Nonlinear rate control techniques for constant bit rate MPEG video coders

Yoo-Sok Saw

A thesis submitted for the degree of Doctor of Philosophy.
The University of Edinburgh.
- March 1997 -
Abstract

Digital visual communication has been increasingly adopted as an efficient new medium in a variety of different fields; multi-media computers, digital televisions, telecommunications, etc. Exchange of visual information between remote sites requires that digital video is encoded by compressing the amount of data and transmitting it through specified network connections. The compression and transmission of digital video is an amalgamation of statistical data coding processes, which aims at efficient exchange of visual information without technical barriers due to different standards, services, media, etc. It is associated with a series of different disciplines of digital signal processing, each of which can be applied independently. It includes a few different technical principles; distortion-rate theory, prediction techniques and control theory. The MPEG (Moving Picture Experts Group) video compression standard is based on this paradigm, thus, it contains a variety of different coding parameters which may result in different performance depending on their values. It specifies the bit stream syntax and the decoding process as its normative parts. The encoder details remain non-normative and are configured by a specific design. This means that the MPEG video encoder has a great deal of flexibility in the aspects of performance and implementation. This thesis deals with control techniques for the data rate of compressed video, which determine the encoding efficiency and video quality. The video rate control is achieved by adjusting quantisation step size depending on the occupancy of a transmission buffer memory which stores the compressed video data for a specific period of time.

Conventional video rate control techniques have generally been based either on linear predictive or on control theoretic models. However, this thesis takes a different view on digital video and MPEG video coding, and focuses on the non-stationary and nonlinear nature of realistic moving pictures. Furthermore, considering the MPEG encoding structure involved in the different disciplines, A series of improvements for video rate control are proposed, each of which enhances the performance of the MPEG encoder. A nonlinear quantisation control technique is investigated, which controls the buffer occupancy with the quantiser using a series of nonlinear functions. Linear and nonlinear feed-forward networks are also employed to control the quantiser. The linear combiner is used as a linear estimator and a radial basis function network as a nonlinear one. Finally, fuzzy rule-based control is applied to exploit the advantages of the nonlinear control technique which is able to provide linguistic judgement in the control mechanism. All these techniques are employed according to two global approaches (feed-forward and feedback) applied to the rate control. The performance evaluation is carried out in terms of controllability over bit rate variation and video quality, by conducting a series of simulations.
Declaration of originality

I hereby declare that this thesis and the work reported herein was composed and originated entirely by myself, in the Department of Electrical Engineering at the University of Edinburgh.

Yoo-Sok Saw
Acknowledgements

I would like to give my great thanks to Prof. Peter M. Grant for his continuous support, invaluable technical comments, encouragement, guidance and for detailed checking published papers and this thesis. I am also truly grateful to Dr. John M. Hannah for his continuous encouragement, detailed and critical technical comments and reviewing published papers and this thesis. Special thanks should go to Dr. Bernard Mulgrew for his precious contribution to this thesis and his kindness.

I also would like to thank the following people for their willing and kind assistance during the course of my PhD:

- Dr. Steve McLaughlin and Dr. Edward McDonnell for their general guidance about phd research.
- Dr. Iain Scott, Mr. Paul Strauch, Dr. Justin Fackrell, Dr. Achilleas Stogioglou and Dr. Eng-Siong Chng for their technical discussion and their time for linear and nonlinear prediction techniques.
- Mr. Paul Antoszczyszyn for his technical discussion and his time for model-based video coding.
- Mr. Jianzhou Wang for his technical discussion and his time for fuzzy logic control.
- Dr. John Thomson and Mr. Stephen Bates for proofreading this thesis.

I would like to thank the following organisations for providing continuous financial support during the course of my PhD:

- *The British Council.*
- *GoldStar Information and Communications Ltd., Korea.*
To my loving wife Eun-Jun and
our lovely son and daughter, Kun-Won and Jae-Won
# Contents

<table>
<thead>
<tr>
<th>List of Figures</th>
<th>vii</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>x</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>xi</td>
</tr>
<tr>
<td>List of Symbols</td>
<td>xiii</td>
</tr>
</tbody>
</table>

## 1 Introduction
1.1 Theme of the thesis ........................................... 1
1.2 Motivation .................................................... 1
1.3 Real-time video compression and video rate control ........... 3
1.4 Nonlinear approaches .......................................... 3
1.5 Thesis organisation ........................................... 5

## 2 Real-time video compression and video traffic management  

2.1 Introduction .................................................. 6
2.2 Real-time video compression techniques ....................... 7
   2.2.1 MC-DCT-based compression ................................ 8
   2.2.2 Fractal-based compression ................................ 9
   2.2.3 Wavelet-based compression ................................ 10
   2.2.4 Contextual-feature-based compression ................... 11
   2.2.5 Parallel video processing and full-frame transform coding 11
2.3 Transmission of compressed video via rate management ....... 12
2.4 A review of video traffic management techniques .............. 13
   2.4.1 A categorisation of video traffic management techniques 13
   2.4.2 Video traffic management techniques and traffic models for VBR 14
   2.4.3 Video rate control and buffering techniques for CBR and ABR 16
2.5 Conclusions ...................................................... 17

## 3 Foundation of video rate control for CBR
3.1 Introduction .................................................. 18
3.2 A comparative review of rate control mechanisms employed in compression standards 19
   3.2.1 MPEG1 ............................................. 19
   3.2.2 MPEG2 ............................................. 19
   3.2.3 MPEG4 and H.263 ................................. 20
   3.2.4 H.261 ............................................. 20
3.3 Rate control via adaptive quantisation ......................... 22
3.3.1 The buffer occupancy of an MPEG encoder ............................................. 24
3.3.2 Rate-distortion analysis on quantisation control ..................................... 25
3.4 Nonlinearity of the MPEG video encoding process ....................................... 25
3.5 Non-stationary nature of real-time video ..................................................... 27
  3.5.1 Classification of scene change features ................................................... 28
  3.5.2 Visual features of video sequences used in simulations ............................ 29
3.6 Two rate control approaches employed in the thesis .................................... 31
3.7 Rate and quantiser control schemes ............................................................. 33
3.8 Configuration for simulation ....................................................................... 35
3.9 Conclusions ................................................................................................. 36

4 Feed-forward rate control using scene change features ..................................... 37
  4.1 Introduction ................................................................................................. 37
  4.2 Variable bit rate video coding ..................................................................... 38
  4.3 Coding parameters associated with the rate control .................................... 41
    4.3.1 Channel rate .......................................................................................... 41
    4.3.2 Buffer occupancy .................................................................................. 42
  4.4 MPEG1 encoder using feed-forward rate control and scene change features .. 42
    4.4.1 Selected scene change features ............................................................. 43
    4.4.2 Video rate balance, \( R_b \) ..................................................................... 46
    4.4.3 Nonlinear quantisation control ............................................................. 48
    4.4.4 Local adaptation of the quantisation scale ............................................. 51
    4.4.5 Computational requirement .................................................................. 51
  4.5 Simulation results ........................................................................................ 53
    4.5.1 Simulation environment ......................................................................... 53
    4.5.2 Performance evaluation ......................................................................... 53
    4.5.3 Discussion .............................................................................................. 59
  4.6 Conclusion .................................................................................................... 59

5 Video rate control based on nonlinear rate estimation ...................................... 61
  5.1 Introduction ................................................................................................. 61
  5.2 An encoder model based on a stochastic interpretation ............................... 62
  5.3 Limitations of the rate control technique employed in the MPEG2 TM5 ........ 63
  5.4 System identification via linear prediction .................................................... 66
    5.4.1 Linear video rate estimator ..................................................................... 67
    5.4.2 Adaptation of the predictor coefficients ................................................ 70
    5.4.3 Simulation results .................................................................................. 71
  5.5 Configuration of the MPEG2 encoder based on a rate estimator ................... 76
    5.5.1 Selection and calculation of the scene change features ........................... 77
    5.5.2 Video rate estimators ............................................................................ 78
    5.5.3 Nonlinear quantisation control .............................................................. 79
  5.6 Rate-Distortion analysis of the quantiser control functions: sigmoidal and unimodal 82
    5.6.1 Background rate-distortion theory ........................................................ 82
    5.6.2 Rate-distortion functions ....................................................................... 83
    5.6.3 Analysis of \( D_x(R) \) and \( D_y(R) \) ......................................................... 84
5.6.4 Controllability over the occupancy fluctuation ........................................... 87
5.6.5 Simulations results ......................................................................................... 89
5.7 RBF rate estimator-based MPEG2 video encoder .............................................. 92
5.8 Simulation studies ............................................................................................. 94
5.8.1 Simulation setup ............................................................................................. 94
5.8.2 The number of RBF-network centres ........................................................... 95
5.8.3 Performance evaluation and discussion ......................................................... 96
5.8.4 Discussions .................................................................................................... 100
5.9 Conclusions ....................................................................................................... 101

6 Video rate control using fuzzy logic controller .................................................... 103
6.1 Introduction ........................................................................................................ 103
6.2 Fuzzy rule-based control for video rate control ................................................. 104
6.2.1 A brief review of major applications of FRC .............................................. 105
6.2.2 A basic FRC configuration for rate control ............................................... 106
6.2.3 Deriving a set of fuzzy rules for rate control .............................................. 107
6.2.4 Fuzzification and defuzzification ................................................................. 109
6.3 The fuzzy control parameters ............................................................................ 111
6.3.1 Scaling factors: $g_e$, $g_d$ and $g_o$ ............................................................... 111
6.3.2 Membership functions ................................................................................. 111
6.3.3 FAMs and control surfaces ......................................................................... 112
6.4 Improved FRC-based rate control algorithms ................................................. 114
6.4.1 Derivation of a dedicated fuzzy associative memory and its control surface 115
6.4.2 FRC using the scene change features (FRC-SC) ...................................... 116
6.4.3 Quality-monitored FRC based on PSNR and the occupancy (FRC-QM) 118
6.5 Simulation studies ............................................................................................. 122
6.5.1 Simulation environment ............................................................................... 122
6.5.2 Performance evaluation ............................................................................... 123
6.6 Conclusion ......................................................................................................... 136

7 Conclusions .......................................................................................................... 138
7.1 Introduction ........................................................................................................ 138
7.2 Major achievements of the work ..................................................................... 138
7.3 Advantages and limitations of the developed schemes .................................. 140
7.4 Limitations of the experiments ...................................................................... 141
7.5 Future of the research ..................................................................................... 141

References ............................................................................................................ 143

A Video sequences used in simulations .................................................................. 152

B Complementary results for Chapters 2 to 6 ...................................................... 158

C List of publications .............................................................................................. 180

D Publications ....................................................................................................... 181
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Classification of video rate and traffic management techniques.</td>
<td>14</td>
</tr>
<tr>
<td>3.1</td>
<td>Feedback buffer-based rate control scheme specified in MPEG.</td>
<td>22</td>
</tr>
<tr>
<td>3.2</td>
<td>Parameters for the buffer-based rate control.</td>
<td>23</td>
</tr>
<tr>
<td>3.3</td>
<td>Quantisation control function</td>
<td>23</td>
</tr>
<tr>
<td>3.4</td>
<td>Simplified video encoding process</td>
<td>26</td>
</tr>
<tr>
<td>3.5</td>
<td>A classification of the scene change features</td>
<td>29</td>
</tr>
<tr>
<td>3.6</td>
<td>Examples of the buffer occupancy fluctuation</td>
<td>31</td>
</tr>
<tr>
<td>3.7</td>
<td>Approach 1: feed-forward predictive scheme</td>
<td>32</td>
</tr>
<tr>
<td>3.8</td>
<td>Approach 2: fuzzy logic control</td>
<td>32</td>
</tr>
<tr>
<td>3.9</td>
<td>Rate control schemes employed in the research</td>
<td>34</td>
</tr>
<tr>
<td>3.10</td>
<td>Configuration of simulation</td>
<td>35</td>
</tr>
<tr>
<td>4.1</td>
<td>Video rate fluctuation at three channel transmission rates.</td>
<td>38</td>
</tr>
<tr>
<td>4.2</td>
<td>Bit rate and PSNR for various quantisation scale values for VBR (“Starwars “) (a) bit rate; (b) standard deviation of the bit rate; (c) PSNR; (d) standard deviation of PSNR</td>
<td>40</td>
</tr>
<tr>
<td>4.3</td>
<td>Bit rate and PSNR variations depending on the quantisation scale (“Starwars “) (a) Coded bits/frame; (b) PSNR</td>
<td>41</td>
</tr>
<tr>
<td>4.4</td>
<td>Feed-forward rate control scheme for MPEG1.</td>
<td>43</td>
</tr>
<tr>
<td>4.5</td>
<td>Frame variances $\var_{org}(k)$ and $\var_{dif}(k)$: (a) Cascaded; (b) Starwars; (c) Adverts</td>
<td>44</td>
</tr>
<tr>
<td>4.6</td>
<td>Directional motion vector function, $MVF_D$: (a) Cascaded; (b) Starwars; (c) Adverts</td>
<td>45</td>
</tr>
<tr>
<td>4.7</td>
<td>Framewise intra/inter decision using variances.</td>
<td>47</td>
</tr>
<tr>
<td>4.8</td>
<td>Occupancy for a 2-frame buffer at different transmission rates (“Football”)</td>
<td>49</td>
</tr>
<tr>
<td>4.9</td>
<td>Sigmoidal quantisation scale control.</td>
<td>50</td>
</tr>
<tr>
<td>4.10</td>
<td>Logarithmic-exponential quantisation scale control.</td>
<td>51</td>
</tr>
<tr>
<td>4.11</td>
<td>Variation in the quantisation scale value selected for 6 feed-forward buffering techniques: (a) Cascaded; (b) Starwars; (c) Adverts</td>
<td>56</td>
</tr>
<tr>
<td>4.12</td>
<td>Variation in the occupancy for 6 feed-forward buffering techniques: (a) Cascaded; (b) Starwars; (c) Adverts</td>
<td>57</td>
</tr>
<tr>
<td>4.13</td>
<td>Variation in PSNR for 6 feed-forward buffering techniques: (a) Cascaded; (b) Starwars; (c) Adverts</td>
<td>58</td>
</tr>
<tr>
<td>5.1</td>
<td>Rate control features in TM5: (a) Picture complexity measured as the product of the quantisation scale and the number of coded bits for the previous picture; (b) Normalised macro block activity and the resulting quantisation scale.</td>
<td>65</td>
</tr>
<tr>
<td>5.2</td>
<td>A basic linear predictor</td>
<td>67</td>
</tr>
<tr>
<td>5.3</td>
<td>Quantiser combined with the linear predictor</td>
<td>67</td>
</tr>
<tr>
<td>5.4</td>
<td>VBR MPEG2 encoder using the linear estimator</td>
<td>68</td>
</tr>
<tr>
<td>5.5</td>
<td>A basic linear estimator (Type I)</td>
<td>68</td>
</tr>
<tr>
<td>5.6</td>
<td>A linear predictor, Type II</td>
<td>69</td>
</tr>
<tr>
<td>5.7</td>
<td>A linear predictor, Type III</td>
<td>69</td>
</tr>
<tr>
<td>5.8</td>
<td>Type III predictor with recursive least square-based weight calculation</td>
<td>70</td>
</tr>
<tr>
<td>5.9</td>
<td>Three inputs, $\var_{org}(k)$, $\var_{dif}(k)$ and $\var_{type}(k)$: (a) Cascaded; (b) Starwars; (c) Adverts</td>
<td>72</td>
</tr>
<tr>
<td>5.10</td>
<td>Linear prediction Type III for “Starwars” (VBR): (a) Coded bits/frame; (b) MSE</td>
<td>73</td>
</tr>
</tbody>
</table>
5.11 Recursive linear prediction for “Starwars” (VBR): (a) Coded bits/frame; (b) MSE. 73
5.12 Linear prediction Type III for “Starwars” at 1280 kbit/s: (a) Coded bits/frame; (b) MSE. .................. 74
5.13 Recursive linear prediction for “Starwars” at 1280 kbit/s: (a) Coded bits/frame; (b) MSE. .................. 74
5.14 The structure of the nonlinear predictive rate control for MPEG2: (a) a functional representation; (b) the signal flow. 76
5.15 Quantiser control based on the nonlinear function surface. 79
5.16 Nonlinear quantiser control surfaces: (a) Sigmoidal ($S_{max} = 5$); (b) Sigmoidal ($S_{max} = 13$); (c) Unimodal ($S_{max} = 5$); (d) Unimodal ($S_{max} = 13$). 81
5.17 3-dimensional representation of $D_5(r)$ and $D_7(r)$. 86
5.18 3-dimensional representation of the video rate fluctuation. 88
5.19 The variation in the buffer occupancy for the quantiser control surfaces. 91
5.20 RBF predictor with 3 inputs and 9 taps. 93
5.21 MSE profile depending on the number of centres. 95
5.22 Performance of rate control algorithms (“Starwars”): (a) occupancy; (b) coded bits/frame; (c) PSNR. 98
5.23 Macroblock-wise variations in the quantisation scale and the buffer occupancy (“Starwars”): (a) TM5; (b) LIN; (c) SIGM; (d) LOGEXP; (e) RLS; (e) RBF. 99
5.24 Perceptual quality comparison (“JFK”): (a) TM5; (b) RBF; (c) adjacent pictures. 102

6.1 A basic model of fuzzy rule-based control. 106
6.2 Configuration of an MPEG2 video encoder with fuzzy logic rate control (FRC-R). 106
6.3 Configuration of the FRC-based rate control (FRC-R). 107
6.4 Fuzzy rule-based control parameters: (a) membership function (Type 0); (b) fuzzy associative memory map for deriving $La$. 109
6.5 A simplified defuzzification process. 110
6.6 Membership functions with irregular spacing: (a) wider spacing around the middle value (Type 1); (b) narrower spacing around the middle value (Type 2). 112
6.7 3-dimensional representation of the fuzzy associative memory: (a) LIN; (b) EXP; (c) LOG; (d) SGM. 113
6.8 Corresponding control surfaces to the fuzzy associative memory in Fig.6.7: (a) LIN; (b) EXP; (c) LOG; (d) SGM. 114
6.9 The fuzzy associative memory and its control surface used for performance improvement: (a) Allocation of fuzzy sets; (b) resulting FAM; (c) 3 dimensional representation; (d) control surface. 116
6.10 Scene change-based FRC for MPEG2 video rate control. 117
6.11 Configuration of the scene change-based FRC (FRC-SC). 117
6.12 Configuration of the FRC based on macro-block-wise video rate balance (FRC-B). 119
6.13 Configuration of the quality-monitored FRC (FRC-QM). 121
6.14 Detailed block diagram of FRC-QM. 122
6.15 Distribution of fuzzy sets used for $ge, gd = 4, 16$ in the FRC-R scheme: (a) “Cascaded”; (b) “Starwars”. 124
6.16 The effect of varying scaling factors in FRC-R (“Starwars”): (a) occupancy; (b) coded bits/frame; (c) PSNR. 125
6.17 Performance comparison for the different membership functions in FRC-R (“Starwars”: (a) occupancy; (b) coded bits/frame; (c) PSNR. 126
6.18 Performance comparison for different fuzzy associative memory-control surfaces in the FRC-R (“Starwars”): (a) occupancy; (b) coded bits/frame; (c) PSNR. 128
6.19 Performance evaluation for FRC-B (“Starwars”): (a) occupancy; (b) coded bits/frame; (c) PSNR. 130
6.20 Performance evaluation for FRC-Q (“Starwars”): (a) coded bits/frame; (b) PSNR. 131
6.21 Performance comparison between FRC-Q and VBR (“Starwars”: (a) coded bits/frame; (b) PSNR. 132
6.22 Performance profiles for FRC-R, FRC-SC and FRC-QM (“Starwars”): (a) occupancy; (b) coded bits/frame; (c) PSNR. 135
6.23 Perceptual quality comparison ("JFK"): (a) FRC-R; (b) FRC-QM; (c) adjacent pictures. ........................................ 137
A.1 Video sequence “Cascaded” (Miss America - Football - Susie) ................................. 153
A.2 Video sequence “Starwars” .................................................................................... 154
A.3 Video sequence “Adverts” ...................................................................................... 155
A.4 Video sequence “JFK” .......................................................................................... 156
A.5 Video sequence “Topgun” ...................................................................................... 157
B.1 Entropies of luminance (Y) difference of video sequences used in simulations .... 158
B.2 Bit rate and PSNR means depending on the quantisation scale for VBR ("JFK") .... 160
B.3 Bit rate and PSNR variations depending on the quantisation scale for VBR
("JFK") ..................................................................................................................... 161
B.4 Number of intra-coded macro blocks in P and B pictures (JIF(k)) ......................... 165
B.5 Mean value of the quantisation scale per slice ......................................................... 166
B.6 The number of coded bits per frame with rate control ........................................... 167
B.7 Examples of buffer occupancy in the MPEG1 video encoder ................................ 168
B.8 The PSNR profile for two different bit rates ........................................................... 169
B.9 Linear prediction for “Adverts” (VBR) ................................................................... 170
B.10 Recursive linear prediction for “Adverts” (VBR) .................................................. 170
B.11 Linear prediction for “Adverts” at 1280 kbit/s ...................................................... 171
B.12 Recursive linear prediction for “Adverts” at 1280 kbit/s ..................................... 171
B.13 Performance of rate control algorithms with linear or nonlinear rate estimators
("Cascaded") .............................................................................................................. 172
B.14 Performance of rate control algorithms with linear and nonlinear rate estimators
("JFK") ....................................................................................................................... 173
B.15 Performance comparison between FRC-Q and VBR (IQ:FQ:BQ = 4:5:12) ("Starwars") ................................................................. 175
B.16 Performance comparison for FRC-R, FRC-SC and FRC-QM ("Cascaded") ........ 177
B.17 Performance comparison for FRC-R, FRC-SC and FRC-QM ("JFK") .................... 178
B.18 Performance comparison for FRC-R, FRC-SC and FRC-QM ("Topgun") ............. 179
# List of Tables

3.1 Summary of visual features of the video sequences. ........................................... 30

4.1 Summary of the performance of the six rate control techniques: (a) Cascaded; (b) Starwars; (c) Adverts. ................................................................. 54

5.1 Comparison of MSE values in dB between the Type III and the RLS-adapted estimator for “Starwars” and “Adverts” sequences. ........................................ 75

5.2 Mean and standard deviation of the performance measures for the quantiser control surfaces ..................................................................................... 90

5.3 Mean MSE and mean variances of video sequences. ........................................... 95

5.4 Mean and standard deviation of performance measures (“Starwars”). .......... 97

6.1 Fuzzy logic variables for the fuzzy control input and resulting output. .......... 108

6.2 The effect of varying scaling factors for “Starwars” in the FRC-R scheme. .... 124

6.3 Performance comparison for different membership functions in FRC-R (“Starwars”). .............................................................................................. 127

6.4 Performance comparison for different fuzzy associative memory-control surfaces in FRC-R (“Starwars”). ................................................................. 127

6.5 Performance comparison between FRC-Q and VBR at different PSNR target values (“Starwars”). ................................................................. 132

6.6 Performance comparison among FRC-R, FRC-SC and FRC-QM (“Starwars”). 133

B.1 Bit rate and PSNR means depending on the quantisation scale for VBR (“Starwars”). .............................................................................................. 159

B.2 Bit rate and PSNR means depending on the quantisation scale (“JFK”). ...... 162

B.3 Computational requirement of the main profile MPEG1 encoder. ............... 163

B.4 Computational requirement for calculating the scene change features. ...... 164

B.5 Mean and standard deviation of performance measures for feed-forward rate control techniques (“Cascaded”). ................................................. 174

B.6 Mean and standard deviation of performance measures for feed-forward rate control techniques (“JFK”). ......................................................... 174

B.7 Performance comparison for FRC-R, FRC-SC and FRC-QM (“Cascaded”). 176

B.8 Performance comparison for FRC-R, FRC-SC and FRC-QM (“JFK”). ......... 176

B.9 Performance comparison for FRC-R, FRC-SC and FRC-QM (“Torgun”). .... 176
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABR</td>
<td>Available bit rate</td>
</tr>
<tr>
<td>ADSL</td>
<td>Asymmetrical digital subscriber loop</td>
</tr>
<tr>
<td>AR</td>
<td>Auto-regressive</td>
</tr>
<tr>
<td>ARMA</td>
<td>Autoregressive moving average</td>
</tr>
<tr>
<td>ATM</td>
<td>Asynchronous transfer mode</td>
</tr>
<tr>
<td>B</td>
<td>Bi-directionally predicted picture</td>
</tr>
<tr>
<td>B-ISDN</td>
<td>Broadband ISDN</td>
</tr>
<tr>
<td>BPF</td>
<td>Coded bits per picture frame</td>
</tr>
<tr>
<td>CBR</td>
<td>Constant bit rate</td>
</tr>
<tr>
<td>CIF</td>
<td>Common intermediate format</td>
</tr>
<tr>
<td>DCT</td>
<td>Discrete cosine transform</td>
</tr>
<tr>
<td>IDCT</td>
<td>Inverse discrete cosine transform</td>
</tr>
<tr>
<td>IQ</td>
<td>Inverse quantiser</td>
</tr>
<tr>
<td>EXP</td>
<td>Exponential quantiser control function/exponential FAM</td>
</tr>
<tr>
<td>FAM</td>
<td>Fuzzy associative memory</td>
</tr>
<tr>
<td>FRC</td>
<td>Fuzzy rule-based control</td>
</tr>
<tr>
<td>HDSL</td>
<td>High-speed digital subscriber line</td>
</tr>
<tr>
<td>HDTV</td>
<td>High definition TV</td>
</tr>
<tr>
<td>I</td>
<td>Intra-coded picture</td>
</tr>
<tr>
<td>IFS</td>
<td>Iterated function system</td>
</tr>
<tr>
<td>ISDN</td>
<td>Integrated Services Digital Networks</td>
</tr>
<tr>
<td>ISO</td>
<td>International Standard Organisation</td>
</tr>
<tr>
<td>ITU</td>
<td>International Telecommunications Union</td>
</tr>
<tr>
<td>IVOD</td>
<td>Interactive VOD</td>
</tr>
<tr>
<td>KLT</td>
<td>Karhunan-Loeve transform</td>
</tr>
<tr>
<td>LIN</td>
<td>Linear quantiser control function/linear FAM</td>
</tr>
<tr>
<td>LOG</td>
<td>Logarithmic quantiser control function/logarithmic FAM</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>Meaning</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>LOGEXP</td>
<td>Logarithmic/exponential quantiser control function</td>
</tr>
<tr>
<td>LOGEXP-A</td>
<td>Adaptive quantiser control function based on LOGEXP</td>
</tr>
<tr>
<td>LS</td>
<td>Least squares</td>
</tr>
<tr>
<td>MA</td>
<td>Moving average</td>
</tr>
<tr>
<td>MC</td>
<td>Motion-compensated</td>
</tr>
<tr>
<td>MIMD</td>
<td>Multiple Instruction and Multiple Data</td>
</tr>
<tr>
<td>MPEG</td>
<td>Motion Picture Experts Group</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean square error</td>
</tr>
<tr>
<td>MUX</td>
<td>Multiplexer</td>
</tr>
<tr>
<td>NFVR</td>
<td>Normalised fluctuation of video rate</td>
</tr>
<tr>
<td>NQC</td>
<td>Nonlinear quantisation control</td>
</tr>
<tr>
<td>OLS</td>
<td>Orthogonal least square</td>
</tr>
<tr>
<td>P</td>
<td>Predicted picture</td>
</tr>
<tr>
<td>PCM</td>
<td>Pulse code modulation</td>
</tr>
<tr>
<td>PDH</td>
<td>Plesiochronous digital hierarchy</td>
</tr>
<tr>
<td>PSNR</td>
<td>Peak signal-to-noise ratio</td>
</tr>
<tr>
<td>Q</td>
<td>Quantiser</td>
</tr>
<tr>
<td>RBF</td>
<td>Radial basis function</td>
</tr>
<tr>
<td>RE</td>
<td>Video rate estimator</td>
</tr>
<tr>
<td>RLC</td>
<td>Run length coding</td>
</tr>
<tr>
<td>RLS</td>
<td>Recursive least squares</td>
</tr>
<tr>
<td>RM</td>
<td>Reference model</td>
</tr>
<tr>
<td>SCC</td>
<td>Scene change calculator</td>
</tr>
<tr>
<td>SDH</td>
<td>Synchronous digital hierarchy</td>
</tr>
<tr>
<td>SIF</td>
<td>Standard input format</td>
</tr>
<tr>
<td>SIGM</td>
<td>Sigmoidal quantiser control function/sigmoidal FAM</td>
</tr>
<tr>
<td>SIGM-A</td>
<td>Adaptive quantiser control based on SIGM</td>
</tr>
<tr>
<td>SIMD</td>
<td>Single Instruction Multiple Data</td>
</tr>
<tr>
<td>SM</td>
<td>Simulation model</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-noise ratio</td>
</tr>
<tr>
<td>STFT</td>
<td>Short-time Fourier transform</td>
</tr>
<tr>
<td>TM5</td>
<td>Test model 5</td>
</tr>
<tr>
<td>TMN5</td>
<td>Test model near 5</td>
</tr>
<tr>
<td>VBR</td>
<td>Variable bit rate</td>
</tr>
<tr>
<td>VBV</td>
<td>Video buffering verifier</td>
</tr>
</tbody>
</table>
Abbreviations

**VLC**  Variable length coding

**VOD**  Video-on-demand

**VM**  Verification model

**WHT**  Walsh-Hadamard transform
List of principal symbols

\( b_{\text{size}} \)  
buffer size

\( b(B) \)  
the number of bits generated by a B picture

\( b(I) \)  
the number of bits generated by a I picture

\( b(P) \)  
the number of bits generated by a P picture

\( \text{BQ} \)  
fixed quantisation scale for B pictures

\( cbf(k) \)  
video rate in bits for the \( k \)th picture

\( \hat{cbf}(k) \)  
estimation of \( cbf(k) \)

\( c_{\text{rate}} \)  
channel rate

\( \text{COL} \)  
the number of pixels per column

\( d(n) \)  
\( e(n) - e(n - 1) \)

\( d_{\text{target}} \)  
the maximum tolerable coding delay

\( D(R) \)  
rate-distortion function

\( D(R)_G \)  
rate-distortion function for Gaussian sources

\( D(R)_N \)  
rate-distortion function for non-Gaussian sources

\( D_S(r, R_b) \)  
rate-distortion function for the sigmoidal function

\( D_U(r, R_b) \)  
rate-distortion function for the unimodal function

\( D_S(R) \)  
rate-distortion function for the sigmoidal function

\( D_U(R) \)  
rate-distortion function for the unimodal function

\( e(n) \)  
error in estimation between \( \hat{y}(n) \) and desired signal \( y(n) \) (Chapter 5) / difference between the target value and the feedback fuzzy control value (Chapter 6)

\( f_{LS}() \)  
fluctuation function for the sigmoidal quantiser control function

\( f_{LU}() \)  
fluctuation function for the unimodal quantiser control function

\( g_{\varepsilon} \)  
the learning rate for \( \kappa \)-means clustering algorithm

\( gd(n) \)  
scaling factor for the fuzzy variable \( d(n) \)

\( gc(n) \)  
scaling factor for the fuzzy variable \( e(n) \)

\( go(n) \)  
scaling factor for the fuzzy variable \( o(n) \)

\( Gd(n) \)  
scaled crisp value of the fuzzy variable \( d(n) \)

\( Ge(n) \)  
scaled crisp value of the fuzzy variable \( e(n) \)
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G_0(n)$</td>
<td>scaled crisp value of the fuzzy variable $o(n)$</td>
</tr>
<tr>
<td>$G_{RLC}$</td>
<td>run length coding gain</td>
</tr>
<tr>
<td>$\gamma_r^2$</td>
<td>spectral flatness measure</td>
</tr>
<tr>
<td>$H_D$</td>
<td>entropy for difference picture</td>
</tr>
<tr>
<td>$H_N(x)$</td>
<td>$N$-block entropy for a vector $x$</td>
</tr>
<tr>
<td>IQ</td>
<td>fixed quantisation scale for I pictures</td>
</tr>
<tr>
<td>$L_d(n)$</td>
<td>linguistic judgement for $Gd$</td>
</tr>
<tr>
<td>$L_e(n)$</td>
<td>linguistic judgement for $Ge$</td>
</tr>
<tr>
<td>$Lo(n)$</td>
<td>linguistic judgement for $Go$</td>
</tr>
<tr>
<td>MBF</td>
<td>the mean number of bits allocated per picture</td>
</tr>
<tr>
<td>MBN</td>
<td>the number of macro blocks per picture</td>
</tr>
<tr>
<td>MBS</td>
<td>the number of macro blocks per slice</td>
</tr>
<tr>
<td>$MV F_D(s)$</td>
<td>directional motion vector function</td>
</tr>
<tr>
<td>$MV F_N(s)$</td>
<td>non-directional motion vector function</td>
</tr>
<tr>
<td>$MV F(s)$</td>
<td>motion vector function at the slice $s$</td>
</tr>
<tr>
<td>$\mu()$</td>
<td>fuzzy logic membership function</td>
</tr>
<tr>
<td>$o(n)$</td>
<td>defuzzified control input to the target system</td>
</tr>
<tr>
<td>$O(k, n)$</td>
<td>buffer occupancy at the frame $k - 1$ and the macro block $n$</td>
</tr>
<tr>
<td>$O_{-set}$</td>
<td>target value of the buffer occupancy</td>
</tr>
<tr>
<td>$p_{-rate}$</td>
<td>picture rate</td>
</tr>
<tr>
<td>$p_{type}(k)$</td>
<td>picture type for the $k$th picture</td>
</tr>
<tr>
<td>PQ</td>
<td>fixed quantisation scale for P pictures</td>
</tr>
<tr>
<td>$q_S$</td>
<td>quantiser step size derived from the sigmoidal control function</td>
</tr>
<tr>
<td>$q_U$</td>
<td>quantiser step size derived from the unimodal control function</td>
</tr>
<tr>
<td>$Q_f$</td>
<td>quantisation parameter defined in MPEG2 TM5</td>
</tr>
<tr>
<td>$Q_s(k, n)$</td>
<td>normalised quantisation scale</td>
</tr>
<tr>
<td>$Q_{-set}$</td>
<td>target PSNR value for the fuzzy logic control</td>
</tr>
<tr>
<td>$r(t)$</td>
<td>the buffer occupancy at time $t$ specified as the reaction parameter in TM5</td>
</tr>
<tr>
<td>$R_b$</td>
<td>video rate balance</td>
</tr>
<tr>
<td>$R_T$</td>
<td>channel rate in bits per pixel</td>
</tr>
<tr>
<td>ROW</td>
<td>the number of pixels per row</td>
</tr>
<tr>
<td>$S_{max}$</td>
<td>torsion factor</td>
</tr>
<tr>
<td>$\sigma_r^2$</td>
<td>short-term source variance</td>
</tr>
<tr>
<td>$\sigma_{r\Delta t}$</td>
<td>short-term source variance</td>
</tr>
<tr>
<td>$\sigma_r^2$</td>
<td>long-term variance</td>
</tr>
<tr>
<td>$\text{var}_{\Delta f}(k)$</td>
<td>variance of the difference picture between $k$ and $k - 1$</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>$\text{var}_\text{org}(k)$</td>
<td>variance of an input picture $k$</td>
</tr>
<tr>
<td>$V R(k)$</td>
<td>variance ratio for the $k$th picture</td>
</tr>
<tr>
<td>$w_i$</td>
<td>$i$th linear weight used in the RLS algorithm</td>
</tr>
<tr>
<td>$\mathbf{x}_j$</td>
<td>the $j$th RBF centre vector</td>
</tr>
<tr>
<td>$y(n)$</td>
<td>desired signal at time $n$</td>
</tr>
<tr>
<td>$\hat{y}(n)$</td>
<td>estimation of $y(n)$</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Theme of the thesis

The evolution of digital communication networks has opened a new era of multiple and simultaneous exchange of audio-visual and computerised information, i.e. multimedia communication. This enables users to exchange or share information using many different media - voice, audio, graphics, documents, image, video, etc. Among these media, video is an emerging information carrier since it can efficiently convey a large amount of information. However, the cost of storing or transmitting video is very high due to its inherently large requirement of data rate, hence the need for high efficiency encoding rises. Video compression or video coding is an economical solution to this. Some people anticipate that the need for data compression will reduce as wide-band fibre optic cables are increasingly deployed for end users. However, video compression is expected to be important, even after the wide realisation of the fibre optic-based network.

Data compression is still required for most broadband ISDN (B-ISDN) services, mobile communication networks and digital storage media. Furthermore, the future high-definition digital television standards need compression due to their considerably higher bandwidth requirement than current digital TV standards. Thus, transmission of digital video inevitably introduces compression requirements, and transmission technology should deal with compressed digital video rather than raw video. In this thesis, the transmission aspects of compressed video are investigated by studying a series of new techniques for achieving better video rate control. The evaluation is conducted via simulations on video traffic comprising MPEG video bit streams.

Rate control techniques for compressed video are primarily discussed and assessed. The main contribution of this work is in the investigation of nonlinear techniques which improve the performance of video rate control algorithms for MPEG (Moving Picture Experts Group) video encoders. Both a radial basis function (RBF) network and fuzzy logic control methods are employed as the nonlinear techniques. A series of nonlinear functions are also proposed for enhancing the quantiser control. All these techniques are designed to operate with scene change features which are used to inform the video encoder of the extent of scene change. In the following chapters, the background of the theoretical analysis and the simulations will be presented.

1.2 Motivation

The infrastructure for visual communication is the digital network [1]. They are capable of providing a wide range of services from narrow bandwidth voice to wider bandwidth variable bit
rate video. Video data transmission is generally viable over constant bit rate (CBR) channels. In this channel mode, a transmission buffer memory is widely used to regulate the video data rate, thus, there exists a possibility of buffer malfunctions such as overflow and underflow due to the variable data rate nature of compressed video. Variable bit rate (VBR) channels can also be used for B-ISDN services [2] in which the buffer is not necessary since the network can accommodate the video source regardless of its data rate. However, network congestion is highly likely due to the peak data rate of video sources exceeding the network capacity. To avoid the buffer malfunction and the congestion, it is desirable to regulate the video rate within a specified range. A variety of audiovisual services demand a certain level of freedom to use channel bandwidth depending on the required data rate for achieving a satisfactory transmission picture quality. In particular, audiovisual conferencing or interactive video-on-demand (IVOD) [3] may often contain video with rapid change, which causes much wider variation in video rate than those of the simpler video telephony systems. The situation could become worse when the IVOD uses low bit rate transmission media such as ADSL (asymmetrical digital subscriber loop) or the HDSL (high-speed digital subscriber line) in comparison to ATM-based B-ISDN channels [4].

Feedback control technique is the standard rate control algorithm employed in major video compression standards. This technique, which uses a fixed size buffer, performs well when the compressed video contains data with stationary statistics. Otherwise, it may experience malfunctions such as buffer overflow, underflow or drastic fluctuation in the buffer occupancy. This case occurs when the buffer cannot adequately accommodate the incoming video rate variation. Attention is focused on dramatic scene changes and rapid motion as the cause of the abnormal buffer operation and subsequent quality degradation in the viewed output video signal. In particular, bi-directional communication is primarily dealt with since this application highly demands effective rate control techniques. Conventional approaches apply linear predictive techniques to designing the buffer-based rate control algorithm. They include representative evaluation models for the video compression standards such as ITU-H.261 Reference Model 8, MPEG2 Test Model 5, MPEG4 Verification Model, etc. Many of recently published schemes also appear to be based on one of these models or their modified versions. In this thesis, the performance of nonlinear techniques is also assessed on video data which possesses non-stationary statistical properties.

Firstly, nonlinear equations are used for the mapping between the buffer occupancy and the quantisation step size since linear methods do not adequately control the quantiser for the rapid scene changes. Secondly, an approach of deriving scene change features from the front end processing of the video encoder is applied so that the rate control mechanism can take early action to handle rapid video rate variation with the resulting quality variation distributed over several adjacent pictures. These two approaches create a linear feed-forward rate control scheme. Thirdly, the nonlinear time-series prediction technique is employed as an alternative to the linear technique. An RBF-network rate estimator is used to perform the rate prediction.
Finally, a fuzzy rule-based control technique is utilised as an advanced feedback control which may outperform conventional feedback control techniques.

1.3 Real-time video compression and video rate control

Real-time video generally refers to moving pictures with television frame rate around 25 or 30 frames/s which is the minimum picture display speed to provide the same smoothness as real-life motion. Real-time processing of this type of video requires low coding and transmission delay in order to avoid objectionable quality degradation, particularly for bi-directional applications. Thus, the delay needs to be as short as possible. In contrast to this, the buffer size, which determines the delay, has to be as big as possible to prevent interruption to normal transmission due to buffer malfunctions such as overflow and the subsequent loss of picture information. These contradictory requirements need to be resolved by tradeoff between the delay and the quality. Thus, the goal of video rate control is to maintain the buffer size or the delay within a specified tolerance range and provide an optimum peak signal-to-noise ratio (PSNR), which is considered as one of video quality measures, within the budget of available transmission bit rate. Since a certain duration of delay is allowed within the specified delay requirement, the buffer size equal to the maximum allowed delay can be used as the fixed buffer size. This thesis adopts this methodology for the permitted buffer size, and buffer malfunctions are prevented and better PSNR is achieved.

1.4 Nonlinear approaches

Three nonlinear approaches are applied to the MPEG video rate control under the condition that the scene change features are provided to the video encoder. The scene change features are intra-frame, inter-frame variances and a series of MPEG video encoding parameters.

Scene change features:
The scene change features are the numerical representation of the change in visual information. They can be used to analyse the complexity of input pictures and also to enhance the performance of rate control techniques. The American high definition TV system ("Grand Alliance") [5] appeared to employ scene change features (the variance of motion compensated frame difference) to improve its rate control performance. It is an MPEG2-based all digital system for high quality terrestrial video and audio transmission over a 6 MHz channel which can deliver transmission bit rates of 17 to 20 Mbits/s. This rate can be achieved by employing 50:1 compression. The video rate after a scene change can be several times greater than the transmission rate in this system. The situation in MPEG video (SIF format) transmission is similar to this in respect of rate control and the scene change features when the transmission is carried out through channels with around 1 Mbits/s rate and the full rate of the SIF video reaches some 30 Mbits/s. The scene change features are considered as effective a priori information measures for the input of a rate control algorithm.
Nonlinear time-series prediction:
Time series prediction is one of the established signal processing areas, which has traditionally been treated in linear analytic ways [6]. A basic assumption here is that the signal falls into the category of linearity, stationarity with second-order statistics, with particular emphasis on Gaussian statistics. However, nonlinear prediction and estimation techniques do not have these assumptions as a premise, since realistic signals often contain very different statistical properties. Like other nonlinear predictors such as Volterra series [7] and neural networks based on multilayer perceptrons [8], it has been shown that the RBF-network can universally approximate any function or time series [9, 10] on the basis of functional approximation theory [11]. Since its first introduction as a means of signal interpolation by Broomhead [12], many applications such as channel equalisation [13], chaotic signal prediction [14], etc. have benefited from RBF-based signal modelling. The approximation capability can also be applied to the prediction of non-stationary video rate time series. Thus, one of the novel architectures applied in this thesis is the nonlinear predictive video rate estimator using the RBF network. The RBF network is often classified as a special form of neural network [15] with a smaller number of hidden layers. It also has an advantage in the aspects of implementation, as it is a massively parallel structure [16–18]. In this thesis the RBF-network is used for rate estimation, which is designed to adaptively predict the video rate depending on the scene change features. It was used as a video rate estimator to anticipate the one-frame-ahead video rate information by using scene change feature data as input.

Fuzzy logic control:
Fuzzy set theory opens a new way of applying linguistic judgement to control engineering systems by introducing nonlinearity associated with the process of human reasoning. Applications such as the fluid level control in a reservoir [19] have already been shown to achieve substantial improvement compared to conventional linear control. In video rate control, the video quality should also be viewed as a critical control variable together with the video rate. To achieve improvement, both quality monitoring techniques and the scene change features were used to control the parameters of the fuzzy logic control scheme. Fuzzy logic control techniques can also be applied to feed-forward approaches such as the adaptive-network-based fuzzy inference systems [15] and fuzzy basis function networks [20, 21]. However, since this feed-forward approach is theoretically equivalent to the RBF-network approach [15], the fuzzy basis function approach is not investigated in this thesis.

Nonlinear quantiser control:
Another nonlinear technique employed in this thesis is quantiser control using nonlinear functions. The quantiser control in MPEG video encoders can be viewed as mapping the buffer occupancy to the quantisation step size. Whilst the conventional quantiser control scheme is based on linear methods, here nonlinear quantiser control is used by employing a sigmoidal and unimodal functions. Analytic and experimental investigations are presented with performance comparisons.
1.5 Thesis organisation

After this brief introduction, an overview of real-time video compression and rate control techniques is given in order to provide an insight into this thesis. In chapter 2 much of the necessary background is presented about representative video compression techniques and video rate and traffic control techniques. The content includes fractal-based, wavelet-based and model-based compression techniques, VBR video traffic modelling and video traffic management techniques.

In Chapter 3, after reviewing rate control algorithms specified in international video compression standards, rate-distortion theoretic aspects are discussed within the framework of the buffer-based rate control. In order to widen the understanding of this thesis, the nonlinear and non-stationary nature of the MPEG video coding is considered. Finally, the approaches adopted in this thesis are briefly introduced.

Chapter 4 explores the MPEG video coding parameters which are used for video rate control and explains the devised adaptive technique based on the linear prediction and the nonlinear quantisation control. The linear prediction is a one-step ahead control using previous history of the video rate and the scene change features. The two nonlinear quantisation control functions are evaluated in simulations. The performance of this new scheme is compared with different quantiser control functions.

Chapter 5 proposes the RBF-network-based rate control algorithm after presenting the analysis of the linear relationship between the scene change features and the video rate using least squares methods. With this linear configuration, the difference of the RBF-based scheme is highlighted in respect of the network parameters. The nonlinear quantiser control functions are revisited by describing their 3-dimensional versions and are analysed in a rate-distortion theoretic way. The performance of the RBF-based scheme is contrasted with that of the linear estimators and some conclusions are drawn.

Chapter 6 describes the fuzzy rule-based video rate control scheme. Firstly, a basic model of fuzzy rule-based control is derived from conventional applications. This model functions as a performance reference. Secondly, the effect of fuzzy logic control parameters is examined on the video rate and PSNR. After selecting appropriate values and conditions the parameters for rate control purposes enhanced algorithms are proposed. Performance is evaluated in comparison to the basic model. In chapter 7 the conclusions of this work are summarised and issues worthy of further work are identified.
Chapter 2

Real-time video compression and video traffic management

2.1 Introduction

Digital video includes a variety of different digitised representations of moving visual information: computer graphics, animation, entertainment videos, etc. Visual representations differ from one another in respect of spatial resolution, quantisation resolution and temporal (picture) rate. They have different data rates depending on the required quality and the mode of access: storage or transmission. Real-time digital video contains moving images which need several tens of Mbits/s or more capacity [22] since it has to be displayed at a specified frame rate. Thus, the need for real-time compression is inevitable for moving images.

The next process after compression is either storage or transmission in which a specific medium is used. Here, a critical parameter, which has a substantial effect on the compression process, is the bandwidth of the medium, that is, channel capacity available for conveying the compressed video bit stream. When the video bit stream is being transmitted, a few different modes of channel connection can be employed in plesiochronous digital hierarchy (PDH) networks and asynchronous transfer mode (ATM)-based synchronous digital hierarchy (SDH) networks [1] using constant bit rate (CBR), variable bit rate (VBR) and available bit rate (ABR), etc. The CBR mode has been used for all PDH channels, ISDN channels and specified services of SDH channels. The VBR and ABR modes are selected to deliver services for inherently variable rate sources such as video and computer communication traffic. The VBR mode is known to transmit variable rate video without quality variation, and the ABR is a newly proposed mode which adaptively alters available channel rate allocated to information sources depending on the network condition.

The highly variable data rate of compressed video requires to be controlled since the storage or transmission medium may not be able to accommodate the peak rate. For the CBR mode, since the channel rate is fixed, the video rate must be constantly controlled and limited within the fixed channel capacity. For the VBR mode, on the other hand, the quality of video is guaranteed by allocating the necessary bandwidth to achieve a given desired quality. However, network congestion may occur if the total bandwidth demand from multiple compressed video sources exceeds the channel capacity [2]. In this case, video rate control strategy must be
considered in order to maintain the video quality in an operational connection. The ABR mode has been introduced to alleviate such a risk of congestion by controlling the input source rate before it arrives at multiplexers installed in the network interface [23]. The video quality of CBR applications is considered to be highly dependent upon video rate management techniques since the channel rate cannot be changed during communication. In this research, rate control techniques were extensively investigated, which are based on a transmission buffer and quantiser for CBR applications. However, the VBR and ABR modes, to a certain extent, also need to adopt similar rate management techniques, in that existing video traffic control techniques have limited performance for non-stationary video traffic.

This chapter aims to

- provide an insight into the background of real-time video compression and the related transmission techniques, and
- broaden understanding of video rate management and buffering techniques, which is the main topic in subsequent chapters.

