Multiple Description Coding for 3D Video

Hezerul Abdul Karim

Submitted for the Degree of
Doctor of Philosophy
from the
University of Surrey

Centre for Communication and Systems Research
School of Electronics and Physical Sciences
University of Surrey
Guildford, Surrey GU2 7XH, UK

May 2008

© Hezerul Abdul Karim 2008
Abstract

In the near future, 3D video is likely to be used to enhance video applications, as it offers a greater sense of immersion. When 3D video is compressed and transmitted over error prone channels, the associated packet loss leads to poor visual quality. Hence, error resilience techniques for 3D video are needed. This thesis aims to improve the error robustness of the compressed 3D video in error prone transmission scenarios.

Firstly, this thesis describes how 3D video can be represented using 2D video information, and depth information. This format can be compressed using tools available in some video coding standards, including Multiple Auxiliary Component (MAC) tool in MPEG-4 version 2, and the use of reduced resolution coding for depth compression. It is observed that the reduced resolution depth compression provides improved 2D video performance. However, the quality of the depth information is limited at high bit rates due to the distortion introduced by down-sampling and up-sampling (DSUS).

Secondly, Multiple Description Coding (MDC), based on even and odd frames is proposed for error resilient 3D video. Improvements are made to the original scheme by adding a controllable amount of side information to improve frame interpolation at the decoder and compression efficiency. The side information is also sent according to the video sequence motion for further improvement. The performances of the proposed MDC algorithms are found to be better than single description coding (SDC) and the original scheme at high error rates with reduced error free coding efficiency.

Finally, the combination of Scalable Video Coding (SVC) and MDC (scalable MDC) for 3D video is investigated for error robustness and scalability. A scalable MDC scheme based on even and odd frames is proposed for H.264 based SVC. Reduced resolution depth compression is then applied to improve the performance. The proposed algorithms provide better 3D video performance than the original SVC in error prone environments and for low bit-rate video.

Key words: stereoscopic 3D video coding, 2D and depth, error resilience, multiple description video coding, scalable multiple description video coding.
Acknowledgments

I would like to thank Allah swt for His guidance and blessing in completing this thesis. Special thank to my supervisor, Dr. S. Worrall for all his kind supervision and valuable suggestions for the works in this thesis. I would also like to thank Prof A. M. Kondoz for all his comments and proposal for the completion of this PhD work. I would like to thank Dr. A. H. Sadka who previously supervised me for large part of work in this thesis. I would like to specially thank my lovely wife, Aduwati Sali and my two kids Hazim and Abdul Haseeb for their patience and support during the completion of the thesis. Special thanks I dedicated to my parents for their prayer, support and love. I would also like to acknowledge my sponsor, Islamic Development Bank (IDB) and my employer, Multimedia University (MMU) for allowing me to further my studies. Some of the work presented here was developed within VISNET, a European Network of Excellence (http://www.visnet-noe.org), funded under the European Commission IST FP6 programme. I also like to specially thank my brothers and sisters in Majlis Syura Muslimun UK and EIRE for their spiritual support and prayer. And last but not least, thank you to my colleague in ILAB and support staffs in CCSR for their support and encouragement.
## Contents

Abstract ............................................................................................................................................... ii
Acknowledgments ................................................................................................................................. iii
Contents ................................................................................................................................................ iv
List of Figures ....................................................................................................................................... ix
List of Abbreviations ............................................................................................................................ xv
List of Mathematical Symbols .............................................................................................................. xviii

1 Introduction ......................................................................................................................................... 1
   1.1 Preamble ....................................................................................................................................... 1
   1.2 Motivation ..................................................................................................................................... 2
   1.3 Application and Scenarios ........................................................................................................... 3
      1.3.1 3DTV Broadcast ............................................................................................................... 3
      1.3.2 Mobile 3DTV ................................................................................................................... 4
      1.3.3 3D Video Streaming ......................................................................................................... 5
      1.3.4 Virtual Collaboration 3D Video Conferencing ................................................................ 5
   1.4 Objectives and Overall Project Description ............................................................................... 6
   1.5 Original Achievements ............................................................................................................... 7
   1.6 Structure of Thesis .................................................................................................................... 8
   1.7 Publications ............................................................................................................................. 10

2 3D Video Coding ................................................................................................................................ 12
   2.1 Introduction .................................................................................................................................. 12
   2.2 3D Video Classification ............................................................................................................. 12
   2.3 Stereoscopic Video ..................................................................................................................... 13
   2.4 Stereoscopic Video from 2D Video and Depth ......................................................................... 16
   2.5 2D Video and Depth Test Video Sequences ........................................................................... 17
   2.6 3D Video Quality Assessment and Performance Evaluation ................................................. 19
   2.7 3D Video Coding ....................................................................................................................... 19
      2.7.1 Two-View Stereoscopic Video Coding ............................................................................ 20
      2.7.2 2D and Depth Stereoscopic Video Coding ..................................................................... 20
   2.8 Reduced Resolution Depth Compression for 3D Video ......................................................... 21
2.8.1 Depth Compression .................................................................................................. 22
2.8.2 Down-Sampling and Up-Sampling Method ............................................................... 22
2.8.3 Results of Down-Sampling and Up-Sampling for Depth Information ................. 24
2.8.4 Discussion .................................................................................................................. 25
2.9 MPEG4-Multiple Auxiliary Component for 3D Video Coding ........................................ 26
2.9.1 MPEG4-Multiple Auxiliary Component Architecture ............................................... 27
2.9.2 MPEG4-Separate Architecture ................................................................................... 28
2.9.3 Experimental Results for MPEG4-MAC and MPEG4-Separate Coding ............... 29
2.9.3.1 Discussion ............................................................................................................ 29
2.9.4 MPEG4-MAC for 3D Video Coding with Down-Sampling and Up-Sampling Application............................................................................................................................. 34
2.10 Conclusion ........................................................................................................................ 36

3 Error Resilience and Multiple Description Coding for 3D Video ................................. 38

3.1 Introduction ....................................................................................................................... 38
3.2 Error Resilience Tools ...................................................................................................... 38
3.2.1 Packet Resynchronisation ........................................................................................... 38
3.2.2 Data Partitioning ......................................................................................................... 39
3.2.3 Reversible Variable Length Codes ............................................................................. 39
3.2.4 Depth Data Error Resilience ....................................................................................... 40
3.2.5 Depth Data Sensitivity and Data Partitioning ............................................................. 40
3.3 Multiple Description Coding ............................................................................................ 42
3.3.1 Redundancies and Mismatch ...................................................................................... 43
3.3.2 Multiple Description Coding Method Based on Quantisation ................................... 44
3.3.3 Multiple Description Coding Method Based on Transform Coding ..................... 45
3.3.4 Multiple Description Coding Method Based on Sub-Sampling ............................... 46
3.3.5 Other Multiple Description Coding Algorithms ......................................................... 46
3.3.6 Even and Odd Frames Based Multiple Description Coding ....................................... 47
3.4 Even and Odd Frames Based Multiple Description Coding for 3D Video ............................. 49
3.4.1 The Proposed Even and Odd Frame Based Multiple Description Coding Algorithm 49
3.4.2 Detailed Structure of the Proposed Multiple Description Coding Even And Odd Algorithm............................................................................................................................. 51
3.4.3 MDC-Even and Odd with Depth Data Partitioning .................................................. 52
3.4.4 Performance Evaluation of the Proposed Multiple Description Coding Even And Odd Algorithm............................................................................................................................. 53
3.4.4.1 Error Free Environment ....................................................................................... 53
3.4.4.2 Frames Lost Environment .................................................................................. 55
4 Even and Odd Frames Based Multiple Description Coding with Side Information for 3D Video .................................................................................................................. 68

4.1 Introduction ....................................................................................................................... 68
4.2 Even and Odd Frames Based Multiple Description Coding with Side Information ........ 69
4.3 Even and Odd Frames Based Multiple Description Coding with Side Information and Prediction ............................................................................................................. 73
4.4 Even and Odd Frames Based Multiple Description Coding using B-Frame .............. 76
4.5 Performance Evaluation of The Proposed MDC-EO, MDC-EOS, MDC-EOSP and MDC-EOSB Algorithms ...................................................................................................................... 78
  4.5.1 Error Free Environment ............................................................................................ 78
  4.5.2 Even and Odd Frames Based Multiple Description Coding using B-Frame
      Performance in a UMTS Network ................................................................................... 79
  4.5.3 Error Free Environment with Varying Side Information .................................................. 83
  4.5.4 Frame Loss Environment .......................................................................................... 83
  4.5.5 Ideal Multiple Description Coding Channel ................................................................ 88
  4.5.6 Average Delta Peak Signal-to-Noise Ratio ................................................................. 88
  4.5.7 Performance of MDC-EOS in a UMTS Network ......................................................... 90
    4.5.7.1 Same Bit Rate Experiment ...................................................................................... 90
    4.5.7.2 Same Quality Experiment ..................................................................................... 91
  4.5.8 Performance of the Proposed MDC-EO, MDC-EOS, MDC-EOSP and MDC-EOSP
      Algorithms in a UMTS Network .................................................................................... 95
  4.5.9 Performance of MDC-EO and MDC-EOS Algorithms in Wireless LAN Network ... 98
4.6 Even and Odd Frames Based Multiple Description Coding with Adaptive Side
      Information ....................................................................................................................... 101
4.7 Performance Evaluation of the Proposed Multiple Description Coding with Adaptive Side
      Information ....................................................................................................................... 103
    4.7.1 Error Free Environment .......................................................................................... 103
    4.7.2 Frame Lost Environment ......................................................................................... 103
    4.7.3 Ideal Multiple Description Coding Channel .............................................................. 110
    4.7.4 Comparison of the Proposed Multiple Description Coding with and without Side
         Information ............................................................................................................... 110
    4.7.5 Average Delta Peak Signal-to-Noise Ratio ............................................................... 116
    4.7.6 Error Prone Environment Rate Distortion ............................................................... 117
    4.7.7 Same Bit Rate Experiment in Error Prone Environment ......................................... 119


## 5 Scalable Multiple Description Coding for 3D Video

5.1 Introduction ................................................................. 126
5.2 Related Work .............................................................. 127
   5.2.1 Scalable Video Coding .............................................. 127
   5.2.2 Review of Scalable Coding Techniques ...................... 127
   5.2.3 Overview of Scalable H.264 ....................................... 128
   5.2.4 Scalable Multiple Description Coding ...................... 129
5.3 Proposed Scalable Multiple Description Coding ............ 130
   5.3.1 Scalable Side Information ........................................... 131
   5.3.2 Scalable Multiple Description Coding Using Scalable Video Coding ................. 132
      5.3.2.1 Performance Evaluation of the Proposed Scalable MDC .................. 136
5.4 Performance Comparison of the Proposed Scalable MDC versus MDC-EOASP ....... 144
   5.4.1 Error Free Environment ........................................... 144
   5.4.2 Error Prone Environment ......................................... 145
      5.4.2.1 Same Quality Experiment .................................... 146
      5.4.2.2 Same Bit Rate Experiment ................................... 147
5.5 Scalable Video Coding with Down-Sampling and Up-Sampling ...................... 150
5.6 Scalable Multiple Description Coding with Down-Sampling and Up-Sampling .... 155
5.7 Performance Comparison with Down-Sampling and Up-Sampling ............... 158
   5.7.1 Error Free Environment ........................................... 158
   5.7.2 Error Prone Environment ......................................... 162
5.8 Conclusion ................................................................. 167

## 6 Conclusion and Future Work

6.1 Overview ................................................................. 171
6.2 Summary of the Work on 3D Video Compression .................. 171
6.3 Summary of the Work on 3D Video Error Resilience ........... 172
6.4 Summary of the Work on 3D Video Scalability ..................... 174
6.5 Original Achievements .................................................. 175
6.6 Areas of Future Research .............................................. 178
   6.6.1 3D Video Coding ................................................... 178
   6.6.2 3D Video Error Resilience ....................................... 178
   6.6.3 3D Video Scalability .............................................. 179

Bibliography 181
# List of Figures

- Figure 1-1: 3DTV display by Philips .......................................................... 4
- Figure 1-2: A Nokia handheld device that can display 2D video is displaying a stereoscopic anaglyph red and blue video .................................................. 4
- Figure 1-3: Stereoscopic 3D video streaming over Internet .......................................................... 5
- Figure 1-4: Virtual collaboration system application ........................................... 6
- Figure 2-1: (a) 2D video and (b) its depth information ........................................ 13
- Figure 2-2: 3D Stereoscopic sequence on Stereoscopic Movie Player .................. 17
- Figure 2-3: Original frames of 2D and depth image sequences at CIF resolution (a) Orbi texture (b) Orbi depth (c) Interview texture (d) Interview depth [30] ................................................ 18
- Figure 2-4: Separate encoding of 2D and depth ................................................... 21
- Figure 2-5: Orbi test sequence, (a) Luminance information, (b) Associated per-pixel depth information ............................................................... 22
- Figure 2-6: Down-sampling before H.264 encoder and up-sampling after H.264 decoder ........ 23
- Figure 2-7: Objective quality (PSNR) of compressed original resolution, DSUS2 and DSUS4 at 64 kbit/s ........................................................................................................... 24
- Figure 2-8: Quantisation parameters of compressed original resolution, DSUS2 and DSUS4 at 64 kbit/s ........................................................................................................... 24
- Figure 2-9: Frame 52 of Orbi depth image sequence, (a) Original, (b) Original compressed at 64 kbit/s, (c) DSUS2 at 64 kbit/s, (d) DSUS4 at 64 kbit/s. The most affected regions are highlighted in boxes ............................................................... 25
- Figure 2-10: MPEG4-MAC architecture ............................................................ 28
- Figure 2-11: MPEG4-Separate architecture ...................................................... 28
- Figure 2-12: (a) PSNR (Y component of the texture) and (b) total bits (texture) for each frame of Orbi at 2Mbit/s ....................................................................................................................... 31
- Figure 2-13: (a) PSNR (Y component of the depth) and (b) total bits (depth) for each frame of Orbi at 2Mbit/s ....................................................................................................................... 32
- Figure 2-14: Luminance subjective quality for Orbi sequence at 2Mbit/s, (a) MPEG4-MAC (b) MPEG4-Separate ............................................................... 33
- Figure 2-15: Depth subjective quality for Orbi sequence at 2Mbit/s, (a) MPEG4-MAC (b) MPEG4-Separate ............................................................... 33
- Figure 2-16: Block diagram of MPEG4-MAC-DSUS2 ........................................ 34
- Figure 2-17: Rate-Distortion curves for 'Orbi' sequence (a) Luminance (b) Depth ............ 35
- Figure 2-18: Rate distortion for depth information of Orbi sequence at the same bit rate 36
List of Figures

Figure 3-1: Data partitioning in MPEG-4 ................................................................. 39
Figure 3-2: Subjective quality texture (left), depth (middle) and 3D video (right) when only... 41
Figure 3-3: Depth data partitioned ............................................................................ 42
Figure 3-4: Block diagram of MDC encoder and decoder ................................. 43
Figure 3-5: Index assignment table for MD quantisation ...................................... 45
Figure 3-6: Coding and decoding process for PCT .............................................. 46
Figure 3-7: Quincunx sub-sampling .................................................................... 46
Figure 3-8: Even and odd frames prediction ....................................................... 48
Figure 3-9: The proposed MDC-EO encoder and decoder block diagram ............. 49
Figure 3-10: Contents of stream 1 and 2 at frame level ...................................... 50
Figure 3-11: Contents of stream at macro-block level ........................................ 50
Figure 3-12: MDC-EO encoder .............................................................................. 51
Figure 3-13: MDC-EO decoder .............................................................................. 51
Figure 3-14: Performance of MDC-EO for Orbi sequence over UMTS channel, ER means the MDC-EO with the error resilience option enabled .................. 52
Figure 3-15: Performance of MDC-EO Interview sequence over UMTS channel, ER means the MDC-EO with the error resilience option enabled .......... 53
Figure 3-16: Rate-Distortion curves for 'Orbi' sequence (a) Colour image sequence (b) Depth image sequence ................................................................. 54
Figure 3-17: Recovered video quality for Interview sequence for single loss (a) luminance (b) depth ........................................................................... 56
Figure 3-18: Recovered video quality for Interview sequence for Burst loss (a) luminance (b) depth ................................................................. 57
Figure 3-19: Recovered video quality for Interview sequence for two burst losses (a) luminance (b) depth ................................................................. 58
Figure 3-20: Mean PSNR vs packet loss for Orbi (a) luminance and (b) depth .......... 61
Figure 3-21: Mean PSNR vs packet loss for Interview (a) luminance and (b) depth .......... 62
Figure 3-22: Frame PSNR for Orbi (a) luminance and (b) depth at 20% packet loss ........ 63
Figure 3-23: Frame PSNR for Interview (a) luminance and (b) Interview depth at 20% packet loss ................................................................. 64
Figure 3-24: Subjective quality - Interview - at 20% packet loss of luminance for (a) SDC and (b) MDC-EO; and depth for (c) SDC and (d) MDC-EO ..................... 65
Figure 3-25: Subjective quality - Interview - at 20% packet loss of stereoscopic 3D video for (a) SDC and (b) MDC-EO ................................................................. 65
Figure 4-1: Frame Interpolation (a) previous frame (b) blurred interpolated frame (especially in the highlighted box) (c) next frame .................................................. 69
Figure 4-2: The proposed MDC-EOS encoder general block diagram ................. 70
List of Figures

Figure 4-3: The proposed MDC-EOS decoder general block diagram .......................................... 70
Figure 4-4: Block diagram of the proposed MDC-EOS encoder .................................................. 72
Figure 4-5: Block diagram of the proposed MDC-EOS decoder ................................................... 73
Figure 4-6: Content of bit stream at macro-block level for (a) Central information and (b) Side information ............................................................................................................................. 73
Figure 4-7: The proposed MDC-EOSP block diagram .................................................................. 74
Figure 4-8: Detailed structure of MDC-EOSP encoder ................................................................. 75
Figure 4-9: The proposed MDC-EOSB block diagram ................................................................. 77
Figure 4-10: Detailed structure of MDC-EOSB encoder ............................................................... 78
Figure 4-11: Rate-Distortion curves for ‘Orbi’ sequence (a) Colour image sequence (b) Depth image sequence ...................................................................................................................... 80
Figure 4-12: Objective quality (frame PSNR) at Eb/No=8dB of luminance (a); and depth (b) .... 82
Figure 4-13: Subjective quality at Eb/No=8dB of luminance for (a) SDC and (b) MDC-EOSB; and depth for (c) SDC and (d) MDC-EOSB ................................................................. 82
Figure 4-14: Rate distortion when side quantisation parameter (Qpside) is varied for Orbi sequence for (a) luminance and (b) depth ................................................................. 84
Figure 4-15: Recovered video quality for Interview sequence for single loss (a) luminance (b) depth ..................................................................................................................................... 85
Figure 4-16: Recovered video quality for Interview sequence for burst loss (a) luminance (b) depth ..................................................................................................................................... 86
Figure 4-17: Recovered video quality for Interview sequence for two burst losses (a) luminance (b) depth ..................................................................................................................................... 87
Figure 4-18: Recovered video quality for Interview sequence when only the even stream is received (a) luminance (b) depth ..................................................................................................................................... 89
Figure 4-19: Same bit rate experiment for Interview sequence in error prone environments for (a) luminance and (b) depth at 1100 kbit/s ..................................................................................................................................... 92
Figure 4-20: Example of frame PSNR at Eb/No=7dB for Texture ..................................................... 93
Figure 4-21: Subjective quality at Eb/No=7dB of luminance for (a) SDC and (b) MDC-EOS; and depth for (c) SDC and (d) MDC-EOS ................................................................. 93
Figure 4-22: Same quality experiment for Interview sequence in error prone, (a) Texture and (b) Depth ..................................................................................................................................... 94
Figure 4-23: Comparison of SDC and MDC algorithms in UMTS channel for Orbi sequence (a) luminance (b) depth ................................................................. 96
Figure 4-24: Comparison of SDC and MDC algorithms in UMTS channel for Interview sequence (a) luminance (b) depth ................................................................. 97
Figure 4-25: Mean PSNR vs packet loss for Orbi (a) luminance and (b) depth ..................................... 99
Figure 4-26: Mean PSNR vs packet loss for Interview (a) luminance and (b) depth ................. 100
Figure 4-27: Variation of A, the proportion of a packet occupied by the first partition, over the (a) Orbi and (b) Interview sequence .......................................................................................................................... 102
Figure 4-28: Rate-Distortion curves for 'Orbi' sequence (a) Colour image sequence (b) Depth image sequence .................................................................................................................... 104
Figure 4-29: Rate-Distortion curves for MDC-EOS and MDC-EOAS for 'Orbi' sequence (a) Colour image sequence (b) Depth image sequence ........................................................................................................... 105
Figure 4-30: Rate-Distortion curves for MDC-EOSP and MDC-EOASP for 'Orbi' sequence (a) Colour image sequence (b) Depth image sequence .......................................................................................................................... 106
Figure 4-31: Recovered video quality for Interview sequence for single loss (a) luminance (b) depth ......................................................................................................................................................... 107
Figure 4-32: Recovered video quality for Interview sequence for burst loss (a) luminance (b) depth ......................................................................................................................................................... 108
Figure 4-33: Recovered video quality for Interview sequence for two burst losses (a) luminance (b) depth ......................................................................................................................................................... 109
Figure 4-34: Recovered video quality for Interview sequence when only even stream is received (a) luminance (b) depth ......................................................................................................................................................... 111
Figure 4-35: Recovered video quality for Interview sequence for two burst losses (a) luminance (b) depth for SDC, MDC-EO, MDC-EOS and MDC-EOAS .......................................................................................................................... 112
Figure 4-36: Recovered video quality for Interview sequence when only even stream is received (a) luminance (b) depth for MDC-EO, MDC-EOS and MDC-EOAS .......................................................................................................................... 113
Figure 4-37: Recovered video quality for Interview sequence for two burst losses (a) luminance (b) depth for SDC, MDC-EO, MDC-EOSP and MDC-EOASP .......................................................................................................................... 114
Figure 4-38: Recovered video quality for Interview sequence when only even stream is received (a) luminance (b) depth for SDC, MDC-EO, MDC-EOSP and MDC-EOASP .......................................................................................................................... 115
Figure 4-39: Rate-Distortion curves for 'Orbi' sequence (a) Colour image sequence (b) Depth image sequence when subjected to Wireless LAN error pattern of signal-to-noise ratio (SNR) of 4.2 .......................................................................................................................... 118
Figure 4-40: Mean PSNR vs packet loss for Interview (a) luminance and (b) depth ......................................................................................................................................................... 120
Figure 4-41: Mean PSNR vs packet loss for Orbi (a) luminance and (b) depth ......................................................................................................................................................... 122
Figure 4-42: Subjective quality – Interview - at 20% packet loss of luminance for (a) SDC and (b) MDC-EO (c) MDC-EOAS and (d) MDC-EOASP .......................................................................................................................... 123
Figure 4-43: Subjective quality – Interview - at 20% packet loss of depth for (a) SDC and (b) MDC-EO (c) MDC-EOAS and (d) MDC-EOASP .......................................................................................................................... 124
Figure 4-44: Subjective quality – Interview - at 20% packet loss of stereoscopic 3D video for (a) SDC (b) MDC-EO (c) MDC-EOAS (b) MDC-EOASP .......................................................................................................................... 125
Figure 5-1: Block diagram of a scalable encoder using a multi-scale pyramid with 3 levels of spatial scalability [106] .......................................................................................................................... 129
List of Figures

