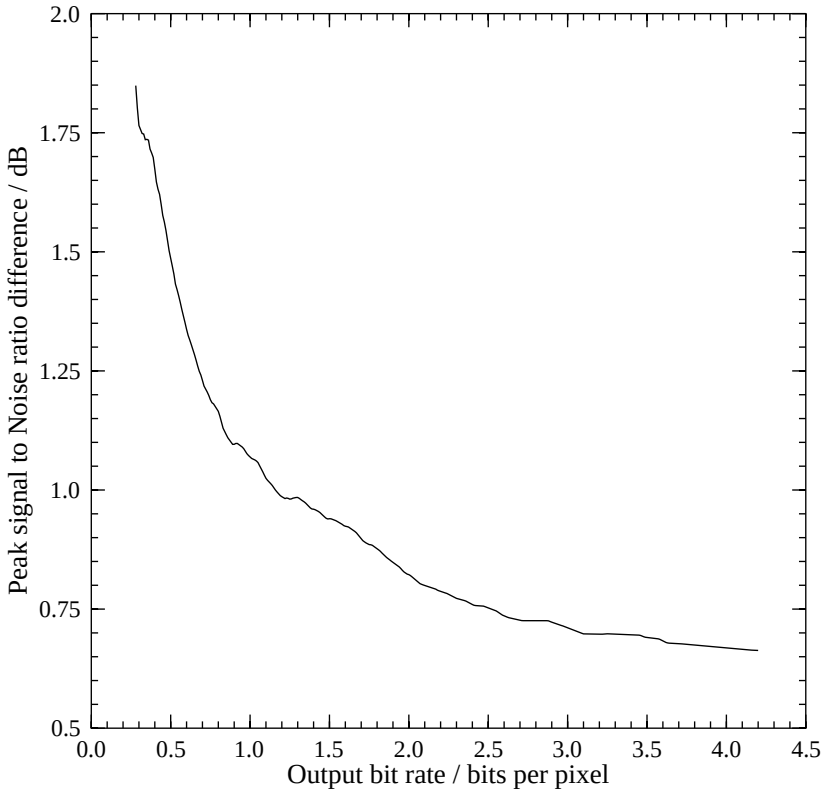


Figure 7.11: Improvement in objective fidelity due to use of the 3D-DCT based scheme with $N = 8$ and fixed quantisation, and the per-plane zigzag scanning strategy rather than the hybrid DPCM/DCT scheme on full colour source data.

Use of local Huffman code tables

Simulations of the 3D-DCT scheme were performed where fixed quantisation tables were used on full colour source data with the per-plane zigzag scanning method where both local and global Huffman code tables were employed. The improvement in rate-distortion performance resulting from use of local code tables may be seen as the difference in reconstructed objective fidelity for both cases, plotted in Figure 7.12. As in the luma only case, the constant overhead due to transmission of the code table in the local case has not been included in output bit rate calculations. As was found in similar tests of the hybrid DPCM/DCT scheme in chapter 6, the local-global difference has the same general form as in the luma channel only case but is smaller in magnitude and range across the bit rate range for which the scheme was tested. Again the minimum difference does not occur at the luma channel code table target bit rate due to greater correlation redundancy in the chrominance channel data than in the luma channel data. The difference varies from 0.14 to 0.58 dB, with a sharp perfor-

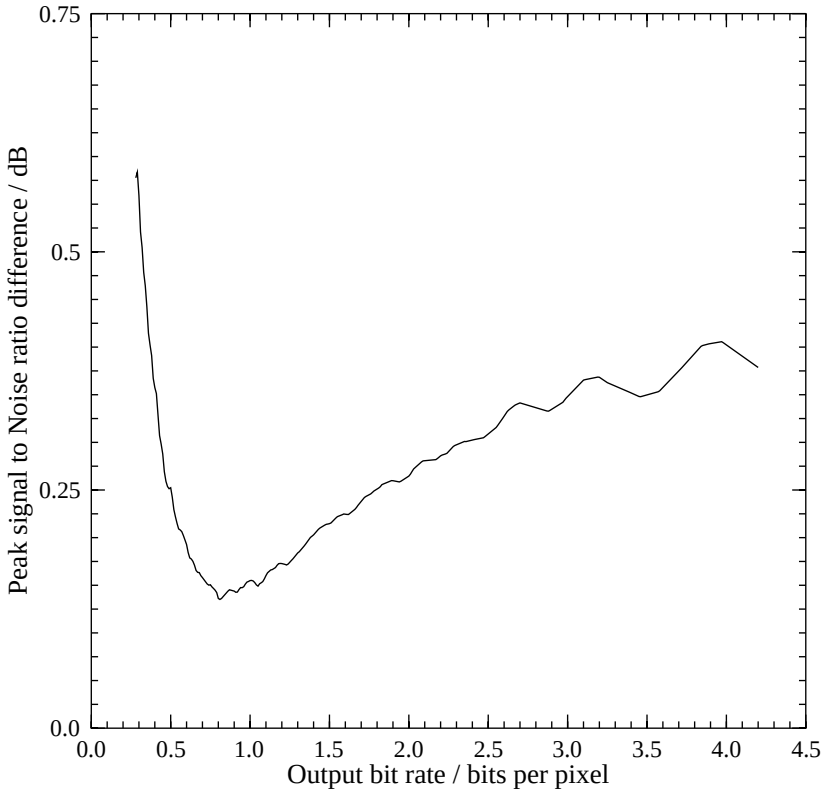


Figure 7.12: Improvement in performance of the 3D-DCT based scheme with fixed quantisation and the per-plane zigzag scanning strategy due to use of a local rather than predefined code table for a group length of 8 on full colour source data.

mance loss occurring at bit rates below approximately 0.75 bpp and a performance decrease with a smaller gradient associated with bit rates above this value. With the 3D-DCT scheme, the raw objective fidelity improvement due to use of local code tables may lead to a net performance benefit, considering the table transmission overhead, for certain applications.

The all-plane per-row improved scanning strategy

Figure 7.13 presents a differential rate-distortion plot illustrating the performance improvement achieved when the all-plane improved scanning strategy is used in place of the per-plane zigzag strategy for full colour source data. Transform group length $N = 8$; fixed quantisation arrays and global Huffman code tables constructed for a luma channel target bit rate of 0.5 bpp were employed. Use of the improved scanning strategy gives an improvement in rate-distortion performance throughout the range of output bit rate values for which the

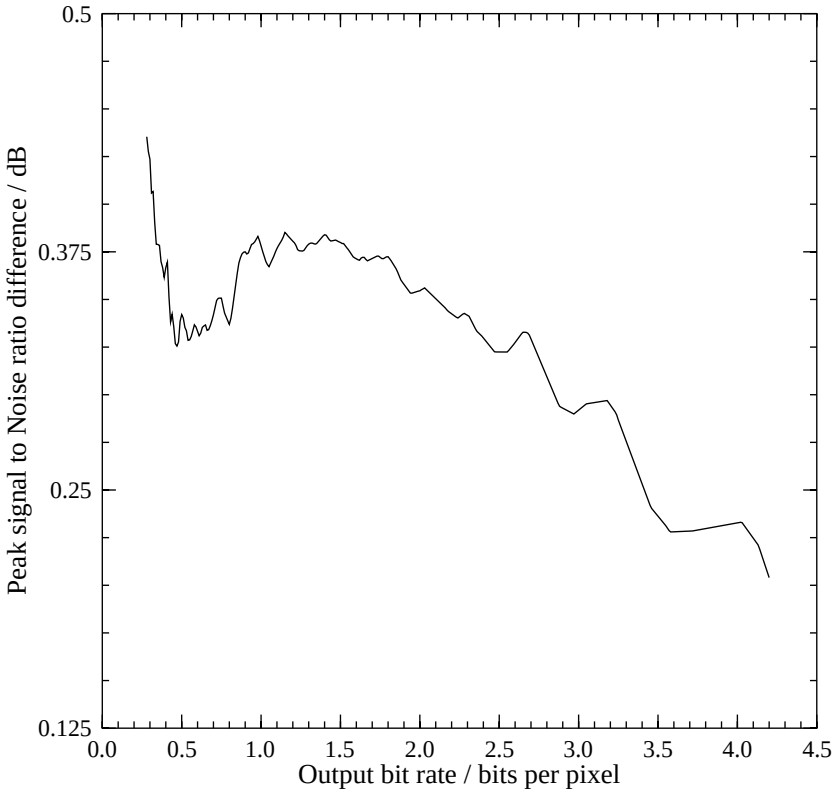


Figure 7.13: Improvement in objective fidelity due to use of the all-plane improved scanning strategy rather than the per-plane zigzag strategy in the 3D-DCT scheme with fixed quantisation for a group length of 8 on full colour source data.

scheme was simulated. The gain in fidelity of between 0.2 and 0.42 dB increases as bit rate is reduced from the maximum value tested (4.2 bpp) to approximately 1 bpp. A local minimum in gain occurs in the region of the global code table target bit rate value, below which gain sharply increases to its maximum value.

Table 7.2 gives a summary of the performance of the 3D-DCT scheme when used to encode three lenticular-integral intensity distributions at three output bit rates. Corresponding summaries for two previously-simulated baseline references are also provided.

7.7.3 Reconstructed intensity distributions

Reconstructed intensity distributions taken at the output of the decoder are now reproduced for the simulations described above. The examples result from full colour coding of the

	Peak signal to Noise ratio / dB		
Encoder	@ 2 bpp	@ 1 bpp	@ 0.4 bpp
Baseline sequential-DCT JPEG	—	27.96	19.86
Best hybrid DPCM/DCT; fixed quant.; global codes	36.66	33.12	27.34
3D-DCT; $N = 8$; fixed quant.; per-plane scan; global codes	37.49	34.19	29.02
3D-DCT; $N = 8$; fixed quant.; per-plane scan; local codes	37.75	34.35	29.37
3D-DCT; $N = 8$; fixed quant.; all-plane scan; global codes	37.84	34.57	29.38

Table 7.2: Summary of performance for full colour 3D-DCT based encoding of three lenticular-integral intensity distributions.

Horseman and *Toyplane* test intensity distributions using the 3D-DCT based scheme with fixed quantisation tables, a group length of 8, the all-plane improved scanning strategy and global code tables generated for a target bit rate of 0.5 bpp. In each display the source intensity distribution is given for comparison.

Images 7.1, 7.2 and 7.3 show the decoded reconstructed image *Horseman* after encoding to bit rates of 1.65 bpp, 0.75 bpp and 0.35 bpp respectively. No differences are visible between the full size reconstructed image coded to 1.65 bpp and the source image, either in intensity distribution form or when replayed as integral 3D images. For the images coded to 0.75 bpp and 0.35 bpp, progressive minor distortion may be seen in the full size intensity distribution in that the transform blocks are just visible around object/background boundaries such as at the top of the horse's head. When these are replayed as integral 3D images, however in the author's opinion there are no visible differences between them and the replayed source images. Figure 7.14 shows magnified sections of the source intensity distribution and the reconstructed intensity distributions of Images 7.2, 7.3 and 6.3 (*Horseman* coded to 0.37 bpp using the base hybrid DPCM/DCT scheme with LVV quantisation). A very small amount of 'blocking' distortion has occurred using the 3D-DCT scheme at 0.75 bpp, which is significantly more noticeable at 0.35 bpp. However, if a comparison is made between Figures 7.14c and 7.14d, significantly more objectionable blocking artifacts are noticeable, in addition to ringing effects, for the hybrid DPCM/DCT case than the 3D-DCT case, despite the slight advantage here of the hybrid DPCM/DCT scheme due to its higher bit rate.

Images 7.4, 7.5 and 7.6 present reconstructed intensity distributions where *Toyplane* has been used as source data. These examples have been coded to similar bit rates to those used to prepare the *Horseman* prints. Here significantly more blocking distortion is visible for each

case than at the near-corresponding bit rates for the *Horseman* image, though they again replay correctly as integral 3D images. The lower rate-distortion performance of the 3D-DCT scheme applied to this image is a result of the significantly higher contrast between many of the objects in this image than for the *Horseman*, *Microscope* or *Tank* images. Due to this the *Toyplane* intensity distribution should be viewed as a slightly unrealistic example and the performance of the coding schemes with which it has been tested as a worst case. In Figure 7.15, magnified sections of the source intensity distribution and the reconstructed intensity distributions of Images 7.5, 7.6 and 6.7 (*Toyplane* coded to 0.41 bpp using the base hybrid DPCM/DCT scheme with LVV quantisation) are presented. The blocking distortion identified can be seen clearly. It was not possible to encode *Toyplane* at approximately the same bit rates for the lowest bit rate 3D-DCT coded example (Figure 7.15c) and the hybrid DPCM/DCT coded example (Figure 7.15d). Despite its small bit rate advantage however, the blocking and ringing distortion observed on the reconstructed hybrid DPCM/DCT coded section is more extensive than for the lowest bit rate 3D-DCT coded section.