First, representative video compression algorithms will be briefly reviewed in Section 2.2. Section 2.3 will highlight several video rate management techniques as milestone schemes. Finally, video traffic modelling and rate control techniques will be introduced in Section 2.4. Section 2.5 concludes this chapter.

### 2.2 Real-time video compression techniques

Real-time video compression begins with digitising analogue video input generally coming from cameras, which forms standard analogue TV signals; PAL, NTSC and SECAM [24], etc. The international standard, ITU-R 601 (formerly CCIR 601), specifies the technical details of the digitisation process of these TV signals. The resulting output meets the requirement of high quality such as studio quality video. The digitised format of the TV signals is the actual input video format for video encoders, which is eventually converted into a common format for video compression and transmission such as ITU-T H.261 CIF (common intermediate format) [25], MPEG SIF (standard input format) [26], etc. The input video format of high definition TV (HDTV) differs from ITU-R 601, and its formats have yet to be standardised, that is, there has been no common digital video format such as ITU-R 601, for various HDTV standards. After formatting digital video, a variety of different compression (or coding) algorithms are used to convert the input video to binary bit streams which can be transmitted or stored via digital media. The compression is the process to reduce statistical redundancy contained in the input video.

The range of applications for video compression is broad. It includes three different service categories; bi-directional communication (video telephony, teleconferencing and video email), uni-directional, i.e. broadcast-type communication (remote lecturing, electronic news gathering and surveillance) and interactive communication (video-on-demand and home shopping [3]).

7
Digital video compression has been applied to commercial audio-visual communication since the 1960’s. For more than three decades, a variety of different compression techniques have been used; motion-compensated coding [27], transform coding [28], adaptive quantisation [29], vector quantisation [30], fractal-based coding [31], wavelet-based coding [32], contextual-feature-based (model-based and object-based) coding [33], etc. Since Schreiber [34] reviewed the properties of the human visual system (HVS) and TV, many review papers have been published; a general review on transform coding [35], colourimetry, TV standards, colour perception models [36], perceptual distortion measures and transform coding [37], a thorough review of waveform-based coding [38], a review of perceptual waveform coding [39], very low bit-rate coding [33], etc.

The nature and environment of digital video applications are known to be heterogeneous [40] in terms of video quality, system platform, communication medium, etc. However, what video services eventually seek is exchange of visual information where compatibility must be guaranteed as the first priority. In order to achieve this, the processes of compression and decompression should be standardised so that communication can be established without technical barriers due to different makes and different services. Such an environment is thought to have led compression algorithms to a common framework: motion-compensated (MC) transform-based coding. The most widely used transform is the discrete cosine transform (DCT). It has been used for many video compression algorithms for several decades, and is currently being adopted in many international standards; JPEG (Joint Photographic Experts Group), MPEG, ITU-T H.260 series and some of the HDTV specifications. Main stream video communication applications (video telephony and video-on-demand) are also based on the MC-DCT algorithm such as ITU-T H.261 [25], MPEG1 [26] and MPEG2 [41]. This section will provide a brief review of these compression techniques.

2.2.1 MC-DCT-based compression

The MC-DCT is a form of hybrid coding which combines DCT and motion compensation. Since digital video has correlation between pictures, the inter-frame motion can be represented by motion vectors to quantify transitional motion, the motion compensation reduces statistical redundancy in the temporal direction. The DCT is then applied to the residual information after the motion compensation [26, 41]. If the temporal correlation is not high enough to be efficiently encoded by the motion compensation, the input picture is coded solely by the DCT without the MC, i.e. intra-only mode which only takes into account the intra-frame redundancy. This implies that the MC-DCT is an adaptive coding technique to switch the operating mode to make the most of the redundancy-based coding scheme. For the MC algorithm, a block matching algorithm and its derivatives [42] are predominantly used in the standards. Different transforms can also be used instead of the DCT, however, the DCT is known to provide superior performance to other non-hierarchical transforms (i.e. pyramidal coding [43] and wavelet transform [44]) such as the discrete sine transform (DST) and Walsh-Hadamard transform (WHT), except for the Karhunen-Loeve transform (KLT) [45].
The performance of the DCT has been extensively evaluated in respect of energy compaction and coding gain [45] compared to other transforms [45, 46] and other compression algorithms such as vector quantisation [30]. Much research work revealed that DCT-based coding was more advantageous than other techniques in terms of coding gain and complexity. As for the block matching algorithm, a variety of different schemes are known to exist. The exhaustive search is known to have the best performance, and there exist several sub-optimal motion search algorithms such as logarithmic, 3-step, etc [26]. Bidirectional motion estimation and half-pixel precision motion vectors, as employed in the MPEG standards [26, 41], are advanced block-based algorithms.

Although, in MC-DCT-based standards, the block matching algorithm and the 8-by-8 DCT are adopted for motion compensation and transformation, respectively, various encoding parameters can be tuned to obtain different performances, motion vector range, motion estimation method, intra/inter control, quantiser control and video rate control. This flexibility exists commonly in main stream video compression standards: ITU-T H.261, H.263 and MPEG (1 and 2). These standards have the same specification [25, 26] for the DCT, furthermore, H.263 and MPEG2 contain more advanced features for MC.

Recent issues in MPEG applications include transcoding of compressed MPEG bit streams [47], high profile MPEG2 layer codings applied to some of HDTV [48, 49], very low bit rate MPEG4 applications for wireless communications [50, 51], etc. Further video rate reduction for compressed video is occasionally required, say from 9 Mbit/s to 5 Mbit/s when a direct broadcast satellite MPEG stream is transmitted through a terrestrial cable with a smaller channel capacity. Transcoding needs to be applied by attaching a cascade of a decoder and an encoder, which reduces the bit rate. Problems here are the increased system complexity and additional quality degradation [47] by the transcoding. The MPEG2 standard has been applied to HDTV systems and super-high-definition imaging systems. Although no commercial HDTV service is currently available, due to the lack of affordable HDTV displays and market interest which is focused more on interactive services, MPEG2 has been successfully implemented in American and European HDTV systems [49, 52]. Furthermore, generalised super-high-definition imaging systems are considered to be achievable with MPEG2 high profile coding [48]. For wireless visual services, narrow channel bandwidth and the strict delay requirement are major technical issues. In order to improve error resilience in the MPEG2 video bit stream, various error concealment techniques have been used [53, 54] such as the method to insert highly protected codes instead of slice re-synchronisation codes into the bit stream [55].

2.2.2 Fractal-based compression

Fractal-based image compression has been controversial. The inventor, Michael Barnsley claimed that this technique can reach far higher compression ratios than existing compression techniques. For example the Barnsley system can compress 45 seconds of full-motion video onto a 1.2 Mbyte
floppy disc, and can compress full screen VGA images into 5800 bytes [56]. The compression ratios are more than 900:1 for the case of PAL video with 720 pixels by 576 lines and more than 50:1 for 640 pixel by 480 lines VGA mode, respectively. But there are some different opinions. According to a recent paper [57], the performance of fractal-based coding is not consistently higher for all kinds of images. It is claimed not to be appropriate for images with Euclidian geometries.

Fractal-based coding is an analysis-and-synthesis-based technique which equates an image with a set of affine (or contractive) equations describing the details in images. If a few mathematical equations can efficiently represent an image, one can achieve an ultra high compression ratio. The contractive transformation and the attractor are the key concepts. The contractive transformation or iterated function systems (IFS) is a mapping process with the property of space preservation, which means that the mapping converges to a final reconstructed image by iteratively applying the IFS. The attractor for an image is a final converging image which is not affected by the transformations [56, 58]. The affine equations can render the image without further quality degradation when it is zoomed in or out, which is called the self similarity property [59]. Examples of IFS forming affine transforms were described by Barnsley [60]. A challenging task here is to find an IFS from images, which is the inverse process of fractal image synthesis, since an image can be too complicated for a limited number of IFS to represent it. For this reason, in one of the representative papers, written by A.E. Jaquin [61], basic features of images are used, such as mean brightness, contrast and directions of edges in blocks which are equally-sized square fractions of an image, not in the whole image.

2.2.3 Wavelet-based compression

The Fourier transform assumes that an input signal is periodic, in the broader sense, stationary. Thus if the signal is composed of a few frequencies, the Fourier transform perfectly decomposes it into corresponding positions on its frequency axis. The Fourier transform for windowed signals is useful to analyse the signal on this basis. However, in practice, the signal may often be neither periodic nor stationary, and this also applies to a video signal. The statistics of video signals vary with time. This raises the necessity for time-frequency analysis by short-time Fourier transform (STFT) so that the signal is observed and processed on the basis of frequency observations with different resolutions over time. In the STFT, however, the time-frequency resolution is fixed over the entire time-frequency plane. This needs to be changed to handle a signal with variable time-frequency resolution, i.e. time-scale. The wavelet transform provides the way to analyse signals with multiple time-frequency resolution [62]. Wavelet-based image compression is fundamentally the same as traditional sub-band coding [31] which decomposes an image into several frequency bands in a pyramidal hierarchy. Instead of using quadrature mirror filters [31] to decompose the image into frequency bands, an orthogonal transform [32] can be used, such as Haar wavelets [63]. Using this scheme, particular frequency bands of interest can be coded with increased bit rates to achieve efficient compression.
2.2.4 Contextual-feature-based compression

Conventional video coding algorithms exploit the statistical redundancy contained in video sources, while model-based and object-based codings [33] deal with visual information in a different domain, i.e. contextual or semantic features such as facial features. This video coding scheme has emerged as a very low bit-rate coding technique. Under the limitation that video is assumed to contain head-and-shoulder motion, 2-D or 3-D face motion features (such as wire frame face model indices) can be transmitted at very low bit rates, e.g. 10 kbits/s [64]. Since the encoder and the decoder have access to the necessary information about the frame model for reconstruction, the compression ratio can reach several hundreds to one with acceptable video quality. Much research work on related problems has been conducted; automatic localisation of semantic features [65], object-oriented analysis synthesis in which no explicit face model is required [66], 3-D motion estimation [67,68] and morphological operation (dilation and erosion)-based edge detection [69], etc. This coding technique is known to be useful only for human face or head-and-shoulder motion. That is, if the input image is a general scene, it cannot be properly coded when the stored information in the encoder and the decoder is designed for the face model only. Similar problems exist also in fractal-based coding and vector quantisation. It is a formidable task to find a generalised IFS data base for fractal-based coding and a generalised code book in vector quantisation as is in this contextual-feature-based coding.

2.2.5 Parallel video processing and full-frame transform coding

Transform-based coding is a well-established video coding technique and it has been adopted in many international standards. In the standards the DCT transform processes images on a block basis (with the same block size, 8 pixels by 8 lines). The reason why the DCT-based coding is widely adopted is that it can be combined efficiently with other compression techniques (motion compensated coding, run length coding, variable length coding, etc.). That is, it functions as part of a hybrid coding scheme. Another reason is that it has advantages in comparison with other techniques in terms of performance and complexity. However, it has a drawback in that the 8 x 8 DCT causes undesirable blocky effects, since it does not take into account the correlation between blocks. Thus, a bigger block size can be applied where real-time processing is not essential. In medical imaging, full-frame DCT transformation is employed to compress medical images with less quality degradation by taking a whole picture as the input to a large 2-D DCT [70-72].

Disadvantages of the full-frame transform are the saturation of coding gain when the size of transform block increases [31] and the heavy computational requirement. However, despite the limit on the coding gain, the full-frame transform can be used for achieving blocky distortion free images. The second drawback can be resolved by using parallel computers. Using massively parallel computers such as a SIMD (single instruction multiple data) machine which
dramatically reduces the processing time for the full-frame transform. When using a sequential computer, each data transfer and every arithmetic operation take several CPU cycles. Eventually, it may take several thousands of CPU cycles to compute an $8 \times 8$ 2-D DCT. However, when using a parallel computer, it can be programmed to do the entire job within only a few cycles [73]. Parallel computers are classified into several categories of architecture according to the topology of controller and processor elements; SIMD, MIMD (multiple instruction and multiple data) [74], etc. Among many different type of parallel computers, Thinking Machines Corporation, Connection Machine [75], appears to be suitable for full-frame transform coding since it has a massively parallel SIMD architecture. Another parallel computer composes of off-the-shelf computers connected over a network [76]. This network of computers can be effectively used as a parallel computer implementation. If the network is efficient to distribute loads to each constituent computer then this parallel computing environment will be cost effective. Wavelet-based coding is also considered a blocky-effect free scheme since it does not partition images into blocks. However, from the view of parallel architecture, it is not advantageous in comparison to a full-frame transform since the hierarchical decomposition of an image is fundamentally a serial process. Hence, the full-frame DCT is thought to be more suitable in view of real-time, parallel implementation.

2.3 Transmission of compressed video via rate management

The purpose of video compression is the cost efficient representation of visual information either for storage or transmission. In this research, attention is focused on transmission of compressed video through public networks. The necessity for video rate management has a varying importance depending on the ratio of video source data rate to channel rate. At very high transmission rates, the rate management function does not have to have a large effect on transmission since the channel capacity is large enough to accommodate the rate variation. In the other extreme case, i.e. at very low transmission rates, large fluctuations of video rate cannot be easily accommodated in the framework of the MC-DCT algorithm. Therefore, concentration is given to the MPEG1 and MPEG2 standards where the rate control technique works as the only rate management function at reasonable channel rates. Since the MPEG SIF is a widely available picture format, the corresponding channel rate is set to around 1 Mbits/s with a 30 frames/s picture rate. In this setting, some 30:1 compression ratio should be achieved.

In MPEG video transmission, efficient buffering and rate control is particularly crucial for CBR applications where the transmission buffer plays a key role in the rate control algorithm. In VBR applications, the buffering is not required in the encoder because the encoded video is transmitted at the rate which the video requires for a specified quality. Thus, the video rate control is carried out on the network management side. If network congestion is assumed not to occur, the variable rate transmission would be guaranteed. On the other hand, in the CBR mode, the encoder buffer occupancy (i.e. filled capacity of the buffer) fluctuates depending on the amount of visual information. The occupancy thus becomes the key variable to be maintained within the specified control range. If the buffer occupancy moves out with the
normal range, i.e. if overflow occurs, error free communication cannot be achieved. Therefore, the first goal of video rate control in the CBR mode is to guarantee that buffer malfunction does not occur. The next goal is to obtain the highest possible video quality. Generally speaking, a large size of buffer is able to prevent the malfunctions, if the buffering delay can be neglected. However, unless an infinite size of buffer can be installed, a rate control technique should be employed. In addition, for delay critical applications, a more effective and intelligent buffering technique is crucial.

2.4 A review of video traffic management techniques

Video traffic management for B-ISDN has been a challenging task in the field of network management in the sense that compressed video is the major bursty information source which makes the overall traffic less predictable and less likely to be modelled in statistical ways. Much research work on statistical traffic source models has been conducted in order to elucidate its statistical properties, and a variety of models and control schemes have been proposed. In many traffic management techniques, the buffering technique is widely used to smooth so-called “bursty traffic” [77] and maintain the traffic level as stable as possible by regulating the incoming source traffic to the buffer. However, in both areas of traffic modelling and buffering, handling the bursty nature of video signals has been a formidable task to be solved, due to its unpredictable nature which causes performance degradation in statistical multiplexing in ATM networks and buffer malfunction in CBR applications.

Much research work on traffic modelling for VBR-ATM networks has been published [77–79], most of which concerned the investigation of statistical modelling, queueing theory-based approaches and ATM protocol-based schemes. Transmission buffering for CBR applications has been investigated in several different ways; adaptive quantisation [80–83] and control theoretic approaches [84–86]. In this section, video rate management techniques are reviewed according to the type of channel: VBR, CBR and ABR.

2.4.1 A categorisation of video traffic management techniques

Recently, a control theoretic classification of ATM traffic management has been published [87]. In this literature, traffic management techniques are classified into two: open loop and closed loop based on feedback control which incorporates both the network management and the source rate control. The most demanding and bursty traffic source is known to be compressed video, therefore, the classification needs to include rate control techniques for single video sources as a branch of traffic management. Fig. 2.1 shows traffic control techniques for CBR, VBR and ABR modes from the view point of video rate management. In the CBR mode, since the channel rate is fixed, the rate control should be performed on the source encoding side, by regulating the buffer occupancy. On the other hand, in the VBR mode, video sources are allowed to transmit at the rate demanded for a specified target quality. Thus, the network should handle
the traffic flow. This may cause network congestion in some cases. Due to this nature of the ATM traffic control, the ABR mode has recently been proposed as a new mechanism which alleviates the congestion on the network side and diverts the congestion to the source coding side. The primary purpose of the ABR mode is to provide the network with more room for real-time sources to exploit the bandwidth by prioritising the sources [88] according to the type of service they require.

![Figure 2.1: Classification of video rate and traffic management techniques.](image)

For VBR MPEG video transmission, modelling of the compressed video appears to be the core task to be tackled [77] since its statistical properties provide a firm base for a strategy to establish an optimum effective bandwidth allocation [78]. Much research work has been carried out to establish an optimum statistical model of compressed video [79, 89–93]. In CBR MPEG transmission, buffering, which stores the compressed video in a first-in-first-out buffer for a specified period of time, is widely used to adapt the variable rate video to a fixed rate channel. The buffering techniques range from statistical and reactive control schemes [23, 80] to preventative control schemes [84, 85].

### 2.4.2 Video traffic management techniques and traffic models for VBR

Much of the existing research has focused on statistical modelling of multiplexed VBR traffic or a single VBR video source. This is mainly based on conventional models such as autoregressive (AR), Markovian and queuing theory models [77]. Video traffic has mostly been characterised by a stationary Markov renewal process [78, 79, 89] or a Markov chain process [94] which also covers a single traffic source. The single video source was also modelled as an autoregressive moving average (ARMA) process based on empirical studies [95]. These models can be categorised into two types on the basis of whether or not the model takes account the temporal correlation of video traffic. A renewal process or Poisson process views an arrival process as an independent, identically distributed process and it does not exploit temporal correlation of video traffic. This is derived from the concept of queueing theory that incoming traffic is generally considered
to be a Poisson process [96]. In the Poisson process there is no memory associated with the input data, that is, the rate of current input data is independent of the previous or future data. However, compressed video traffic has memory, i.e. it possesses Markovian characteristics. Hence, Markovian models are considered to be more appropriate and more widely acceptable.

Statistical multiplexing [2] is crucial for the ATM-based B-ISDN. The gain from statistical multiplexing is the increased network utilisation by multiplexing more VBR sources whose sum of peak rates may exceed the channel capacity, while maintaining the video quality requirement [97]. Video traffic can be classified into single sources and aggregated sources, i.e. a multiplexed stream of multiple sources. Multiplexing, in general, makes the bursty traffic less fluctuating [78] when the multiplexing gain is adequately achieved. However, although the bit streams are multiplexed, effective bandwidth allocation is difficult to achieve when the traffic is bursty and non-stationary since the allocation must cover the worst case, i.e. the peak video rate [98]. Thus, dynamic bandwidth allocation rather than static allocation is more desirable [79]. In a case of congestion, subsequent packet loss due to dropping ATM cells is the major cause of performance degradation. Queuing delay is an important secondary degradation of performance [99]. Thus, traffic management should aim to maintain the traffic level as stable as possible and meet the quality requirement.

The major problem in applying statistical models to video traffic is the burstiness (the ratio of peak rate to mean rate of sources) of video traffic, when measured as the number of bits or ATM cells. The congestion may occur when multiple non-stationary compressed video sources are transmitted through the network. When several sources transmit at their peak rates simultaneously, buffer overflow in the statistical multiplexer can only be avoided by dropping parts of the traffic in the queue. Other solutions adopt a priority mechanism [100-102]. It flags cells which are essential for minimum quality of service. Dropping cells takes place only on the cells flagged when buffer overflow is anticipated [103] as they are less essential for the quality requirement. In another approach, preventative congestion control can be applied in order to maintain the traffic within a safe level in advance of the congestion in which the video encoder takes action to reduce the emitting rate of compressed video. Neural network-based traffic management [104] schemes have recently emerged, which include traffic parameter prediction [105] and adaptive call admission control [106] as such approaches. Recently, a new approach to tackle the burstiness has been proposed [107] in which the video traffic information is modelled based on camera operations. The camera movement is represented by “epochs” which means discernible events: static camera, zoom, pan, etc. Global and local content descriptors are used to represent a picture. The descriptors consist of correlation, mean and variances of pictures. This technique looks promising in that it takes into account the camera movements which predominantly cause scene changes and entail a large number of coded bits when realistic video is being encoded. Detecting scene changes or camera operation is another task which requires investigation. A scene change detection technique based on bit rate change between two pictures has also been published as a relatively simple algorithm [108].
2.4.3 Video rate control and buffering techniques for CBR and ABR.

Representative research work for the CBR mode are mostly based on conventional linear predictive models [80, 84, 86, 93, 109] or control theoretic models [83, 110]. The prediction models are autoregressive models with orders up to 2 or 3. These approaches adopt the adaptive quantisation technique using global and local image features. It is common to use feedback control rather than feed-forward control in these approaches. The former reactively controls the encoding parameters such as quantisation step size according to the buffer occupancy. For example, Zdebski’s [80] buffering scheme adopts feedback control; when buffer overflow takes place, the dropping mechanism is activated so that picture line scan rate reduces to half by 2:1 subsampling. A nonlinear control curve for quantisation step size was also used. Leduc’s [85] work, which is based on adaptive control theory, employed feedback control under a consideration of the non-stationary nature of video. Cheng’s method [84] adopted an adaptive scheme for the quantiser step size using the variation of the actual number of coded bits from the estimated bits rather than only using the buffer occupancy. This scheme also employs feedback buffering and is for studio quality video compression of up to 5 Mbits/s. Lee’s scheme [86] using estimated bits of each picture and each macro block (16 lines × 16 pixels) can also be classified as feedback buffering. Keesman [93] has proposed a feed forward technique, viewing the video traffic as a non-stationary process. In this technique an adaptive quantisation is applied by calculating the ratio of local activity of a macro block to average activity over a picture.

Recent approaches to CBR rate control include forward analysis-based rate control, iterative search techniques, rate-quantisation models for optimum quantisation scales, parallel rate control for multiple video buffers, etc. The forward analysis scheme for incoming video is implemented in the American HDTV system, “Grand Alliance”. It anticipates scene complexity by use of coarse motion vectors already computed before the picture reaches the encoding loop. It adapts the quantisation scale according to the buffer occupancy as well as the input video and motion vectors [52]. The iterative search for an optimum quantisation scale uses rate-distortion ratio changes until an optimum quantisation scale is found for a whole picture [111]. In order to specify the relationship between the bit rate and the quantisation, rate-quantisation models have been proposed [109, 112]. The common feature in the proposed quantisation models is that a quadratic term is applied to specify the relationship between the video rate and the quantisation. In HDTV video encoding systems, parallel processing is one of the big technical challenges which aims to alleviate the requirement on processing speed by distributing the computational loads to several identical processing units [113]. Accordingly the rate control task is also distributed to several buffers [114] and handling system complexity becomes the issue. Although various different rate control algorithms have been proposed, the complexity-based or activity-based rate control scheme is currently the mainstream rate control technique which is similar to MPEG2 TM5 or its modified versions [115, 116]. In variable bit rate video transmission network congestion must be avoided so that the quality of service requirement can be met [117]. Among many congestion management techniques, rate-based schemes [23, 118, 119]
are very similar to rate control techniques for the CBR applications in the sense that the video encoder regulates its data rate depending on the network status.

The ABR mode has been proposed specifically for data applications such as computer-generated traffic. ABR traffic has access to bandwidth only when no CBR or VBR traffic is waiting for transmission [120]. The following rationale is the starting point: “ATM data traffic often requires no firm guarantee of bandwidth, but, instead can be sent at whatever rate is convenient for the network, i.e. rate availability. ABR traffic gives the network the opportunity to offer guarantees to high priority, and divide the remaining bandwidth among ABR connections” [88]. In order to prioritise the sources and allocate available bit rates to them, it is essential to feed the network status back to the sources so that the sources control their emission rate depending on the available channel rate. A credit-based scheme [88, 121] has been proposed for this ABR connection, which changes its buffer capacity allocation in the network side depending on the network status. Thus, from the view of the video encoder, this is an adaptive channel rate allocation scheme in the sense that the output rate of the buffer varies.

2.5 Conclusions

In this chapter, a brief review of real-time video compression has been presented in two aspects: algorithmic features and the transmission environment. Among state-of-the-art video compression techniques, the performance of fractal-based and contextual-feature-based coding has been claimed to be higher in comparison to the motion-estimated DCT coding scheme, in terms of compression ratio and video quality. However, many problems remain unsolved; comprehensive representations (IFS and face model) of general scenes, transmission aspects such as optimised variable length codes for IFS and face model parameters, etc. These problems preclude the emergence of standardised products and require more research on those techniques. On the other hand, the MC-DCT scheme is well established in that it can be used in a wide range of video coding applications and it also has a competitive performance over the others.

It was indicated that the new ABR connection mode for ATM networks has been considered as a solution to network congestion. Due to the time-varying nature of compressed video traffic, it is necessary to control the video rate on the source side as well as the network side. How the video rate is controlled in the source encoder becomes the next question. The buffer-based rate control technique is thought to be applicable to the VBR connection as well as conventional CBR applications, since it may provide an effective solution to network congestion in the VBR mode. Although ATM cells can be discarded depending on network status in the VBR mode to avoid network overloading, if the video rate is controlled adequately in the source side, the network congestion can be avoided more efficiently. The buffer-based rate control technique can contribute to both the CBR and VBR modes. In subsequent chapters attention is focused on video rate control techniques for the CBR mode.
Chapter 3

Foundation of video rate control for CBR

3.1 Introduction

In Chapter 2, real-time video compression techniques and video traffic management schemes for ATM networks were briefly reviewed. It was shown that the transmission buffering technique is widely used for traffic management as well as video rate control. In this chapter, first, algorithmic features of the buffer-based rate control technique will be investigated, which are specified in international standards. This will be followed by a review of quantisation aspects: adaptive quantisation and rate-distortion theory. The quantiser is the function which actually controls the data rate. The buffering technique is associated with the quantiser by changing the step size depending on the buffer occupancy. As a compressed video stream from an MPEG video encoder is inherently variable rate data there should be a buffer to convert the variable bit rate to constant bit rate in the CBR mode. This configuration is common to the standards, thus, the effect of improvements of this rate control technique can be generalised for any buffer-controlled quantisation scheme. Rate-distortion theory is inevitably involved since the buffer-based rate control scheme operates via the quantiser which introduces distortion. On this basis, a brief review of the rate-distortion analysis on the adaptive quantisation will be presented within the framework of the CBR mode.

Conventional approaches to video rate control and traffic management have been based on traditional linear techniques under the assumption that video rate can be treated as a time series with stationary statistics. However, as indicated in Chapter 2, many of those approaches do not adequately accommodate non-stationary variations in video rate. Thus, a different point of view focused in this research is that the statistical nature of realistic video is non-stationary, hence, it is thought to need nonlinear approaches. Unlike standard video sequences, which generally do not contain camera movements and realistic scene changes, movies, sports and TV advertisements do exhibit many dramatic changes in visual context, even over a short period of time. This fact needs to be taken into account in a realistic environment. An analytical review will deal with nonlinearity of the video encoder and non-stationary properties of digital video, in the second half of this chapter. This will provide an insight into the nature of digital video and video encoding.
In this chapter, we:

- review the rate control techniques specified in the international standards,
- investigate the non-stationary nature of digital video and hence the nonlinear signal processing in the video encoder.

This chapter is organised as follows: Section 3.2 provides a review of the rate control algorithms described in the standards. Section 3.3 deals with quantisation and its rate-distortion analysis. Sections 3.4 and 3.5, respectively, discuss the nonlinearity of the video encoder and the non-stationary nature of real-time video. Sections 3.6 to 3.8 summarise the proposed schemes and the adopted configuration for an evaluation test bed. Finally, Section 3.9 concludes this chapter.

3.2 A comparative review of rate control mechanisms employed in compression standards

In the four international standards (ISO11172 MPEG1, ISO13818 MPEG2, ITU-T H.263 and ITU-T H.261), the buffer-based quantisation technique is adopted as a standard rate control mechanism. The standards specify only the core factors of the buffer-based video transmission; the quantisation step size range, the buffer size and the conditions for the buffer occupancy. On the other hand, their evaluation models describe detailed procedures. They are referred to differently; the simulation model (SM) for MPEG1, the test model (TM) for MPEG2, the verification model (VM) for MPEG4 and the reference model (RM) for ITU-T H.261. These evaluation models are designed to verify a specific video coding implementation. The reference model (RM1 to RM8) of ITU-T H.261 is the earliest version of these evaluation models. It functioned as a yardstick to compare new coding elements and/or improvements during the standardisation process [122]. The same approach has been applied to the three SMs and five TMs of MPEG (1 and 2) [123]. The MPEG4 standardisation (now underway), is based on the same platform, the VM [124]. The buffering techniques of the evaluation models aim to control the output data rate of compressed video from the video encoder, to accommodate it in a fixed rate channel. Since the size of the buffer is constant and limited due to the delay requirement, if the buffer mechanism fails to maintain the video rate within the specified range, buffer malfunction may occur. The worst case is overflow when a dramatic increase in the video rate cannot be accommodated in the buffer. By controlling the video rate and avoiding the buffer malfunctions, normal video transmission should be achieved without interruption. In this section the algorithmic features of rate control techniques are explored, which are specified in the standards and the evaluation models.

3.2.1 MPEG1

The MPEG1 standard, ISO11172-2 [26], was developed for video compression and transmission with constant bit rates of up to 1.5 Mbit/s which corresponds to primary rate access [1]. The
buffer control mechanism, defined in MPEG1, gives basic guidelines to avoid interruptions to transmission and abrupt changes of video quality. MPEG1 provides a few qualitative guidelines as follows:

- Allocate coded bits based on the picture type,
- Control quantiser step size in proportion to the buffer occupancy and picture detail,
- Discard high frequency DCT coefficients when the buffer overflow is about to occur,
- Insert stuffing words in the coded bit stream when the buffer underflows.

The quantisation step size of MPEG1 is controlled adaptively by the buffer occupancy and picture details. While the ITU-T H.261 for videotelephony has no predefined picture types, the MPEG standards have 3 different picture types, i.e. I, P, and B pictures; intra-coded picture (I), predicted picture (P) and bi-directionally predicted picture (B). Each picture is partitioned into equal-size macro blocks each containing 16 pixels by 16 lines. All macro blocks of an I picture are coded with intra mode, and those of P and B pictures are coded with either intra or motion-compensated inter mode. Since the picture type regularly repeats such as IBPBPBBBPBBB... or IBPBPPPPBBB... the buffer occupancy may thus fluctuate more in MPEG encoders than H.261 for the same video input due to the picture-type-based coding.

### 3.2.2 MPEG2

The MPEG2 standard, ISO13818-2, contains a very similar rate control algorithm to MPEG1, although the B picture is specified as an optional picture type. This standard sets up the baseline for the MPEG2 video encoder interfaced with a video buffering verifier (VBV) for a constant bit rate channel. The VBV is a rate control scheme based on the buffering technique. The standard specifies the size of the VBV buffer suitable for CBR channels, desired buffer states, etc. However, it describes global system aspects rather than the details of the rate control process itself, since the detailed control mechanism depends on the design of the encoding system. On the other hand, TM5 specifies the details of the rate control algorithm which is based on one-step ahead estimation using picture complexity measures. It allocates a certain number of bits to each type of picture with equations for the target number of bits, using the previous encoding results. The buffer size can be determined flexibly depending on the application so that a large buffer size can be used, in order to prevent buffer overflow. However, the actual quantisation scale (step size) is determined by the occupancy with respect to twice the mean number of bits per frame, regardless of the actual buffer size. This is employed so that the maximum delay of the video encoding remains shorter than two frame periods.

According to the standard, the delay, for low-delay mode communications, must be less than 150 ms. The delay requirement includes decoding as well as encoding. Therefore, if the encoder
and the decoder have the same size of buffers and no excessive delay exists in the encoding and decoding processes, the maximum tolerable delay in each side becomes 75 ms which is approximately a 2-frame period. The buffer occupancy should thus be controlled to stay within a 75 ms delay in low-delay mode. In this thesis the 2-frame delay is taken as a maximum occupancy value regardless of channel rate as is in TM5 (except Chapter 4). Since the delay is the most critical parameter to be controlled, measuring the occupancy as a percentage cannot represent the delay unless the size of buffer is specified. While ISO11172-2 specifies a single type of quantisation scale, ISO13818-2 provides two different types which are linear and nonlinear, respectively. In a strict sense, the latter is piecewise linear and globally exponential. The values of the nonlinear quantisation scale are fixed, i.e. non-adaptive.

3.2.3 MPEG4 and H.263

Initially, the video coding technique for MPEG4 was expected to be standardised on a framework of object-based coding. However, since the ITU-T has proposed its low bit-rate coding standard H.263 [125] as one of the MPEG4 audiovisual coding tools [40]. H.263 appeared to be taking its place as an established low bit-rate video coding standard. It employs a similar rate control scheme to MPEG1 and MPEG2, however, it contains a smaller number of constraints on the buffer size and the occupancy states than in MPEG1 and MPEG2. The noteworthy difference in H.263 is that it specifies a new coding mode, PB-frame, which encodes the P and the B pictures together by forming a macro block with 12 constituent blocks rather than 6 which is common to MPEG1 and MPEG2. This mode is designed to increase the frame rate with less increase in bit rate. However, from the view point of rate control, it is expected to entail a much larger number of coded bits in the buffer since two pictures are being coded in the same coded picture, PB.

As explained in the previous section, TM5 is the evaluation model for the MPEG2 standard. A widely used counterpart for the H.263 standard is the test model near (TMN) 5 [124] proposed by Telenor (Norwegian telecommunication authorities), which is similar to MPEG2 TM5. However, its rate control scheme differs considerably from TM5 in that it does not use the picture complexity measures and sets its initial occupancy as four times the mean number of bits allocated per frame. The quantisation step size is determined with respect to its one-step previous value and bit rate values by a linear equation. The rate control scheme of TMN5 guarantees no underflow but may result in overflow as does TM5.

3.2.4 H.261

The rate control algorithm specified in the RM8 is based only on buffer occupancy [122] which is mapped to the quantisation step size by a linear equation. Therefore, the magnitude of the quantisation step size is linearly proportional to the occupancy. This is considered as the
The simplest form of quantiser control function. The buffer size is set to a tenth of the output data rate of the buffer. This means that the buffer delay is inversely proportional to the frame rate for a given channel rate. For example, if the frame rate is 30 frames/s, the delay will be three frames period. However, if it is 15 frames/s, the delay becomes 1.5 frames.

### 3.3 Rate control via adaptive quantisation

The rate control scheme is now primarily discussed, which is suitable for MPEG1 or MPEG2 video encoders. Fig. 3.1 shows the configuration. When the buffer is used as storage for compressed video, it is vital to use a reactive mechanism which controls the quantisation step size by the feedback information of the buffer occupancy. Since the quantisation step size plays a key role to control the occupancy, the goal of rate control is how effectively the occupancy can be controlled by the quantisation step size specified in the MPEG standards. Several different quantiser control functions have been proposed. They can be classified into linear [25, 42, 86, 126], piecewise linear [41, 85, 127], and nonlinear [80, 83, 128]. Recently, two representative nonlinear functions have been proposed, which are sigmoidal [128] and unimodal [129] and will be described in Chapters 4 and 5. The nonlinear quantiser control looks more promising than the conventional linear method in that the linear method appeared unable to adequately cope with video rate fluctuations caused by video containing wide variation of visual context.

![Feedback buffer-based rate control scheme specified in MPEG.](image)

Fig. 3.2 depicts the parameters for rate control. The channel rate, $c_{\text{rate}}$, should be determined first since it may have different values depending on the application. The maximum tolerable coding delay, $d_{\text{target}}$, and the buffer size, $b_{\text{size}}$, are set to constant values. In the CBR mode, these three parameters remain unchanged once a channel connection has been established. The picture rate, $p_{\text{rate}}$, may vary according to users’ selections, however, it maintains the same value for a long period of time throughout the connection. Therefore, the only independent variable becomes the source entropy [31] of the input video, i.e. the amount of visual information.
In mid 80’s, an adaptive quantisation technique, which is one of the earliest forms of the buffer-based rate control, was proposed [127]. In this scheme, the quantiser is controlled in a way in which the buffer occupancy is maintained at 50% of the full state by using the following relationship:

$$\Delta n = \Delta f(Z_{n-1} - b_{-size}/2) \quad (3.1)$$

where $Z_{n-1}$ and $\Delta n$, respectively, represent the occupancy and the corresponding quantisation step size. The function $f()$ has the following relations:

$$f = 1 + \frac{a}{2} \quad x \geq U$$

$$f = \frac{a}{U - L} x + 1 \quad L < x < U$$

$$f = 1 - \frac{a}{2} \quad x \leq L \quad (3.2)$$

This piecewise linear function is known as soft limiter [127]. Fig. 3.3.

![Quantisation control function](image)

Here, $L$ and $U$ are threshold occupancies to detect underflow and overflow, respectively. The constant, $a$, determines the slope of the section between $L$ and $U$ depending on the buffer size.
Conceptually, this function is similar to the sigmoidal function which will be discussed in later chapters.

3.3.1 The buffer occupancy of an MPEG encoder

A larger-size buffer can absorb the bursty video rate transition. Hence, the large size may be a merit except for the consequent buffering delay. However, in interactive video communication such as video conferencing, the delay should be kept under a limit, e.g. a 4-frame period [24] because a longer delay may bring annoyance to viewers. Since the buffer size has direct influence on the coding delay, the delay requirement is the critical limitation in dimensioning the buffer.

The buffer operation is often described as follows; when the occupancy is low, then the encoder can generate more coded bits to increase the occupancy; if the occupancy is high, then the encoder must generate less coded bits to decrease the occupancy. When the input video contains drastic scene changes and rapid motion, the resulting fluctuation in video rate may not be accommodated. In this case, buffer overflow can take place, which interrupts normal video transmission. When the video contains extremely low activity, on the other hand, underflow may occur. However, this can be managed so as not to interrupt normal transmission by inserting justification bits into bit streams (bit stuffing). Farvardin [82] specifies average terminal time, $E_{n_2}(B, 1)$, which represents the duration until the first overflow or underflow occurring for a memoryless Gaussian source with a finite size buffer, as follows:

$$E_{n_2}(B, 1) \cong \frac{B_0}{\bar{m}} + \frac{B - B_0}{\bar{m} - \hat{m}}, \quad \bar{m} - \hat{m} > 0 \quad (3.3)$$

where $\bar{m}$ and $\hat{m}$ represent the average input rate and the output rate of the buffer in bits/s, respectively. $B$ and $B_0$ are the buffer size and the initial buffer occupancy in bits, respectively. Thus, $E_{n_2}(B, 1)$ depends linearly on the buffer size and inversely on the net data rate ($\bar{m} - \hat{m}$) into the buffer. The net rate becomes the critical variable, which is dominated by the source entropy. It is known that the source statistics cannot be easily modelled. Thus, the above assumption, “memoryless Gaussian” can hardly be applied to non-Gaussian sources which is a more general case. Imperfect knowledge of source statistics can lead to a quantiser output (after variable length coding) that is different from the design values, in such a way that the average VLC code length of the encoder output is longer than the buffer output rate. In this case, buffer overflow may occur [127]. For example, for a video with scene changes, the MPEG video encoder encodes the video with fixed length codes rather than variable length codes. The short-term average code length may exceed the buffer output rate, leading to overflow.
3.3.2 Rate-distortion analysis on quantisation control

Rate-distortion theory [31] is the foundation for the analysis of the entropy of a source signal and its distortion caused by the encoding process. In MPEG-style video coding, the source data is processed by a cascade of statistical redundancy reduction processes (DCT, quantisation, run-length coding, Huffman coding, etc.). The quantiser is controlled by the occupancy control mechanism which monitors the buffer. For a block transform quantisation scheme operating on 1st-order stationary Gauss-Markov sources, it is difficult to solve the equation of the relationship between the quantised transform coefficient entropy and the corresponding distortion. Since the entropy and the distortion are related to each other through the quantiser (quantisation step size), in general, it is known that no explicit expression is available. For this reason, it is not straightforward to apply rate-distortion theory to the source and the resulting distortion in this framework.

Due to this reason, much analysis work has focused on Gaussian sources [31] and linear source models [29, 130] which are mathematically tractable. Analysis of the transform coding was also based on the assumption that the transform coefficients have a Gaussian distribution [131]. Quantisers for DCT-based coding have been typically designed under the assumption of Gaussian statistics for the transform coefficients. The theoretical justification for this assumption is the central limit theorem [96] which states that a summation of Gaussian random variables also has a Gaussian probability density function (pdf). Since the linear transform (DCT) is a process of weighted summation, the pdf of its coefficients is close to Gaussian, too [132]. However, in practice, the distribution of the majority of DCT components may be quite different from Gaussian statistics. Therefore, the rate-distortion function can only be solved for a few simple examples where the video has Gaussian statistics. As digital video often contains non-Gaussian statistics, however, the rate-distortion function, \( D(R) \), can be interpreted as the upper bound to the actual performance of the DCT-based video coding [133]. It is known that \( D(R) \) of non-Gaussian sources and Gaussian sources have the relationship, \( D(R)_N \leq D(R)_G \) [31]. With this practical use of the rate-distortion function, two different quantisation control functions will be analysed in Chapters 4 and 5.

3.4 Nonlinearity of the MPEG video encoding process

A definition of a nonlinear system (Hilborn, 1994) [134] can be given as follows: “A nonlinear system is a system whose time evolution equations are nonlinear. That is, the dynamic variables describing the properties of the system (e.g. position, velocity, acceleration, pressure, etc.) appear in the equations in a nonlinear form. The key significance of nonlinearity is that a small change in a parameter can lead to sudden and dramatic changes in both the qualitative and quantitative behaviour of the system.”

Thus, to study nonlinearity in rate control requires an investigation of its time evolution equations and a search for “sudden and dramatic changes” in video rate responding to a small
variation in its inputs. In what follows, an idea is developed which shows the nonlinearity of video encoding by considering the time evolution equations. Dramatic variations in video rate will be exemplified in next section.

A video encoder can be simplified without the motion estimation/compensation function, as shown in Fig. 3.4 in which the motion compensator sends zero values for all motion vectors. Then, the encoder consists of four functions: DCT, quantiser (Q), run length coder (RLC) and variable length coder (VLC). The video is encoded with intra or inter mode. $Q_s$ represents the quantisation step size which is controlled by a linear function.

$$\frac{d c_b}{dt} = \text{bit rate}$$
$$\frac{d^2 c_b}{dt^2} = \alpha c_b$$

(3.4)

where the second derivative becomes the video rate fluctuation and $\alpha$ is a constant. The second derivative is interpreted that the fluctuation is linearly proportional to $c_b$. Since the DCT and Q are linear processes, Eqn. 3.4 holds for the output data rate of the quantiser (Q). However, the subsequent RLC is a nonlinear process. The RLC coding gain is given by a nonlinear equation [31],

$$G_{RLC} \leq \frac{1}{H_Q(X)}$$

(3.5)

where $H_Q(X)$ is the source entropy of the quantiser output. The output bit rate of the VLC is given by a linear equation [31]

$$R_{VLC} = H_R(X) + \epsilon \quad \text{bits/sample}$$

(3.6)

where $H_R(X)$ is the source entropy of the RLC output. If the VLC encoder is properly designed, the positive real constant $\epsilon$ goes to 0. In practice, the quantiser may be controlled by a nonlinear
function, thus Eqn. 3.4 may not hold even for the quantiser output. In addition, due to the subsequent nonlinear coding process, the time evolution equation, Eqn. 3.4, of the video rate fluctuation of the video encoder is not linear.

3.5 Non-stationary nature of real-time video

Non-stationarity is a fundamental premise of nonlinear signal processing. What ‘non-stationary’ means can be clarified by considering the following definition of stationarity from Priestley, 1988 [135] which provides a clear view: “Linear models are generally described with stationary process. Any stationary process with a purely continuous spectrum - which means that its spectral density function exists for all frequencies - can be described by a general linear model

\[ X_t = \sum_{n=0}^{\infty} r_n \epsilon_{t-n}. \]

Here, \( \epsilon_{t-u} \) is uncorrelated with the process \( X_t \). Conventional time series analysis is heavily dependent on the twin assumptions of linearity and stationarity”.

Therefore, statistical processes which do not fall into this definition can be viewed non-stationary in a spectral analytic sense. It is known to be reasonable to apply nonlinear approaches to non-stationary sources [135]. Thus, the question here becomes: Does digital video possess non-stationary statistics? Even by intuition, one can answer this question positively. When the video contains the small motion of a news presenter with a constant background, one may say that the statistics of the video are stationary. On the other hand, when the video shows sports scenes with frequent camera operations (i.e. scene changes) or action movies, it cannot be said that the statistics are stationary. This judgement can be justified from the definition of stationarity. A wide sense or covariance stationary process, which is stationary with the order up to 2, can be defined as follows [7, 136]:

\[
\begin{align*}
E(X_t) &= \text{constant for all } t, \\
Var(X_t) &= \text{constant for all } t, \\
Cov(X_t, X_{t-h}) &= Cov(h) = Cov(-h).
\end{align*}
\]  

(3.7)

where \( X_t \) is a random variable and \( h \) is the shift in time. Let \( X_t \) be a difference picture. The mean, \( E(X_t) \), and the variance, \( Var(X_t) \), are not constant for \( t \) and the covariance, \( Cov(X_t, X_{t-h}) \), may not be \( Cov(h) \), when the video contains rapidly changing visual information. The entropy profiles for test sequences demonstrate this by showing inter-frame difference, Fig. B.1, Appendix B. If the conditions specified in Eqn. 3.7 hold for \( X_t \), its entropy will be constant for all \( X_t \). However, this cannot be the case for realistic video containing large variation in visual context. One can see the non-stationary nature of realistic video Fig. B.1(b) and (c). After the reduction of statistical redundancy of realistic video, the data rate of the compressed video may often have a similar profile to the source entropy since the coding process is not capable of suppressing the non-stationary rate variation in the compressed video.
3.5.1 Classification of scene change features

In this research, it is considered that the non-stationary nature of digital video and its compressed bit stream originates from scene changes. A scene change represents any distinctive difference between two adjacent picture frames. Thus, it includes rapid motion of moving objects as well as a change to different visual content. Without the scene change, video would be treated as a statistically stationary source such as head-and-shoulder video telephony. The scene change is induced by moving objects and/or camera movement, and can be enumerated from Class i to Class vi as follows, in the order of increasing scene change:

- (Class i) general scene: no rapid motion, no camera motion (video telephony scenes).
- (Class ii) rapid motion: still background and rapid object motion,
- (Class iii) panning: slow, steady and transitional (horizontal, vertical or diagonal) camera movement [24],
- (Class iv) zooming (in or out): rapid change in background and objects by changing camera viewing range,
- (Class v) scene cuts: an abrupt change by switching cameras or by changing visual context,
- (Class vi) continuous scene change: continuous change of visual context.

Among those classes the continuous scene changes, Class vi, contribute most to the non-stationary and bursty video rate variation since it contains the most frequent scene changes and eventually causes larger variations in the source entropy than other scene change features. This often occurs, particularly in television advertisements. Zooming is considered generally to entail more encoded bits than panning since the former alters scenes in all directions but the latter does only in the direction of the camera movement. Realistic video may have more complicated forms of scene change, e.g. a combination of the above-mentioned 6 classes over a short period of time. Fluctuating video rate originates from this unpredictable nature of the visual content. Therefore, in order to resolve the problem of rate control, different approaches are required rather than relying only on purely statistical methods.

Another classification is made via statistical and contextual definitions as shown in Fig. 3.5. The camera-induced scene change is included in the contextual class. In this thesis, the statistical scene change features up to second-order are used as they are computationally simple and the goal of the devised rate control techniques can be achieved.
### 3.5.2 Visual features of video sequences used in simulations

Five video sequences were used in simulations: “Cascaded”, “Starwars”, “Adverts”, “JFK” and “Topgun” (Fig. A.1 to Fig. A.5, Appendix A). The image format is MPEG SIF (CCIR 601 4:2:0) where the picture size is 352 pixels (row) by 240 lines (column). All the video sequences are digitised at a 30 frames/s rate. The edited versions of the five sequences are shown in Appendix A.

“Cascaded” is a cascaded video sequence which comprises the three standard sequences; “Miss America”, “Football” and “Susie”. While “Miss America” and “Susie” show relatively smaller motion, “Football” has rapid and large motion throughout the sequence with a still background. This sequence has three different sections each of which has 90 frames. Since “Football” is located in the middle there come out two dramatic scene changes plus rapid motion embedded in the “Football” section. “Starwars” is widely used in the field of video traffic modelling [89, 90]. It often exhibits dark background, slow motion, less colour change and many artificially synthesised scenes throughout the sequence. The sequence was captured from a part with relatively rapid motion and dramatic scene changes using a televised version of the movie. “Advert” consists of three television advertisements each of which has rapid motion and frequent camera movement. This also has many distinctive scene changes between and within advertisements. “JFK”, which was digitised from an edited version for a television advertisement, has more dramatic scene changes, e.g. transitions between coloured and monochrome scenes and rapid zoomings. “Topgun” also contains many scene changes from a wide range of the scene change classes.