Figure 5-2: Artifacts from scalable MDC proposed in [99] under ideal MDC channel ....................... 130
Figure 5-3: Block diagram of MDC-FGS ...................................................................................... 132
Figure 5-4: Rate-Distortion curves for 'Orbi' sequence (a) Colour image sequence (b) Depth 
image sequence .......................................................................................................................... 133
Figure 5-5: Coding texture and depth information using SVC encoder ......................................... 134
Figure 5-6: Proposed scalable MDC .............................................................................................. 135
Figure 5-7: Contents of stream 1 and 2 at frame level ................................................................. 135
Figure 5-8: Example of error in one of MDC stream ..................................................................... 136
Figure 5-9: Rate distortion of MDC vs SDC for Interview (average PNSR_lumin) ..................... 138
Figure 5-10: Rate distortion of MDC vs SDC for Orbi (average PSNR_lumin) ......................... 138
Figure 5-11: Subjective results for frame 80 of interview sequence when subjected to 20% packet 
loss: (a) SDC; (b) MDC-repeat; (c) MDC-interpolate ................................................................ 140
Figure 5-12: Performance of scalable MDC compared to SDC in WiMAX for Orbi ..................... 142
Figure 5-13: Performance of scalable MDC compared to SDC in WiMAX for Interview ............. 142
Figure 5-14: Subjective results for frame 23 of the Orbi sequence for 2D video: (a) SDC; (b) 
MDC; and for stereoscopic 3D video: (c) SDC; (d) MDC when subjected to 10 dB SNR 
WiMAX channel ...................................................................................................................... 143
Figure 5-15: Rate distortion for Orbi sequence for SDC, MDC and MDC-EOASP .................... 145
Figure 5-16: Rate distortion for Interview sequence for SDC, MDC and MDC-EOASP .......... 145
Figure 5-17: Error prone performance of the algorithms using Table 5-7 configuration for Orbi 146
Figure 5-18: Error prone performance of the algorithms using Table 5-8 configuration for 
Interview ................................................................................................................................... 147
Figure 5-19: Error prone performance of the algorithms using Table 5-9 configuration for Orbi 148
Figure 5-20: Error prone performance of the algorithms using Table 5-10 configuration for 
Interview ................................................................................................................................... 149
Figure 5-21: Block diagram of SVC encoder with DSUS .............................................................. 150
Figure 5-22: Rate distortion for the depth information of Orbi image sequence ......................... 151
Figure 5-23: Disocclusion artefacts from DIBR implementation on Interview sequence (Left-
image) ........................................................................................................................................ 152
Figure 5-24: Rate distortion for the Orbi image sequence (average left-and-right PSNR)........... 152
Figure 5-25: Rate distortion for the Interview image sequence (average left-and-right PSNR).. 153
Figure 5-26: Rate distortion comparison of DSUS application to SVC for Orbi sequence ........... 154
Figure 5-27: Rate distortion comparison of DSUS application to SVC for Interview sequence. 154
Figure 5-28: Block diagram of SVC-MDC encoder with DSUS (SVC-MDC-DSUS) ............... 155
Figure 5-29: Rate distortion for the depth information of Orbi image sequence ......................... 156
Figure 5-30: Rate distortion for the depth information of Interview image sequence .................. 157
Figure 5-31: Rate distortion for the Orbi image sequence (average left-and-right PSNR)........... 157
Figure 5-32: Rate distortion for the Interview image sequence (average left-and-right PSNR).. 158
Figure 5-33: Rate distortion for the luminance information of the Orbi image sequence (average PSNR versus overall bit rates) ................................................................. 159
Figure 5-34: Rate distortion for the depth information of the Orbi image sequence (average PSNR versus overall bit rates) ................................................................. 159
Figure 5-35: Rate distortion for the depth information of the Orbi image sequence (average PSNR versus depth bit rates) ................................................................. 160
Figure 5-36: Rate distortion for the Orbi image sequence (average left-and-right PSNR)........ 161
Figure 5-37: Rate distortion for the Interview image sequence (average left-and-right PSNR) .. 161
Figure 5-38: Rate distortion for Orbi sequence at 3% packet loss ....................................... 163
Figure 5-39: Rate distortion for Orbi at 5% packet loss ......................................................... 163
Figure 5-40: Rate distortion for Orbi sequence at 10% packet loss ....................................... 164
Figure 5-41: Rate distortion for Orbi sequence at 20% packet loss ....................................... 164
Figure 5-42: Rate distortion for Interview sequence at 3% packet loss .................................. 165
Figure 5-43: Rate distortion for Interview sequence at 5% packet loss .................................. 165
Figure 5-44: Rate distortion for Interview sequence at 10% packet loss.............................. 166
Figure 5-45: Rate distortion for Interview sequence at 20% packet loss.............................. 166
Figure 5-46: Stereoscopic 3D video of the Interview sequence, (a) Original, and at qp=35 for 5% packet loss, (b) SVC-Org (c) SVC-DSUS (d) SVC-MDC-Org and (e) SVC-MDC-DSUS 168
Figure 5-47: Packet loss performance for the Orbi sequence at qp=35 ................................. 169
Figure 5-48: Packet loss performance for the Interview sequence at qp=35 ........................... 169
Figure A-6-1: UMTS Physical layer model ................................................................. 191
Figure B-6-2: Baseband system model of IEEE 802.11g .................................................... 192
Figure C-6-3: Physical layer of IEEE802.16e standard ........................................... 193
List of Abbreviations

2D  Two dimensions
3D  Three dimensions
3G  3rd Generations
3DTV  3D Television
AIR  Adaptive Intra Refresh
B-frame  Bidirectional predicted frame
CC  Convolutional Code
CCSR  Centre for Communication System Research
CDN  Content Delivery Network
CIF  Common Intermediate Format
CIR  Cyclic Intra Refresh
CRC  Cyclic Redundancy Check
CTC  Convolutional Turbo Coding
DCT  Discrete Cosine Transform
DIBR  Depth Image Based Rendering
DSUS  Down-Sampling and Up-Sampling
DL  Down Link
DLP  Digital Light Processing
DVB  Digital Video Broadcast
DVB-C  Digital Video Broadcast over Cable
DVB-S  Digital Video Broadcast over Satellite
DVB-T  Digital Video Broadcast over Terrestrial
DVC  Distributed Video Coding
EO  Even and odd frames
FEC  Forward Error Correction
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FGS</td>
<td>Fine Granularity Scalability</td>
</tr>
<tr>
<td>HDTV</td>
<td>High Definition Television</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>ITU</td>
<td>International Telecommunication Union</td>
</tr>
<tr>
<td>LOS</td>
<td>Line Of Sight</td>
</tr>
<tr>
<td>MAC</td>
<td>Multiple Auxiliary Component</td>
</tr>
<tr>
<td>MCS</td>
<td>Modulation Coding Scheme</td>
</tr>
<tr>
<td>MDC</td>
<td>Multiple Description Coding</td>
</tr>
<tr>
<td>MDC-EO</td>
<td>Even and odd frame based MDC</td>
</tr>
<tr>
<td>MDC-EOS</td>
<td>Even and odd frames based MDC with side information</td>
</tr>
<tr>
<td>MDC-EOAS</td>
<td>Even and odd frames based MDC with adaptive side information</td>
</tr>
<tr>
<td>MDC-EOSB</td>
<td>Even and odd frames based MDC using B-frame</td>
</tr>
<tr>
<td>MDC-EOSP</td>
<td>Even and odd frames based MDC with side information and prediction</td>
</tr>
<tr>
<td>MDC-EOASP</td>
<td>Even and odd frames based MDC with adaptive side information and prediction</td>
</tr>
<tr>
<td>MPEG</td>
<td>Moving Picture Expert Group</td>
</tr>
<tr>
<td>MPT</td>
<td>Multiple Path Transport</td>
</tr>
<tr>
<td>NALU</td>
<td>Network Abstraction Layer Unit</td>
</tr>
<tr>
<td>P2P</td>
<td>Peer-to-Peer</td>
</tr>
<tr>
<td>PDA</td>
<td>Personal Digital Assistant</td>
</tr>
<tr>
<td>PSNR</td>
<td>Peak Signal to Noise Ratio</td>
</tr>
<tr>
<td>PUSC</td>
<td>Partially Used Sub Channelisation</td>
</tr>
<tr>
<td>QAM</td>
<td>Quadrature Amplitude Modulation</td>
</tr>
<tr>
<td>QCIF</td>
<td>Quarter Common Intermediate Format</td>
</tr>
<tr>
<td>QP</td>
<td>Quantisation Parameter</td>
</tr>
<tr>
<td>RVLC</td>
<td>Reversible Variable-Length Codes</td>
</tr>
<tr>
<td>SDC</td>
<td>Single Description Coding</td>
</tr>
<tr>
<td>SEI</td>
<td>Supplemental Enhancement Information</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal to Noise Ratio</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>SVC</td>
<td>Scalable Video Coding</td>
</tr>
<tr>
<td>TV</td>
<td>Television</td>
</tr>
<tr>
<td>UMTS</td>
<td>Universal Multimedia Telecommunication Services</td>
</tr>
<tr>
<td>VLC</td>
<td>Variable Length Coding</td>
</tr>
<tr>
<td>WiMAX</td>
<td>Worldwide Interoperability for Microwave Access</td>
</tr>
<tr>
<td>WLAN</td>
<td>Wireless Local Area Network</td>
</tr>
</tbody>
</table>
List of Mathematical Symbols

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$</td>
<td>Viewing distance</td>
</tr>
<tr>
<td>$Ec(n)$</td>
<td>Residual error for current frame</td>
</tr>
<tr>
<td>$Ecq(n)$</td>
<td>Coded residual error</td>
</tr>
<tr>
<td>$Ee(n)$</td>
<td>Side information residual error</td>
</tr>
<tr>
<td>$Eeq(n)$</td>
<td>Coded side information residual error</td>
</tr>
<tr>
<td>$F'(n-1)$</td>
<td>Previous reconstructed odd frame</td>
</tr>
<tr>
<td>$Fc(n)$</td>
<td>Current frame</td>
</tr>
<tr>
<td>$Fc'(n)$</td>
<td>Predicted current frame</td>
</tr>
<tr>
<td>$fd(i,j)$</td>
<td>Down-sampled image</td>
</tr>
<tr>
<td>$Fe'(n)$</td>
<td>Current reconstructed even frame</td>
</tr>
<tr>
<td>$Fe'(n-2)$</td>
<td>Previous reconstructed even frame</td>
</tr>
<tr>
<td>$Fo'(n)$</td>
<td>Current reconstructed odd frame</td>
</tr>
<tr>
<td>$Fip$</td>
<td>Interpolation of $Fe'(n)$ and $Fe'(n-2)$</td>
</tr>
<tr>
<td>$Fip'$</td>
<td>Decoded $Eeq(n) + Fip$</td>
</tr>
<tr>
<td>$Fp(n-1)$</td>
<td>B-frame interpolation of $Fe'(n)$ and $Fe'(n-2)$</td>
</tr>
<tr>
<td>$fru(i,j)$</td>
<td>Up-sampled reconstructed image</td>
</tr>
<tr>
<td>$fr(i,j)$</td>
<td>Reconstructed image</td>
</tr>
<tr>
<td>$h$</td>
<td>Screen parallax</td>
</tr>
<tr>
<td>$I(x,y)$</td>
<td>An image at coordinate $(x,y)$</td>
</tr>
<tr>
<td>$I_p(i,j)$</td>
<td>Frame to be interpolated at pixel location $(i,j)$</td>
</tr>
<tr>
<td>$I_{prev}(i,j)$</td>
<td>Previous frame</td>
</tr>
<tr>
<td>$I_{fut}(i,j)$</td>
<td>Future frame</td>
</tr>
<tr>
<td>$k_{near}$</td>
<td>Range of depth behind camera</td>
</tr>
<tr>
<td>$k_{far}$</td>
<td>Range of depth in front camera</td>
</tr>
<tr>
<td>$m$</td>
<td>Value of depth map</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>$N_{pix}$</td>
<td>Number of horizontal pixels of the display</td>
</tr>
<tr>
<td>$N_x$</td>
<td>Width of the video frame</td>
</tr>
<tr>
<td>$N_y$</td>
<td>Height of the video frame</td>
</tr>
<tr>
<td>$Q$</td>
<td>Quantiser</td>
</tr>
<tr>
<td>$X_\theta$</td>
<td>Eye separation</td>
</tr>
<tr>
<td>$X_{SIZE}$</td>
<td>Vertical size of image</td>
</tr>
<tr>
<td>$Y_{SIZE}$</td>
<td>Horizontal size of image</td>
</tr>
<tr>
<td>$Z$</td>
<td>Depth map</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Preamble

Communication using three dimensional (3D) video is an enhanced representation method for current two dimensional (2D) video. 3D video differs from 2D in the sense that it also accounts for depth information. It will allow users to feel the presence of the persons they are communicating with or be truly immersed in the event they are watching.

Stereoscopic video is the simplest form of 3D video and can be adapted in communication applications with the support of existing video technologies. Stereoscopic video renders two views for each eye, which facilitates depth perception of the 3D scene. A new form of stereoscopic video is developed in [1], which renders ordinary 2D video from an ordinary 2D video camera and depth information from a laser range or depth camera [2], producing two views to the user, one for each eye, creating the 3D impression. This type of 3D video (2D video plus depth information) can be compressed using standard video codecs such as MPEG4 and H.264 [3].

Compressed 3D video is intended for transmission over networks such as Internet, wireless networks and broadcast (e.g. terrestrial and satellite). Since coded 3D video data is highly sensitive to information loss and channel bit errors, error resilience techniques are needed. Many error resilience methods are available in the literature. One of them is known as MDC. The objective of the MDC is to encode a source into two or more bit streams for transmission over a communication system with multiple channels. A high-quality reconstruction is decodable when all bit streams are received correctly, while a lower, but still acceptable quality reconstruction is achievable if only one stream is received. MDC has been applied to 2D video to provide error resilience. In this thesis, the application of MDC to 3D video (specifically 2D video plus depth information) is investigated over error free and error prone environment.
1.2 Motivation

This PhD research is motivated by applications that require 3D video, for example in entertainment (such as movies and video games), training (such as military/non-military pilot and car driving training), in amusement park (such as virtual roller coaster) and communication (such as 3D video conferencing and mobile 3D television). 3D video is not much used in communication as it consumes more bandwidth compared to 2D video. Hence it is more prone to transmission errors. It also needs special displays or special glasses that can be expensive and inconvenient for users. However, recently, research on 3D-TV and 3D audio and video communication such as [4], [15] and [11] is being conducted to provide 3D television contents to the users and to explore the possibilities of 3D video communication.

Over the years, many 3D displays have been developed and users do not have to wear special glasses anymore [5]. 3D displays for 3D television (3DTV) are also available in the market such as Philips (Philips 42” 3D Intelligent Display Solutions), which offer auto-stereoscopic 3D display. However, commercial 3DTV with special glass is still available for example the Samsung 3D HDTV, which offers 3DTV with shutter glass. 3D mobile phone is also being built, by companies, such as Dynamic Digital Depth [6], allowing us to communicate in 3D environment. Some companies in film industries are also producing 3D movies to enhance theatre experience [7]. Consumer products, visualised using 3D, are likely to grab customers’ attention and bring more businesses [8].

Combining 3D video with MDC is especially promising for 3D video applications where retransmission is unacceptable (such as interactive real time 3D video phone and 3D video conferencing). The retransmission is often not acceptable as it incurs long delay. In this situation, MDC can provide adequate quality without requiring retransmission of any lost packets, as long as one of the MDC descriptions is received by the receiver. Thus, MDC can simplify network design because feedback is not necessary needed and all the packets can be treated equally.

MDC also has advantage over layered video coding or scalable video coding that produces a base layer and one or more enhancement layers. The base layer must be delivered error free to guarantee a basic level of quality. In this case, differential treatments are required from the network to the base layer and the enhancement layers. The base layer may need to be retransmitted which can cause unacceptable delay. In MDC, the descriptions can be treated equally by the network. Either one of the MDC descriptions can provide adequate quality, whereas scalable video coding decoder must have the base layer to provide acceptable quality.

MDC also can effectively relieve congestion at hotspots and increase the overall utilisation of the network resources. These advantages can be made possible by enabling the traffic dispersion and the load balancing in the network, which can be achieved by combining MDC with multiple path
transport (MPT). It is mentioned in [57] that MDC has been effectively integrated with MPT for 2D video transport over different networks, namely, wireless adhoc networks, content delivery networks (CDN) and peer-to-peer networks.

In the wireless adhoc networks simulation, it was found in [57] that transmitting packets from the MDC coder with MPT over two paths has better performance than using MDC coder over single path. In the CDN, MDC is coupled with server diversity to provide video streaming service. It was shown in [57] that the MDC-CDN setup can provide significant performance benefits over a conventional single description coding CDN, even in a CDN that is designed for delivering single description video. In the peer-to-peer network in [57], the video is encoded using MDC and stored in a dedicated server. Many cooperating peers are connected to the server and are used to distribute the live or pre-encoded video to alleviate the overload at the server caused by special event. The MDC video descriptions are distributed over a multiple distribution trees that are spanned over the participating peer nodes.

MDC can be used to improve the error resilience of the compressed 3D video. However, error resilience is not the only issue in modern video transmission. Modern video transmission are typically characterised by a wide range of access network bit rates and receiving devices characteristics that require some sort of 3D video scalability. Scalable video coding is an attractive solution for the issues posed by these scenarios. In this thesis, the combination of scalable coding and multiple description coding (scalable MDC) for 3D video is also investigated to improve error robustness, and at the same time provides adaptability to bandwidth variations and receiving device characteristics.

### 1.3 Application and Scenarios

Example of scenarios where 3D video and MDC can be beneficial includes 3D television (3DTV) broadcast, mobile 3DTV, 3D video streaming and virtual collaboration 3D video conferencing.

#### 1.3.1 3DTV Broadcast

3DTV displays already available commercially in the market. Example of 3DTV display from Philips is shown in Figure 1-1. With 3D content, soon consumers will be able to watch 3DTV in the comfort of their home. The 3DTV may be enabled using Digital Video Broadcast (DVB) technologies that include digital television signal broadcast over satellite (DVB-S, DVB-S2 and DVB-SH), cable (DVB-C) and terrestrial (DVB-T).
Chapter 1. Introduction

3DTV over DVB-T was demonstrated at the International Broadcast Convention (IBC) in 2004. In [16], MDC in combination with scalable video coding (MDC-SVC) is used to provide error resilience 2D video content over DVB-T/H and WiMAX channel. MDC-SVC will be capable of delivering 2D video and in the future 3D video content over best-effort error prone networks. Due to its scalable erasure-resilient compression capabilities, MDC-SVC with 3D video should be able to (1) meet the users' requirement in term of quality and resolution; (2) dynamically adapt the rate to the available channel capacity; (3) provide robustness to data losses as retransmission is often impractical, and (4) provide 3D for immersive entertainment and communication.

1.3.2 Mobile 3DTV

Current handheld devices are already able to record and display 2D video. One example of these handheld devices is shown in Figure 1-2. With the recent advance in camera and display technologies, 3D capturing and 3D auto-stereoscopic display for handheld devices should be possible. Mobile TV involves bringing TV services to the mobile phones. It combines the services of a mobile phone with television content.
The mobile TV service is already available in Korea and recently in United Kingdom. Technically, there are currently two main ways of delivering mobile TV. The first is via a two-way cellular network (e.g. 3G networks) and the second is through a one-way dedicated broadcast network (e.g. digital video broadcasting-handheld (DVB-H)). Example includes the mobile TV service offered by T-Mobile in United Kingdom [9]. Most of the mobile network operators started to use IP network successfully to offer wireless video services. Extension of the service to 3D video content seems to be possible. One way of delivering the 3DTV content over wireless and mobile IP network is by using peer-to-peer (P2P) networks [17]. As mentioned before, MDC can be used to deliver error resilience 2D video over P2P networks. Thus, the MDC method can be extended to 3D video for improved performance.

1.3.3 3D Video Streaming

Figure 1-3: Stereoscopic 3D video streaming over Internet

Figure 1-3 shows a snapshot of 3D video streaming over Internet which uses IP network. In [17], the 3DTV streaming architecture can be classified as: 1) server unicasting to a single client; 2) server multicasting to several clients; 3) P2P unicasting; where each peer forwards packets to another peer; and 4) P2P multicasting, where each peer forwards packet to several other peers. These types of architecture may use multiple path transport (MPT) to ease the network load and relieve congestion at hotspots. Combining MDC with MPT to deliver 3D video streaming over IP networks should be able to offers both error resilience and load balancing, and is applicable in both wired and wireless networks.

1.3.4 Virtual Collaboration 3D Video Conferencing

A virtual collaboration system scenario, as shown in Figure 1-4, consists of a large, fixed-terminal acts as the main control/commanding point and serves a group of co-located users. This may be
the headquarters of the organization and consists of communication terminals, shared desk spaces, displays and various user interaction devices to collaborate with remotely located partners. The remotely located users with a small, fixed terminal will act as the local contact and provide the local information. Mobile units (distribution, surveying, marketing, patrolling, etc) of the organization will use mobile terminals, such as mobile phones and PDAs, to collaborate with the headquarters.

![Virtual Collaboration System Application](image)

**Figure 1-4: Virtual collaboration system application**

In such system, numbers of users are varying, each with their own time varying data throughput requirements, adaptively share network resources, resulting in time varying connection qualities. The users possess a variety of devices with different capabilities, ranging from cell phones with small screens and restricted processing power, to high-end PCs with high-definition displays. One example of such system is the one developed by Thales Research & Technology UK Limited in iLAB, CCSR, University of Surrey [10].

In [18], MDC in combination with scalable video coding provides a robust video coding solution for this virtual collaboration application. A key aspect of the collaboration is audiovisual conferencing. 3D scalable MDC video can provide significant benefits for the videoconferencing in terms of: 1) 3D: a more immersive communication experience; 2) scalability: adaptability to different terminal types; and 3) MDC: improved robustness to packet losses.

### 1.4 Objectives and Overall Project Description

The research in this thesis involves 3D visual data compression for transmission over error prone networks. Issues such as resilience and scalability have to be taken into account to mitigate the effects of transmission errors and information loss on the compressed 3D video stream.
Chapter 1. Introduction

The main research objective of this thesis is to tackle error resilience and scalability issues of 3D video signals for transmission over networks. For this purpose, the compression of 3D video signals has to be investigated. Compression tools available in recent video coding standards such as MPEG4 and H.264 are explored to compress the 3D video. The compressed 3D video signal is then transmitted over simulated wired/wireless networks.

Since the coded 3D video data is highly sensitive to information loss and channel bit errors in the wireless network, error resilience techniques are needed. MDC has emerged as a promising approach to enhance the error resilience of 2D video delivery system over multiple paths. Hence, one of the objectives of this research is to investigate the application of MDC to 3D video data. Beside video error robustness issues, video scalability is also an important issue in a modern video transmission as it can provide adaptability to bandwidth variations and receiving device characteristics. Thus, to develop a scalable MDC is also one of the objectives of this research. Therefore, the objectives are summarised into following points:

- Apply existing method of 2D video compression, such as MPEG4 to the 3D video.
- Develop, apply and test a method of error resilience, namely MDC on 3D video data, in error free and error prone conditions.
- Improve the developed MDC technique.
- Investigate the scalability aspect of the MDC and its application to 3D video.

1.5 Original Achievements

The research in this thesis has produced a number of publications, which are listed at the end of this section. The work involved extensive studies on the video coding standard, multiple description video coding and scalable multiple description video coding for 3D video application. The work which is believed to be original is listed below, categorised into three main areas:

- 3D video compression
  - Compression of depth information using DSUS method (Chapter 2).
  - Compression of 2D and depth using MPEG4-MAC (Chapter 2).
  - Application of DSUS to the depth information in MPEG4-MAC (Chapter 2).

- 3D video error resilience
  - Extension of RVLC and data partitioning to depth information (Chapter 3).
  - Application of even and odd frame based MDC to 3D video (Chapter 3).
Chapter 1. Introduction

- Development of even and odd frame based MDC with side information (Chapter 4).
- Development of even and odd frame based MDC with adaptive side information (Chapter 4).

- 3D video scalability
  - Development of even and odd frame based MDC with scalable side information using Fine Granularity Scalability (FGS) (Chapter 5).
  - Development of scalable MDC base on even and odd frames MDC for SVC (Chapter 5).
  - Application of DSUS to the depth information of scalable video coding (Chapter 5).
  - Application of DSUS to the depth information of scalable multiple description video coding (Chapter 5).

1.6 Structure of Thesis

Chapter 2 presents background information on 3D video, specifically the stereoscopic video in terms of its production, source material, quality assessment and coding. Classification of 3D video is presented and stereoscopic video generated from 2D video and depth information is described. It also describes ways to view stereoscopic video. This chapter also mentions about the quality evaluation of 3D video. Some background information on coding the 3D video is presented in this chapter. Furthermore, this chapter presents our contribution in coding the 2D video and depth information using the available video coding standard by compressing the depth information at reduced resolution. Compression of the 2D video and depth information (and later reduced resolution depth information) using MPEG4-MAC is also investigated in this chapter.

In Chapter 3, the compressed 3D video (2D plus depth video) are intended for transmission over communication networks. Error resilience tools are needed to suppressed the effect of channel errors on the compressed 3D video. Many error resilience methods are available in the literature. This chapter briefly describe the error resilience tools included by the MPEG-4 version 2 standard. The error resilience tools are extended to the depth data in 3D video. Next, this chapter introduces MDC as a form of error resilience tools. Finally, an even and odd frame based MDC method (MDC-EO) was developed and applied to 3D video.

Chapter 4 proposes a novel multiple description coding for 3D video based on even and odd frames MDC developed in Chapter 3. The even and odd frames based MDC is improved by
adding variable redundancies in the form of side information (MDC-EOS). The redundant side information consists of the difference between the interpolated frame and the locally reconstructed frame that can be quantised, hence, the redundancies can be controlled by the quantisation parameter. This method is improved by including the side information in the central prediction and using the concept of multiple predictions (MDC-EOSP).