[Original]



Image 7.1: *Horseman* compressed using the 3D-DCT scheme with fixed quantisation; output bit rate is 1.65 bpp using the all-plane scanning strategy. [SCREEN B]



[Original]



Image 7.2: *Horseman* compressed using the 3D-DCT scheme with fixed quantisation; output bit rate is 0.75 bpp using the all-plane scanning strategy. [SCREEN B]



[Original]



Image 7.3: *Horseman* compressed using the 3D-DCT scheme with fixed quantisation; output bit rate is 0.35 bpp using the all-plane scanning strategy. [SCREEN B]

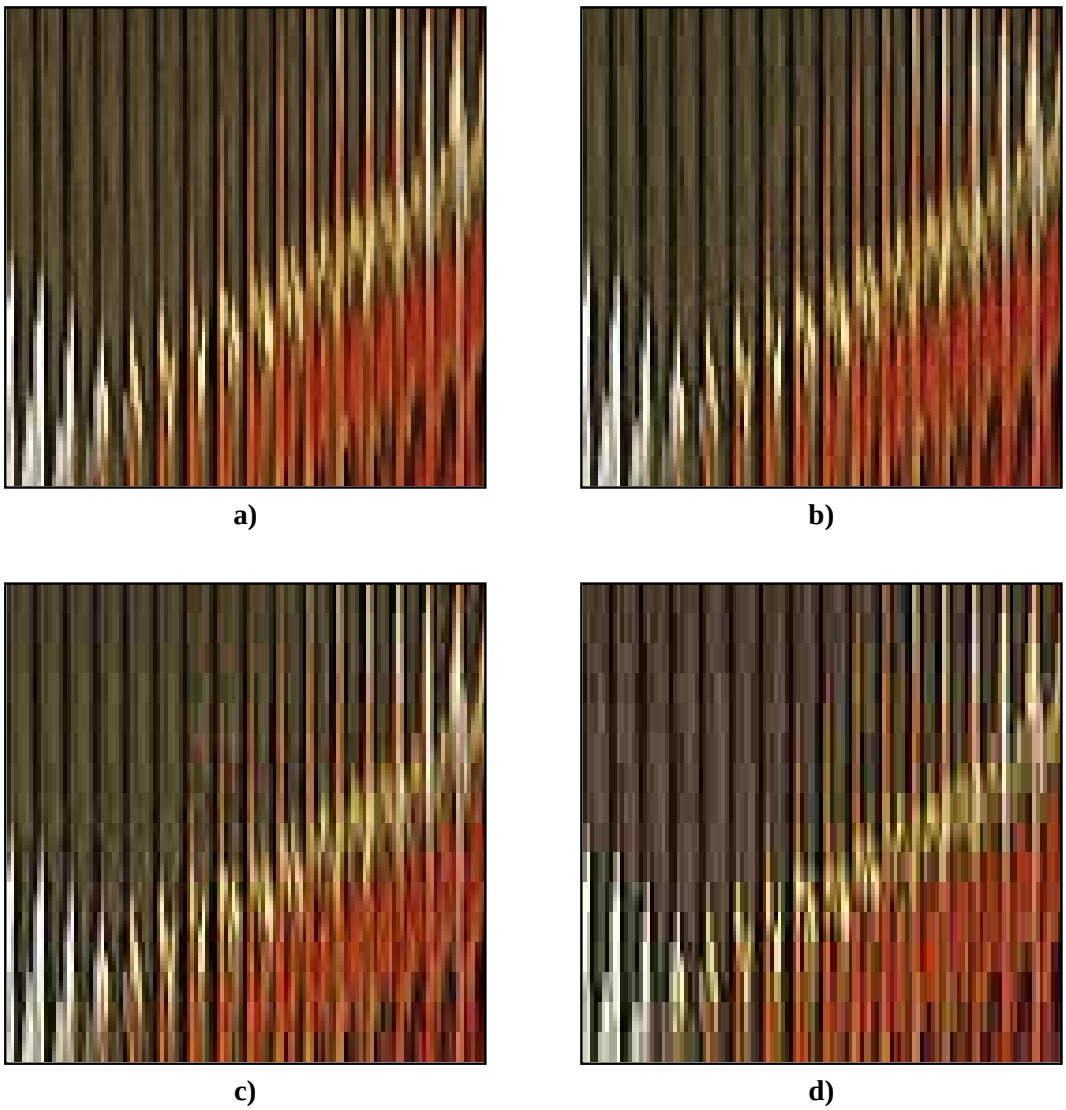


Figure 7.14: Magnified 16×16-subimage sections of *Horseman* resulting from coding using the 3D-DCT and base hybrid DPCM/DCT schemes.
a) Source.
b) Reconstructed, 3D-DCT with fixed quantisation at 0.75 bpp (all-plane scanning).
c) Reconstructed, 3D-DCT with fixed quantisation at 0.35 bpp (all-plane scanning).
d) Reconstructed, base hybrid DPCM/DCT with LVV quantisation at 0.37 bpp.



[Original]



Image 7.4: *Toyplane* compressed using the 3D-DCT scheme with fixed quantisation; output bit rate is 1.74 bpp using the all-plane scanning strategy. [SCREEN A]



[Original]



Image 7.5: *Toyplane* compressed using the 3D-DCT scheme with fixed quantisation; output bit rate is 0.76 bpp using the all-plane scanning strategy. [SCREEN A]



[Original]



Image 7.6: *Toyplane* compressed using the 3D-DCT scheme with fixed quantisation; output bit rate is 0.39 bpp using the all-plane scanning strategy. [SCREEN A]

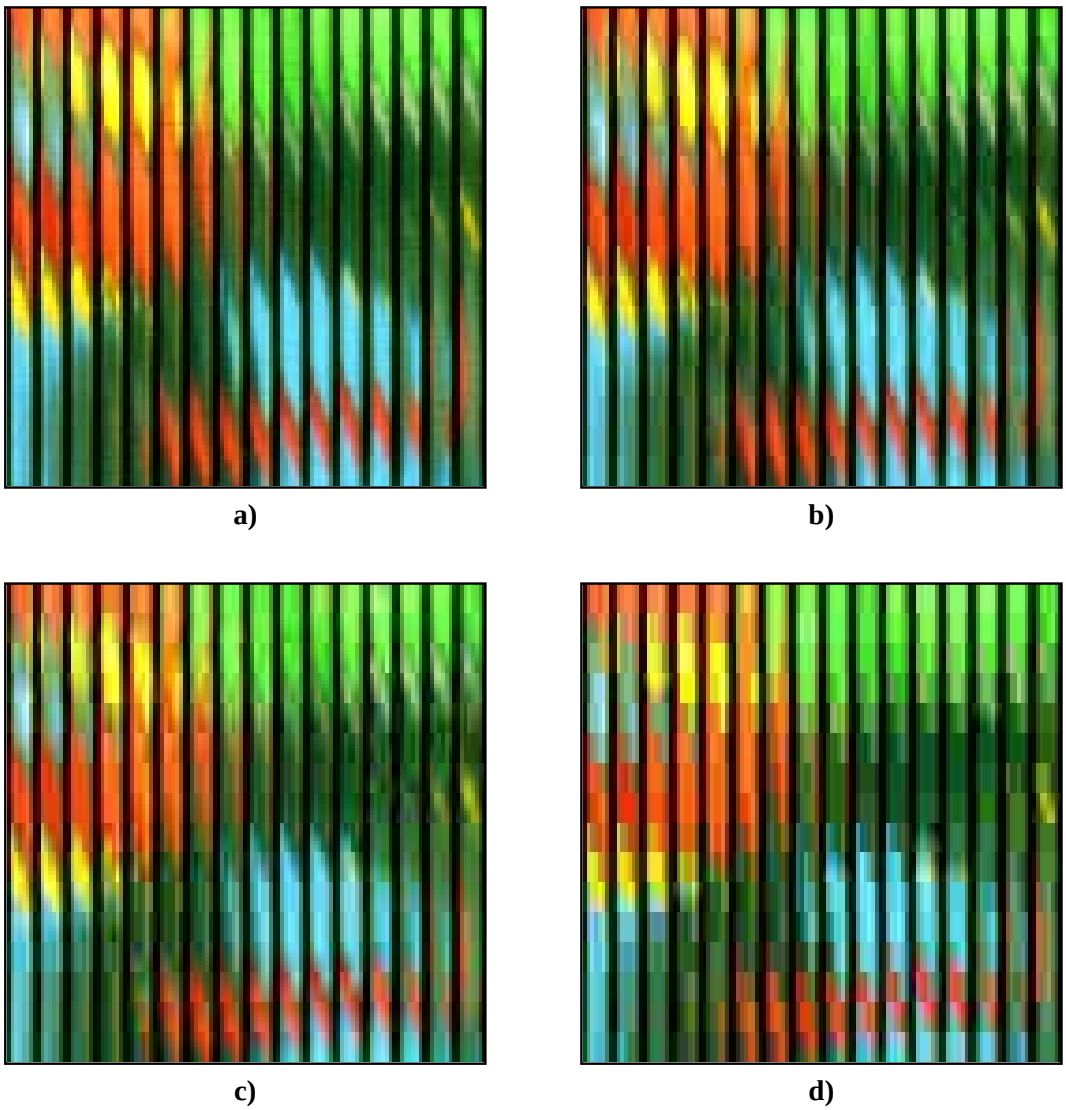


Figure 7.15: Magnified 16×16-subimage sections of *Toyplane* resulting from coding using the 3D-DCT and base hybrid DPCM/DCT schemes.
a) Source.
b) Reconstructed, 3D-DCT with fixed quantisation at 0.76 bpp (all-plane scanning).
c) Reconstructed, 3D-DCT with fixed quantisation at 0.39 bpp (all-plane scanning).
d) Reconstructed, base hybrid DPCM/DCT with LVV quantisation at 0.41 bpp.

Summary

In this chapter, the basis of a second lenticular-integral 3D image compression scheme has been described. The core of the scheme is a three dimensional DCT, which decorrelates a *group* of contiguous subimages from the source intensity distribution data in both inter-subimage and intra-subimage dimensions simultaneously. A quantisation stage processes each coefficient in the transformed group using an individual scalar quantiser with a given step size, and the resulting quantised group is scanned, mapped to coder symbols and Huffman entropy coded to produce the final raw output bit stream.

As in the hybrid DPCM/DCT schemes, prior to entropy coding each quantised coefficient array is 'scanned' to produce a linear sequence. In the 3D-DCT based scheme, initially this quantised coefficient ordering process was based on a 'zigzag' scanning of each inter-subimage spatially variant plane in the group, followed by mapping to coder symbols and coding of each sequence using an independent Huffman coder. A modified scanning and coding process was developed, in which a single scan is performed throughout the entire group. The scan is carried out in such a way as to produce a sequence with decreasing coefficient variance along the scan. The resulting single linear sequence is then mapped to symbols and entropy coded using a single Huffman stage. Use of this improved method was found to decrease bit rate for a given level of distortion in the reconstructed image, and also reduces transmission overhead in that only a single Huffman code table must be communicated to the decoder.

The 3D-DCT based scheme has been implemented for simulation using both of the above scanning/coding methods. Intensity distribution data from the four test integral 3D images reproduced in appendix A have been used as source data, though for the secondary simulations where the source data have been coded in full colour, three of the four test images have been used. The step sizes for each coefficient position in the quantisation process were identical for the simulations described in this chapter, that is fixed quantisation was used throughout. The 3D-DCT based scheme has been tested using both local (per-image) and global (predefined) Huffman code tables. The fidelity of reconstructed decoded images has been assessed objectively using the peak signal to noise ratio metric and subjectively both by examining unwanted coding artifacts and by assessing the quality of their replay as integral 3D images.

When global Huffman code tables were used, when coded using the 3D-DCT scheme with the per-plane zigzag scanning strategy and a transform group length of 8 planes, the mean im-

provement in peak signal to noise ratio of reconstructed full colour test lenticular-integral intensity distributions was found to be between 0.83 and 1.68 dB compared with the best result at each bit rate for the hybrid DPCM/DCT scheme. The greater gains were achieved at low bit rates. When the improved all-plane scanning strategy was used in the 3D-DCT scheme, the corresponding improvements compared with the best case for the hybrid DPCM/DCT scheme were between 1.18 and 2.14 dB with further low bit rate gain. A rate-distortion gain was noted due to use of the improved scanning strategy at all values of output bit rate tested. It is therefore concluded that it leads to better rate-distortion performance than use of the per-plane zigzag method in lenticular-integral image coding and should be chosen as the scanning strategy in future use of the 3D-DCT scheme. As was observed for the hybrid DPCM/DCT schemes, the rate-distortion performance of the scheme using global code tables has a maximum value within a small distance of the bit rate for which the code tables were produced. It is hence important that code tables are generated for a particular application which requires operation in a given output bit rate range. When local rather than global Huffman code tables are used, an improvement in rate-distortion performance is achieved for all bit rate values tested and the variability of performance with bit rate seen when global tables are used does not occur. However as with all cases of local code table use, the benefit in performance obtained must be compared with the loss expected as a result of the overhead caused by transmission of the tables to the decoder. The hybrid approach outlined in the summary for chapter 6 is likely to be most suitable in a motion video coding application.

A subjective assessment of replayed decoded integral 3D images coded to various bit rates confirms the observation made for use of the hybrid DPCM/DCT schemes, in that three dimensional replay is highly tolerant of blocking distortion introduced as a result of excessively coarse quantisation. The effects noted in replay were again an overall loss of detail and inaccurate colour reproduction, though in the author's opinion these effects occurred at lower bit rates than where the hybrid DPCM/DCT schemes were used. For coding to bit rates above approximately 0.8 bpp (compression ratios below 30 : 1), the reconstructed data replayed as an integral 3D image was found to be indistinguishable from the source image. The blocking and ringing artifacts noted in the raw reconstructed intensity distribution were found to be significantly less objectionable than where the hybrid DPCM/DCT scheme was used to code at the same output bit rate.

As previously, the target compression ratios for transmission are 6.9 : 1 and 36.8 : 1 for standard and high definition resolution respectively, for a source bit rate of 24 bpp. For the standard definition case, the 3D-DCT scheme reconstructs lenticular-integral image data without noticeable distortion. Due to the improved rate-distortion performance of the 3D-DCT

scheme, up to three 3D television pictures could be transmitted at standard definition resolution with even higher quality reconstruction than achieved by the hybrid DPCM/DCT scheme, or up to four pictures could be transmitted with similar quality to that achieved previously. For high definition resolution in the same capacity channel, a bit rate of 0.652 bpp is required. The 3D-DCT scheme, with global code tables optimised for this bit rate value, is capable of compressing the source to this bit rate, therefore one high definition resolution picture could be transmitted. Use of the scheme in a motion video environment is projected to give rise to reconstructed 3D frames indistinguishable from the original source frames, and the possibility of transmission of a greater number of video streams per channel. The further bit rate reduction would be possible since the encoder would take account of the temporal domain (inter-frame) pixel correlation, which is expected to be high.

The parameters of the 3D-DCT scheme which have a fundamental effect on rate-distortion performance are the quantiser step sizes used, as it is at this stage that distortion is introduced. In this chapter the fixed quantisation table has been used exclusively to provide baseline information on the performance of the scheme. Chapter 8 presents investigations into the design of these quantisation parameters to provide optimal and near-optimal rate-distortion performance while allowing a more accurate selection of output bit rate and distortion than allowed by a simple linear scaling of a given quantisation array.

8. EFFICIENT QUANTISATION IN THE 3D-DCT-BASED COMPRESSION SCHEME

Introduction

Lossy coding of a source inevitably involves a compromise between the level of distortion tolerable in the decoded information and the bit rate reduction potential available. The bit rate reduction achieved by a given binary compaction code is directly related to the coarseness of quantisation via the reduction in entropy provided by the quantisation process, while simultaneously distortion worsens due to increasing misrepresentation of the source by the quantiser outputs. The design goal is the minimisation of the rate-distortion criterion; appropriate selection of quantisation parameters is the prime consideration.