A summary of the video features is shown in Table 3.1. “Adverts” appears to have the highest entropy values, while “Cascaded” has the lowest. “Adverts” and “JFK” contain all the six scene change classes. The entropy is given in bits for difference pictures [31]:

![Diagram](image_url)
\[ H_D = - \sum_{k=1}^{N} p_k \log_2 p_k \]  

(3.8)

where \( p_k \) is the probability of the pixel difference value \( k \) among total \( N \) values.

As shown in Fig. B.1 (a), (b), and (c), “Starwars” and “Adverts” have more fluctuations in entropy, i.e., more time-varying level of visual information, particularly in “Adverts”. While “Cascaded” maintains an occupancy stable in each constituent sequence and shows high correlation between frames, the two others frequently show abrupt changes even within a short period of time, e.g., a 3 or 4 frames period.

<table>
<thead>
<tr>
<th>Scene change class</th>
<th>Cascaded</th>
<th>Starwars</th>
<th>Adverts</th>
<th>JFK</th>
<th>Topgun</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class i</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Class ii</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Class iii</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Class iv</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Class v</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Class vi</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

Table 3.1: Summary of visual features of the video sequences.

The dramatic fluctuation in video rate is caused by various kinds of scene changes and camera movement. As shown in Appendix A, many scene changes are observed in the movie and advertisement sequences, ranging from Class i to Class vi. On the other hand, although the “Football” sequence section included in “Cascaded” has rapid motion, the entropy level is less spiky than the other ones due to its constant background. To see the effect of the scene change, in Fig. 3.6, the buffer occupancy of the three video sequences is shown when they are encoded with the MPEG2 TM5 encoder. At a channel rate of 1024 kbits/s and a 30 frames/s frame rate, the mean number of bits allocated per frame becomes 34133.3 bits. The buffer size is thus 68266 bits. “Adverts” exhibits much wider fluctuations and a couple of buffer overflows, while “Starwars” shows a flatter profile. The mean entropy of “Cascaded” is smaller than the two others, however, the two dramatic scene changes appear to cause significant variations in the occupancy.
3.6 Two rate control approaches employed in the thesis

In the previous sections the background to video rate control was presented. The non-stationary and nonlinear nature of the video and the encoding process was also discussed. In the following sections, the approaches devised for the non-stationary nature of video signal will be introduced. The two different approaches will be described in a schematic way. The constituent functions of the approaches will be explained in Section 3.7, showing the connection with the MPEG video encoder. The simulation configuration will also be presented in Section 3.8. The detail of the approaches will then be explored in Chapters 4 through 6.

The two rate control approaches applied in this thesis can be summarised as a feed-forward approach and a feedback approach. The feed-forward approach, Chapters 4 and 5, comprises video rate estimators and the scene change feature calculator. The estimated video rate and the current buffer occupancy are used to derive a quantisation scale value via nonlinear functions to map the occupancy. The feedback approach, which will be discussed in Chapter 6, consists of a fuzzy rule-based feedback control and adaptive scaling factors which are determined by scene change features. Furthermore, a fuzzy logic control algorithm, which considers PSNR as a video quality measure as well as the buffer occupancy, is also investigated.

Fig. 3.7 depicts the feed-forward approach based on the video rate estimator, e.g. a RBF-network. The buffer-based quantiser control (Q/Buffer) is controlled by the rate estimator and the nonlinear quantiser control function. The scene change calculator is supplied with the input video signal and it generates the scene change features, $scf_1(k)$, $scf_2(k)$, etc. for each input picture. The video rate estimator predicts the video rate value, $\hat{v}_r(k)$ for the $k$th picture in either a linear or a nonlinear way using the scene change feature information. The
estimated video rate, \( \hat{v}(k) \), is used by the quantiser control function together with the actual video rate, \( v(k) \), to generate the quantisation scale, \( Q(k) \), used by the quantiser Q. The error signal, \( e(k) \), between the actual and estimated video rate signals is used to update the estimator weights in order to adapt the rate estimator to the scene changes in the input video.

Fig. 3.8 shows the configuration of the feedback approach to regulate the Q/Buffer by employing a fuzzy logic control technique. This scheme controls the quantisation scale, \( Q(k) \), by monitoring the error, \( e(k) \) between the video rate, \( v(k) \) and the target video rate, \( V_r \), which is a desired resulting video rate value set to a constant. In this scheme, the adaptive quantiser control is achieved by tuning the scaling factors of the fuzzy logic control. The scaling factors, \( S_{ge}(k) \) and \( S_{gd}(k) \), are derived from the scene change features, \( scf_i(k) \), by using a mapping equation. The performance of the fuzzy logic control-based scheme is improved by adopting a PSNR monitor function to calculate the accumulated frame quality, \( Q(k) \), at each macro block. The PSNR difference, \( q(k) \), is derived from \( Q(k) \) and the target quality, \( QT \). Thus, the fuzzy logic control outputs \( Q(k) \) using both \( e(k) \) and \( q(k) \) as control inputs.

![Figure 3.7: Approach 1: feed-forward predictive scheme.](image1)

![Figure 3.8: Approach 2: fuzzy logic control.](image2)

\(^1\)In later chapters the variables \( e(k) \) and \( Q(k) \) will be used instead of \( e(k) \) and \( Q(k) \) in order to represent the video rate variables by a more explicit notation.
In the above two approaches, the rate control techniques employed in this research aim to:

- smooth the fluctuation of the buffer occupancy,
- avoid overflow without objectionable degradation in video quality,
- achieve better subjective and objective video quality.

### 3.7 Rate and quantiser control schemes

In this section the interface of each rate control function is described in relation to the MPEG video encoder. Fig. 3.9 shows the block diagram of the approaches in the framework of the generalised MPEG video encoder boxed by the dotted line. For the feed-forward approach, the scene change calculator (SCC) and the video rate estimator (VRE) are associated with the nonlinear quantiser control (NQC). The input for the SCC comes with the SIF picture format. The estimated quantisation scale, $Q_{sf}(k)$, for the feed-forward approach, is used by the quantiser. The fuzzy logic control (FLC) and the SCC constitute the feedback approach. The resulting quantisation scale is $Q_{af}(k)$ for the feedback approach. These two quantisation scales are selectively used according to the particular approach. These four functional blocks operate as integrated supervisory functions for rate control. The feed-forward approach uses the scene change features and the rate estimator to estimate the future buffer occupancy. They are represented by variances and MPEG encoding parameters such as picture type and motion vector values. The NQC is based on nonlinear functions which are adaptively controlled in association with the estimated video rate. The fuzzy logic control is located in between the buffer and the quantiser. This configuration improves its performance by controlling its scaling factors depending on the scene change features.

The operations of the boxed MPEG encoder are as follows: the analogue TV signal (PAL or NTSC) is first converted and digitised into the CCIR 601 format before being reformatted into the progressive SIF format. The picture encoding order is changed to $L_0P_0B$ for the motion estimation and compensation. The input format for the actual encoding loop (DCT–Q–IQ–IDCT) is the macro block which consists of four $8 \times 8$ luminance blocks and two chrominance blocks when the 4:2:0 format is used. The intra/inter decision is made after the motion compensated macro block, fetched from the frame memory, has been compared with the input macro block. The block data is then processed by the DCT and the quantiser. The inverse quantiser and the IDCT reconstruct the transformed and quantised data for encoding next input pictures. The quantised block data is run-length coded and variable-length coded before being multiplexed with header information. Finally, the compressed bit stream is stored in the transmission buffer for a short-period of time according to the delay requirement. The output of the buffer is transmitted through a CBR channel.
3.8 Configuration for simulation

For precise evaluation of the rate control schemes, the MPEG1 and MPEG2 software video encoders were verified at the lowest level, which comprises coded bit data. All the coded bit streams were tested using MPEG decoders which are available in the public domain; Berkeley [137] and MPEG Simulation Group [138]. An MPEG1 encoder and a decoder were developed with C++ programming language and they were verified by the Berkeley decoder. The simulation results presented in Chapter 4 are obtained from this encoder. All other simulations (Chapters 5 and 6) were conducted using the MPEG Simulation Group MPEG2 encoder and the decoder. The overall configuration is shown in Fig. 3.10.

![Configuration diagram](image)

Figure 3.10: Configuration of simulation.

A series of performance measures were used in the simulation; the occupancy, the number of coded bits per frame (BPF), peak signal-to-noise ratio (PSNR) and the normalised fluctuation of video rate (NFVR). The buffer occupancy was measured in terms of delay (ms) rather than percentage in order to take into account the delay requirement. Strictly speaking, the percentage occupancy does not properly represent delay information, as one cannot easily assess the incurred delay which is critical to interactive video communication. However, in this thesis, the percentage is appropriate since the buffer size is unchanged in the simulations. The BPF is given in terms of the number of encoded bits for a single frame. It is used to represent instantaneous variation of bit statistics in each picture. The PSNR [28, 31] is a widely accepted quality measure even though it is known not to be the best measure of visual quality. The PSNR is given:
\[
\text{PSNR [dB]} = 10 \log_{10} \frac{255^2}{\text{MSE}}
\]  
(3.9)

where MSE represents mean square error. The value 255 is the maximum grey level value when a picture has an 8-bit resolution. Thus, the PSNR represents the distortion in reference to the maximum grey level. Another measure is the NFVR (Chapter 5) which represents the overall video rate fluctuation for an entire simulated video sequence.

3.9 Conclusions

In this chapter, the background to video rate control was investigated, in particular, attention was focused on the buffer-based rate control and related quantisation techniques for CBR applications. The framework of the rate control algorithms, which are specified in the MPEG and ITU-T H.260 series standards, appear to be based on the same techniques, i.e. linear feedback control. The buffer occupancy is linearly mapped on to the quantisation step size and fed back to the quantiser. Rate-distortion analysis is not easily applied to MPEG-style codings due to the existence of the quantiser and other nonlinear encoding processes. Bearing this in mind, nonlinear and non-stationary signal analysis was discussed for estimating non-stationary video rate. In Chapters 4 to 6, these nonlinear approaches will be investigated and performance evaluation will also be conducted to assess their relative performance.
Chapter 4

Feed-forward rate control using scene change features

4.1 Introduction

In this chapter an efficient adaptive buffering scheme is described which takes a priori information representing the variation of input video into consideration. The a priori information is given by a set of scene change features. This scheme is based on a feed-forward control technique to handle the non-stationary nature of video. The scene change features are selected and applied to the adaptation of the quantisation scale.

The main objectives in this chapter are to extract a series of significant coding parameters which influence the video rate and video quality, and to examine their relationship with the video rate control scheme. An advanced rate control scheme is also aimed to be devised which takes into account the non-stationary properties of video rate. In particular, attention is focused on:

- The scene change features which can be used effectively in the developed video rate control scheme.
- A predictive rate control technique which exploits the scene change features.
- Quantisation control functions which respond effectively to dramatic rate changes.

Thus, the relationship of the video coding parameters will be described, which is associated with rate control and also discuss the properties of the selected scene change features from the rate control point of view. The performance of the quantisation control functions will be evaluated in terms of the occupancy and the video quality.

This chapter is organised as follows: In Section 4.2 a series of fundamental video coding parameters related to the quantisation process are investigated with respect to selected performance measures for variable bit rate codings. Section 4.3 describes the three critical parameters associated with the video rate control. In Section 4.4 the configuration of the scene change-based rate control scheme is described. Section 4.5 highlights the performance of the proposed scheme and Section 4.6 concludes this chapter.
4.2 Variable bit rate video coding

Although this research is focused on CBR applications, the VBR statistics on MPEG encoding need to be obtained first, in order to examine the effect of fixed quantisation step size on the video rate and to use it as a reference for performance monitoring purposes. In the CBR mode, the quantisation scale is controlled predominantly by the buffer occupancy and picture detail. If an input video sequence has a large amount of detail within a frame, or rapid motion between frames, the occupancy increases. Another important factor is the picture type. Since the picture type repeats, the buffer occupancy is liable to fluctuate more in the MPEG encoder than in the ITU-T H.261 encoder which has no such picture type. In variable bit rate transmission the quantisation step size may be fixed for a picture type or for all picture types. There is no need for using buffers since the output data can be transmitted at the required data rate in the VBR mode. However, in the CBR mode this picture type configuration may cause abnormal buffer operations, i.e. overflow or abrupt occupancy transitions. Fig. 4.1 shows the bit rate variation of the MPEG1 encoder when encoding the sport video sequence “Football”. It is shown at three transmission rates with a fixed buffer size of 327680 bits, which is the maximum size allowed in the MPEG1 standard for sub-primary rate communication [26].

![Graph showing video rate fluctuation at three channel transmission rates.](image)

Figure 4.1: Video rate fluctuation at three channel transmission rates.

The “Football” sequence has a large amount of motion throughout. At every I picture, a relatively large change in occupancy may occur since it subsequently generates the resulting number of coded bits. P pictures have smaller peaks since the encoding of them generates a smaller amount of data than I pictures. As most of the macro blocks of a B picture are coded with the bi-directional motion compensation mode, the data to be transmitted (i.e. prediction error) is smaller and hence troughs in the rate appear at B pictures, Fig. 4.1.

In order to investigate the relationship between the bit rate and the video quality, various combinations of the quantisation scales for the VBR mode were evaluated. Firstly, a series
of linear increases in the quantisation scale were applied to all three quantisation scales for I, P and B pictures. Let IQ, PQ and BQ be framewise quantisation scales for I, P and B pictures, respectively. All the three quantisation scale values are set to the same value such as 4, 7 or 10, etc. It is known that the number of coded bits varies depending on the picture type [26] when the input video only contains small motion. It is recommended in the MPEG1 standard that a larger number of bits are allocated to an I picture than P or B pictures since the picture type, I, generates the largest amount of video data. Thus, the B picture is allocated the largest quantisation scale value so that it may generate the smallest video rate, under the assumption that the input video contains no large scene variation. In the simulation, three different combinations of the quantisation scale values were tested, i.e. IQ:PQ:BQ = 2:3:6, 4:5:12 and 8:10:25. For IQ:PQ:BQ = 2:3:6 the mean quantisation scale value is 4.25 since the number of B pictures is 2. That is, for an I and a P picture followed by 2 B pictures, the mean value 4.25 is calculated by \((2+3+6 \times 2)/4\). Likewise, the mean values of 4.5:12 and 8:10:25 are 8.25 and 17.0, respectively.

Fig. 4.2 shows the profile of the VBR video rate and the corresponding peak signal-to-noise ratio (PSNR) for the simulated quantisation scale combinations for “Starwars”. The curves in Fig. 4.2 show the performance for identical quantisation scale values for the I, P and B pictures as given in the z-axis. For example, the value 7 in the x-axis represents the case of IQ:PQ:BQ = 7:7:7. The arrowheads show the performance when using different quantisation scale values such as 2:3:6 for the pictures as given in the descriptions in the top-right corner of each graph. The bit rate and its standard deviation decrease exponentially as the quantisation scale increases linearly, Fig. 4.2 (a) and (b). This nonlinear change originates from the nature of the discrete cosine transform (DCT). The magnitude of the DCT coefficients in a block is known to exponentially decrease as the spectral frequency approaches the highest value [45], i.e. from the DC component at top-left hand corner through the AC components down to the bottom-right corner in a \(8 \times 8\) DCT block. Therefore, a linear increase in the quantisation scale discards more and more of the DCT coefficients, decreasing the video rate rapidly. Subsequently, the PSNR curves show a similar profile, Fig. 4.2(c) and (d), and its standard deviation increases gradually. For the different combinations of the quantisation scales marked as 1, 2 and 3 with the arrowheads (e.g. IQ:PQ:BQ = 2:3:6) in Fig. 4.2, mean values of the video rate and the PSNR are nearly the same as their counterparts in the plots, e.g. IQ:PQ:BQ = 4:4:4. However, the standard deviations are far higher. Allocating a different quantisation scale to each picture type is expected to perform well for video with stationary statistics. However, for realistic video, such as “Starwars” and “JFK”, it degrades the encoding performance by causing large variations in the video rate and the quality. It appears that the allocation of the quantisation scale, based on the picture type information, is not appropriate for video quality control due to wide variations in PSNR. It results in large fluctuations both in the video rate and in PSNR as shown in Fig. 4.3. The precise values are shown in Table B.1 in Appendix B. The statistics on the VBR simulations for the “JFK” video sequence are also shown in Appendix B. See Fig. B.2, Fig. B.3 and Table B.2.
Figure 4.2: Bit rate and PSNR for various quantisation scale values for VBR (‘Starwars’). (a) Bit rate; (b) standard deviation of the bit rate; (c) PSNR; (d) standard deviation of PSNR.
Figure 4.3: Bit rate and PSNR variations depending on the quantisation scale (“Starwars”): (a) Coded bits/frame; (b) PSNR.

4.3 Coding parameters associated with the rate control

4.3.1 Channel rate

The channel rate specified in the MPEG1 standard may be used generally up to 1.5 Mbits/s. The standard can also be applied to higher channel rates, however, the optimum channel rate appears to be lower than 2 Mbits/s since the functionalities and the performance of MPEG1 are focused on low bit rates. In this research 1 Mbits/s and 1.2 Mbits/s channel rates are extensively tested for MPEG1. This channel rate is also used in the MPEG2 encoding in order to evaluate the performance of the rate control algorithms at the same bit rate, though the
MPEG2 algorithm is devised for higher bit rates. If the channel rate is set high enough to cover the peak rate of the VBR video, the necessity for rate control eventually diminishes. Therefore, it is set low enough to examine the performance of the rate control algorithms.

### 4.3.2 Buffer occupancy

The buffer occupancy fluctuates depending on short-term video rate. The smallest unit of the video rate is the number of bits per frame and per macro block. The framewise buffer occupancy at the time $k$, $O(k)$, is expressed as follows:

$$
O(k) = \frac{c(k) - c_{rate}/p_{rate}}{b_{size}} + O(k - 1) \text{ if } k > 0
$$

$$
O(0) = \frac{c(0)}{b_{size}} \text{ if } k = 0, \quad c(0) = b_{size}/2
$$

where $c(k)$ and $O(k - 1)$ are the number of coded bits for the $k$th picture and the previous occupancy, respectively. $c_{rate}$ and $p_{rate}$ represent the channel rate (bits/s) and the frame rate (frames/s), respectively. $b_{size}$ stands for the buffer size in bits. The initial condition of the occupancy is usually set to 50%, i.e. the half full state, before the encoding process starts.

### 4.4 MPEG1 encoder using feed-forward rate control and scene change features

An effective buffering and rate control technique is expected to accommodate large variations in the buffer occupancy without buffer malfunction, maintaining good subjective and objective video quality. In order to achieve these goals, an adaptive scheme is introduced which is a combination of a predictive technique and nonlinear quantisation control. Fig. 4.4 shows the block diagram of the buffering and rate control algorithm. In the proposed scheme three function blocks (the scene change calculator, rate estimator and Q scale control) are added to the MPEG1 video encoder. The encoder compresses the input video to two main coded information representations; motion vector, $mv(k, n)$, and quantised DCT coefficients, $dq(k, n)$. Here, $n$ is the macro block time index. These data are stored in the transmission buffer as the final coded bits, $Cv(k, n)$, after being multiplexed by the video mux. The scene change calculator estimates the extent of scene change by calculating two measures, $\text{var}_{-\text{org}}(k)$ (Eqn. 4.2) and $\text{var}_{-\text{dif}}(k)$ (Eqn. 4.3) to represent the variances of the input picture, $\text{SIF}(k)$ and its difference picture. The video rate estimator takes the current buffer occupancy as an input, and estimates the occupancy using these variances. It outputs the rate balance, $R_b$ (Eqn. 4.9), to the nonlinear quantisation scale control block which calculates the quantisation scale value using nonlinear (logarithmic/exponential and sigmoidal) equations. The nonlinear quantiser control block is also supplied with the current occupancy, $O(k - 1, n)$, and the motion vector function, $MVF(s)$ (Eqns. 4.4 and 4.5). Here $s$ is the slice (horizontal stripe of a picture with an equal size) time index. This scheme is based on both the scene change features and the feed-forward control. It exploits the information from the encoder, $\text{SIF}(k)$, motion vectors, $mv(k, n)$, and buffer occupancy, $O(k - 1, n)$. The rate control estimator uses the ratio of the
variances, \( V_R(k) \) (Eqn. 4.8), to quantify the scene change depending on the picture type. The ratio is used to estimate the number of coded bits in the current picture, in conjunction with previous bit rate statistics. The resulting predictive occupancy is eventually used to calculate the rate balance for a short period of time. The motion vector function, \( MVF(s) \), is the ratio of the summation motion vector in a slice. The nonlinear quantisation scale control calculates a scale value by changing the shape of its curve and locally tuning it with \( MVF(s) \). The final scale value, \( Q_s(k, n) \), is determined by \( O(k - 1, n) \). All the detailed operations for each rate control block will be explained in the following sections.

![Diagram of MPEG1 video encoder](image)

Figure 4.4: Feed-forward rate control scheme for MPEG1.

### 4.4.1 Selected scene change features

Three scene change features are used to estimate the buffer occupancy. They are the intra-frame variance of an input picture, \( var_{\text{org}}(k) \), the variance between two pictures, \( var_{\text{dif}}(k) \), and the summation of motion vector values in a slice of a picture, \( MVF(s) \). \( var_{\text{org}}(k) \) and \( var_{\text{dif}}(k) \) are given by:

\[
var_{\text{org}}(k) = \frac{\sum_{i=0}^{ROW-1} \sum_{j=0}^{COL-1} (org(k, i, j) - \mu_{\text{org}})^2}{(ROW \times COL)}
\]  \hspace{1cm} (4.2)

\[
var_{\text{dif}}(k) = \frac{\sum_{i=0}^{ROW-1} \sum_{j=0}^{COL-1} (dif(k, i, j) - \mu_{\text{dif}})^2}{(ROW \times COL)}
\]  \hspace{1cm} (4.3)

where the indices \( m \) and \( n \) represent the pixels in the row and column directions, respectively. \( \mu_{\text{org}} \) and \( \mu_{\text{dif}} \) are mean values of the original input picture, \( org() \), and its difference picture,
\( \text{dif}(k, i, j) \), respectively. \( \text{ROW} \) and \( \text{COL} \), respectively, represent the numbers of pixels of a row and a column. The \( \text{dif}(k, i, j) \) is obtained by subtracting \( \text{org}(k - 1, i, j) \) from \( \text{org}(k, i, j) \) on a pixel-by-pixel basis in the order of encoding pictures, i.e. IPBBPBB. Fig. 4.5 shows the variances.

![Graphs showing variances for different scenes](image)

Figure 4.5: Frame variances \( \text{var}_\text{org}(k) \) and \( \text{var}_\text{dif}(k) \): (a) Cascaded; (b) Starwars; (c) Adverts.

Two motion vector functions for the slice index \( s \), \( \text{MV} F_D \) and \( \text{MV} F_N \) are defined:

\[
\text{MV} F_D(s) = \frac{\sum_{n=0}^{\text{MBS}-1} \text{MV}_x(s, n) + \sum_{n=0}^{\text{MBS}-1} \text{MV}_y(s, n)}{\text{MBS} \times (\text{MVX} + \text{MVY})}
\]  

(4.4)
where MBS is the number of macro blocks in a single slice. \( m_{vx}(s, n) \) and \( m_{vy}(s, n) \) represent motion vectors of macro block \( n \) for row and column directions in the slice \( s \), respectively. MVX and MVY are the maximum values of motion vectors for a specified motion search range. \( MVF_D(s) \) and \( MVF_N(s) \) specify directional and non-directional motion vector functions, respectively. The subscripts “D” and “N” stand for “directional” and “non-directional”, respectively. Fig. 4.6 shows the two motion vector functions.

\[
MV F_N(s) = \frac{\sum_{n=0}^{MBS-1} |m_{vx}(s, n)| + \sum_{n=0}^{MBS-1} |m_{vy}(s, n)|}{MBS \times (MVX + MVY)}
\] (4.5)

Figure 4.6: Directional motion vector function, \( MVF_D \): (a) Cascaded; (b) Starwars; (c) Adverts.
The variances, \( \text{var}_\text{org}(k) \) and \( \text{var}_\text{dif}(k) \), shown in Fig. 4.5, represent the framewise variances for the three video test sequences used here. While \( \text{var}_\text{dif}(k) \) shows large fluctuations at scene changes, \( \text{var}_\text{org}(k) \) shows relatively smooth transition since it contains the variance in a frame which has less variations. This implies that \( \text{var}_\text{dif}(k) \) and \( \text{var}_\text{org}(k) \) can adequately represent scene change. “Cascaded” shows two distinctive \( \text{var}_\text{dif}(k) \) values at two scene changes, since its “football” section is cascaded with two completely different video sequences which only possess small motion. While “Adverts” shows many frame positions indicating scene changes, “Starwars” shows a small number of dramatic changes in \( \text{var}_\text{dif}(k) \) since it contains less scene change, in luminance and chrominance, throughout the sequence. The directional motion vector function shown in Fig. 4.6 is used to adjust the quantisation step size for a slice. Although the \( MVF_D(s) \) plots do not precisely follow the variance plots of Fig. 4.5, it can effectively be used for adaptively changing the value of the quantisation scale value for a slice. The non-directional motion vector function, \( MVF_N(s) \), shows wider variation in its plots, however, they are not presented since it is not used in this feed-forward rate control scheme.

### 4.4.2 Video rate balance, \( R_b \)

The predicted occupancy is used to select a suitable quantisation scale value in advance of encoding an input picture. It is calculated by estimating the number of coded bits for the current picture using previous video rates and the ratio of framewise variances. The shape of the control curves depends on the short-term change in the video rate, i.e. the video rate balance.

First, parameters for estimating the video rate balance and the predictive occupancy are defined as follows:

\[
\frac{c_{\text{rate}}}{p_{\text{rate}}} = \text{MBF (bits)}
\]  

(4.6)

where MBF stands for mean bits allocated per frame. The frame rate is set to the maximum, i.e. 25 Hz for the PAL standard and 30 Hz (precisely, 29.97 Hz) for the NTSC standard, respectively. Let \( C_{\text{bit}}(k) \) be the number of coded bits for \( k \)th picture.

\[
C_{\text{bit}}(k) = \sum_{i=k-L}^{k-1} C_{\text{bit}}(i)
\]

(4.7)

where \( C_{\text{bit}}(k - 1) \) represents the cumulative number of bits coded during the previous \( L \) frame period.

The decision to change the control curve is based on a comparison of framewise variances, i.e. intra/inter decision at the picture level, which provides an efficient measure for scene change. The variance is a framewise extension of macro-block variance specified in ISO11172-2 (MPEG1)
The visual content of each macro block can be considerably different from one to another in a picture, if the picture has complicated detail. However, since framewise ergodicity is assumed here and the scene change features are calculated framewise, it is practically feasible to adopt a framewise intra/inter decision. The framewise ergodicity introduced here is an application-specific concept which is appropriate for video rate control. Since a part of a picture, i.e. a macro block, can be very different from the others this concept may not be significant if the video rate prediction is carried out on a block-by-block basis. However, the prediction employed in this approach is considered in a framewise manner since rate control algorithms need to optimise the bit rate and PSNR relationship over a complete picture which represents the smallest unit in rate control. Thus, the framewise intra/inter decision is used here. Using the two variances \( \text{var}_{\text{org}}(k) \) and \( \text{var}_{\text{dif}}(k) \), the whole area for the variance comparison is divided into four regions as shown in Fig. 4.7. “A” represents the area with no dramatic scene change and no subsequent abrupt changes in the video rate since the \( \text{var}_{\text{dif}}(k) \) is too small. “B” is the area with higher \( \text{var}_{\text{dif}}(k) \), however, no dramatic video rate change occurs in this area since \( \text{var}_{\text{org}}(k) \) is much smaller than \( \text{var}_{\text{dif}}(k) \). In this case, as the input picture appears to have simpler visual content, the DCT encodes macro blocks mainly with the intra mode and its performance is believed to be better than encoding them with the inter mode. The areas “C”, and “D” may cause more dramatic increases in video rate. For the areas “C” and “D” the proposed buffering scheme begins to operate. They are separated by the line of the equation \( y = x \).

![Figure 4.7: Framewise intra/inter decision using variances.](image)

Estimating the number of coded bits of the \( k \)th input frame is based on the ratio of \( \text{var}_{\text{dif}}(k) \) and \( \text{var}_{\text{org}}(k) \) and previous statistics of coded bits per frame, i.e.:

\[
\tilde{C}_{\text{bits}}(k) = \frac{\text{var}_{\text{dif}}(k)}{\text{var}_{\text{org}}(k)} \times N_{\text{bits}}(P(k)) \quad \text{for the area “C”}
\]

47
\[
\tilde{C}_{bit}(k) = \frac{\text{var}_- \text{org}(k)}{\text{var}_- \text{diff}(k)} \times N_{bit}(I(k)) \quad \text{for the area “D”}
\]

where \(I(k)\) and \(P(k)\) represent the nearest I and P pictures involved in estimating the coded bits for the \(k\)th picture, respectively. \(N_{bit}(P(k))\) and \(N_{bit}(I(k))\) represent the number of coded bits for the nearest P and I pictures from the \(k\)th picture. Either \(\tilde{C}_{bit}^p(k)\) or \(\tilde{C}_{bit}^q(k)\) is used for all three picture types since the input picture included in the area of either “C” or “D” is assumed to have similar encoding statistics to those of the previous I or P picture. The ratio of the two variances, \(\text{var}_- \text{diff}(k)/\text{var}_- \text{org}(k)\) or \(\text{var}_- \text{org}(k)/\text{var}_- \text{diff}(k)\), corresponds to \(VR(k)\), Fig. 4.4. In the case of a large scene change, the statistics of an input picture, e.g. the number of coded bits, follows that for the previous I or P picture rather than that of the B picture since the normal statistics of a B picture (i.e. the lowest profile in coded bits) appears only when encoding a video with high correlation between pictures. As the pictures in both the areas “C” and “D” may generate a bursty compressed video, it is necessary to choose either \(\tilde{C}_{bit}^p(k)\) or \(\tilde{C}_{bit}^q(k)\) according to the decision on whether the image falls into the area “C” or “D” in Fig. 4.7. After this estimation is carried out at each frame start, using the afore-mentioned definitions, the video rate for a short period of time, e.g. a series of pictures, BBP or BBI, can be characterised, as the video rate balance, \(R_b\):

\[
R_b = \frac{C_{bit}(k-1) + \tilde{C}_{bit}(k)}{(L + 1) \times \text{MBF}}
\]  

(4.9)

where \(\tilde{C}_{bit}(k)\) is either \(\tilde{C}_{bit}^p(k)\) or \(\tilde{C}_{bit}^q(k)\) depending on the decision. The constant \(L\) is the number of frames in the series of pictures, i.e. for the picture type repetition of BBP or BBI, \(L\) becomes 3. The value of the video rate balance, \(R_b\), can be interpreted as follows:

\[
R_b \simeq 1; \quad \text{balanced video rate, (short-term rate } \simeq \text{ MBF)}
\]

\[
\gg 1; \quad \text{overflowed video rate, (short-term } \gg \text{ MBF)}
\]

\[
\ll 1; \quad \text{underflowed video rate, (short-term } \ll \text{ MBF)}
\]

(4.10)

\(R_b\) represents the short-term (a few frame periods) history of the video rate. Since \(R_b\) contains the buffer occupancy information for a few frames rather than a single frame, the buffer fluctuation due to scene changes in a shorter period than the \(L\) frame period is negligible. \(R_b\) is used to determine a quantiser control curve which is applied to a whole picture. The current occupancy, \(O(k-1, n)\), and the motion vector function, \(MVF(s)\), are used along with \(R_b\) to select the final quantisation scale value.

### 4.4.3 Nonlinear quantisation control

Linear control of the quantisation scale works satisfactorily under normal operating conditions in a large buffer if the occupancy remains around 50%. However, in cases of lower bit rates or abrupt scene changes, the occupancy may increase drastically. As shown in Fig. 4.8, at the lowest two bit rates, the buffer cannot accommodate the incoming video rate, and buffer
overflow soon occurs.

At 768 kbit/s the buffer does not overflow for the first 100 video frames (1500 slices shown in Fig. 4.8). However, if the input video has more information coming in than for the previous frames, the buffer will overflow at a future time point. While the output bit rate of the transmission buffer is constant, the input rate is highly variable. If the input bit rate is equal to the output rate, then the occupancy does not change and keeps a specific, fixed value. However, the occupancy changes continuously since the input video data and the picture type vary. If the change in the input bit rate is constant and greater than the output rate, then the occupancy increases linearly. Furthermore, if it is not constant, e.g. increasing, the occupancy may increase in a nonlinear way. For example, the four plots of lower bit rates in Fig. 4.8 show a nonlinear increase in occupancy up to the slice number 400. All plots in Fig. 4.8 display the results of linear control via the equation, \( Q_s(k, n) = O(k - 1, n) \), where \( Q_s(k, n) \) and \( O(k - 1, n) \) are the normalised quantisation scale and the occupancy values, respectively.

![Figure 4.8: Occupancy for a 2-frame buffer at different transmission rates ("Football").](image)

The goal of rate control in the CBR mode is to effectively control the quantisation step size depending on the buffer occupancy. Several different quantiser control functions have been investigated and published in the literature. They can be classified into linear [25, 42], piecewise linear [41, 85], and nonlinear [80, 83, 128]. The nonlinear control is an adaptive way to allocate a quantisation scale value to compensate for the dramatic changes in the occupancy. In this chapter two nonlinear control functions, the sigmoidal [128] and logarithmic/exponential [83] are investigated. In the sigmoidal function, two nonlinear equations form a set of nonlinear function curves with the shape of a skewed S, Fig. 4.9.

1 In later chapters, the logarithmic/exponential control function is also called unimodal function. Both functions are conceptually equivalent but the equations are different.
Let $Q_s(k, n)$ and $O(k - 1, n)$ be the normalised step size and the current occupancy, respectively.

$$Q_s(k, n) = \alpha \left( \frac{1}{\alpha} O(k - 1, n) \right)^{R_b} \quad 0 \leq O(k - 1, n) < \alpha$$

$$= 1 - (1 - \alpha) \left( \frac{1}{1 - \alpha} \left( 1 - O(k - 1, n) \right) \right)^{R_b} \quad \alpha \leq O(k - 1, n) \leq 1 \quad (4.11)$$

where the video rate balance $R_b$ becomes a steepness factor which determines the shape of the curve. $\alpha$ is a control factor to determine the symmetry of each curve, e.g. if $\alpha$ is 0.5, the curve shows a symmetrical shape whose upper half and lower half are the same shape. The combination of two curves forms the sigmoidal curve. If $R_b = 1$ then $Q_s(k, n) = O(k - 1, n)$, and linear control is achieved. When $R_b$ is smaller than 1, the curve with $R_b$ is also used.

![Figure 4.9: Sigmoidal quantisation scale control.](image)

In the logarithmic/exponential control curve the quantisation scale is mapped to a set of logarithmic and exponential curves. A control curve for the quantisation scale is selected based on $R_b$ for a specific period of time and the current occupancy, in the same way as Eqn. 4.11, i.e.:

$$Q_s(k, n) = \alpha(R_b) \log_{10}(\beta(R_b)O(k - 1, n) + 1), \quad R_b > 1$$

$$= \rho(R_b) \exp(\mu(R_b)O(k - 1, n)) - \mu(R_b), \quad 0 < R_b < 1 \quad (4.12)$$

where $\alpha(R_b)$, $\beta(R_b)$, $\rho(R_b)$ and $\mu(R_b)$ are real values to determine the steepness of the logarithmic and exponential curves so that they can be fitted into the normalised occupancy and step size ranges. Since a limited number of logarithmic/exponential curves are used, the number of the real values is equal to the number of curves. For computational simplicity, they are stored in a lookup table so that they can be selected by the $R_b$ value. If the $R_b$ value is far bigger or smaller than 1, a more nonlinear curve is selected. Otherwise, if $R_b$ stays close to 1, the selected curve becomes linear, Fig. 4.10.
4.4.4 Local adaptation of the quantisation scale

Although framewise ergodicity is assumed in the previous section, in practice, a part of a picture may have a very different amount of visual information from the other parts of the picture. Therefore, the value of $Q_s(k, n)$ needs to be changed to the quantisation scale for each slice. In order to adapt $Q_s(k, n)$ to the quantisation scale for each slice, the motion vector function defined in Eqn. 4.4 is used. If the motion in a slice has a tendency towards a uniform direction, $MV_F_D(s)$ and $MV_F_N(s)$ will have a similar value. If the motion is random or small in a slice, $MV_F_D(s)$ tends to have a far smaller value than $MV_F_N(s)$, since $MV_F_D(s)$ does not include an absolute operation for each motion vector. On the other hand, as $MV_F_N(s)$ sums absolute values of motion vectors, it represents the total amount of motion in a slice regardless of the direction of the motion. A slice with a higher $MV_F_D(s)$ value is assumed to generally entail a larger amount of data and $MV_F_D(s)$ properly works as the measure for the local adaptation, thus $MV_F_D(s)$ is used for the slice-wise adaptation of the quantisation scale. $MV_F_D(s)$ ranges from 0 to 1, thus in order to make $MV_F_D(s)$ a scaling factor, it is multiplied by 2 and applied to $Q_s(k, n)$, i.e.:

$$Q_{ss}(k, n) = 2 \times MV_F_D(s) \times Q_s(k, n)$$ (4.13)

where $Q_{ss}(k, n)$ is the quantisation scale for the slice $s$. If the $MV_F_D(s)$ value is 0.5, $Q_{ss}(k, n)$ becomes equal to $Q_s(k, n)$.

4.4.5 Computational requirement

In this section the computational burden of the proposed scene change-based buffering scheme is assessed with respect to the MPEG1 encoder. Since the functions employed in the MPEG1
encoder can be implemented in a variety of different ways, no unique calculation can be provided. In this section the computational requirement of the scene change features is approximately assessed by only considering the major functional blocks of the video encoder such as motion estimation, DCT, quantisation, etc. The calculation is based on the software encoder used in simulations, which has the exhaustive search motion estimation and the DCT without using a fast algorithm. For an I picture, the intra/inter decision is not required since all the macro blocks are coded in the intra mode. However, the I picture from the intra/inter calculation is not disregarded since the number of I pictures in a sequence is generally far smaller than those of P and B pictures. Table B.3 and Table B.4 in Appendix B summarise the number of additions and multiplications required in encoding a single picture. Here, SWX and SWY stand for the size of the search window for the motion estimation, e.g. 46 pixels by 46 lines. MB and BL represent the sizes of a macro block and a block, i.e. 16 by 16 and 8 by 8, respectively. MBN represents the number of macro blocks in a picture, which is 330 in the simulation. \( \mu \) and \( \delta \), as shown in Table B.4, are the mean values for \( \text{var}_\text{org}(k) \) and \( \text{var}_\text{diff}(k) \), respectively. \( \text{ROW} \) and \( \text{COL} \) are set to 240 lines and 352 pixels, respectively. \( \text{DCT}[l, i] \) represents the DCT basis function. \( \text{QMAT}(k, l) \) is the quantisation matrix used in the quantisation and the inverse quantisation processes. The asterisk shown in Table B.4 means that the corresponding calculation does not entail significant computational burden to the encoder system, hence they are not included in the computational requirement. The motion vector function, \( MV_F_D(s) \), also involves a few additions in a picture, however, its computational load is negligible, thus it is not included.

The order of computation for the MPEG1 encoder appeared to be far larger than that for the calculation of the scene change features as compared in the Tables B.3 and B.4. The total load of computation for the encoder is represented by the order of the computation using a logarithmic scale for complexity comparison, i.e.:

\[
\log_{10}(\text{# of multiplications}) = \log_{10}(361,912,200) = 8.5586
\]
\[
\log_{10}(\text{# of additions}) = \log_{10}(8,384,640) = 6.9235
\]

On the other hand the total computational requirement for calculating the two variances is far smaller:

\[
\log_{10}(\text{# of multiplications}) = \log_{10}(168,960) = 5.2278
\]
\[
\log_{10}(\text{# of additions}) = \log_{10}(506,880) = 5.7049
\]

For multiplication, the computational load required for the two variances is smaller than 1% (addition, lower than 10%) of that for the complete MPEG1 encoder. This signifies that the scene change features do not entail much computational burden to the encoding system. Therefore, the proposed scheme can be implemented for the encoder in a cost effective way.
4.5 Simulation results

4.5.1 Simulation environment

The software MPEG1 video encoder and the decoder, described in Section 3.7, were used to conduct the verification of the rate control techniques discussed in this chapter. The encoder has been rigourously verified to confirm the full compatibility with the ISO/IEC11172 standard. The encoder accepts 352 pixels by 240 lines colour pictures and generates a compressed bit stream file. It also outputs a series of statistics files to show the performance of the encoder. The compressed file can be decoded by any standard MPEG1 decoders and contains visual information with PSNR fugures at a specific bit rate. All compressed files have been successfully decoded by the Berkeley decoder [137]. The decoder decodes a bit stream file and generates various statistics files to analyse the encoding parameters encapsulated in the coded bit stream. All the video sequences were encoded at 1024 kbits/s and at 30 frames/s. The occupancy is measured in percentage for the fixed buffer size of 327680 bits specified in the MPEG1 standard. At the chosen channel rate the buffer can contain approximately up to 8 frames. In order to observe the encoder performance for realistic video signals three different types of video sequences were tested which contain manipulated and natural scene changes. Fig. A.1, A.2, and A.3 show sample pictures of the sequences; “Cascaded”, “Starwars”, and “Adverts” (See Appendix A).

4.5.2 Performance evaluation

Six buffering methods were simulated for the three video sequences. The buffering schemes can be classified into adaptive and non-adaptive. The adaptive scheme has been applied to sigmoidal and logarithmic/exponential quantiser control curves which change the steepness of the curves depending on the video rate balance. The non-adaptive method uses a single control curve. The linear control (LIN) uses a linear relationship which is used as a performance benchmark in the simulation. In the logarithmic (LOG) and the exponential (EXP) techniques, the curves A and B shown in Fig. 4.10 are used, respectively. The sigmoidal control (SIGM) is based on the curve R shown in Fig. 4.9. These four schemes are non-adaptive in that the steepness remains constant. The adaptive algorithm LOGEXP-A forms a combined control composed of the logarithmic and the exponential curves with the adaptation explained in Section 4.4. SIGM-A is also an adaptive scheme to employ multiple sigmoidal curves, which adaptively changes the steepness.

Table 4.1 summarises the performance of the schemes according to the performance measures. The mean occupancy of LOG and LIN appears to be much lower than EXP and SIGM performs in between EXP and LOG/LIN. Since EXP takes lower quantisation scale values at low occupancy, it generates more coded bits at the low and the initial occupancies. This results in very high average occupancies. SIGM exhibits better capability to control the occupancy by
changing the quantisation scale with wider variation thus its standard deviation is slightly larger than the other non-adaptive ones. SIGM-A and LOGEXP-A appear to have better performance in comparison to the non-adaptive techniques since they adaptively change the control curves according to the previous history of the occupancy. LOGEXP-A controls the quantisation scale in a more adaptive way, as shown in the standard deviations (std. dev.) of the quantisation scale in Table 4.1, hence it maintains the occupancy with less fluctuation. The PSNR profile does not show noticeable difference over the four non-adaptive methods. This implies that the overall PSNR value is not influenced significantly since the average quantisation scale maintains a similar value. For “Starwars” and “Adverts” video sequences, the PSNR value and its profile of LOGEXP-A and SIGM-A keep some 1 dB lower than other non-adaptive methods due to wider-range quantiser control.

<table>
<thead>
<tr>
<th>Cascaded</th>
<th>LIN</th>
<th>LOG</th>
<th>EXP</th>
<th>SIGM</th>
<th>SIGM-A</th>
<th>LOGEXP-A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantisation scale</td>
<td>mean</td>
<td>10.92</td>
<td>10.94</td>
<td>10.85</td>
<td>10.95</td>
<td>11.73</td>
</tr>
<tr>
<td></td>
<td>std. dev.</td>
<td>8.37</td>
<td>8.41</td>
<td>8.48</td>
<td>8.50</td>
<td>8.32</td>
</tr>
<tr>
<td>Occupancy (%)</td>
<td>mean</td>
<td>32.44</td>
<td>20.67</td>
<td>58.90</td>
<td>42.42</td>
<td>42.55</td>
</tr>
<tr>
<td></td>
<td>std. dev.</td>
<td>26.92</td>
<td>19.81</td>
<td>22.08</td>
<td>12.79</td>
<td>12.12</td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>mean</td>
<td>34.01</td>
<td>33.92</td>
<td>34.07</td>
<td>33.97</td>
<td>33.43</td>
</tr>
<tr>
<td></td>
<td>std. dev.</td>
<td>8.33</td>
<td>8.28</td>
<td>8.35</td>
<td>8.35</td>
<td>8.02</td>
</tr>
</tbody>
</table>

(a)

<table>
<thead>
<tr>
<th>Starwars</th>
<th>LIN</th>
<th>LOG</th>
<th>EXP</th>
<th>SIGM</th>
<th>SIGM-A</th>
<th>LOGEXP-A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantisation scale</td>
<td>mean</td>
<td>10.98</td>
<td>11.03</td>
<td>10.98</td>
<td>10.96</td>
<td>11.83</td>
</tr>
<tr>
<td></td>
<td>std. dev.</td>
<td>2.18</td>
<td>2.39</td>
<td>2.28</td>
<td>2.69</td>
<td>2.40</td>
</tr>
<tr>
<td>Occupancy (%)</td>
<td>mean</td>
<td>32.54</td>
<td>16.52</td>
<td>64.91</td>
<td>44.36</td>
<td>35.16</td>
</tr>
<tr>
<td></td>
<td>std. dev.</td>
<td>6.93</td>
<td>4.67</td>
<td>6.71</td>
<td>4.03</td>
<td>3.18</td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>mean</td>
<td>33.16</td>
<td>33.10</td>
<td>33.16</td>
<td>33.08</td>
<td>32.66</td>
</tr>
<tr>
<td></td>
<td>std. dev.</td>
<td>4.98</td>
<td>4.97</td>
<td>4.99</td>
<td>5.20</td>
<td>4.86</td>
</tr>
</tbody>
</table>

(b)

<table>
<thead>
<tr>
<th>Adverts</th>
<th>LIN</th>
<th>LOG</th>
<th>EXP</th>
<th>SIGM</th>
<th>SIGM-A</th>
<th>LOGEXP-A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantisation scale</td>
<td>mean</td>
<td>13.12</td>
<td>13.20</td>
<td>13.15</td>
<td>13.35</td>
<td>13.81</td>
</tr>
<tr>
<td></td>
<td>std. dev.</td>
<td>4.81</td>
<td>4.93</td>
<td>5.06</td>
<td>5.36</td>
<td>4.94</td>
</tr>
<tr>
<td>Occupancy (%)</td>
<td>mean</td>
<td>39.45</td>
<td>22.13</td>
<td>68.99</td>
<td>47.01</td>
<td>39.61</td>
</tr>
<tr>
<td></td>
<td>std. dev.</td>
<td>15.45</td>
<td>10.69</td>
<td>13.66</td>
<td>7.67</td>
<td>6.92</td>
</tr>
<tr>
<td>PSNR (dB)</td>
<td>mean</td>
<td>32.53</td>
<td>32.47</td>
<td>32.53</td>
<td>32.42</td>
<td>32.13</td>
</tr>
<tr>
<td></td>
<td>std. dev.</td>
<td>6.74</td>
<td>6.73</td>
<td>6.78</td>
<td>6.76</td>
<td>6.51</td>
</tr>
</tbody>
</table>

(c)

Table 4.1: Summary of the performance of the six rate control techniques: (a) Cascaded; (b) Starwars; (c) Adverts.
Fig. 4.11 and 4.12 show critical parts of the simulation results for the six methods, in terms of the quantisation scale and the occupancy, which highlight sections of 90 slices (6 pictures) in each video sequence. Fig. 4.13 shows PSNR profiles where the rate control algorithms should operate effectively (Frame number corresponds to slice number divided by 15). The corresponding sections to Fig. 4.11 and Fig. 4.12 are marked by two vertical lines. SIGM-A shows more stable variations in occupancy, however, at dramatic changes LOGEXP-A appears to be more effective since it responds quickly to the occupancy increase. This is achieved by a rapid control of the quantisation scale, as shown in Fig. 4.11. The four non-adaptive methods exhibit no significant performance difference. Degradation in PSNR is inversely proportional to the occupancy. Fig. 4.13, since the larger quantisation step sizes degrade the picture quality.