It was found that the side information is quite large which affects the coding efficiency of the MDC coder. Hence, an attempt is made to reduce the side information using B-frame interpolation (MDC-EOSB). The simulation results for MDC-EOS, MDC-EOSP and MDC-EOSB are presented and discussed in this chapter. The side information is also reduced by making it adaptive according to the motion information and is applied to MDC-EOS and MDC-EOSP. With the adaptive side information MDC-EOS is renamed to MDC-EOAS and MDC-EOSP is renamed to MDC-EOASP. The simulation results in error free and error prone environment for MDC-EOAS and MDC-EOASP are also presented and discussed.

In Chapter 5, the combination of scalable coding and multiple description coding (scalable MDC) is proposed to improve error robustness, and at the same time provides adaptability to bandwidth variations and receiving device characteristics. For more immersive communication experiences, stereoscopic 3D video content will be used as a source for the scalable multiple description coder. Related works on the scalable video coding, the scalable extension of the H.264/AVC (SVC) and the available scalable MDC video coding schemes are firstly reviewed in this chapter. It is then followed by a description of our proposed scalable MDC to 3D video, including results and discussion. Two methods of scalable MDC are proposed, one is based on the MPEG4-MAC-MDC and the other one is based on the SVC. This chapter also discusses the application of DSUS from Chapter 2 to the scalable video coding and the scalable MDC. Performance comparison in error free environment of SVC, SVC with DSUS, scalable MDC and scalable MDC with DSUS is analysed. The performance of these algorithms in error prone channel is also presented in this chapter.

The last Chapter presents the overall conclusion of the thesis leading to some suggestions for future work.
1.7 Publications

A list of publications produced during the research work is as follows:


Chapter 2

3D Video Coding

2.1 Introduction

This chapter presents background information on 3D video, specifically the stereoscopic video in terms of its production, source material, quality assessment and coding. A classification of 3D video is presented in Section 2.2. Stereoscopic video generated from 2D video and depth information is described in Section 2.3. It also describes ways to view stereoscopic video. Generation of stereoscopic video from 2D video and depth is described in Section 2.4 and examples of 2D video and depth image sequences are given in Section 2.5. The quality evaluation of 3D video is discussed in Section 2.6. The coding of stereoscopic 3D video using the available video coding standard is presented in Section 2.7. Our contribution in term of compression of depth information at reduced resolution is investigated in Section 2.8. Furthermore, our contribution in term of compression of 2D video and depth using MPEG4-MAC is explored in Section 2.8. Section 2.9 concludes the chapter and suggests the way forward.

2.2 3D Video Classification

There are many variations of 3D video in the literature, for example, the MPEG 3DAV AdHoc researchers categorized 3D video according to its applications, namely omni-directional video, free viewpoint video and interactive stereoscopic video [11].

Omni-directional video will allow users to have look-around view and zoom from one viewpoint of a scene. Free viewpoint video is generated using multiple cameras such that a scene can be viewed from any direction depending on the cameras configurations. The scene may be composed of one or more 3D video object. Interactive stereoscopic video is essentially the free viewpoint video that uses two cameras. In stereoscopic video two views are created from the two cameras and each view is displayed separately to our eye. Interactivity can also be achieved in stereoscopic video if the parameters relating to the 3D impression can be adjusted such as adjusting the depth perception. Currently our research interest is in the stereoscopic video since it is the simplest form of 3D video and many materials and tools on stereoscopic video are already available. It is also most likely to be the first type of 3D video used by the user.
2.3 Stereoscopic Video

Currently, stereoscopic video contents can be produced by using a stereo camera pair, or a depth range camera, or by converting existing 2D video into stereoscopic 3D [12]. The stereo camera will result in separate left and right eye view video.

The depth camera [2] produces 2D video plus depth map information. The depth information tells us how far each pixel is from the camera. This depth information together with the ordinary 2D video data can be manipulated using a Depth Image Based Rendering (DIBR) technique [13] to produce two views to the user, one for each eye, creating a 3D impression.

Figure 2-1 shows an example of 2D video data and its associated depth information taken from the ATTEST project [15]. Notice that the depth image has different level of darkeners. The darkest colour shows that the object is farthest from the camera. The brightest colour shows that the object is nearer to the camera.

The depth camera generates the luminance and chrominance component of the scene and also produces the depth information. The depth camera sends out pulsed light or light wall to the object. As the light hits the objects it is reflected back towards the camera carrying an imprint of the objects. The depth sensor at the camera measures the exact distance of each pixel in the scene from the reflected light. This distance from the camera can be used to reconstruct a depth map as shown in Figure 2-1 (b).

![Figure 2-1: (a) 2D video and (b) its depth information](image)

Some of the advantages of using 2D video and its associated depth information compared to the classical approach of stereoscopic video are [19]:

- The associated depth information contains a lot of low frequency components. Therefore, it can be compressed much more efficiently. It was reported in [3] that the depth information may be compressed to below 10-20% of the basic colour video bit rate without significantly affecting the quality.
• Since the 3D impression is related to how much the depth is rendered using the depth image based rendering technique, the viewer can adjust the reproduction of the 3D depth to suit his/her personal preferences. This is because there is difference in depth appreciation over age groups [20]. For example, older adults are less sensitive than younger adults to perceive stereoscopic depth.

• As both views are synthesized from the same original image in 2D plus depth camera, photometrical asymmetries such as colour and brightness, between the left- and right-eye views in two-views-camera are eliminated. The photometrical asymmetries can destroy the stereoscopic 3D sensation

• Disparity maps between left and right view produced in traditional stereoscopic video is not needed for 2D plus depth video. This will save computation time as the encoder does not need to perform disparity estimation. However, DIBR has to be implemented at the decoder. Disparity estimation, which produces the disparity map, is a process to determine homologous image points between left and right stereo view [21]. Hence, only the left image and the disparity map are compressed. At the decoder, the right image is reconstructed using the reconstructed left image and the reconstructed disparity map [22].

The disadvantages of using the 2D video and depth camera are:

• The quality of the 3D view surely depends on the accuracy of the depth information. Depth-image artifacts resulting from compression and transmission have to be investigated

• Objects in the virtual left- and right-eye views may be partially occluded in the original image. Suitable hole-filling techniques have been developed to counter the occlusion problem

• Real-time DIBR algorithm is required for real time playback applications.

In our research, we focused on 3D material from the depth cameras. The 2D+depth image sequences are available at [4]. There are also other stereo video sequences available on the internet (e.g. in [23]). These stereo videos can be converted to 2D plus depth image using, for example a machine learning algorithm as described in [24].

Stereoscopic view/display technology is needed to obtain 3D impression from the stereo video. The objective of the display technology is to display one view from the stereo video to each eye. The existing stereoscopic view/display technology may be categorised as follows [25]:

1. Anaglyph colour filter glasses:
This glass is normally used in old 3D movies. It comes in different colours and styles. Both
eyes have a different colour filter in front of them. Red+blue (or cyan) and yellow+blue are
the most common colour combination. The two views of the same subject are overlaid in
contrasting colours (red and blue for red+blue glasses). The glasses filter out one of the views.
These types of glasses will be used in some of our research simulations to view the 3D video
sequence. It is also the cheapest technology to view stereoscopic 3D video.

2. Polarizing glasses:
The glasses have two lenses which have their polarization directions adjusted to be 90 degrees
different. The 3D material is typically projected using two projectors with polarizing lenses in
front of each projector. The polarizing lenses are adjusted to meet the polarization directions
of the glasses. Hence, the left eye can only view material polarized for left eye and vice versa
to the right eye. This technology is used in 3D movies theatres such as IMAX and is also
known as passive stereo.

3. Shutter glasses:
In this method, the left and right images are alternated rapidly on monitor screen and the
viewer looks through shuttering eyewear. Each shutter is synchronized to the monitor screen
so that the left eye sees only the left view and the right eye only the right view. The shutter
glasses technology is also known as active stereo. These types of glasses will be used in some
of our research simulations to view the 3D video sequence. An example of display with shutter
glasses is installed in ILAB, University of Surrey, UK. It is a 7x2.5 metre semi-cylindrical
rear-projected active stereo display [26].

4. Head mount display (HMD):
HMD is usually a helmet like device that provides separate display for each eye. It is
normally installed with head motion feedback to provide the feelings of inside the 3D world.
It is used in many virtual reality applications.

5. Auto-stereoscopic display:
3D video can also be viewed without the viewing aids above. However a special 3D monitor
or display is required. One example is a 3D display that uses optical filter to differentiate
visual information belonging to one view from that of another view. It then directs each view
to each of a viewer eye generating a 3D impression in viewer mind. Another 3D display
example is by Philips (Philips 42" 3D Intelligent Display Solutions), which offer auto-
stereoscopic 3D display, allowing multiple users to view 3D content at the same time. The
interface to the Philips 3D monitor is based on 2D plus depth information.
2.4 Stereoscopic Video from 2D Video and Depth

As mentioned in Section 2.3, DIBR technique can be used to produce stereoscopic video from the 2D video data and the depth information. The 2D video plus depth format was recently standardised in ISO/IEC 23002-3 (MPEG-C part 3) ([27] and [28]). A format to represent an auxiliary video data map is specified in the MPEG-C part 3 standard which allows encoding depth as conventional 2D sequences. Alpha planes and depth maps are examples of the auxiliary video data map.

In DIBR, given an image \( I \) at location \((x, y)\) and depth map \( z \), a new image \( I' \) at location \((x, y+h)\) on the display can be created by shifting the viewpoint. The resulting pixel shifting on the display, \( h \), is called “screen parallax”. From [28], the pixel parallax, \( h \), is given by Equation (2-1).

\[
h = -x_B \frac{N_{\text{pix}}}{D} \left( \frac{m}{2^N} (k_{\text{near}} + k_{\text{far}}) - k_{\text{far}} \right)
\]

(2-2)

where \( x_B \) is the eye separation and the typical value is 6 cm, \( N_{\text{pix}} \) is the number of horizontal pixels of the display, \( D \) is the viewing distance and the typical living room viewing distance is 250 cm, \( m \) is the value from the depth map, \( z \), represented by an unsigned N-bit value (typically from 0 to 255), \( k_{\text{near}} \) and \( k_{\text{far}} \) specify the range of the depth information respectively behind and in front of the picture relative to the screen width. The default value of \( k_{\text{near}} \) is 2 and \( k_{\text{far}} \) is 8. This is to accommodate the need for a depth range behind the picture that is larger than the range in front of the picture. If we assumed that the image \( I(x, y) \) is the left image, then the image \( I'(x, y+h) \) will be the right image. Displaying these two images separately to both eyes will create the stereoscopic 3D impression. More detailed explanation about the equation can be found in [13], [28] and [29].

The equation was implemented in C programming language software and is used to produce a left and right image sequence from 2D image and its associated depth information of any standard image size. The left and right image sequence can then be played using a Stereoscopic Player to produce a 3D stereoscopic video [30] as shown in Figure 2-2 and can be viewed using one of the 3D viewing glasses mentioned before. The stereoscopic image in Figure 2-2 can be viewed using a red and blue glass.
2.5 2D Video and Depth Test Video Sequences

The MPEG 3D AV Adhoc group has specified some 2D video and depth test video sequences for use in the evaluation of 3DTV coding techniques [31]. Two of the sequences were mainly used for the simulation in this thesis, namely Orbi and Interview. There are also other sequences such as Cup&Glasses, Breakdance and Room3D. Orbi shows a table with a number of static toy objects (a box, a mask of a human, etc) captured by a moving camera in an indoor environment. Interview is a scene of two actors playing police officers, discussing something captured by a static camera in an indoor environment.

For every frame of each sequence, the basic colour information and an 8-bit per-pixel depth map was provided. The grey level in the depth map is inversely proportional to the depth. The depth data was captured using an infrared range camera and was post-processed with a 2-D Gaussian low-pass filter to smooth out large discontinuities. The original image size has a resolution of 720 pixels by 576 lines, a temporal resolution (frame rate) of 25 frames per second, raw YUV 4:2:0, a length of 5 seconds and was created for 3DTV ATTEST project [15]. In this PhD project, the image size is down-sampled to CIF (352x288) and QCIF (176x144) image size using the
algorithm in Section 2.8 for usage in mobile applications that requires reduced bandwidth and portability. To produce longer sequences (30 seconds), when the original sequence ends, the sequences are repeated in reverse order to guarantee a smooth transition and avoid scene change. The original sequences are available at the 3DTV project website [30]. Some original frames extracted from each of the two sequences are depicted in Figure 2-3.

Figure 2-3: Original frames of 2D and depth image sequences at CIF resolution (a) Orbi texture (b) Orbi depth (c) Interview texture (d) Interview depth [30]
2.6 3D Video Quality Assessment and Performance Evaluation

The 3D video quality can be assessed objectively and subjectively. PSNR (peak-signal-to-noise ratio) as defined below can be used for objective measurement.

\[
PSNR = 10 \log_{10} \frac{255^2}{\frac{1}{N_x \times N_y} \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} [I(i,j) - I'(i,j)]^2}
\]  

(2-3)

where \( N_x \) and \( N_y \) are respectively the width and height of the video frame and \( I(i,j) \) and \( I'(i,j) \) are respectively the original and reconstructed pixel luminance or chrominance at position \( (i,j) \). The PSNR for both the luminance of 2D video and depth can be measured using Equation (2-4) resulting in two PSNRs, PSNR_lumin (PSNR for luminance) and PSNR_depth (PSNR for depth). For PSNR_depth, \( I(i,j) \) and \( I'(i,j) \) are respectively the original and reconstructed pixel depth at position \( (i,j) \).

In [29], PSNR for depth is used to measure the coding efficiency of MPEG video codecs on the depth information. It was shown that the depth-images can be compressed to target rates below 10-20 % of the basic colour video bit rate while still maintaining high PSNR values. In this thesis, PSNR_lumin and PSNR_depth are used to measure the objective quality of the 2D video and depth data respectively.

The subjective quality of the 3D video can be evaluated in two ways, first is by viewing the 2D video and depth separately and rate them using for example the double-stimulus subjective test method specified in [29]. With this method, any impairment on the 2D video and depth can be rated by the viewers as for example imperceptible, perceptible and so on. The second method is by viewing the combined 2D video and depth image, which is the 3D stereoscopic image, and rating them during human factor experiments to measure the perceived visual quality and viewing comfort [29]. However, in this thesis, the 3D stereoscopic image is evaluated using pair comparison [32], where the original 3D stereoscopic sequence and the decoded sequence are displayed side by side. For error resilience performance evaluation, the original 3D stereoscopic sequence is replaced with error-free decoded ones to show the error performance of the coder.

2.7 3D Video Coding

In this section, the compression of 3D video in the form of stereo video will be investigated. It is a great challenge to compress the stereo video because it requires double the bandwidth of normal video. However, there are large redundancies in stereo video that can be removed without degrading quality, as the left view and right view are highly correlated (the left- and right-view are offset by the position of the camera that is next to each other). The coding of stereo video
Chapter 2. 3D Video Coding

from a two-view camera is presented in Section 2.7.1. The depth information is the second component to be compressed in the 2D video and depth data. The redundancies in the depth data can be removed similarly to redundancies in 2D video. Section 2.7.2 describes the coding of 2D video and its depth.

2.7.1 Two-View Stereoscopic Video Coding

The straightforward approach to code stereo-views is to code the two views separately into two video streams. Another method is to combine the left and right views in one frame and code into a single bit stream. Only spatial redundancies in the combined left and right views are considered in this method but not the temporal redundancies between the two views. The 2-view sequence can also be considered as interlaced materials where the left view picture is the top field and the other view is the bottom field. Then, the interlaced video can be coded using the existing interlace coding method as available in the standard video codec. However it is not suitable to be decoded by devices without 3D display functionality. These three methods are briefly described in [33].

Two view stereoscopic video can also be coded using MPEG4 temporal scalable coding. In this method, the left view is the base layer and the right view is the enhancement layer. This method was investigated by MPEG 3DAV Adhoc group in Exploration Experiment 3 (EE3) [34]. They basically compared this method with coding the stereoscopic video using Multiple Auxiliary Component (MAC) in MPEG4. In basic MPEG4-MAC stereoscopic video coding, the left image is compressed as the normal video. Disparity map between left in right image is generated by pixel-by-pixel basis using disparity estimation algorithm [21]. This map is then assigned to a component in MAC.

MAC will be described more in Section 2.9. The single bit stream from MPEG4-MAC encoder is decoded by MPEG4-MAC decoder. The reconstructed disparity map is added to the reconstructed left image to form the reconstructed right image. Extension of MPEG4-MAC is also proposed in [34]. Nevertheless, as reported in EE3, MPEG4 temporal scalable coding is so far understood to be the best coding tools for left-right stereoscopic video coding.

2.7.2 2D and Depth Stereoscopic Video Coding

In this thesis, our research interest is on stereoscopic video from the depth camera as explained in Section 2.3. Compression of 2D video and depth can be achieved using the standard video codecs such as MPEG4 and H.264 [35] and [3]. One way to code the 2D and depth information is to code them separately using two video encoders as shown in Figure 2-4.
For example in [3], the depth information obtained from the depth range camera is encoded using standard video codecs, with H.264 giving superior performance in terms of rate distortion compared to other video codecs such as MPEG4 and MPEG2. Alternatively, the MAC tool in MPEG4 can be used to code the depth data. The MAC can also be used to code the disparity map ([etri] and [wam14.5]) as in classical stereoscopic video coding. Thus, it provides us with options whether to use 2D video plus depth or two-view stereoscopic video as the input material to the encoder.

In [36], the usage of MPEG4-MAC to transmit the 2D video and depth has been investigated, but the input to the encoder is not the data from the depth camera. Furthermore, the background in the test sequence is suppressed and not encoded which is not a real-life scenario. Although compression of depth information using standard video codecs is reported in [3], the performance compared to MPEG4-MAC is not considered.

In Section 2.9, we further investigate the use of MPEG4-MAC to efficiently compress and transmit the 2D video and the depth data. Performance of MPEG4-MAC compared to separately encoding the depth and 2D video information using MPEG4 (MPEG4-Separate) is also evaluated.

### 2.8 Reduced Resolution Depth Compression for 3D Video

This section investigates our contribution in term of the compression of depth information, but at a reduced resolution for low bit rate applications, by down-sampling the video frames prior to H.264 encoding and up-sampling them after the H.264 decoding. Application of DSUS for image compression has been investigated in [37]. It is shown that at low bit rates, the down-sampled image, when compressed and later extrapolated after reconstruction outperforms the original resolution image both subjectively and objectively.
2.8.1 Depth Compression

Reduced resolution for 2D video is used for example in mobile applications with the aim to reduce bandwidth and provide portability. The depth image sequence from the depth range camera is in European digital TV format (720x576 luminance pixels, 8-bit per pixel, 25 Hz) [3], which produces an uncompressed bit rate of about 83 Mbit/s. Figure 2-5 shows an example of one of the frame in the sequence, the basic luminance information and the associated per-pixel depth information. In practice, the depth image sequence must be transmitted together with the luminance and colour information, hence the need for compression.

![Example frames](image)

**Figure 2-5: Orbi test sequence, (a) Luminance information, (b) Associated per-pixel depth information**

There are several ways to compress depth image sequences presented in the literature. In [38], depth images are compressed by using region-of-interest coding and reshaping the dynamic range of the depth map. This is currently being extended to the depth image sequence. Another method for depth image compression is presented in [39]. Predictive image coding is used to code the depth image based on context modelling. In [40], mesh based method is used to moderately compress the depth image sequence targeting real time rendering on a standard PC. Compression of layered depth image, which is an extension of the depth image, is proposed in [41]. Each component in the layered depth image is compressed separately using a video object wavelet codec. In [36], [42], and [43], a simple depth image is compressed using the multiple auxiliary components already standardized within MPEG4 and a median filter is used to improve the performance.

2.8.2 Down-Sampling and Up-Sampling Method

We propose to compress the depth image sequence by using image down-sampling prior to standard H.264 encoding and up-sampling after H.264 decoding as shown in Figure 2-6.
Image down-sampling can be performed in the pixel domain and in discrete cosine transform (DCT) domain [44]. A simple way to down-sample an image by a factor of two is by using sub-sampling. If \( f(i,j) \) is the pixel value of an image at location \((i,j)\), then the down-sampled image,

\[
fd(i/2,j/2) = f(i,j)
\]  \hspace{1cm} (2-5)

For \( i=0,2,4,...\ N, \) \( j=0,2,4,...\ M \)

\( N \) is the vertical size and \( M \) is the horizontal size of the image to be down-sampled. It should be noted that three neighbourhood pixels to \( f(i,j) \), which are \( f(i,j+1), f(i+1,j) \) and \( f(i+1,j+1) \) can also be used. In this thesis, the pixel \( f(i,j) \) plus the three neighbourhood pixel values are averaged as below to produce the down-sampled image by a factor of two (DSUS2), \( fd() \),

\[
fd(i/2,j/2) = \frac{f(i,j)+f(i,j+1)+f(i+1,j)+f(i+1,j+1)}{4}
\]  \hspace{1cm} (2-6)

For \( i=0,2,4,...\ N-1, \) \( j=0,2,4,...\ M-1 \)

Equation (2-4) is repeated for every frame in the 720x576 depth image sequence resulting in a 360x288 depth image sequence. The latter sequence has to be cropped to 352x288 (CIF resolution) size to make it suitable for H.264 encoder operation. Note that equation (2-4) can be used again for down-sampling by a factor of four (DSUS4) to produce a QCIF image sequence.

After the H.264 decoder, the reconstructed depth image sequence is up-sampled to its original size,

\[
fru(i,j) = fru(i,j+1) = fru(i+1,j) = fru(i+1,j+1) = fr(u, v)
\]  \hspace{1cm} (2-7)

For \( i=0,2,4,...\ N-1, \) \( j=0,2,4,...\ M-1 \)

where \( fr() \) is the reconstructed image from H.264 decoder and \( fru() \) is the up-sampled reconstructed image. This is again repeated for every frame in the depth image sequence.

It is expected that the up-sampling distortion will reduce the quality of the decoded image sequence. Our simulation results show that the up-sampling distortion is only noticeable at high bit rate, but not at low bit rate.
2.8.3 Results of Down-Sampling and Up-Sampling for Depth Information

H.264 codec (version JM 7.5b) is used to produce the simulation results. Encoding parameters include: Initial QP = 24, IPPP… coding format (only the first frame is intra-coded) and 124 frames of depth Orbi sequence. Figure 2-7 shows the PSNR level obtained for every frame at 64 kbit/s. Figure 2-8 shows the quantisation parameters obtained when the target bit rate is set to 64 kbit/s. Subjective results obtained for one of the frames in the depth image sequence are shown in Figure 2-9. Table 2-1 shows the objective results (average PSNR for luminance) of the H.264 reconstructed depth image sequence at the original resolution, as well as the DSUS2 and DSUS4 sequences. Fixed target bit rates are used during the H.264 encoding process as shown in the table.

![Figure 2-7](image1)

**Figure 2-7:** Objective quality (PSNR) of compressed original resolution, DSUS2 and DSUS4 at 64 kbit/s

![Figure 2-8](image2)

**Figure 2-8:** Quantisation parameters of compressed original resolution, DSUS2 and DSUS4 at 64 kbit/s
Figure 2-9: Frame 52 of Orbi depth image sequence, (a) Original, (b) Original compressed at 64 kbit/s, (c) DSUS2 at 64kbit/s, (d) DSUS4 at 64 kbit/s. The most affected regions are highlighted in boxes.

Table 2-1: Average PSNR for luminance value of the depth information at various bit rates

<table>
<thead>
<tr>
<th>Bit Rate (kbit/s)</th>
<th>Average PSNR (Y) Original</th>
<th>DSUS2</th>
<th>DSUS4</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>35.05</td>
<td>35.39</td>
<td>34.16</td>
</tr>
<tr>
<td>64</td>
<td>30.94</td>
<td>34.47</td>
<td>33.88</td>
</tr>
<tr>
<td>32</td>
<td>N/A</td>
<td>32.39</td>
<td>33.3</td>
</tr>
<tr>
<td>16</td>
<td>N/A</td>
<td>28.7</td>
<td>31.44</td>
</tr>
<tr>
<td>8</td>
<td>N/A</td>
<td>N/A</td>
<td>27.29</td>
</tr>
</tbody>
</table>

2.8.4 Discussion

In Figure 2-7, the original resolution sequence starts with high PSNR values and eventually drops to lower values than DSUS2 and DSUS4 because it cannot meet the target bit rate with the same
starting quality. DSUS2 and DSUS4 have almost constant values throughout the sequence. DSUS2 and DSUS4 perform better because the up-sampling distortions in the form of averaging and “zooming in” effects are much less damaging to the objective quality than the compression distortions of the original resolution pictures at this low bit rate. It can also be seen from Figure 2-7 that DSUS2 performs better than DSUS4 in most of the frames since the up-sampling process is performed twice in DSUS4, resulting in greater up-sampling distortion in the reconstructed image sequence. Performing DSUS by four once may give the same results as DSUS4.