The 3D-DCT based integral 3D image coding scheme described in chapter 7 has initially used fixed quantisation arrays due to there being no psychovisual precedent for use of arrays with different step sizes per coefficient. This chapter describes work carried out to characterise a likely general form for integral 3D-DCT quantisation arrays to further improve the rate-distortion performance of the scheme, by statistical analysis of 3D-DCT coefficients and image-orientated optimisation using the technique of simulated annealing. The optimisation technique described also allows more precise selection of output bit rate than allowed by simple linear scaling of the quantisation array.

8.1 Statistical Analysis of 3D-DCT Coefficients

In quantisation table design for the hybrid DPCM/DCT schemes, the variability of decorrelated subimage coefficients was studied to determine the significance of each coefficient position. This was carried out by computing the standard deviation from the mean of each

decorrelated coefficient across several sample intensity distributions. A similar experiment is now described for the case of 3D-DCT decorrelation. The standard deviation for the decorrelated coefficient position (u,v,w) is $\sigma_{u,v,w}$, evaluated as the square root of the variance determined by equation 7.5.

The standard deviation per coefficient position was computed for the decorrelated coefficient groups of the luma channels of the intensity distributions *Microscope*, *Tank*, *Horseman* and *Toyplane* and the chrominance-blue/chrominance-red channels of *Tank*, *Horseman* and *Toyplane* with group length $N = 4$. Figure 8.1 gives the resulting plots of $\sigma_{u,v,w}$ against coefficient position in each plane, for all planes $w = 0 \ldots 3$.

The variability of coefficients in the horizontal spatial frequency dimension is higher than that of corresponding coefficients in the vertical dimension for each (u,v)-plane in the transformed coefficient group. This result is to be expected, considering the observations from the similar study for hybrid DPCM/DCT decorrelated subimage data in chapter 6. Following that experiment it was concluded that within each subimage vertically-running features are less significant than horizontally-running features. Again the form of the plots for luma and combined chrominance channels are very similar with the standard deviation values for chrominance being between 1.5 and 5 times lower than for the luma case.

Two observations are made regarding the third transform dimension (w). Firstly, the variability and hence spatial significance of the DC coefficient in each plane diminishes relative to the surrounding AC coefficients as w is increased from 0 to 3. Also the variabilities of all low vertical index coefficients ($v = 1 \ldots 3$) both reduce and become more even relative to each other as w is increased.

8.2 Use of Optimisation Techniques in Quantisation Array Design

The short analysis described above gives qualitative information on the likely general form of a suitable quantisation array, but the results can not be used directly to determine practical quantiser step size parameters. For this purpose the derivation of quantisation parameters which lead to optimal rate-distortion performance has been carried out using a numerical optimisation technique. The resulting optimal quantisation parameters have been used to derive a general form of quantisation array suitable for use with the 3D-DCT coding scheme applied to lenticular-integral intensity distributions, which leads to improved rate-distortion performance over use of linearly scaled fixed quantisation arrays. In all cases the all-plane,

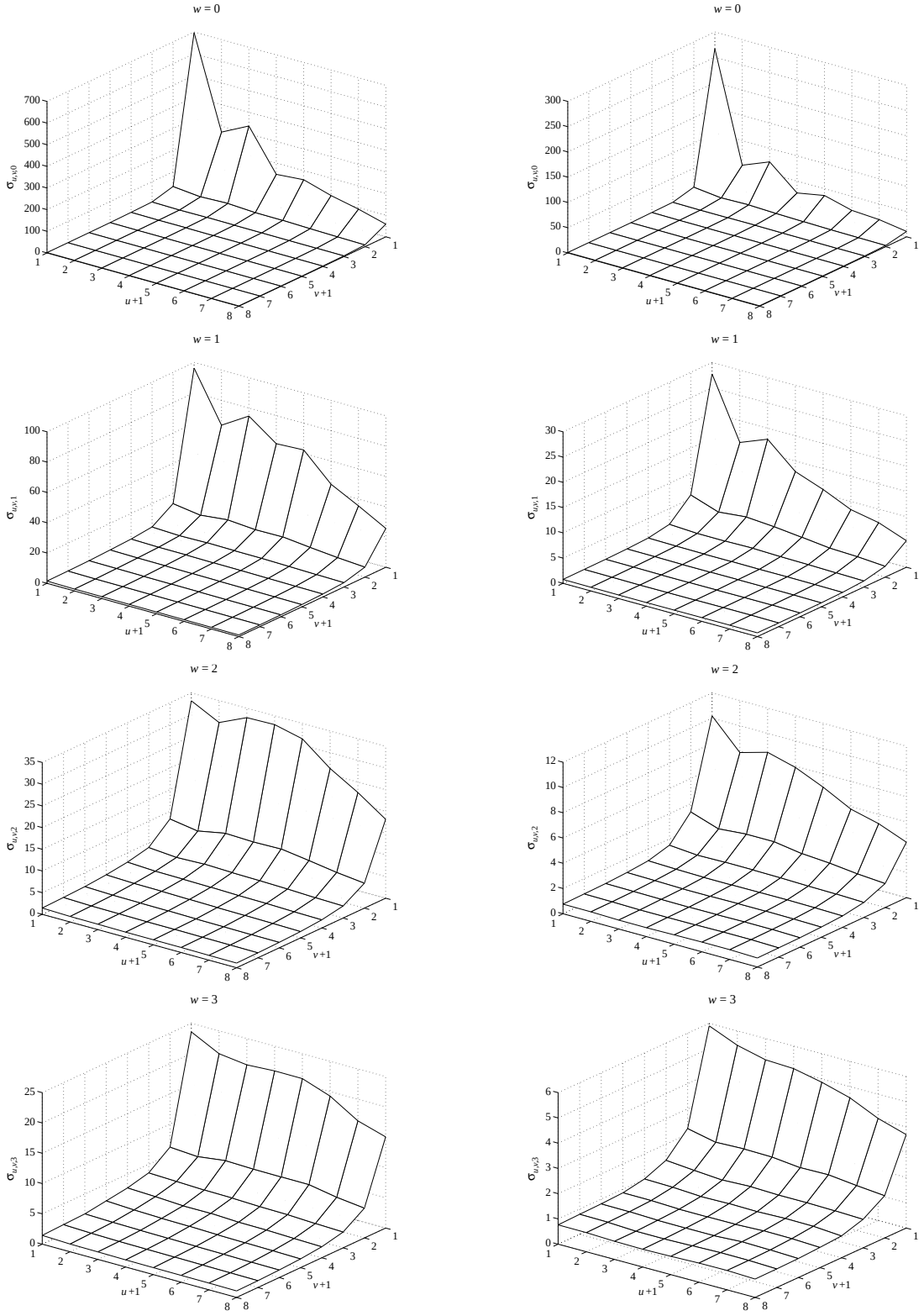


Figure 8.1: Standard deviation from the mean of 3D-DCT decorrelated subimage group coefficients for a group length of 4. a) Luma channel. b) Combined chrominance channels.

per-row coefficient scanning method has been used in the 3D-DCT scheme. The luma channel Huffman code table used was that derived for a target bit rate of 0.25 bpp, since this had been found to give rise to better low bit rate performance than where the target was 0.5 bpp when compared with the optimum (local) code table case. The optimisation problem is formulated in broad terms as depicted by Figure 8.2.

Generally, an iteration proceeds as follows. The optimisation subsystem generates a set of quantiser parameters (the quantisation array) which is fed to the 3D-DCT coder, operating on a specified source intensity distribution. The bit rate resulting from coding and the objective fidelity of the decoded intensity distribution are returned to the optimisation subsystem, which then makes a decision by some defined criterion on whether the bit rate-distortion pair constitute a 'better' solution and hence quantisation array than the previous 'best' solution. The meanings of 'better' and 'best' in this context must be defined. This process continues until a certain predefined number of iterations have taken place. The number of iterations before termination is determined by trial runs of the system as that required to produce bit rate and fidelity values which satisfy some condition.

Research has previously been carried out on producing rate-distortion optimal quantisation tables for use with the baseline sequential-DCT JPEG conventional still image coding scheme [56]. The method of *simulated annealing* was used as the optimisation subsystem in this work. This technique is suitable for optimisation of the parameters of complex nonlinear systems in which the output values used to judge the quality of a solution typically possess many local extrema. Simulated annealing is described in further detail below. The system

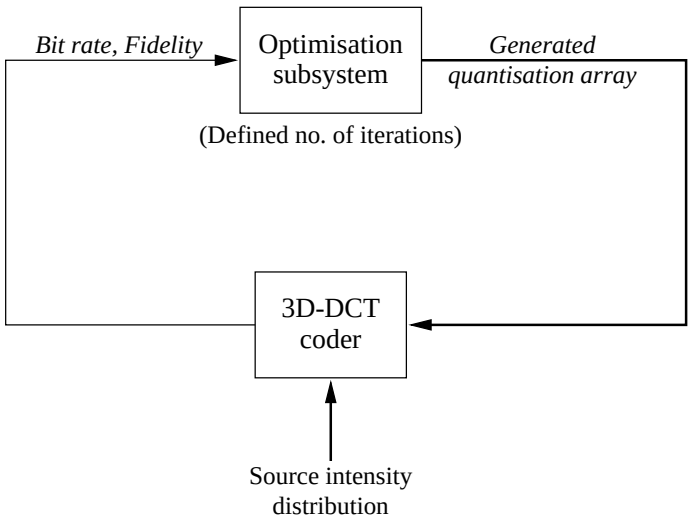


Figure 8.2: Derivation of rate-distortion optimal quantisation arrays.

was initially used to seek a rate-distortion optimal quantisation array directly, with the optimisation subsystem modifying and providing each of the 64 quantiser step sizes required individually to the JPEG coder model. Empirical models were then fitted which were capable of generating the optimal state found by direct optimisation. These reduced parameter models were then employed in further optimisation runs, leading to a decrease in computational effort required to obtain the optimal state. In order for the optimisation subsystem to determine the quality of a solution, a composite function was defined whose global minimum was to be sought. For the evaluation of a particular generated quantisation table, the goal was minimisation of the objective function:

$$F = E + \lambda(C - C_0)^2, \tag{8.1}$$

where E is the mean square error in the reconstructed decoded image, C is the compression ratio obtained, and C_0 is a predefined desired compression ratio. When this function is minimised, the overall effect is the minimisation of distortion while keeping the compression ratio obtained close to the desired compression ratio. The multiplier λ was set by experiment to produce an optimal state with C as close as possible to C_0. A value for λ of around 1 was reported to give the best results [56].

8.2.1 Simulated annealing

Many optimisation problems of practical interest involve finding values for the parameters of complex systems which can not be adequately represented by straightforward linear or quasi-linear models. Typically when such problems must be solved it is found that the objective function possesses complex terrain with many local extrema which may be at entirely different scales than the desired global extremum. In such cases the use of *local search* optimisation algorithms such as the simplex method [53, 57] is highly unlikely to produce the global extremum, but rather a local extremum which is in the neighbourhood of the starting point. This problem is illustrated by Figure 8.3 which represents a hypothetical univariate cost function with two local minima and one global minimum. If a local search technique were used, and its randomly chosen initial solution was at point 'A', it would descend toward the local minimum directly to its left, the true global minimum remaining undiscovered. It has been found that the problem of deriving optimal quantisation arrays for image coding schemes is of this class of 'hard' problems, and although some success has been obtained using local search methods for their solution [58], the use of global search methods which avoid termination at local minima has produced superior results.

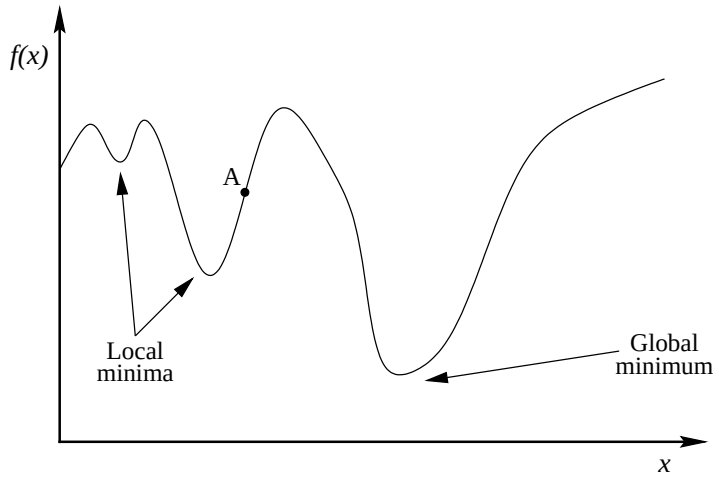


Figure 8.3: Global and local minima in a hypothetical cost function.

Simulated annealing (SA) is a global search method which makes use of stochastic variation to avoid terminating in local extrema [59], and has proved extremely useful in locating the global extrema of objective, or *cost* functions derived from complex nonlinear systems. The technique was first used by its developers to obtain optimum layouts of electronic circuit elements on silicon substrates in very large scale integrated circuit design [60]. It has since been applied to many diverse problems in a wide range of disciplines, including the derivation of optimal quantisation parameters for JPEG coding outlined previously. The method carries a statistical guarantee that the global minimum for a given problem will be found [61], albeit at considerable computational expense. An outline of the basis of the technique will now be given.

The SA algorithm is based on an analogy with metallurgical annealing, a process whose objective is to cause a solid to reach a minimal energy state with a highly regular crystalline structure. This is achieved through *slow* cooling to allow ample time for the redistribution of the particles of the sample. Faster cooling (*quenching*) typically results in a suboptimal configuration with imperfections in the crystal lattice and a *local* energy minimum. The first step towards the development of general purpose SA optimisation algorithms was the simulation of the evolution of states a metal undergoes as it reaches thermal equilibrium, at constant temperature (the Metropolis algorithm) [62]. This algorithm was then extended to include the effect of reducing temperature, and generalised to non-specific optimisation problems [63]. Basically, the analogy is formed as follows:

1. A method for generating random changes in the state space is defined;

2. The objective (cost) function of the problem is equivalent to energy in the mechanical process;

3. A control parameter is defined ('temperature') which is lowered in accordance with an *annealing schedule*;

4. An *acceptance distribution* is defined, based on a suitable physical law which determines, at a given 'temperature', whether a given generated state which does not lead to a lower cost is to be accepted.