For the non-adaptive methods SIGM shows a better performance since the quantisation scale increases quickly as the occupancy increases rapidly. The performance of LOG is better than LIN. However, both are poorer than SIGM since their response at high and low occupancies is slower due to their curvature. The occupancy of EXP resides between SIGM and LOG. However, since it keeps lower scale values at low occupancy than SIGM and LOG, it generates more coded bits at low and initial occupancies. The EXP technique outputs far smaller scale values at either a lower occupancy level or when the coding starts from an empty buffer. Though EXP works properly at higher occupancy, the previous lower scale values at lower occupancy allow the encoder to generate a larger amount of coded bits. Due to this feature, its performance in the occupancy control is inferior to the other non-adaptive methods, as shown in Fig. 4.12. Therefore, EXP is not appropriate for bursty video. SIGM and LOG show relatively stable occupancy fluctuations. While SIGM maintains the occupancy around 50% of the buffer with a smaller variance, the occupancy of LOG remains below that of SIGM. Generally, SIGM shows smaller variation in the occupancy. In a case without adaptation, SIGM is better than LOG at the same steepness of the quantisation control curves.

SIGM-A and LOGEXP-A outperform the four non-adaptive methods since they adaptively change the quantiser control curves depending on the previous occupancy. SIGM-A shows smaller fluctuations in the occupancy on the whole, however, at dramatic scene changes LOGEXP-A performs better since it responds more quickly to the occupancy increase. For all three video sequences, SIGM-A maintains a flatter overall occupancy level than the others. However, in terms of the capability to accommodate an abrupt increase in occupancy, LOGEXP-A is better than SIGM-A. SIGM-A and LOGEXP-A show similar PSNR profiles as they have similar performance on a global scale, Table 4.1.
Figure 4.11: Variation in the quantisation scale value selected for 6 feed-forward buffering techniques: (a) Cascaded; (b) Starwars; (c) Adverts.
Figure 4.12: Variation in the occupancy for 6 feed-forward buffering techniques: (a) Cascaded; (b) Starwars; (c) Adverts.
Figure 4.13: Variation in PSNR for 6 feed-forward buffering techniques: (a) Cascaded; (b) Starwars; (c) Adverts.
4.5.3 Discussion

When the video sequence has no dramatic scene changes or its motion is small, a majority of macro blocks are coded using the inter mode. However, if the motion is large enough to turn a macro block into the intra mode, a whole P or B picture can be virtually identical to an I picture, since only a few macro blocks are coded with the inter mode. This is a typical case for realistic video sequences such as “Starwars” and “Adverts” (See Fig. B.4 in Appendix B). For “Cascaded”, the segment of “Miss America” has no noticeable scene changes, “Susie” has smaller number of intra macro blocks in the middle of the segment. On the other hand, “Football” has higher peaks throughout the segment. Above all, there are two distinctive peaks at each scene change where the number of the intra macro blocks is greater than 300 out of a total of 330 macro blocks. Such scene change can be detected by setting a threshold value and the information can be used for rate control.

“Starwars” and “Adverts” show more significant peaks than “Cascaded” since they have more scene changes and their motion is more rapid. Figs. B.6, B.7 and B.8 in Appendix B show the bit statistics, the corresponding occupancy and PSNR profiles for the linear quantiser control at two different channel rates of 1280 and 1024 kbits/s. While “Cascaded” shows higher correlation within each segment of the sequence with the repetition of a series of bit values depending on picture type, i.e. IBBPBBB..., the others show virtually no such repetition since they contain many dramatic scene changes and a longer period of rapid motion. Two occupancy plots at different channel rates show important properties of occupancy change. As indicated in Fig. B.7 (a) and (c), the rate of increase at higher levels of occupancy is far greater than that at lower levels (A > B) despite the corresponding increase of quantisation scale, Fig. B.5. This implies that the linear quantiser control does not effectively handle the occupancy increase at higher occupancy levels, and illustrates the nonlinear nature of the relationship between occupancy and quantisation scale. On the other hand, “Starwars” shows no noticeable difference in occupancy change throughout the sequence, since the occupancy remains at a lower level.

4.6 Conclusion

In this chapter an investigation was conducted on a feed-forward buffering scheme for video rate control using a series of scene change features and non-linear quantisation. A series of critical statistics and features of the encoding process have also been investigated, which are closely related to buffering and rate control for the constant bit-rate applications of the MPEG1 video encoder. An advanced algorithm for buffering and rate control has been proposed and verified by a software encoder and decoders. The framewise variances and the motion vector value in a slice have been effectively used as scene change features, to provide the predictive quantisation scale control, as the scene change detection is performed before encoding the pictures. Since the video rate is not stationary and occasionally bursty, a feed-forward control scheme using scene change information has appeared to be a promising approach for video rate control.
control. Multiple quantisation control curves were applied by a selection criterion according to a short-time history of occupancy, the rate balance. This adaptive technique appears to be more effective than using a single control curve in which the steepness is fixed. The performance of several different quantisation control curves (linear, sigmoidal, logarithmic and exponential) has been evaluated. Among those various combinations, the adaptive scheme using combined logarithmic/exponential curves (LOGEXP-A) shows better performance than the others in terms of stability of the occupancy. This scheme regulates the incoming video rate and adapts it to the transmission buffer, hence the encoder is able to control effectively abnormal buffer operation, such as overflow. The scheme does not rely either on expanding the size of the buffer or discarding input pictures, and the coding delay of the encoder remains unchanged, which is a critical parameter in bi-directional communications. Thus, this scheme improves the performance of buffering and rate control mechanism, with only a minimal degradation in PSNR.
Chapter 5

Video rate control based on nonlinear rate estimation

5.1 Introduction

In Chapter 4 the contribution of the scene change features and the nonlinear quantisation control functions was evaluated in respect of the proposed video rate control scheme. The technique was a linear adaptive control which exploited the predictive nature of the video rate using scene change feature information.

The approach presented in this chapter starts from the widely accepted assumption that the video rate is a correlated time series [80, 91]. This implies that the video rate can be estimated by using linear prediction techniques. However, linear prediction performs effectively only when the video contains small motion, i.e. no dramatic scene change occurs. A more effective rate control technique needs to be applied when video contains rapid motion or frequent scene changes. Although one can achieve a certain level of performance by using linear techniques which exploits short-term correlation of video, the performance can be further improved if nonlinear predictive technique is used. That is, the estimation and control of the video rate is considered to be ineffective when only using conventional linear prediction techniques, since realistic video has rather different properties from video with no large scene variation. In contrast to the linear technique adopted in Chapter 4, a nonlinear feed-forward prediction technique is used in this chapter, in order to achieve the further improvement. The scene change features used in Chapter 4 are also used as inputs to video rate estimators described in this chapter.

The main objective in this chapter is to assess the feasibility of the estimator-based rate control scheme, the performance of the conventional techniques and an emerging nonlinear technique based on radial basis function (RBF) networks. Thus four main topics in this chapter are:

- Identifying a linear or nonlinear relationship between scene change features and the video rate time series.

- Evaluating the performance of the nonlinear quantisation control technique which operates in the MPEG2 video encoder.
• Developing an estimator-based MPEG2 video rate control scheme which employs a linear or a nonlinear prediction technique.

• Comparing the performance of linear moving average (MA) models and the RBF rate estimator in terms of the buffer occupancy and the video quality.

In this chapter, the performance of the rate control algorithms will be evaluated with heuristic, linear and nonlinear rate estimators on the MPEG2 Test Model 5 video encoder, by observing performance measures such as the buffer occupancy, number of coded bits per picture and peak signal-to-noise ratio (PSNR).

This chapter is organised as follows: Section 5.2 explains the background assumption of MPEG2 video encoding in a descriptive way. Section 5.3 highlights the fundamental limitation of the rate control described in the MPEG2 standards. In Section 5.4 system identification is discussed, presenting the results of linear prediction. Section 5.5 describes the configuration of the rate estimator-based MPEG2 encoder, including nonlinear quantisation control. Section 5.6 presents an analysis of the nonlinear quantiser control functions in a rate-distortion theoretic way. In Section 5.7 the configuration of the RBF rate estimator is explained. Section 5.8 includes the simulation results and discusses performance aspects between the RBF and linear estimators. Finally, Section 5.9 concludes this chapter.

5.2 An encoder model based on a stochastic interpretation

Given that video rate is non-stationary around scene changes, a feed-forward, estimator-based rate control scheme is proposed which is supplied with the scene change information and combined with the MPEG2 video encoder. The scene change information is used to estimate the video rate. The estimated video rate is then used for future buffer occupancy estimation and enables the encoder system to adapt the quantisation step size. The future occupancy and the current occupancy are finally exploited in the nonlinear quantiser control scheme to select an appropriate quantiser control curve between the two occupancies and the quantisation step size.

It is assumed, in this chapter, that the MPEG2 video encoder preserves the statistical properties between its input (video) and output (compressed bit stream). If the input video has more redundancy, the encoder provides a higher compression ratio and outputs a virtually fixed rate bit stream. However, for realistic video, dramatic scene changes may degrade the performance of the encoder and result in fluctuations of the video rate. A linear change in the statistical properties of the input produces a similar outcome. However, since the MPEG video encoding is a nonlinear process, in that it includes several nonlinear functions, as explained in Section 3.4, the statistics of the compressed bit stream may often show nonlinear properties. Thus, digital video is viewed as a signal with non-stationary statistics and the MPEG video encoder as a nonlinear process. In the process of the MPEG video encoding, the stationary properties of the digital video are assumed to be preserved. On the other hand, the linear properties may lead to
Chapter 5: Video rate control based on nonlinear rate estimation

a different outcome. The linear change in the encoding parameters is assumed not necessarily to result in a corresponding linear change in the video rate. A linear decrease in channel rate may cause a nonlinear increase in the video rate, and subsequently, the buffer occupancy (See Fig. B.7).

Digital video has short-term correlation in the temporal direction between adjacent frames as well as the spatial correlation within a picture. It is feasible to exploit these correlations when video compression is performed [77]. This is considered to be applicable in the case of video with small motion or an aggregated video bit stream from multiple video sources, which is transmitted through a high capacity channel. Here, the correlation is used in conjunction with scene change features on the basis that large variation in video content co-exists with it. In order to achieve the desired control performance, three scene change features are used to indicate how dramatically a scene has altered. These features represent the 1st and 2nd-order scene change features: intra-frame variance, inter-frame variance and picture type value. These variances were already used in Chapter 4, however, the picture type is introduced in this chapter, instead of the motion vector function, \( MVF_p(\cdot) \) in Eqn. 4.4, to achieve improved performance.

Buffer-based rate control, as described in Chapter 4, is the technique which smoothes burstiness in the data rate of compressed video. The buffer occupancy value at a specific time informs the quantiser which step size value should be used for the next encoding operation. In this chapter the mechanism for controlling the occupancy is enhanced by predicting the future occupancy. The future occupancy is assessed by a one-frame-ahead buffer capacity estimate before encoding the next picture. The quantiser then changes the step size in advance of the arrival of the picture data.

Rate estimators with an improved performance were designed and evaluated. First, linear estimators were tested, i.e. moving average (MA) estimators trained with least squares and the RLS algorithms, which are widely used in the field of adaptive control. A radial basis function (RBF) network was then adopted as a nonlinear estimator. The RBF network is known to have better estimation performance than linear predictors and has recently been used successfully in several engineering applications such as channel equalisation [139]. Although the RBF-network is classified as a neural network, it is computationally simpler than general neural networks which have multiple hidden layers [8] since it has only a single hidden layer, i.e. the RBF layer.

5.3 Limitations of the rate control technique employed in the MPEG2 TM5

The ISO13818-2 MPEG2 standard sets up the baseline for the MPEG2 video encoder interfaced with the video buffering verifier (VBV) rate control scheme based on feedback buffer control for constant bit rate channels. It specifies the size of the VBV buffer for constant bit rate channels, providing desired buffer occupancies and quantisation scale values as critical parameters. However, it describes global system aspects rather than the details of the rate control process itself. While the ISO11172-2 MPEG 1 standard specifies a single type of quantisation scale,
ISO13818-2 provides two different types; linear and nonlinear. In a strict sense, the latter is piecewise linear and globally exponential.

The test model [42] specifies the detail of the rate control as a performance reference for other algorithms. The rate control scheme employed in the Test Model (TM5) specifies transmission parameters such as bit rate, frame rate, picture complexity and virtual buffer occupancy to determine the quantisation scale. This rate control algorithm is based on a one-step ahead estimation using the picture complexity measure. It allocates a certain number of bits to each type of picture according to the equations for the number of target bits. This process consists of three steps: firstly, estimating target bits for the next frame by using the complexity measure which is the product of the number of coded bits and a mean quantisation scale for the previous picture. Secondly, the rate control is achieved by calculating the quantisation step size for a whole picture using the current buffer occupancy. Finally, local adaptation of the quantisation step size is obtained for each macro block using a normalised activity. The buffer occupancy is calculated by subtracting the actual number of coded bits from the allocated target bits. The quantisation reference parameter, $Q_j$, for the quantisation scale value is calculated by the following equation [42]:

$$ Q_j = \frac{d_j}{rp} \times 31 $$

(5.1)

where $d_j$ is the buffer occupancy and $rp$ is the reaction parameter defined by $rp = 2 \times c\_rate/p\_rate = 2 \times MBF$, Eqn. 4.6. The integer value 31 sets $Q_j$ in the range from 1 to 31. This is the legal range of the quantisation scale value specified in the MPEG standards. $rp$ is used in calculating the actual quantisation scale. $c\_rate$ is expressed in bits/s and $p\_rate$ in frames/s. Eqn. 5.1 can also be written as:

$$ Q_j = \left( \frac{d_j}{2 \times MBF} \right) \times 31 $$

(5.2)

where $\frac{d_j}{2 \times MBF}$ is interpreted as the normalised buffer occupancy with the buffer size, $2 \times MBF$. Eqn. 5.2 is the key to controlling the video rate and is applied to all channel rates irrespective of the actual buffer size. In other words, the quantisation step size is derived from the occupancy based on the delay requirement rather than a specific buffer size. The occupancy is determined on a picture-type basis, that is, the occupancy of an I picture differs from those of B and P pictures since it is calculated for the same picture type only. The buffer size specification, $vbv\_buffer\_size$, defined in the TM5 is not involved in determining the quantisation step size. If the video contains large scene variation, the number of allocated target bits will differ from the number of the actual coded bits. Likewise, the activity changes in the macroblock may often result in large quantisation scale variations. This may cause large fluctuations in both the occupancy and the quantisation value. The number of bits allocated to each picture depends on the picture complexity measure which is given as:
complexity = \( Q_j \times S_{t,p,b} \) \hspace{1cm} (5.3)

where \( S_{t,p,b} \) is the number of coded bits of the previous I, P or B picture. If the input video contains a great deal of variation in visual content, the corresponding complexity measure will exhibit fluctuations as shown in Fig. 5.1.

![Complexity graph](image)

**Figure 5.1:** Rate control features in TM5: (a) Picture complexity measured as the product of the quantisation scale and the number of coded bits for the previous picture; (b) Normalised macro block activity and the resulting quantisation scale.

Fig. 5.1(a) exhibits a very fluctuating profile, Fig. 5.1(b) depicts enlarged plots of the quantisation scale, \( Q_s \), and its constituent terms. \( Q_j \) is initially calculated using the buffer occupancy as given in Eqn. 5.2 and it shows gradual change. However, since \( N_{-actj} \), which represents local activity in a macro block, has large variations, \( Q_s \) shows the magnified variation of \( N_{-actj} \) which is given as:

\[
N_{-actj} = \frac{V_{\text{MB}}}{AV_{PP}}
\]  \hspace{1cm} (5.4)

where \( V_{\text{MB}} \) and \( AV_{PP} \) are the average variance of the current macro block \( MB \) and the average variance of the previous picture, \( PP \), respectively. This rate control scheme is considered to
be appropriate either for video with stationary statistics or for variable bit rate applications. However, it does not perform adequately for realistic video [42] since it controls the video rate in accordance with the picture activity rather than the buffer occupancy.

5.4 System identification via linear prediction

Before applying a linear estimator to the rate control technique, it is necessary to investigate how closely the estimator predicts the video rate in a supervised training environment [8]. This is a process called system identification in which the linear relationship between the input and the output is assessed. Through this process, one can judge whether or not using a linear estimator is appropriate for the rate control purpose.

The main objective of using a linear approach [140] (linear combiner) is to make the most of the scene change information, which is considered to have a direct influence on the compressed video rate. The rationale for using a finite impulse response (FIR) filter instead of an infinite impulse response (IIR) filter is that the non-stationary statistics of video rate depends considerably on variations in the input (scene change) features rather than on its previous values. The variation in the input features results in corresponding changes in the video rate. In the IIR filter-based rate estimator, the output of the rate estimator is fed back to the estimator input together with the scene change features. When a dramatic change occurs in the input, the IIR estimator may not correctly reflect the input change into the output, due to the existence of the feedback input, which differs largely from the variation in the scene change features. Feedback of the video rate estimator output is considered not to be effective. Thus, the FIR approach is employed for feed-forward rate control.

For system identification purposes, a generic linear combiner structure with a single input channel [141] is used:

\[ \hat{y}(k) = \sum_{i=0}^{N-1} h_i x(k - i) \]  

(5.5)

where \( \hat{y}(k) \) is the predicted value of the actual output \( y(k) \). \( x(k - i) \) are the time lagged input sample values with \( N \) samples and the predictor coefficients of the linear combiner, \( h_i \).

The linear combiner expressed in Eqn. 5.5 is depicted as shown in Fig. 5.2. It outputs the value of \( \hat{y}(k) \) using selected estimator coefficients. The error \( e(k) \) represents the difference between \( y(k) \) and \( \hat{y}(k) \). The \( |e(k)|^2 \) is minimised in the least square sense.

The predicted video rate can now be written as:

\[ \tilde{c}_{bf}(k) = \sum_{i=0}^{N-1} h_{i1} var_{org}(k - i) + \sum_{i=0}^{N-1} h_{i2} var_{df}(k - i) + \sum_{i=0}^{N-1} h_{i3} ptype(k - i) \]  

(5.6)
where $h_{1i}$, $h_{2i}$, $h_{3i}$ and $ptype(k - i)$ represent the estimator coefficients and the picture type, respectively. The estimated number of coded bits per frame, $\hat{cbf}(k)$, is the output of the linear combiner. In this approach it is assumed that there is a linear relationship between the scene change features and the video rate. We can determine the performance of the linear model by looking at how effectively the linear technique predicts the video rate.

![Linear predictor](image)

**Figure 5.2:** A basic linear predictor.

### 5.4.1 Linear video rate estimator

Fig. 5.3 shows the configuration of the linear predictor coupled with the quantiser in the MPEG2 video encoder. The predictor is supplied with three frame-wise scene change features and outputs a single $\hat{cbf}(k)$ value in bits. When $\hat{cbf}(k)$ is added to the current occupancy, $O(k - 1)$, it estimates the new occupancy at frame $k$, $\hat{O}(k)$. Both current and future occupancies, $O(k - 1)$ and $\hat{O}(k)$ are used to determine $\hat{Q}(k)$.

![Quantiser combined with linear predictor](image)

**Figure 5.3:** Quantiser combined with the linear predictor.

The configuration of a VBR MPEG2 encoder combined with the estimator is shown in Fig. 5.4, and is also used for the purpose of system identification. Since the encoder operates in the VBR mode, the buffer is not necessary. Therefore, the feedback path from the buffer is not connected
to the quantiser. This implies that the configuration only takes into account the effect of the variation in the scene change on the estimated video rate, \( \hat{cbf}(k) \), since the quantisation scale is fixed. That is, the linear and nonlinear relationships are intended to be observed between scene change features and the predicted video rate without the influence from the buffering.

![Diagram of VBR MPEG2 encoder using the linear estimator.](image)

**Figure 5.4: VBR MPEG2 encoder using the linear estimator.**

Three types of linear estimators are considered on the criterion that a better estimator should show enhanced performance in terms of mean square error (MSE) between \( cbf(k) \) and \( \hat{cbf}(k) \). Their configurations are shown in Fig. 5.5 (Type I), Fig. 5.6 (Type II) and Fig. 5.7 (Type III).

![Diagram of a basic linear estimator (Type I).](image)

**Figure 5.5: A basic linear estimator (Type I).**

Type I is a typical linear predictor, Eqn. 5.6, with three input channels, without the time lagged inputs. On the other hand, both Type II and Type III have time lagged input samples in each input channel. For each input channel two delays are used since the MPEG2 video encoder is set to process two B pictures between two P pictures. The difference between the Type II and
the Type III is the way the coefficients, \( h_a \), are trained. In Type II, three filter taps from each input are used in training the coefficients (\( h_{11}, h_{22} \) and \( h_{33} \)) separately with respect to the error \( e(k) \). The estimated outputs \( \hat{cbf}_1(k) \), \( \hat{cbf}_2(k) \) and \( \hat{cbf}_3(k) \) are multiplied by \( 1/3 \), respectively, on the assumption that each output has an equal contribution to the final output, \( \hat{cbf}(k) \). On the other hand, in Type III, the nine inputs are treated as a single set of inputs as shown in Fig. 5.7. The estimated outputs of Type II and Type III are equivalent when the three input data are orthogonal to one another. Thus, except for the difference in the way of training coefficients, both types are basically equivalent.

![Diagram](image)

**Figure 5.6:** A linear predictor, Type II.

![Diagram](image)

**Figure 5.7:** A linear predictor, Type III.

The RLS algorithm, Fig. 5.8, is used to adapt the predictor coefficients. It updates the predictor coefficients, \( h_i(n) \), depending on changes in the input. The RLS algorithm is known...
to be appropriate for an input whose statistical properties vary [140] since it can change the coefficients during the operation. In the simulation the RLS-based estimator was tested which has the same input configuration as Type III.

Figure 5.8: Type III predictor with recursive least square-based weight calculation.

### 5.4.2 Adaptation of the predictor coefficients

For training the predictor coefficients, the LS algorithm was first used to obtain the coefficients \( h \) in the linear equation expressed in matrices, \( Xh = y \) where \( X \) is the input matrix and \( y \) is the predicted output as in Eqn. 5.5 using vector (lower case bold face) and matrix (upper case bold face) notations. The coefficient vector \( h \) is estimated by the singular value decomposition (SVD) technique [140, 142] using a finite number of input and output time series which are assumed to represent the overall linear property of the input. The SVD technique is known to be an effective tool for avoiding the problem of ill-conditioning - where the matrix \( X \) can often be a singular matrix or nearly singular in that its determinant is very close to zero [143]. This technique was applied to the predictor Types I, II and III.

In the RLS algorithm, Fig. 5.8 and Eqn. 5.7, the coefficients, \( h(k) \), are updated by constituent terms, \( K(k)e(k) \). The error signal, \( e(k) \), is the difference between the actual output and the predicted output in the previous step. The product term, \( K(k) \), works as a scaling term which varies depending on the input \( u(k) \). In short, the RLS algorithm updates the coefficients \( h(k) \) in accordance with the variation in the input, i.e. in a recursive manner:

\[
\begin{align*}
    h(k) &= h(k-1) + K(k)e(k) \\
    e(k) &= e(k) - hf'(k) - h^T(k-1)x(k) \\
    K(k) &= \frac{\Phi^T(k)}{\kappa(k)} \\
    \Phi(k) &= x^T(k)P(k-1) \\
    \kappa(k) &= \lambda + \Phi(k)x(k) \\
    P(k) &= \frac{1}{\lambda}(P(k-1) - P^T(k-1)) \\
    P^T(k-1) &= K(k)\Phi(k)
\end{align*}
\]

(5.7)
where $\lambda$ is the ‘forgetting factor’ determining the ‘memory’ of the algorithm. The soft-constrained initialisation [140] was applied to set the initial condition for the recursion as:

\[
P(0) = \delta^{-1}I \\
h(0) = 0
\]  

(5.8)

where $\delta$ is the small positive constant determined with respect to the variance of the input $x(k)$, which is responsible for the initial convergence speed, and $I$ is the identity matrix. The magnitude of $\delta$ does not have critical influence for a large data length [140]. This algorithm was applied to the predictor Type III.

### 5.4.3 Simulation results

In this section, the simulation results for the linear predictors are presented. Type I does not use time lagged inputs and the coefficients of Type II are trained separately. Thus, they are not considered in the performance evaluation, since they do not possess sufficiently good performance to compare with Type III and its RLS-adapted version. The objective of this evaluation is to examine the performance of linear predictors with given input and output data, rather than to evaluate the actual rate control performance in a running MPEG2 encoder. The input and output data are given by encoding the test video sequences at 1280 kbits/s and at variable bit rate. In the non-recursive technique using Type III, 200 input samples in the mid range are taken for training among 300 available data samples. The RLS forgetting factor $\alpha$ was chosen experimentally. A small value of $\alpha$, say, smaller than 0.7, generally results in large variation in predicted signals since the ‘memory’ of the algorithm reduces. Thus, a large value 0.95 was consistently used in order to maintain the predicted signal less noisy. Fig. 5.9 shows the three inputs used for the simulation, repeating and extending Fig. 4.5 on a logarithmic vertical scale since the dynamic range of each input is considerably different from the others. The relative performance is assessed by using the mean square error (MSE) in dB, which is given by:

\[
\text{MSE} = 10 \log_{10} \frac{\text{error power}}{\text{signal power}}
\]

(5.9)

where the signal power is the mean value of the squared sum of the actual output signal and the error power represents that of the error between the actual and the estimated signals. Figs. 5.10 and 5.11 show prediction results (i.e. the predicted signal and the actual output signal in bits per frame) for the VBR mode, which are obtained by the Type III and the RLS-adapted Type III, respectively. Figs. 5.12 and 5.13 show the results for the CBR mode at 1280 kbits/s. These results were obtained simulating the “Starwars” video sequence and summarised in Table 5.1.
Figure 5.9: Three inputs, $\text{var}_{\text{org}}(k)$, $\text{var}_{\text{dif}}(k)$ and $\text{ptype}(k)$: (a) Cascaded; (b) Starwars; (c) Adverts.
Figure 3.11: Recursively linear prediction for "Sharrows". (a) Coded bits/frame. (b) MSE.

Figure 3.10: Linear prediction Type III for "Sharrows". (a) Coded bits/frame. (b) MSE.
Figure 3.12: Recursive error prediction for "Stewarts" at 1280 Ki/s:
(a) Coded bits/frame;
(b) MSE.
The actual output signal in the CBR mode results (Figs. 5.12 and 5.13) does not show abrupt changes as appeared in the VBR mode results, and the resulting predicted signals show similar characteristics to the actual output. However, in the VBR mode the actual output signal contains a dramatic increase and a decrease around the frame numbers 140 and 180, respectively, maintaining its mean value higher than other signal values outside the frame number range. For this sudden change in the actual output signal, the RLS-adapted estimator appears to predict the actual output signal with more accuracy, as shown in Fig. 5.11(a) since it adaptively changes the predictor coefficients. The RLS-adapted estimator appears to have better performance than the Type III for VBR transmission with the MSE -23.40 dB versus -23.94 dB ("Starwars"). On the other hand, for the 1280 kbits/s case where the video rate is controlled by the buffer, the performance of the Type III is better than the RLS-based estimator in terms of the MSE, i.e. -24.25 dB versus -22.65 dB ("Starwars"). For "Adverts", the RLS-based estimator shows better MSE figures in both VBR and CBR modes. This may be interpreted as follows: the RLS-based estimator is better able to track the short-term, dramatic changes in the time series than the Type III while the Type III works well with a time series with less dramatic changes such, as is the case for 1280 kbits/s. Similar results are shown in Fig. B.9, Fig. B.10, Fig. B.11 and Fig. B.12 in Appendix B for "Adverts".

The variations in \( \text{var}_{-\text{org}}(k) \) and \( \text{var}_{-\text{dif}}(k) \), Fig. 5.9(b), appear similar to those in \( \text{cbf}(k) \) over the whole frame number range for the VBR simulations as shown in Fig. 5.10, Fig. 5.11. On the other hand, in the CBR simulations, Fig. 5.12 and Fig. 5.13, the profile of \( \text{cbf}(k) \) shows flatter overall variations compared to the VBR since the video rate is controlled so that the buffer occupancy variation can be reduced. However, this outcome from the CBR mode simulations is the result of the rate control rather than a cause of degradation in the prediction performance. Therefore, these prediction results do not imply that the inputs are necessarily suitable for the VBR mode. On the contrary, the variance inputs, \( \text{var}_{-\text{org}}(k) \) and \( \text{var}_{-\text{dif}}(k) \), may be more suitable for the CBR mode as they are employed to improve the rate control performance by being used for estimating video rate within the framework of buffer control.

<table>
<thead>
<tr>
<th>MSE [dB]</th>
<th>VBR</th>
<th>CBR 1280 kbits/s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Linear</td>
<td>RLS</td>
</tr>
<tr>
<td>Starwars</td>
<td>-23.40</td>
<td>-23.94</td>
</tr>
<tr>
<td>Adverts</td>
<td>-22.47</td>
<td>-23.07</td>
</tr>
</tbody>
</table>

Table 5.1: Comparison of MSE values in dB between the Type III and the RLS-adapted estimator for "Starwars" and "Adverts" sequences.
5.5 Configuration of the MPEG2 encoder based on a rate estimator

In this section the configuration of the improved MPEG2 encoder (Fig. 5.14) is described as a development of Fig. 3.7.

![Diagram](image)

Figure 5.14: The structure of the nonlinear predictive rate control for MPEG2: (a) a functional representation; (b) the signal flow.

The improved MPEG2 encoder composes of three main functions, Fig. 5.14(a): the scene change...
calculator (SCC), the video rate estimator (VRE) adapted with the recursive least square algorithm and the nonlinear quantisation scale control (NQC). The RLS-based adaptation updates predictor coefficients at every frame start with the updated error signal, \( e(k) \). In view of the non-stationary nature of the inputs, the RLS algorithm is consistently used, since the coefficients should be updated according to input changes in the MPEG video encoder. The scene change calculator outputs the two variances, \( \text{var}_\text{org}(k) \) and \( \text{var}_\text{dif}(k) \), as explained in the previous section and the picture type information, \( \text{ptype}(k) \), as the inputs for the rate estimator. The predicted video rate, \( \hat{cf}(k) \), is added to the current buffer occupancy value, \( O(k-1, n) \), and subtracted from the number of mean allocated bits per picture, MBF, to estimate the predicted occupancy, \( \hat{O}(k) \). The nonlinear quantiser control finally outputs the quantisation scale value, \( Q_s(k, n) \). The quantiser and the MPEG video encoder process the video data to output a coded bit stream with the video rate, \( cf(k) \). Finally, the bit stream goes to the buffer for the transmission through the channel. Fig. 5.14(b) depicts the configuration in terms of feed-forward control in order to provide an insight into the overall structure. Here, the MPEG 2 video encoder is viewed as a finite impulse response filter (H) which accepts \( Q_s(k) \) as an input, and outputs \( cf(k) \) and \( O(k) \). The transmission buffer is treated as a delay \( (z^{-1}) \) since it stores coded bits for a specific frame period. The a priori information from the pre-processing side of the video encoder (\( \text{var}_\text{org}(k), \text{var}_\text{dif}(k) \) and \( \text{ptype}(k) \)) is fed to the video rate estimator which adaptively changes its coefficients. The predicted video rate, \( \hat{cf}(k) \), and the current occupancy finally supply the inputs, \( \hat{O}(k) \) and \( O(k-1) \), to the nonlinear quantiser control which provides the quantiser with the scale, \( Q_s(k, n) \). The video rate estimator coefficients are updated at every frame start in a recursive least square sense with the error signal, \( e(k) \).

### 5.5.1 Selection and calculation of the scene change features

The scene change calculator operates in advance of actual encoding in order to estimate the \( cf(k) \) signal. The motion vector function, \( MVF_D(s) \), defined in Chapter 4, was not attractive as an input, as it does not possess a strong linear relationship with the actual output \( cf(k) \). On the other hand, the picture type appeared to have a significant effect on the performance in terms of the MSE. If the input video is encoded in the order of the picture type, the variation in the video rate is subsequently proportional to the picture type. Under this premise, an I picture will give rise to the largest amount of video data, and a P picture generates more coded bits than a B picture. Thus, the resulting video rate exhibits a periodic time series as the picture types repeat (See Fig. 4.1). The bit rate ratio for I, P, and B pictures varies in accordance with the amount of the visual information captured in the video. Therefore, a nominal value of the ratio is significant only under the condition of video with small motion. In addition, it depends on a specific bit allocation technique. For example, the MPEG 1 standard recommends a bit allocation ratio depending on the picture type. Let \( b(B), b(P) \), and \( b(I) \) be the numbers of bits generated by each type of picture, respectively. The ratio is given by [26]:

\[
b(P) = (2 \sim 5) \times b(B)
\]
\[ b(I) = 3 \times b(P) \] (5.10)

where \( 2 \sim 5 \) represents any value in the range. In this chapter, the ratio, \( b(I):b(P):b(B) \) is set to 6:2:1. That is, for the I, P and B picture types the integers 6, 2 and 1 are allocated, respectively. Thus, any \( p_{type}(k) \) value has one of these integers for the corresponding picture type, and it forms a periodic time series as \( k \) increases as shown in Fig. 5.9. This \( p_{type}(k) \) specification is based on the ratio of quantisation scale values for VBR channels shown in Fig. 4.2 and Fig. 4.3 in Chapter 4.

Several other scene change features were also considered such as \( MVF_D(s) \) and \( IIF(k) \), as shown in Fig. 4.6 and Fig. B.4 in Appendix B. \( MVF_D(s) \) is the summation of all motion vector values in a slice, as defined in Eqn. 4.4; \( IIF(k) \) is the number of intra-coded macro blocks in a P or a B picture. \( MVF_D(s) \) was tested instead of \( p_{type}(k) \) as one of the rate estimator inputs in the initial simulation. However, it appeared to be less effective and it complicated the encoder. While \( var_{-org}(k) \) and \( var_{-dif}(k) \) come from the scene change calculator, \( MVF_D(s) \) comes from inside the encoder, thus, it makes the configuration of the scene change calculator more complex. In addition, it requires more computation than \( p_{type}(k) \) since \( MVF_D(s) \) can only be calculated after encoding a slice, it represents the value for the previous slice, which is less effective on the current slice. Thus, for an improved prediction performance, \( p_{type}(k) \) is adopted along with \( var_{-org}(k) \) and \( var_{-dif}(k) \) for the rest of this thesis.

### 5.5.2 Video rate estimators

The video rate estimator, Fig. 5.14, outputs the predicted rate value, \( \hat{ebf}(k) \), which will be compared to the actual rate value, \( ebf(k) \). \( ebf(k) \) is subtracted from \( \hat{ebf}(k) \) after completing the encoding of the \( k \)th frame. The performance depends entirely on the relevance of the scene change features to the video rate. If the scene change features have a certain linear or nonlinear relationship with the video rate, the latter will be effectively estimated by the predictor with adequately selected coefficients. The parameters of the linear combiner are set based on the configuration shown in Fig. 5.8. The forgetting factor, \( \lambda \), is set to .95. The initial values of the weights were set to zero. This initialisation complies with Eqn. 5.8. The predictor takes a certain length of time for the coefficients to converge after the initialisation, depending on the number of weights. This is a negligibly short time for the video encoder to handle the initial transient state before the actual communication starts.

The RBF-network predictor has the same input configuration as the linear combiner. However, it has a different predictor structure which is a combination of a nonlinear function layer and a linear predictor layer. It includes the linear combiner as its linear part and the radial basis function as its nonlinear part. This structure is known, in general, to have better prediction performance [8] in comparison to conventional linear predictors, for non-stationary signals.
A heuristic rate estimator, LOGEXP-A in Chapter 4, is also examined in order to compare its performance with the linear predictors. The heuristic rate predictor performs an estimation using scene change features and the video rate balance factor [129]. The rate balance is determined by calculating the ratio of the number of bits for coded frames to the available number of bits, as expressed in Eqn. 4.9. It is used to select a quantiser control curve from the nonlinear surface. This scheme can also be combined with the sigmoidal surface, however, the unimodal surface was used since it appeared to show better performance than the sigmoidal surface.

5.5.3 Nonlinear quantisation control

The nonlinear quantiser control block is shown in Fig. 5.14. Rate control regulates the occupancy by selecting an appropriate quantisation step size. The quantisation step size - determined by the quantisation scale value - is the dominant parameter used to achieve rate control. The nonlinear quantiser control technique makes use of both current and future buffer occupancies rather than only using the current occupancy to select the curve for the quantisation scale for a picture, as shown in Fig. 5.15.

![Figure 5.15: Quantiser control based on the nonlinear function surface.](image)

Let \( O(k-1, n) \) and \( \hat{O}(k) \) be current occupancy and predicted occupancy, respectively, as defined earlier. \( n \) is the macro block index which ranges from 0 to 329 for the 352 pixels \( \times \) 240 lines SIF image format. The occupancy and the quantisation scale value vary on a macro-block-by-macro-block basis. The quantisation scale, \( Q_s(k, n) \), is calculated with the following equations:

\[
Q_s(k, n) = f(O(k-1, n), \hat{O}(k)) \\
\hat{O}(k) = O(k-1, n) + \hat{c}bf(k) - \text{MBF}
\]  

(5.11)

The quantisation scale \( Q_s(k, n) \) is determined by both \( O(k-1, n) \) and \( \hat{O}(k) \). The function \( f() \) is a surface derived from a nonlinear equation to adaptively map the two occupancies to the quantisation scale.
The nonlinear quantisation control surface adaptively changes its steepness depending on the predicted occupancy. In this chapter two different control surfaces (sigmoidal and unimodal) will be examined, which are 3-dimensional extensions of those described in Chapter 4, with different dynamic properties as shown in Fig. 5.16. \( \tilde{O}(k, n) \) selects a curve from the surface for a picture and \( O(k - 1, n) \) indicates the current occupancy to be mapped to the quantisation scale value for the macro block. If a dramatic change in the occupancy is predicted, then it makes the control curve more nonlinear, otherwise, it selects a more linear curve. The sigmoidal surface is formed by changing the curvature of a sigmoidal function. The unimodal surface consists of a combination of an exponential part and a logarithmic part.

The 2-dimensional sigmoidal nonlinearity function \([8]\) is given \([128]\):

\[
y = \frac{1}{1 + \exp(-ax)}
\]

where \( a \) determines its steepness. The 3-dimensional form can be given as:

\[
z = \frac{1}{1 + \exp(-axy)}
\]

where \( y \) determines curvature of the surface depending on its value. However, the 3-D shape of Eqn. 5.13 is not perfectly fitted to the occupancy space shown in Fig. 5.16. Thus, a modified sigmoidal function is used instead (See Eqn. 4.11 for the 2-D function), which directly maps the occupancies to the quantisation scale as follows:

\[
\text{SIGM: } f(O(k - 1, n), \tilde{O}(k)) = \frac{1}{\alpha}O(k - 1, n)\left(1 + \alpha - O(k - 1, n)\right) \times \text{trunc}(1 + \alpha - O(k - 1, n)) \\
+ \left(1 - (1 - \alpha)\left(1 - O(k - 1, n)\right)\right) \times \text{trunc}\left(\frac{O(k - 1, n)}{\alpha}\right)
\]

A direct conversion of the unimodal surface can be expressed (See Eqn. 4.12 for the 2-D function):

\[
z = \begin{cases} 
\alpha \log_{10}(\beta xy + 1) & \text{if } x > 0.5 \\
\alpha \exp^{\beta xy} - \beta & \text{if } x < 0.5
\end{cases}
\]

where \( \alpha \) and \( \beta \) are constants to change the shape of the function. The variable \( x \) and \( y \) correspond to the two occupancies. However, the actual quantisation scale value is determined by the following equation which maps both occupancies to the quantisation scale in a computationally efficient way:

\[
\text{UNIM: } f(O(k - 1, n), \tilde{O}(k)) = O(k - 1, n)^{C/(S\tilde{O}(k) + 1)}
\]

80
In Eqs. 5.14 and 5.16, $S$ is a steepness factor which represents the extent of nonlinearity, and 
**trunc** stands for truncation function to truncate an input to 1 or 0 depending on its value. 
The torsion factor $S_{max}$, which is the maximum value of $S$, varies with channel rates. There 
exists an inversely proportional relationship between the torsion factor $S_{max}$ and the channel 
rate, expressed as follows:

$$S_{max} = \frac{A}{e \cdot \text{rate}} \quad (5.17)$$

where $A$ is a constant which determines the extent of torsion. When the channel rate is high, 
e.g. 5 Mbits/s or higher for the SIF video format, the expanded channel capacity can handle 
the video rate fluctuation, hence, a small $S_{max}$ can be used. For a lower channel rate a higher 
value is assigned to $S$ to give the surface a larger torsion. Figs. 5.16(b) and (d) correspond to 
the case of large $S_{max}$ value and Figs. 5.16(a) and (c) show the case of small $S_{max}$ value.

![Graphs showing nonlinear quantiser control surfaces](image)

Figure 5.16: Nonlinear quantiser control surfaces: (a) Sigmoidal ($S_{max} = 5$); (b) Sigmoidal 
($S_{max} = 13$); (c) Unimodal ($S_{max} = 5$); (d) Unimodal ($S_{max} = 13$).

The constants $\alpha$ and $C$ are balancing factors to make the surfaces balanced or unbalanced. If 
$\alpha$ is 0.5, the lower part and the upper part of the sigmoidal surface form a symmetrical sigmoid
function, Fig. 5.16(a) and (b). If \( C = S_{\text{max}} / 2 \), both exponential and logarithmic parts remain balanced in shape as shown in Fig. 5.16(c) and (d). For the unimodal nonlinear control curves in both logarithmic and exponential sides appear in the ranges close to the two extreme values of \( \hat{O}(k) \), i.e. 0 and 1.

These nonlinear quantisation functions are capable of allocating an adequate quantisation scale value depending on both current and previous occupancies. This feature is considered to be effective particularly for dramatic video rate variations. A detailed rate-distortion analysis of these two control functions is presented in the following section.

5.6 Rate-Distortion analysis of the quantiser control functions: sigmoidal and unimodal

In this section the performance of the two quantisation control functions is analysed (sigmoidal and unimodal) in a rate-distortion theoretic way. The content explained here is presented as a background analysis in conjunction with Sections 3.3 and 4.4.

5.6.1 Background rate-distortion theory

In the context of Section 3.3.2, the \( N \)-block entropy for a vector \( \mathbf{x} \), \( H_N(\mathbf{x}) \), is given by [31]:

\[
H(\mathbf{x}) = \lim_{n \to \infty} H_N(\mathbf{x})
\]

\[
H_N(\mathbf{x}) = -\frac{1}{N} \sum_{i=1}^{N} \cdots \sum_{a \in \mathbf{x}} p(\mathbf{x}) \log_2 p(\mathbf{x})
\]  

(5.18)

where \( p(\mathbf{x}) \) is the probability of the source vector \( \mathbf{x} \). It is known that a Markov source with a finite memory has the property that the limiting value of \( H_N(\mathbf{x}) \) is reached for a finite \( N \) [31], i.e. \( H(\mathbf{x}) = H_N(\mathbf{x}) \). This implies, in practice, that the entropy of the vector \( \mathbf{x} \) with a large number of samples can represent that of vectors with an arbitrary number \( N \). In general, the following inequality describes the nature of the required information to encode the vector \( \mathbf{x} \) [31]:

\[
H(\mathbf{x})_{\text{source with memory}} < H(\mathbf{x})_{\text{source without memory}} \leq \log_2 K
\]  

(5.19)

where \( K \) is the number of symbols per sample, i.e. the number of grey levels per pixel. Less information is thus required to transmit the source with memory than without memory.

As addressed in Section 3.3.2, the distribution of DCT data may differ considerably from Gaussian pdf particularly for realistic video. However, the following rate-distortion function [31] is used for the DCT data specified in the MPEG standards under the assumption that their pdf is Gaussian.

\[
D(R)_G = \gamma^2 \frac{1}{2} - 2R \gamma^2
\]  

(5.20)

82
where $R$ is the available bit rate. $D(R)_{IC}$ is the rate-distortion function for a memoryless zero-mean Gaussian source $N(0, \sigma^2_z)$ with the variance $\sigma^2_z$. $\gamma^2_z$ is the spectral flatness measure which represents the correlation in the source data. For simplicity, it is set to 1, i.e. implying a memoryless source since it is assumed to be constant. In the following sections, two performance criteria will be examined; the rate-distortion functions for the sigmoidal and the unimodal functions and the bit rate fluctuation property of these two functions will be examined.

### 5.6.2 Rate-distortion functions

**Proposition 1: the buffer occupancy**

Let $r$ be the buffer occupancy, with a range of real values 0 to 1, be represented as a function, $r = f(B, R_T, \sigma^2_z)$. Here, $B$ and $R_T$, respectively, represent the buffer size in bits and a given data rate in bits per pixel which are set to constant values in most CBR applications. Therefore, it can be simplified to $r = f(\sigma^2_z)$, which means that the buffer occupancy $r$ is the function of a sole independent variable, the source variance, $\sigma^2_z$, of the DCT data. The short-term source variance, $\sigma^2_z_{\Delta t}$, is introduced which represents the variance during the short period of time, $\Delta t$, in relation to the long-term variance, $\sigma^2_z$. For a specific $R_T$ value, Eqn. 5.20 reduces to $D(R)_{IC} = 2^{-2R_T}\sigma^2_z = H\sigma^2_z$ where $H$ is a constant. This signifies that the distortion is given by the only variable $\sigma^2_z$, i.e. $D = g(\sigma^2_z)$ where $g(\ )$ is an arbitrary function. Let $r(\Delta t)$ be the short-term occupancy. Then, the occupancy, $r$, can be described with respect to the variance $\sigma^2_z$, if $\sigma^2_z_{\Delta t} > \sigma^2_z$ then $r(\Delta t)$ increases, if $\sigma^2_z_{\Delta t} < \sigma^2_z$ then it decreases, otherwise, it remains unchanged. Thus, the occupancy equation can be given as follows:

$$r = r_0 \left( \frac{\sigma^2_z_{\Delta t}}{\sigma^2_z} \right)^b \quad (5.21)$$

where $r_0$ is the initial occupancy at $t = t_0$. The index $b$ represents the behaviour of the buffer occupancy, which is determined by the buffer size. If the buffer size is small the $b$ value should be large enough to correctly describe the variation of $r$. Note that the initial occupancy $r_0$ is multiplied by the variance ratio rather than added to it. Since $\sigma^2_z_{\Delta t} = k\sigma^2_z$, the real value $k$ should be positive. If $k$ is close to 1, Eqn. 5.21 results in $r = r_0$ for an arbitrary value of $b$. Thus, if $r_0 = 0.5$, i.e. 50% full, then $r$ remains at that state. However, if $k > 1$ or $k < 1$, $r$ either increases or decreases. The value of $b$ should thus be set to be larger than or equal to 1. If it is smaller than 1 (say 0.5), the occupancy, $r$, may saturate towards a specific value lower than 100%, even for a very large magnitude of $\sigma^2_z_{\Delta t}$. However, this cannot be accommodated in practice, since $r$ eventually reaches the buffer full state as $\sigma^2_z_{\Delta t}$ increases.

**Proposition 2: the relationship between the bit rate and the occupancy**

Let $R$ and $q_\ast$ be the available bit rate depending on the occupancy $r$ and the quantisation step size for the picture $\ast$, respectively. For arbitrary functions, $f(\ )$ and $g(\ )$, the relationship
between the bit rate and the quantisation step size can be given as follows:

\[ R = f(q_s) = f(g(r)) \]

\[ = R_T \frac{1}{q_s^2}, \quad (\alpha \geq 1) \]  

(5.22)

where \( \alpha \) is a positive integer to determine the nonlinearity of Eqn. 5.22. Note that \( R \) is different from \( R_T \); \( R_T \) is a constant channel rate, however, \( R \) is the available bit rate determined by the quantiser step size, Eqn. 5.22. If Eqn. 5.22 has a quadratic term, it will be given as follows [109]:

\[ R = R_T \left( \frac{u}{q_s^2} + \frac{v}{q_s^2} \right) \]  

(5.23)

where \( u \) and \( v \) are the parameters which must be optimised. Furthermore, higher order terms, e.g. \( q_s^3 \), can also be associated with Eqn. 5.23. However, Eqn. 5.22 is dealt with rather than Eqn. 5.23 or higher order terms since \( \alpha \) is assumed to vary depending on \( \sigma_{\Delta}^2 \). Bigger \( \alpha \) values introduce more nonlinear relationship between \( R \) and \( q_s \). However, since the resulting rate-distortion function is given by the similar variable \( r \), \( \alpha \) is set to 1 for simplicity. The function, \( g(r) \), specifies the relationship between the occupancy and the quantiser step size. This relationship can be either linear or nonlinear. In what follows, two nonlinear functions: sigmoidal and unimodal will be discussed.