From the subjective quality evaluation in Figure 2-9, the compressed original resolution (Figure 2-9(b)) has poorer image quality compared to DSUS2 and DSUS4 because of the heavy compression performed to meet the target bit rates. The encoder uses large quantisation parameter (coarser quantisation) to compress the original resolution resulting in blocking artefacts in the image sequence. This usage of large quantisation parameters for original resolution can be seen from Figure 2-8. Figure 2-8 plots the resulting quantisation parameters obtained at 64 kbit/s for the three cases. Another reason why DSUS2 and DSUS4 have performed better is because of the smoothing filtering effect of the up-sampling process as shown in Figure 2-9(c) and (d). Although DSUS4 has the lowest quantisation parameters values as depicted in Figure 2-8, it does not necessarily mean that it has the highest quality (see Figure 2-7). This is because of the twice up-sampling process in DSUS4 as explained in the previous paragraph.

From Table 2-1, at 128 kbit/s, performance of the three resolutions is comparable. At lower bit rates, especially at 64 kbit/s, DSUS2 and DSUS4 perform much better than the original resolution by about 3.5 dB and 3 dB on average respectively. Table 2-1 also shows that using the H.264 version 7.5b encoder, with the available rate control method, a bit rate of 32 kbit/s and below are not achievable with the original resolution sequence. Other parameters have to be adjusted to achieve the target bit rate, such as the frame skip values.

At 32 and 16 kbit/s, DSUS4 performs better than DSUS2 by about 1 dB and 3 dB respectively. At these bit rates, compression and up-sampling distortions of DSUS4 are less annoying than the compression and up-sampling distortions of DSUS2. Furthermore, DSUS4 managed to achieve 8 kbit/s with reasonable PSNR values, which was not possible to achieve with the original resolution and DSUS2 sequences using the available rate control method.

### 2.9 MPEG4-Multiple Auxiliary Component for 3D Video Coding

Our contribution in coding the 2D video and depth data is presented in this section. The objective of this section is to further investigate the use of Multiple Auxiliary Component (MAC) in MPEG4 (MPEG4-MAC) to efficiently compress and transmit the 2D video and the depth data. Performance of MPEG4-MAC compared to separately encoding the depth and 2D video
information using MPEG4 (MPEG4-Separate) is also evaluated. The MPEG4-MAC architecture is explained in Section 2.9.1 followed by MPEG4-Separate architecture in Section 2.9.2. Results of the experiments are presented in Section 2.9.3 and are discussed in Section 2.9.3.1. Section 2.9.4 applies the DSUS method to the depth information in MPEG4-MAC.

2.9.1 MPEG4-Multiple Auxiliary Component Architecture

The MAC is introduced in version 2 of MPEG4 and is also known as the alpha channels. Basically, the MAC is the greyscale shape that is not only used to describe transparency of the video object, but can also be defined in more general way to describe shape, disparity shape of multi-view video objects, depth shape (obtained for example from laser range finder in depth camera and disparity analysis) and infra-red or other secondary texture [45]. The MPEG4-MAC has been used to code the disparity for two-view stereoscopic video in [47] and [48].

The MAC usually has the same shape and resolution as the texture component. The depth information obtained from the depth camera has the same resolution as the 2D video information. Therefore, any of the multiple auxiliary components can be used for the depth information. The encoding of the auxiliary components is similar to the texture component, which employs motion compensation and DCT. It uses the same motion vectors of the luminance for the motion compensation of the auxiliary components.

The configuration of MPEG4-MAC to encode the 2D video and the depth information is shown in Figure 2-10. In order to use the MPEG4-MAC, shape information that has the same resolution as the luminance has to be supplied to the encoder. This is because the MAC is initially developed to encode greyscale alpha, which indicates the amount of transparency of an image. To encode the greyscale alpha, shape information is needed to preserve the object shape information inherited in the data [46]. However, in this project the shape information is just used as the input to the encoder but is not compressed and transmitted.

In our simulation, all the pixels values of the shape information are set to 255 (not shown in Figure 2-10). This will notify the MPEG4-MAC to code all the objects and background in the 2D and depth image. Otherwise, the MPEG4-MAC will encode according to the shape information provided so as to produce a segmented output.
2.9.2 MPEG4-Separate Architecture

In this configuration, the 2D video is encoded separately from the depth data, as shown in Figure 2-11. Using an MPEG4 video codec, the depth data is supplied as the luminance channel. Another MPEG4 video codec is used to encode the ordinary 2D video. Two parallel video codecs can be utilized if the speed of the compression is an important factor. Alternatively, one video codec can be used to code the 2D video first, followed by the depth data. One of the advantages of MPEG4-Separate architecture is that the shape information does not have to be supplied as input as in MPEG4-MAC. However, if the shape is needed, it can also be supplied to MPEG4-Separate architecture.
2.9.3 Experimental Results for MPEG4-MAC and MPEG4-Separate Coding

The experiments are performed using the configuration shown in Figure 2-10 and Figure 2-11. 2D Orbi image sequence of size 720x576 (125 frames) and the depth information (720x576, 8 bits/pixel) are used for the simulation. A CIF size image sequence (352x288), namely Interview is also used for the simulation. The image sequences are downloadable from [4]. It is to be noted that the depth sequence is made into YUV 4:2:0 format for MPEG4-Separate encoding by setting the chrominance value (U and V) to a constant value. After decoding, the 2D video and the depth data are combined using the depth image based rendering technique proposed in [13]. The output sequence is displayed as a colour anaglyph format that can be viewed using colour (blue and red) anaglyph glasses to get the 3D impression.

The objective results in terms of average PSNR for Orbi and Interview are shown in Table 2-2 and Table 2-3 respectively. As an example, the PSNR and total bits for texture and depth of each frame of the Orbi image sequence at 2 Mbit/s are depicted in Figure 2-12(a) and Figure 2-12(b). The subjective results for texture and depth are shown in Figure 2-14 and Figure 2-15 respectively. The bit rates used for texture and depth are also shown in the figures. In the simulation, high bit rate is used to accommodate for the large size 2D image sequence and the depth information. The bit rates for MPEG4-MAC and MPEG4-Separate are the overall bit rates, which includes all overheads bits, texture, colour, depth and motion vectors.

2.9.3.1 Discussion

It is shown in Table 2-2 (for Orbi) that the objectives results for encoding the texture using MPEG4-MAC and MPEG4-Separate are comparable at the same target bit rates. MPEG4-

Table 2-2: Comparison of average PSNR for MPEG4-MAC and MPEG4-Separate for Orbi image sequence

<table>
<thead>
<tr>
<th>Bit rate Mbit/s</th>
<th>Texture</th>
<th>Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAC</td>
<td>Separate</td>
</tr>
<tr>
<td>1</td>
<td>34.06</td>
<td>35.89</td>
</tr>
<tr>
<td>1.5</td>
<td>36.84</td>
<td>37.52</td>
</tr>
<tr>
<td>2</td>
<td>38.19</td>
<td>38.65</td>
</tr>
<tr>
<td>2.5</td>
<td>39.02</td>
<td>39.36</td>
</tr>
</tbody>
</table>
Table 2-3: Comparison of average PSNR for MPEG4-MAC and MPEG4-Separate for Interview image sequence

<table>
<thead>
<tr>
<th>Bit rate kbit/s</th>
<th>Texture MAC</th>
<th>Texture Separate</th>
<th>Depth MAC</th>
<th>Depth Separate</th>
</tr>
</thead>
<tbody>
<tr>
<td>750</td>
<td>39.74</td>
<td>38.93</td>
<td>39.74</td>
<td>40.53</td>
</tr>
<tr>
<td>800</td>
<td>38.23</td>
<td>38.62</td>
<td>41.24</td>
<td>41.87</td>
</tr>
<tr>
<td>900</td>
<td>38.45</td>
<td>38.84</td>
<td>41.9</td>
<td>43.12</td>
</tr>
<tr>
<td>1000</td>
<td>39.14</td>
<td>39.37</td>
<td>41.79</td>
<td>43.12</td>
</tr>
</tbody>
</table>

Separate encoding is slightly better than MPEG4-MAC because of the separate encoding of the texture. It can be seen from Figure 2-12(a) that at about the same target bit rate, MPEG4-MAC performance is comparable to MPEG4-Separate performance in terms of PSNR of the Y-component for Orbi at 2 Mbit/s. This is confirmed by Figure 2-14 that shows the subjective quality of the luminance. Table 2-3 (for Interview) also shows that MPEG4-MAC is as good as MPEG4-Separate in encoding the texture component.

It can also be seen from Table 2-2 that the depth data for Orbi is slightly better encoded by MPEG4-Separate. Figure 2-13 gives the PSNR value and bits/frame for all the frames in Orbi sequence. At about the same bit rate, MPEG4-MAC codes the depth at constant quality (Figure 2-13(a)). This is because the QP is fixed for the depth data in MPEG4-MAC. The QP is set to be 20 for I-frame and 16 for P-frame. This QP can be adjusted to improve the depth information quality. Since the QP is fixed for the depth data in MPEG4-MAC, the target bit rate can be achieved by adjusting the QP value for the luminance component. The depth PSNR values of MPEG4-Separate (Figure 2-13(a)) are varying and not as constant as the depth PSNR values of MPEG4-MAC. This is because the depth data is fed into the luminance channel of MPEG4-Separate, where the QP is varied to obtain the target bit rates. Although there is difference in encoding the depth using MPEG4-MAC and MPEG4-Separate, the subjective quality is almost similar as can be seen from Figure 2-15.

The reason MPEG4-Separate operates slightly better is because it uses motion estimation to code the depth giving better estimate of the reconstructed depth image. MPEG4-MAC uses motion vector of the luminance to code the depth, therefore the reconstructed depth image is not as good as the MPEG4-Separate. Nevertheless, the 3D subjective quality between the two architectures shows comparable performance. For the Interview sequence, the performance of MPEG4-MAC in coding the depth is again comparable to MPEG4-Separate. For example at 750 kbit/s, an average
Figure 2-12: (a) PSNR (Y component of the texture) and (b) total bits (texture) for each frame of Orbi at 2Mbit/s

PSNR of 39.74 is obtained by MPEG4-MAC and 40.53 is obtained by MPEG4-Separate. This result clearly shows that MPEG4-MAC can be used to code the 2D-video and the depth at comparable performance as MPEG4-Separate.

It is observed from Table 2-2 and Table 2-3 that the depth data is better or slightly better coded using MPEG4-Separate, but the performance of MPEG4-MAC in coding the depth can be improved by fixing the quantisation parameter for depth to a lower value. In our simulation, the QP is set to 20 for I-frame and 16 for P-frame. Therefore, fixed quality depth data and variable quality 2D video can be obtained from MPEG4-MAC, while variable quality depth data and 2D video are produced by MPEG4-Separate.

After decoding, the 2D video and the depth data are combined and displayed as a colour anaglyph format that can be viewed using colour (blue and red) anaglyph glasses to get the 3D impression. The 3D impression is successfully generated for both MPEG4-MAC and MPEG4-Separate. It is found that MPEG4-MAC 3D subjective performance is more or less similar to MPEG4-Separate. The difference in quality of the 3D impression for both architectures is also unnoticeable.
Furthermore, MPEG4-MAC has some advantages over MPEG4-Separate. In term of architecture, MPEG4-MAC is less complex than MPEG4-Separate because it only uses one encoder/decoder at one time. Although shape has to be encoded in MPEG4-MAC, the chrominance component of the depth (U and V component) is not needed in MPEG4-MAC. The chrominance component has to be supplied to MPEG4-Separate architecture together with the depth, which is supplied as the luminance channel. In addition to that, two motion vectors, one for the 2D video and one for the depth, need to be transmitted if MPEG4-Separate is used, whereas only one motion vector (for the 2D video) needs to be transmitted in MPEG4-MAC. The depth data uses the 2D video motion vector for its motion compensation.

Another advantage of MPEG4-MAC is that the overall bit rate generated from MPEG4-MAC already includes the bit rate for the depth, whilst the bit rate for depth data has to be taken into consideration for the overall bit rate in MPEG4-Separate. Beside those advantages, researchers in [3], [36] and [43] suggested MPEG4-MAC can be used to encode the depth data. Improved
(a) (b)

Figure 2-14: Luminance subjective quality for Orbi sequence at 2Mbit/s, (a) MPEG4-MAC (b) MPEG4-Separate

(a) (b)

Figure 2-15: Depth subjective quality for Orbi sequence at 2Mbit/s, (a) MPEG4-MAC (b) MPEG4-Separate

Performance is obtained with non-linear post filtering and in-loop median filter as contributed in [36] and [43].

Finally, MPEG4-MAC has been chosen to encode the 2D video and the depth data because it provides one output bit stream at the encoder. This one bit stream can be fed into a Multiple Description Coding (MDC) encoder to increase its error resilience. The output of the MDC encoder is two or more bit streams that are equally important. If only one of the bit streams is received at the decoder, the 2D or 3D video can be reconstructed with acceptable quality. Receipt of more than one bit stream can provide an enhanced level of quality to the 2D or 3D video. Work in this area is described in Chapter 3 and 4.
2.9.4 MPEG4-MAC for 3D Video Coding with Down-Sampling and Up-Sampling Application

The DSUS technique in Section 2.8 can be used to reduce the coding bit rate for the depth information of high resolution sequence such as the original Orbi and Interview sequences. This section proposes the combination of the DSUS with MPEG4-MAC. The idea is to down-sample the depth information before the MAC encoding and up-sample it back after the MAC decoding. It is expected that the performance of the MPEG4-MAC-DSUS is better than MPEG4-MAC at low bit rate range. The block diagram of the proposed method is shown in Figure 2-16.

The method is tested to code 125 frames of the Orbi sequence at 720x576 resolution and at 30 frames/s. Only DSUS2 of the depth information is investigated in this section to show the performance of DSUS on MPEG4-MAC. Performance of DSUS4 is not investigated as the up-sampling distortion on the depth information is more than DSUS2. However, DSUS4 could be used to further improve the coding of the 2D video using MPEG4-MAC as it can produce lower bit rates than DSUS2 at the expense of the depth quality. The rate distortion for luminance and depth is plotted in Figure 2-17 showing the average PSNR for one vertical axis and the bit rate on horizontal axis. The bit rate shows the overall bit rate including the texture and the depth information. Fixed quantisation parameters are used to obtain the bit rates and only the first frame is encoded as I-frame. The error resilience options are disabled.

![Figure 2-16: Block diagram of MPEG4-MAC-DSUS2](image)

The bit rate of the MPEG4-MAC-DSUS is found out to be less than MPEG4-MAC due to coding the depth information at reduced resolution. Hence, rate distortion of the luminance for the MPEG4-MAC-DSUS is better than the MPEG4-MAC. The rate distortion for the depth in Figure 2-17 (b) indicates a better performance by the MPEG4-MAC-DSUS compared to the MPEG4-MAC especially at low bit rate range (less than 1 Mbit/s). For example, at bit rate of 500 kbit/s, MPEG4-MAC-DSUS achieves average depth PSNR of 31.5 dB, while MPEG4-MAC achieves average depth PSNR of 29.5 dB. This indicates that MPEG4-MAC-DSUS performs better than MPEG4-MAC for depth information at the low bit rate range.
Figure 2-17: Rate-Distortion curves for ‘Orbi’ sequence (a) Luminance (b) Depth

The same bit rate experiment is performed for Orbi sequence at low bit rate range (less than 1 Mbit/s). In this experiment, the quantisation parameter for the depth of MPEG4-MAC-DSUS is decreased for this bit rate range so that the overall bit rate of MPEG4-MAC-DSUS matches the overall bit rate of MPEG4-MAC and the result is plotted in Figure 2-18.

It can be seen from Figure 2-18 that MPEG4-MAC-DSUS performed better by about 0.5 dB to 1 dB than MPEG4-MAC for coding the depth at bit rate less than 950 kbit/s due to the used of lower quantisation parameter for the depth. Above 1 Mbit/s, performance of the MPEG4-MAC
Chapter 2. 3D Video Coding

Figure 2-18: Rate distortion for depth information of Orbi sequence at the same bit rate for the depth is better than MPEG4-MAC-DSUS as the up-sampling distortion begins to show its effect.

2.10 Conclusion

In this chapter, background on 3D video is presented. The 3D video is classified according to MPEG 3D AV AdHoc group. Stereoscopic video from 2D video and depth data is the focus of the work because it offers backward compatibility and is widely used in the 3D-TV project. Furthermore, the 2D video plus depth format was recently standardised in ISO/IEC 23002-3 (MPEG-C part 3) ([27] and [28]). Moreover, commercial 3D display such as Philips 3D display recognises 2D video plus depth information as reliable 3D stereoscopic source. This chapter also describes the available 3D stereoscopic displays technology. The DIBR method to convert 2D and depth video into left and right views, is also described in this chapter. The 2D and depth video test sequences and the stereoscopic 3D video performance evaluation used in this thesis are also discussed. Furthermore, the 3D video coding of two-view stereoscopic video and 2D and depth stereoscopic video is presented.

This chapter also presents one of our contributions in terms of the use of down-sampling and up-sampling of the depth data. The use of down-sampling prior to encoding and up-sampling after decoding introduces up-sampling distortions beside the quantisation errors. However, using the simulation results, we have demonstrated that if the resolution of the depth image sequence is reduced prior to encoding and up-sampled back to its original resolution after decoding, far better
objective and subjective quality could be achieved compared to compressing using the original resolution of the depth image for low bit rate 3D video applications.

Next, the chapter proposes to use the coding tool in MPEG4, namely MAC, to code the 2D video and depth data. Objective and subjective performance of MPEG4 to code the 2D video and depth have been explored. Two architectures of MPEG4, namely MPEG4-MAC and MPEG4-Separate are used for performance comparison. Simulation results with 2D video plus depth image sequences show that both perform at par in coding the 2D video at the same target bit rate. MPEG4-Separate is slightly better in coding the depth information. More importantly, the resulting 3D perception of both architectures does not seem to show much difference in terms of the subjective quality achieved. Hence, MPEG4-MAC is an attractive solution to compress the 3D video. The one bit stream from MPEG4-MAC can be fed into an MDC encoder to increase its error resilience. The proposed MPEG4-MAC-MDC will be discussed in Chapter 3.

Finally, the application of DSUS to the depth information in MPEG4-MAC is proposed in this chapter. The results show that if the resolution of the depth image sequence is reduced prior to MAC encoding and up-sampled back to its original resolution after decoding, far better luminance performance could be achieved compared to compressing using the original resolution of the depth image. This is due to the overall reduced bit rate obtained by down-sampling the depth information. However, performance of the depth information degrades as the bit rate increases beyond 1 Mbit/s because of the up-sampling distortion. Further investigation of DSUS on a scalable video coding and a scalable MDC will be discussed in Chapter 5.
Chapter 3

Error Resilience and Multiple Description Coding for 3D Video

3.1 Introduction

When the compressed 3D video is transmitted over communication networks, the associated packet loss can lead to poor visual quality. Error control schemes, known as error resilience techniques are applied at the encoder to make the compressed video more resilience to channel errors. Many error resilience methods for 2D video are available in the literature. In Section 3.2, the error resilience tools included by the MPEG-4 standard for 2D video are briefly described. The error resilience tools are then extended to the depth data in 3D video. Section 3.3 introduces MDC as a form of error resilience tool. In Section 3.4 an even and odd (EO) frame based MDC method was developed on top of the MPEG-4 version 2 standard and applied to 3D video. The performance of the developed MDC-EO is also investigated in error free and error prone environment. Section 3.5 concludes the chapter.

3.2 Error Resilience Tools

Errors are introduced in the bit stream when the compressed video data is transmitted over noisy communication channels. MPEG-4 has adopted the following error resilience tools to provide basic error robustness, namely packet resynchronisation, data partitioning and reversible variable-length-codes (RVLC) [49].

3.2.1 Packet Resynchronisation

A video decoder that is decoding an erroneous bit stream will lose synchronisation with the encoder if errors are encountered causing the quality of the decoded video degrades rapidly. One remedial action is for the encoder to insert resynchronisation markers in the bit stream at various locations. When an error is detected, the decoders can then search for the next resynchronisation marker and regain resynchronisation. The MPEG-4 encoder has the option of dividing the image into video packets, each made up of an integer number of consecutive macro-blocks. At the
beginning of each video packet, the MPEG-4 encoder inserts a resynchronisation marker. The size of the video packet can be determined by the user in terms of K bits. This ensures that the decoder can effectively localise the error depending on the content of the images. In addition to inserting the resynchronisation markers at the beginning of each video packet, the MPEG-4 encoder removes all data dependencies that exist between the data belonging to two different video packets within the same image. This is required because if one packet is in error, the other packet can still be decoded by the decoder.

3.2.2 Data Partitioning

In this mode, the data within a video packet is partitioned into two parts, a motion part and a texture part separated by a unique motion boundary marker as shown in Figure 3-1. The data is partitioned according to their sensitivity to errors. The texture part contains DCT data which is less sensitive information. All elements that are sensitive to errors, mostly the motion-related information, shape information and the administrative bits, are placed in the motion partition. This scheme allows the decoder to restore error-free motion of a video packet when errors corrupt only the less sensitive texture part. Errors on the texture part can usually be concealed successfully, resulting in little visible distortion. Furthermore, to improve its error resilience performance, unequal error protection can be applied to the two partitions by using more powerful error protection schemes on the first partition and less powerful error control schemes on the second partition [50].

![Figure 3-1: Data partitioning in MPEG-4](image)

3.2.3 Reversible Variable Length Codes

Reversible Variable Length Codes (RVLC) is used to confine the effect of the damage occurred in the compressed bit stream. It allows decoding in the reverse direction. If an error is detected in forward direction, the decoder stops it operation and search for the next synch word skipping segments of the bit stream. The decoder resumes its operation from this synch word in the reverse direction, decoding the bits that were initially skipped in the forward direction. To enable this mode of operation, the bit stream must be coded using RVLC, a variable-length codes that can be decoded in both the forward and reverse directions and produce the same output.
3.2.4 Depth Data Error Resilience

The compressed depth data from the 3D video can also be subjected to errors when transmitted over communication channels. Therefore, it is necessary to provide error resilience tools for the depth data to mitigate the effect of the channel errors. In this thesis, the depth data is transmitted via one of the auxiliary channels in MAC of MPEG-4. Unfortunately, error resilience tools in MPEG-4 are not available for these auxiliary channels (or so called alpha plane). In this section, data partitioning and RVLC are applied to the depth data to enhance its performance in error prone condition. Before data partitioning is applied to the depth data, the depth data error need to be investigated. This is important so that the depth information is properly partitioned when data partitioning mode is enabled.

3.2.5 Depth Data Sensitivity and Data Partitioning

To investigate the sensitivity of depth to error, bit errors are simulated using the MPEG-4 version 2 standard decoder reference software. The performance of the 3D video is evaluated when the simulated errors hit the bit stream according to the following conditions: (a) Only the texture part of the bit stream is corrupted (b) Only the motion part of the bit stream is corrupted (c) Only the depth part of the bit stream is corrupted. A bit error rate of $10^{-3}$ is simulated in the experiment using random number generator in C source code. It is applied to the reference software at the decoder side.

Interview sequence at 720x576 format is used in the simulation and was encoded using MPEG-4 encoder using $QP = 5$. Format of the sequence is IPPP, which means only one I frame is inserted at the start of the encoded sequence. Temporal error concealment algorithm is turned on in the experiment. PSNR_lumin and PSNR_depth (introduced in Section 2.6) are not measured here because the interest is only on the impact of the errors on the 3D video. Figure 3-2 shows the subjective quality of the texture and depth at a reduced resolution. It also shows the 3D stereoscopic video when the texture and depth are combined using DIBR technique described in Chapter 2.
It was found from the experiment that the depth information is less sensitive than texture and motion vectors. This is clearly shown in the subjective quality of the stereoscopic 3D video in Figure 3-2. The 3D video is shown as overlaid red and blue images, which can be viewed using the red and blue glass. The resulting 3D video is not much affected if the error only hits the depth data. This may be due to the nature of the depth data, which contains a lot of grey and black colour. If the errors hit a macro-block in the background of the depth data, which is mostly black in colour, it can be easily concealed with corresponding black macro-block in the previous frame. The 3D video quality is worst if the errors hit the motion part of the bit stream. This result is reconfirmed by experiments in [51] and [52]. Hence the depth data is put at the end of the data.
partitioned block as shown in Figure 3-3. Simulation results with this configuration are presented in Section 3.4.3.

![Figure 3-3: Depth data partitioned](image)

### 3.3 Multiple Description Coding

A good review of MDC and its variation is given in [55] and [56]. In this section a review on available MDC methods is briefly presented. MDC can be used to encode a source for transmission over a communication system with multiple channels. Its objective is to encode a source into two or more bit streams. The streams or the descriptions are correlated and equally important. This means that a high-quality reconstruction is decodable from the received bit streams together, while a lower, but still acceptable quality reconstruction is achievable if only one stream is received.

With MDC, re-transmission is not required hence it is suitable for applications that require low delay. MDC assumes there are more than one channel between transmitter and receiver (Channel 1, 2, ... n). Either channel may fail with probability \( p_i \), \( i = 1, 2, ... n \). It also assumes independent error and failure events of the two or more channels. Moreover, it is assumed that the probability of all channels failing at one time is low.