Solutions which evaluate to a cost lower than the previous best are always accepted. However, the key to the capability of the algorithm to escape from local minima is the inclusion of the acceptance test; this allows acceptance of a solution with a higher cost than the previous lowest with a probability which reduces with temperature.

Practical SA implementations contain various extensions to the procedure outlined to allow faster convergence to the global minimum sought, and improved flexibility by providing for 'tuning' of the algorithm for application to a wide variety of problems. The work described in this chapter has been carried out using a system called 'Adaptive Simulated Annealing' (ASA) [61, 64, 65]. ASA can work with systems with an arbitrary number of arbitrarily constrained parameters and provides several very useful enhancements over the basic SA algorithm which render it of practical utility for a wide range of problems. It allows extensive tuning of many of its internal parameters, generally required when approaching a complex problem about which little is known. It is significantly faster than the basic SA algorithm due to its use of a modified annealing schedule, though this does not violate the statistical guarantee of finding the global minimum given suitable ASA internal parameter choices. Additionally, *reannealing* is performed at certain stages in the search to take into account the different and changing sensitivities of the cost function value to each parameter of the problem throughout the search.

8.2.2 Problem formulation: The parameter space

The goal of the optimisation problem is to find a quantisation array the use of which will lead to optimal rate-distortion performance when the 3D-DCT based coding scheme is applied to a typical lenticular-integral intensity distribution. Trivially therefore, the parameter space is a $64N$-dimensional array of quantiser step sizes where the coder is operating with a transform group length N. For most of the simulation results presented in chapter 7, a group of length 8

was employed as a compromise between overall DCT decorrelation performance and ease of implementation of the transform. For the simulations carried out by the optimisation process described in this chapter a group length of 8 is again used, and hence 512 parameters are required to represent the quantisation array. Using such a large parameter space with necessarily wide bounds, it was found that the results were highly sensitive to appropriate tuning of internal ASA parameter values, and the search was prohibitively slow. A generating function was therefore used to produce the quantisation array indirectly. A function constructed from a third degree polynomial for each spatial frequency dimension was trialled initially. This was used on the assumption that a section through a suitable quantisation array in each spatial frequency dimension would be monotonic, but imposing no specific profile on any of the three dimensions. A quantisation array generated in this way is:

$$\begin{aligned} Q(r-1,s-1,t-1) = 1 + |\lfloor & a_1 r^3 + a_2 r^2 + a_3 r \\ & + b_1 s^3 + b_2 s^2 + b_3 s \\ & + c_1 t^3 + c_2 t^2 + c_3 t + d \rfloor | \end{aligned} \tag{8.2}$$

$$\text{for } r,s,t = 1 \ldots 8.$$

The parameter space for the optimisation problem now consists of the ten real coefficients of Equation 8.2. The floor ($\lfloor \cdot \rfloor$) and absolute value operations and the unary addition are present to ensure that $Q \geq 1$. The spatial frequency variables used were in the range $1 \ldots 8$ as use of the range $0 \ldots 7$ would cause a result of zero for the first point in each dimension, thus giving a less flexible generating function.

In order that no unfair bias is placed on the structure of an optimal quantisation array generated using Equation 8.2, the bounds on the parameters must be relatively wide. This presents a larger solution space to the SA algorithm than is desirable, tending to increase the time required to find the global optimum. To reduce the scale of the problem, it was felt necessary to develop a generating function which presents a smaller solution space, achievable by reducing the number of parameters or their allowable ranges, or both. In research previously carried out to obtain optimal quantisation arrays for the JPEG conventional image coding scheme [56], a three parameter model of the following form was employed:

$$Q(u,v) = a + bz^c \tag{8.3}$$

where a, b and c are the model parameters to be found and $z = u + v$ is the Manhattan distance of a coefficient at position (u,v) from $(0,0)$ with $u,v = 0 \ldots 7$. It had been found in testing Equation 8.2 that generally one polynomial term in each spatial frequency dimension

was dominant in each case, and therefore a function similar to Equation 8.3 with a single variable power term would be sufficient to generate an efficient quantisation array. The final generating function is therefore:

$$Q(r-1,s-1,t-1) = 1 + |\lfloor a_1 r^{b_1} + a_2 s^{b_2} + a_3 t^{b_3} + c \rfloor| \quad \text{for } r,s,t = 1 \ldots 8. \tag{8.4}$$

The parameter space now consists of only the seven parameters of this function. Table 8.1 lists the initial values of the parameters and their allowable ranges. The initial parameter values were set in order that the previously known best quantisation strategy, the fixed quantisation array, was generated, where q_0 is the quantiser scale which produces the array leading to an output bit rate which is nearest to the desired bit rate for which the optimisation process is to be run.

The extents of the parameter ranges were set such that the generating function was capable of producing quantisation arrays with suitably widely varying step sizes. The constant term, c, was constrained within a range slightly wider than q_0.

8.2.3 Problem formulation: The cost function

The cost function defines the overall objective of the problem and must produce a single value which is to be minimised by the SA algorithm. This value must represent accurately the quality of a solution such that the SA algorithm will tend to converge on the global optimum solution. The problem is the determination of a set of parameters for Equation 8.4 which cause it to generate a rate-distortion optimal quantisation array. The cost of the solution is therefore derived from the bit rate and distortion values which a given set of parameters produced. In practical terms, the problem requires one of the two output objective values, either bit rate or distortion, to be fixed and the other allowed to move towards its optimum.

Parameter	a_1	b_1	a_2	b_2	a_3	b_3	c
Initial value	0	0	0	0	0	0	(q_0-1)
Range	0 to 5	0 to 5	0 to 5	0 to 5	0 to 5	0 to 5	0 to $(c+5)$

Table 8.1: Constraints and initial conditions for the quantisation array generating function.

In this instance the output bit rate was fixed.

The design of an appropriate cost function is largely concerned with deciding on the manner in which the two objectives of the problem are combined to produce a single cost value in such a way that both are satisfied. It was found that use of an additive cost function similar to Equation 8.1 tended to allow too much individual variation of the two terms representing distortion and compression ratio error, resulting in a solution where at termination C was some considerable distance from C_0; the error was typically between $1:1$ and $3:1$. A cost function based on the quotient of the two objectives was found to avoid this problem, terminating with bit rate far closer to the desired value. The objective is the maximisation of the function

$$H = \frac{P}{1 + |R - R_0|}, \tag{8.5}$$

where P is the peak signal to noise ratio measuring the objective fidelity of the reconstructed intensity distribution, R is the actual output bit rate and R_0 is the desired bit rate in bits per pixel. To transform the problem to one of minimisation as required by the SA algorithm, the sign is reversed such that the cost function,

$$F = -H. \tag{8.6}$$

Use of this cost function generally resulted in an optimum with an extremely small bit rate error, typically below 0.0001 bpp. Unfortunately, however, when solving for low bit rates ($R_0 < 0.25$ bpp) the bit rate error on termination was significantly greater, of the order of 0.05 bpp. The cause of this was determined to be the large value of the derivative of the P–R relationship at low bit rates, as may be observed in the result plots from the baseline simulations performed of the coding scheme given in Chapter 7. This leads to a situation in this bit rate range in which the solution cost is lower for some value of R greater than R_0 than for $R = R_0$. The latter case should always evaluate to the lowest cost to ensure that the final bit rate error is minimised. It was noted that ASA allows a validity flag to be set on cost function exit; if this flag has the value zero, that particular solution is discarded and is not subject to the ASA acceptance test. To alleviate the cost function problem, the validity flag was set to zero if the bit rate R resulting from a given solution was greater than R_0, thus limiting the search to bit rate values $R \leq R_0$. The cost function may then be simplified slightly to:

$$F = -\frac{P}{1 + R_0 - R}. \tag{8.7}$$

Additionally due to the imposed limit on bit rate values, the initial values of the coefficients of Equation 8.4, the parameters for the problem, must produce a fixed quantisation array which leads to a bit rate nearest to but less than R_0.

It was found in testing of Equation 8.7 that it was often the case that a new solution with a value of P higher than that associated with the best solution obtained at a given point in the search would not be accepted as the new best solution due to a significant deviation of R from R_0. This imposes an unfair limitation on the solution space. The general P-R trend is that of a monotonically increasing P with increasing R, as evidenced by the performance plots presented for the coding scheme in chapter 7. Taking this into account, and noting that all bit rate deviations in the cost function are now negative since any solutions with $R > R_0$ are rejected, it can be seen that any value of P greater than the previous best value represents a better solution. A significant simplification of the cost function results from this observation, such that the problem now has a single objective, that is

$$F = -P. \tag{8.8}$$

Due to the general form of the P-R relationship, a successful optimisation run is expected to lead to a final bit rate extremely close to the desired bit rate.

8.2.4 The optimisation procedure

The overall optimisation problem is depicted in Figure 8.4. For the evaluation of a specific set of generated parameters, a 32×32-subimage (256×256-pixel) section of the luma channel of the intensity distribution *Toyplane* was input as the source data to the coder. This section, shown in Figure 8.5, was chosen to be representative of the general structure observed in the luma channels of the full range of intensity distributions examined. The sample, taken from the central area of the full intensity distribution, is composed mainly of information representing a high detail area together with a smaller portion with no detail.

8.2.5 Single-point optimisation example

This subsection describes the application of the above procedure for a sample required output bit rate. For assessment of the performance benefit from use of an optimal quantisation array, a comparison is made with the results from applying a fixed quantisation array. For this pur-

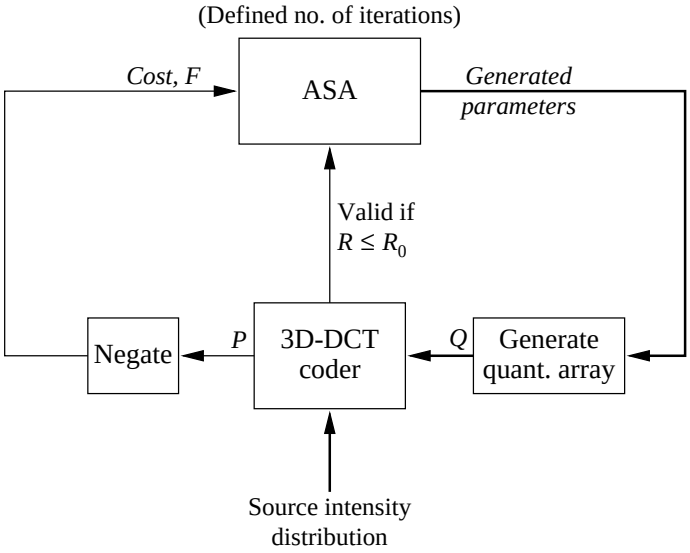


Figure 8.4: Environment for derivation of optimal quantisation parameters for the 3D-DCT based coding scheme.

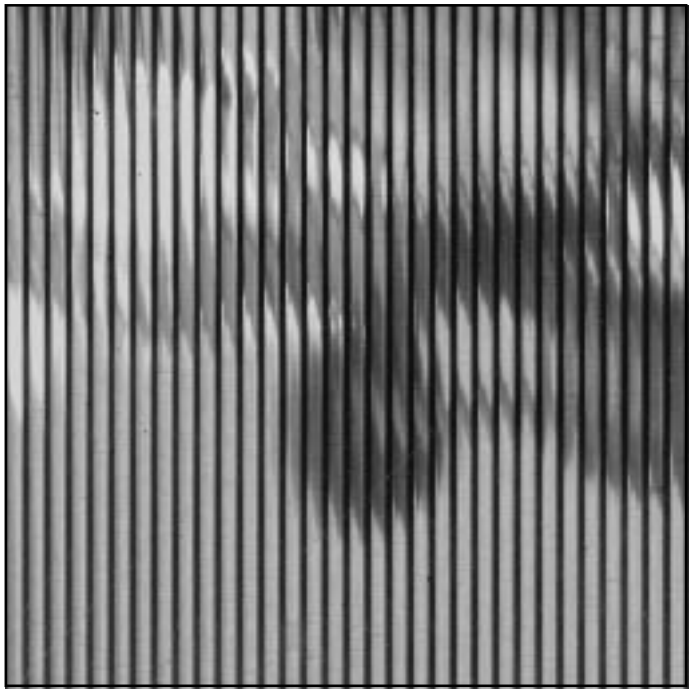


Figure 8.5: Intensity distribution section used in derivation of optimal quantisation parameters.

pose, the optimisation process has been carried out for an output bit rate of 0.250061 bits per pixel, as this can be produced by a fixed quantisation array with scale $q = 69$, thereby allowing a direct comparison between quantisation strategies. The initial value for the parameter c is 68, generating the fixed quantisation array as above. This parameter is constrained within the range $0-73$. The initial values and bounds for the remaining parameters are as given in Table 8.1.

It was found necessary to modify the annealing temperature schedule from its default such that more time was spent at relatively high temperatures. This causes a greater number of higher-than-best-cost solutions to be accepted, tending to allow the process to escape from the many local minima in the solution space. If this was not done, the optimisation process was found to terminate relatively soon after starting in a local minimum. A suitable value for the appropriate ASA internal parameter was found by experiment. The procedure was run to a total of $100,000$ generated states; this value was again set by experiment as the number of states after which very few better solutions were being found. The total execution time on a Digital AlphaStation 500/266 was approximately ten CPU hours. Both ASA and the 3D-DCT coder were implemented in the C language under the UNIX operating system.

The final best cost obtained was -30.62521, corresponding to a peak signal to noise ratio of 30.62521 dB. The output bit rate was precisely 0.250061 bits per pixel. The associated generating function parameters are listed in Table 8.2, and the first (u,v)-plane of the generated optimal quantisation array is shown in Figure 8.6. Note that the high precision of the analysis results has been preserved to ensure that the optimally-determined quantisation array may be regenerated with the highest possible level of accuracy. Coefficients along the vertical spatial frequency axis are quantised with far greater step sizes than those at corresponding points on the horizontal axis. This confirms that the coefficients with the highest significance in high quality reconstruction of the source data are those at low vertical frequency points; this is in agreement with the coefficient standard deviation surfaces presented in section 8.1. The profile of the increase in step size with vertical frequency index can be seen to be almost linear, but with a shallower gradient at low values of v. This was gener-

Parameter	a_1	b_1	a_2	b_2	a_3	b_3	c
Final value	1.145325	1.021189	2.876126	1.725473	1.344206	1.728969	46.4251

Table 8.2: Final parameter values for the quantisation array generating function after an ASA run with $R_0 = 0.250061$ bpp.