5.6.3 Analysis of \( D_S(R) \) and \( D_U(R) \)

From Eqn. 5.22 the quantiser step size, \( q_s \), is replaced with \( q_s \) or \( q_u \) according to the definitions of sigmoidal [128] and unimodal [129] functions as follows:

\[ q_s = \frac{1}{1 + \exp(-\beta(r_s - 0.5))} \]  

(5.24)

\[ q_u = r_u^{1/\beta} \]  

(5.25)

where \( r_s \) and \( r_u \) are corresponding occupancy variables to the sigmoidal and unimodal functions, respectively. \( \beta \) determines the curvature of the control functions. Their graphical representations are shown in Fig. 4.9 and 4.10, respectively. The curvature parameter, \( \beta \), is substituted for a transmission parameter of rate balance, \( R_b \), which represents the ratio of the number of bits transmitted for a short period of time to the allocated number of bits for the pictures according to the channel rate (Refer to Section 4.4).

\[ R_b = \frac{\text{coded}\_bits[\Delta][\text{bits}] \times \text{picture}\_rate[\text{frame/s}]}{\text{channel}\_rate[\text{bit/s}] \times \text{number}\_of\_pictures[\Delta][\text{frames}]} \]  

(5.26)

A constant scaling factor \( c \) is multiplied by \( R_b \), Eqn. 5.24, i.e. \( cR_b \) is used instead of \( R_b \), in order to make the shape of the sigmoidal function appropriate for quantiser control. Here, \( c \) is set to 10 so that the curvature of the sigmoidal functions appears to be a skewed “S” shape. When \( R_b > 1 \), more bits were encoded than the channel rate. On the other hand, when \( R_b < 1 \),
less bits were coded for the period. Thus, it provides crucial information about the previous video rate so that the quantiser control function may take subsequent control action on the quantiser.

In Fig. 4.9 and Fig. 4.10 the curves are shown for representative $R_b$ values 1 to 4. The $R_b$ value must be greater than 0 so that buffer underflow can be prevented. The upper bound of $R_b$ depends on the buffer size. For delay critical applications it needs to be smaller than 2 since the buffer size is generally set to twice as large as the number of bits allocated per picture. In this case, if $R_b$ is bigger than 2, buffer overflow may occur. In Fig. 4.9 one can also consider an inverse sigmoidal function which is the inverse of the sigmoidal for the same $R_b$ value. There can be other inverse sigmoidal curves for different $R_b$ values. However, these curves are considered as inappropriate for quantiser control due to their resulting quantiser step size, $q_*$, for the occupancy change. For example, despite the occupancy, $r_\sigma$, increases from 0.2 to 0.8, the $q_*$ value remains in the range between 0.4 to 0.6. This type of inverse control, in contrast to the sigmoidal function, may exacerbate the occupancy fluctuation. Thus, the inverse sigmoidal functions are not used. Note that they do not represent the sigmoidal curves with $R_b$ values smaller than 1 due to this reason. When the $R_b$ value approaches zero the shape of the curve becomes closer to the straight diagonal. On the other hand, in the unimodal function, the curvature changes in reference to the $R_b = 1$ as shown in Fig. 4.10. When $R_b$ increases with larger values than 1, the curvature resembles logarithmic function. When $R_b$ decreases (smaller than 1), the curvature gets close to exponential function.

From Eqns. 5.20, 5.22, 5.24 and 5.25, the rate-distortion functions, $D_\sigma (R)$ and $D_u (R)$, for the sigmoidal and unimodal functions, are given in terms of $r_{\sigma, u}$ and $R_b$ as follows:

$$D_\sigma (r_\sigma, R_b) = 2^{-2R_b[1+\exp(-cR_b(r_\sigma - 0.5))]} \sigma_\sigma^2 \Delta t$$

$$D_u (r_u, R_b) = 2^{-2} \left( \frac{r_u}{\sigma_u} \right)^2 \sigma_u^2 \Delta t$$

where $r_\sigma$ and $r_u$ represent the occupancies determined by the sigmoidal and unimodal control functions, respectively. From Eqn. 5.22, $R = \frac{R_b \sigma}{4}$, and by replacing $\beta$ with $cR_b$ in Eqns. 5.24 and 5.25,

$$q_\sigma = \frac{1}{1 + \exp(-cR_b(r_\sigma - 0.5))}$$

$$q_u = \frac{1}{r_u^{1/R_b}}$$

(5.29)

(5.30)

Using Eqns. 5.20, $D_\sigma (R)$ and $D_u (R)$ are given in $q_\sigma$ and $q_u$, respectively, i.e.:

$$D_\sigma (q_\sigma) = 2^{-2[R_b \sigma]} \sigma_\sigma^2$$

$$D_u (q_u) = 2^{-2[R_b \sigma]} \sigma_u^2$$

(5.31)

(5.32)

Substituting $q_\sigma$ and $q_u$ in Eqns. 5.31 and 5.32 for those in Eqns. 5.29 and 5.30 yields $D_\sigma (r_\sigma, R_b)$ and $D_u (r_u, R_b)$, i.e. Eqns. 5.27 and 5.28.
Since this derivation procedure aims to represent the rate-distortion functions in terms of the occupancy and the rate balance, the short-term variance, \( \sigma^2_{\Delta t} \), in Eqns. 5.27 and 5.28 needs to be substituted for these variables. In Eqn. 5.21 the long-term variance, \( \sigma^2_{\Delta t} \), and the initial occupancy, \( r_0 \), are set to 1 for simplicity and combining Eqns. 5.27 and 5.28 eventually achieves the following rate-distortion functions expressed in terms of \( r_s, u \) and \( R_b \):

\[
D_s(r_s, R_b) = 2^{-2R_s[1+\exp(-\epsilon r_s r_s][1])^{1/\delta}}
\]

\[
D_b(r_s, R_b) = 2^{-2\left(\frac{R_b}{r_s R_b}\right)} r^{1/\delta}
\]

where \( R_T \) is a constant determined by the channel rate and \( b \) is also given a constant value. For example, the \( b \) value can be set to 2 under the assumption that the occupancy, \( r_s, u \), is proportional to the variance ratio in Eqn. 5.21 in a quadratic sense. This type of rate-distortion function is derived to serve as a yardstick to assess the curvature the quantiser control function on the resulting distortion. Fig. 5.17 shows in 3 dimensions the resulting rate-distortion functions depending on \( r \) and \( R_b \).

![Diagrams showing rate-distortion functions](image)

(a) 0.202 bpp (front)  
(b) 0.202 bpp (rear)  
(c) 0.539 bpp (front)  
(d) 0.539 bpp (rear)

Figure 5.17: 3-dimensional representation of \( D_s(r) \) and \( D_b(r) \)
Chapter 5: Video rate control based on nonlinear rate estimation

The z-axis represents the relative distortion with respect to $\sigma_z^2$. Fig. 5.17(a) and (b) refer to the transmission with a low bit rate, 768 kbit/s, i.e. 0.202 bits per pixel (bpp). (c) and (d) depict the case of a high bit rate transmission, 2048 kbit/s, i.e. 0.539 bpp. The area of interest in these plots is the region with lower $R_b$, say, smaller than 2 and the occupancy, $r$, less than 80%. Fig. 5.17(a) and (c) (front view), since an effective rate control algorithm should be able to maintain steady occupancy with the resulting $R_b$ close to 1. In this area $D_u(r_u, R_b)$ appears to lie below $D_u(r, R_b)$. This suggests that the unimodal function can provide better quality than the sigmoidal function, i.e. achieve reduced distortion. On the other hand, if $R_b$ is higher than 2, the sigmoidal function shows a better distortion profile, Fig. 5.17(b) and (d) (rear view), where the occupancy is lower than 50%. However, when the rate balance reaches such large values, the occupancy may not remain at such low values. In addition, buffer may overflow. Thus, this case is not considered as normal coding operation. The results appear identical for two different transmission rates, 768 and 2048 kbit/s. When the $R_b$ approaches zero, buffer underrflow may occur. However, this is generally avoided by inserting justification or dummy bits into the bit stream, so it is not the focus of attention, here.

### 5.6.4 Controllability over the occupancy fluctuation

Fluctuation in the buffer occupancy is the second important performance criterion, since it has direct influence on the performance of rate control. The occupancy fluctuation can be defined as instantaneous variation of the occupancy, i.e. the second derivative of its transmission parameters; the rate balance and quantiser step size. The first derivative of the occupancy is the average occupancy for a relatively large interval of the parameter. Since the quantiser step size was expressed as a function of $r_u$ and $R_b$, Equs. 5.24 and 5.25, the fluctuation, $fL_u$ for the sigmoidal and $fL_u$ for the unimodal, becomes the second partial derivatives as follows:

$$\frac{\partial^2 r_s}{\partial R_b^2} = fL_s(R_b), \quad \frac{\partial^2 r_u}{\partial R_b^2} = fL_u(R_b)$$

$$\frac{\partial^2 r_s}{\partial q_u^2} = fL_s(q_u), \quad \frac{\partial^2 r_u}{\partial q_u^2} = fL_u(q_u) \quad (5.35)$$

Although the quantiser step size and the rate balance are coupled with each other, they are treated as separate variables for analysis. To derive Eqn. 5.35, Equs. 5.24 and 5.25 for the occupancy should be re-written as:

$$r_s = \frac{1}{cR_b} \ln \left( \frac{1}{q_s} - 1 \right) + 0.5$$

$$r_u = \frac{R_b}{q_u} \quad (5.36)$$

Thus their second derivatives represent the occupancy fluctuation:

$$fL_s(R_b) = -c \ln \left( \frac{1}{q_s} - 1 \right) 2R_b^{-3} \quad (5.37)$$

$$fL_u(R_b) = (\ln q_u)^2 \exp (R_b \ln q_u) \quad (5.38)$$

87
\[ f_{l_s}(q_s) = \frac{1}{c R_b} \left(2q_s - 1\right) \]  \[ f_{l_u}(q_u) = R_b \left(R_b - 1\right)q_u^2 \] (5.39) (5.40)

These fluctuation equations can also be depicted with 3-dimensional plots, as shown in Fig. 5.18. The z-axis represents the relative fluctuation in terms either of the rate balance or of the quantiser step size. The \( R_b \) value ranges from 0.5 to 4. The z-axis values are theoretical ones which are useful to assess the performance of these functions.

For the \( R_b \) fluctuations, \( f_{l_s}(R_b) \) and \( f_{l_u}(R_b) \), the unimodal function exhibits much smaller variation than the sigmoidal functions for the same ranges of the rate balance and the quantisation scale. Note that the z-axis ranges of \( f_{l_s}(R_b) \) and \( f_{l_s}(q_s) \) (sigmoidal) are much wider than those of the unimodal. For the quantisation scales, \( q_s \) and \( q_u \), \( f_{l_u}(q_u) \) appears to have a steady and smaller fluctuation profile in comparison to \( f_{l_s}(q_s) \). Thus, this leads to a judgement that the unimodal control function should offer superior control of the video rate fluctuation.
5.6.5 Simulations results

For verification, the quantisation parameter equation of the MPEG2 TM5 was replaced by the two quantisation control equations, i.e. Eqns. 5.24 and 5.25. The rate balance, \( R_b \), is determined by:

\[
R_b = \frac{c_{\text{bits}_n}}{t_{\text{bits}_n}}
\]  

(5.41)

where \( c_{\text{bits}_n} \) and \( t_{\text{bits}_n} \) represent the numbers of coded and target bits allocated for the \( n \)th picture, respectively. This is used for Eqns. 5.24 and 5.25. The value of \( R_b \) was clipped at 2 to see the effect for the low range. This also prevents the quantisation scale from drastic change which may cause malfunction to the bit allocation process of the TM5. Three different channel rates are tested: 768, 1024 and 1536 kbits/s. Three video sequences are used: “Adverts”, “Starwars” and “JKF”.

Table 5.2 summarises the performance, showing mean and standard deviation (std. dev.) for the measures: the PSNR and the occupancy for the luminance signal. TM5 represents the MPEG2 TM5 evaluation model, and Sigm. and Unim. represent the modified TM5 versions based on the sigmoidal and the unimodal control surfaces (Fig. 5.16(b) and (d)), respectively. The peak occupancy greater than 100% (with tick marks) means that buffer overflow has occurred. Unim. exhibits slightly higher PSNR than the other two methods but its occupancy is much lower. Note that TM5 shows the worst performance of controllability over occupancy fluctuation. Unim. outperforms the Sigm. for all the channel rates by showing far smaller fluctuation in the occupancy with the same video quality. This agrees to the analysis that Unim. has better controllability than Sigm. over the video rate fluctuation. Fig. 5.19 shows the profile of the buffer occupancy for the three methods at the 1024 kbits/s rate. Each plot shows the critical part of the video sequence where dramatic scene change occurs. For all the three video sequences, Unim. exhibits quicker response to the drastic fluctuation in the video rate and hence the unimodal surface was selected for use in the subsequent nonlinear predictor simulations.
<table>
<thead>
<tr>
<th>Video</th>
<th>Channel rate</th>
<th>Rate control</th>
<th>PSNR (Y) [dB]</th>
<th>Occupancy [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>mean</td>
<td>std. dev.</td>
<td>mean (peak)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>768</td>
<td>TM5</td>
<td>31.55</td>
<td>4.51</td>
<td>97 (309) ✔</td>
</tr>
<tr>
<td></td>
<td>Sigm.</td>
<td>31.37</td>
<td>4.46</td>
<td>174 (559) ✔</td>
</tr>
<tr>
<td></td>
<td>Unim.</td>
<td>31.55</td>
<td>4.45</td>
<td>66 (190) ✔</td>
</tr>
<tr>
<td>1024</td>
<td>TM5</td>
<td>32.96</td>
<td>4.49</td>
<td>53 (119) ✔</td>
</tr>
<tr>
<td></td>
<td>Sigm.</td>
<td>32.94</td>
<td>4.48</td>
<td>52 (92)</td>
</tr>
<tr>
<td></td>
<td>Unim.</td>
<td>32.94</td>
<td>4.49</td>
<td>37 (68)</td>
</tr>
<tr>
<td>1536</td>
<td>TM5</td>
<td>35.04</td>
<td>4.36</td>
<td>33 (65)</td>
</tr>
<tr>
<td></td>
<td>Sigm.</td>
<td>34.99</td>
<td>4.37</td>
<td>42 (58)</td>
</tr>
<tr>
<td></td>
<td>Unim.</td>
<td>35.02</td>
<td>4.37</td>
<td>37 (56)</td>
</tr>
<tr>
<td>Starwars</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>768</td>
<td>TM5</td>
<td>31.67</td>
<td>2.62</td>
<td>59 (119) ✔</td>
</tr>
<tr>
<td></td>
<td>Sigm.</td>
<td>31.64</td>
<td>2.60</td>
<td>54 (96)</td>
</tr>
<tr>
<td></td>
<td>Unim.</td>
<td>31.65</td>
<td>2.61</td>
<td>40 (65)</td>
</tr>
<tr>
<td>1024</td>
<td>TM5</td>
<td>33.27</td>
<td>2.69</td>
<td>41 (82)</td>
</tr>
<tr>
<td></td>
<td>Sigm.</td>
<td>33.03</td>
<td>2.68</td>
<td>46 (69)</td>
</tr>
<tr>
<td></td>
<td>Unim.</td>
<td>33.04</td>
<td>2.69</td>
<td>33 (54)</td>
</tr>
<tr>
<td>1536</td>
<td>TM5</td>
<td>34.99</td>
<td>2.78</td>
<td>29 (58)</td>
</tr>
<tr>
<td></td>
<td>Sigm.</td>
<td>34.96</td>
<td>2.78</td>
<td>40 (55)</td>
</tr>
<tr>
<td></td>
<td>Unim.</td>
<td>34.98</td>
<td>2.78</td>
<td>25 (43)</td>
</tr>
<tr>
<td>JFK</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>768</td>
<td>TM5</td>
<td>34.10</td>
<td>4.17</td>
<td>61 (273) ✔</td>
</tr>
<tr>
<td></td>
<td>Sigm.</td>
<td>34.10</td>
<td>4.08</td>
<td>79 (356) ✔</td>
</tr>
<tr>
<td></td>
<td>Unim.</td>
<td>34.04</td>
<td>4.27</td>
<td>39 (123) ✔</td>
</tr>
<tr>
<td>1024</td>
<td>TM5</td>
<td>35.70</td>
<td>4.13</td>
<td>39 (145) ✔</td>
</tr>
<tr>
<td></td>
<td>Sigm.</td>
<td>35.65</td>
<td>4.11</td>
<td>46 (133) ✔</td>
</tr>
<tr>
<td></td>
<td>Unim.</td>
<td>35.66</td>
<td>4.13</td>
<td>30 (68)</td>
</tr>
<tr>
<td>1536</td>
<td>TM5</td>
<td>37.82</td>
<td>3.86</td>
<td>24 (70)</td>
</tr>
<tr>
<td></td>
<td>Sigm.</td>
<td>37.76</td>
<td>3.90</td>
<td>37 (65)</td>
</tr>
<tr>
<td></td>
<td>Unim.</td>
<td>37.78</td>
<td>3.89</td>
<td>22 (51)</td>
</tr>
</tbody>
</table>

Table 5.2: Mean and standard deviation of the performance measures for the quantiser control surfaces
Figure 5.19: The variation in the buffer occupancy for the quantiser control surfaces
These simulations were conducted by simply replacing the TM5's quantisation equation with the sigmoidal and the unimodal functions. However, they may perform more effectively when an integral rate control scheme is applied with them. The feed-forward, nonlinear rate estimator-based scheme is thought to be a more appropriate technique than the feedback rate control employed in the TM5.

The unimodal and sigmoidal control surfaces are their 3-D extensions of the 2-D equations which are originally defined in Eqns. 4.12 and 4.11. Similar results to Table 5.2 are also shown in Table 4.1. For the video sequences “Adverts” and “Starwars”, much smaller occupancy fluctuation is achieved in the unimodal function (LOGEXP-A) than in the sigmoidal function (SIGM-A) at the channel rate 1024 kbits/s, without further degradation in PSNR. The results shown in Table 5.2 were obtained without using the video rate predictors. Hence, further improvement can be achieved by the rate prediction. The detailed comparison for the rate prediction is given in Table 5.4.

5.7 RBF rate estimator-based MPEG2 video encoder

The RBF-network rate estimator is viewed as a nonlinear (universal) functional approximator [9,10], which approximates the video rate signal which is a real-valued function of the scene change feature vector \( \mathbf{x} \). From the equations described in [9],

\[
    \hat{c}bf(\mathbf{x}') = < \mathbf{x}', \mathbf{w} > + b
\]  

(5.42)

where \( \mathbf{x}' \) and \( \mathbf{w} \) are the processed version of the input vector \( \mathbf{x} \) and a vector of weights, respectively, and \( b \) is a scalar bias. The term \( < \mathbf{x}', \mathbf{w} > \) represents inner product. The input vector \( \mathbf{x} \) can be transformed into \( \mathbf{x}' \) by any continuous non-constant function. Thus, if this function is replaced with a radial basis function, e.g. Gaussian function, the resulting network becomes the radial basis function network which includes the Gaussian hidden layer.

The RBF network consists of centres with the radial basis function and linear weights given:

\[
    \hat{c}bf(\mathbf{x}) = \sum_{i=1}^{N} u_i \phi(||\mathbf{x} - \mathbf{x}_i||)
\]

\[
    \phi(||\mathbf{x} - \mathbf{x}_i||) = \exp\left(-\frac{||\mathbf{x} - \mathbf{x}_i||^2}{2\sigma^2}\right)
\]

(5.43)

where \( \hat{c}bf(\mathbf{x}) \) is the output of the RBF network, \( u_i \) is the linear weight, \( \mathbf{x} \) is an input vector containing scene change features, and \( \mathbf{x}_i \) represents the \( i \)th selected centre to represent a large number of input vectors in terms of Euclidian distance. \( \phi() \) is the Gaussian function which outputs RBF layer values determined by Euclidian distance between the vectors \( \mathbf{x} \) and \( \mathbf{x}_i \). The RBF network, shown in Fig. 5.20, may have as many centres as required by selecting input vectors, and it calculates the contribution of each input using the Gaussian function. For network efficiency, however, the RBF centres are usually selected by the orthogonal least square
(OLS) algorithm [144] or by clustering algorithms [8]. The OLS algorithm selects representative RBF centres when supervised learning is used. However, in the case of the MPEG2 encoder, supervised learning cannot be used properly, since statistical properties of the network input are not known and are non-stationary. Supervised learning is known to be effective to time series prediction where the signal is stationary or the statistical properties are known. Thus, unsupervised learning needs to be used in the RBF-network-based MPEG encoder. The k-means clustering algorithm is used for the unsupervised learning, as for RBF-network-based channel equalisation [145,146] applications. The centres are updated as follows [139]:

$$x_i(k) = x_i(k - 1) + g_c (c_{bf}(k) - x_i(k - 1))$$

(5.44)

where $x_i$ is the $i$th centre and the constant $g_c$ controls the learning rate in the $k$-means clustering algorithm. This clustering technique updates centres in reference to a single previous centre since the time index for the previous centre is set to $k - 1$. More previous centres can be taken into account to update the next centre. However, the value $k - 1$ is particularly useful when centres are changing rapidly due to dramatic scene changes for a short period of time. Thus, the $k - 1$ time index for centres is consistently used for all the RBF-network simulations.

![Diagram of RBF predictor with 3 inputs and 9 taps.](image)

Figure 5.20: RBF predictor with 3 inputs and 9 taps.

The number of selected centres is 9. As shown in Fig. 5.20, the input consists of the three scene change features, each of which has three taps. The number of time lags equals the number
of B pictures between P pictures. It is assumed that the cyclic repetition of the video rate
is determined by the number of B pictures (see Fig. 4.1) and the correlation of the video rate
varies depending on the interval between two B pictures. Thus, the three inputs with three taps
each form a nine samples per input vector. That is, the input dimension is 9. In general, the
number of RBF centres is much more than the input dimension. However, in this application
the number of centres is the same as the input dimension. The number of the RBF centres can
be set equal to the number of B pictures. This originates from the nature of the MPEG encoder.
The video rate variation depends largely on the picture type as well as the variances. Since
two B pictures are encoded between two P pictures, the video rate is related to the number
of B pictures. That is, it is assumed that 9 or 10 input vectors can represent the whole set of
input vectors when video is being encoded with the picture type order, e.g. IPBBPBBPBB...
This implies that the system complexity of RBF-network-based rate control can be considerably
reduced by selecting appropriate values for the video encoding parameters.

5.8 Simulation studies

The performance of the rate predictors was evaluated on the test bed of the MPEG2 software
encoder. The new configuration of the MPEG2 encoder shown in Fig. 5.14 was verified with a
variety of different settings of encoding parameters. In this section we present representative
simulation results for three different video sequences and six different rate control schemes.

5.8.1 Simulation setup

Three video sequences, “Cascaded”, “Starwars” and “JFK”, are used to give frequent scene
changes and input video with large scene variation to the encoder. See Fig. A.1, A.2 and A.4
in Appendix A. The video encoder was set to operate at a channel rate at 1024 kbits/s and a
frame rate at 30 frames/s. It has a buffer with the size of twice of MBF. All the encoded bit
streams were decoded with the MPEG2 decoder described in Section 3.7. For the nonlinear
quantiser control surfaces, $S_{max}$, defined in Section 5.5 is set to 7.

Since we focus on the CBR mode, simulation results for six rate control schemes for a fixed rate
channel are presented:

- TM5: as specified in TM5, Section 5.3,
- LIN: with a single linear quantiser control function,
- SIGM: with the sigmoidal control surface plus the heuristic prediction (SIGM-A) defined
  in Chapter 4
- LOGEXP: with the unimodal control surface plus the heuristic prediction (LOGEXP-A)
  defined in Chapter 4
- RLS: with the Type III linear rate estimator adapted by the RLS algorithm and the unimodal control surface

- RBF: with the RBF adaptive rate estimator and the unimodal control surface

### 5.8.2 The number of RBF-network centres

Different numbers of RBF centres - up to 50 - were simulated and their performance was compared in terms of the MSE, as shown in Fig. 5.21. The MSE is based on the same definition as Eqn. 5.9. For all four video sequences, the 9 centres appeared to exhibit very close performance to cases of larger number of centres. The performance appears to be better for the sequences with smaller variance values, e.g. “Cascaded” than those with larger variances, e.g. “JFK”. These are expected results since the video rate is easier to estimate when the overall scene change is small. (See Fig. 5.9.).

![Graph showing MSE profile depending on the number of centres.](image)

**Figure 5.21: MSE profile depending on the number of centres.**

<table>
<thead>
<tr>
<th></th>
<th>MSE[dB]</th>
<th>var_org(k)</th>
<th>var_diff(k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cascaded</td>
<td>-24.49</td>
<td>1224.8</td>
<td>579.6</td>
</tr>
<tr>
<td>Starwars</td>
<td>-22.44</td>
<td>1671.2</td>
<td>713.4</td>
</tr>
<tr>
<td>Adverts</td>
<td>-20.67</td>
<td>2888.5</td>
<td>1143.9</td>
</tr>
<tr>
<td>JFK</td>
<td>-15.33</td>
<td>3329.9</td>
<td>1914.7</td>
</tr>
</tbody>
</table>

**Table 5.3: Mean MSE and mean variances of video sequences.**
5.8.3 Performance evaluation and discussion

Three performance measures are evaluated; the coded bit rate, the buffer occupancy and PSNR. The key criterion to judge the performance is to assess the standard deviation of the measures. For all three performance measures, maintaining stable profiles is critically important. A better scheme should be able to achieve variations in the bit rate and the occupancy as small as possible. Also, for PSNR, a smaller variation can provide a constant level of video quality, if a high mean value is achievable.

Table 5.4 shows the mean and the standard deviation for each of the performance measures. Fig. 5.22 shows the performance of the six rate control schemes for the “Starwars” video sequence. The left hand plots correspond to TM5, linear and other control surfaces while the right hand plots to the linear and the nonlinear (RBF) estimators, which use the unimodal control surface (See also Fig. B.13, B.14, Table B.5 and Table B.6 in Appendix B for “Cascaded” and “JFK”). The NFVR in the middle column stands for normalised fluctuation of the video rate, which represents overall flatness of $cbf(k)$. It is expressed in the following equation:

$$NFVR = \frac{\sigma}{1 + \sigma}$$

where $\sigma^2 = E\left[\left(\frac{cbf(k)}{MBF} - 1\right)^2\right]$ (5.45)

and $\frac{cbf(k)}{MBF}$ represents instantaneous fluctuation. $E[\cdot]$ is the statistical expectation operator.

LIN shows better performance than TM5 in that it maintains a lower standard deviation for all three measures - though it has a simpler control process for quantisation scale - since it more effectively reflects the buffer occupancy without large PSNR degradation. The mean value of its PSNR is slightly better than TM5 while it keeps the occupancy significantly lower. SIGM shows virtually the same performance as LIN, as there is no noticeable difference in the occupancy and PSNR. However, its standard deviation in the occupancy is smaller than that of LIN. SIGM appeared to be capable of maintaining the occupancy in the middle range. LOGEXP controls the occupancy far lower than the previous rate control schemes with little degradation in PSNR, and it is able to reduce the rate fluctuation well in advance of the major scene changes, as shown in Fig. 5.22(a) but it shows much larger variation in the bit rate and NFVR than SIGM. The schemes based on rate estimators, RLS and RBF, outperform the others in terms of the flatness of these performance measures. The standard deviation of RLS is slightly smaller than those of LOGEXP without further degradation in PSNR. RBF appeared to be capable of maintaining the occupancy lower with a smaller standard deviation in comparison to RLS. In terms of PSNR, RBF has similar PSNR to LIN and SIGM. LOGEXP results in undesirable abrupt fluctuations which are marked as “A”, Fig. 5.22(a), showing a series of oscillations around frame number 100. This technique is not based on prediction technique as are RLS and RBF, thus, it may introduce such a degrading effect.
Fig. 5.23 shows the detailed variations in the quantisation scale and the corresponding occupancy over the frames from 142 to 148 for the six rate control schemes simulated. TM5 exhibits the worst occupancy performance, often reaching buffer overflow for “JFK” and “Cascaded”. The resulting abrupt change in the quantisation scale causes wider variations in the buffer occupancy. Consequently, it shows dramatic PSNR changes. While the x-axis of Fig. 5.22 is given in frame number, that of Fig. 5.23 is represented in macro block number. Thus, the latter shows detailed variation in the quantisation scale and the resulting occupancy. The mark “A” indicated in Fig. 5.23(a) shows dramatic changes in the occupancy as the picture type switches since TM5 uses a different occupancy for each picture type. LIN and SIGM shows similar variations depending on the occupancy. Their quantisation scale values do not exhibit such dramatic changes since a single buffer occupancy measure is applied to all the picture types. SIGM appeared to be more capable of reducing occupancy after the occupancy reaches 60% as indicated by “B” and “C”. LOGEXP has better adaptivity in controlling the quantisation scale, however, it may cause a rapid increase in the occupancy (“D” and “E”) when it changes the quantisation control curve from the logarithmic to the exponential, or the other way. Fig. 5.23(e) and (f) show that RBF has better performance than RLS when the occupancy dramatically increases, as indicated by “F” and “G”. While RLS shows 27% occupancy difference, RBF achieves 22% and shows stable occupancy for the frames 146 and 147. This signifies that RBF has the best rate prediction performance since it exhibits an extended capability, in comparison to the linear estimator.

<table>
<thead>
<tr>
<th>Starwars</th>
<th>Occupancy(%)</th>
<th>coded bits / frame (bits)</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean(max.)</td>
<td>std.dev.</td>
<td>NFVR</td>
</tr>
<tr>
<td>TM5</td>
<td>41 (75)</td>
<td>10.8</td>
<td>0.285</td>
</tr>
<tr>
<td>LIN</td>
<td>33 (51)</td>
<td>7.4</td>
<td>0.138</td>
</tr>
<tr>
<td>SIGM</td>
<td>35 (58)</td>
<td>6.6</td>
<td>0.129</td>
</tr>
<tr>
<td>LOGEXP</td>
<td>17 (36)</td>
<td>5.9</td>
<td>0.214</td>
</tr>
<tr>
<td>RLS</td>
<td>18 (38)</td>
<td>4.5</td>
<td>0.127</td>
</tr>
<tr>
<td>RBF</td>
<td>12 (34)</td>
<td>4.3</td>
<td>0.124</td>
</tr>
</tbody>
</table>

Table 5.4: Mean and standard deviation of performance measures (“Starwars”).
Figure 5.22: Performance of rate control algorithms ("Starwars"): (a) occupancy; (b) coded bits/frame; (c) PSNR.
Figure 5.23: Macroblock-wise variations in the quantisation scale and the buffer occupancy (“Starwars”): (a) TMS; (b) LIN; (c) SIGM; (d) LOGEXP; (e) RLS; (f) RBF.
Although the mechanism of LIN is far simpler than TM5, it shows superior profiles in the assessed performance measures. From our simulations it is confirmed that the bit allocation depending on the picture type is not appropriate for CBR applications which inevitably require short buffering delay. Though the linear estimator is inferior to RLS in tracking the variations in the scene change features, its linear response maintains a steadier occupancy value than LOGEXP.

Finally, for a subjective evaluation of the video quality, luminance pictures are shown in Fig. 5.24. The pictures show the case where TM5 results in the worst PSNR value. The pictures (a) and (b) are from the results of the TM5 and RBF rate control schemes, respectively. The frame number is 199 among 330 frames. The cascaded pictures of (c) are the adjacent ones of the 199th picture in the “JFK” sequence. They are shown in order to provide the scene change information in adjacent pictures. The scene change of the 199th picture is different from general ones in that the previous scenes show less variation in picture detail without colour image data. The very different scene (the 199th frame) follows the preceding monochrome images, i.e. the frames 198 and earlier ones. This causes rapid increase in the occupancy and the resulting PSNR decrease for the TM5 scheme as shown in Fig. B.14 (See the picture number around 200). The PSNR difference is about 4.3 dB between the two pictures; (a) and (b) in Fig. 5.24. The blocky artifacts clearly appear around the edges of the face, Fig. 5.24(a). On the other hand, the Fig. 5.24(b) exhibits much clearer edges in the corresponding locations.

5.8.4 Discussions

The profiles of the three input parameters appear to be very similar to those of the entropies of the original input pictures and difference pictures. It is reasonable to use entropies for inputs instead of variances. However, the entropies defined in Chapter 2 do not sufficiently accurately represent the scene change information since the value of the entropy usually ranges within 8 bits per pixel. Thus, it reveals a definite limitation in representing the scene change information. On the other hand, variances are able to represent various scene changes with efficiency. Therefore, the variances and $pType(k)$ are believed to be more suitable input features than the entropy.

Two different time lag values (0 and 2) of $i$ defined in Eqn. 5.6 have been tested on the rate predictors. For $i=0$, the rate predictor does not take account of the correlation between samples in a single scene change feature. It appeared to have better performance than the case of $i=2$ for the video containing more dramatic scene changes such as “Cascaded”. For an extremely dramatic scene change, increasing the number of taps does not achieve an improvement in prediction performance. This is highlighted in the scene cut from ”Miss America” to ”Football” where the change is too dramatic to be accommodated. However, setting $i=2$ makes use of the temporal correlation and generally shows better performance for realistic videos such as “Starwars” and “JFK” (Appendix B).
5.9 Conclusions

In this chapter the performance of conventional linear predictors has been examined on an off-line simulation, i.e. the prediction for given input and output without running the MPEG2 encoder. Through this simulation, it is confirmed that there exists a linear relationship between the scene change features and the video rate and that the linear estimator can be used effectively for video rate control. The nonlinear quantisation control was developed as a technique combining the scene change-based rate estimator with the MPEG2 quantiser. The effect on controlling the video rate was then verified for the linear estimator and the RBF-network estimators on the MPEG2 encoder platform. The predictive rate control approach using scene change features appeared to be effective for scene changes and it forms a successful mechanism for rate control in fixed rate video coders. This chapter has clearly demonstrated that using a rate predictor for the rate control system provides the encoder with a higher capability to manage and control the compressed video. As for the performance of the rate estimators, the RBF-network estimator appeared to be better than the linear estimator in terms of the video rate and the video quality. It shows virtually no degradation in PSNR in comparison to TM5 and other linear rate estimators. The nonlinear quantiser control technique is also confirmed to provide more substantial improvements in combination with the RBF rate estimator. The rate-distortion investigation is significant in a sense that the rate-distortion functions of nonlinear quantiser control are expressed as a generic form with respect to the buffer occupancy and the rate balance. This provides a useful approach to analytic evaluation of buffer-quantiser control algorithm. The developed rate control scheme was tested intensively for the CBR mode, however, since it is also crucial to regulating the video rate in the VBR mode in case of a congested network, it may be useful to achieve an effective network congestion control scheme.
Figure 5.24: Perceptual quality comparison (“JFK”): (a) TM5; (b) RBF; (c) adjacent pictures.
Chapter 6

Video rate control using fuzzy logic controller

6.1 Introduction

In Chapters 4 and 5, nonlinear predictive rate control techniques were examined. They appeared to be effective in terms of the stability of the video rate without objectionable PSNR degradation. However, some cases of dramatic rate fluctuation were observed, which require more effective techniques. In this chapter, fuzzy logic control is applied, which is known to outperform conventional linear control techniques in feedback control [147]. It is also known to be effective in conveying heuristic judgement of humans, which means introducing the meaning of linguistic decisions on the system status to the target system [148]. A series of conventional fuzzy logic control algorithms were applied to the CBR MPEG2 video coding. Furthermore, adaptive and supervisory fuzzy logic control schemes have also been applied and their performance has been evaluated.

In this chapter, conventional fuzzy rule-based control (FRC) is first examined in which the number of control variables is one, and its performance is evaluated for various settings of the fuzzy control parameters. A scene changed-assisted FRC scheme was developed as an adaptive algorithm and its performance was compared to the conventional FRC which will be referred as FRC-R (reference FRC). A series of simulations revealed that the FRC exhibits good performance for abrupt fluctuations in the video rate. However, a side effect associated with PSNR was observed. The PSNR measure was lower - by 2 dB in some cases - compared with the predictive techniques investigated in previous chapters. It is considered that conventional FRC does not effectively control the two variables (buffer occupancy and video quality) which have a trade-off relationship. That is, PSNR and the video rate may not be optimised simultaneously in the conventional FRC. Thus, it is necessary to introduce new techniques to deal with this problem. It appeared to be advantageous to use scene change features in a way which controls the quantisation step size more flexibly, depending on the extent of scene change. First, scene change features are used in the scene change-based FRC scheme (FRC-SC) to improve the performance of the FRC-R. For an advanced approach, a quality-monitored FRC scheme (FRC-QM), which takes both the occupancy and PSNR as fuzzy control variables, was applied. By investigating the FRC-based rate control schemes, this chapter is aimed to;

- Investigate the contribution of each fuzzy control parameter to rate control performance.
- Assess the performance of scene change features in fuzzy logic control.

- Improve the performance of conventional FRC by introducing a video quality measure applied to the quality-monitored FRC rate control scheme.

The FRC-R scheme consists of a widely used set of fuzzy rules and common fuzzification and defuzzification processes. For this scheme different values of three critical parameters were tested: scaling factors, membership functions and fuzzy control surfaces. Although an analogy can be found in other industrial applications, video rate control using fuzzy logic is not as simple as controlling the level of liquid in a tank or the pressure of water in a reservoir which have been treated as a typical prior FRC application [19]. Here the occupancy of the video buffer has a direct influence on video quality while the level or the pressure is irrelevant to the output fluid quality. Therefore, it is necessary to introduce a measure which effectively represents video quality in the fuzzy logic-based buffering. We consider both bit rate and PSNR as key control variables throughout this chapter. In the FRC-SC scheme, the scene change parameters were used to adaptively change the scaling factors. In the FRC-QM scheme a quality measure was derived and used as the control variable of buffer occupancy. The performance of the rate control schemes (FRC-R, FRC-SC and FRC-QM) was highlighted by presenting the profiles of the occupancy, the number of coded bits per frame and the PSNR.

This chapter is organised as follows: Section 6.2 briefly reviews fuzzy logic applications in relevant areas and describes a basic FRC configuration for video rate control. Section 6.3 discusses matters of major fuzzy logic control parameters. Section 6.4 presents configurations of the scene change-assisted FRC and the quality-monitored FRC. Section 6.5 includes the simulation results and discussions on performance. Finally, Section 6.6 concludes this chapter.

6.2 Fuzzy rule-based control for video rate control

In this section, a brief overview of fuzzy rule-based control techniques, which have been implemented during the past three decades, shows how fuzzy logic control has been applied to applications other than video rate control. This will provide an insight into the configuration of most common methods of using FRC. This configuration will be used to test the performance of what is viewed as a reference FRC as a basis for developing the improved FRC schemes examined in later sections in comparison to FRC-R.

Fuzzy logic and fuzzy set theory has been used extensively, particularly for industrial and commercial control applications [149, 150] since its concept [151] was first published. Although the controversy between conventional control and fuzzy logic control [152, 153] and orthodox set theory and fuzzy set theory [148] emerged occasionally in the early stages, successful applications of these techniques now range from medicine to home appliances in daily use, such as washing machines, air conditioners and camcoders [152].

104
6.2.1 A brief review of major applications of FRC

Applications of fuzzy logic control cover many different industrial and commercial products [154], mechanical alignment [155, 156], automated carrier trucks for factory automation [147], aircraft wing surface control [157], pH (acidity) level tuning [158], motor speed control [150, 160], liquid reservoir level control [19], etc. Many applications in a variety of fields of industry have raised a few fundamental issues, in particular, setting up appropriate rules and selecting membership functions, which are believed to dominate the performance of fuzzy logic control [161]. Automated formation of fuzzy logic rules has recently become the core of the research, with the emergence of neural network-based approaches (neuro-fuzzy) [147, 162] and genetic algorithm-based approaches [163].

Fuzzy logic control has also been used effectively in image processing [164–168] and computer vision [169]. Recently, FRC has been applied to video sequence coding algorithms such as JPEG [170, 171] and ITU-T H.261 [172]. These techniques aim to improve the video rate control performance for these standards by adaptively controlling the quantiser and the buffer occupancy. A FRC-based ATM traffic control technique [173] has also been published. The fuzzy rule-based control techniques used in those applications have the same technical basis in that they follow a series of common processes; fuzzification, decision making and defuzzification. However, they differ from one another in terms of the number of rules and input variables, the layout of membership functions, etc. They can also be differentiated by the process of defuzzification, which is also called fuzzy (or approximate) reasoning [15].

The fuzzy rule-based control technique can be summarised as shown in Figure 6.1. The output of the target system is controlled by feedback to the fuzzy rule-based control which is supplied with the error between output and the target value of the output, the set point. The entire FRC block consists of three processes; fuzzification, a rule base (decision making) and defuzzification. The FRC projects the meaning of the linguistic judgement defined in the fuzzy sets onto a numerical value of error using the fuzzification and the rule base. The numerical value is often called a crisp value in the sense that no human judgement is involved [15]. Then, the defuzzification process converts the linguistic meaning to the control input which is used in the target system. Output is often not the actual system output, which is generally used by other systems. It is a variable or a measure which indicates the current state of the actual output. In video rate control, the actual output is a compressed video bit stream, and the output corresponds to its rate or the buffer occupancy. This configuration is used for all the FRC models described in this chapter.
6.2.2 A basic FRC configuration for rate control

Generally, the FRC has control over a specific system parameter to be regulated within a predefined range or at a preset target value. The objective of video rate control is to keep the video rate of the encoder as stable or constant as possible in compliance with a given channel capacity. The FRC operates between the buffer and the quantiser. Figure 6.2 shows the configuration of a FRC-based video rate control which takes the buffer occupancy as its only input. The occupancy from the buffer is subtracted from the set point which represents required occupancy or delay. The resulting error becomes a crisp input value for the FRC to generate the quantisation step size at the end of its process. Thus, the FRC relates only to the buffering and the quantiser as depicted in Fig. 6.3.

![Diagram](image)

**Figure 6.1:** A basic model of fuzzy rule-based control.

**Figure 6.2:** Configuration of an MPEG2 video encoder with fuzzy logic rate control (FRC-R).

The MPEG video encoder section of interest includes the DCT, the quantiser, the VLC/MUX and the buffer. Here, the variable extracted from the encoder is the occupancy, \( O(n) \), and the variable applied to the encoder is the quantisation scale, \( Qs(n) \). The error signal between \( O_T \) (target occupancy) and \( O(n) \) is then \( e(n) \) which is fed to the fuzzification process together with the differential error, \( d(n) \), which is given by \( e(n) - e(n-1) \) for the time index \( n \). The functional block, \( \Delta e(n) \), represents the process to calculate \( d(n) \). The crisp values, \( e(n) \) and \( d(n) \) are multiplied by the scaling factors, \( ge \) and \( gd \), before being processed by the fuzzification. The linguistic judgements, \( Le \) and \( Ld \), are generated by mapping the scaled crisp values, \( Ge \) and \( Gd \), onto the pre-defined membership functions. In the decision making process, a linguistic judgement for the required action, \( Lo \), is made for inputs, \( Le \) and \( Ld \), based on the defined
decision rules. In the defuzzification process, $Lo$, is converted into the numerical value, $o(n)$, which is multiplied by the scaling factor, $go$. The final value, $Go$, i.e. $Qs(n)$ is used by the quantiser. In this FRC-R configuration, all the scaling factors ($ge$, $gd$ and $go$) are constants which are generally tuned by expert knowledge.

![Diagram](image)

Figure 6.3: Configuration of the FRC-based rate control (FRC-R).

### 6.2.3 Deriving a set of fuzzy rules for rate control

The first step of designing a FRC scheme is to transform the expert knowledge into a set of rules comprising linguistic expressions [154] so that a decision about the error, $e(n)$, can be made. A fuzzy set is expressed in a non-numerical form carrying linguistic meanings, e.g. very big, big, small or very small, etc. Three fuzzy logic variables are listed in Table 6.1, each of which consists of 7 fuzzy sets.

The fuzzy logic variable for occupancy (FVO) is used to represent the meaning of the buffer state, $e(n)$, with the central fuzzy set (HF: half full) and two extremes (FL: full and ET: empty). The fuzzy logic variable for the differential occupancy (FVD) represents the meaning of the differential value of $e(n)$, i.e. $d(n)$. FVD has been widely used in various fuzzy logic applications (See [155, 158, 167, 168]) as it signifies generally acceptable meanings of control variables. FVO and FVD exhibit an equivalent capability to represent fuzzy logic meanings since they have the same number of fuzzy sets and similar connotations. However, FVO is considered to be more suitable, particularly for representing the buffer states. For example, if the buffer occupancy approaches the full state, say 85% full or higher, close to full (CF) looks more appropriate than positive medium (PM) to express the buffer state since the former provides an more appropriate linguistic meaning. It is also beneficial in designing the membership function.

FVO, which is used for $Ge$, takes 50% occupancy, half full (HF), as a central value for a given

---

1The concepts of fuzzy set and variable are explained in [174]
delay requirement. Therefore, for a 2-frame delay the fuzzy set HF represents a delay of 1 frame. If the occupancy is higher than 50% but lower than 65%, it can be said to be in the HH state, if it is higher than 65% and lower than 85%, close to a full (CF) buffer. When it is higher than 85%, it may soon reach the buffer full state (FL). The low occupancy states can be defined in a similar manner. The fuzzy logic variable FVQ is used for the quantisation scale, i.e. control input, Fig. 6.1, as it can sufficiently express the meaning of its value to be used for a certain condition of input variables FVO and FVD.

<table>
<thead>
<tr>
<th>Fuzzy logic variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occupancy (FVO)</td>
</tr>
<tr>
<td>FL</td>
</tr>
<tr>
<td>CF</td>
</tr>
<tr>
<td>HH</td>
</tr>
<tr>
<td>HF</td>
</tr>
<tr>
<td>LH</td>
</tr>
<tr>
<td>CE</td>
</tr>
<tr>
<td>ET</td>
</tr>
</tbody>
</table>

Table 6.1: Fuzzy logic variables for the fuzzy control input and resulting output.

A fuzzy set generally corresponds to a predefined membership function ($\mu( )$) with a triangle, trapezoid or a Gaussian function [147] shape. We elected to use a triangular shape, Fig. 6.4(a), for the membership function as it is computationally simpler and has been shown to possess similar capabilities to the other shapes of membership function [147]. This membership function, which is mapped onto a normalised range from -.5 to .5, is used for two inputs (Ge and Gd) and the FRC output (Go) under the assumption that all the control variables have similar dynamic properties associated with the membership functions.

The complete representation of the rules can be given either in IF ... THEN statements or in the fuzzy associative memory (FAM) [147] tabular form, as shown in Fig. 6.4(b). The FAM representation is known to be more efficient in handling a complicated organisation of the rules [147]. Each fuzzy set assigned in the FAM represents the linguistic judgement of the corresponding output, Lo, according to the inputs, Le and Ld. The FAM can be configured by applying a method to locate fuzzy sets in the diagonal direction. MD is assigned on the diagonal positions. In this case the extreme value of either Le or Ld is compensated by the other. For example, if Le is FL and Ld is NB, the corresponding fuzzy variable for the output will be MD. By the same token, others in the MD diagonal represent such input pairs. The diagonal alignment is also applied to other fuzzy sets for the output. In the upper diagonal section,
SM, SL and TN are allocated along the diagonal direction with a descending order of signal magnitude. In the lower diagonal section LM, LG and HG possess an ascending order. Thus, the FAM provides an organisation of the fuzzy rules for the FRC-R model which corresponds to the configuration shown in Fig. 6.3.

\[ \mu(\cdot) \]

(a)

(b)\[
\begin{array}{cccccccc}
| Ld \times Le | \hline 
| ET \quad CE \quad LH \quad HF \quad HH \quad CF \quad FL | \hline 
| NB \quad NM \quad NS \quad ZE \quad PS \quad PM \quad PB | \hline 
| ET | TN | TN | SL | SL | SM | SM | MD | \\
| CE | TN | SL | SL | SM | SM | MD | LM | \\
| LH | SL | SL | SM | SM | MD | LM | LM | \\
| HF | SL | SM | SM | MD | LM | LM | LG | \\
| HH | SM | SM | MD | LM | LM | LG | LG | \\
| CF | SM | MD | LM | LM | LG | LG | HG | \\
| FL | MD | LM | LM | LG | LG | HG | HG |
\end{array}
\]

Figure 6.4: Fuzzy rule-based control parameters: (a) membership function (Type 0); (b) fuzzy associative memory map for deriving \( L_0 \).

6.2.4 Fuzzification and defuzzification

The decision making process is the key part to achieve the fuzzy logic control. However, the meaning of the fuzzy sets cannot be used directly for the target system without conversion to numerical values since the target system is in practice capable of only accepting numerical values. The conversion should be carried out from crisp values to fuzzy sets for the same reason that the decision making process can only accept linguistic judgements. Thus, there exists a necessity for using the fuzzification and the defuzzification processes as shown in Fig. 6.3. For fuzzification, the error and differential error, \( e(n) \) and \( d(n) \), are calculated as follows:

\[
\begin{align*}
 e(n) &= O(n) - O_T \\
 d(n) &= e(n) - e(n-1)
\end{align*}
\]

where \( O_T \) is generally set to 50%. The fuzzification converts crisp input values of \( Gc \) and \( Gd \) - which are scaled by \( gc \) and \( gd \) - to fuzzy sets according to the pre-defined FAM. The
membership functions, \( \mu() \) of the seven fuzzy sets shown in Fig. 6.4(a) are overlapped with adjacent ones, hence, a single crisp value normally intersects two rules except for the leftmost and the rightmost rules (ET and FL). For example, if \( Gc \) is .10 it involves the rules HF and HH with differing significance, marked as \( \alpha \) and \( \beta \) in Fig. 6.4(a). In this case \( Gc \) is much closer to HH, however, HF also has some significance in the linguistic judgement for \( Gc \). Thus, the resulting fuzzy sets are HF and HH. Likewise, if \( Gd \) is -0.08, the fuzzy sets NS and ZE are selected. As the fuzzification accepts two inputs (\( Gc \) and \( Gd \)), the maximum number of fuzzy sets involved for \( Go \) becomes 4.