Figure 3-4 shows the general block diagram of MDC encoder and decoder for two descriptions. In general, it can be extended to more than two descriptions. In Figure 3-4, the two descriptions created by the encoder are sent separately across two channels. The total bit rate is \( R = R_1 + R_2 \), where \( R_1 \) and \( R_2 \) are the bit rates used to send each description.
In the case of two channels, if El and E2 are received, the decoder invokes Decoder 0 to decode El and E2, and produces a high-quality reconstruction with central distortion D0. If only El is received, Decoder 1 is invoked to decode signal from Channel 1 producing a lower, but still acceptable quality reconstructions with side distortions D11. Decoder 2 is invoked when only E2 is received, producing acceptable quality reconstructions with side distortions D12. A balanced design is achieved when R1=R2 and D1=D11=D12.

A single description coder (SDC) minimises D0 for a fixed total rate R, and rate-distortion function R(D0) is used to measure the performance. MDC has contradictory requirements to simultaneously minimise both D0 and D1. At one extreme, minimising D0 by simply alternating the R bits of an SDC bit stream into each description will have unacceptably high D1. At the other extreme, minimising D1 by simply duplicating the SDC bit stream with rate R into each description will have a large D0 because it uses 2R bits to achieve D0.

### 3.3.1 Redundancies and Mismatch

Two main concerns of MDC are the redundancies generated by the two or more bit streams and the prediction mismatch between encoder and decoder when only a single description is received. MDC has less coding efficiency due to the redundancies, but the redundancies are acceptable if it can provide good quality in error prone environment. The challenge is to maximise robustness to channel errors at permissible level of redundancy.

The redundancy $r$ is defined as the additional bit rate required by the MDC coder, $r = R - R^*$, where $R$ is the total bit rate of the two streams in MDC, and $R^*$ is the reference bit rate from (SDC).
Chapter 3. Error Resilience and Multiple Description Coding for 3D Video

A condition called “mismatch” occurs whenever an encoder uses a signal for prediction that is unavailable to the decoder because of transmission loss. In Figure 3-4, depending on which descriptions are received, the decoder has three possible states, Decoder 0, Decoder 1 and Decoder 2. However, the encoder can never know which of these states is present. If the encoder uses a predictor that depends on state not available at the decoder, the encoder and decoder states will be mismatched. This potential mismatch and the subsequent error propagation present a fundamental design concern for MDC video coders. In Figure 3-4, the mismatch occurs when only E1 or E2 is received and only Decoder 0 is available. Decoder 0 is designed to decode both E1 and E2 together. It fails to decode when only E1 or E2 is received. The mismatch can be eliminated by having Decoder 1 and 2 specially designed to decode E1 and E2 respectively.

In [57], the mismatch can be completely avoided by the video redundancy coding (VRC) method, where even and odd frames are coded into two separate descriptions namely E1 and E2 if we refer to Figure 3-4 as reference. Each frame in E1 and E2 is predicted from the previous even frame and odd frame respectively. If only even frames (or E1) are received for example, the VRC decoder will decode the video at a reduced temporal resolution using previous even frames in the prediction. Hence, the prediction mismatch between the MDC encoder and Decoder 0 is eliminated. The mismatch can be also partially encoded and hence the redundancies are controlled as in [54] and [59].

MDC algorithms in the literature can be broadly categorised into three methods, MDC quantisation, MDC transform coding and MDC sub-sampling.

3.3.2 Multiple Description Coding Method Based on Quantisation

MDC quantisation splits the quantized coefficient into two or more streams. In a simple implementation of MDC quantisation algorithm, the multiple descriptions are produced by using two quantisers whose decision regions are offset by one-half of a quantisation interval of each other [56]. The first quantiser partitions the input, for example, into 15 quantisation bins (0-14) as shown in Figure 3-5.

The output of the first quantiser is assigned two or more indexes, one for each description. Based on the received indexes, the MDC decoder estimates the reconstructed signals. If both descriptions are available, for example index 0 for side quantiser 1 and index 0 for side quantiser 2, quantisation bin 0 is chosen to be decoded by the first quantiser. If only index 0 for side quantiser 1 is available, then there are two quantisation bins (0 and 1) available to be decoded. Hence range of values to be decoded by the first quantiser increases and side reconstruction quality decreases. The redundancy and the corresponding side distortion introduced by this MDC algorithm are controlled by the assignment of indexes to each quantisation bin.
Chapter 3. Error Resilience and Multiple Description Coding for 3D Video

3.3.3 Multiple Description Coding Method Based on Transform Coding

In MDC transform coding, the multiple descriptions in the form of transform coefficients are produced from the output of a transform coded block. For two descriptions, a controlled amount of redundancy between two sets of coefficients is introduced. For maximum coding efficiency, coefficients within each description should be uncorrelated, but coefficients between the two descriptions are correlated. Missing coefficients at the decoder can be estimated from the received description.

In [59], pair-wise correlating transform (PCT) is proposed to transform a set of coefficients into two sets of correlated coefficients with controlled redundancy and side distortion. Figure 3-6 shows the basic coding and decoding process for PCT for a single pair of A and B DCT coefficients. The coefficients are firstly quantised and then transformed using forward PCT. Next, C' and D' are individually entropy coded and their resulting bit streams are sent on two separate channels. Inverse PCT and inverse quantisation are applied at the decoder to obtain the reconstructed coefficients pair, A' and B'. In [77], cascaded correlating transform as extension to the pair-wise correlating transform is proposed. This transform performs the same function as pair-wise correlating transform, but uses two transformers in parallel processing order.
Chapter 3. Error Resilience and Multiple Description Coding for 3D Video

Figure 3-6: Coding and decoding process for PCT

In wavelet transform based MDC such as in [71] and [72], the MDC streams are respectively produced from the arranged wavelet coefficients and the partitioned transform domain of the signals. Researchers in [73] investigated another wavelet based MDC, which produces multiple descriptions from the wavelet representation following a checker-board pattern. An MDC scheme and its application in multiple path transport have been investigated in [75]. Lapped orthogonal transform is used in the transform stage and the transformed coefficient is splitted into two descriptions using a checker-board pattern.

3.3.4 Multiple Description Coding Method Based on Sub-Sampling

In MDC sub-sampling, the original signal is decomposed into subsets, either in spatial, temporal or frequency domain, where each description corresponds to different subsets. This algorithm takes advantage of the correlation of the spatially or temporally adjacent video data samples. Examples of algorithms include temporal frame interleaving [60] and spatial pixel interleaving on image samples [61] or motion vectors using quincunx sub-sampling [70].

Figure 3-7: Quincunx sub-sampling

In [70] for example, the MDC streams are generated by encoding the motion vector field into two description using quincunx sub-sampling process. More on temporal frame interleaving MDC, specifically even and odd frames based MDC will be discussed in Section 3.3.6 as it is the basis of the proposed MDC techniques in this thesis.

3.3.5 Other Multiple Description Coding Algorithms

Beside the three types MDC, there are also other MDC algorithms. An implementation of MDC without modification to the source codec is developed, for example in [76], [78] and [79]. In [76], the method adds redundancy by padding zeros in the DCT domain, which results in interpolation.
of original frame and increases correlation between pixels. The padded DCT coefficients are then sub-sampled and coded independently.

The directional MDC scheme developed in [78] is based on one dimensional (either horizontal or vertical) image expansion. It has better performance than zero padding techniques. Pre-processing schemes using spatial-based Least-predictable vectors and temporal approach are investigated in [79]. In this algorithm, a video sequence is divided into two sub-sequences, to be encoded by a standard encoder. Proper pre-processing is used to obtain the sub-sequences so that a controllable amount of redundancy is inserted between the two. A matching pursuit video coder [74] can also produce two descriptions from the two set of atoms generated from the residue of original frame and motion-compensated frame.

### 3.3.6 Even and Odd Frames Based Multiple Description Coding

Many even and odd frames based MDC methods are investigated in the literature due to its simplicity in producing multiple streams. It also introduces no mismatch when only one of the descriptions is received because the decoder uses the same prediction signal as the encoder for each generated description. Compatibility with the existing video coding standard is another advantage for the even and odd frames MDC as the descriptions from this MDC can be decoded by the standard decoder, provided the descriptions are encoded using the standard encoder.

The even and odd MDC basically includes the even and odd video frames into description one and two respectively [60]. An odd frame is predicted from previously reconstructed odd frame only as shown in Figure 3-8. Prediction of the even frame is also similar to the odd frame. For both descriptions in Figure 3-8, frame 0 is actually the first intra frame. It is needed to for description 2 in case description 1 is completely lost. It is important to note that the two descriptions are independently coded so that when only a single stream is received at the decoder, it can be decoded with acceptable quality at lower temporal resolution. This has the advantage of not having to code the mismatch information as in other methods of MDC. The redundancy in even and odd MDC comes from the longer temporal prediction distance compared to standard video coder, which uses the nearest past frame for prediction. Hence, its coding efficiency is reduced. This method is similar to the video redundancy coding (VRC) proposed in [58].
Chapter 3. Error Resilience and Multiple Description Coding for 3D Video

The VRC has a fixed redundancy operating point because the temporal prediction distance is fixed for a chosen encoding frame rate. The multi-state coder proposed in [60] also has a fixed redundancy except that the decoder recovers the missing frames using advanced motion compensated frame interpolation. A comparison of multi-state coding and layered coding is given in [63]. The approach by [64] is built upon this multi-state coder to achieve a flexible unbalance rate of the two streams.

For a practical MDC scheme, it is necessary to control the redundancy to match the network conditions. The MDC method in [55] is similar to the VRC method, but the mismatch between the predicted frames at the encoder and decoder is also coded and appended in both descriptions. The predictor and the mismatch signal quantiser can be varied to control the redundancies. A pre- and post-processing MDC scheme is proposed in [62] for video streaming application. At the pre-processing stage, extra odd frames are inserted in the even stream and vice versa for the odd stream. If single description is received, the post-processing at the decoder attempts to predict the lost frames using the received and extra frames. In [62], performance in error prone environment is not investigated.

Researchers in [61] proposed a MDC system working in 3D wavelet transform (3D DWT) domain. The original sequence is split into even and odd frames. The even and odd frames are grouped in Group of Picture (GOP) and 3D DWT transformed. In [80], two streams of lower-resolution pictures are added to the multiple state video streams. In case one of the streams is lost, a spatial-temporal hybrid interpolation is used to recover the missing frames. Although more bits are required to encode the extra pictures, it has many desirable features for video streaming in peer-to-peer network.

Performance of the multi-state video encoder proposed in [60] is improved by [82] using multi-hypothesis motion prediction at the encoder. Small additional block motion information is introduced, which helps fast error recovery at the decoder. Multi-state video encoder with side information is proposed in [81]. The side information, which is calculated offline at the encoder, will tell the decoder which reconstruction method will give optimal quality. This method
outperforms the original multi-state encoder up to 1dB depending on the loss rates of transmission channels.

3.4 Even and Odd Frames Based Multiple Description Coding for 3D Video

All the MDC discussed before were applied to 2D video to provide error resilience, but so far it has not been applied to 3D video. A method of multiple description coding based on even and odd MDC (MDC-EO) and frame interpolation is extended to 3D video (specifically 2D plus video plus depth information). The MDC method is based on even and odd frame as in [60]. Although MDC has less coding efficiency than SDC in error free environment, simulation results show that the performance of the proposed algorithm out performs SDC especially in high packet loss environments.

3.4.1 The Proposed Even and Odd Frame Based Multiple Description Coding Algorithm

The general block diagram for even and odd frame based MDC for 3D video is shown in Figure 3-9. It is built upon the existing MPEG-4 video coding standard that has Multiple Auxiliary Component tools to support depth information. There are texture part that includes luminance (Y) and chrominance (U and V) components, and also depth part (also called alpha plane) for each macro block in an even/odd video frame. The even and odd frames are predicted from previous even and odd frame respectively as in multi-sate encoder [60]. The even and odd frames are encoded into streams 1 and 2 respectively.

![Figure 3-9: The proposed MDC-EO encoder and decoder block diagram](image)

The content of the two bit streams at the frame level for texture and depth information is shown in Figure 3-10. Streams 1 and 2 contains even and odd frames respectively. The content of the two
Chapter 3. Error Resilience and Multiple Description Coding for 3D Video

bit streams at the macro-block level for texture and depth information is shown in Figure 3-11. The alpha information is actually the depth information.

Even frames

Odd frames

Figure 3-10: Contents of stream 1 and 2 at frame level

Figure 3-11: Contents of stream at macro-block level

If both of the even and odd streams are received, the decoder can reconstruct the coded sequence at full temporal resolution. If only one stream is received, the decoder can still decode the received stream at half the original temporal resolution. Since the even frames are predicted from previous even frames (independent from odd frames), there will be no mismatch if one of the streams are lost at the decoder. Additionally, in the case of one stream is received, the decoder can decode at full resolution by interpolating between the received frames as in [60]. Frame interpolation is performed using the equation below:

\[ I_{\text{ip}}(i, j) = \frac{I_{\text{prev}}(i, j) + I_{\text{fut}}(i, j)}{2} \]  

(3-1)

where \( I_{\text{ip}}(i, j) \) is the frame to be interpolated at pixel location \((i, j)\), \( I_{\text{prev}}(i, j) \) is the previous frame and \( I_{\text{fut}}(i, j) \) is the future frame. This average frame interpolation is used in the simulation when there are errors in a frame. Motion compensated frame interpolation can also be used to obtain improved performance as in [60] but at the expense of decoder complexity.
3.4.2 Detailed Structure of the Proposed Multiple Description Coding Even And Odd Algorithm

MPEG-4-MAC described in Chapter 2 is used in the simulation experiment in this chapter. The even and odd MDC is developed on top of the MPEG-4-MAC codec. The block diagrams of the MPEG-4-MAC-MDC encoder and decoder are shown in Figure 3-12 and Figure 3-13 respectively. A frame buffer is used to store the previous \((n-2)\) reconstructed frame, \(F'(n-2)\). If the input is an even frame, then the coded residual, \(Ecq(n)\) is appended into stream 1 and vice versa for the odd frame. At the decoder, if stream 2 (even frames) is missing, the decoder can either decode at reduced temporal resolution or use the frame interpolation to decode at full resolution.
3.4.3 MDC-Even and Odd with Depth Data Partitioning

In 3.2.5, the depth data is put at the end of the data partitioned block to increase its error resilience. To test the error resilience tools capabilities, the compressed 3D video is transmitted over a Universal Multimedia Telecommunication Services (UMTS) wireless channel (Appendix A) [52]. The UMTS parameters used are Convolutional Code 1/3, Vehicular A and velocity of 50km/h with spreading factors of 8. Eb/No is used to quantify the errors. Mean PSNR, which is the average PSNR at each Eb/No over 20 simulations, is used to measure the quality of the decoded sequence. Readers are referred to [52] for more detailed information on these parameters. The MDC-EO encoder with depth data partitioning is used in the experiment.

Two 2D and depth sequences (Orbi and Interview) at QCIF (176x144) resolutions are compressed by the encoder to achieve bit rate of 300 kbit/s at 10 frames/s. QCIF resolution at 10 frames/s is used in this experiment because it gives good performance for 2D video transmission over UMTS [52]. The large bit rate is used to accommodate the depth data and side information from the MDC encoder. The sequence resolution, bit rate and frame rate chosen are not the ideal one, but the experiment’s objective is just to show the effectiveness of using the error resilience tools on the 2D and depth data. The compressed bit streams are then subjected to error patterns from UMTS simulator ranging from 7 dB Eb/No to 12 db Eb/No, with 12 dB being the least error. Figure 3-14 and Figure 3-15 show the results for Orbi and Interview sequences respectively.

![Figure 3-14: Performance of MDC-EO for Orbi sequence over UMTS channel, ER means the MDC-EO with the error resilience option enabled](image)
Figure 3-15: Performance of MDC-EO Interview sequence over UMTS channel, ER means the MDC-EO with the error resilience option enabled

The results clearly demonstrate the effectiveness of using the error resilience tools on the 2D video and depth under error prone condition. For example, the mean PSNR of luminance for Orbi sequence with error resilience tools is about 5 dB higher than without the error resilience tools and about 10 dB higher for the depth information at Eb/No = 7 dB. The results for Interview sequence also show almost similar performance with the Orbi sequence. Hence, the error resilience options in the encoder will be turned on for all the simulation results in error prone environment in the subsequence chapters.

3.4.4 Performance Evaluation of the Proposed Multiple Description Coding

Even And Odd Algorithm

3.4.4.1 Error Free Environment

To evaluate the coding performance of the encoder in error free environments, a rate distortion curve is plotted for the Orbi sequence. The tests are carried out using CIF format (352x288). The basic encoding parameters are: 300 frames, IPPP... sequence format (only the first frame is encoded as an I-frame and all others are encoded as P-frames), 30 frames/s original frame rate, variable length coding (VLC) and without error resilience. The Quantisation parameter (QP) in the configuration file is varied to obtain the bit rate range shown in the rate-distortion curves. The rate distortion curves show the image quality measured in PSNR (Peak-Signal-to-Noise Ratio) against the resulting bit rate when both of the MDC streams are received in error free, also known as central distortion. Figure 3-16 shows the rate-distortion curves for Orbi colour and depth sequences based on SDC and MPEG-4-MAC-MDC-EO (short form MDC-EO) configurations.
It can be seen that the SDC coder performs better at all the target bit rates than the MDC-EO scheme. This is because MDC-EO has more redundancy than SDC that comes from the inefficient
prediction for frame $n$ from previous $n-2$ frame in MDC-EO. Hence, in error free environments, MDC-EO requires more bits to achieve the same quality as SDC at the same bit rate.

Although a decrease in coding efficiency is a drawback, MDC-EO has performance advantages in error prone environments. In the case of two channels, if one of the channels is error prone (bad channel) and the other channel is good, an MDC-EO decoder will be able to reconstruct the heavily corrupted data from the bad channel, using the less corrupted data from the good channel, at acceptable quality.

### 3.4.4.2 Frames Lost Environment

A number of experiments according to [83] were performed to examine the effectiveness of the proposed MDC-EO for 3D video versus the SDC with both operating at the same bit rate. In the experiments, it is assumed that one frame lost correspond to one packet loss. Interview sequence at CIF (352x288) format is used in the experiments and is coded at 512 kbit/s and 30 frames/s using SDC encoder and MDC-EO encoder. Fixed QP is used in the simulation to obtain the target bit rate.

Three different types of frame losses were investigated: (1) single loss - loss of a single entire frame (frame 80), which corresponds to one packet loss, (2) burst loss - loss of three frames (starting at frame 80), which corresponds to burst losses of 100 ms duration, (3) two burst losses - loss of three frames in two locations spaced apart by 0.667 seconds (starting at frame 80 and frame 100), which corresponds to two burst losses of 100 ms duration, spaced apart by 0.667 seconds.

In the MDC-EO system, independent loss on each channel is assumed. Specifically, in case (2), three frames are lost in the even sequence, and in case (3), three frames are lost in even sequence and the other three frames are lost on odd sequence. In these tests, the MDC-EO use frame interpolation to recover from frame lost and the SDC replaces the lost frame with the previous frame. Figure 3-17, Figure 3-18 and Figure 3-19 illustrate the performance of SDC and MDC-EO in term of frame PSNR under single loss, burst loss and two burst losses respectively for both luminance and depth.
Figure 3-17: Recovered video quality for Interview sequence for single loss (a) luminance (b) depth

With the Interview sequence and the SDC, a single frame loss (Figure 3-17) leads to approximately a 4 dB loss in quality for luminance and 8 dB loss for depth. For MDC-EO, the loss is about 1 dB for luminance and 2.5 dB for depth. It can be seen that in error free, SDC performs better than MDC-EO. In error prone, SDC is more vulnerable to losses compared to MDC-EO.
The 100 ms burst loss (Figure 3-18) has a minimal affect on MDC-EO compared to SDC. It leads to an additional of 4 dB loss of luminance and 4 dB loss of depth for SDC. For MDC-EO, the additional quality loss is minimal for both luminance and depth. As long as the other channel is correctly received, MDC-EO is largely immune to the duration of lost in one channel.
Figure 3-19: Recovered video quality for Interview sequence for two burst losses (a) luminance (b) depth

The effect of a significant loss on a system while it is still recovering from previous loss is shown in Figure 3-19. It is illustrated as two 100 ms burst losses. It can be seen that quality of the SDC is largely affected by the two losses compared to MDC-EO. In MDC-EO, as long as both streams are not lost simultaneously, at least one of the streams can assist in recovering the other.
3.4.4.3 Error Prone Environment

The compressed 3D video is transmitted over a simulated Wireless LAN channel [84]. Same bit rate experiments have been conducted to show the performance of MDC-EO coder compared to SDC. It is when the aggregate bit rate of both MDC-EO streams is the same as the bit rate of the SDC. With this setting, it is expected that at large error rates, MDC-EO will be better than SDC, and at small error rates, SDC will be better than MDC-EO. The bit rate chosen is 512 kbit/s. The QP parameters of the encoder are manually adjusted to achieve 512 kbit/s for SDC and MDC-EO.

Basic parameters are the same as in error free environment, but with error resilience options turned on. To reduce error propagation, adaptive intra refresh (AIR) of 9 and cyclic intra refresh (CIR) of 1 is used in both SDC and MDC-EO. Basic error concealment, which is to replace corrupted macro block with the corresponding macro block in previous frame, is applied to SDC. In the MDC-EO algorithm, if an error occurs in a frame of one stream, the frame is replaced by the interpolated frame from the other stream.

Packet error trace from a Wireless LAN channel simulator developed in SUIT (Scalable, Ultra-fast and Interoperable Interactive Television) project (Appendix B) [84] is used the simulation. The system parameters for the WLAN IEEE802.11g are presented in Table 3-1. In the Wireless LAN channel simulator, combination of channel coding and modulation schemes produces several transmission modes with different data rate. The combination is also called Modulation and Coding Scheme (MCS) mode. In this thesis, packet error trace from MCS mode 0 as shown in Table 3-2 is used to corrupt the compressed 2D and depth video sequences.

For SDC, the 2D and depth video sequences are compressed using the original MPEG4 version 2 reference software. For MDC-EO, modified MPEG4 version 2 reference software is used to compress the sequences. At the reference software decoder, packets are dropped according to the packet error trace generated from the Wireless LAN channel simulator.

<table>
<thead>
<tr>
<th>System Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrier Modulation</td>
<td>OFDM</td>
</tr>
<tr>
<td>FFT Size</td>
<td>64</td>
</tr>
<tr>
<td>Carrier Frequency</td>
<td>2.4 GHz</td>
</tr>
<tr>
<td>Sampling Rate</td>
<td>20 MHz</td>
</tr>
<tr>
<td>Channel Coding</td>
<td>Punctured Convolutional Coding</td>
</tr>
<tr>
<td>Modulation</td>
<td>BPSK, QPSK, 16QAM, 64QAM</td>
</tr>
</tbody>
</table>
Chapter 3. Error Resilience and Multiple Description Coding for 3D Video

Table 3-2: Modulation and Coding Scheme 0 (MCS 0)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modulation</td>
<td>BPSK</td>
</tr>
<tr>
<td>Physical layer bit rate</td>
<td>6 Mbps</td>
</tr>
<tr>
<td>Coded bits per OFDM symbol</td>
<td>48</td>
</tr>
<tr>
<td>Data bits per OFDM symbol</td>
<td>24</td>
</tr>
</tbody>
</table>

Packet loss is used to quantify the errors and is measured in percentage. One video packet per RTP packet is assumed in the simulation [85]. The size of one video packet is set to be 1000 bits. Mean PSNR, which is the average PSNR at each packet loss over 10 simulations, is used to measure the quality of the decoded sequence. The QP used in the simulation for MDC-EO and SDC to achieve 512 kbit/s and its corresponding error free PSNR is shown in Table 3-3 for Orbi sequence and Table 3-4 for Interview sequence. Figure 3-20 and Figure 3-21 show the results for the experiments for the Orbi and Interview sequence respectively. As an example, the frame PSNR for the Orbi and Interview sequence at 20% packet loss is shown in Figure 3-22 and Figure 3-23 respectively.