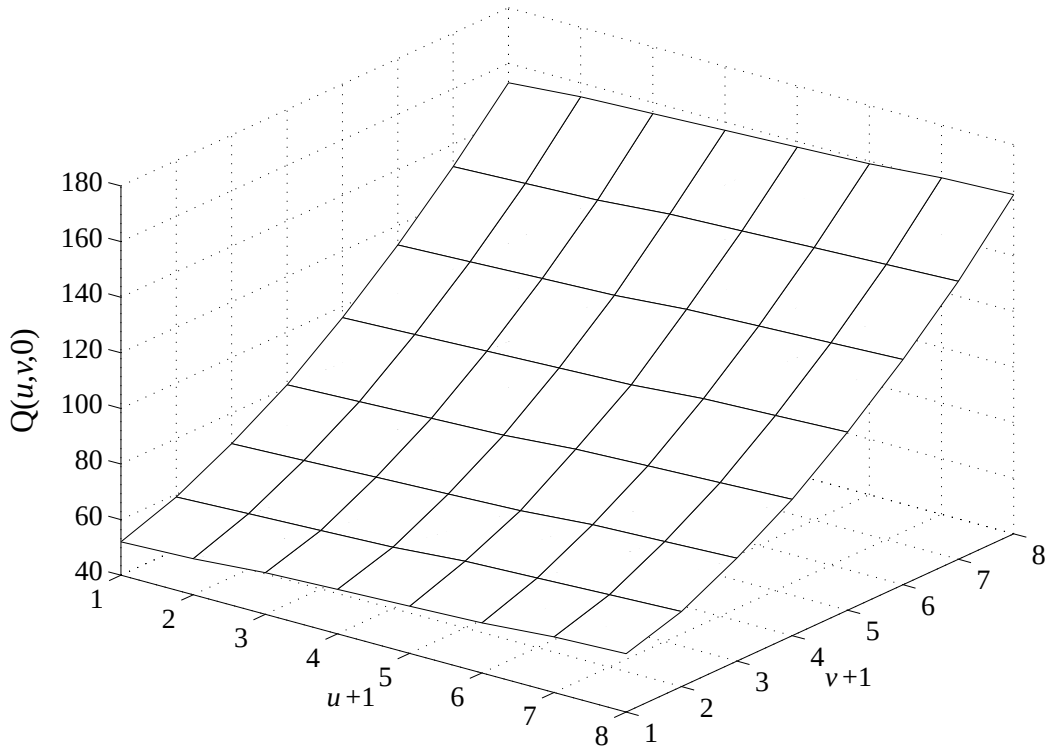


Figure 8.6: First (u,v)-plane of optimal quantisation array obtained for a bit rate $R_0 = 0.250061$ bpp.

ally found to be the case in testing the procedure with various values of R_0 between 0.125 and 1 bpp.

Use of the single-point optimised quantisation array led to an increase in peak signal to noise ratio of 0.165 dB over use of fixed quantisation. The fact that such a performance improvement is possible confirms that to obtain optimal performance, a unique quantisation array structure should be used in the 3D-DCT lenticular-integral image coding scheme.

8.2.6 Development of a multi-point quantisation array generation method

The previous subsection outlined the derivation of an optimal quantisation array to give rise to a predefined coder output bit rate when applied to a specific intensity distribution. The development of a method for production of general purpose quantisation arrays which include variation of output bit rate is now described. This has been carried out by fitting empirical models to the seven generating function parameters which produce optimal quantisation

arrays across a range of values of R_0. The models derived are then used to produce quantisation arrays for use at other intermediate output bit rates. Since this process is carried out with one particular intensity distribution as source data, when a general-purpose quantisation array produced from the model equations is applied to any other source intensity distribution data the output bit rate is highly unlikely to be equal to the bit rate requested. The requested bit rate parameter is therefore effectively an arbitrary scale factor with no numerical significance, as was the parameter q in previous use of fixed quantisation arrays. The procedure is now described in detail.

Optimisation

Initially the single-point optimisation procedure was run for values of R_0 from 0.125 to 1 bpp in steps of 0.125 bpp. The adjustment of the ASA internal parameter modifying the annealing schedule as described in the previous subsection, was made. It was found that in order to obtain satisfactory optima at all points it was necessary for the process to spend more time at high temperatures at lower values of R_0 (0.125–0.375 bpp) than at intermediate or higher values. The resulting eight sets of generating function parameters corresponding to optimal quantisation arrays at each value of R_0 tested were plotted against R_0 and examined for trends in order that empirical models could be fitted. Trends were identified in each case though a number of points fell some distance from them. This was believed to be the case due to interaction between the multiplicative and exponential parameters in each spatial frequency dimension resulting in a number of quantisation array profiles which were similar in the highly significant low frequency regions, hence giving rise to similarly high quality optima. Second stage optimisation runs were carried out as necessary with narrower bounds on these parameters to cause each to more closely follow the overall identified trends. The solutions were deemed satisfactory provided that they led to optima which were the same as or better than those produced during the first stage optimisation runs. It was necessary to insert two extra points at bit rates of 0.3125 and 0.4375 bpp to confirm the validity of a trough observed in the parameter trends at 0.375 bpp. Table 8.3 lists the final output bit rate R_f and peak signal to noise ratio P_f obtained. The peak signal to noise ratio expected from use of fixed quantisation at R_f is also given in each case, produced by linear interpolation between points on the associated rate-distortion performance curve. The mean improvement in peak signal to noise ratio across the range tested is 0.147 dB. Figure 8.7 shows each generating function parameter plotted against output bit rate for the final ten optimal quantisation arrays.

Desired bit rate R_0 / bpp	*Final bit rate* R_f / bpp	*Final PSNR* P_f / dB	*Fixed quant. PSNR* P_{fixed} / dB	*PSNR gain* $P_f - P_{fixed}$ / dB
0.1250	0.124985	27.058443	26.898277	0.160166
0.2500	0.250000	30.617353	30.459164	0.158189
0.3125	0.312469	31.779984	31.605852	0.174132
0.3750	0.374969	32.694618	32.559680	0.134938
0.4375	0.437500	33.481133	33.353223	0.127910
0.5000	0.499939	34.165676	34.015075	0.150601
0.6250	0.624985	35.327255	35.160499	0.166756
0.7500	0.749878	36.246910	36.132822	0.114088
0.8750	0.874954	37.032940	36.895202	0.137738
1.0000	0.999939	37.657547	37.509310	0.148237

Table 8.3: Comparison of objective fidelity between use of fixed quantisation and optimal quantisation arrays.

Parameter model fitting

Empirical models were fitted piecewise to each parameter curve; model fitting was performed using a general least squares fit program 'fitls', part of the 'GLE' data and function manipulation and plotting system [66]. Between three and five models were used per parameter; the number of models necessary for each parameter and the form of each model were determined by inspection such that each had a maximum of one point of inflection. Most models were of polynomial form of second and third degree. The following represent the final models used for each parameter:

$$a_1 = \begin{cases} -2.34343R^2 - 0.982619R + 5.08354 & \text{if } 0 < R < 0.374969 \\ \frac{0.388477}{R-0.292573} - 0.336799 & \text{if } 0.374969 \leq R < 0.749878 \\ 1.47921R^3 - 0.0147098R^2 - 2.6664R + 1.95195 & \text{if } 0.749878 \leq R \leq 0.999939 \end{cases} \tag{8.9a}$$

$$b_1 = \begin{cases} 15.67R^2 - 9.1106R + 1.81023 & \text{if } 0 < R < 0.374969 \\ 23.9867R^3 - 2.14246R^2 - 13.8641R + 4.8298 & \text{if } 0.374969 \leq R < 0.499939 \\ -0.57164R^2 + 1.21367R - 0.103537 & \text{if } 0.499939 \leq R < 0.749878 \\ 0.0701631R^3 - 0.0421988R^2 - 0.167313R + 0.604731 & \text{if } 0.749878 \leq R \leq 0.999939 \end{cases} \tag{8.9b}$$

$$a_2 = \begin{cases} -35.6708R^2 + 4.63937R + 3.51476 & \text{if } 0 < R < 0.312469 \\ -17.7012R^2 + 11.2388R - 0.308662 & \text{if } 0.312469 \leq R < 0.499939 \\ -12.6741R^2 + 13.0838R - 2.49439 & \text{if } 0.499939 \leq R < 0.749878 \\ -6.61594R^3 + 1.06438R^2 + 15.0071R - 8.87221 & \text{if } 0.749878 \leq R \leq 0.999939 \end{cases} \tag{8.9c}$$

$$b_2 = \begin{cases} 16.1228R^2 - 9.21436R + 4.6654 & \text{if } 0 < R < 0.312469 \\ 28.8841R^2 - 24.5005R + 8.19586 & \text{if } 0.312469 \leq R < 0.437500 \\ 0.0169072R + 2.99795 & \text{if } 0.437500 \leq R < 0.624985 \\ -0.292082R + 3.19114 & \text{if } 0.624985 \leq R < 0.749878 \\ 9.47488R^3 - 1.73225R^2 - 22.0505R + 16.4861 & \text{if } 0.749878 \leq R \leq 0.999939 \end{cases} \tag{8.9d}$$

$$a_3 = \begin{cases} -21.8126R^2 + 0.975296R + 3.14592 & \text{if } 0 < R < 0.374969 \\ -51.8961R^3 + 5.24703R^2 + 31.2884R - 9.32927 & \text{if } 0.374969 \leq R < 0.499939 \\ -2.37742R^2 + 0.47554R + 1.49631 & \text{if } 0.499939 \leq R < 0.749878 \\ -1.29942R^3 - 0.0396764R^2 + 2.33876R - 0.667498 & \text{if } 0.749878 \leq R \leq 0.999939 \end{cases} \tag{8.9e}$$

$$b_3 = \begin{cases} -9.01335R^2 + 1.9703R + 1.52131 & \text{if } 0 < R < 0.312469 \\ -10.8046R^2 + 7.45234R - 0.0167603 & \text{if } 0.312469 \leq R < 0.437500 \\ 0.571908R^2 - 2.25571R + 2.0598 & \text{if } 0.437500 \leq R < 0.749878 \\ 3.02304R^3 - 0.473161R^2 - 6.7524R + 4.74726 & \text{if } 0.749878 \leq R \leq 0.999939 \end{cases} \tag{8.9f}$$

$$c = \begin{cases} \frac{7.78404}{R - 0.0491284} + 1.55002 & \text{if } 0 < R < 0.624985 \\ 13.5331R + 5.21174 & \text{if } 0.624985 \leq R < 0.749878 \\ 27.8748R^3 - 9.0847R^2 - 75.972R + 65.6842 & \text{if } 0.749878 \leq R \leq 0.999939 \end{cases} \tag{8.9g}$$

This set of equations may be used in conjunction with the generating function (Equation 8.4) to produce quantisation arrays for general use with the 3D-DCT lenticular-integral image coding scheme.

Model-derived quantisation array evaluation

This procedure was carried out with R ranging from 0.13 to 0.99 in steps of 0.01. The resulting quantisation arrays were applied to coding of the luma channels of the four test intensity distributions *Horseman*, *Microscope*, *Tank* and *Toyplane* using the 3D-DCT scheme. In the tests the scheme used all-plane per-row scanning and coding and a global Huffman code ta-

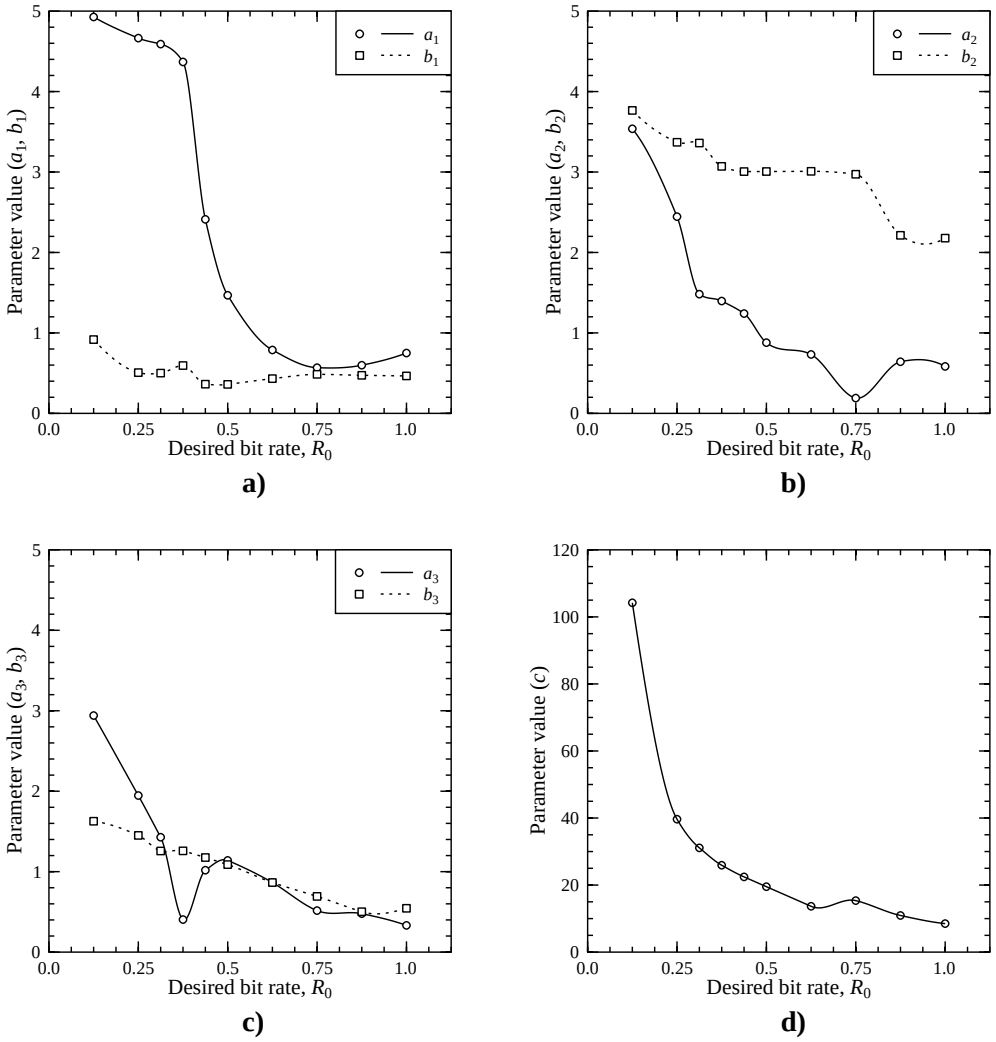


Figure 8.7: Plot of final ten-point generating function parameters against output bit rate. a) *u*-dimension (a_1, b_1). b) *v*-dimension (a_2, b_2). c) *w*-dimension (a_3, b_3). d) Constant term (c).

ble generated for a target output bit rate of 0.5 bpp, to allow direct comparison with the fixed quantisation case simulated previously. Figure 8.8 shows the mean improvement in objective fidelity obtained when the quantisation arrays used have been produced from the models developed above rather than being fixed. An improvement in rate-distortion performance is noted across the bit rate range tested. However, it is observed that the improvement in performance increases as output bit rate is lowered below approximately 0.35 bpp, partially compensating for the loss in performance in this region due to use of a global Huffman code table.