Defuzzification, which converts \( Gc \) and \( Gd \) to an output value \( o(n) \), is carried out using the fuzzy rules determining \( Lo \) and the membership function values derived from input values (\( Gc \) and \( Gd \)). In the previous example, since \( Lo \) relates to the rules (SM, MD and LM) with two input values, 4 membership function values - \( \mu_{HF}(Gc = .10), \mu_{HH}(Gc = .10), \mu_{NS}(Gd = -.08) \) and \( \mu_{ZE}(Gd = -.08) \) - come out by mapping each input value to corresponding membership function. In the next step, two dominant membership function values are selected based on the set theoretic rule, either intersection or union for the two inputs, \( Gc \) and \( Gd \). Here, the intersection (the product rule) is used. For \( Gc \), since \( \mu_{HF}(Gc = .10) \) is smaller than \( \mu_{HH}(Gc = .10) \), \( \mu_{HH}(Gc = .10) \) is selected. Likewise, \( \mu_{NS}(Gd = -.08) \) is chosen. Each selected membership value is used to defuzzify the output into the numerical value, \( o(n) \), using the centre of gravity method [147] or one of its simplified versions (Mamdani’s model or Larsen’s product operation rule) [15,153] as depicted in Fig. 6.5. \( \mu_{HH} \) is mapped onto the two membership function values, \( \mu_{SM} \) and \( \mu_{MD} \), of \( o(n) \). In the same way, \( \mu_{NS} \) is mapped onto \( \mu_{MD} \) and \( \mu_{LM} \). Fig. 6.5 shows a simplified case for only two fuzzy sets SM and LM. \( A_i \) and \( A_j \) represent the areas covered by shaded triangles bounded by the trapezoids. \( Z_i \) and \( Z_j \) are normalised values of \( o(n) \). The final defuzzified output, \( o(n) \), is given:

\[
\begin{align*}
A & = A_i + A_j \\
W & = A_i Z_i + A_j Z_j \\
o(n) & = \frac{W}{A} \quad \text{(6.2)}
\end{align*}
\]

Figure 6.5: A simplified defuzzification process.
6.3 The fuzzy control parameters

In this section three major fuzzy control parameters are discussed in relation to rate control performance. While maintaining the basic configuration of the FRC-R shown in Fig. 6.3, various different parameter settings are examined to see how they influence the performance measures. The parameters are the scaling factor, the membership function and the FAM.

6.3.1 Scaling factors; \( g_e, g_d \) and \( g_o \)

The scaling factors can be tuned depending on the dynamic ranges of corresponding inputs, \( e(n) \) and \( d(n) \) and the output, \( o(n) \). The larger the values they take, the quicker the steady state response will be achieved, as in adaptive control. In video rate control \( g_o \) may be fixed at 1.0 since the actual output, \( Q_s \), is multiplied by 31 to adjust it onto the specified range of MPEG2 quantisation scale values. Even a small change of \( g_o \) can cause wide fluctuation in the quantisation scale value. \( g_e \) and \( g_d \) can be set to specific values suitable for the expected dynamic ranges of the \( e(n) \) and \( d(n) \) signals. The scaling factors can also be controlled adaptively by a supervisory function [147].

6.3.2 Membership functions

The membership function can take different shapes, overall formation, inter-rule spacing, etc. The formation can be asymmetrical since the positive and negative sections of the crisp values shown in Fig. 6.4(a), can have different significance. In video rate control, however, both sections here are assumed to have unbiased linguistic interpretation since neither overflow nor underflow is desired. Thus, the membership functions shown in Fig. 6.6 are symmetrical with respect to the centre value 0.

The spacings between fuzzy sets contribute to controlling the rapidity of the response to change in the inputs. The narrower spacing at the mid-range, Type 2, (Fig. 6.6(b)) has a quicker response to a crisp value change in comparison to Type 1 (Fig. 6.6(a)). For the same variation in the input crisp value, \( d(n) \), e.g. 0 to 0.25, the Type 1 associates only with HF and HH. On the other hand, the Type 2 involves the 4 fuzzy sets HF, HH, CF and FL. The latter is able to take early action for large values of \( d(n) \) so that the buffer occupancy remains at a stable level, although it is known to have problems related to stability; overshoot, undershoot and oscillation. Three types of spacing used in this chapter are summarised as follows;

- regular spacing (Type 0): equal spacing for both the middle range and end ranges, Fig. 6.4(a).
- wider-middle spacing (Type 1): wider spacing in the middle range and narrower spacing in end ranges.
**narrower-middle spacing (Type 2): wider spacing in end ranges and narrower spacing in the middle range.**

\[ \mu(0) \]

![Diagram](image)

**Figure 6.6:** Membership functions with irregular-spacing: (a) wider spacing around the middle value (Type 1); (b) narrower spacing around the middle value (Type 2).

### 6.3.3 FAMs and control surfaces

The 3-dimensional formation of a FAM reflects the dynamic property of a given organisation of the rules and membership functions since the FAM has a large influence on the performance of the entire FRC. For the same inputs, \( c(n) \) and \( d(n) \), a different FAM reacts with a different response, the resulting control surfaces have different formations, as shown in Fig. 6.7 and Fig. 6.8. Four FAMs - linear (LIN), logarithmic (LOG), exponential (EXP) and sigmoidal (SGM) - have been examined here in the FRC-R scheme, in order to introduce an analogy to the nonlinear quantisation control surface defined in Chapter 5. In the 3-dimensional representations in Fig. 6.7, \( FS(c) \) and \( FS(d) \) represent the corresponding fuzzy sets for the inputs \( c(n) \) and \( d(n) \) by assigning integers. The number 7 represents the fuzzy sets PB and FL, and the fuzzy sets NB and ET are represented by the number 1. The numbers 2 to 6 are assigned to the other fuzzy sets. The fuzzy sets for the output, \( o(n) \), are represented in the same manner. The overall appearance of the FAMs looks similar to that of each named function, e.g. the fuzzy associative memory EXP appears similar to an exponential function. The corresponding control surfaces resemble their 3-dimensional representation but their nonlinear change does not properly appear in the FAMs. Hence, the control surface is presented to provide a better insight into the dynamic properties of the designed FRC. The shapes of the control surfaces, LIN and SGM appear very similar, however, the formation of SGM is different from LIN at the bottom-left and the top-right corners. The nonlinear change is found in SGM, which does
not appear well in LIN. The slope of SGM in the middle section also differs from LIN. SGM shows a stiffer slope. LIN has the same organisation as the FAM shown in Fig. 6.4(b). The other three nonlinear control surfaces can be referred to as distorted versions of LIN based on logarithmic, exponential and sigmoidal functions.

\[
\begin{array}{cccccccc}
ET & TN & TN & SL & SL & SM & SM & MD \\
CE & TN & SL & SL & SM & MD & SM & LM \\
LH & SL & SL & SM & SM & MD & LM & LM \\
HF & SL & SM & SM & MD & LM & LM & LG \\
HH & SM & SM & MD & LM & LM & LG & LG \\
CF & SM & MD & LM & LG & LG & HG & HG \\
FL & MD & LM & LM & LG & LG & HG & HG \\
\end{array}
\]

\[
\begin{array}{cccccccc}
ET & TN & TN & SL & SL & SL & SM & SM \\
CE & TN & TN & SL & SL & SM & MD & LM \\
LH & SL & SL & SL & SL & SL & SM & SM \\
HF & SL & SL & SL & SL & SM & MD & SM \\
HH & SL & SL & SL & SL & SM & SM & MD \\
CF & SM & SM & SM & SM & MD & LM & LG \\
FL & SM & SM & SM & MD & LM & LG & HG \\
\end{array}
\]

Figure 6.7: 3-dimensional representation of the fuzzy associative memory: (a) LIN; (b) EXP; (c) LOG; (d) SGM.
Chapter 6: Video rate control using fuzzy logic controller

Figure 6.8: Corresponding control surfaces to the fuzzy associative memory in Fig. 6.7: (a) LIN; (b) EXP; (c) LOG; (d) SGM.

6.4 Improved FRC-based rate control algorithms

The improvement goal of the occupancy-based FRC is how effectively PSNR is used as a quality variable. Two different approaches were applied, aiming to obtain better PSNR figures in comparison to the previous FRC-R conventional approach. In this section the modified configurations of the FRC are compared to FRC-R; the FRC with adaptive scaling factors controlled by the scene change features (FRC-SC) and the quality-monitored FRC (FRC-QM). The former exploits the predictive property of the scene change features to control the scaling factors, $ge$ and $gd$. The latter takes into account two separate control variables for both occupancy and PSNR when it outputs the quantisation scale. The scene change features are the two variances, $var_{org}(k)$ and $var_{diff}(k)$, defined in Chapter 4 and the picture type, $ptype(k)$, used in Chapter 5.
6.4.1 Derivation of a dedicated fuzzy associative memory and its control surface

In the previous section, the 4 generalised FAMs (LIN, EXP, LOG and SGM) were employed to assess their effect on performance. It is well known that the formation of the FAM and the resulting control surface critically influence the entire performance. Thus, in this section, a derivation of a FAM, which is more suitable for the FRC-based video rate control, is undertaken.

Fig. 6.9 illustrates the procedure for forming the FAM and the resulting control surface. Fig. 6.9(a) depicts the significance of each boxed area in the FAM. [MD] resides in the centre position of the FAM where \( L_e \) is HF and \( L_d \) is ZE, that is, the occupancy, \( e(n) \), remains in half full range, and its variation is very small. In this condition, it is not required to take drastic control action to the quantisation scale, thus, the medium range values are used. The areas [A] [B] [C] and [D] represent the input conditions where either \( L_e \) or \( L_d \) remains stable. In [A] and [B] since \( L_d \) is ZE, fuzzy sets for \( o(n) \) are allocated depending on the fuzzy sets of \( L_e \), as shown in Fig. 6.9(b). Since \( L_e \) is HF in [C] and [D] areas, \( L_d \) affects the allocation of fuzzy sets for \( o(n) \) in the same way. The areas in the corners [1] [2] [3] and [4] correspond to the input conditions where both \( L_e \) and \( L_d \) influence the allocation of fuzzy sets for \( o(n) \). The position [1] signifies that the occupancy has decreased dramatically to the empty state since \( L_d \) is NB (negative big) and \( L_e \) is ET (empty). In this condition, a corresponding action should be taken to recover the occupancy quickly. Accordingly, the upper triangular area in the FAM (Fig. 6.9 (b)) is filled with TN, i.e. the smallest quantisation scale values. On the other hand, in the position [2] the occupancy has increased dramatically to the full state since \( L_d \) is PB (positive big) and \( L_e \) is FL (full). Therefore, the occupancy should reduce rapidly to the normal state (HF) by applying larger quantisation scale values, e.g. HG. This results in the lower triangular area of Fig. 6.9(b) being filled with HG. The areas [3] and [4] have exceptional physical meanings associated with actual buffer operations. [5] means that the occupancy has reached the empty state by adding the high level of occupancy (PB) to the previous occupancy. However, this is highly unlikely to occur in practice because there is no lower occupancy level than the empty state. Likewise, the area [4] also corresponds to a similar situation, since the occupancy cannot reach the full state by reducing it from further higher occupancy, which does not exist. Thus, corresponding areas are filled with the same fuzzy sets in the diagonal direction as the FAMs in Fig. 6.7. Considering all these conditions, the FAM configuration appears as shown in Fig. 6.9(b) and (c). The resulting control surface, Fig. 6.9(d), whose shape is similar to SGM, shows more nonlinear change in the shape than SGM. The membership function shown in Fig. 6.6(a) and the corresponding control surface, Fig. 6.9(d), are used for all the schemes investigated in the sections.
Figure 6.9: The fuzzy associative memory and its control surface used for performance improvement: (a) Allocation of fuzzy sets; (b) resulting FAM; (c) 3 dimensional representation; (d) control surface.

6.4.2 FRC using the scene change features (FRC-SC)

The scene change features used in Chapters 4 and 5 appeared to bring considerable improvement to the rate control schemes described in Chapters 4 and 5 by providing the video encoder with scene change information in advance of encoding input pictures. Here, they are coupled with the non-adaptive scaling factors ($gc$ and $gd$) of FRC-R as shown in Fig. 6.10 in order to adaptively change the inputs, $Ge$ and $Gd$. The scene change calculator (SCC) gives the three features ($var_{org}(k)$, $var_{dif}(k)$ and $ptype(k)$) to FRC-SC, in the same way described in Section 5.5, which generates scaling factor values for given scene change features. The rest of the FRC-SC operation is the same as the FRC-R depicted in Fig. 6.2. A scaling function is added to the FRC-R as shown in Fig. 6.11. It uses an equation with the variables scaled logarithmically as follows:
\[ g_e(k) = g_d(k) = \left[ \log_{10} \text{var}_{-\text{org}}(k) / \log_{10} \text{var}_{-\text{dif}}(k) \right] \times p_{\text{type}}(k) \]  

(6.3)

where \( k \) is the picture frame index. The logarithm is taken in order to scale down the dynamic range of the variances which often contain extremely large values. The ratio of \( \text{var}_{-\text{org}}(k) \) to \( \text{var}_{-\text{dif}}(k) \) is interpreted as follows; if \( \text{var}_{-\text{org}}(k) \) is bigger than \( \text{var}_{-\text{dif}}(k) \) - which means a picture has a greater amount of intra-frame detail than inter-frame scene change, the area “A” in Fig. 4.7, - then \( g_e(k) \) and \( g_d(k) \) become larger, since such pictures are assumed to be likely to require a large amount of coded bits. On the other hand, a larger \( \text{var}_{-\text{dif}}(k) \) is assumed to contribute less to the coded bits since a smaller \( \text{var}_{-\text{org}}(k) \) corresponds to a picture with less intra-frame detail, the area “B” in Fig. 4.7. The output scaling factor \( g_o(k) \) is set to 1 according to the stability rationale mentioned in the previous section.

![Figure 6.10: Scene change-based FRC for MPEG2 video rate control.](image)

![Figure 6.11: Configuration of the scene change-based FRC (FRC-SC).](image)
6.4.3 Quality-monitored FRC based on PSNR and the occupancy (FRC-QM)

Although video rate control aims to maintain the amount of coded bits per unit period as close as possible to the available channel rate, it should also take into account video quality. Fuzzy logic control algorithms FRC-R and FRC-SC take a single input (the occupancy) which is converted to $c(n)$ and $d(n)$. Since the rules and their control action are designed to control the occupancy, they are incapable of controlling the quality. Thus, FRC-R may cause considerable degradation in PSNR. In order to improve video quality, PSNR is introduced to FRC-R, and this constitutes the quality-monitored FRC (FRC-QM), which aims to achieve both rate control and quality enhancement simultaneously. Before dealing with FRC-QM, first, two single-input FRCs are evaluated to separately investigate the effect of the two variables; macro-block-wise bit-rate balance (for the FRC-B scheme) and frame-wise PSNR (for the FRC-Q scheme). These two FRCs are discussed one at a time. The quality issue is also discussed in relation to the VBR MPEG2 encoding.

FRC-B is supplied with its input from the rate balance, $B(n)$, as shown in Fig. 6.12. The balance is calculated by monitoring the accumulated number of bits, $ab(n)$, which is the summation of all the coded bits generated from the macro block 0 to $n - 1$:

$$
c(n) = B_T - \frac{(MBF - ab(n))/(MBN - n)}{MBF/MBN} \quad (6.4)
$$

$$
d(n) = c(n) - c(n - 1) \quad (6.5)
$$

The rate balance, $B(n)$ is conceptually equivalent to $R_b$ in Eqn.4.9. While $R_b$ is a frame-wise measure, $B(n)$ is macro block-wise. In these equations $B_T$ is the target value of the balance, which is set to 1. $MBN$ is the total number of macro blocks in a picture. The denominator $MBF/MBN$ represents mean bits allocated per macro block. The term $MBN - n$ ranges from $MBN$ to 1 as $n$ ranges from 0 to $MBN - 1$. Thus, the term $MBF - ab(n)$ becomes the bits to be allocated to the remaining macro blocks $MBN - n$. The error signal, $c(n)$, finally becomes the macro block rate balance. When $n$ is 0, $ab(n)$ is set to the number of remaining bits of the last macro block in the previous picture. As the macro block index $n$ increases, $c(n)$ reflects the accumulated result of $ab(n)$ to subsequent macro blocks. If a large number of bits are coded at earlier indices of $n$, the shortage of the remaining number of bits will be distributed to later ones. This will inform FCR-B that the quantisation scale should increase. On the other hand, if a smaller number of bits are consumed for the earlier macro blocks, the remaining ones are quantised with a smaller quantisation scale value in order to supply the macro blocks with more bits. This fuzzy logic control scheme is considered to be effective in maintaining the average video rate close to MBF. Hence, FRC-B is functionally the same as FRC-R since both of them aim to control the video rate or the occupancy. However, each one has its own advantages. FRC-R is capable of controlling the occupancy at a certain level by adjusting $O_{set}$ value or the target value of delay, maintaining mean video rate equal to MBF. However, if $O_{set}$ is set to a low value, e.g. 30%, this scheme may not accommodate video rate fluctuation since the
occupancy is controlled within the reduced range compared to the 50% by being 20% closer to the empty buffer. On the other hand, FRC-B is more powerful than FRC-R for controlling the bit rate itself since it directly monitors the number of bits by setting the value of $B_T$. Even if the $B_T$ value is set lower than 1, the FRC-B is capable of distributing the allocated bits to all the macro blocks in a picture.

![Figure 6.12: Configuration of the FRC based on macro-block-wise video rate balance (FRC-B).](image)

The FRC-R and FRC-B schemes are designed to control the buffer occupancy and the bit rate - which are macro block variables to be measured and controlled at every macro block start - while PSNR is a framewise variable. An occupancy value of a macro block includes a previous history for the specific period equivalent to a size of buffer. The occupancy has causality in that the previous occupancy subsequently influences the current occupancy since the current occupancy includes the number of remaining bits of the previous picture. On the other hand, PSNR is not causal in the sense that the PSNR value for a macro block or a frame is not associated with the previous counterparts. Although two adjacent PSNR values may appear close to each other due to the inter-frame correlation, the previous value does not contribute to the next one. That is, a high PSNR value for the previous picture does not necessarily result in the same outcome for the following one. Thus, a PSNR value of a macro block is not a suitable measure unless it is considered on a framewise basis. That is, FRC-Q scheme considering video quality needs to be designed so that the framewise PSNR may improve. In order to do so, it is required to determine a set point value for the framewise PSNR value (PSNRF) as follows:

$$Q_{set} = 255^2/10^{(PSNRF/10)} \times 256$$

$$TD = Q_{set} \times MBN$$

(6.6)

where $Q_{set}$ is obtained by calculating MSE for a given PSNR value in Eqn. 3.9. The resulting MSE represents the distortion at the pixel level. Since the macro block-level distortion is needed, the MSE should be multiplied by the number of pixels in the macro block. The multiplication
term, 256, represents the number of pixels in a macro block (for luminance, 16 \times 16) and MBN the total number of macro blocks in a frame. Thus, the total target distortion, TD, of a frame is given in Eqn. 6.6. The accumulated distortion at the macro block \( mb_n \), \( Ad(n) \), is given.

\[
Ad(n) = \sum_{i=0}^{n} \left( \left( \sum_{l=0}^{15} \sum_{m=0}^{15} (mb_i(l, m) - \hat{mb}_i(l, m))^2 \right) \right)
\]  

(6.7)

where \( i \) is the macro block indices, and \( l \) and \( m \) are pixel indices, respectively. The PSNR balance measure, \( Q(n) \), is defined in the same way as Eqn. 6.4. Thus, the two inputs, \( e_q(n) \) and \( d_q(n) \), are calculated as follows:

\[
e_q(n) = Q_T - Q(n) \\
Q(n) = \frac{(TD - Ad(n))/(MBN - n)}{Q_{set}} \\
d_q(n) = e_q(n) - e_q(n - 1)
\]

(6.8)

where the term \( TD - Ad(n) \) represents the remaining distortion which is allowed to occur in the subsequent encoding of the macro blocks. The entire process to obtain \( Q(n) \) is carried out in the distortion calculation block (DC), Fig. 6.13, for a given PSNR value and the two macro blocks, \( mb_{n+1} \) and \( \hat{mb}_{n+1} \). The relative video quality, \( e_q(n) \), is represented in reference to the set point value, \( Q_T \), which is set to 1, and the differential quality, \( d_q(n) \), by \( e_q(n) - e_q(n - 1) \). Using Eqns. 6.6 to 6.8, PSNR control can be achieved by distributing the distortion incurred in the preceding macro blocks to the subsequent macro blocks. Thus PSNR is controlled not just for a macro block but for a whole picture. The configuration of FRC-Q is the same as FRC-R shown in Fig. 6.3, except that the PSNR variables, \( e_q(n) \) and \( d_q(n) \), are used as the inputs to FRC-Q instead of \( e(n) \) and \( d(n) \). The fuzzy control input, \( Q_q(n) \) (the quantisation scale derived by FRC-Q), is used when FRC-Q is being used on a stand-alone basis.

For the quality-monitored FRC (FRC-QM), shown along with FRC-Q in Fig. 6.13, the same fuzzy rules and FAM are used in combination with the above-mentioned PSNR control formulae Eqns. 6.6 to 6.8. This scheme also exploits the adaptive scaling factor used in FRC-SC. The quantiser is supplied with DCT coefficients and the quantisation scale, \( Q_s(n) \). The quantiser outputs the quantised DCT coefficients for the inverse quantiser (IQ) and the inverse DCT (IDCT). The quantised DCT data are also encoded by the variable length coder (VLC) and the multiplexer (MUX). Two scaling factors are introduced to FRC-QM: \( e_q(n) \) from FRC-Q and \( g_e(k) \) from the scene change calculator, Fig. 6.11. Note that \( g_e(k) \) is supplied every picture while \( e_q(n) \) is every macro block. They are multiplied by \( e_o(n) \) and \( d_o(n) \) as expressed in Eqn. 6.9 and shown in Fig. 6.14:

\[
e_o(n) = (O(n) - O_T) \\
G_e = e_o(n)g_e(k) \\
d_o(n) = e_o(n)(e_q(n) + 1) - e_o(n - 1)
\]

(6.9)
where $O(n)$ and $O_T$ represent the buffer occupancy at the macro block $n$ and the target occupancy, respectively. The term $e_q(n) + 1$ functions as a scaling factor for $d_q(n)$, representing the required PSNR value given by FRC-Q. It ranges from 0 to 2 since $e_q(n)$ varies from -1 to 1 as given in Eqn. 6.8.

The quantisation scale, $Q_q(n)$, is not used for FRC-QM since it is associated solely with FRC-Q. Combining both quantisation scale values, $Q_q(n)$ and $Q_s(n)$, are considered with another fuzzy rule-based control to output the finalised scale value. However, our simulations verified that such a hierarchical approach did not contribute to performance improvement in terms of either PSNR or video rate reduction. This is because, for FRC-Q or FRC-QM, the final quantisation scale value turns out to be abnormally controlled, since it is neither the value of FCR-Q nor that of FCR-QM. Thus, both FRC-Q and FRC-QM receive false feedback values of $Q(n)$ and $O(n)$. This caused performance degradation rather than improvement. Therefore, the output quantisation scale of FRC-QM is used directly for the quantiser so that the resulting occupancy can be fed back properly to FRC-QM without additional processing.

Figure 6.13: Configuration of the quality-monitored FRC (FRC-QM).
6.5 Simulation studies

In this section simulation results are included for the fuzzy rule-based control models examined in the previous sections; FRC-R, FRC-SC, FRC-B, FRC-Q and FRC-QM. For FRC-R, we evaluated the performance for several different parameters; scaling factors ($ge$ and $gd$), membership functions and control surfaces. A comparative evaluation among the schemes was also presented in terms of the performance measures; the occupancy, coded bits/frame and PSNR for a CBR channel at 1024 kbits/s as in previous chapters.

6.5.1 Simulation environment

The basic FRC model (FRC-R), described in Section 6.2 and Section 6.3, has considerable flexibility to change the fuzzy control parameters. First, the scaling factors were assessed by applying 5 different values ranging from 2 to 16. The same values were allocated to the input scaling factors, $ge(k)$ and $gd(k)$. The scaling factor for the output $go(k)$ was set to 1. A comparative review was undertaken on the distribution of fuzzy sets used in the FAM during the decision making process. The target occupancy was set to 30%, in order to examine the delay values which are comparable to the rate control techniques used in Chapter 5. The performance for the different membership functions and the control surfaces, described in Figs. 6.6 and Fig. 6.8, was also evaluated. Three types of membership function profiles were simulated; Type 0, Type 1 and Type 2 as explained in Section 6.3. The performance of the four different control surfaces (LIN, EXP, LOG and SGM) were compared in relation to the membership function Type 1, to achieve better PSNR performance.

The FRC model based on the scene change features (FRC-SC) was assessed in comparison to FRC-R since the difference between the two models is in the scene change calculation block attached to FRC-R as shown in Fig. 6.10. Before evaluating the quality-monitored FRC (FRC-
QM), the scheme based on the bit-rate balance (FRC-B) and the PSNR-derived FRC scheme (FRC-Q) were first tested, which are controlled separately by the bit rate balance and quality criteria, respectively. Then the FRC-QM scheme was compared to FRC-R and FRC-SC. The target value of the PSNRF was set at 30 dB, 35 dB or 40 dB in FRC-Q and FRC-QM. All other settings are the same as FCR-SC.

6.5.2 Performance evaluation

For FRC-R, the effect of changing the FRC parameters was observed. The variations given to \( ge(k) \) and \( gd(k) \) appear to cause a shift in the distribution of fuzzy control rules invoked during the decision making process, as shown in Fig. 6.15. The two horizontal axes represent the fuzzy sets for the input variables, \( e(n) \) and \( d(n) \), respectively, given by the numbers as shown in Fig. 6.7. The vertical scale represents the number of hits (Nh) of the fuzzy rules used when the decision making process outputs the judgement for Lo. When the scaling factors \( ge(k) \) and \( gd(k) \) are set to 4, the distribution resides in the middle of the range of \( d(n) \) values. On the other hand, when the scaling factors are 16, the distribution is concentrated in the bottom-left corner where both \( e(n) \) and \( d(n) \) are small. Each circled area (marked as ‘*’) represents the distribution of rules used for large \( d(n) \) values. For larger \( d(n) \) values, large scaling factors appear to bring a quick response. Eventually, the overall distribution converges to the region representing smaller fluctuation of \( e(n) \) and \( d(n) \) values, i.e. the bottom-left corner. The arrows represent the shift of the distribution from the left-hand side to the right-hand side as the scaling factor values increase. This implies that the quantiser subsequently takes two extreme quantisation scale values - very large and vary small - for larger scaling factors.

Table 6.2 and Fig. 6.16 show encoding results for the different scaling factor values. The scaling factor value 16 shows the closest mean occupancy value to the target value (30%), while others have smaller ones. Accordingly, the variation for the scaling factor value 16 appeared to be smaller for the occupancy and the NFVR (defined in Eqn. 5.45, Chapter 5). While the standard deviation of the PSNR is fairly constant, the mean value decreases as scaling factors increase. For simplicity, only two extreme cases are displayed in Fig. 6.16(b) and (c). The bigger values of \( ge \) and \( gd \) (e.g. 16) exhibit far better control capability over the bits/frame and the occupancy variation is reduced. However, the PSNR profile is noticeably lower than the case of smaller scaling factor values (e.g. 2) since, in FRC-R, the quality is not taken into account, thus a relatively large distortion can be incurred. The MPEG2 video coding scheme makes use of a set of previous pictures as a reference for encoding, hence, a low PSNR value of the reference pictures (I or P picture) will propagate through subsequent pictures. In FRC-R, pictures which require a large amount of bits will be given much stronger control action. When they are I or P pictures, the distortion caused by rate control will affect the following ones. For this reason, as shown in Fig. 6.16(c), PSNR is low when scaling factors are large. Note in Fig. 6.16(c) the sudden increase in PSNR at frame 178. This is due to the scene change, at 4.33 sec in Fig. A.2, moving to simple image frames with less high frequency context and hence increased PSNR.
Fig. 6.17 shows the performance for the different membership functions. Type 0, shown in Fig. 6.4(a), causes the widest fluctuation in the occupancy and the coded bits/frame. For occupancy control performance Type 2, Fig. 6.6(b), shows a stable profile than the others. It is capable of responding rapidly to the variation in input values in comparison to Type 1 which shows similar results to Type 0 in terms of the coded bits/frame. While Type 0 and Type 1 exhibit similar PSNR performance, the PSNR of Type 2 is far lower due to its control action. Table 6.3 summarises the performance. Although Type 2 exhibits the smallest std. dev. in the occupancy, the best mean PSNR appears in Type 1 with a similar std. dev.

![Diagram](ge(k), gd(k) = 4)

![Diagram](ge(k), gd(k) = 16)

![Diagram](ge(k), gd(k) = 4)

![Diagram](ge(k), gd(k) = 16)

Figure 6.15: Distribution of fuzzy sets used for ge, gd = 4, 16 in the FRC-R scheme: (a) “Cascaded”; (b) “Starwars”.

<table>
<thead>
<tr>
<th>ge, gd</th>
<th>Occupancy(%)</th>
<th>Coded bits / frame</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean(max.)</td>
<td>std.dev.</td>
<td>NFVR</td>
</tr>
<tr>
<td>2</td>
<td>14 (31)</td>
<td>2.8</td>
<td>0.098</td>
</tr>
<tr>
<td>4</td>
<td>22 (32)</td>
<td>1.8</td>
<td>0.080</td>
</tr>
<tr>
<td>8</td>
<td>26 (31)</td>
<td>1.0</td>
<td>0.067</td>
</tr>
<tr>
<td>12</td>
<td>27 (31)</td>
<td>0.9</td>
<td>0.063</td>
</tr>
<tr>
<td>16</td>
<td>28 (31)</td>
<td>0.5</td>
<td>0.061</td>
</tr>
</tbody>
</table>

Table 6.2: The effect of varying scaling factors for “Starwars” in the FRC-R scheme.
Figure 6.16: The effect of varying scaling factors in FRC-R ("Starwars"): (a) occupancy; (b) coded bits/frame; (c) PSNR.
Figure 6.17: Performance comparison for the different membership functions in FRC-R ("Star-wars"): (a) occupancy; (b) coded bits/frame; (c) PSNR.
<table>
<thead>
<tr>
<th>Membership functions</th>
<th>Occupancy(%)</th>
<th>Coded bits / frame</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean(max.)</td>
<td>std.dev.</td>
<td>NFVR</td>
</tr>
<tr>
<td>Type 0</td>
<td>22 (32)</td>
<td>1.8</td>
<td>0.080</td>
</tr>
<tr>
<td>Type 1</td>
<td>19 (29)</td>
<td>1.8</td>
<td>0.076</td>
</tr>
<tr>
<td>Type 2</td>
<td>23 (24)</td>
<td>0.2</td>
<td>0.060</td>
</tr>
</tbody>
</table>

Table 6.3: Performance comparison for different membership functions in FRC-R ("Starwars").

Table 6.4 and Fig. 6.18 show the performance for the different control surfaces shown in Fig. 6.7. EXP and LOG appear to be the two extreme cases; the shape of the EXP control surface causes sudden increases around scene changes since it allocates smaller quantisation scale values for the middle and low ranges of the occupancy than the others. Thus, it generates spiky rate variations, and maintains a higher PSNR value than LOG which is designed to perform in the opposite way. Accordingly, LOG shows the smallest fluctuation in the occupancy and the coded bits/frame, and the lowest picture quality. This result is similar to the case explained in Chapter 4 - the nonlinear quantisation control functions LOG and EXP. The control surfaces LIN and SGM rank between EXP and LOG in the performance profile. As for the PSNR figures, LIN and SGM are very close to EXP. SGM appears to achieve a better rate control capability than the others since it maintains the occupancy much lower than LIN and EXP, showing the close PSNR value to them. The overall shape of the LIN control surface is more similar to SGM than LOG or EXP, thus, resulting PSNR values (33.15 and 33.02 dB) are very close to the PSNR of EXP (33.53 dB).

<table>
<thead>
<tr>
<th>Control surfaces</th>
<th>Occupancy(%)</th>
<th>Coded bits / frame</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean(max.)</td>
<td>std.dev.</td>
<td>NFVR</td>
</tr>
<tr>
<td>LIN</td>
<td>19 (29)</td>
<td>1.8</td>
<td>0.076</td>
</tr>
<tr>
<td>EXP</td>
<td>22 (38)</td>
<td>2.6</td>
<td>0.093</td>
</tr>
<tr>
<td>LOG</td>
<td>19 (22)</td>
<td>0.4</td>
<td>0.062</td>
</tr>
<tr>
<td>SGM</td>
<td>19 (24)</td>
<td>1.0</td>
<td>0.066</td>
</tr>
</tbody>
</table>

Table 6.4: Performance comparison for different fuzzy associative memory-control surfaces in FRC-R ("Starwars").
Figure 6.18: Performance comparison for different fuzzy associative memory-control surfaces in the FRC-R (“Starwars“): (a) occupancy; (b) coded bits/frame; (c) PSNR.
The improved fuzzy control approach (FRC-QM) will now be discussed. First, the two constituent schemes based on the macro-block-wise bit-rate balance (FRC-B) and the PSNR quality measure (FRC-Q) have been evaluated, in order to compare them with FRC-R.

Fig. 6.19 shows the results for FRC-B. The buffer occupancy appears to maintain the level of 45% to 50%; its mean value and the peak value are 46% and 53%, respectively, with the std. dev. of 2.06%. The PSNR values show promising results; the mean value of PSNR is 33.21 and its std. dev. is 2.58 dB. This signifies that the FRC-B scheme can provide a level of performance close to the membership function Type 1 or the LOG fuzzy associate memory. The FCR-B scheme does not use the occupancy value but uses the bit rate balance. Hence, when the scheme attempts to control the balance around the target value, the occupancy maintains similar levels to the previous simulations. Therefore, the occupancy fluctuates around the initial value, as shown in Fig. 6.19(a). Here, the initial occupancy is set to 50%. The profile of the coded bits/frame appears very stable except the frames 140 to 160 where dramatic scene changes occur. This technique is considered to be effective in delay critical applications since it can maintain a very stable occupancy level. However, it may not be efficient in that it does not exploit the occupancy margin below 40%.

Fig. 6.20 and Table 6.5 show the performance of FRC-Q for three different PSNR target values for the VBR mode encoding; the bits/frame profile has fluctuations, since the scheme only controls PSNR without considering the video rate. Although the increment in PSNR is 5 dB, the corresponding bit rate increase is not linearly proportional, due to the nature of the PSNR definition. The pattern of fluctuation for the 45-dB case appears to resemble the other cases. It is confirmed that PSNR is properly controlled by FCR-Q as shown in Table 6.5 as well as Fig. 6.20(b) without noticeable PSNR variation. Only the 35-dB case shows slightly larger variations, since the required PSNR is far lower than the other two values. In this case, the average quantisation scale value is larger than that of the 40 or 45 dB case, so that the quality requirement is satisfied. Higher quantisation scale values cause abrupt change in PSNR, as macro blocks with considerable picture detail cannot be reconstructed with the given quality requirement. The occupancy plots are not shown, because the FRC-Q scheme does not take account of occupancy. That is, it has no significance to compare the occupancy profile in this scheme.

Fig. 6.21 and Table 6.5 give a comparison between FRC-Q with $Q_T = 40$ dB and a VBR encoding which has a bit rate close to the FRC-Q case. Here, VBR(8) represents the VBR encoding with a fixed quantisation scale value 8 throughout the video sequence “Starwars”. The mean value of PSNR of FRC-Q is very close to that of VBR(8), and Fig. 6.21(a) shows a similar profile, while a large difference appears in the PSNR profile, Fig. 6.21(b). Generally, VBR video encoding is known to provide constant video quality without large variation [28,79,93]. However, its performance appears to be inferior to FRC-Q, since the quantisation step size is fixed in VBR encodings, disregarding quality variation caused by picture complexity change. On the other hand, FRC-Q takes the quality measure, PSNR, into account to maintain PSNR at a
constant level. Thus, the PSNR std.dev. of the VBR appears three times larger than that of FRC-Q, i.e. 0.14 and 0.47, respectively, in Table 6.5. Both FRC-Q and VBR(8) schemes do not take account of the occupancy since they only control the video quality. For the same reason as Fig. 6.20 and Table 6.5, the occupancy plots are not shown in Fig. 6.21.

Figure 6.19: Performance evaluation for FRC-B (“Starwars”): (a) occupancy; (b) coded bits/frame; (c) PSNR.
Figure 6.20: Performance evaluation for FRC-Q (“Starwars”): (a) coded bits/frame; (b) PSNR.

The MPEG2 standard [41] recommends that video encoders allocate different quantisation scale values to each picture type, e.g. 4:5:12 for I, P and B pictures, respectively. The I picture generally contains the most significant picture details as a reference picture for the other types of pictures. Thus, the smallest value (4) is allocated. However, simulations show that the picture-wise allocation of the quantisation scale appears to have worse performance than VBR(8) due to the nature of the video sequence containing a considerable amount of variation in picture details, Fig. B.15, VBR(4:5:12) (see Appendix B). The MPEG2 TM5 is known to be appropriate for video with small scene variation where inter-frame correlation is high. This is the case under the assumption that intra-coded macro blocks are very few in P and B pictures. However, since many macro blocks may be encoded in the intra mode in P and B pictures when many scene changes occur, this scheme does not benefit from such allocation. Hence, this scheme is
considered to be sub-optimal for realistic video.

![Graph](image)

**Figure 6.21**: Performance comparison between FRC-Q and VBR (“Starwars”): (a) coded bits/frame; (b) PSNR.

<table>
<thead>
<tr>
<th>Starwars</th>
<th>Coded bits / frame</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>std.dev.</td>
</tr>
<tr>
<td>35 dB</td>
<td>40092.5</td>
<td>13071</td>
</tr>
<tr>
<td>40 dB</td>
<td>99095.1</td>
<td>21424</td>
</tr>
<tr>
<td>45 dB</td>
<td>207759.2</td>
<td>30014</td>
</tr>
<tr>
<td>VBR (8)</td>
<td>94125.5</td>
<td>17578</td>
</tr>
</tbody>
</table>

**Table 6.5**: Performance comparison between FRC-Q and VBR at different PSNR target values (“Starwars”).
Finally, the performance of the improved schemes (FRC-SC and FRC-QM) are discussed and compared with FRC-R. As described in Section 6.3, the scaling factors of FRC-SC are controlled in conjunction with the scene change features. The FRC-QM scheme has been improved by introducing the quality measure, PSNR, to its fuzzy control section. Table 6.6 summarises the performance. The performance of TM5 is given for reference. Fig. 6.22 shows a critical part of the encoded results where the frame number ranges from 130 to 200. FRC-QM ranks between FRC-R and FRC-SC in terms of the occupancy and the coded bits/frame. While FRC-R is superior to FRC-SC in controlling the video rate or the occupancy, the latter shows far wider fluctuations since its scaling factors change according to Eqn. 6.3. The occupancy profile of FRC-SC may vary dramatically depending on the scene change, resulting in far lower occupancy than the target value, i.e. 30%. This is because the scaling factors of FRC-SC vary in accordance with the scene change features while those of FRC-R are fixed. The profile of the coded bits/frame exhibits similar changing patterns. FRC-R is powerful in controlling the occupancy but not in achieving higher video quality. While FRC-SC can achieve a better PSNR value than FRC-R, it’s occupancy shows a much more fluctuating profile. On the other hand, the FRC-QM exhibits nearly the same PSNR value as FRC-SC and the smaller occupancy variation than FRC-R. Thus, FRC-QM is considered to perform better than FRC-R and FRC-SC in terms of PSNR as well as the occupancy by exploiting both scene change features and the quality control measure.

<table>
<thead>
<tr>
<th>Starwars</th>
<th>Occupancy(%)</th>
<th>Coded bits / frame</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean(max.)</td>
<td>std.dev.</td>
<td>NFVR</td>
</tr>
<tr>
<td>TM5</td>
<td>41 (75)</td>
<td>10.8</td>
<td>0.285</td>
</tr>
<tr>
<td>FRC-R</td>
<td>27 (32)</td>
<td>1.5</td>
<td>0.086</td>
</tr>
<tr>
<td>FRC-SC</td>
<td>19 (42)</td>
<td>5.3</td>
<td>0.146</td>
</tr>
<tr>
<td>FRC-QM</td>
<td>22 (38)</td>
<td>4.29</td>
<td>0.128</td>
</tr>
</tbody>
</table>

Table 6.6: Performance comparison among FRC-R, FRC-SC and FRC-QM (“Starwars”).

In the simulation, the scaling factors are clipped at the maximum value (8) of their dynamic range to see the performance of controlling the occupancy. For FRC-SC and FRC-QM, the scaling factors are changing within a 1-to-8 range depending on the scene change features according to Eqn. 6.3. In FRC-SC, the factors $gc(k)$ and $gd(k)$ scale up the error signal $e(n)$ and the differential error signal $d(n)$. Fig. 6.10, to adaptively change the actual input of $e(n)$. The scaling factors are generally smaller than the maximum value of 8, and the resulting performance of the occupancy control in FRC-SC and the FRC-QM appears inferior to the FRC-R. That is, FRC-SC and FRC-QM show more fluctuating profiles. As a result of wider variations in the occupancy, FRC-SC may reach buffer underflow, as shown in Fig. B.18 for the
video sequence “Topgun” (See Appendix B). However, the scene change-based scaling factors, $ge(k)$ and $gd(k)$, are capable of improving PSNR by making the most of the occupancy margin. In FRC-QM, this capability is exploited effectively by applying the quality measure, $Q(n)$, Fig. 6.13. The differential quality, $e_q(n)$, in Eqn. 6.8 functions as a regulating factor for $e_o(n)$. After an abrupt scene change, $e_q(n)$ becomes negative since the accumulative quality measure, $Ad(n)$, exceeds the target distortion, TD. Hence, the term $e_q(n)+1$ becomes smaller than 1, Fig. 6.14. This prevents the occupancy from a drastic fall resulting in an empty buffer. This is the main advantage of the FRC-QM scheme over the other two schemes. As mentioned in Chapter 3 and Chapter 5, the video rate and the quality measure are coupled with each other through the buffer-based quantisation in a rate-distortion theoretic way. A change in either quantity is reflected onto the other quantity. Thus, this nature was considered in devising the FRC-QM scheme in a way in which higher PSNR values can be achieved and the occupancy is also in control in the targeted range. See Tables B.7, B.8 and B.9 and Figs. B.16, B.17 and B.18 in Appendix B for the results for the other video sequences.

For a subjective evaluation of the video quality, luminance pictures are shown in Fig. 6.23 for the largest quality difference. The pictures (a) and (b) are from the results of the FRC-R and the FRC-QM rate control schemes, respectively. The selected frame is the 199th among 330 frames, as compared in Chapter 5. The cascaded pictures of (c) are the adjacent ones of the 199th picture in the “JFK” sequence. The blocky artifacts clearly appear around the edges in the boxed areas, Fig. 6.23(a). On the other hand, the Fig. 6.23(b) exhibits much clearer details in the corresponding locations.
Figure 6.22: Performance profiles for FRC-R, FRC-SC and FRC-QM (“Starwan”): (a) occupancy; (b) coded bits/frame; (c) PSNR.
6.6 Conclusion

In this chapter fuzzy rule-based control schemes were applied to video rate control and their performance has been evaluated for a variety of different parameter settings. The basic fuzzy logic control model FRC-R was examined, using the occupancy as a fuzzy control variable. It was used to investigate the effect of fundamental fuzzy logic variables: scaling factors, membership functions and control surfaces. With this model as a performance reference, four different fuzzy logic control schemes were evaluated; the rate-balance-based scheme (FRC-B), the video quality-driven (FRC-Q), the scene change feature-assisted (FRC-SC) and the quality-monitored (FRC-QM). All the four schemes were compared in terms of the mean and standard deviation of the performance measures; the occupancy, the coded bits/frame and the PSNR. For the occupancy and the coded bits/frame, fuzzy rule-based control schemes appear to achieve the goal of video rate control, i.e. a stable bit rate profile no matter how dramatic the scene change is. However, the necessity of enhancing video quality has arisen since the FRC-R or FRC-B scheme does not take the quality factor into account. Thus, the FRC-Q scheme was examined, in order to combine it with FRC-R. The scene change features, used in Chapter 5, were also applied so that PSNR was improved by allowing occupancy fluctuation.

As video rate control aims at effective management of the bit rate as the top priority, the FRC-R scheme was adopted to form a basic scheme. The accumulative quality measure and the scene change features were then introduced to FRC-R to improve the video quality. The comparative review was given to the two enhanced schemes; FRC-SC and FRC-QM. FRC-SC exhibited better PSNR figures than the FRC-R, however, occasionally its occupancy fluctuation was unacceptably wide and dramatic. The FRC-QM scheme, which is monitored by the quality measure and controlled by the scene change-based scaling factors, appeared to be superior to the other schemes in respect of both occupancy and video quality. Simulations for different video sequences showed similar results. The basic fuzzy rule-based model (FRC-R) is recommended for occupancy (or delay) critical applications such as bi-direction communications. However, if a certain extent of occupancy variation is accepted, the occupancy margin can be efficiently utilised as shown in the FRC-SC and the FRC-QM schemes.
Figure 6.23: Perceptual quality comparison (“JFK”): (a) FRC-R; (b) FRC-QM; (c) adjacent pictures.
Chapter 7

Conclusions

7.1 Introduction

The work investigated in this thesis is primarily concerned with the data rate control of compressed digital video signals transmitted through constant bit rate channels. In particular, the work has focused on the buffer-based quantiser control technique for MPEG video coding. An effort was made to improve performance of rate control techniques, in terms of video rate and PSNR, by devising a supervisory monitoring and control system. In order to achieve this goal, a series of nonlinear signal processing techniques were applied to the MPEG video encoder; radial basis function network, fuzzy logic control and nonlinear quantiser control. The performance of these techniques was evaluated in both analytical and experimental ways. In this chapter, the main conclusions are highlighted. Advantages and limitations of the techniques are also discussed and contrasted. Finally, future directions of the work are discussed.

7.2 Major achievements of the work

The MPEG video rate control mechanism is basically a feedback control which adjusts the current video data rate in reference to previous history of the video rate, depending on the transmission buffer state. The major problem here was that the mechanism is completely blind to the future variation in the video rate which may cause buffer malfunctions; such as overflow, underflow or drastic fluctuations in buffer occupancy. The developed techniques functioned as a supervisory system which informs the feedback mechanism of the predicted future variation in the video rate. This enabled the quantiser, which determines the video rate and the quality, to take required control action before quantising the video data. The improved schemes appeared to outperform the MPEG2 evaluation model TM5. This implies, in a global sense, that the new schemes have succeeded in realising an effective feed-forward rate control system with improved performance. They have also demonstrated how feed-forward networks such as RBF-networks and fuzzy logic control can operate for the MPEG rate control, when proper scene change features are provided.

The first step in the introduction of the feed-forward approach was to confirm that there is a linear relationship between the scene change features and the video rate. In this process conventional linear predictors were used such as the least squares algorithm. For the 1st and 2nd-order scene change features, they demonstrated improved performance when estimating
the time series of video rate. The RBF-network was employed to evaluate whether it brings further improvement in estimation performance. For the same scene change features, the RBF-network estimator exhibited enhanced performance particularly for non-stationary transition in the video rate.

The way to make use of the estimated video rate information is to utilise an appropriate control function for the buffer occupancy, the estimated video rate and the quantisation step size. Two nonlinear functions were predominantly analysed and tested; the unimodal and the sigmoidal functions. Both functions were proved to have better rate control performance than linear functions. The unimodal shows better performance than the sigmoidal. This was also shown in the rate-distortion analysis by investigating the effect of the two functions on both video rate and the quality.

As an advanced feedback control technique, the fuzzy rule-based control was adopted; this operates in between the buffer and the quantiser. Since this technique is able to model the quantiser step size decisions by employing a series of non-linear and non-mathematical models rather than linear mathematical models, it is more effective in controlling the video rate. However, this is a feedback control which takes the buffer occupancy into account as the only variable. If the fuzzy rule-based control is designed only to control the occupancy, video quality cannot be improved, as the quality measure is not a control variable. Furthermore, no feed-forward information was provided to it. Thus, the scheme appeared to be effective in controlling the occupancy but the resulting PSNR value was lower than the RBF-network-based scheme due to the configuration of the fuzzy rule-based control. Improvements on this scheme were achieved by applying two techniques; the feed-forward adaptive scaling factors and the quality-monitoring. The two modified fuzzy rule-based control schemes exhibited improved PSNR results, also with the similar video rate performance.