Table 3-3: Quantisation parameter for Orbi sequence and the corresponding error free PSNR

<table>
<thead>
<tr>
<th>Encoder</th>
<th>Texture</th>
<th>Average Depth</th>
<th>Average PSNR</th>
<th>Depth</th>
<th>Average PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDC</td>
<td>4</td>
<td>8</td>
<td>36.56</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>MDC-EO</td>
<td>9</td>
<td>12</td>
<td>34.00</td>
<td>14</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 3-4: Quantisation parameter for Interview sequence and the corresponding error free PSNR

<table>
<thead>
<tr>
<th>Encoder</th>
<th>Texture</th>
<th>Average Depth</th>
<th>Average PSNR</th>
<th>Depth</th>
<th>Average PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDC</td>
<td>12</td>
<td>7</td>
<td>35.30</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>MDC-EO</td>
<td>5</td>
<td>9</td>
<td>34.16</td>
<td>8</td>
<td>12</td>
</tr>
</tbody>
</table>
Figure 3-20: Mean PSNR vs packet loss for Orbi (a) luminance and (b) depth
Chapter 3. Error Resilience and Multiple Description Coding for 3D Video

![Image: Graph showing Mean PSNR vs packet loss for Interview (a) luminance and (b) depth]

Figure 3-21: Mean PSNR vs packet loss for Interview (a) luminance and (b) depth
Figure 3-22: Frame PSNR for Orbi (a) luminance and (b) depth at 20% packet loss
The subjective quality for Frame 120 of Interview sequence for both texture and depth is shown in Figure 3-24 for SDC and MDC-EO. Figure 3-25 shows the resulting stereoscopic 3D video of Frame 120 for Interview from the DIBR of texture and depth. Figure 3-25 can be viewed using the red and blue glass.
Figure 3-24: Subjective quality – Interview - at 20% packet loss of luminance for (a) SDC and (b) MDC-EO; and depth for (c) SDC and (d) MDC-EO

Figure 3-25: Subjective quality – Interview - at 20% packet loss of stereoscopic 3D video for (a) SDC and (b) MDC-EO
Chapter 3. Error Resilience and Multiple Description Coding for 3D Video

It can be seen from Figure 3-20 that the results are as expected for both texture and depth. For Orbi sequence, at high packet loss, MDC-EO outperforms SDC by about 0.6dB mean PSNR for luminance and about 1 dB for depth. At low packet loss, SDC is better than MDC-EO because SDC's fixed PSNR quality in error free conditions is higher than MDC-EO at the same bit rate. For Interview sequence (Figure 3-21), MDC-EO is slightly better than SDC at 20% packet loss for luminance, and is better than SDC by more than 1 dB for depth at the same packet loss.

The PSNR per frame plot for Orbi in Figure 3-22 also shows that MDC-EO is better than SDC in error prone conditions. Orbi sequence is a complex sequence with camera movement and multiple objects, hence there is greater problem with error propagation with SDC. For example, at 20% packet loss, the average PSNR per frame of MDC-EO for texture is 27.82 dB compared to SDC, 26.93 dB. The difference is larger for the depth data where the average PSNR for MDC-EO is 29.45 and SDC is 26.81 (the same case for Interview sequence). This is because the nature of the depth image that contains mostly black and grey components, which can be concealed easily using frame interpolation. For example, when errors occur in Frame 7 (Figure 3-22(b)) of the depth image, it can easily and quickly be concealed by frame interpolation of MDC-EO, before the errors propagate further, but not for SDC. SDC uses corresponding macro block in the previous frame to conceal a corrupted macro block. This shows that using frame interpolation for depth data in MDC-EO is better than replacing macro block in SDC. SDC also have to depend on the intra refresh to stop error propagation. When errors occur in Frame 7 (Figure 3-22(a)) of the texture image, it could not be easily concealed as the depth image by the frame interpolation. Therefore, on average PSNR, the MDC-EO performs largely better than SDC in depth image, but slightly better than SDC in texture image.

Sometime, SDC perform better than MDC-EO for example between frames 199 to 210 of Orbi (Figure 3-22) and frames 200 to 240 of Interview (Figure 3-23) due to the intra refresh. Figure 3-24 shows the improved texture and depth subjective quality obtained by MDC-EO for Interview sequence in error prone conditions. The achieved stereoscopic 3D video quality in Figure 3-25 from texture and depth also shows improved performance with MDC-EO. Figure 3-25 can be viewed using the red and blue glass.
3.5 Conclusion

Transmitting 3D compressed video over error prone channels is a challenging task. This chapter briefly presents the error resilience tools included by the MPEG-4 standard. The error resilience tools are extended to the depth data in 3D video. It is demonstrated that the performance of the 2D video, additionally the depth data improved under error prone condition with data partitioning and RVLC.

Furthermore, this chapter introduces MDC as a form of error resilience tools. Various MDC algorithms are described in four categories, namely MDC quantisation, MDC transform coding, MDC sub-sampling and other type of MDC. Finally, MDC-EO method is developed and applied to 3D video. The performance is investigated in error free and error prone environment.

Although a decrease in coding efficiency is a drawback, MDC-EO has performance advantages in error prone environments. In frame loss situation, SDC is more vulnerable to errors compared to MDC-EO. SDC's PSNR quality drops is larger than MDC-EO. As long as both of the MDC-EO streams are not corrupted at the same time, the uncorrupted stream can be used to speed up the recovery from error.

Simulation results in an error prone wireless channel, specifically Wireless LAN, show that MDC-EO is objectively better than SDC for both texture and depth data at very high error rates, especially for the depth data. Due to its content, the corrupted frame of depth data that is concealed or replaced using the frame interpolation in MDC-EO is better than the corrupted macro block in a frame of SDC that is replaced with the corresponding macro block in the previous frame. This factor makes the average PSNR of MDC-EO better than SDC for depth information, but slightly better than SDC for texture information in high error rates. In terms of the 3D stereoscopic video, the subjective quality achieved by MDC-EO is shown to be better than SDC at high error rates.
Chapter 4

Even and Odd Frames Based Multiple Description Coding with Side Information for 3D Video

4.1 Introduction

This chapter proposes a novel multiple description coding technique for 3D video based on the even and odd frames MDC developed in Chapter 3. The even and odd frames based MDC is improved by adding variable redundancy in the form of side information (MDC-EOS). The redundant side information consists of the difference between the interpolated frame and the locally reconstructed frame that can be quantised, hence, the redundancies can be controlled by the quantisation parameter. This method is described in Section 4.2. The method is improved in Section 4.3 by including the side information in the central prediction and using the concept of multiple predictions (MDC-EOSP).

The side information is found to be quite large, which affects the coding efficiency of the MDC coder. Hence, in Section 4.4, an attempt is made to reduce the side information using B-frame interpolation (MDC-EOSB). The simulation results for MDC-EOS, MDC-EOSP and MDC-EOSB are presented and discussed in Section 4.5. The side information is also reduced by making it adaptive according to the motion information and is applied to MDC-EOS and MDC-EOSP in Section 4.6. With the adaptive side information, MDC-EOS is renamed to MDC-EOAS and MDC-EOSP is renamed to MDC-EOASP. The simulation results for MDC-EOAS and MDC-EOASP are also presented and discussed in Section 4.6. Section 4.7 concludes this chapter.
4.2 Even and Odd Frames Based Multiple Description Coding with Side Information

MDC-EO in previous chapter performs better than SDC in a high error rate situation. If for example, one stream is corrupted, it can be replaced with the interpolated frames of the other stream provided that the other stream is not in error. The interpolation produces a blurred image, especially if the difference between the frames used in the interpolation is large as shown in the example in Figure 4-1. It also produces large PSNR variation between frames when errors occur as shown in the results in the previous chapter (example in Figure 3-17). The frame PSNR is low for the interpolated frame and high for the uncorrupted frame in the other stream. The frame PSNR for the following frames predicted from the interpolated frame are also affected by the error.

![Figure 4-1: Frame Interpolation](image)

To reduce the PSNR variation and the blurring effect, it is proposed to send controllable side information on top of the MDC-EO at the expense of reduced coding efficiency in error free environments. The general block diagram for our proposed MDC for 3D video (MDC-EOS) is shown in Figure 4-2 (encoder) and Figure 4-3 (decoder). The even and odd frames are encoded into streams 1 and 2 respectively. Each frame contains texture, motion and depth data. Side information for even and odd stream frames is also appended to their corresponding streams.
In Figure 4-2, each even frame is predicted from the previous even frame and each odd frame is predicted from the previous odd frame. Hence, streams 1 and 2 contain even and odd frames respectively along with their corresponding side information. The side information consists of the difference between the interpolated frame and the locally reconstructed frame that can be quantised, hence, the redundancies can be controlled by the quantization parameter (QPside).

Figure 4-3: The proposed MDC-EOS decoder general block diagram
In Figure 4-3, if only one stream is received at the decoder, for example the even stream, the odd frames can be reconstructed from the side information by adding the result of frame interpolation at the decoder to the quantized side information. Depending on the quantisation parameter, the quality of the interpolated frame and the side information can be varied.

It is important to have variable redundancy introduced by MDC so that it can be controlled according to the conditions of the network [55]. To avoid mismatch and at the same time have variable redundancy, side information is appended to the even and odd streams. The detailed operation of the MDC encoder and decoder shown in Figure 4-4 and Figure 4-5 is described as follows:

1. At the encoder, the central encoder is used to produce even or odd frames. The frame buffer is used to store the reconstructed frames, \( F'(n-1) \) and \( F'(n-2) \). Even frames are predicted from previous reconstructed even frames and vice versa for odd frames.

2. Side encoder 1 and 2 are used to produce the side information for even and odd stream respectively. In side encoder 1, frame interpolation is performed between the current reconstructed even frame, \( F_e'(n) \) and the previous reconstructed even frame, \( F_e'(n-2) \). Equation (3-1) is used to produce the interpolated frame. Only side encoder 1 is shown in Figure 4-4, but side encoder 2 has similar structure.

3. The interpolated frame is subtracted from the previous reconstructed frame, \( F'(n-1) \) and the difference, \( E_e(n) \), which is the side information, is coded using DCT and quantisation. Hence, the redundancy introduced can be controlled by varying the quantisation parameter (Q1) of the side information.

4. At the decoder in Figure 4-5, the central decoder is used to decode the central information (even or odd frames). If for example only an even stream is received, side decoder 1 is invoked to recover the odd frame, \( F_o'(n) \). The results of frame interpolation of the previous reconstructed even frame \( F_e'(n-2) \), and previous reconstructed frame \( F_e'(n) \), is added to the decoded side information, \( E_e'(n) \), to get \( F_o'(n) \).
Chapter 4. Even and Odd Frames Based Multiple Description Coding with Side Information for 3D Video

Figure 4-4: Block diagram of the proposed MDC-EOS encoder

In this way, if the quantisation parameter of the side information (Q1) is low, a high quality interpolated frame is produced at the decoder at the expense of higher redundancies. On the other hand, if Q1 is high, a reduced quality interpolated frame is produced but at lower redundancies. These features allow us to control the amount of redundancies needed for the MDC coder.

These operations are extended to include the depth data. The content of the two bit streams at the macro block level for texture and depth information is shown in Figure 4-6(a). The alpha information is actually the depth information. The content of the two bit streams at the macro block level for the side information is shown in Figure 4-6(b). Simulation results for MDC-EOS are discussed in Section 4.5.
4.3 Even And Odd Frames Based Multiple Description Coding with Side Information and Prediction

It is mentioned in [56] that one of the redundancies in a predictive multiple description video coder is any bit-rate used to describe side information in excess of that used by an SDC. For the MDC-EOS method, this extra signal is called \( Ee(n) \) (\( Eon(n) \) for odd frame), which is the

![Diagram](image-url)

Figure 4-5: Block diagram of the proposed MDC-EOS decoder

![Diagram](image-url)

Figure 4-6: Content of bit stream at macro-block level for (a) Central information and (b) Side information
difference between the reconstructed and the interpolated frames. $Ee(n)$ is used to reconstruct the odd frames when only even frames are received. However, in error free conditions, $Ee(n)$ is not used. In other words, this side information is ignored when both descriptions are received, similar to [55].

In both MDC-EO and MDC-EOS method, frame $n$ is predicted from frame $n-2$, causing a decrease in coding efficiency in the central prediction due the usage of predictor that is less efficient than the SDC predictor (in SDC, frame $n$ is predicted from frame $n-1$). Hence we proposed to use the side information, $Ee(n)$, to improve the central prediction in error free conditions. Figure 4-7 shows the general block diagram of the proposed scheme. The decoded $Ee(n)$ is added to the interpolated frame to obtain $F_ip'$ and is used for the prediction of frame $n$. Using the idea of multiple predictions as in [55], frame $n$ is predicted from the superposition ('S' in Figure 4-7) of $F_ip'$ and $F'(n-2)$ frame.

For $n \geq 4$, $n$ is predicted from $P$, which is defined in Equation (4-1),

\[
P = a_1F_{ip'} + a_2F'(n-2)
\]

where $a_1$ and $a_2$ are the weighting factors for $F_{ip'}$ and $F'(n-2)$ respectively. $F_{ip'}$ is the interpolation of frames $n-2$ and frames $n-4$ plus the difference of the interpolated frame and the reconstructed odd frame.
Chapter 4. Even and Odd Frames Based Multiple Description Coding with Side Information for 3D Video

The sum of \(a_1\) and \(a_2\) must be equal to 1 to following the approach of leaky predictor in [55]. Note that \(F_{ep}'\) is equal to the reconstructed \(n-3\) frame if quantisation and inverse quantisation block are absent. Basically, for \(n=4\), the prediction comes from a weighted sum of reconstructed frames \(n-3\) and \(n-2\). The prediction is applied to frame \(n>=4\) because the interpolated frame of \(n-3\) is available only from \(n=4\). As an example, for \(n=3\) frame, the interpolated frame is frame \(n=0\), which is not available. The \(a_1\) and \(a_2\) values can be adjusted to provide different weighted sums of prediction. It is found from experiments, that \(a_1=0.1\) and \(a_2 = 0.9\) gives the best result in terms of PSNR and total bit rate, which means more weight to frame \(n-2\).

\[
P = a_1 F_{ep}' + a_2 F'(n-2)
\]

**Central encoder**
- \(F_{c}(n)\) - current frame
- \(F_{c}'(n)\) - predicted current frame
- \(E_{c}(n)\) - residual error for current frame
- \(E_{eq}(n)\) - coded residual error
- \(F_{ip}'\) - decoded \(E_{eq}(n) + F_{ip}\)
- \(F'(n-2)\) - previous reconstructed even frame

**Side encoder 1**
- \(F_{e}'(n)\) - current reconstructed even frame
- \(F_{e}'(n-2)\) - previous reconstructed even frame
- \(F_{ip}\) - interpolation of \(F_{e}'(n)\) and \(F_{e}'(n-2)\)
- \(F'(n-1)\) - previous reconstructed odd frame
- \(E_{e}(n)\) - side information residual error
- \(E_{eq}(n)\) - coded side information residual error

**Side encoder 2**
- \(F_{e}(n)\) - current frame
- \(F_{e}'(n)\) - predicted current frame
- \(E_{eq}(n)\) - coded residual error
- \(P\) - prediction

**Figure 4-8: Detailed structure of MDC-EOSP encoder**

The detailed block diagram of MDC-EOSP is shown in Figure 4-8. Compared to MDC-EOS, there is a new block called \(P\) in the central prediction section. The output of \(P\) is obtained from equation (4-1). The side encoder section performs frame interpolation between the current even frame, \(F_{e}(n)\), and the previous even frame, \(F_{e}'(n-2)\), to produce \(F_{ip}\). The difference between the interpolated frame, \(F_{ip}\), and the previous reconstructed frame (or the odd frame), \(F'(n-1)\), is coded using DCT and Q1 (side information quantiser) to produce \(E_{eq}(n)\). Decoded \(E_{eq}(n)\) is added.
back to $F_{ip}$ to form $F_{ip'}$. Ideally, $E_e(n)$ should be added back to $F_{ip}$, but to avoid mismatch prediction at the decoder, decoded $E_{eq}(n)$ is used. In other words, $E_e(n)$ is not available at the decoder, but decoded $E_{eq}(n)$ is available to be added to $F_{ip}$.

The difference between this method and [55] is block $P$ is located before the motion compensation process. Also with the proposed configuration, there is no motion vector sent as side information, and no mismatch signal needs to be coded as the even and odd frames are separately predicted. Simulation results for the proposed MDC-EOSP are presented in Section 4.5.

### 4.4 Even and Odd Frames Based Multiple Description Coding using B-Frame

It is found from the rate distortion curve in Section 4.5, that the side information for MDC-EOS is quite large, which affects the coding efficiency of the MDC coder. The side information is the difference between the interpolated frame and the reconstructed frame. One way to reduce the side information is by using better interpolation techniques such as proposed in [60]. In MPEG-4, bi-directional prediction from previous and future frames can be used to produce a B-frame, which also uses frame interpolation in one of its operation modes. Hence, in this section, it is proposed to reduce the side information using B-frame techniques built in MPEG-4 (MDC-EOSB). The overview of the proposed MDC architecture for 3D video using a B-frame is shown in Figure 4-9. Streams 1 and 2 contain even and odd frames respectively along with their corresponding side information. The side information is produced similarly to MDC-EOS and MDC-EOSP, but the B-frame method of prediction is used.

The redundancy in the proposed MDC comes from the fact that the even and odd frames are predicted from previous even and odd frame respectively as in multi-state video coding methods [60]. Additional redundancy is caused by the side information, which consists of the B-frame's difference and motion vectors.

In the proposed scheme, the future, past and current frames are used as shown in Figure 4-9. The difference between MPEG-4 B-frame coding and the proposed technique is that the B-frame is sent as side information. In the example shown in Figure 4-9, the B-frame for frame 1 is sent as the side information in the even stream. Hence, the side information contains the difference between the current frame (reconstructed frame 1) and the B-predicted current frame and the motion vectors. This side information can be quantized using the quantisation parameter of the side information ($QP_{side}$) to control the amount of redundancy introduced.
Chapter 4. Even and Odd Frames Based Multiple Description Coding with Side Information for 3D Video

If only one stream is received at the decoder, for example the even stream, the odd frames can be reconstructed from the decoded B-frame side information as in MPEG-4 B-frame decoding.

Detailed operation of the B-frame MDC encoder is shown in Figure 4-10. The central encoder produces either even or odd streams based on the frame number. For example, after an even frame (the future frame) is encoded, the side encoder 1 is invoked to generate side information, $Ee(n)$, from the future frame (current reconstructed even frame, $F_e'(n)$), past frame (previous reconstructed even frame, $F_e'(n-2)$) and current frame (previous reconstructed frame, $F'(n-1)$). Side encoder 2 is invoked to produce side information for the odd stream. In B-frame generation, bi-directional motion estimation and compensation is employed. It involves forward, backward and bi-directional prediction.

One of the advantages of B-frame is that for a fixed bit rate it has higher PSNR and subjective quality as more information (2 frames) is available for prediction purposes compared to P-frame prediction. Therefore, $F_p(n-1)$ is predicted more accurately than with interpolation using averaging, as in MDC-EOS. Hence, the difference, $Ee(n)$, between the locally reconstructed frame, $F'(n-1)$, and the predicted frame, $F_p(n-1)$, is small and consequently this reduces the redundancy of the side information. Although the motion vectors for the B-frame also need to be sent as side information, experiment has shown that this B-frame method has less redundancy than MDC-EOS at certain bit rate. Simulation results for the proposed MDC-EOSB are presented in Section 4.5

Figure 4-9: The proposed MDC-EOSB block diagram
Chapter 4. Even and Odd Frames Based Multiple Description Coding with Side Information for 3D Video


In this section simulation results for SDC, MDC-EO, MDC-EOS, MDC-EOSP and MDC-EOSB are compared in error free environments, ideal MDC channel and frame loss situations. In ideal MDC channels, it is assumed only one of the streams is received and the MDC decoder will try to reconstruct the full sequence.

4.5.1 Error Free Environment

In error free environment, similar experiment setup as in Section 3.4.4.1 is used. Figure 4-11 shows the rate-distortion curves for Orbi colour and depth sequences for SDC, MDC-EO, MDC-EOS, MDC-EOSP and MDC-EOSB. For MDC-EOS, MDC-EOSP and MDC-EOSB, the quantisation parameter for the side information is set to 20.
Chapter 4. Even and Odd Frames Based Multiple Description Coding
with Side Information for 3D Video

It can be seen that the SDC coder performs better at all the target bit rates than all the MDC schemes for both luminance and depth with MDC-EOSB being the worst especially at bit rates less than 1 Mbit/s. This is because MDC-EOSB has the most redundancy that comes from inefficient prediction for frame \( n \) from previous frame \( n-2 \) and the side information which contains B-frame residuals and motion vectors. It can also be seen from Figure 4-11, that MDC-EOSP and MDC-EOS has comparable performance with MDC-EOSP performing better at bit rates less than 500 kbit/s. It is use of the side information in the central prediction which improves its prediction compared to just using frame \( n-2 \).

Although a decrease in coding efficiency is a drawback, MDC-EOS, MDC-EOSP and MDC-EOSB have performance advantages in ideal MDC channels. In the case of two channels, if one of the channels is error prone (bad channel) and the other channel is good, the MDC decoder will be able to reconstruct the heavily corrupted data from the bad channel, using the less corrupted data from the good channel, at acceptable quality. Compared to MDC-EO, the MDC schemes above have reduced coding efficiency in error free environments, but in frame loss situations and ideal MDC channel, these MDC algorithms can use the redundant side information to produce better interpolated frames than MDC-EO.

4.5.2 Even and Odd Frames Based Multiple Description Coding using B-Frame Performance in a UMTS Network

It was anticipated that in MDC-EOSB, the B-frame residual should be less than the interpolation residual in MDC-EOS because of the better interpolation technique, however, the rate distortion curves in Figure 4-11 prove that MDC-EOSB has more redundancy than MDC-EOS due to the inclusion of B-frame motion vectors in the side information. Hence, in an ideal MDC channel, frame loss situation and Wireless LAN error prone environment that uses 512 kbit/s as a target bit rate, MDC-EOSB is not evaluated due to its very large redundancies at this bit rate in error free.

Nevertheless, MDC-EOSB and MDC-EOS has comparable performance at high bit rates (more than 1 Mbit/s) as the B-frame motion vector bit rate is negligible at this bit rate range [82]. Table 4-1 show the error free performance for MDC-EOSB and MDC-EOS. Quantisation of the central information (QPc) is set to 5 and QPside is also set to 5. 3D sequence, Orbi at CIF resolution (352x288, 125 frames, 10 frames/s) is used in this experiment. It can be seen that at the same quality, the MDC-EOSB requires less bit rate than MDC-EOS.
Figure 4-11: Rate-Distortion curves for ‘Orbi’ sequence (a) Colour image sequence (b) Depth image sequence
Chapter 4. Even and Odd Frames Based Multiple Description Coding with Side Information for 3D Video

Table 4-1: Comparison between MDC-EOS and MDC-EOSB

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Average PSNR (dB) Luminance</th>
<th>Depth</th>
<th>Bit Rate (Mbit/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDC-EOS</td>
<td>39.09</td>
<td>41.27</td>
<td>1.74</td>
</tr>
<tr>
<td>MDC-EOSB</td>
<td>39.09</td>
<td>41.27</td>
<td>1.63</td>
</tr>
</tbody>
</table>

MDC-EOSB performance compared to SDC at high bit rate in a controlled error prone UMTS wireless environment (1 frame lost) [86] is shown in Figure 4-12. The compressed 3D video is transmitted over a Universal Multimedia Telecommunication Services (UMTS) wireless channel (Appendix A) [52].

The UMTS parameters used are Convolutional Code 1/3, Vehicular A and velocity of 50km/h with spreading factors of 16 for both proposed MDC and SDC and at Eb/No = 8dB. 3D sequence, Orbi at QCIF resolution is used in the experiment. The calculated bit rate for both proposed MDC-EOSB and SDC is 2 Mbit/s.

Figure 4-12 (a) and (b) show the frame PSNR results for the luminance and depth respectively for the Orbi sequence. The average luminance PSNR for MDC-EOSB and SDC is 39.53 and 29.05 respectively. The average depth PSNR for MDC-EOSB and SDC is 39.10 and 16.01 respectively.

It can be seen that the proposed MDC-EOSB is much better than SDC when there is error. In Figure 4-12 (a), the error occurs at around frame 20 for both MDC-EOSB and SDC. MDC-EOSB then used the side information of the other stream to replace the corrupted frame. The subsequent frame is then predicted from this new frame. Although there is a large drop in PSNR for this new frame, it is still better than SDC. The proposed MDC-EOSB also manages to quickly recover from the error as in multiple state video coding.

For SDC, when the error occurs at around frame 20, normal error concealment is performed. However, it does not succeed to remove the errors and the errors propagate to the subsequent frames. Figure 4-13 shows the improved texture and depth subjective quality obtained by MDC-EOSB for Orbi sequence in error prone conditions for both texture and depth. The achieved 3D perceptual video quality from texture and depth also shows improved performance with MDC-EOSB.
Chapter 4. Even and Odd Frames Based Multiple Description Coding
with Side Information for 3D Video

Figure 4-12: Objective quality (frame PSNR) at Eb/No=8dB of luminance (a); and depth (b)

Figure 4-13: Subjective quality at Eb/No=8dB of luminance for (a) SDC and (b) MDC-EOSB; and depth for (c) SDC and (d) MDC-EOSB
4.5.3 Error Free Environment with Varying Side Information

The redundant side information in MDC-EOS consists of the difference between the interpolated frame and the locally reconstructed frame that can be quantised, hence, the redundancies can be controlled by the side quantisation parameter, QPside (qps). Figure 4-14 shows the rate distortion when the QPside is varied (qps 5, qps 10 and qps 30) for MDC-EOS and MDC-EOSP of the Orbi sequence coded at 15 frames/s.