Optimal model-derived quantisation arrays were also applied to coding of the chrominance

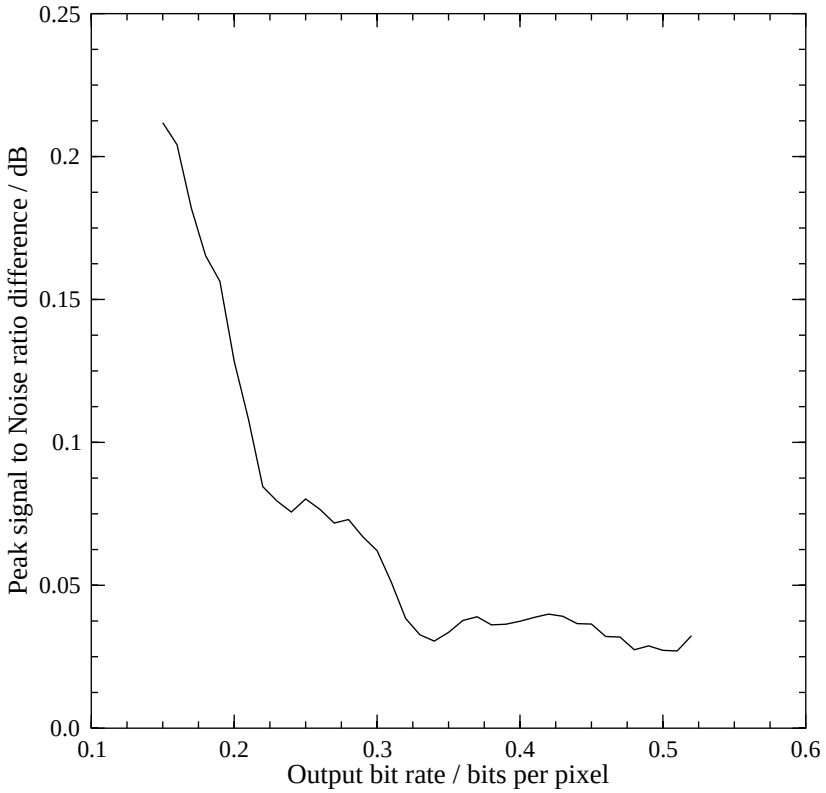


Figure 8.8: Improvement in objective fidelity due to use of optimum model-derived quantisation arrays rather than fixed quantisation arrays; all-plane scanning strategy and code table generated for 0.5 bpp were used on the luma channel only of the source data.

channels of the intensity distributions *Horseman*, *Tank* and *Toyplane*. In full colour coding a performance loss resulted compared with use of fixed quantisation. Since the model-derived arrays were developed based on optimisation using luma channel data only, this result indicates that for full colour coding a different quantisation model is required to be applied to the chrominance data from that applied to the luma data. Such a model would be developed using the multi-point optimisation procedure described above but providing the coder with a small section of the chrominance-blue and chrominance-red channels of the intensity distribution from which the luma section was taken.

Images 8.1 and 8.2 are the reconstructed luma channels of the intensity distributions *Horseman* and *Toyplane* respectively, having been coded and decoded using the 3D-DCT scheme with optimal model-derived quantisation arrays developed above. The results of coding to the same bit rate where fixed quantisation arrays were used are also presented to enable a comparison to be made in each case. When operating at such low bit rates, blocking arti-

facts become highly noticeable both in the subimage and group domains. Where optimum model-derived quantisation arrays have been used, however, the group blocking evident in the fixed quantisation case, particularly in background areas of the images, has been reduced. This may be seen in the magnified sections of the above intensity distributions presented in Figure 8.9. This observation demonstrates that the optimisation/model based design process used has yielded a quantisation scheme which provides improved rate-distortion performance over use of the scaled fixed quantisation array. The process provides the basis for a general quantisation strategy suitable for use with the 3D-DCT scheme when compressing lenticular-integral image data. For use in a practical implementation, its robustness should be improved by the use of a selection of intensity distribution sections during optimisation, and it should be extended to the case of chrominance channel quantisation for colour integral image coding.



[Fixed quantisation case]



Image 8.1: Luma channel of *Horseman* compressed using the 3D-DCT scheme with optimal model-derived quantisation; output bit rate is 0.14 bpp. [SCREEN B]



[Fixed quantisation case]



Image 8.2: Luma channel of *Toyplane* compressed using the 3D-DCT scheme with optimal model-derived quantisation; output bit rate is 0.16 bpp. [SCREEN A]

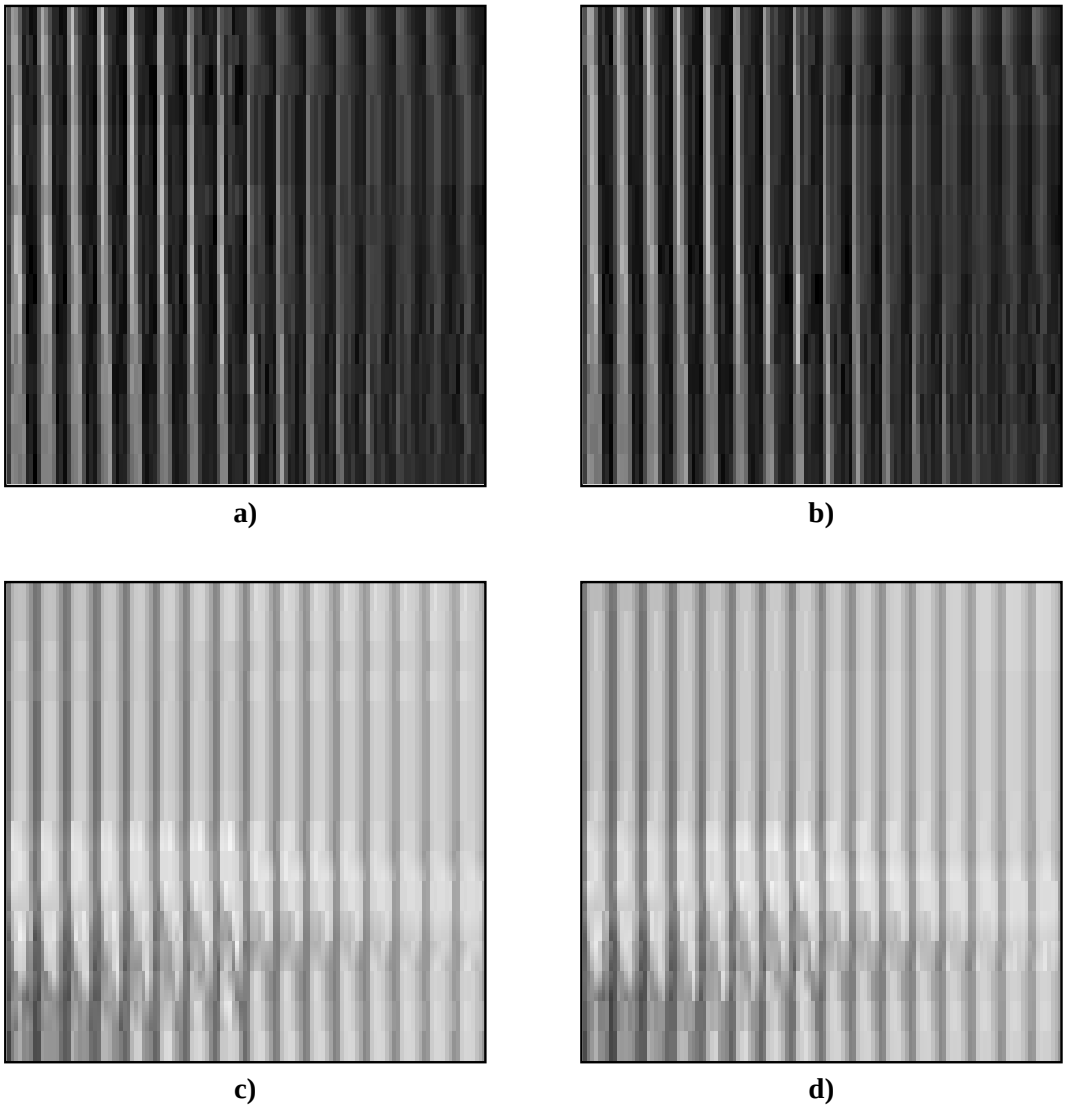


Figure 8.9: Magnified 16×16-subimage sections of luma channels of *Horseman* and *Toyplane* resulting from coding using the 3D-DCT scheme with fixed and optimal model-derived quantisation.
a) *Horseman*, 3D-DCT with fixed quantisation at 0.14 bpp.
b) *Horseman*, 3D-DCT with optimal model-derived quantisation at 0.14 bpp.
c) *Toyplane*, 3D-DCT with fixed quantisation at 0.16 bpp.
d) *Toyplane*, 3D-DCT with optimal model-derived quantisation at 0.16 bpp.

Summary

This chapter has described the development of quantisation arrays for the 3D-DCT based coding scheme which are more closely suited to the statistical characteristics of transformed lenticular-integral 3D image data than the fixed quantisation arrays used previously. A statistical evaluation led to the conclusion that the spatial significance of 3D-DCT coefficients in the intra-subimage vertical dimension is generally lower than that in the intra-subimage horizontal dimension. The method of simulated annealing was employed, together with model fitting techniques, to design a method for applying quantisation which is optimal in a rate-distortion sense at any point within a continuous range of output bit rate values. The form of the resulting optimal quantisation arrays confirmed the earlier statistical observations, with the comparatively low significance intra-subimage vertical 3D-DCT coefficients being quantised with larger step sizes than corresponding intra-subimage horizontal coefficients. The general procedure developed requires only the evaluation of a set of equations to produce a quantisation array for direct use by the encoder. Simulations using the luma channels of four intensity distributions resulted in improved rate-distortion performance compared with use of fixed quantisation arrays. An increase in objective fidelity, using the peak signal to noise ratio metric, of between 0.05 and 0.2 dB was achieved using optimal model-derived quantisation arrays, and subimage group orientated blocking artifacts, particularly noticeable in image backgrounds, were substantially reduced. The investigation indicates that improved performance results if the quantisation arrays used by the 3D-DCT scheme are tailored to lenticular-integral 3D image data, taking account of the unique spectral characteristics of their intensity distributions.

9. CONCLUSIONS

9.1 Research Undertaken

9.1.1 Acquisition and Pre-processing of Integral 3D Image Data

For digital domain processing to be possible, a photographically recorded integral three dimensional image must first be scanned electronically. The resulting data, in the form of an $R'G'B'$-space pixel map, is subjected to processing operations for correction of geometric distortions introduced on the photographic/electronic interface at the scanner, and in order to scale the resulting data to a common pixel resolution for compression. A filtering process, based on a one dimensional median filter, has been developed to reject any high spatial frequency noise which may be introduced at the scanning stage. For final display of the image, the data reconstructed at the receiver's decoder are scaled to the display system's output resolution.

9.1.2 The Nature of the Digital Integral 3D Image Representation

Although the captured integral 3D image data are in the standard format of an $R'G'B'$-space pixel map, it is inappropriate to refer to the data as an image. Since they result from a complex mapping of three dimensional space to a planar form carried out by a lenticular or microlens array, the pixel map is referred to as an integral intensity distribution. The intensity distribution possesses a structure unique to the integral imaging method, giving rise to statistical characteristics unlike those typical to a conventional two dimensional natural image. Significant cross-correlation is present between corresponding pixels from neighbouring lenticular bands or microlenses. Additionally, significant correlation is present between pixels from each lenticular band or microlens due to low spatial frequencies being dominant.

9.1.3 Coding Techniques Developed for Use with Integral 3D Image Data

Two coding scheme classes have been developed which exploit the unique statistical properties of electronically captured integral 3D image intensity distributions. The first scheme uses a hybrid DPCM/DCT mapping structure and the second a 3D-DCT based mapping structure. Both use scalar uniform symmetric quantisation stages and run-length/Huffman entropy coders. The results of simulations using fixed quantisation arrays indicate that transmission of up to three standard definition resolution pictures is possible using the hybrid DPCM/DCT scheme, and up to four standard definition or one high definition picture using the 3D-DCT scheme, such that the reconstructed decoded image forms a high quality 3D display indistinguishable from the source image. These values assume a channel using 64-QAM modulation, a baseband shaping filter with a roll-off factor of 0.15 and an analogue transmission bandwidth of 8 MHz. Such a channel is practicable within a cable transmission system. Additionally, since the base coding schemes have been developed for still image compression, the significant inter-frame correlation present in motion video sequences is not exploited, leading to underestimation in the performance of the schemes for video coding; see subsection 9.2.3.