The two approaches - the RBF-network and the fuzzy logic - have different structures and properties, thus, performance comparison is not straightforward. However, a comparison can be conducted from the user’s point of view. As far as the controllability over the occupancy is concerned, the fuzzy rule-based control scheme shows better performance than the RBF-network approach with virtually the same video quality. Although a detailed comparison in different aspects was not presented in this thesis, the fuzzy logic approach is considered to be computationally simpler and less complicated, which is critical to implementation. However, the simulation results show that the RBF-network scheme exhibits much lower mean buffer occupancy than the fuzzy rule-base control schemes. This feature is advantageous for low-delay applications since shorter delay can be achieved by maintaining low buffer occupancy. The RBF-network scheme also appears to better exploit the occupancy margin by allowing more occupancy variation than the fuzzy rule-based control. Furthermore, the parameters of the RBF-network estimator examined in this thesis were not selected by an analytical approach. The number of centres, which determines the complexity, was only determined by an experimental method. More intelligent selection of the centres, i.e. the training problem, is another
important issue to be investigated. Hence, it is likely that further improvement can be achieved from the RBF-network-based scheme by enhancing the training process, tuning the parameters and introducing more adaptivity. An in-depth research on these matters is a topic for future research.

However, a conclusion can be drawn in respect of system complexity. In the RBF-network-based scheme, supplementary processing is required on the estimated video rate information to derive a corresponding quantisation step size to achieve the predictive rate control function. On the other hand, the fuzzy rule-based control can accomplish the equivalent task by using a set of rules, a pair of control variables and adaptive scaling factors. This configuration is generally simpler than the approach for predictive scheme. Hence, as the performance is very similar, the fuzzy logic approach is more promising in terms of system complexity than the RBF-network-based scheme.

In this thesis, the 1st and 2nd-order scene change features were used, instead of the other classes or the higher order features described in Chapter 3, since they were more appropriate for the method of time series analysis and mathematical treatment. The variances of an original picture and its difference picture are thought to provide the rate control function with the required information. The motivation for using the variance is that the video rate information is in bits per frame, which is proportional to its source entropy or variance. The scene change features were shown to work effectively for the feed-forward video rate estimation techniques.

7.3 Advantages and limitations of the developed schemes

Advantages
The MPEG2 evaluation model TM5 specifies its rate control algorithm in a computationally expensive way. It employs the three-step bit allocation and rate control method, however, simulations did not show good performance of the TM5 in comparison with the techniques developed in this thesis. In particular, the techniques developed appeared to effectively control dramatic scene changes and video rate fluctuations. Their advantages can be summarised as follows:

- They are scalable in the aspects of implementation. The RBF-network-based method can be implemented in a variety of different ways depending on required complexity. The fuzzy rule-based control can be simplified by reducing the number of fuzzy rules.

- They can be applied to various different video coding schemes, independent of standards in which the buffer-based quantiser control is employed.

- They can also be used for the variable bit rate and the available bit rate modes for ATM networks, where transmission rate is negotiable and adjustable depending on its bandwidth availability.

Limitations
The developed algorithms also have limitations as follows:

- The system complexity may increase in the RBF-network-based technique, depending on its parameters: the number of centres, training methods, adaptation schemes used, etc.
- The capability of the RBF-network to trace scene changes may degrade in the case of consecutive drastic scene changes, due to the limitation of convergence speed of the \( k \)-means clustering algorithm used in this thesis.
- In the fuzzy rule-based control, the scene change information is used for scaling factors only. However, a more effective usage may improve the system performance further.

### 7.4 Limitations of the experiments

All experimental results described in this thesis were obtained from software simulations, in which software encoders and decoders process digital video sequences on Unix-based workstations. This environment for experiment has the following limitations:

- Due to shortage of computational power and storage space, only a limited number of pictures were tested, which does not fully represent the entire movie sequence.
- Since the software encoder relies on the simulated transmission setting, the results may differ slightly from real-time MPEG encoders. However, this difference does not alter the judgement on the relative performance of the developed techniques.
- The video quality is evaluated only by comparing the PSNR. More effective ways of subjective assessment - such as perceptual entropy and perceptual distortion rate function [39] - need to be introduced for a more complete evaluation.

### 7.5 Future of the research

The feed-forward video rate control approach, using scene change features, looks promising, in that it is capable of accommodating the non-stationary nature of realistic video. The RBF-network is suitable for this approach, however, similar techniques can also be used, such as other nonlinear time series prediction techniques and neural networks. An imminent problem to be solved is the optimisation of the RBF-network by selecting the most appropriate parameters for a particular application so that it has improved performance. In the fuzzy rule-based control, the video quality was considered by introducing a quality measure as one of the scaling factors. However, one can also employ a quality measure and the video rate as two fuzzy control variables with equal importance. This can be achieved by using another fuzzy rule base which accepts two quantisation step sizes each of which is calculated separately by the two earlier fuzzy rule bases. A critical view of this scheme is focused on how to compose the fuzzy rule base for two contradictory fuzzy variables. This problem has emerged as one of the challenging tasks in the field of fuzzy logic control. The scene change features used in this thesis are 1st and 2nd-order statistical quantities. As indicated in Chapter 3, there are other classes of scene change features, and they have the potential to be used effectively for the algorithms developed.
Further improvement can be brought by adequately quantifying the different scene change classes. A general problem throughout the simulations was how to select more appropriate quality measures, to evaluate the predicted significance of the simulations. Small variation, or degradation in PSNR does not necessarily lead to subjective quality degradation, since perceptual quality may be different, depending on the complexity of an image and/or of motion between pictures. Thus, if the developed algorithms can be evaluated by a widely accepted subjective evaluation measure, more improvement in the performance of these rate control algorithms can be achieved.
References


References


References


[137] Univ. of Berkeley, mm-fip.cs.berkeley.edu, 1993.


Appendix A

Video sequences used in simulations

MPEG SIF video sequences used in simulations: 352 pixels × 240 lines

- “Cascaded”
- “Starwars”
- “Adverts”
- “JFK”
- “Topgun”

In the following video sequence figures, the interval between two pictures is 1/3 s since decimation by 10 was applied. The blue triangles represent noticeable scene changes. While “Cascaded” is made of the standard sequences, others contain specific parts of movies and television advertisements.
Figure A.1: Video sequence “Cascaded” (Miss America - Football - Susie)
Figure A.2: Video sequence “Starwars”
Figure A.3: Video sequence “Adverts”
Figure A.4: Video sequence “JFK”
Appendix A: Video sequences used in simulations

Figure A.5: Video sequence "Topgun"
Fig. B.1: Envelope of luminance (Y) difference of video sequences used in simulations.

Chapters 2 to 6

Complementary results for Appendix B
<table>
<thead>
<tr>
<th>IQ:PQ:BQ</th>
<th>Bit rate [Mbits/s]</th>
<th>Coded bits / frame [bits/frame] (std. dev.)</th>
<th>PSNR [dB] (std. dev.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:1:1</td>
<td>10.896</td>
<td>44467</td>
<td>48.97 (0.55)</td>
</tr>
<tr>
<td>4:4:4</td>
<td>5.699</td>
<td>28472</td>
<td>44.43 (0.29)</td>
</tr>
<tr>
<td>7:7:7</td>
<td>3.230</td>
<td>19159</td>
<td>40.65 (0.38)</td>
</tr>
<tr>
<td>10:10:10</td>
<td>2.246</td>
<td>15441</td>
<td>38.39 (0.64)</td>
</tr>
<tr>
<td>13:13:13</td>
<td>1.650</td>
<td>12908</td>
<td>36.87 (0.79)</td>
</tr>
<tr>
<td>16:16:16</td>
<td>1.317</td>
<td>11303</td>
<td>35.71 (0.93)</td>
</tr>
<tr>
<td>19:19:19</td>
<td>1.093</td>
<td>10040</td>
<td>34.84 (1.02)</td>
</tr>
<tr>
<td>22:22:22</td>
<td>0.946</td>
<td>9307</td>
<td>34.13 (1.11)</td>
</tr>
<tr>
<td>25:25:25</td>
<td>0.836</td>
<td>8646</td>
<td>33.54 (1.17)</td>
</tr>
<tr>
<td>28:28:28</td>
<td>0.752</td>
<td>8031</td>
<td>33.01 (1.22)</td>
</tr>
<tr>
<td>31:31:31</td>
<td>0.688</td>
<td>7556</td>
<td>32.56 (1.26)</td>
</tr>
<tr>
<td>2:3:6</td>
<td>5.183</td>
<td>71944</td>
<td>43.45 (2.44)</td>
</tr>
<tr>
<td>4:5:12</td>
<td>2.752</td>
<td>54958</td>
<td>39.33 (2.84)</td>
</tr>
<tr>
<td>8:10:25</td>
<td>1.389</td>
<td>31667</td>
<td>35.52 (2.51)</td>
</tr>
</tbody>
</table>

Table B.1: Bit rate and PSNR means depending on the quantisation scale for VBR (“Star-wars”).
Figure B.2: Bit rate and PSNR results depending on the quantisation scale for VBR ("JFK").
Figure B.3: Bit rate and PSNR variations depending on the quantisation scale for VBR (“JFK”).
<table>
<thead>
<tr>
<th>IQ:PQ:BQ</th>
<th>M bit/s</th>
<th>bits/frame (sd)</th>
<th>PSNR[dB] (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:1:1</td>
<td>8.943</td>
<td>50838</td>
<td>49.25 (0.52)</td>
</tr>
<tr>
<td>4:4:4</td>
<td>4.214</td>
<td>39337</td>
<td>44.87 (0.36)</td>
</tr>
<tr>
<td>7:7:7</td>
<td>2.300</td>
<td>30966</td>
<td>41.60 (0.66)</td>
</tr>
<tr>
<td>10:10:10</td>
<td>2.246</td>
<td>26382</td>
<td>39.65 (0.95)</td>
</tr>
<tr>
<td>13:13:13</td>
<td>1.650</td>
<td>22338</td>
<td>38.30 (1.16)</td>
</tr>
<tr>
<td>16:16:16</td>
<td>1.317</td>
<td>19566</td>
<td>37.22 (1.33)</td>
</tr>
<tr>
<td>19:19:19</td>
<td>1.093</td>
<td>17422</td>
<td>36.39 (1.46)</td>
</tr>
<tr>
<td>22:22:22</td>
<td>0.775</td>
<td>15974</td>
<td>35.69 (1.58)</td>
</tr>
<tr>
<td>25:25:25</td>
<td>0.699</td>
<td>14685</td>
<td>35.07 (1.66)</td>
</tr>
<tr>
<td>28:28:28</td>
<td>0.640</td>
<td>13552</td>
<td>34.53 (1.74)</td>
</tr>
<tr>
<td>31:31:31</td>
<td>0.594</td>
<td>12678</td>
<td>34.05 (1.80)</td>
</tr>
<tr>
<td>2:3:6</td>
<td>3.904</td>
<td>77162</td>
<td>44.08 (2.32)</td>
</tr>
<tr>
<td>4:5:12</td>
<td>2.024</td>
<td>52333</td>
<td>40.66 (2.46)</td>
</tr>
<tr>
<td>8:10:25</td>
<td>1.041</td>
<td>31036</td>
<td>37.42 (2.38)</td>
</tr>
</tbody>
</table>

Table B.2: Bit rate and PSNR means depending on the quantisation scale ("JFK").
### Motion estimation

\[
mae(k, l) = \left( \sum_{i=k}^{j=(i+1)(MB-1)} \sum_{j=l}^{(j+1)(MB-1)} |mb_{curr}(i,j)-mb_{prev}(i,j)| \right)_{i=k+MBN, j=l+MBN}
\]

<table>
<thead>
<tr>
<th>mae</th>
<th>Addition</th>
<th>Multiplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>sub-total</td>
<td>( SWY \times SWX \times (MB \times MB \times 2) \times MBN = 357,519,300 )</td>
<td>none</td>
</tr>
</tbody>
</table>

### Intra-Inter decision

\[
\begin{align*}
var_{mbo}(k, l) &= \sum_{i=k}^{j=(i+1)(MB-1)} \sum_{j=l}^{(j+1)(MB-1)} (b(i,j)-\mu)^2 \\
var_{mbd}(k, l) &= \sum_{i=k}^{j=(i+1)(MB-1)} \sum_{j=l}^{(j+1)(MB-1)} (d(i,j)-\delta)^2
\end{align*}
\]

<table>
<thead>
<tr>
<th>var_mbo</th>
<th>var_mbd</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition</td>
<td>Multiplication</td>
<td></td>
</tr>
<tr>
<td>sub-total</td>
<td>((MB \times MB \times 4) \times MBN = 337,920)</td>
<td>((MB \times MB \times 2) \times MBN = 168,960)</td>
</tr>
</tbody>
</table>

### DCT, IDCT

\[
DCT(k,l) = \sum_{i=0}^{BL-1} b(k,i) \times DCT(l,i)
\]

<table>
<thead>
<tr>
<th>for row</th>
<th>for column</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Addition</td>
<td>Multiplication</td>
<td></td>
</tr>
<tr>
<td>sub-total</td>
<td>((BL \times BL \times BL) \times 6 \times 6 \times MBN = 4,055,040)</td>
<td>((BL \times BL \times (BL+2)) \times 6 \times 6 \times MBN = 8,110,080)</td>
</tr>
</tbody>
</table>

### Quantisation, I.Quantisation

\[
\begin{align*}
DCTQ(k,l) &= DCT(k,l) / (Q\_MAT(k,l) \times mquant) \\
DCTR(k,l) &= (DCTQ(k,l) \times Q\_MAT(k,l) \times mquant) / 16
\end{align*}
\]

<table>
<thead>
<tr>
<th>Quantisation</th>
<th>I.Quantisation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiplication / Division</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sub-total</td>
<td>((BL \times BL) \times 2 \times MBN)</td>
<td>((BL \times BL) \times 3 \times MBN)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>sub-total</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>((BL \times BL) \times 5 \times MBN = 105600)</td>
<td></td>
</tr>
</tbody>
</table>

Table B.3: Computational requirement of the main profile MPEG1 encoder.
<table>
<thead>
<tr>
<th></th>
<th>Addition</th>
<th>Multiplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>$var_{\text{org}}$</td>
<td>$\text{COL} \times \text{ROW}$</td>
<td>$\text{COL} \times \text{ROW}$</td>
</tr>
<tr>
<td>$\text{mean}$</td>
<td>$\text{COL} \times \text{ROW}$</td>
<td>*</td>
</tr>
<tr>
<td>$\text{var}_{\text{org}}$</td>
<td>$\text{COL} \times \text{ROW} \times 2$</td>
<td>$\text{COL} \times \text{ROW}$</td>
</tr>
<tr>
<td>sub-total</td>
<td>$\text{COL} \times \text{ROW} \times 3$</td>
<td>$\text{COL} \times \text{ROW}$</td>
</tr>
<tr>
<td></td>
<td>$= 253,440$</td>
<td>$= 84,480$</td>
</tr>
</tbody>
</table>

$$var_{\text{dif}} = \frac{\sum_{i,j}^{\text{ROW},\text{COL}} (\sum_{i,j}^{\text{ROW},\text{COL}} (b(i,j) - \mu)^2)}{(\text{ROW} \times \text{COL})}$$

<table>
<thead>
<tr>
<th></th>
<th>Addition</th>
<th>Multiplication</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{mean}$</td>
<td>$\text{COL} \times \text{ROW}$</td>
<td>*</td>
</tr>
<tr>
<td>$\text{var}_{\text{org}}$</td>
<td>$\text{COL} \times \text{ROW} \times 2$</td>
<td>$\text{COL} \times \text{ROW}$</td>
</tr>
<tr>
<td>sub-total</td>
<td>$\text{COL} \times \text{ROW} \times 3$</td>
<td>$\text{COL} \times \text{ROW}$</td>
</tr>
<tr>
<td></td>
<td>$= 253,440$</td>
<td>$= 84,480$</td>
</tr>
</tbody>
</table>

Table B.4: Computational requirement for calculating the scene change features.
Figure B.4: Number of intra-coded macro blocks in P and B pictures ($IIF(k)$).
Figure B.5: Mean value of the quantisation scale per slice.
Figure B.6: The number of coded bits per frame with rate control.
Figure B.7: Examples of buffer occupancy in the MPEG1 video encoder.
Figure B.8: The PSNR profile for two different bit rates.
Figure B.10: Recursive linear prediction for "Advers" (VBR).

Figure B.11: Linear prediction for "Advers" (VBR).
Figure B.11: Linear prediction for "Adverts" at 1280 kbit/s.

Figure B.12: Recursive linear prediction for "Adverts" at 1280 kbit/s.
Figure B.12: Performance of rate control algorithms with linear or nonlinear rate estimators.
<table>
<thead>
<tr>
<th>Cascaded</th>
<th>Occupancy(%)</th>
<th>coded bits / frame (bits)</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean(max.)</td>
<td>std.dev.</td>
<td>NFVR</td>
</tr>
<tr>
<td>TM5</td>
<td>35 (106)</td>
<td>27.33</td>
<td>0.397</td>
</tr>
<tr>
<td>LIN</td>
<td>29 (77)</td>
<td>22.48</td>
<td>0.161</td>
</tr>
<tr>
<td>SIGM</td>
<td>31 (70)</td>
<td>18.57</td>
<td>0.168</td>
</tr>
<tr>
<td>LOGEXP</td>
<td>18 (54)</td>
<td>16.99</td>
<td>0.161</td>
</tr>
<tr>
<td>RLS</td>
<td>41 (54)</td>
<td>15.33</td>
<td>0.149</td>
</tr>
<tr>
<td>RBF</td>
<td>14 (56)</td>
<td>14.75</td>
<td>0.146</td>
</tr>
</tbody>
</table>

Table B.5: Mean and standard deviation of performance measures for feed-forward rate control techniques (“Cascaded”).

<table>
<thead>
<tr>
<th>JFK</th>
<th>Occupancy(%)</th>
<th>coded bits / frame (bits)</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean(max.)</td>
<td>std.dev.</td>
<td>NFVR</td>
</tr>
<tr>
<td>TM5</td>
<td>39 (172)</td>
<td>23.00</td>
<td>0.378</td>
</tr>
<tr>
<td>LIN</td>
<td>28 (84)</td>
<td>13.88</td>
<td>0.198</td>
</tr>
<tr>
<td>SIGM</td>
<td>31 (76)</td>
<td>11.72</td>
<td>0.188</td>
</tr>
<tr>
<td>LOGEXP</td>
<td>16 (60)</td>
<td>11.79</td>
<td>0.217</td>
</tr>
<tr>
<td>RLS</td>
<td>17 (76)</td>
<td>10.33</td>
<td>0.188</td>
</tr>
<tr>
<td>RBF</td>
<td>11 (61)</td>
<td>10.07</td>
<td>0.169</td>
</tr>
</tbody>
</table>

Table B.6: Mean and standard deviation of performance measures for feed-forward rate control techniques (“JFK”).
Figure B.15: Performance comparison between FRC-Q and VBR (IQ:PQ:BQ = 4:5:12) (“Star-wars”).
### Appendices

#### Appendix B: Complementary results for Chapters 2 to 6

<table>
<thead>
<tr>
<th>Cascaded</th>
<th>Occupancy(%)</th>
<th>Coded bits / frame</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean(max.)</td>
<td>std.dev.</td>
<td>NFVR</td>
</tr>
<tr>
<td>TM5</td>
<td>37 (102)</td>
<td>26.9</td>
<td>0.397</td>
</tr>
<tr>
<td>FRC-R</td>
<td>27 (41)</td>
<td>3.6</td>
<td>0.102</td>
</tr>
<tr>
<td>FRC-SC</td>
<td>22 (50)</td>
<td>13.8</td>
<td>0.149</td>
</tr>
<tr>
<td>FRC-QM</td>
<td>22 (46)</td>
<td>10.8</td>
<td>0.141</td>
</tr>
</tbody>
</table>

Table B.7: Performance comparison for FRC-R, FRC-SC and FRC-QM (“Cascaded”).

<table>
<thead>
<tr>
<th>JFK</th>
<th>Occupancy(%)</th>
<th>Coded bits / frame</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean(max.)</td>
<td>std.dev.</td>
<td>NFVR</td>
</tr>
<tr>
<td>TM5</td>
<td>39 (172)</td>
<td>23.0</td>
<td>0.378</td>
</tr>
<tr>
<td>FRC-R</td>
<td>27 (38)</td>
<td>2.6</td>
<td>0.098</td>
</tr>
<tr>
<td>FRC-SC</td>
<td>18 (61)</td>
<td>9.7</td>
<td>0.185</td>
</tr>
<tr>
<td>FRC-QM</td>
<td>21 (56)</td>
<td>8.0</td>
<td>0.169</td>
</tr>
</tbody>
</table>

Table B.8: Performance comparison for FRC-R, FRC-SC and FRC-QM (“JFK”).

<table>
<thead>
<tr>
<th>Topgun</th>
<th>Occupancy(%)</th>
<th>Coded bits / frame</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean(max.)</td>
<td>std.dev.</td>
<td>NFVR</td>
</tr>
<tr>
<td>TM5</td>
<td>38 (114)</td>
<td>30.0</td>
<td>0.321</td>
</tr>
<tr>
<td>FRC-R</td>
<td>27 (34)</td>
<td>2.7</td>
<td>0.095</td>
</tr>
<tr>
<td>FRC-SC</td>
<td>17 (47)</td>
<td>13.6</td>
<td>0.154</td>
</tr>
<tr>
<td>FRC-QM</td>
<td>19 (42)</td>
<td>10.9</td>
<td>0.142</td>
</tr>
</tbody>
</table>

Table B.9: Performance comparison for FRC-R, FRC-SC and FRC-QM (“Topgun”).
Figure B.16: Performance comparison for FRC-R, FRC-SC and FRC-QM (“Cascaded”).
Figure B.17: Performance comparison for FRC-R, FRC-SC and FRC-QM (“JFK”).
Figure B.18: Performance comparison for FRC-R, FRC-SC and FRC-QM (“Topgun”).
Appendix C

List of publications

Papers published or accepted

(i). Yoo-Sok Saw, Peter M Grant and John M Hannah, “Reduced storage transmission buffer
designs for an MPEG video encoder”, IEE IPA’95, pp. 608 - 612, Edinburgh, UK, 4th -

(ii). Yoo-Sok Saw, Peter M Grant and John M Hannah, “Feed-forward Buffering and Rate
Control based on Scene Change Parameters for MPEG Video Coder”, European Signal
Processing Conference (EUSIPCO) ’96, V. 2, 10 - 13, September, 1996, pp. 727-730.

(iii). Yoo-Sok Saw, Peter M Grant, John M Hannah and Bernard Mulgrew, “A radial basis
function video rate estimator for constant bit rate MPEG coders”, IEE Electronics Letters,

(iv). Yoo-Sok Saw, Peter M Grant, John M Hannah and Bernard Mulgrew, “Non-linear pre-
dictive rate control for constant bit rate MPEG video coders”, to appear IEEE ICASSP
97, April, Munich, Germany.

(v). Peter M Grant, Yoo-Sok Saw, John M Hannah and Bernard Mulgrew, “Feed-forward
nonlinear network approaches for MPEG video rate prediction”, to appear EURASIP
ECSAP’97, June, Prague, Czech Republic.

(vi). Peter M Grant, Yoo-Sok Saw and John M Hannah, “Fuzzy rule-based MPEG video rate
prediction and control”, to appear EURASIP ECSAP’97, June, Prague, Czech Republic.

Papers submitted

(i). Yoo-Sok Saw, Peter M Grant, John M Hannah and Bernard Mulgrew, “Video Rate Control
using a Radial Basis Function Rate Estimator and Scene Change Features for Constant Bit
Rate MPEG Coders”, submitted to EURASIP Signal Processing: Image Communication.
Appendix D

Publications

Papers listed in Appendix C, from i to vi.
REDUCED STORAGE TRANSMISSION BUFFER DESIGNS FOR AN MPEG VIDEO CODER

Y S Saw  J M Hannah  P M Grant

Electrical Engineering, University of Edinburgh, Scotland

Abstract
In MPEG video encoding, efficient buffering and rate control is especially crucial for constant bit rate (CBR) applications such as non-ATM (Asynchronous Transfer Mode) channels and satellite communication channels. In a CBR environment, compressed video data, which is inherently variable in terms of bit rate, should be throttled to a channel with fixed rate by managing the buffer operation. At lower transmission rates, or in the case of an abrupt scene change, a dramatic increase in buffer occupancy or a buffer overflow occurs and this may cause an interruption to normal encoding and consequent degradation of video quality. An intelligent buffering algorithm is proposed for preventing buffer overflow and smoothing out the occupancy fluctuation. The algorithm exploits major system parameters which have direct influence on one another in the MPEG encoder. The performance of the proposed algorithm has been verified on an MPEG1 encoder.

1 Buffering in MPEG
MPEG, ISO (1) is used for video compression and transmission with constant bit rates of up to 1.5 Mbit/s which is usually called the sub-primary transmission rate. As the compressed video stream from an MPEG video encoder comprises variable rate traffic there has to be a buffer for converting the variable rate traffic to constant rate traffic in a way which controls the amount of incoming video traffic. Research on buffering and adaptive quantisation has been carried out, Chen and Wong (2), Lee et al (3), Kawashima (4). Most of this research focuses on adaptive quantisation itself e.g. adaptive bit allocation, Geresh and Gray (5). MPEG is based on inherent system parameters such as picture type, direction of motion compensation, quantisation scale, and quantisation matrices, etc. These parameters are inter-related and this affects the operation of the encoder. Hence, in this research, more attention is given to these inherent system parameters. The key parameters which control the traffic rate are quantisation scale and quantisation matrices, ISO (1), especially, quantisation scale which directly controls the buffer occupancy. As the number of coded bits of the encoder fluctuates with the other parameters such as picture type as well as picture details, the encoder generates a compressed video bit stream with periodic changes. Eventually, the occupancy shows repetitive variation. This is an important system parameter to be considered in buffering. The buffer control mechanism defined in MPEG gives basic guidelines for avoiding interruptions to buffer operation or abrupt change of video quality. They are a few qualitative guidelines as follows;
- allocating coded bits based on picture type
- controlling quantiser step size proportional to buffer occupancy
- controlling quantiser step size inversely proportional to the amount of data in pictures
- discarding high frequency DCT (discrete cosine transform) coefficients when overflow occurs

However, the actual buffering algorithm depends on the designer as there are no quantitative specifications in the MPEG standard. Figure 1 shows the schematic configuration for buffering in MPEG

Quantisation scale (Qscale) should reflect the current occupancy. However, the details of mapping between occupancy and Qscale are design decisions. Linear mapping is usually used. The linear mapping works properly in normal operation without excessive incoming traffic. However, in the case of overflow or a dramatic increase in occupancy, it is not able to track such change. In addition, the current quantisation scale is derived from the occupancy which results from a cumulative sum of previous coded bits. Any excessive change in occupancy can only be monitored after such an event has occurred since the mapping process itself cannot identify it in advance. An intelligent buffering algorithm should better map the buffer occupancy to the quantisation step size, which is the major parameter controlling the rate of the output bitstream, also, it should be able to predict the excessive change.

2 Features of Buffer Occupancy
The quantisation scale is adaptively controlled mainly by buffer occupancy and picture details. If an input
video sequence has large amounts of detail within frames or significant motion between frames, then the occupancy increases. An another important factor is picture coding type. While the CCITT H.201 for videotelephony has no specific order of predefined picture types, MPEG has 3 different picture types i.e. I, P, and B pictures. All macro blocks of an I picture are coded with intra mode, and those of P and B pictures are coded with either intra or motion-compensated inter mode. The difference between P and B pictures is that while the P picture is motion compensated in the forward direction, the B picture is motion compensated in both forward and backward directions. Since the picture type repeats such as BBBBBBBBBBBBBBBBBB... or PBBBBBBBBBBBBBBBBBBBBBB... the buffer occupancy fluctuates more in MPEG than H.201 for constant bit rate transmission. In variable bit rate transmission a quantisation step size is fixed for a picture type, Pancha and Zarki (6). Also, this picture type configuration may have influence on abnormal buffer operations i.e. overflow or abrupt occupancy transitions. Figure 2 shows the buffer occupancy of an encoder when encoding 100 video frames of a rapidly moving sports video "football". It is shown at six separate transmission rates with a fixed buffer size of 327680 bits which is the maximum size for sub-primary rate communication. As each picture consists of 15 slices the graphs are plotted over 1500 slices for better illustration.

![Graph showing buffer occupancy vs. occupancy rate for different transmission rates](image)

Figure 2: An example of Buffer occupancy in an MPEG video encoder

The "football" video sequence has a large amount of motion throughout the sequence but the background is quite steady. Thus the occupancy graphs in Figure 2 have I, B, P spiky changes, and the buffer overflows at the two lowest bit rates. At every I picture, as indicated in Figure 2, a relatively large change in occupancy occurs. There are slightly smaller peaks at P pictures since coding P pictures generates more data than coding B pictures. As most of the macro blocks of a B picture are coded with bidirectional motion compensation, the amount of prediction error data to be transmitted is smaller, that is, troughs appear at B pictures. There can be overflows at higher bit rates also, if the video sequence has more dramatic some changes.

## 3 Proposed Buffering and Rate Control Scheme

For more efficient buffering and rate control our target is to:
- smooth out the fluctuation of the buffer occupancy,
- avoid overflow without objectionable degradation
- achieve better subjective and objective video quality

Figure 3 shows the configuration of the proposed buffering and rate control algorithm.

![Block diagram of the proposed algorithm](image)

The proposed algorithm is composed of 3 major functional blocks which are combined with the encoder system. A video encoder, variable length coder and transmission buffer are the major functional blocks of the entire encoder. After video is compressed by the video encoder, which consists of motion compensation, DCT transform, Clarke (7), and quantiser, Jayant and Noll (8), the video multiplexer (video MUX) encodes the incoming data to variable length codes and multiplexes this with header data. The output of the video MUX transfers to the transmission buffer, regulating the bit rate. In the proposed system the video encoder quantises the DCT coefficients with both variable and fixed steps. The DCTF and the DCT represent the DCT coefficients quantised by the fixed and the variable quantisation step sizes, respectively. While DCTF is encoded by the video MUX for actual transmission, DCT is used for monitoring the amount of data generated by the video encoder only. The traffic monitor calculates the total bits of data per frame, and the information is sent to the predictor. The predictor collects B(n) and O(n-1) for predicting the value of O(n) which determines the value of quantisation step size QVS. Quantisation step size is regulated by two parameters i.e. quantisation scale and quantisation matrices.

\[
DCTV(u,v) = int\left( \frac{8 \times DCT(u,v)}{QVS(u,v)} \right)
\]

where \( u \) and \( v \) are integers ranging from 0 to 7, representing \( x \) and \( y \) coordinates in the DCT domain.
DCT\((u,v)\) is the output of the DCT transform and DCTV represents the quantised version of the DCT. Quantisation step size \(QS(u,v)\) is a product of the quantisation scale and the matrixes i.e. \(QS(u,v) = Qo(i,j) \times \text{sqm}(i,j)\) or \(QS(u,v) = Qo(i,j) \times \text{nqm}(i,j)\). \(Qo(i,j)\) is the quantisation scale ranging from 1 to 31, which reflects the buffer occupancy, and \text{sqm}(i,j)\) and \text{nqm}(i,j)\) are the intra quantisation matrix and the non-intra quantisation matrix, respectively. The \text{sqm}(i,j)\) covers macro blocks coded with intra mode, and \text{nqm}(i,j)\) is used for those coded with inter mode. \(i\) and \(j\) indicate the \(x\) and \(y\) coordinates of a block. As the intra-mode macro blocks and the inter-mode macro blocks have different amounts of information, it is necessary to apply a more optimised quantisation matrix to each type of macro block. For applications with 1850 kbit/s or lower bit rates, MPEG restricts the major parameters to specific values. In this case quantisation matrices are set to defaults. Thus the video traffic is controlled solely by the \(Qo(i,j)\) which may be updated at every macro block interval. Hence, the value of \(QS\) is constant as \(Qo(i,j)\) has a fixed value. The traffic monitor calculates the amount of bits by applying the \(QoS\) to the DCT coefficients. The three functional blocks operate to generate better \(Qo(i,j)\) values which are robust against dramatic changes in occupancy. A higher order autoregressive (AR) model is adopted for the predictor and set of logarithmic equations are used for allocating quantisation scale values as a nonlinear mapping.

4 Nonlinear Quantisation Scale Mapping

Linear mapping of \(Qo(i,j)\) works satisfactorily under normal operating condition in a large buffer with the occupancy fluctuating around 50%. However, in cases of lower bit rates or abrupt scene changes, it simply increases the occupancy. As shown in Figure 2, at the lowest two bit rates the buffer cannot cope with the incoming traffic. At 768 kbit/s the buffer does not overflow for the first 100 video frames, however, if the input video has more information than the previous frames, the buffer may overflow at any future point. While the output bit rate of the transmission buffer is constant, the input rate is highly variable. If the input bit rate is equal to the output rate, then the occupancy does not change and keeps a specific, fixed value. However, the occupancy changes all the time since the input video data and/or the picture type vary. Therefore, if the change in input bit rate is constant and greater than the output rate, then the occupancy increases linearly. On the other hand, if it is not constant e.g. increasing, then the occupancy increases nonlinearly e.g. exponentially. For example, the four plots of higher bit rates in Figure 2 show a nonlinear increase in occupancy around the slice number 400. All plots in Figure 2 display the results of linear mapping by the diagonal straight in Figure 4.

Nonlinear mapping is an adaptive way to allocate a quantisation scale value to compensate for the dramatic changes in occupancy. In this study logarithmic mapping is used as shown in Figure 4. A value of quantisation scale is selected based on the rate of occupancy change and the current occupancy for \(\Delta O > 0\) i.e.

\[ Qo(\Delta O, O) = a(\Delta O)Qo(\Delta O, O + 1) \]

where \(O\) is the current occupancy, \(\Delta O\) is the differential occupancy i.e. the change in occupancy, and \(a(\Delta O)\) and \(\Delta(\Delta O)\) are the coefficients to form logarithmic plots. Thus, the quantisation scale \(Qo(\Delta O, O)\) is determined according to the current and the differential occupancies. If the \(\Delta O\) value is bigger, a more nonlinear curve is selected, and the plot maps a current occupancy to a more appropriate \(QoS\) value. Otherwise, if \(\Delta O\) is smaller, the plot selected gets closer to the straight line or the straight line itself. This adaptive scheme provides greater capability to cope with abnormal buffer operation.

5 Autoregressive modeling

The buffer occupancy fluctuates with the sum of the bits per frame. The number of coded bits depends on the picture type i.e. I, P, and B. As the picture type repeats regularly e.g. "BBBBBBBBBBBB", the number changes periodically as shown in Figure 5. As this behaviour is virtually the same irrespective of the bit rate, the characteristic of the occupancy change is suited to an autoregressive (AR) model, Press et al (9), as used in speech waveform prediction, Jayant and Noll (8). Current buffer occupancy, \(O(n)\) is the sum of the previous occupancy, \(O(n - 1)\) and the current number of coded bits, \(B(n)\) i.e.,

\[ O(n) = O(n - 1) + B(n) \]

(3)

\[ B(n) = \sum_{j=1}^{N} d_j B(n-j) + x_n \]

(4)

\(B(n)\) can be predicted with a set of previous \(B(n-j)\)'s and a correction term \(x_n\) where \(n\) is picture number,
Appendix D: Publications

\[ N \] represents the order of AR model, and \( d_j \) are the coefficients of an AR model. The order of the AR model has a close relationship with the number of B pictures between two P pictures. In a typical case the number of B pictures is 2, and that of P pictures between I pictures is 3, and there are 12 pictures in a pattern of picture types. This means that \( B(n) \) has a close correlation with multiple values of previous \( B(n-j) \) rather than a single value of \( B(n-j) \), that is, a higher order AR model is more appropriate for predicting \( B(n) \), and, finally, \( O(n) \).

6 Simulation results

A software MPEG video encoder has been implemented which has customisable buffering and rate control functions in order to evaluate various aspects of MPEG video encoding performance. The encoder has been rigorously verified to obtain full compatibility with the ISO/IEC11172 standard. The encoder accepts 352 pixels by 240 lines color video and generates a compressed bit stream file. The compressed file can be decoded by any regular MPEG decoder and contains video information which displays the video quality at a specific bit rate. All compressed files have been satisfactorily decoded by the Berkeley's decoder, Fogg (19), which is one of the most widely used decoders available on the Internet. Our own decoder has also been developed and tested. This decodes a bit stream file and generates various statistics files and its performance has been shown to be identical to the Berkeley decoder.

Two different types of video sequences have been taken as input to the encoder. As the "football" sequence has a lot of motion and dramatic scene changes between adjacent video frames, it is a typical input with a large amount of data. "Susie" is a normal head-and-shoulder video sequence which has still background and slight motion. However, among the 150 video frames there is 30-frames worth of fast motion in the middle of the sequence. Figure 6 shows two representative frames of "football" and "Susie".

Figure 7 displays the improvement of the proposed adaptive quantisation scheme. The peak occupancy is reduced by about 20%.

The fall in PSNR at the transient state in Figure 7 (b) is caused by temporary oscillation of the quantisation scale. Since the initial state of the buffer is empty, the occupancy increases sharply, and a large \( Qo(\Delta O, O) \) is selected, consequently, the PSNR falls. At the next frame, as the occupancy change is relatively small, \( Qo(\Delta O, O) \) becomes smaller, and the PSNR recovers. For further video frames, the occupancy increases but not as sharply as at the beginning, thus, the difference between the two PSNR's gradually reduces. To obtain a reduction in occupancy means losing some video quality around the video frames where the occupancy control works. The degradation in PSNR is approximately 3 dB. However, this has no effect on the video frames coming afterwards, when both PSNR plots maintain the same level but the occupancy is lower, as avoiding overflow.

The result for "Susie" is shown in Figure 8. Since "Susie" has smaller amount of data, a lower bit rate is simulated, which ensures buffer overflow. While "football" shows an exponential increase in occupancy, the occupancy of "Susie" increases linearly as the motion is smaller than "football" on the whole. In this case the proposed scheme selects \( Qo(\Delta O, O) \) values which are close to the values on the straight line, thus, the degradation in PSNR is not so dramatic and spreads out up to frame number 30 to 40. The maximum PSNR loss is approximately 2 dB. The buffer overflow which
7 Conclusion

In this paper important statistics and features have been investigated which are closely related to buffering and rate control for constant bit rate communication of an MPEG video encoder. An intelligent algorithm for buffering and rate control has been proposed and verified by software encoders and decoders. The algorithm consists of a nonlinear mapping of quantisation scale and a higher-order AR model predictor. Since the scheme regulates the incoming video traffic to the transmission buffer the encoder is able to better control abnormal buffer operation such as overflow. This approach maintains more stable control of its buffer occupancy. Since the scheme does not rely on expanding the size of the buffer, the coding delay of the encoder is unchanged which is a critical parameter in bidirectional communication. Also, this reduced storage buffer design does not need to increase the buffer size and provides a saving in buffer storage itself. Thus, this scheme improves the performance of buffering and rate control mechanism employed in the MPEG video encoder, for only minimal picture performance degradation. Further work on the motion adaptive parameters and the predictor will be carried out to improve the rate control.

8 Acknowledgement

This research is sponsored by the British Council and GoldStar Information and Communications Ltd.

9 References

FEED-FORWARD BUFFERING AND RATE CONTROL BASED ON SCENE CHANGE FEATURES FOR MPEG VIDEO CODER

Yoo-Sok Saw, Peter M. Grant, and John M. Hannah
Dept of Electrical Engineering, University of Edinburgh, Edinburgh, EH9 3JL, UK.
Tel: +44 131 6505655; fax: +44 131 650 6554
e-mail: ys@ee.ed.ac.uk

ABSTRACT
Video traffic management has been a challenging task in the fields of network management and multi-media communication. Transmission buffering is widely used to smooth bursty traffic and maintain a steady traffic level by adapting the incoming source traffic to the buffer. This paper describes an efficient adaptive buffering scheme which is based on feed-forward control to adaptively handle the non-stationary nature of bursty video traffic. The performance of a series of quantisation scale mapping curves is presented in terms of occupancy and video quality.

1 INTRODUCTION
A digital video compression algorithm such as MPEG generates a coded bit stream with variable rate. The term “traffic” is used to indicate the time-varying non-stationary nature of the video bit stream, since the video can alter from still to moving pictures.

In applications using transmission buffer, overflow and congestion may take place in bursty traffic or a low bit-rate CBR (constant bit rate) channel. While much research work on traffic modeling for ATM (Asynchronous Transfer Mode) networks has been investigated [1], less research work has been carried out on buffering and rate control [2, 3, 4, 5, 6] in the video encoder. Thus, we concentrate on how effectively the buffer occupancy can be controlled within a given channel rate. We introduce features representing the amount of scene change which are used for calculating predictive buffer occupancy. ISO/IEC 11172 (MPEG1) and 13818 (MPEG2) are representative video compression standards to which the buffering scheme can ultimately be applied.

2 CONFIGURATION OF BUFFERING IN MPEG ENCODER
In an MPEG video encoder, quantisation step size can be adaptively controlled by buffer occupancy and picture details [7, 8]. When a buffer is used for traffic control, it is vital to use a reactive mechanism which controls the quantisation step size by feedback information on buffer occupancy. Exploitation of picture details can also be applied to control the picture quality. Occupancy-based reactive control becomes a dominant function, since maintaining a constant occupancy is the main objective. Figure 1 shows a diagram of the reactive buffering scheme.

3 MAPPING QUANTISATION STEP SIZE
The target in rate control is to effectively map the buffer occupancy to the quantisation step size. Several different mapping curves have been investigated. They can be classified into linear [9, 10], piecewise linear [8, 3], and non-linear [11, 5, 6]. Two non-linear mapping curves, sigmoidal [11] and logarithmic [6] have been investigated.

Figure 1: Reactive buffering scheme specified in MPEG

Figure 2: Sigmoidal quantisation scale mapping
The reference simulation software [9, 8] uses linear mapping i.e., \( Qs(O) = O \) where \( Qs(O) \) and \( O \) are the...
normalised quantisation scale and occupancy. In sigmoidal mapping, two nonlinear equations form a nonlinear mapping curve with the shape of a skewed S. Κ is a steepness factor and α is a control factor to determine the symmetry of each curve, e.g., if α = 0.5, the curve shows a symmetrical shape whose upper half and lower half form images. The combination of two curves forms a skewed sigmoidal mapping curve, Figure 2.

In logarithmic / exponential mapping the quantisation scale maps to a set of logarithmic and exponential curves. A value of quantisation scale is selected based on mean occupancy \( O^- \) for a specific period of time and the current occupancy, i.e.,

\[
Q_s(O^-, O) = \alpha(O^-) \log_a(\beta(O^-)O + 1)
\]

where \( O^+ \) is the current occupancy, \( O^- \) is the mean occupancy, and \( \alpha(O^-), \beta(O^-), \rho(O^-), \) and \( \mu(O^-) \) are the coefficients which form the logarithmic and exponential curves. If the \( O^- \) value does not remain in a predefined occupancy range, a more nonlinear curve is selected. Otherwise, if \( O^- \) stays in the range, the curve selected becomes close to a straight line, Figure 3.

\[
Q_s(O^+, O) = \rho(O^-) \exp(\mu(O^-)O) - \mu(O^-)
\]

Figure 3: Logarithmic quantisation scale mapping

4 SCENE CHANGE FEATURES

This buffering scheme takes advantage of the frame delays for re-ordering picture types since MPEG encodes a P or I picture first, instead of having B pictures preceding them. Hence, if a scene change occurs in a B picture, the encoder realises this in advance of encoding the P or I picture. This can effectively be used for advance adjustment of quantisation scale. Figure 4 shows a detailed block diagram of feed-forward buffering combined with the MPEG main profile encoder. NQS (non-linear quantisation scale mapping) selects a non-linear curve to estimate future occupancy. The SCF (scene change function) and MVF (motion vector function) calculate the ratio of the variance of a difference picture frame to the variance of an input frame and the mean value of the motion vector function in a slice, respectively. Output values of NQS, SCF, and MVF go to the Q scale control block to determine the best quantisation scale value for a slice. No quantisation scale adaptation is performed for a macro block or smaller scale of block since we assume that macro blocks have high correlation.

5 PREDICTIVE BUFFER OCCUPANCY

Frame-wise variances form an appropriate measure for scene change. It is an extension of the intra / inter decision for a macro-block specified in ISO11172-2 [7]. Two variances, \( \text{var}_\text{org} \) and \( \text{var}_\text{diff} \) are defined. \( \text{var}_\text{org} \) is the variance of input picture and \( \text{var}_\text{diff} \) is the variance of the difference picture between current input picture and previous input picture. Using these two variances, the graph is divided into four regions, Figure 5. “A” represents the area with no dramatic scene changes and no subsequent abrupt changes in traffic since the \( \text{var}_\text{diff} \) is small. “B” is the area with higher \( \text{var}_\text{diff} \), however, there would be no dramatic traffic level change at this picture since \( \text{var}_\text{org} \) is smaller than \( \text{var}_\text{diff} \). The DCT encodes macro blocks in intra mode and its performance is considered to be better than encoding them in inter mode. Areas “C” and “D” may cause more dramatic increases in traffic. “C” and “D” are separated by the partition \( \text{var}_\text{diff} = \text{var}_\text{org} \). Estimating the required number of coded bits (\( C_{\text{bits}}^p(k) \) and \( C_{\text{bits}}^i(k) \)) of kth input frame is thus based on \( \text{var}_\text{diff}, \text{var}_\text{org}, \) and previous statistics of coded bits per frame, i.e.,

\[
C_{\text{bits}}^p(k) = \frac{\text{var}_\text{org} \times C_{\text{bits}}(P(k-n))}{\text{var}_\text{diff}}
\]

\[
C_{\text{bits}}^i(k) = \frac{\text{var}_\text{diff} \times C_{\text{bits}}(I(k-n))}{\text{var}_\text{org}}
\]

where \( C_{\text{bits}}(P(k-n)) \) and \( C_{\text{bits}}(I(k-n)) \) are the number of bits of the previously encoded I or P picture.

Figure 5: Plot for frame-wise inter / intra decision

As the pictures in both “C” and “D” may generate bursty traffic, it is necessary to choose either \( C_{\text{bits}}^p(k) \) or \( C_{\text{bits}}^i(k) \) when the value of \( \text{var}_\text{diff} / \text{var}_\text{org} \) is greater...
than 1. Once \( C_{\text{bit}}(k) \) is determined, current buffer occupancy \( O(k) \) is checked, i.e.,
\[
O(k) = \text{int} \left( \frac{D_T \times F R^{-1} - O_{\text{max}}(k)}{F R^{-1}} \right),
\]
where the term \( D_T \times F R^{-1} \) represents the maximum delay in ms and \( O_{\text{max}}(k) \) stands for current occupancy in ms. Therefore, \( O(k) \) represents current buffer capacity in the number of frames. \( D_T \) is the delay target representing the maximum tolerance of coding delay in the number of frames, and \( F R^{-1} \) is the reciprocal of frame rate. If \( O(k) > 1 \), i.e., if the remaining capacity of the buffer is able to accept bits of a complete frame, the following equation applies with each \( C_{\text{bit}}^+(k) \) or \( C_{\text{bit}}^-(k) \).
\[
O_p(k) = \frac{O(k) + C_{\text{bit}}^+(k)}{D_T \times \text{MBF}}
\]
\( O_p(k) \) represents the predictive buffer occupancy and is used for selecting quantisation scale for the \( k \)th frame, \( QSF(k) \), function,
\[
QSF(k) = \phi(T_b, O_p)
\]
\( QSF(k) \) is a function of the predictive occupancy \( (O_p(k)) \) and the traffic balance \( (T_b) \) defined as follows.
\[
T_b = \frac{C_{\text{bit}}(\Delta k)}{L \times \text{MBF}}
\]
where \( L \) is the number of frames in a set of pictures, e.g., for BBP or BBI \( L \) becomes 3, and MBF is the mean bits per frame. The function \( \phi \) is one of the logarithmic / exponential mapping curves described in the previous section. Thus, the quantisation scale for a whole input frame is determined by short-term traffic history and future occupancy.

6 LOCAL ADAPTATION OF THE QUANTISATION SCALE

Parts of a picture may have different amount of visual information from the other parts of the picture. The value of \( QSF(k) \) can be adaptively changed at each picture slice or horizontal stripe. In order to adapt \( QSF(k) \) to the quantisation scale for each slice, motion vector values are used. The motion vector function, \( MVFD \) is defined as follows:
\[
MVFD(s) = \frac{\sum_{i=0}^{MBN-1} \text{mvx}(k) + \sum_{i=0}^{MBN-1} \text{mvy}(k)}{MBN \times (\text{MVX}_{\text{max}} + \text{MVY}_{\text{max}})}
\]
where \( s \) is slice number and \( MBN \) is the number of macro blocks in a slice, hence, \( \text{mvx}(k) \) and \( \text{mvy}(k) \) become motion vectors of macro block \( k \). \( \text{MVX}_{\text{max}} \) and \( \text{MVY}_{\text{max}} \) are the maximum values of motion vectors for a specified motion search range. \( MVFD(s) \) is a directional motion vector function. As \( MVFD(s) \) ranges from 0 to 1, in order to make \( MVFD(s) \) a scaling factor of \( QSF(k) \) it is multiplied by 2, and applied to \( QSF(k) \), i.e.,
\[
QSS(s) = 2 \times MVFD(s) \times QSF(k)
\]
where \( QSS(s) \) is the quantisation scale for the slice \( s \).