It can be seen from Figure 4-14(a) that as QPside for MDC-EOS and MDC-EOSP increases, the amount of side information decreases, thus improving the error free rate distortion performance of the luminance data. For example, the rate distortion of MDC-EOS and MDC-EOSP is very close to MDC-EO when QPside = 30 due to the small amount of redundant side information coded at this quantisation parameter. Similar performance for the depth data is observed in Figure 4-14(b). Hence, these features allow the user to control the amount of redundancy needed for the MDC-EOS and MDC-EOSP coder.

4.5.4 Frame Loss Environment

A number of experiments as in Section 3.4.4.2 were performed to examine the effectiveness of the proposed MDC-EOS and MDC-EOSP for 3D video versus the SDC and MDC-EO when both operating at the same bit rate. The Interview sequence in CIF (352x288) format is used in the experiments and is coded at 512 kbit/s and 30 frames/s using SDC, MDC-EO, MDC-EOS and MDC-EOSP encoder. A fixed QP as shown in Table 4-2 is used in the simulation to obtain the target bit rate.

Table 4-2: QP and QP-Side for SDC, MDC-EO, MDC-EOS and MDC-EOSP to achieve about 512 kbit/s

<table>
<thead>
<tr>
<th>Encoder</th>
<th>Texture</th>
<th>Average PSNR</th>
<th>Depth</th>
<th>Average PSNR</th>
<th>QP Side</th>
<th>Bit rates (kbit/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame</td>
<td>I</td>
<td>35.30</td>
<td>I</td>
<td>38.21</td>
<td>N/A</td>
<td>512.83</td>
</tr>
<tr>
<td>SDC</td>
<td>12</td>
<td>34.16</td>
<td>8</td>
<td>35.83</td>
<td>N/A</td>
<td>512.21</td>
</tr>
<tr>
<td>MDC-EO</td>
<td>9</td>
<td>32.67</td>
<td>12</td>
<td>35.44</td>
<td>20</td>
<td>511.88</td>
</tr>
<tr>
<td>MDC-EOS</td>
<td>6</td>
<td>33.62</td>
<td>10</td>
<td>36.67</td>
<td>20</td>
<td>512.70</td>
</tr>
<tr>
<td>MDC-EOSP</td>
<td>10</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Chapter 4. Even and Odd Frames Based Multiple Description Coding
with Side Information for 3D Video

Figure 4-14: Rate distortion when side quantisation parameter (Qpside) is varied for Orbi sequence
for (a) luminance and (b) depth
Figure 4-15: Recovered video quality for Interview sequence for single loss (a) luminance (b) depth

In these tests, the MDC-EOS and MDC-EOSP decoders use the side information and frame interpolation to recover from frame loss and the SDC replaces the lost frame with the previous frame. Figure 4-15, Figure 4-16 and Figure 4-17 illustrate the performance of SDC, MDC-EO, MDC-EOS and MDC-EOSP in term of frame PSNR under single loss, burst loss and two burst losses respectively for both luminance and depth.
Chapter 4. Even and Odd Frames Based Multiple Description Coding with Side Information for 3D Video

Figure 4.16: Recovered video quality for Interview sequence for burst loss (a) luminance (b) depth

With Interview sequence and the SDC, a single frame loss (Figure 4-15) leads to approximately a 4 dB loss in quality for luminance and 8 dB loss for depth. For MDC-EO, the loss is about 1 dB for luminance and 2.5 dB for depth. For both MDC-EOS and MDC-EOSP, the PSNR drops by less than 1dB for texture and depth, with MDC-EOSP having better quality than MDC-EOS. However, MDC-EOS and MDC-EOSP feature more redundancies than MDC-EO and SDC that make their PSNR curves lower than others. An exceptional case is for MDC-EOSP, where its error free performance is better than MDC-EO for depth, due to usage of lower quantisation.
Figure 4-17: Recovered video quality for Interview sequence for two burst losses (a) luminance (b) depth

parameter. Note that MDC-EO luminance PSNR is better than MDC-EOSP in error free conditions. These observations are also depicted in the burst loss and two burst losses situation in Figure 4-16 and Figure 4-17 respectively. It can be seen that in error free, SDC performs better than the MDCs. In error prone environments, SDC is more vulnerable to losses compared to the MDCs.
4.5.5 Ideal Multiple Description Coding Channel

In this experiment, it is assumed only even MDC streams are received by the decoder. The MDC-E0 decoder reconstructs the full sequence using only frame interpolation while MDC-EOS and MDC-EOSP use frame interpolation and add the interpolated frame to the received side information. It is also assumed that the side information in the even stream is received by the decoder. The same experiment settings as in Section 4.5.4 are used. Performance is not compared with SDC because it has only one stream. Figure 4-18 shows the frame PSNR performance for MDC-E0, MDC-EOS and MDC-EOSP when only even stream is received.

It can be seen in Figure 4-18, from frame 0 up to about frame 70, where there is less motion, MDC-E0 performs better than MDC-EOS and MDC-EOSP for luminance. For depth, MDC-EOSP is the best for the same reason explained in the previous section. This shows that using only interpolation without the side information does not cause much degradation in the performance when motion is insignificant. Above frame 70, where the motion is large, MDC-E0 gives very large PSNR variation between frames compared to the other proposed MDCs. It can be deduced that the side information in MDC-EOS and MDC-EOSP is useful when the motion in the sequence is significant. Hence, it is proposed in Section 4.6 to send the side information adaptively according to the motion parameter in the sequence.

4.5.6 Average Delta Peak Signal-to-Noise Ratio

Table 4-3 indicates the PSNR variation for the results in Section 4.5.4 and Section 4.5.5 measured using average delta PSNR. The absolute differences between frames when the errors occur are averaged to obtain the average delta PSNR. The table shows the performance for SDC, MDC-E0, MDC-EOS and MDC-EOSP for the luminance and the depth for the single loss, burst loss, two burst losses and the ideal MDC channel. The number that has the lowest value for a particular loss type is highlighted in bold.

<table>
<thead>
<tr>
<th>Loss type</th>
<th>Luminance</th>
<th>Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SDC</td>
<td>EO</td>
</tr>
<tr>
<td>Single loss</td>
<td>0.58</td>
<td>0.92</td>
</tr>
<tr>
<td>Burst loss</td>
<td>0.52</td>
<td>0.61</td>
</tr>
<tr>
<td>Two burst losses</td>
<td>0.52</td>
<td>0.71</td>
</tr>
<tr>
<td>Ideal MDC channel</td>
<td>N/A</td>
<td>0.64</td>
</tr>
</tbody>
</table>
Figure 4-18: Recovered video quality for Interview sequence when only the even stream is received
(a) luminance (b) depth

It can be seen that most of the times MDC-EOS has the lowest average delta PSNR when there are errors. This is due to the use of the side information that is added to the interpolated frame. MDC-EO has the highest average delta PSNR because of the blurring effect of the interpolated frame. MDC-EOSP has higher average delta PSNR than MDC-EOS but still lower that MDC-EO, except for single loss for luminance where MDC-EOSP performance is about the same as MDC-EOS. Probably the use of the side information for the central prediction is not effective for the
tested loss type. In the next section, performance of MDC-EOS is investigated in the error prone UMTS networks.

### 4.5.7 Performance of MDC-EOS in a UMTS Network

Two experiments have been conducted to show the error prone performance of MDC-EOS coder compared to SDC for 3D video transmission in UMTS network (Appendix A) [87]. The first one is when the aggregate bit rate of both MDC-EOS streams is the same as the bit rate of the SDC (Same bit rate). With this setting, it is expected that at large error rates, MDC-EOS will be better than SDC, and at small error rates, SDC will be better than MDC-EOS.

The second experiment is when both MDC-EOS and SDC have the same quality in error free conditions (Same quality). It is expected that: i) performance of the MDC-EOS will be better than the SDC at large error rates, ii) the MDC-EOS and the SDC performances are similar at small error rates.

The UMTS parameters used are Convolutional Code 1/3, Vehicular A and velocity of 50km/h with spreading factors of 16 for both MDC-EOS and SDC. Eb/No is used to quantify the errors. Mean PSNR, which is the average PSNR at each Eb/No over 20 simulations, is used to measure the quality of the decoded sequence. Readers are referred to [52] for more detailed information on these parameters.

#### 4.5.7.1 Same Bit Rate Experiment

Figure 4-19(a) and Figure 4-19(b) show the results for the same bit rate experiments for the Interview sequence at 720x576 resolutions. For MDC-EOS, fixed quantisation parameter (QP) of 16 is used for all the frames, resulting in 1101 kbit/s, and for SDC, QP=11 is used to get 1113 kbit/s. The same QP is used for the depth data.

The average PSNR for MDC-EOS and SDC in the error free case is 32.30 and 35.03 respectively. As an example, the frame PSNR for the Interview sequence at Eb/No=7dB is shown in Figure 4-20. The subjective quality for Frame 41 for both texture and depth is shown in Figure 4-21(a), (b) and Figure 4-21(c), (d) respectively.

It can be seen from Figure 4-19 that the results are as expected for both texture and depth. At low Eb/No, MDC-EOS outperforms SDC by about 3-4dB mean PSNR, and at high Eb/No, SDC is better than MDC-EOS because SDC's fixed PSNR quality in error free conditions is higher than MDC at the same bit rate.

The per frame PSNR plot for Interview in Figure 4-20 also shows that MDC-EOS is better in error prone conditions. For example, an Eb/No of 7dB, the average PSNR per frame of MDC-
EOS for texture is 10.56dB compared to SDC, 8.80dB. This is also true for the depth data where the average PSNR for MDC is 5.52 and SDC is 3.06. The frame PSNR in Figure 4-20 is very low because there are many corrupted bits at Eb/No=7db. In fact, every frame has errors. Furthermore, the error resilience options in MPEG4 such as data partitioning and reversible variable length coding are disabled during the experiment for both MDC-EOS and SDC cases.

The first I-frame (Frame 0) and the next 29 frames are corrupted by errors. The sudden increase in PSNR at frame 30 is due to the I-frame (which was less affected by errors) inserted every 30 frames. The very large PSNR variation at about frame 33 to 78 is due to more errors affecting one of the MDC-EOS streams compared to the other stream. So, the PSNR per frame is high (about 14dB) for one stream that is less affected by the errors and low (about 10dB) for the other stream that is largely affected by the errors. Figure 4-21 shows the improved texture and depth subjective quality obtained by MDC-EOS for Interview sequences in error prone conditions for the same bit rate experiment. The achieved 3D perceptual video quality from texture and depth also shows improved performance with MDC-EOS.

4.5.7.2 Same Quality Experiment

Figure 4-22 shows the results for the same quality experiments for the Interview sequence. For MDC-EOS a fixed quantisation parameter (QP) of 18 is used for all the frames, resulting in 1006 kbit/s, and for SDC, QP=25 is used to get 505 kbit/s. The same QP is used for the depth data. The average PSNR for MDC-EOS and SDC in error free environments is 31.67 and 31.26 respectively.

For the same quality experiment, the results are as expected for both texture and depth as seen in Figure 4-22. MDC-EOS performs better than SDC at high error rates and performs similarly to SDC at low error rates. For Interview sequences, the gain achieved by MDC-EOS at high error rates is approximately 3-5dB.

Simulation results in this section show that MDC-EOS is objectively better than SDC for both texture and depth data at very high error rates (8-7dB Eb/No). Moreover, the texture, depth and 3D subjective quality achieved by MDC-EOS is also shown to be better than Single Description Coding (SDC) at high error rates. However, there are still disturbing errors in the MDC-EOS 3D subjective quality. For Interview sequence, the average PSNR at 7dB Eb/No is also quite low (about 13 dB for luminance and 9 dB for depth), although it is higher than the SDC (about 8 dB for luminance and 3 dB for depth).
Figure 4-19: Same bit rate experiment for Interview sequence in error prone environments for (a) luminance and (b) depth at 1100 kbit/s
Figure 4-20: Example of frame PSNR at Eb/No=7dB for Texture

Figure 4-21: Subjective quality at Eb/No=7dB of luminance for (a) SDC and (b) MDC-EOS; and depth for (c) SDC and (d) MDC-EOS
Chapter 4. Even and Odd Frames Based Multiple Description Coding with Side Information for 3D Video

Figure 4-22: Same quality experiment for Interview sequence in error prone, (a) Texture and (b) Depth.
4.5.8 Performance of the Proposed MDC-EO, MDC-EOS, MDC-EOSB and MDC-EOSP Algorithms in a UMTS Network

In this section, the performance of SDC, MDC-EO (EO), MDC-EOS (EOS), MDC-EOSB (EOSB) and MDC-EOSP (EOSP) is investigated in UMTS channel (Appendix A). The UMTS channel parameters are: spreading factor 8, turbo code 1/3, Vehicular A at 50 km/hr and for circuit switch radio bearers. The sequences used in this simulation are Orbi and Interview at QCIF resolution, coded at 300 kbit/s. Data partitioning and reversible variable length coding is enabled during the encoding to ensure error resilience. Adaptive Intra Refresh of 9 and Cyclic Intra Refresh of 1 are used. Error free average PSNR obtained by using the quantisation parameter for the algorithms are shown in Table 4-4 and Table 4-5. The average PSNR is obtained at 300 kbit/s bit rate. Figure 4-23 and Figure 4-24 show the mean PSNR obtained versus the error rates in Eb/No for Orbi and Interview respectively.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Texture</th>
<th>Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDC</td>
<td>41.4</td>
<td>40.99</td>
</tr>
<tr>
<td>EO</td>
<td>36.1</td>
<td>43.19</td>
</tr>
<tr>
<td>EOS</td>
<td>31.44</td>
<td>33.66</td>
</tr>
<tr>
<td>EOSB</td>
<td>32.18</td>
<td>33.98</td>
</tr>
<tr>
<td>EOSP</td>
<td>32.79</td>
<td>35.75</td>
</tr>
</tbody>
</table>

Table 4-5: Error free average PSNR for Interview sequence

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Texture</th>
<th>Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDC</td>
<td>40.25</td>
<td>43.48</td>
</tr>
<tr>
<td>EO</td>
<td>40.11</td>
<td>37.06</td>
</tr>
<tr>
<td>EOS</td>
<td>31.84</td>
<td>34.58</td>
</tr>
<tr>
<td>EOSB</td>
<td>32.49</td>
<td>34.83</td>
</tr>
<tr>
<td>EOSP</td>
<td>37.28</td>
<td>38.14</td>
</tr>
</tbody>
</table>

For Orbi sequence, it can be seen from Figure 4-23 that EOSP and EOS performance is almost similar over the error rate range for both luminance and depth. At 10 to 12 dB Eb/No, EOSP performs slightly better than EOS. For the luminance, EO performs worst than the others followed by SDC. For the depth, SDC performance is the worst, followed by EO. This is probably because of the nature of the depth image that contains mostly black and grey components, which can be concealed easily using frame interpolation in EO, compared to concealing the luminance information. In both luminance and depth, EOSB performs better than SDC and EO although the error free PSNR is much less than SDC and EO.

For Interview sequence, performance of EOSP seems to be comparable with EOS, and is slightly higher than EO and EOSB. For the depth information, SDC performs worst that the others.
Chapter 4. Even and Odd Frames Based Multiple Description Coding with Side Information for 3D Video

Figure 4-23: Comparison of SDC and MDC algorithms in UMTS channel for Orbi sequence (a) luminance (b) depth

The reason of these variations is probably due to the usage of different quantisation parameters to obtain the target bit rate of 300 kbit/s. It can be seen from Table 4-4 and Table 4-5 that there are some variations in the error free average PSNR of the algorithms. As an example, for Orbi sequence, average PSNR for luminance SDC is 41.40 dB and depth 40.99 dB, while EO has 36.10 dB for luminance and 43.19 dB for the depth. These factors make the comparison quite difficult. Nevertheless, a rough idea of the algorithms performance can be seen in Figure 4-23 and Figure 4-24.
Figure 4-24: Comparison of SDC and MDC algorithms in UMTS channel for Interview sequence (a) luminance (b) depth

As a summary, although the error-free PSNR for the MDC algorithms are much less than the SDC algorithm, the MDC algorithms outperform SDC in error prone condition by quite huge gain in UMTS environment. This shows that MDC with or without side information is very useful in the error prone environment. It can also be noted from the results that the average PSNR obtained is quite low at high error rate (less than 10 dB Eb/No). In addition to that, the bit rate, frame rate and the size of the sequence are not the ideal one (Bit rate: 300 kbit/s, frame rate: 10 frames/s, frame size: QCIF). These parameters are selected to match the throughput requirement of the UMTS simulation channel.
Chapter 4. Even and Odd Frames Based Multiple Description Coding with Side Information for 3D Video

With spreading factor of 8 (SF8) and 1/3 turbo coding, about 300 kbit/s throughput is available for video data, but there are lots of error with SF8 as can be seen in the low average PSNR value in Figure 4-23 and Figure 4-24. Hence, in further simulations (including MDC-EO in Chapter 3), a different error pattern will be used namely Wireless LAN. Different size of image will also be used (CIF) as the 3D effect is more appealing in CIF image rather than the small QCIF image. The encoding frame rate is also changed from 10 frames/s to 30 frames/s for better temporal resolution. These changes will increase the target bit rates and indeed the target bit rates is changed from 300 kbit/s to 512 kbit/s. But this will not be a problem as the Wireless LAN can support up to 6 Mbit/s of physical layer bit rates.

4.5.9 Performance of MDC-EO and MDC-EOS Algorithms in Wireless LAN Network

The performance of MDC-EOS in a controlled error prone environment for the Wireless LAN (Appendix B) is shown in Figure 4-25 for the Orbi sequence and Figure 4-26 for the Interview sequence. In this environment, the error is confined so that it will not corrupt both the MDC streams at the same time.

The results in this section are obtained using Orbi and Interview sequence with CIF resolution at frame rate of 30 frames/s and bit rate of 512 kbit/s. Wireless LAN error pattern as firstly described in Chapter 3 is used in the simulation. The error is confined so that it does not corrupt both the MDC streams simultaneously. The side information in MDC-EOS is also not corrupted. For the Orbi sequence, the performance of the MDC-EOS is better than SDC at high packet loss. However, MDC-EO outperforms MDC-EOS in this experiment due to its superior performance in error free situation. Similar results are obtained with the Interview sequence. These results show that the side information in MDC-EOS may not be very useful in certain condition in error prone environment. In this experiment, frame interpolation alone (without side information) is enough to provide acceptable quality as high frame rate (30 frames/s) is used. It means that the difference between frames is not large, compared to the 10 frames/s used in the UMTS environment in Section 4.5.8. Therefore, frame interpolation can work well without the side information especially when there is not much motion in the sequence. Furthermore, the side information contributes to the high redundancies of MDC-EOS in error free environment. In Section 4.6, the side information is sent adaptively according to the motion in the sequence.
Figure 4-25: Mean PSNR vs packet loss for Orbi (a) luminance and (b) depth
Figure 4-26: Mean PSNR vs packet loss for Interview (a) luminance and (b) depth
4.6 Even and Odd Frames Based Multiple Description Coding with Adaptive Side Information

The reduced coding efficiency of MDC-EOS and MDC-EOSP is due to the large redundancies in the side information. Hence, it is proposed in this section to send the side information adaptively according to the motion in the sequence. If the motion is large and bigger than a threshold, side information is appended to the bit stream. If not, no side information is sent. This is because, at low motion, using just interpolation does not cause much degradation as shown in Section 4.5.5.

A method in [88] is used to estimate the amount of motion between frames. It exploits the data partitioning mode of MPEG-4 that placed the motion in first partition of the video packet. A value named ‘A’, which is the proportion of the video packet size occupied by the first partition, can be related to the amount of motion. ‘A’ can be expressed as:

\[ A = \frac{Y_{MB}}{X_{MB} + Y_{MB}} \]  

(4-2)

where \( Y_{MB} \) is the average number of bits per macro block in the first partition, and \( X_{MB} \) is the average number of bits per macro block in the second partition. Figure 4-27 shows the variation in ‘A’ over the Orbi and Interview sequences used in this thesis for 300 frames. The period of high motion can be detected through the large values in ‘A’. In the Interview sequence for example, this period is after about frame 70 when the two subjects shake their hand.

The side information for MDC-EOS and MDC-EOSP is then sent according to this ‘A’ value. The MDC-EOS and MDC-EOSP are now known as MDC-EOAS and MDC-EOASP respectively. If the value of ‘A’ is bigger than a pre-determined threshold, then the side information is sent. The threshold value is determined from Figure 4-27. It was selected so that only minimum needed amount of side information is sent. For Orbi the threshold value is set to 0.34 and for Interview threshold value is 0.15.
Figure 4-27: Variation of $A$, the proportion of a packet occupied by the first partition, over the (a) Orbi and (b) Interview sequence.
4.7 Performance Evaluation of the Proposed Multiple Description Coding with Adaptive Side Information

4.7.1 Error Free Environment

In error free environment, similar experiment setup as in Section 3.4.4.1 is used. Figure 4-28 shows the rate-distortion curves for Orbi colour and depth sequences for SDC, MDC-EO, MDC-EOAS and MDC-EOASP. For MDC-EOAS and MDC-EOASP, the quantisation parameter for the side information is set to 20.

As can be seen from Figure 4-28, the error free performance for MDC-EOAS and MDC-EOASP have improved compared to MDC-EOS and MDC-EOSP in Figure 4-11. The rate distortion curve is quite close to MDC-EO because most of the side information is not sent as it is below the threshold. Hence, more bits are available to send the central information using a lower quantisation parameter. Figure 4-29 and Figure 4-30 show the improvement obtained by MDC-EOAS and MDC-EOASP respectively. At the same bit rate, the MDC algorithms with adaptive side information are about 1 to 2 dB better than without the adaptive side information.

4.7.2 Frame Lost Environment

Similar experiments setup, as in Section 3.4.4.2, is performed to examine the effectiveness of the proposed MDC-EOAS and MDC-EOASP for 3D video. Fixed QP as shown in Table 4-6 is used in the simulation to obtain the target bit rate.

In these tests, the MDC-EOAS and MDC-EOASP decoder uses the side information adaptively according to motion and frame interpolation to recover from frame lost. Figure 4-31, Figure 4-32 and Figure 4-33 illustrate the performance of SDC, MDC-EO, MDC-EOAS and MDC-EOASP in terms of frame PSNR under single loss, burst loss and two burst losses respectively for both luminance and depth.
Chapter 4. Even and Odd Frames Based Multiple Description Coding with Side Information for 3D Video

Figure 4-28: Rate-Distortion curves for 'Orbi' sequence (a) Colour image sequence (b) Depth image sequence
Chapter 4. Even and Odd Frames Based Multiple Description Coding
with Side Information for 3D Video

Figure 4-29: Rate-Distortion curves for MDC-EOS and MDC-EOAS for ‘Orbi’ sequence (a) Colour image sequence (b) Depth image sequence
Figure 4-30: Rate-Distortion curves for MDC-EOSP and MDC-EOASP for ‘Orbi’ sequence (a)
Colour image sequence (b) Depth image sequence
Figure 4-31: Recovered video quality for Interview sequence for single loss (a) luminance (b) depth
Figure 4-32: Recovered video quality for Interview sequence for burst loss (a) luminance (b) depth
From Figure 4-31, it can be seen that the performance of the MDC algorithms with adaptive side information is better than SDC and MDC-EO when error occurs. The error free performance of the MDC algorithms with adaptive side information is comparable to MDC-EO for both luminance and depth. After a single frame loss, MDC-EOAS and MDC-EOASP have less frame...
PSNR variation than MDC-EO. As an example, when frame 80 is lost in the Interview sequence, MDC-EO loses about 1 dB and 2 dB for luminance and depth respectively. MDC-EOAS and MDC-EOASP loses less than 1 dB for luminance and depth. It can be concluded that in error free conditions, SDC performs better than the MDCs and in frame lost situation, MDCs recover more quickly than SDC. These observations are also true in single burst loss and two burst losses situation as shown in Figure 4-32 and Figure 4-33 respectively. The performance for MDC-EOAS and MDC-EOASP compared to MDC-EOS and MDC-EOASP respectively is given in Section 4.6.4 and Section 4.6.5.

4.7.3 Ideal Multiple Description Coding Channel

In this experiment, the MDC-EO decoder reconstructs the full sequence using frame interpolation while MDC-EOAS and MDC-EOASP use frame interpolation and add the interpolated frame to the adaptively received side information. The same experiment settings as in Section 4.5.5 are used. Figure 4-34 shows the frame PSNR performance for MDC-EO, MDC-EOAS and MDC-EOASP when only the even stream is received.