9.1.4 Coding Scheme Quantisation Stage Optimisation

Improved rate-distortion performance is possible in the 3D-DCT scheme by making use of quantisation arrays which are tailored specifically to the characteristics of lenticular-integral 3D image intensity distributions. An optimisation process has been used to derive a model for quantisation within a continuous range of output bit rate values. An improvement in objective fidelity over use of fixed quantisation has been observed. Measured by the peak signal to noise ratio of the reconstructed intensity distribution, this is between 0.05 and 0.2 dB where the luma channel only of the source is taken into account. Examinations of reconstructed intensity distributions were carried out where the quantisation array was designed in this way. A reduction in group domain blocking artifacts was noted when compared with the use of fixed quantisation, particularly when coding to low bit rate values, giving rise to improved reconstructed image fidelity. The optimisation and modelling technique provides a general design methodology for quantisation strategies suitable for use with the 3D-DCT based integral image coding scheme. In general, the use of such a tailored quantisation array is important in achieving optimal rate-distortion performance.

Figures 9.1 and 9.2 present summary performance data for the various schemes developed

in the form of staged bar charts, for luma only and full colour cases respectively. In both the luma and full colour cases, the greatest rate-distortion performance increase occurs in moving from baseline sequential-DCT JPEG to the base hybrid DPCM/DCT scheme. This demonstrates the importance of including processes to exploit the inter-subimage correlation which occurs in integral intensity distribution data. The enhanced hybrid DPCM/DCT scheme provides a further minimal performance increase without further distortion for a given output bit rate. The second most significant improvement occurs when the 3D-DCT scheme is applied, particularly at low bit rates. This occurs due to the improved decorrelation effectiveness of the DCT over the differencing of neighbouring subimages as is used in the hybrid DPCM/DCT schemes. The use of the all-plane scanning strategy, matched closely to the statistical characteristics of lenticular-integral image data, provides a further

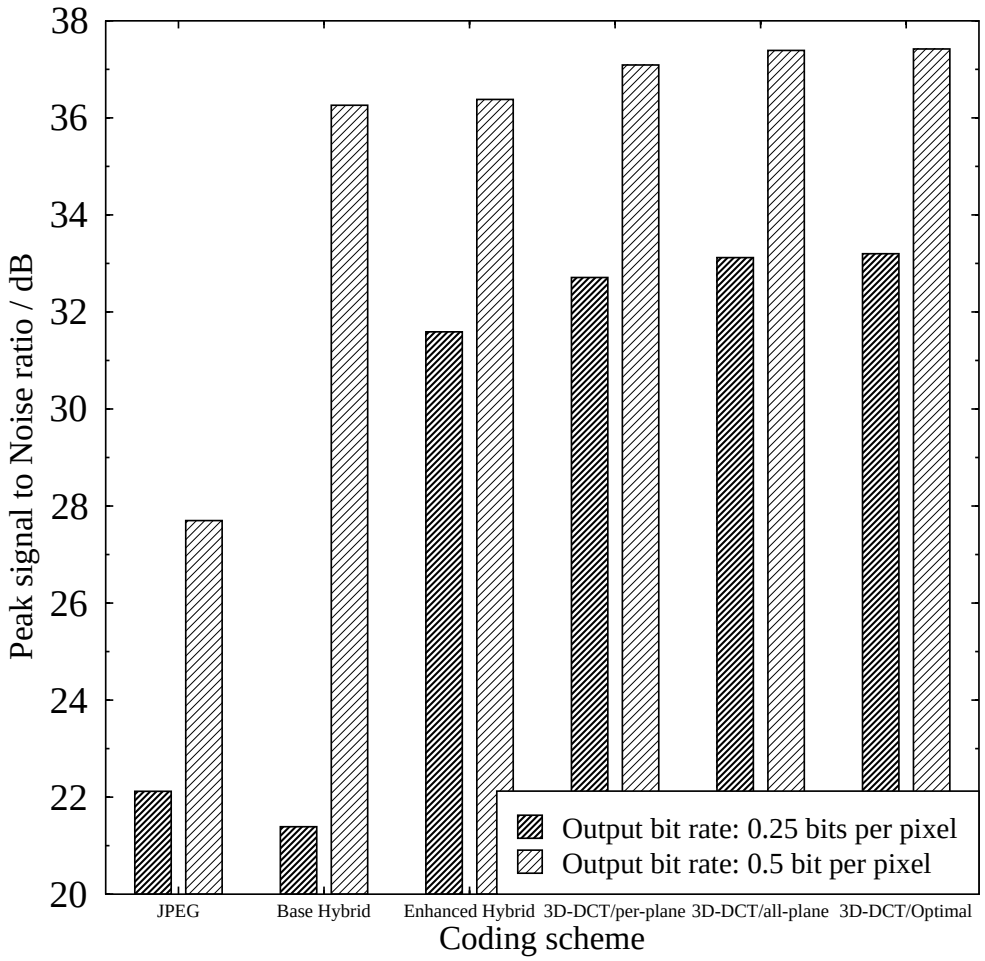


Figure 9.1: Staged performance data for six coding schemes applied to lenticular-integral intensity distribution data (luma channel only).

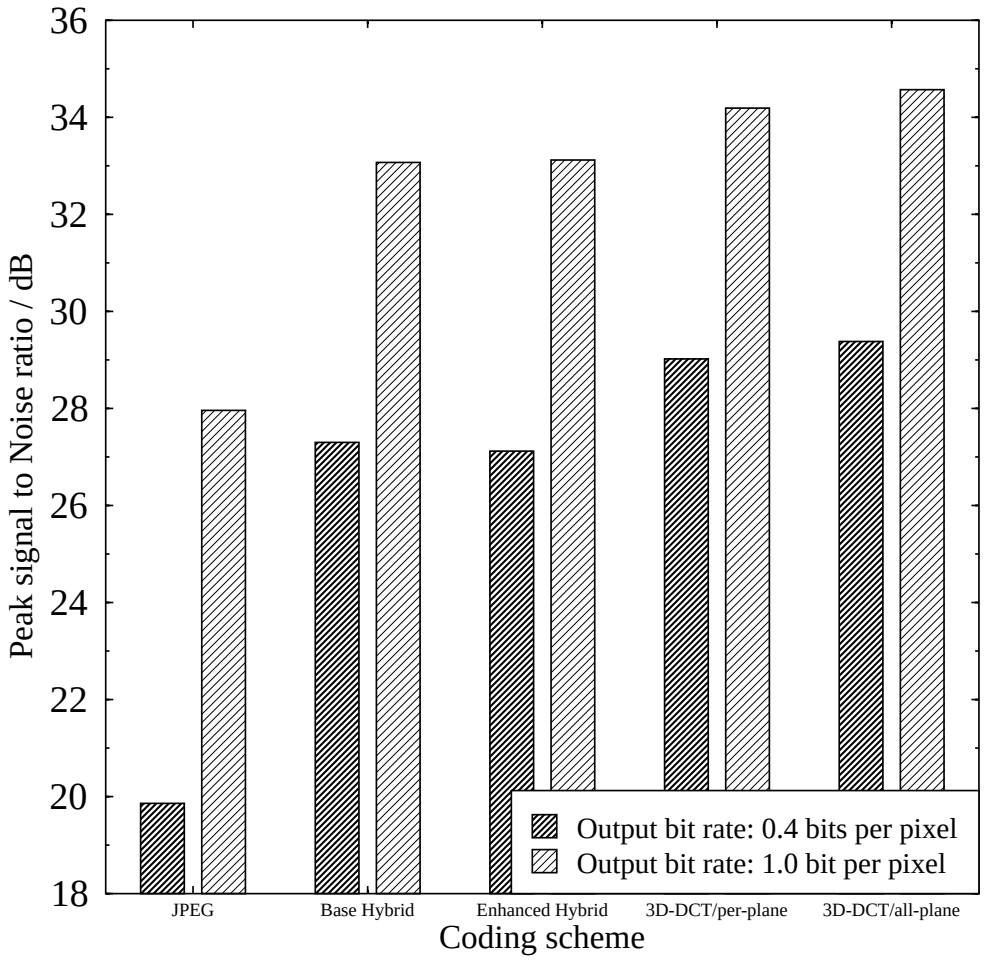


Figure 9.2: Staged performance data for five coding schemes applied to lenticular-integral intensity distribution data (full colour coding).

reduction in bit rate for a given fidelity level. Finally, a small improvement is seen when the quantisation stage is optimised for a bit rate of 0.25 bits per pixel and a model derived to produce near-optimal rate-distortion performance across the bit rate range. Note that optimal model-derived quantisation arrays have been produced and tested only for the case of luma channel only coding.

In summary, the highest performance scheme developed, that is the 3D-DCT based scheme with optimal model-derived quantisation and tailored coefficient scanning order, has a performance level allowing high quality reconstruction of encoded integral three dimensional images, and transmission of four such images at standard definition or one such image at high definition resolution. This performance level is not possible using conventional coding

schemes such as JPEG.

9.2 Suggestions for Further Work

9.2.1 Use of Microlens Captured Full Integral Image Data

The research carried out for this thesis has made use of integral 3D images captured using lenticular sheet material (*lenticular-integral* images). Images of this type possess continuous parallax in replay only in the horizontal dimension. Use of two-dimensional microlens arrays, which allow display with continuous parallax in all directions, has been found in laboratory trials to lead to an improved sense of depth and realism to the viewer. An intensity distribution which represents an integral 3D image captured using a microlens array as the encoding element (termed a *full integral* image) consists of a two dimensional array of subimages. This arrangement may be envisaged as one in which a lenticular encoding sheet has been overlaid with a second rotated by 90° relative to the first. The resulting intensity distribution is expected to possess significant inter-subimage correlation in both axes rather than in the horizontal axis only as is the case for a lenticular-integral intensity distribution.

The 3D-DCT based coding scheme could be applied to such data, exploiting inter-subimage correlation in both dimensions of the microlens array. Figure 9.3 shows a possible strategy for assembly of a subimage group of length 4. An additional consideration arises in the group assembly problem where a full integral intensity distribution is concerned, that is the arrangement of source subimages into the transform group. For example, in the figure assembly takes place in a clockwise sense beginning at subimage ‘A’, that is in order ‘ABCD’. In this case, inter-subimage correlation between subimages ‘A’ and ‘D’ is not exploited in the transform computation. If this were higher than that between, for example ‘A’ and ‘B’, more effective decorrelation may be obtained by beginning the assembly at subimage ‘B’, resulting in order ‘BCDA’. Additionally the sense of rotation in assembly may be reversed. For the 3D-DCT to provide maximally efficient decorrelation of a group, the group assembly strategy must be chosen based on the statistical characteristics of the source data.

Modifications to the quantisation step size array, quantised coefficient scanning order and entropy coder Huffman code table will be necessary to suit the horizontal/vertical symmetrical statistical characteristics associated with microlens array-encoded integral images. These will also be influenced by the group assembly arrangement chosen.

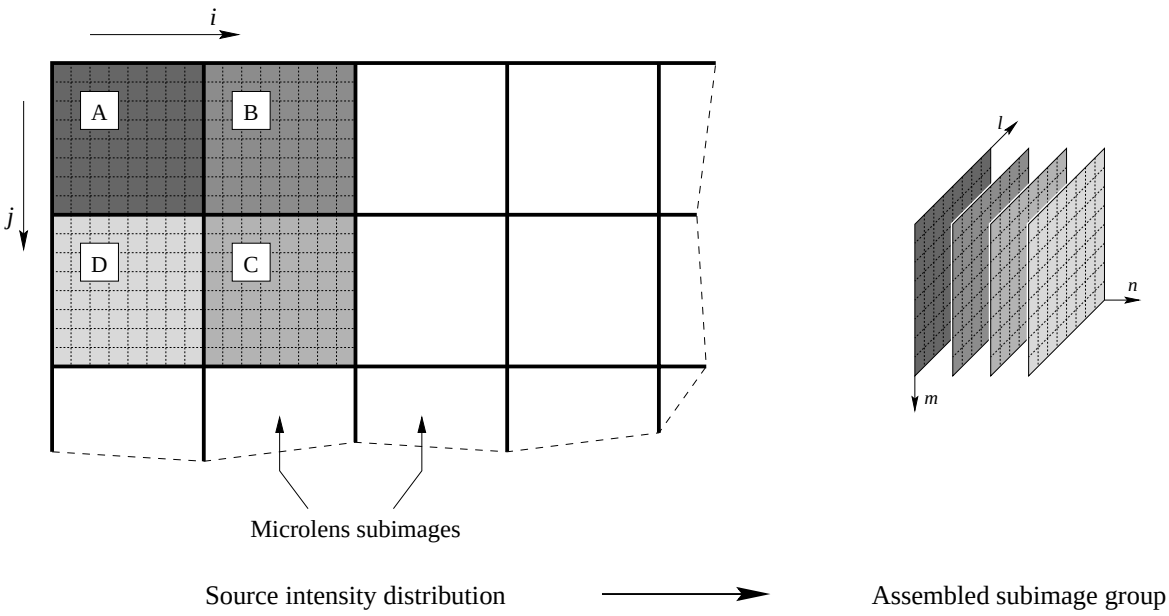


Figure 9.3: Assembly of a subimage group for 3D-DCT based full integral 3D image coding.

9.2.2 Addition of Local Image-Adaptive Features

The inclusion of methods to allow the 3D-DCT coding scheme to exploit the local statistical features of a given intensity distribution more efficiently would be expected to lead to improved rate-distortion performance. This could be achieved by adaptively modifying parameters of the scheme based on measurements made on the source intensity distribution. Of chief importance is the group assembly strategy, since this determines the effectiveness of decorrelation performed by the transform stage. For individual regions of the source intensity distribution, it is suggested that group length, assembly start point and order of subimage assembly are chosen independently by an adaptive algorithm using information obtained from a segmentation of a *mean image*. The mean image is a pixel map comprised of the mean value of each subimage in the source data, and effectively forms a low resolution image of the object scene. The quantisation and entropy coding stage parameters will require adaptation to the group assembly structure in use in a particular region of the intensity distribution.

9.2.3 Extension to Motion Video Integral Image Data

For application of the 3D-DCT based coding scheme to integral 3D motion video, methods should be developed to take account of the inter-frame cross-correlation existing between

corresponding pixels and also to encode the inter-frame movement of regions of the object scene efficiently. It is suggested that a DPCM loop be used for inter-frame decorrelation, the resulting differential frame data then being input to the 3D-DCT based still integral image coding scheme. A motion estimation scheme suitable for use with integral intensity distributions and the 3D-DCT scheme could be based on a full search within the mean images formed from each frame (described in the previous subsection). Use of the two-dimensional mean images reduces the computational complexity of the motion estimation algorithm, though it will be necessary to assess fully the implications of making use of the relationship between source frames and their associated mean images in this fashion.