7 SIMULATION RESULTS

Six buffering schemes have been simulated over three video sequences “MAFSU”, “Star Wars”, and “Advertisements”. “MAFSU” is a cascaded video sequence of “Miss America”, “Football”, and “ иметь”. “Star Wars” and “Advertisements” are taken from the movie “Star Wars” and television advertisements, respectively. The buffering schemes can be classified into adaptive and non-adaptive. The adaptive scheme has been applied to sigmoidal and logarithmic / exponential mapping curves, which adaptively change the steepness of the curves according to traffic balance. The non-adaptive method uses a single curve. The linear mapping scheme (LIN) takes a straight line as a mapping curve, as a performance benchmark. In the logarithmic mapping (LOG) and the exponential mapping (EXP), the curves A and B of Figure 3 are used, respectively. Sigmoidal mapping (SGM) uses the curve R of Figure 2. Adaptive logarithmic / exponential mapping (LOGEXP-A) forms a combined scheme composed of LOG and EXP with
adapation. Adaptive SIGM (SIGM-A) is an adaptive scheme of SIGM.

Figure 6 shows the simulation results of the six methods both in terms of occupancy and peak SNR for these video sequences. In (a) there is a major scene change at slice 1300. The performance of LOG and LIN are lower than SIGM. EXP shows medium performance between SIGM and LOG. However, since it keeps lower quantisation scale values at low occupancy than SIGM and LOG, it generates more coded bits at low and initial occupancy. SIGM-A and LOGEXP-A show better performance since they adaptively change the mapping curves depending on the previous occupancy. SIGM-A shows steadier fluctuation in occupancy, however, at dramatic changes LOGEXP-A performs better since it can more quickly respond to the occupancy increase. Quality degradation, by controlling the quantisation scale, is seen in Figure 6 to be inversely proportional to the occupancy. The larger quantisation step sizes clearly lower the picture quality from the encoder, as shown in the PSNR plots.

8 CONCLUSION
In this paper we have investigated a feed-forward buffering scheme for video traffic using scene change features. Frame-wise variances and motion vectors in a slice have been effectively exploited to provide some change information to control quantisation scale value. The performance of several different quantisation mapping curves (linear, sigmoidal, logarithmic, and exponential) has been evaluated. The adaptive scheme using combined logarithmic/exponential curves shows superior performance in terms of buffer occupancy.

9 ACKNOWLEDGEMENTS
Yoo-Sok Saw acknowledges the support of The British Council for this research.

References
A radial basis function video rate estimator for constant bit rate MPEG coders

Yoo-Sok Saw, Peter M. Grant, Fellow, IEE, John M. Hannah, and Bernard Mulgrew

Abstract—In this letter a radial basis function (RBF) network has been applied to constant bit rate control for an MPEG2 video encoder. The non-stationary property of the video data has been exploited effectively by using the RBF network as a rate estimator in a feed-forward rate control algorithm. The performance of this scheme is evaluated in comparison with Test Model 5 (TMD), by measuring video rate and picture quality.

Indexing terms
MPEG, video rate estimator, radial basis function network.

I. VIDEO RATE CONTROL BASED ON A RBF-NETWORK ESTIMATOR

The rate control mechanism of an MPEG2 video encoder controls its output video rate by maintaining buffer occupancy within a specified delay constraint for constant bit rate channel applications. The most widely used method to control the occupancy is to change the quantisation step size depending on instantaneous video rate [1]. This scheme is a feedback control which uses the video rate from the previous frame as a prediction of the current frame's rate.

In this letter, a feed forward control technique, based on a RBF network, has been incorporated so that the quantisation step size can be adaptively increased in order to enclose incoming video. The RBF non-linear network [2] functions as a rate estimator to inform the quantiser which quantisation step size should be used for a particular picture.

Fig. 1 shows the modified MPEG video rate control structure. The control scheme comprises three main functions: the RBF rate estimator, some change calculator and non-linear quantiser control. The RBF takes three inputs from the change calculator which outputs two variances (var_org(k) and var_dif(k)) and the picture type information (ptype(k)), to represent variations in visual information. The two variances are derived from a single input picture and its difference picture from the previous picture, respectively. ptype(k) is a cyclic signal containing constants for specific picture types. The predicted video rate, coded bits per frame cf(k) is added to the current occupancy (O(k - 1)) to form the predicted occupancy (O(k)) used by the non-linear quantisation control which finally outputs the quantisation scale value (Q(k)). The error signal (e(k)) is fed back to the RBF in order to adaptively change its linear weights. MBF stands for mean value of bits per frame, the mean video rate.

Fig. 1. Video rate estimator combined with MPEG video encoder

The RBF network, Fig. 2, consists of centres with a radial basis function and linear weights defined as:

\[
\text{cf}(k) = \sum_{i=1}^{N} w_i \phi(||x - \mu_i||)
\]

(1)

\[
\phi(||x - \mu_i||) = \exp\left(-\frac{||x - \mu_i||^2}{2\sigma^2}\right)
\]

where \text{cf}(k) is the output of the RBF network, \(w_i\) is the linear weight, \(x\) is an input vector containing some change features, and \(\mu_i\) represents the \(i\)th selected centre among \(N\) centres in total. The radial function \(\phi()\) is Gaussian. The Euclidean distance \(||x - \mu_i||\) between the input and the centre determines the output value of the RBF layer. \(\sigma^2\) represents the variance of \(x\).

Fig. 2. RBF estimator with 3 inputs with 3 taps each

For computational efficiency, the RBF centres are usually selected by an orthogonal least square (OLS) algorithm

---

Yoo-Sok Saw is with Signal and Systems Group, Electrical Engineering, University of Edinburgh, Edinburgh EH9 3JL, Scotland, UK. E-mail: yss@ee.ed.ac.uk. Phone: 031 - 650 - 563, Fax: 031 - 660 - 6554

---

Appendix D: Publications
II. Simulated encoder results

The movie sequence, “JFK”, was used to provide non-stationary video test data. This is a 330-frame sequence with rapid motion and dramatic scene changes. The video encoder was set to operate with a channel rate at 1024 kbits/s and a frame rate at 30 frames/s. The MBF is thus 34133 bits/s and the buffer size is 64206 bits. The performance of the RBF presented here is based on a 9-node network. The learning rate of the k-means clustering algorithm [5] is set to 0.4.

Three performance measures were evaluated: the number of coded bits per frame \( cbf(k) \), frame-wise buffer occupancy and peak signal-to-noise ratio (PSNR), as shown in Fig. 3. Table I shows the mean and standard deviation (std.dev.) of these performance measures.

All bit streams have the same MBF values since the transmission rate is fixed. The normalised fluctuation of the video rate (NFVR) is represented by variations in \( cbf(k) \). It is expressed by the following equation:

\[
NFVR = \frac{\sigma}{1 + \sigma}
\]

\[
\sigma^2 = E \left( \frac{cbf(k)}{MBF} - 1 \right)^2
\]

where \( \frac{cbf(k)}{MBF} \) represents the instantaneous fluctuation. In Fig. 3, TM5 exhibits far more fluctuations in \( cbf(k) \) and occupancy than the RBF-based rate control algorithm. TM5 overflows the buffer since it allocates bits based on picture types and previous picture complexity which are suitable only for encoding stationary video data. Subsequently, the PSNR profile shows considerable variation. Table I shows that the RBF controls the video rate with far smaller variation, without seriously degrading the video quality (PSNR). The std.dev. of PSNR shows that the RBF causes a slightly greater variation than TM5. Thus, it is confirmed that the RBF approach has better performance in both local and global scales in terms of rate control measures. The RBF-based scheme has also been shown to perform equally well on other test video sequences with dramatic scene changes.

III. Conclusion

Video rate control using a RBF rate estimator has been investigated for constant bit rate control of MPEG2 video data. The RBF network non-linear estimator has been combined with scene change features to implement a non-linear quantiser control technique. The RBF shows superior performance in achieving a more constant video rate and occupancy without noticeable quality degradation, in comparison with TM5.

References

NONLINEAR PREDICTIVE RATE CONTROL FOR CONSTANT BIT RATE MPEG VIDEO CODERS

Yoo-Sok Saw, Peter M. Grant, John M. Hamman and Bernard Malgras
Dept of Electrical Engineering, University of Edinburgh, Edinburgh, EH9 3JL, UK.
Tel: +44 131 6505655; fax: +44 131 650 6554
e-mail: ysoo@ed.ac.uk

ABSTRACT

A nonlinear predictive approach has been employed in MPEG (Moving Picture Experts Group) video transmission in order to improve the rate control performance of the video encoder. A nonlinear prediction and quantisation technique has been applied to the video rate control which employs a transmission buffer for constant bit rate video transmission. A radial basis function (RBF) network has been adopted as a video rate estimator to predict the rate value of a picture in advance of encoding. The quantiser control surfaces are based on nonlinear equations which map both estimated and current buffer occupancies to a suitable quantisation step size, hence to achieve higher responses to dynamic video rate variation. This scheme aims to accurately accommodate non-stationary video in the limited capacity of the buffer. Performance has been evaluated in comparison to the MPEG2 Test Model 5 (TM5) in terms of the buffer occupancy and picture quality.

1. INTRODUCTION

The rate control algorithm of TM5 is based on the previous history of video rate, global and local picture complexity measures. This technique is known to be inappropriate for non-stationary videos with frequent scene changes or rapid motion [1], since the statistical properties are changing accordingly. Therefore, for such video, a different approach is required. Recently, we have developed a feed-forward video rate control technique using scene change features [2, 3]. The main advantage of the technique is the rate estimation of the video rate value is derived from a series of scene change features. The employed prediction technique is based on a one-step ahead linear prediction using previous video rate data in a heuristic way. This paper a nonlinear estimation technique is applied in order to more effectively control dynamic scene changes. The RBF-network [4] was designed to estimate the video rate using the scene change features of the input video so that the quantisation step size can be adjusted in advance of encoding the picture. The scene change features are frame-wise variances and picture type information. The nonlinear quantiser control surface changes the quantisation step size depending on the estimated video rate and the current buffer occupancy. Three performance evaluation measures were used; number of coded bits per frame, buffer occupancy and peak signal-to-noise ratio (PSNR).

2. A RBF-NETWORK RATE ESTIMATOR

Before describing the detail of the RBF-network estimator, an example of performance of the linear predictive video rate control techniques described in [2] is shown in Figure 1.

![Figure 1. A performance example of linear predictive rate control techniques (a) buffer occupancy, (b) Peak SNR](image_url)

TM5 represents the video rate control technique employed in the MPEG2 TM5. The other three methods (LIN, SIGM and LOGEXP) are based on the same linear predictive method [2] but a different quantisation control function is applied to each method. A linear function and a sigmoidal function are used for the methods, LIN and SIGM, respectively. For LOGEXP, a combination of logarithmic and exponential functions is employed, which is collectively named “animaloid” in later sections of this paper, instead of the linear or sigmoidal function. The results shown in Figure 1 are obtained from the MPEG2 encoding of “FIR” motion sequence at the 128 kbit/s channel rate and the 30 frames/s frame rate. TM5 shows the worst performance of less reducing the buffer full state, also showing wider variations in the occupancy and the PSNR alike. Particularly, LOGEXP shows the most stable occupancy profile with the very similar quality to LIN and SIGM.

The RBF-network video rate estimator aims to further improve the performance by applying its nonlinear predic-
tive properties to non-stationary signals. The innovated MPEG2 encoder contains three additional rate control functions as shown in Figure 2: the scene change calculator, the rate estimator and the non-linear quantiser control. The scene change calculator outputs the two variances, $\text{var}_1$ and $\text{var}_2$, and the picture type information, $\text{ptype}(k)$, as inputs for the rate estimator. The predicted video rate, $\hat{\text{r}_v}(k)$, is added to the current occupancy, $O(k-1, n)$, to form the predicted occupancy, $\hat{O}(k, n)$, used by the non-linear quantisation control, which finally outputs the quantisation scale value, $Q(k, n)$. $\text{var}_1$, $\text{var}_2$, $\text{r}_v(k)$ and $\text{ptype}(k)$ represent the variance within an input picture and the variance between the input picture and the previous picture, respectively. $\text{ptype}(k)$ has a single integer for a particular picture type (I, P and B), thus it forms a curve as $k$ increases such as 0, 4, 2, 2, 4, 2, ..., for I,P,B,P,B,...

![RBF network](image)

**Figure 2.** A RBF rate estimator-based MPEG2 video encoder.

A RBF network consists of centres with a radial basis function and linear weights, Figure 3, defined in the following equations:

$$\hat{\text{r}_v}(x) = \sum_{i=1}^{N} w_i \phi(||x - x_i||)$$

$$\phi(||x - x_i||) = \exp\left(-\frac{||x - x_i||^2}{2\sigma^2}\right)$$

where $\hat{\text{r}_v}(x)$ is the output of the RBF network, $w_i$ is the linear weight, $x$ is an input vector containing scene change features, and $x_i$ represents the selected centre. The radial function $\phi$ is a Gaussian function. The Euclidean distance between the input and the centre $||x - x_i||$ determines the output value of the RBF layer, $\sigma$ represents the variance of $x$. The RBF centres may be selected by the orthogonal least square (OLS) algorithm [5]. The OLS algorithm selects representative RBF centres when supervised learning is used. However, in the case of the running MPEG2 encoder, supervised learning cannot be used properly since the nature of realistic video is non-stationary. A supervision of some change features for a short period of time does not provide an entire insight into the whole properties of the non-stationary video. Thus the $k$-means clustering algorithm is used, which adaptively updates the RBF centres depending on variations in scene change features. The centres are updated as follows [6]:

$$x_j(k) = x_j(k-1) + g_c(\hat{\text{r}_v}(k) - x_j(k-1))$$

where $x_j$ is the $j$th centre and the constant $g_c$ controls the learning rate. The linear weights, $w_i$, are optimised recursively in a least square sense (RLS) [7].

3. Quantisation Control Surfaces Based on Non-linear Equations

The quantisation step size is the core parameter which controls the occupancy. The goal of the buffer-based rate control technique is to effectively map the occupancy to the quantisation step size specified in the MPEG2 standard. Several different control functions have been proposed. They can be classified into linear, piecewise linear and non-linear functions [8, 2]. This paper focuses on the non-linear control functions. The non-linear quantiser control, as shown in Figure 4, uses both the current $(\hat{O})$ and the predicted buffer occupancies (n) to select a quantiser control curve for the quantisation scale ($Q$). It changes between linear and non-linear curves depending on the predicted occupancy. If a dramatic change is the occupancy is predicted, then it changes the shape of the curve towards a more distorted one, otherwise, it selects a curve close to the linear function. The final quantisation scale value is determined by the current occupancy. In this paper two non-linear mapping surfaces are examined, sigmoidal and unimodal as shown in Figure 5.

![Non-linear quantisation step size mapping](image)

**Figure 4.** Non-linear quantisation step size mapping.

The sigmoidal surface (SIGM) is formed by changing the steepness of a sigmoidal function. The unimodal surface (UNIM) consists of a combination of an exponential part and a logarithmic part. The quantisation scale for the macro block $x$, $Q(k, n)$, is calculated as follows:

$$Q(k, n) = f(O(k-1, n), \hat{O}(k))$$

$$\hat{O}(k) = O(k-1, n) + \hat{r}_v(k) - \text{MBF}$$

where MBF is the target video rate given by the mean value of bits per picture. $f()$ is one of the non-linear mapping surfaces, and a value of $Q(k)$ for the next macro block is determined for given $O(k-1)$ and $\hat{O}(k)$. The two surfaces, shown in Figure 5, are expressed in equations of $f_{SIGM}(O(k-1, n), \hat{O}(k))$ and $f_{UNIM}(O(k-1, n), \hat{O}(k))$, which represent surfaces of SIGM and UNIM, respectively:

$$f_{SIGM}(x) = \alpha \left( \frac{1}{1 + \alpha} O(k-1) \right)^{1/\alpha - 1}$$

$$f_{UNIM}(x) = \alpha \left( 1 + \alpha - O(k-1) \right) \left( 1 - \alpha \right) \left( \frac{1}{1 - \alpha} - 1 \right)^{1/\alpha - 1}$$

194
\[ f_U(\bullet) = O(k-1) e^{\gamma v(k+1)} \]  

where \( f_{\text{trunc}} \) is a truncation function to output 0 or 1 depending on its input value.

The torsion factor, \( T \), represents the shape distortion of the control surfaces, ranging from 1 to \( T_{\text{max}} \) which represents its maximum value, varying with channel rates, as shown below:

\[ T_{\text{max}} = \frac{A}{\text{channel rate}} \]

where \( A \) is a constant. When the channel rate is high, the expanded channel capacity can handle the video rate fluctuation, hence, a smaller \( T_{\text{max}} \) can be used. For a lower channel rate, a higher value is assigned to provide the surface with a larger torsion. The constants, \( \alpha \) and \( C \), are balancing factors forming the surfaces in a balanced or an unbalanced shape. Figure 5 shows two extreme cases of \( T_{\text{max}} \) for specific values 3 and 13. The surfaces with a larger \( T_{\text{max}} \) value exhibit more torsion.

**4. SIMULATION RESULTS**

Two video sequences, “Starwars” and “JFK,” were used in simulations to give frequent scene changes and non-stationary input video data to the encoder. The sequence we used contains 390 frames captured from parts with rapid motion and dramatic scene changes. “JFK” has more dramatic scene changes; transitions between colored and monochromatic scenes and rapid zooming. The video encoder is set to operate at a channel rate of 1024 Kbps and a frame rate of 30 frames/s. It has a buffer with the size of twice of MBF. For each value of \( p_{\text{type}}(i) \), the integers 10, 8 and 6 are assigned to 1, P and B pictures, respectively. We first assessed the performance of non-linear surfaces depending on the values of \( T_{\text{max}} \), as shown in Table 1. NFVR in the middle column stands for normalized fluctuation of the video rate which represents the total amount of MFB fluctuations:

\[ \text{NFVR} = \frac{\sigma^2}{1 + \sigma^2} \quad \sigma^2 = E \left( \frac{\text{MBF} - 1}{\text{MBF}} \right)^2 \]
where \( \phi(t) \) represents instantaneous fluctuation. Both surfaces show better rate control performance with reduced variance as \( T_{\text{max}} \) increases. While SIGM exhibits less fluctuations in video rate, UNIM appears superior in terms of mean PSNR with the standard deviation (std. dev.) close to SIGM.

### Table 1. Effect of changing the torsion factor.

Two rate control schemes were evaluated in comparison to TM5, a linear rate estimator optimised with the recursive least square (RLS) algorithm, which has no RBF layer and the RBF-network estimator shown in Figure 3. Both schemes employed UNIM surface for better video quality. For the nonlinear quantisation mapping surfaces, \( T_{\text{max}} \) is set to 7. Figure 6 shows profiles of the three schemes for frames 180 to 250 where dramatic scene changes occur. Table 2 summarises the performance for all 300 frames.

![Figure 6. Performance of rate control algorithms ("JFK" sequence): (a) buffer occupancy, (b) PSNR.](image)

<table>
<thead>
<tr>
<th>Starwars</th>
<th>Occupancy(%)</th>
<th>coded bits / frame (bits)</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JFK</td>
<td>Occupancy(%)</td>
<td>coded bits / frame (bits)</td>
<td>PSNR (dB)</td>
</tr>
<tr>
<td>TMAX</td>
<td>mean</td>
<td>std.dev.</td>
<td>NFVR</td>
</tr>
<tr>
<td>TM5</td>
<td>41</td>
<td>10.78</td>
<td>0.285</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>0.54</td>
<td>0.030</td>
</tr>
<tr>
<td>UNIM</td>
<td>26</td>
<td>4.89</td>
<td>0.122</td>
</tr>
<tr>
<td>7</td>
<td>51</td>
<td>0.39</td>
<td>0.027</td>
</tr>
<tr>
<td>RLS</td>
<td>18</td>
<td>4.17</td>
<td>0.117</td>
</tr>
<tr>
<td>9</td>
<td>51</td>
<td>0.33</td>
<td>0.023</td>
</tr>
<tr>
<td>RBF</td>
<td>13</td>
<td>4.16</td>
<td>0.111</td>
</tr>
<tr>
<td>11</td>
<td>51</td>
<td>0.26</td>
<td>0.022</td>
</tr>
<tr>
<td>13</td>
<td>10</td>
<td>3.69</td>
<td>0.106</td>
</tr>
<tr>
<td>UNIM</td>
<td>51</td>
<td>0.24</td>
<td>0.020</td>
</tr>
<tr>
<td>7</td>
<td>3.43</td>
<td>0.099</td>
<td>3755</td>
</tr>
</tbody>
</table>

Table 2. Mean and standard deviation of performance measures.

TM5 exhibits inferior control capability to the two other schemes in terms of both occupancy. Although the std. dev. in PSNR appeared smaller for TM5 than for RLS and RBF, the average PSNR of TM5 is slightly lower than the two others. RBF appeared to be capable of keeping the occupancy lower with a smaller std. dev. than RLS, without quality degradation. Note that the NFVR and the std. dev. of coded bits/frame are considerably smaller than those of RLS, and that the performance is better than Figure 1.

5. CONCLUSION

The MPEG2 video rate control technique, which is based on a nonlinear predictor and quantisation control, has been investigated for a constant bit rate transmission. The RBF network rate estimator appeared to improve the rate control performance in terms of video rate and video quality, when it is used in combination with the nonlinear quantisation technique employing the unimodal function. This signifies that the nonlinear predictive technique may substantially enhance the performance of the rate control mechanism when processing non-stationary video.

6. ACKNOWLEDGEMENTS

Yoo-Suk Saw acknowledges the support of The British Council and GoldStar Information and Communications Ltd., Korea.

REFERENCES

FEED-FORWARD NONLINEAR NETWORK
APPROACHES FOR MPEG VIDEO RATE PREDICTION

Peter M. Grant, Yoo-Sok Saw, John M. Hannah and Bernard Mulgrew
Dept. of Electrical Engineering, University of Edinburgh, Edinburgh, EH9 3JJ, UK.
Tel: +44 131 650 5655; fax: +44 131 650 6554
E-mail: pmsg@ee.ed.ac.uk

ABSTRACT

Conventional approaches generally assume that the compressed video has high correlation so that linear predictive methods can be applied. However, for realistic videos such as movies, sports and advertisements, there can be many exceptions since the correlation may be abnormally low. In this paper, we developed a feed-forward network-based rate control scheme which effectively accommodates dramatic variations in video rate by introducing a nonlinear predictive technique. We investigated a radial basis function (RBF) network nonlinear video rate estimator. As a constituent quantiser control method, nonlinear functional surfaces were also employed. The performance of the scheme was compared with the MPEG2 Test Model (TM) 5 and was assessed in terms of buffer occupancy, video data rate (bits/frame) and peak signal-to-noise ratio (PSNR).

1 INTRODUCTION

An effective rate control technique becomes more demanding when the video contains a long duration of rapid motion or frequent scene changes. A noteworthy fact for this type of video is that the occurrence of abrupt changes in visual information can hardly be predicted in purely stochastic ways. Hence, as a solution, it is necessary to introduce apriori knowledge to control the buffer and the quantiser in the MPEG video encoder. In this approach, we employed a nonlinear feed-forward predictive scheme in order to exploit short-term correlation of video and to improve the estimation performance. Attention was focused on the RBF network as a nonlinear predictor [1]. A good rate control algorithm will keep the occupancy as steady as possible within the desired delay limit - by controlling the number of coded bits per unit time - with little quality degradation in PSNR.

This paper is organised as follows: Sec. 2 gives an overview on the feed-forward video rate estimation approach. Sec. 3 describes the configuration of the nonlinear video rate estimation scheme. Sec. 4 discusses the nonlinear functional surfaces for the quantiser control. Sec. 5 includes the simulation results and discussions on the performance. Finally, Sec. 6 concludes this chapter.

2 FEED-FORWARD VIDEO RATE ESTIMATION FOR MPEG

Any linear combiner can be used as a video rate estimator which takes some change features as its input. First, we tested the linear estimator trained with recursive least squares (RLS) algorithm, which is widely used in the field of adaptive control. A radial basis function (RBF) network was then adopted as a nonlinear estimator. The RBF network is known to have better estimation performance than linear predictors for non-stationary signals and has recently been used successfully in several engineering areas such as channel equalisation [2]. Though the RBF is classified as a neural network, it is computationally simpler than other neural networks with multiple hidden layers [3], i.e. multi-layer perceptron. Fig. 1 shows the configuration of the generalised feed-forward video rate estimation. This scheme comprises three main functions: some change calculator (SCC), rate estimator (RE) and quantiser control (QC) based on nonlinear functional surfaces.

The RLS algorithm updates estimator weights at every frame start with the updated error signal e(k). Considering the non-stationary nature of the inputs, the RLS-based linear estimator is used instead of the non-recursive least square predictor since the weights should be updated depending on changes in input statistics. The scene change calculator operates in advance of actual encoding in order to estimate the frame-wise video data rate, cfb(k). It outputs scene change features by calculating the variances, var prog(k) and var diff(k), which represent the variance of the input picture and the variance between the input picture and the previous picture, respectively.

The scene change calculator also passes picture type information, ptype(k), for an input picture on to the rate estimator. ptype(k) gives vital information on the video rate to the rate estimator since the MPEG video encoder processes picture frames in different coding modes according to the repetitive picture type [4]. Each ptype(k) value has a single value for the corresponding picture type, thus it forms a cyclic time series as k proceeds.

The predicted video rate, $\tilde{c}fb(k)$, is added to the current occupancy, $\tilde{O}(k-1, n)$, to form the predicted occupancy, $\tilde{O}(k)$. The nonlinear quantiser control finally outputs the quantisation scale value, $Qz(k, n)$. We view the
MPEG2 video encoder as a finite impulse response filter (H) which accepts Qa(k) as an input, and outputs cbf(k). The transmission buffer is treated as a delay ($T^{-1}$) since it stores coded bits for a specific frame period. The a priori information from SCC (var_org(k), var_dif(k) and pframe(k)) is fed to the rate estimator which adaptively changes its coefficients. MBF is a constant determined by the current channel rate, which represents the mean bits per frame.

Fig. 1: The structure of the feed-forward network-based rate control for MPEG.

3 A NONLINEAR ESTIMATOR, RADIAL BASIS FUNCTION NETWORK

The RBF-network rate estimator is viewed as a nonlinear (universal) functional approximator [5] to approximate the video rate which is a function of scene change feature vector $x$. From the equations described in [5],

$$\hat{d}f(x) = (x, w) + b$$  \hspace{1cm} (1)

where $x$ and $w$ are input vector and weights respectively, and $b$ is scalar bias. The term $(x, w)$ represents inner product, thus, this is replaced with a network with a radial function, i.e. Gaussian layer plus a linear combiner.

The RBF network, Fig. 2, consists of centres with the radial basis function and linear weights expressed as:

$$\hat{d}f(x) = \sum_{i=1}^{N} w_i \phi(||x - x_i||)$$  \hspace{1cm} (2)

where $\phi(||x - x_i||) = \exp(- \frac{||x - x_i||^2}{2\sigma^2})$

where $\hat{d}f(x)$ is the output of the RBF network, $w_i$ represents linear weights which are trained in a recursive least square method, $x$ is an input vector containing scene change features, and $x_i$ represents the $i$th selected centre among $N$ centres in total. $\phi(N)$ is a Gaussian function which outputs RBF layer values.

Fig. 2 RBF predictor with 3 inputs and 9 taps.

The RBF network may have as many centres as required by selecting from the input, and it calculates the closeness of each input over the selected centres using the Gaussian function. For computational efficiency, however, the RBF centres are usually selected by the orthogonal least square (OLS) algorithm [6]. The OLS algorithm selects representative RBF centres by using supervised learning. However, in the case of the running MPEG encoder, the supervised learning cannot be efficiently used, thus the k-means clustering algorithm is used, as for the RBF channel equalisation [7, 8] application where the centres are updated as follows:

$$x_i(k) = x_i(k-1) + g \cdot \hat{d}f(k) - x_i(k-1)$$  \hspace{1cm} (3)

where the constant $g$ controls the learning rate.

The number of RBF centres is set to 9. As shown in Fig. 2, the input consists of 3 scene change features, each of which has 3 taps. The number of taps is 1 less than the number of B pictures between P pictures. It is assumed that the cyclic repetition of the video rate is determined by the number of B pictures, and the correlation of the video rate varies with an interval of two pictures. Different numbers of centres - up to 50 - were also simulated and their performance was compared in NMSE, as shown in Fig. 3 and Table 1. We used realistic video sequences, whose features will be discussed in Sec. 5, to see the effect of the RBF network estimator. For all four video sequences tested, the 9 centres appeared to exhibit the same performance as cases of the larger number of centres. It is considered that the number of RBF centres may be set in conjunction with the number of the B pictures. This implies that the system complexity of a RBF-network-based rate control can be dramatically reduced by selecting appropriate number of centres from video encoding parameters.
MSE[dB] | \(\text{var}_\text{org}(k)\) | \(\text{var}_\text{dif}(k)\)
---|---|---
Cascaded | -24.49 | 1224.8 | 579.6
Starwars | -22.44 | 1671.2 | 713.4
Adverts | -20.67 | 2888.5 | 1143.9
JFK | -15.33 | 3329.9 | 1914.7

**Table 1: Mean NMSE and Mean Variances of Video Sequences.**

### 4 Quantisation Control by Nonlinear Functional Surfaces

Rate control adjusts the buffer occupancy by selecting an appropriate quantisation step size. The quantisation step size - determined by the quantisation scale value - is the dominant parameter which is used to achieve the goal of rate control, as shown in Fig. 4.

\[ Q_s(k, n) = f(\hat{O}(k - 1, n), \hat{O}(k, n)) \]

where the quantisation scale \(Q_s(k, n)\) is determined by both \(\hat{O}(k - 1, n)\) and \(\hat{O}(k, n)\), which is the summation of the current occupancy and the predicted video rate \(\hat{d}(k)\) as shown in Fig. 4. The function \(f()\) is a nonlinear functional surface to adaptively map the occupancies to the quantisation scale.

In this paper two different nonlinear functions were used; sigmoidal and unimodal, as shown in Fig. 5. The way the two surfaces work is the same as described in Fig. 4.

The lower the channel rate is, the more nonlinear the surface becomes, such as Fig. 5 (b) and (d).

### 5 Simulation Results

Four video sequences, "Cascaded", "Adverts", "Starwars" and "JFK", are used to give frequent scene changes and rapid motion videos to the encoder. "Cascaded" is composed of three standard sequences in the order of "Miss America", "Football" and "Susie". "Adverts" contains three television advertisements which exhibit rapid motion and dramatic scene changes. "Starwars" and "JFK" were digitised from their televised versions. "JFK" possesses dramatic scene changes, i.e. transitions between coloured scenes and monochrome scenes and rapid zooming, etc. The MPEG2 video encoder based on TM5 was set to operate at a channel rate of 1024 kbps and a frame rate at 30 frames/s. It has a buffer with the size of twice MBF [4]. The unimodal function was used as the quantiser control surface since it possesses better performance for video rate control [9].

Fig. 6 shows the performance of the three rate control schemes for the critical part of "JFK" video sequence.
Table 2 shows the mean and the standard deviation for each of the performance measure. NFVR in the middle column stands for normalised fluctuation of the video rate which represents flatness of $ch/(MFB)$. It is expressed in the following equation:

$$NFVR = \frac{\sigma}{1 + \sigma} \quad \sigma^2 = E \left[ \left( \frac{ch(k)}{MFB} - 1 \right)^2 \right]$$

where $ch(k)/MFB$ represents instantaneous fluctuation.

![Graphs](image)

Fig. 6 Performance of rate control algorithms (“JFK”): (a) occupancy, (b) coded bits/frame, (c) PSNR

<table>
<thead>
<tr>
<th>JFK</th>
<th>Occupancy (%)</th>
<th>coded bits/frame (bits)</th>
<th>PSNR (dB)</th>
<th>mean</th>
<th>std.dev</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TM5</td>
<td>10 (172)</td>
<td>0.378</td>
<td>2017</td>
<td>36.08</td>
<td>3.34</td>
</tr>
<tr>
<td>RLS</td>
<td>17.70</td>
<td>0.188</td>
<td>7687</td>
<td>36.49</td>
<td>3.83</td>
</tr>
<tr>
<td>RBF</td>
<td>11.61</td>
<td>0.169</td>
<td>6661</td>
<td>36.48</td>
<td>3.83</td>
</tr>
</tbody>
</table>

Table 2: Mean and standard deviation of performance measures (“JFK”).

TM5 exhibits the worst occupancy performance, often reaching buffer overflow, Fig. 6 (a). The schemes based on predictors, RLS and RBF, show better performance than TM5. RBF appeared to be capable of maintaining the occupancy lower with a smaller standard deviation in comparison to RLS. RBF also exhibits a similar PSNR value to RLS. This result implies that the nonlinear rate estimator, RBF-network, works more effectively for non-stationary video with many scene changes and rapid motion without further degradation in video quality.

6 CONCLUSION

We developed a feed-forward network-based technique for MPEG video rate control by employing a rate estimator which is either the linear network trained with the RLS algorithm or the RBF network. In order to efficiently combine the rate estimator with the quantiser control, the nonlinear functional surface was adopted. The unimodal surface appeared to have better performance than sigmoidal surface when it was coupled with the rate estimator. We focused our attention on the performance of the two rate estimators. In comparison with TM5, the feedforward approach appeared to have substantial improvement in video rate controllability and the video quality. The nonlinear RBF-network rate estimator exhibited better performance than the linear network for video with more scene changes. Hence, we conclude that the nonlinear predictive approach is better suited to video rate control to permit dramatic video rate fluctuation to be accommodated effectively.

7 ACKNOWLEDGEMENTS

Yoo-Suk Saw acknowledges the support of The British Council and GoldStar Information and Communications Ltd., Korea.

References

FUZZY RULE-BASED MPEG VIDEO RATE PREDICTION AND CONTROL

Peter M. Grant, Yoo-Sok Saw and John M. Hannah
Dept. of Electrical Engineering, University of Edinburgh, Edinburgh, EH9 3JL, UK.
Tel: +44 131 650 6555; fax: +44 131 650 6554
e-mail: psg@ee.ed.ac.uk

ABSTRACT

In this paper we investigated a fuzzy logic-based video rate control technique which aims to regulate the data rate of compressed video at a constant transmission rate without objectionable quality degradation. It is considered that conventional fuzzy rule-based control (FRC) does not effectively control the two variables (video data rate and video quality) which are mutually contradictory. The primary reason for this is that it is not easy to effectively project the control variables onto the fuzzy rules due to the contradiction on a rate-distortion theoretic basis. However, it is clear that the video quality should be considered as an equally important variable as well as the video rate. We developed a FRC scheme which also takes the video quality into account by using feed-forward scaling factors whose inputs are scene change features. The performance of this scheme was compared with the MPEG-2 Test Model (TM) 5 and the FRC scheme without the quality consideration, in terms of the buffer occupancy, the number of coded bits per frame and peak signal-to-noise ratio (PSNR).

1 INTRODUCTION

Fuzzy logic and fuzzy set theory has been extensively used, particularly for industrial and commercial control applications [1, 2] since its concept [3] was first published. Fuzzy logic control is known to be effective in conveying the meaning of linguistic variables to target systems [4]. Although an analogy can be found in other industrial applications, video rate control does not just fall into a matter of controlling the level of liquid reservoir which has been treated as a typical FRC application. The occupancy of video buffer has direct influence on video quality while the liquid level is irrelevant to the output fluid quality. In order to improve the FRC performance for video rate control, we introduced an adaptation induced by scaling factors in a way which they change depending on the scene change features. First, we examined the conventional FRC where the number of fuzzy variables is one, and its performance was evaluated in various settings of the fuzzy control parameters. As an adaptive algorithm, a FRC scheme with feed-forward scaling factors was designed and its performance was compared to the conventional FRC.

This paper is organised as follows: Sec. 2 describes the basic FRC configuration for video rate control. Sec. 3 and Sec. 4, respectively, discuss the design process of fuzzy rules and matters of fuzzy control parameters. In Sec. 5 we present the configuration of the FRC with the feed-forward scaling factors (FRC-FS). Sec. 6 includes the simulation results and discussions on the performance. Finally, Sec. 7 concludes this chapter.

2 FUZZY RULE-BASED CONTROL FOR VIDEO RATE CONTROL

Recently, a few leading researches applied the FRC to video sequence coding algorithms such as JPEG and ITU-T H.261. The techniques aim to improve the video rate control performance for JPEG [5, 6] and H.261 [7] by adaptively controlling the quantiser and the buffer occupancy. The fuzzy rule-based control techniques used in these researches have the same technical base in that they appeared to follow a series of common processes: fuzzification, decision making and defuzzification.

Figure 1 shows the configuration of the FRC-based video rate control (FRC-V) which takes the buffer occupancy as its only input. The control input, $Q_{e}$, which is used as the quantisation step size, $Q_{e}(n)$, is the input to the encoder for the time index $n$. The step size is used for the next macro block, $mb(n+1)$, by the quantiser, which outputs the corresponding compressed bit stream $Cb(n+1)$. The occupancy, $O(n)$, can be seen as the output of the encoder. The error signal, $e(n)$, which is the difference between $O_{t}$ (target occupancy) and $O(n)$, becomes the input of the FRC. The process of the FRC begins with calculating the error value, $e(n)$, and its differential error value, $d(n)$, which is the difference between the current error value, $e(n)$, and the previous error value, $e(n-1)$. The entire process to generate the output, $go$, proceeds with the two inputs, $Ge$ and $Gd$, which are translated into linguistic expressions $Le$ and $Ld$. In the decision making process a linguistic judgement, $Lo$, is determined based on a predetermined set of rules. The defuzzification process calculates, $o(n)$, by combining the membership function of $Lo$ and those of $Le$ and $Ld$ in a set theoretic way, e.g. intersection. Its final output is obtained by a series of arithmetic opera-
tions, e.g., centre of gravity [8]. All scaling factors ($g_c$, $g_d$ and $g_o$) are constants which are generally tuned by expert knowledge.

![Diagram of FRC-based rate control](image)

**Fig. 1:** Configuration of the FRC-based rate control.

### 3 DESIGNING FUZZY RULE-BASED VIDEO RATE CONTROL

The first step of designing FRC is initiated by transforming a series of expert knowledge into a set of rules comprising linguistic expressions [9] so that a decision for the output value can be made. A fuzzy set is expressed in a non-numerical form carrying linguistic meanings, e.g., very big, big, small or very small. Three fuzzy variables are listed in Table 1, each of which consists of 7 fuzzy sets.

<table>
<thead>
<tr>
<th>Fuzzy logic variables</th>
<th>Occupancy (FVQ)</th>
<th>Differential scale (FVQ)</th>
<th>Quantisation scale (FVQ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FL</td>
<td>FL</td>
<td>PB</td>
<td>HG</td>
</tr>
<tr>
<td>CF</td>
<td>close to full</td>
<td>positive big</td>
<td>large</td>
</tr>
<tr>
<td>HH</td>
<td>higher than HH</td>
<td>positive medium</td>
<td>LG</td>
</tr>
<tr>
<td>MF</td>
<td>half full</td>
<td>positive small</td>
<td>LM</td>
</tr>
<tr>
<td>LH</td>
<td>lower than HH</td>
<td>negative small</td>
<td>SM</td>
</tr>
<tr>
<td>NE</td>
<td>close to empty</td>
<td>negative medium</td>
<td>SL</td>
</tr>
<tr>
<td>ET</td>
<td>empty</td>
<td>negative big</td>
<td>TN</td>
</tr>
</tbody>
</table>

**Table 1:** Fuzzy logic variables for the fuzzy control input and the resulting output.

A triangular membership function, Fig. 2(a), which is mapped on to a normalised range from -0.5 to 0.5, is used for two inputs ($g_c$ and $g_d$) and the output ($g_o$) under the assumption that all the control variables have similar dynamic property associated with the membership functions. The complete representation of the rules can be given either in **IF ... THEN** statements or in a tabular form which is usually called fuzzy associative memory (FAM) [10]. The FAM representation, Fig. 2(b), is known to be more efficient in handling a complicated organisation of the rules [10]. The organisation of FAM follows a common method which locates fuzzy variables in the 45-degree diagonal. Defuzzification, which converts $g_c$ and $g_d$ to a crisp output value $o(n)$, is performed using the rules representing $L_0$ and the membership function values derived from input values ($g_c$ and $g_d$). Each selected membership value is used to defuzzify the output into the crisp value $o(n)$ using the centre of gravity method or one of its simplified versions (Mamdani’s model or Larsen’s product operation rule) [11, 8]

### 4 THE FUZZY CONTROL PARAMETERS

The scaling factors can be tuned depending on the dynamic ranges of corresponding inputs, $c(n)$ and $d(n)$, and output, $o(n)$. The bigger values they take, the quicker response can be achieved. In video rate control $g_o$ is fixed at 1.0 since the actual output $Q_o$ is multiplied by 31 for adjusting it to the legal range of MPEG2 quantisation scale. A small change of $g_o$ can cause wide fluctuation in the quantisation scale. $g_c$ and $g_d$ can be set to specific values suitable for the dynamic ranges of $c(n)$ and $d(n)$. Alternatively, the scaling factors can be adaptively controlled by a supervisory or adaptive control function.

![Diagram of fuzzy control parameters](image)

**Fig. 2:** Fuzzy rule-based control parameters for video rate control. (a) membership function, (b) FAM, (c) 3-dimensional representation of the FAM, (d) the resulting control surface.

The membership function can take different shapes, different inter-rule spacing, etc. It is well known that using non-triangular shapes does not provide substantial difference in the performance [4], hence, the triangular shape is used in this paper. The formation can be asymmetrical since the positive section of crisp value can have different significance from the negative section. In video rate control, however, both sections here are assumed to have unbiased linguistic interpretation. Thus, the membership functions shown in Fig. 2 (a) are symmetrical with respect to the centre value 0.

The 3-dimensional representation of FAM, Fig. 2 (c), reflects the dynamic property of an organisation of rules and membership functions since it has large influence on the performance of the entire FRC. Fig. 2(d) shows the resulting control surface for the FAM configuration of Fig.
2(b) and (c).

5 FEED-FORWARD SCALING FACTORS USING SCENE CHANGE FEATURES

Scene change features provide vital information about the resulting number of bits for an incoming picture in advance of encoding it. Here, they are incorporated with the non-adaptive scaling factors \( g_e, g_d \) and \( g_0 \) of FRC-R, as shown in Fig. 3, in order to adaptively scale inputs of the fuzzyization process, \( e(n) \) and \( d(n) \). This configuration forms the adaptive FRC, i.e. FRC-FS. Three scene change features, \( \var{ge}(k), \var{gd}(k) \) and \( \var{ptype}(k) \), are supplied to the scaling factor calculation block (Mapping function) which generates time varying scaling factor values, \( g_e(k) \) and \( g_d(k) \). \( \var{ge}(k) \) and \( \var{gd}(k) \) are variances of the input picture and the difference picture between the current and the previous ones. \( \var{ptype}(k) \) is the picture type information which has an integer value for the time index \( k \) depending on the picture type, i.e. I, P and B [12]. The rest of the FRC-SC operation is the same as FRC-R depicted in Fig. 1. In the mapping function the equation maps the scene change features to the scale factors as follows:

\[
g_e(k) = g_d(k) = \frac{\log_{10} \var{ge}(k)}{\log_{10} \var{gd}(k)} \times \var{ptype}(k) \tag{1}
\]

Fig. 3: Configuration of the FRC-FS.

6 SIMULATION RESULTS

The basic FRC model (FRC-R), described in Sec. 2 to 4, possesses a considerable flexibility to change the fuzzy control parameters. The scaling factors were assessed by applying 8 different values ranging from 2 to 16 increasing by 2. The scaling factor for the output \( g_o(k) \) was set to 1. The target occupancy was set to 30%, in order to observe the effect of lower value than 50%. As a test sequence, “Starwars” movie was used. The pictures (300 frames) were digitised from a part of the movie, with rapid motion and frequent scene changes. Fig. 4 shows encoding results for different scaling factor values.

For simplicity, only two extreme cases are displayed in Fig. 4 (b) and (c). Bigger values of \( g_e(k) \) and \( g_d(k) \) (e.g. 16) exhibit far better control capability over the occupancy and the bits/frame. However, the PSNR is noticeably lower than the case of smaller scaling factor values since, in FRC-R, the quality is not taken into account of its control process. Pictures, which entail a large amount of bits, will be given much stronger control action. When they are I or P pictures, the distortion caused by rate control will affect the next coming pictures. For this reason, as shown in Fig. 4 (c), PSNR remains low when scaling factors are large. The performance difference can be found easily around the frame number 130 to 180.

Fig. 4: Performance of FRC-R depending on scaling factors (“Starwars”).

The FRC model assisted by the feed-forward scaling factors (FRC-FS), Sec. 5, was assessed with respect to FRC-R. Fig. 5 shows a critical part of the encoded results where the frame number ranges from 130 to 200. Table 2 summarises the performance for the entire video sequence. The normalised fluctuation of the video rate (NFVR) is represented by variations in the coded bits per frame \( \var{df}(k) \). It is expressed by the following equation:

\[
NFVR = \frac{\sigma}{\mu + \sigma}, \quad \sigma^2 = E \left[ \left( \frac{\var{df}(k)}{\var{MBF}} - 1 \right)^2 \right] \tag{2}
\]

where \( \var{df}(k) \) represents the instantaneous fluctuation. The operator \( E \) is the statistical expectation. The performance of TM5 is given for reference. While FRC-R is superior in controlling the video rate or the occupancy to FRC-FS, the latter shows wider variations since its scaling factors change depending on scene change. Hence, profiles of the occupancy and video rate may dramatically vary depending on the scene change. The video rate profile exhibits similarly changing patterns.

FRC-R consists of fuzzy control rules based only on the occupancy-related rules. Therefore, it appeared to be powerful in controlling the occupancy. However, it
Appendix D: Publications

does not take account of the quality, and it shows lower figures in PSNR. In the simulation, the scaling factors are set to 8 to see the performance of controlling the occupancy. For FRC-FS, they are allowed to change within a 1-to-8 range depending on the scene change features according to Eqn. 1. In FRC-FS, the scaling factors \( g_c(k) \) and \( g_d(k) \) scale up the error signal, \( e(n) \). Fig. 3, to adaptively change the actual input of \( e(n) \). The scaling factors are generally smaller than 8, and the resulting performance of the occupancy control appear to be inferior to FRC-R. Accordingly, FRC-FS shows more fluctuating profiles. However, the feed-forward scaling factors \( g_c(k) \) and \( g_d(k) \) improve the video quality by making the most of the occupancy margin below 30%. Therefore, it is concluded that FRC-FS performs better in terms of PSNR achieved as well as the occupancy.

![Performance profiles for FRC-R and FRC-FS (Starwars).](image)

**Fig. 5. Performance profiles for FRC-R and FRC-FS ("Starwars").** (a) occupancy, (b) coded bits/frame, (c) PSNR.

<table>
<thead>
<tr>
<th>Starwars</th>
<th>Occupancy (%)</th>
<th>Coded bits/frame</th>
<th>PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TM5</td>
<td>40 (75)</td>
<td>1.10</td>
<td>33.70</td>
</tr>
<tr>
<td>FRC-R</td>
<td>27 (32)</td>
<td>1.5</td>
<td>33.40</td>
</tr>
<tr>
<td>FRC-SC</td>
<td>16 (42)</td>
<td>5.3</td>
<td>33.86</td>
</tr>
</tbody>
</table>

**Table 2: Performance Comparison between FRC-R and FRC-FS ("Starwars").**

7 CONCLUSION

In this paper the fuzzy rule-based control scheme was applied to video rate control and its performance evaluated for several different scaling factor values. A basic fuzzy logic control model was examined, considering the buffer occupancy as the only fuzzy control variable. It was used to investigate the effect of scaling factors in fuzzy logic control. With this model as a performance reference, an improved fuzzy logic control scheme was evaluated, which was the FRC with feed-forward scaling factors (FRC-FS).

The two schemes were compared in terms of mean and standard deviation of the performance measures (the occupancy, the coded bits/frame and PSNR). As for the occupancy and the coded bits/frame, fuzzy rule-based control schemes appeared to achieve a goal of video rate control - with a steady bit rate no matter how dramatic the scene change. However, the necessity of enhancing video quality has risen since the occupancy-driven scheme does not take the quality factor into account. In FRC-FS, the scene change factors were introduced so that the quality was improved by allowing the occupancy to fluctuate between the target value and the empty state. The FCR-SC scheme, which is controlled both by the video rate factor and the scene change-based factor, exhibited better controllability over the occupancy as well as better video quality. Simulations for different video sequences showed similar results.

8 ACKNOWLEDGEMENTS

Yoo-Sok Saw acknowledges the support of The British Council and GoldStar Information and Communications Ltd., Korea.

References