9.2.4 Design Prototyping of the 3D-DCT based Scheme

For evaluation of the computational complexity of the 3D-DCT coding scheme, a design prototyping operation should be carried out. This will facilitate broad assessment of the hardware requirements and implementation issues associated with the scheme for compression of integral 3D motion video streams.

9.2.5 Integration of Application-Orientated Data Structures and Syntax

Application of the 3D-DCT integral 3D image and video coding schemes as an element of a broadcast system will require the development of a formal syntax for the encoded data stream and formats for transmission of ancillary data such as quantisation arrays and Huffman code tables. This is necessary for correct operation of a receiver/decoder unit in terms of synchronisation of both the incoming stream and the decoding mode with the encoder. Full specification of the format will also enable straightforward integration of error correction coding processes which are essential for correct operation in the broadcast environment. To facilitate this, redundancy will inevitably be introduced into the final broadcast stream due to the requirement for signalling code words. The effect of this on transmitted bit rate will necessitate re-evaluation of the rate-distortion performance of the coding schemes. It is anticipated that the syntax and organisation of the data stream and ancillary information will be based on that used by the MPEG-2 video coding system.

APPENDICES

A. TEST INTEGRAL IMAGE INTENSITY DISTRIBUTIONS

The intensity distributions presented in this appendix are those which have been used in simulations of the compression schemes described in this thesis. They have been reproduced at the correct scale for three dimensional replay using the lenticular sheets provided. The appropriate lenticular sheet for each intensity distribution is given in its caption.



Image A.1: *Toyplane* ($p = 1270\ \mu m$). [SCREEN A]



Image A.2: *Microscope* ($p = 1270\ \mu m$). [SCREEN A]



Image A.3: *Tank* ($p = 600\ \mu m$). [SCREEN B]



Image A.4: *Horseman* ($p = 600\ \mu m$). [SCREEN B]

B. CONVENTIONAL TEST IMAGES FOR STATISTICAL COMPARISONS

The following conventional two dimensional luma channel images were used for comparison purposes as part of the statistical evaluations of lenticular-integral intensity distribution data described in chapter 5.



Figure B.1: *Lena*.



Figure B.2: *Airplane*.



Figure B.3: *Peppers*.



Figure B.4: *Sailboat.*

C. PUBLICATIONS

1. Brewin, M., Forman, M., and Davies, N., ‘Electronic capture and display of full parallax 3D images,’ SPIE/IS&T Electronic Imaging ’95, San Jose, USA. Proc. SPIE, vol. 2409, pp. 118–124, 1995.

2. Forman, M., Aggoun, A., and McCormick, M., ‘Compression of integral 3D TV pictures,’ 5th International Conference on Image Processing and its Applications (IPA ’95), Edinburgh, UK., Conf. Pub. IEE, vol. 410, pp. 584–588, 1995.

3. Forman, M.C., and Aggoun, A., ‘Compression of full parallax integral 3D-TV image data,’ SPIE/IS&T Electronic Imaging ’97, San Jose, USA. Proc. SPIE, vol. 3012, pp. 222–226, 1997.

4. Forman, M.C., Aggoun, A., and McCormick, M., ‘A novel coding scheme for full parallax 3D-TV pictures,’ Proceedings of the 1997 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP ’97), Munich, Germany, 21–24 April, pp. 2945–2947.

5. Forman, M.C., and Aggoun, A., ‘Quantisation strategies for 3D-DCT-based compression of full parallax 3D images,’ 6th International Conference on Image Processing and its Applications (IPA ’97), Dublin, Ireland, Conf. Pub. IEE, vol. 443, pp. 32–35, 1997.

6. Forman, M.C., Aggoun, A., and McCormick, M., ‘Simulated Annealing for Optimisation and Characterisation of Quantisation Parameters in Integral 3D Image Compression,’ 2nd IMA Conference on Image Processing: Mathematical Methods, Algorithms and Applications, De Montfort University, Leicester, UK., 23–25 September 1998. Published in: (IMA) Institute of Mathematics and its Applications, ‘Image Processing II Mathematical Methods, Algorithms and Applications’, J.M. Blackledge and M.J. Turner eds., pp. 399–413, Horwood Publishing, 2000 (ISBN 1-898563-61-6).

REFERENCES

[1] Pennebaker W.B., Mitchell J.L., *JPEG: Still Image Data Compression Standard*, (Van Nostrand Reinhold, New York, 1993).

[2] ISO/IEC 10918-1,2,3 | ITU-T Recs. T.81,T.83,T.84, 'Information technology — digital compression and coding of continuous-tone still images,' 1994–1996.

[3] Motoki T., *et al.*, 'Present status of three-dimensional television research,' Proc. IEEE, vol. 83, pp. 1009–1021, 1995.

[4] Moor J.R., *et al.*, 'The implementation of a multi-view autostereoscopic display,' in *IEE Colloq. 'Stereoscopic Television' 1992/173*, pp. 4/5–4/16, 1992.

[5] Isono H., Yasuda M., 'Autostereoscopic 3-D television display,' NHK Laboratories Notes, , no. 398, December 1991.

[6] Davies N., *et al.*, 'Design and analysis of an image transfer system using microlens arrays,' Optical Eng., vol. 33, no. 11, pp. 3624–3633, November 1994.

[7] Davies N., *et al.*, 'Three dimensional imaging systems: A new development,' Appl. Optics, vol. 27, pp. 4520–4528, 1988.

[8] Lippmann G., 'Epreuves reversibles,' Comptes rendus hebdomadaires des Séances de l'Académie des Sciences, vol. 146, pp. 446–451, March 1908.

[9] Ives H.E., 'Optical properties of a Lippmann lenticulated sheet,' J. Opt. Soc. Amer., vol. 21, pp. 171–176, 1931.

[10] Okoshi T., *Three Dimensional Imaging Techniques*, (Academic Press, London, 1976).

[11] Burckhardt C.B., 'Optimum parameters and resolution limitation of integral photography,' J. Opt. Soc. Amer., vol. 58, pp. 71–76, 1968.

[12] Brewin M., *et al.*, 'Electronic capture and display of full parallax 3D images,' in *Proc. SPIE*, vol. 2409, pp. 118–124, 1995.

[13] St-Hilaire P., *et al.*, 'Advances in holographic video,' in *Proc. SPIE*, vol. 1914, pp. 188–196, 1993.

[14] McCormick M., Davies N., '3-D worlds,' Physics World, pp. 42–46, June 1992.

[15] Eastman Kodak Company, European Patent 0 520 179 A2, 'Autostereoscopic photography system with electronically interpolated images,' 1992.

[16] ITU-R Recommendation 601-2, 'Encoding parameters of digital television for studios,' 1982–1986–1990.

[17] ITU-R BT.709-3, 'Parameter values for the HDTV standards for production and international programme exchange,' 1998.

[18] EBU Technical Statement D62-1998, 'Choice of an HDTV production standard for Europe,' 1998.

[19] Hutson G.H., *et al.*, *Colour Television: System Principles, Engineering Practice and Applied Technology*, second edn., (McGraw-Hill, 1990).

[20] Schwartz M., *Information Transmission, Modulation and Noise*, fourth edn., (McGraw-Hill, 1990).

[21] Aghvami A.H., 'Digital modulation techniques,' in *Digital Systems Reference Book*, eds. B. Holdsworth, G.R. Martin, (Butterworth-Heinemann, 1991).

[22] ETSI standard EN 300 429, 'Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for cable systems,' 1997.

[23] Mitchell J.L., *et al.*, eds., *MPEG Video Compression Standard*, (Chapman & Hall, New York, 1997).

[24] ISO/IEC 13818-2 | ITU-T Rec. H.262, 'Generic coding of moving pictures and associated audio information: Video,' 1995.

[25] Netravali A.N., Haskell B.G., *Digital Pictures: Representation and Compression*, (Plenum Press, New York, 1988).

[26] Gonzalez R.C., Wintz P., *Digital Image Processing*, second edn., (Addison Wesley, 1987).

[27] Blahut R.E., *Principles and Practice of Information Theory*, (Addison-Wesley, 1987).

[28] Rao K.R., Hwang J.J., *Techniques and Standards for Image, Video and Audio Coding*, (Prentice-Hall, 1996).

[29] Hung S.H.Y., 'A generalization of DPCM for digital image compression,' IEEE Trans. on Pattern Anal. and Machine Intel., vol. PAMI-1, no. 1, pp. 100–109, January 1979.

[30] Ahmed N., *et al.*, 'Discrete cosine transform,' IEEE Trans. Computers, vol. C-23, pp. 90–93, January 1974.

[31] Guy C.G., Bunch D.J., 'Codes and coding for digital systems,' in *Digital Systems Reference Book*, eds. B. Holdsworth, G.R. Martin, (Butterworth-Heinemann, 1991).

[32] Li W., Zhang Y.Q., 'Vector-based signal processing and quantisation for image and video compression,' Proc. IEEE, vol. 83, no. 2, pp. 317–335, February 1995.

[33] Rabbani M., Jones P.W., *Digital Image Compression Techniques*, vol. TT7 of *Tutorial Texts in Optical Engineering*, (SPIE Press, Washington, 1991).

[34] Huffman D.A., 'A method for the construction of minimum redundancy codes,' Proc. IRE, vol. 40, no. 10, pp. 1098–1101, 1952.

[35] Lohscheller H., 'A subjectively adapted image communication system,' IEEE Trans. Comm., vol. COM-32, pp. 1316–1322, December 1984.

[36] Peterson H.A., *et al.*, 'Quantization of color image components in the DCT domain,' in *Proc. SPIE*, vol. 1453, pp. 210–222, 1991.

[37] Campbell F.W., Robson J.G., 'Application of Fourier analysis to the visibility of gratings,' J. Physiol., vol. 197, pp. 551–566, August 1968.

[38] McCormick M., *et al.*, 'Examination of the requirements for autostereoscopic full parallax 3D-TV,' in *Conf. Pub. IEE*, vol. 397, pp. 477–482, 1994.

[39] Cristy J., 'ImageMagick,' 1999,
Internet WWW: http://www.wizards.dupont.com/cristy/ImageMagick.html.

[40] Poskanzer, J. and other contributors, 'NetPBM software,' 1994,
Internet FTP: wuarchive.wustl.edu:/graphics/graphics/packages/NetPBM/.

[41] Paeth A.W., 'A fast algorithm for general raster rotation,' in *Proc. Graphics Interface '86*, pp. 77–81, May 1986.

[42] Schumacher D., 'General filtered image rescaling,' in *Graphics Gems III*, ed. D. Kirk, (Academic Press, 1992).

[43] Khoral Research, Inc., 'Advanced Khoros 2.2,' 1997,
Internet WWW: http://www.khoral.com/.

[44] Budrikis Z.L., 'Visual fidelity criterion and modelling,' Proc. IEEE, vol. 60, pp. 771–779, July 1972.

[45] Mannos J.L., Sakrison D.J., 'The effects of a visual fidelity criterion on the encoding of images,' IEEE Trans. Inf. Theory, vol. IT-20, pp. 525–536, July 1974.

[46] Sproson W.N., *Colour Science in Television and Display Systems*, (Adam Hilger, 1983).

[47] Ford A.M., *et al.*, 'Effects of image compression on image quality,' J. Photo. Science, vol. 43, no. 2, pp. 52–57, 1995.

[48] Durelli A.J., Parks V.J., *Moire Analysis of Strain*, (Prentice-Hall, 1970).

[49] CCITT Study Group XV, 'Recommendations of the H-Series (Report R37),' August 1990.

[50] Carr M.D., 'Video codec hardware to realise a new world standard,' Br. Telecom Technol. J., vol. 8, no. 3, July 1990.

[51] ISO/IEC 11172-2, 'Information technology: Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s: Video,' 1993.

[52] LeGall D., 'MPEG: A video compression standard for multimedia applications,' Comm. ACM, vol. 34, no. 4, pp. 59–63, April 1991.

[53] Press W.H., *et al.*, *Numerical Recipes in C: The Art of Scientific Computing*, second edn., (Cambridge University Press, 1992).

[54] Swarztrauber P.N., 'FFTPACK: A package of Fortran subprograms for the fast Fourier transform of periodic and other symmetric sequences,' National Center for Atmospheric Research, Boulder, Colorado, USA., 1985,
Internet WWW: http://www.scd.ucar.edu/softlib/FFTPACK.html.

[55] Colonnese S., *et al.*, '3D-DCT pipe coding,' in *Proc. SPIE*, vol. 2298, pp. 107–118, 1994.

[56] Monro D.M., Sherlock B.G., 'Optimal quantisation strategy for DCT image compression,' IEE Proc.: Vision, Image & Signal Processing, vol. 143, no. 1, pp. 10–14, 1996.

[57] Beveridge G.S.G., Schechter R.S., *Optimization: Theory and Practice*, (McGraw-Hill, 1970).

[58] Hung A.C., Meng T.H.Y., 'Optimal quantizer step sizes for transform coders,' in *Proc. IEEE ICASSP '91*, pp. 2621–2624, 1991.

[59] Aarts E., Korst J., *Simulated Annealing and Boltzmann Machines*, (John Wiley & Sons, 1989).

[60] Vecchi M.P., Kirkpatrick S., 'Global wiring by simulated annealing,' IEEE Trans. Computer-Aided Design, vol. CAD-2, pp. 215–222, 1983.

[61] Ingber L., 'Simulated annealing: Practice versus theory,' Mathematical & Computer Modelling, vol. 18, no. 11, pp. 29–57, 1993.

[62] Metropolis N., *et al.*, 'Equation of state calculations for fast computing machines,' J. Chem. Phys., vol. 6, pp. 1087–1092, 1953.

[63] Kirkpatrick S., *et al.*, 'Optimization by simulated annealing,' Science, vol. 220, pp. 671–680, 1983.

[64] Ingber L., 'Adaptive simulated annealing (ASA): Lessons learned,' Control and Cybernetics, vol. 25, no. 1, pp. 33–54, 1996.

[65] Ingber L., 'Adaptive Simulated Annealing (ASA) software,' 1993–1998,
Internet WWW: http://www.ingber.com/#ASA-CODE.

[66] Pugmire C., 'Graphics Layout Editor (GLE) 3.3h,' 1991,
Internet Email: chrisp@marc.cri.nz.