It is found that MDC-EOAS and MDC-EOASP managed to achieve comparable performance with MDC-EO in the low motion part of the sequence (frame 0 to frame 70) because no side information is sent for these frames. In the high motion part of the sequence (for example frame 70 to frame 90), where the side information is sent, frame PSNR variation for MDC-EOAS and MDC-EOASP are less than MDC-EO due to the better interpolated frame with side information. MDC-EOASP performance is slightly better than MDC-EOAS for the depth information.

4.7.4 Comparison of the Proposed Multiple Description Coding with and without Side Information

In this section performance of MDC algorithms with adaptive side information is compared MDC algorithm with side information. As an example, results for only two burst losses and ideal MDC channel loss types are shown. Figure 4-35 and Figure 4-36 show the performance of MDC-EOAS compared to MDC-EOS, MDC-EO and SDC for the two burst losses and the ideal MDC channel respectively. Figure 4-37 and Figure 4-38 show the performance of MDC-EOASP compared to MDC-EOSP, MDC-EO and SDC for the two burst losses and the ideal MDC channel respectively.
Chapter 4. Even and Odd Frames Based Multiple Description Coding with Side Information for 3D Video

Figure 4-34: Recovered video quality for Interview sequence when only even stream is received (a) luminance (b) depth
Figure 4-35: Recovered video quality for Interview sequence for two burst losses (a) luminance (b) depth for SDC, MDC-EO, MDC-EOS and MDC-EOAS
Figure 4-36: Recovered video quality for Interview sequence when only even stream is received (a) luminance (b) depth for MDC-EO, MDC-EOS and MDC-EOAS

It can be seen from Figure 4-35 and Figure 4-36 that performance of MDC-EOAS is better than MDC-EOS for luminance and comparable for depth in error free and error prone frames. not sent for MDC-EOAS as it is calculated as the low motion (Figure 4-35, frame 73).
Figure 4-37: Recovered video quality for Interview sequence for two burst losses (a) luminance (b) depth for SDC, MDC-EO, MDC-EOSP and MDC-EOASP
Figure 4-38: Recovered video quality for Interview sequence when only even stream is received (a) luminance (b) depth for SDC, MDC-EO, MDC-EOSP and MDC-EOASP

It can also be seen from Figure 4-37 and Figure 4-38 that performance of MDC-EOASP is better than MDC-EOSP for luminance in error free and error prone frames. However, MDC-EOASP performance is less than MDC-EOSP for depth in error free and error prone frames due to the use
of lower quantisation parameter for MDC-EOSP as shown in Table 4-6. Hence, MDC-EOSP is better than MDC-EOASP in error free and error prone frames for the depth information.

Table 4-6: Quantisation parameter for Interview sequence and the corresponding error free PSNR to achieve 512 kbit/s

<table>
<thead>
<tr>
<th>Encoder</th>
<th>Texture</th>
<th>Average PSNR</th>
<th>Depth</th>
<th>Average PSNR</th>
<th>QP Side</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SDC</td>
<td>I</td>
<td>35.30</td>
<td>I</td>
<td>38.21</td>
<td>N/A</td>
</tr>
<tr>
<td>MDC-EO</td>
<td>I</td>
<td>34.16</td>
<td>P</td>
<td>35.83</td>
<td>N/A</td>
</tr>
<tr>
<td>MDC-EOS</td>
<td>P</td>
<td>32.67</td>
<td>I</td>
<td>35.44</td>
<td>20</td>
</tr>
<tr>
<td>MDC-EOSP</td>
<td>P</td>
<td>33.62</td>
<td>P</td>
<td>36.67</td>
<td>20</td>
</tr>
<tr>
<td>MDC-EOAS</td>
<td>P</td>
<td>33.99</td>
<td>I</td>
<td>34.98</td>
<td>15</td>
</tr>
<tr>
<td>MDC-EOASP</td>
<td>P</td>
<td>33.99</td>
<td>I</td>
<td>35.14</td>
<td>15</td>
</tr>
</tbody>
</table>

4.7.5 Average Delta Peak Signal-to-Noise Ratio

A better way to measure the performance the MDC algorithm performance is by using average delta PSNR to see the PSNR variation when error occurs. Table 4-7 indicates the PSNR variation measured using average delta PSNR. The absolute differences between frames when the errors occur are averaged to obtain the average delta PSNR. The table shows the performance for SDC, MDC-EO, MDC-EOS, MDC-EOSP, MDC-EOAS and MDC-EOASP for the luminance and the depth for the single loss, burst loss, two burst losses and the ideal MDC channel. The number that has the lowest value for a particular loss type is highlighted in bold.

Table 4-7: PSNR variation for SDC, MDC-EO, MDC-EOS, MDC-EOSP, MDC-EOAS and MDC-EOASP

<table>
<thead>
<tr>
<th>Loss type</th>
<th>Luminance</th>
<th>Depth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SDC</td>
<td>EO</td>
</tr>
<tr>
<td>Single loss</td>
<td>0.58 0.92 0.46 0.45 0.50 0.2</td>
<td>1.08 1.64 0.59 0.66 0.65 0.37</td>
</tr>
<tr>
<td>Burst loss</td>
<td>0.52 0.61 0.36 0.44 0.41 0.29</td>
<td>0.67 0.85 0.43 0.55 0.52 0.47</td>
</tr>
<tr>
<td>Two burst losses</td>
<td>0.52 0.71 0.38 0.44 0.46 0.31</td>
<td>0.65 0.84 0.43 0.57 0.51 0.42</td>
</tr>
<tr>
<td>Ideal MDC channel</td>
<td>N/A</td>
<td>0.64</td>
</tr>
</tbody>
</table>
Chapter 4. Even and Odd Frames Based Multiple Description Coding with Side Information for 3D Video

It can be seen that most of the times MDC-EOASP has the lowest average delta PSNR when there are errors. This is due to the use of the adaptive side information that is added to the interpolated frame and also the use of side information in the central prediction. MDC-EO has the highest average delta PSNR because of the blurring effect of the interpolated frame. For burst loss, MDC-EOASP has comparable performance with MDC-EOS.

MDC-EOS has the lowest average delta PSNR for ideal MDC channel because of the addition of the side information to the interpolated frame. MDC-EOAS and MDC-EOASP are not that effective in ideal MDC channel as the side information is only sent when there are high motions. Nevertheless, in the next two sections, the performance of MDC-EOAS and MDC-EOASP are investigated in the error prone wireless LAN networks. MDC-EOAS and MDC-EOASP have better error free rate distortion performance than MDC-EOS and MDC-EOSP respectively as shown in Section 4.6.1.

4.7.6 Error Prone Environment Rate Distortion

The compressed 3D video is transmitted over a simulated Wireless LAN channel (Appendix B). In this section, the rate distortion of the corruptly decoded sequence is plotted for the Orbi sequence, which is encoded at CIF resolution, at 30 frames/s. As an example, Wireless LAN error pattern of SNR 4.2 is used in the simulation, which corresponds to about 12% packet loss. It is assumed that one packet loss corresponds to one frame lost. The side information in MDC-EOAS and MDC-EOASP is not corrupted. Error resilience options are switched off for all algorithms and only the first frame is coded as I-frame. The SDC uses frame copy for error concealment while the MDCs interpolate and add the side information when error occurs. Figure 4-39 shows the results of the simulation.

It can be seen that MDC-EO, MDC-EOAS and MDC-EOASP are better than SDC in the error prone situation. MDC-EOAS and MDC-EOASP better than MDC-EO at bit rate more than 4 Mbit/s. A similar result is obtained for the depth information. There are big gap between SDC and the MDC algorithms because the MDC algorithms use the redundant information to recover from error. Furthermore, in MDC-EOAS and MDC-EOASP, the side information is not corrupted. In addition to that, the error resilience options are not enabled in the simulation and no other intra refresh frame is available except the first frame. Hence, the error propagates more in SDC, while the MDC manage to reduce the error propagation through the frame interpolation and the uncorrupted side information. Results with side information corrupted are presented in next section for a fixed bit rate of 512 kbit/s.
Chapter 4. Even and Odd Frames Based Multiple Description Coding with Side Information for 3D Video

Figure 4-39: Rate-Distortion curves for ‘Orbi’ sequence (a) Colour image sequence (b) Depth image sequence when subjected to Wireless LAN error pattern of signal-to-noise ratio (SNR) of 4.2
4.7.7 Same Bit Rate Experiment in Error Prone Environment

For this section, the compressed 3D video is transmitted over a simulated Wireless LAN channel (Appendix B) \[84\] as in Section 3.4.4.3. Similar encoding parameters are used. If an error occurs in a frame of one stream of the MDC-EOAS and MDC-EOASP algorithm, the frame is replaced by the interpolated frame from the other stream plus the adaptively received side information. In this section the side information may be corrupted by the error.

The QP used in the simulation for SDC, MDC-EO, MDC-EOAS and MDC-EOASP to achieve 512 kbit/s and its corresponding error free PSNR is shown in Table 4-8 for Orbi sequence and Table 4-9 for Interview sequence. Figure 4-40 and Figure 4-41 show the results for the experiments for the Interview and Orbi sequence respectively.

Table 4-8: Quantisation parameter for Orbi sequence and the corresponding error free PSNR

<table>
<thead>
<tr>
<th>Encoder</th>
<th>Texture</th>
<th>Average PSNR</th>
<th>Depth</th>
<th>Average PSNR</th>
<th>QP Side</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame</td>
<td>I</td>
<td>36.56</td>
<td>I</td>
<td>8</td>
<td>38.66</td>
</tr>
<tr>
<td>SDC</td>
<td>4</td>
<td>34.00</td>
<td>14</td>
<td>14</td>
<td>35.90</td>
</tr>
<tr>
<td>MDC-EO</td>
<td>9</td>
<td>33.67</td>
<td>5</td>
<td>17</td>
<td>35.21</td>
</tr>
<tr>
<td>MDC-EOAS</td>
<td>15</td>
<td>32.56</td>
<td>25</td>
<td>25</td>
<td>33.24</td>
</tr>
<tr>
<td>MDC-EOASP</td>
<td>15</td>
<td>32.56</td>
<td>25</td>
<td>25</td>
<td>33.24</td>
</tr>
</tbody>
</table>

Table 4-9: Quantisation parameter for Interview sequence and the corresponding error free PSNR

<table>
<thead>
<tr>
<th>Encoder</th>
<th>Texture</th>
<th>Average PSNR</th>
<th>Depth</th>
<th>Average PSNR</th>
<th>QP Side</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame</td>
<td>I</td>
<td>35.30</td>
<td>I</td>
<td>13</td>
<td>38.21</td>
</tr>
<tr>
<td>SDC</td>
<td>12</td>
<td>34.16</td>
<td>8</td>
<td>12</td>
<td>35.83</td>
</tr>
<tr>
<td>MDC-EO</td>
<td>5</td>
<td>33.99</td>
<td>6</td>
<td>15</td>
<td>34.98</td>
</tr>
<tr>
<td>MDC-EOAS</td>
<td>3</td>
<td>33.99</td>
<td>4</td>
<td>16</td>
<td>35.14</td>
</tr>
<tr>
<td>MDC-EOASP</td>
<td>4</td>
<td>33.99</td>
<td>4</td>
<td>16</td>
<td>35.14</td>
</tr>
</tbody>
</table>
Figure 4-40: Mean PSNR vs packet loss for Interview (a) luminance and (b) depth
Chapter 4. Even and Odd Frames Based Multiple Description Coding with Side Information for 3D Video

From the mean PSNR results, it can be seen that for the Interview sequence, MDC-EOASP result is comparable to MDC-EOAS for luminance and slightly better for depth. MDC-EOAS and MDC-EOASP is also better than SDC and MDC-EO at packet loss above 10%. At 20% packet loss, MDC-EOASP mean PSNR is about 0.5 dB better than SDC for luminance and about 3 dB better than SDC for depth. The small gain in luminance achieved by MDCs algorithms in this error prone environment is probably due to the corruption of both MDC streams at the same time, which, violate MDC assumptions.

Nevertheless, more gain is achieved in depth than luminance. Due to its content, the corrupted frame of depth data that is concealed or replaced using frame interpolation and the side information in MDC-EOASP is better than corrupted macro block in a frame of SDC that was replaced with the corresponding macro block in the previous frame. This factor makes the average PSNR of MDC-EOASP is largely better than SDC for depth information, but slightly better than SDC for luminance information in high error rates.

Error free performance of MDC-EOAS and MDC-EOASP are comparable to MDC-EO because their coding efficiency is quite close as the side information is adaptively sent to the decoder. A similar pattern of results can be observed in the Orbi sequence in Figure 4-41. However, the error free performance for MDC-EOASP and MDC-EOAS (for depth) is not quite so close to MDC-EO due to the usage of higher quantisation parameter than with Interview as shown in Table 4-8 to achieve the target bit rate. Hence, the error prone performance is not much different from MDC-EO. The side information, which is adaptively sent for Orbi is also less than Interview as shown in Figure 4-27, therefore it could not help much when error occurs during large motion in the Orbi sequence.

The luminance subjective quality of frame 78 for the Interview sequence when subjected to 20% packet loss is shown in Figure 4-42. The luminance PSNR for that frame is 26.79 dB, 28.54 dB, 31.17 dB and 31.33 dB for SDC, MDC-EO, MDC-EOAS and MDC-EOASP respectively. The depth frame PSNR for Figure 4-43 is 24.24 dB, 27.47 dB, 32.18 dB and 32.10 dB for SDC, MDC-EO, MDC-EOAS and MDC-EOASP respectively. The 3D stereoscopic video quality in Figure 4-44 also shows an improved performance with MDC-EOAS and MDC-EOASP. As before, Figure 4-44 can be viewed using the red and blue glass.

4.8 Conclusion

In this chapter, MDC algorithms were developed to mitigate the effect of channel errors on 3D stereoscopic video sequences. Novel MDC-EOS and MDC-EOSP are proposed as improvements to MDC-EO. It is found that their coding efficiency is reduced compared to MDC-EO and SDC in
Figure 4-41: Mean PSNR vs packet loss for Orbi (a) luminance and (b) depth
error free and error prone environments due to the large amount of side information that is introduced. Nevertheless, in frame loss situations and in ideal MDC channels, MDC-EOS and MDC-EOSP feature lower PSNR variation compared to MDC-EO and SDC. MDC-EOS also performs better than SDC in a controlled error prone wireless LAN environment and in a UMTS channel.

MDC-EOSB is also developed in this chapter with the objective of reducing the amount of side information by using B-frame interpolation method. Comparable rate distortion performance with MDC-EOS is achieved at high bit rate, but at low bit rates, MDC-EOSB performance deteriorate because of inefficient coding of the B-frame motion vector in the side information. On the other hand, MDC-EOSB performance is better than SDC at high bit rates in an error prone wireless environment.

Figure 4-42: Subjective quality – Interview - at 20% packet loss of luminance for (a) SDC and (b) MDC-EO (c) MDC-EOAS and (d) MDC-EOASP
A comparison of SDC, MDC-EO, MDC-EOS, MDC-EOSB and MDC-EOSP in an error prone UMTS channel is also presented in this chapter. Although the error-free PSNR for the MDC algorithms are much less than the SDC algorithm to achieve the same bit rate, the MDC algorithms out perform SDC in error prone condition by quite huge gain. This shows that MDC with or without side information is very useful in the error prone environment.

The side information in MDC-EOS and MDC-EOSP contributes to the high redundancy of these algorithms. Hence, it is also proposed in this chapter to send the side information adaptively according to the motion in the sequence. Large motion will make the algorithm send the side information and no side information is sent if the motion is low. Novel MDC-EOAS and MDC-EOASP are developed to send the adaptive side information. The coding efficiency of these two algorithms is better than MDC-EOS and MDC-EOSP and very close to MDC-EO. Their frame PSNR variation in frame lost and ideal MDC channel are less than MDC-EO. In frame lost situation, the performance is better than SDC, and better than MDC-EO if the side information is
received by the decoder. The error prone performance of MDC-EOAS and MDC-EOASP is better than SDC and MDC-EO at high packet loss objectively and subjectively. Similar to MDC-EO, the gain achieved by MDC-EOAS and MDC-EOASP for depth is larger than the gain achieved for luminance as in Chapter 3.

![Subjective quality - Interview - at 20% packet loss of stereoscopic 3D video for (a) SDC (b) MDC-EO (c) MDC-EOAS (b) MDC-EOASP](image)

As a conclusion, MDC with side information is promising approach to combat channel errors for stereoscopic 3D video transmission, but the side information should be carefully sent as it can cause huge redundancies. It can be sent adaptively according to motion, and ideally, according to the network condition. The novelty in this chapter is about sending and using the side information for stereoscopic 3D video. Trade off between central distortion (rate distortion performance in error free) and side distortion (rate distortion performance in ideal MDC channel) also need to be taken into account in developing an effective MDC algorithm.
Chapter 5

Scalable Multiple Description Coding for 3D Video

5.1 Introduction

Modern video transmission and storage systems are typically characterised by a wide range of access network connection qualities and receiving devices. Examples of applications include video conferencing in a virtual collaboration system scenario, where a large, fixed-terminal acts as the main control point and serves a group of remote users. The remotely located users possess small, fixed/mobile terminals connected via an error prone network.

Scalable video coding is an attractive solution for these application scenarios. Scalable coding produces a number of hierarchical descriptions that provide flexibility in terms of adaptation to user and network requirements. However, in error prone environments, the loss of a lower layer in the hierarchical descriptions prevents all higher layers being decoded, even if the higher layers are correctly received, which means that a significant amount of correctly received data must be discarded in certain channel conditions. On the other hand, an error resilient technique, namely MDC, divides a source into two or more correlated layers. This means that a high-quality reconstruction is available when the received layers are combined and decoded, while a lower, but still acceptable quality reconstruction is achievable if only one of the layers is received.

In this chapter, the combination of scalable coding and multiple description coding (scalable MDC) is proposed to improve error robustness, and at the same time provides adaptability to bandwidth variations and receiving device characteristics. For more immersive communication experiences, stereoscopic 3D video content will be used as a source for the scalable multiple description coder.

Section 5.2 provides related works on the scalable video coding, the scalable extension of the H.264/AVC (SVC) and the available scalable MDC video coding schemes. It is then followed by a description of our proposed scalable MDC to 3D video in Section 5.3, including results and discussion. Two methods of scalable MDC are proposed in this section, one is based on the MPEG4-MAC-MDC and the other one is based on the SVC. Performance of the scalable MDC
based on the SVC is analysed in error free environment, ideal MDC channel and error prone WiMAX channel. Sections 5.4 discusses the application of DSUS from Chapter 2 to the scalable video coding, followed by application of DSUS to the scalable MDC in Section 5.5. Performance comparison in error free environment of scalable video coding, scalable MDC and scalable MDC with DSUS is analysed in Section 5.6. The section also investigates the performance of scalable video coding and scalable MDC with DSUS in error prone WiMAX channel. The chapter is finally concluded in Section 5.7.

5.2 Related Work

5.2.1 Scalable Video Coding

SVC allows decoding of appropriate subsets of bit-stream to generate complete pictures of size and quality dependent on the proportion of the total bit-stream decoded. Universal media access can be made possible with SVC by coding the video only once to achieve a scalable coded stream. This stream can then be accessed ‘anytime’, from ‘anywhere’ using any access network such as wireless and internet, and by any terminal complexity.

The scalable properties come at the expense of decrease coding efficiency as it often introduces a degree of data redundancy. In general, SVC can be divided into three basic types namely quality scalability (Signal-to-noise-ratio (SNR) scalability), spatial scalability and temporal scalability. Combination of three scalable types can also be achieved to produce hybrid scalability.

SVC produces video layers consist of a base layer and a number of enhancement layers. The base layer is essential for the video decoder to produce a basic acceptable output. The enhancement layers help to improve the visual quality if received by the decoder. Without the base layer, enhancement layers only are useless to the decoder.

5.2.2 Review of Scalable Coding Techniques

A review of SVC can be found in [89]. Early video compression standards, namely MPEG-2, MPEG-4 and extension of H.263 support scalable coding. In this section, the basic methods of SNR, spatial and temporal SVC are briefly described. For SNR SVC, the base layer is produced by using the standard motion-compensated encoder. For frame \( n \), the base layer encoder will produce a reconstructed frame \( n' \). The residual error between frame \( n \) and frame \( n' \) can then be coded as the enhancement layer with the same or different quantisation parameter. The enhancement residual can be added to the base layer reconstructed frame \( n \) at the decoder to improve its visual quality.
In spatial SVC, the original video is down-sampled spatially to obtain a lower resolution video. The lower resolution video is coded as the base layer using standard motion-compensated DCT encoder and the original resolution video is coded as the enhancement layer. To improve coding efficiency of the enhancement layers, inter-layer prediction mechanisms have been introduced. One example of inter-layer prediction mechanism is the re-use of base layer motion vectors for the enhancement layer coding.

To produce temporal SVC, the base layer can be produced at reduced frame rate. The enhancement layer consists mainly of B-pictures that are bi-directionally predicted using the two surrounding pictures of a lower temporal layer as a reference. The temporal resolution of the base layer can be improved if the enhancement layer is received at the decoder.

Fine granularity scalability is also another type of SVC that allows truncating of the enhancement video bit stream to provide partial enhancement proportional to the number of bits decoded [90]. This can be achieved by coding the enhancement layer DCT coefficients using bit-plane coding.

5.2.3 Overview of Scalable H.264

SVC developed by JVT is the latest video coding standard which supports spatial, temporal and quality scalability for video [91]. Scalability of SVC is provided at a bit-stream level, which means bit-streams for a reduced spatial and/or temporal resolution can be obtained by discarding NAL units that are not required for decoding from a global SVC bit stream.

Temporal scalable bit stream can be provided by using hierarchical prediction structures. In this structure, key pictures are coded at regular intervals by using only previous key pictures as references. The pictures between the key pictures are the hierarchically B pictures which are bi-directionally predicted from the key pictures. The base layer contains a sequence of the key pictures at the coarsest supported temporal resolution, while the enhancement layers consists of the hierarchically B pictures. A low-delay coding structure is also possible by restricting the prediction of the enhancement layer pictures from only previous frame.

Figure 5-1 shows a block diagram of a spatially scalable encoder. SVC achieved spatial scalability by using over-sampled pyramid approached. Each spatial layer of a picture is independently coded using motion-compensated prediction. Inter-layer motion, residual or intra prediction mechanisms can be switched on to improve the coding efficiency of the enhancement layers. In inter-layer motion prediction for example, the up-scaled base layer motion data is employed for the spatial enhancement layer coding.

SVC also supports SNR scalability using coarse-grain scalability (CGS) and fine-grain scalability (FGS). CGS is achieved using spatial scalability concepts with exclusion of the up-sampling operations in the inter-layer prediction mechanisms. FGS is supported by introducing progressive
refinement (PR) slices that represent refinement of the residual signal that corresponds to bisection of the quantisation step size.

Figure 5-1: Block diagram of a scalable encoder using a multi-scale pyramid with 3 levels of spatial scalability [106]

5.2.4 Scalable Multiple Description Coding

Scalable MDC has been proposed to improve error resilience of the transmitted video over unreliable networks and at the same time provide adaptation to bandwidth variations and receiving device characteristics [92]. It can be categorized into two categories, first one starts with MDC coder and the MDC descriptions are then made scalable ([54] and [93]). Second one starts with SVC coder and the SVC layers are then mapped into MDC descriptions ([94], [95], [96], [97], [98] and [99]).

In [93] for example, a single MDC is split into base and enhancement layers. In [96], a non-standardised scalable wavelet encoder is used to provide MDC streams that can vary the number
of descriptions, rate of each individual description and redundancy level of each description using post encoding. Another example of non-standardised wavelet scalable MDC encoder is presented in [107]. It is based on motion-compensated scheme derived from the Haar multi-resolution analysis, which produce temporally correlated descriptions. Comparison of several redundant wavelet decompositions for temporal scalable MDC is provided in [108] with an aim to reduce the redundancy and provide better side reconstruction.

Other type of scalable MDC includes embedded MDC proposed in [100]. It uses embedded multiple description scalar quantisers and a combination of motion compensated temporal filtering and wavelet transform. In this approach the transmission of motion data is assumed to occur without errors. Scalable MDC based on SVC is proposed in [99]. It generates two descriptions for each enhancement layer of an SVC coded stream by embedding in each description only half of the motion and texture information of the original coded stream with minimum redundancy. However, noticeable blocky holes artifacts as shown in Figure 5-2 are observed in an ideal MDC channel due to motion vector approximation from neighbouring blocks.

![Figure 5-2: Artifacts from scalable MDC proposed in [99] under ideal MDC channel](image)

All the scalable MDC above has been applied to 2D video to provide error resilience but so far it has not been applied to 3D video. Non-scalable MDC schemes were proposed for 3D stereoscopic left and right view in [102] using spatial scaling and multi-state coding. In this chapter, the application of scalable MDC to stereoscopic 3D video (specifically 2D video plus depth information) is investigated.

### 5.3 Proposed Scalable Multiple Description Coding

This chapter involves the combination of two video coding aspects: Multiple Description Coding and scalable video coding. MDC schemes provide a number of independent descriptions of the same content. Their advantages have been described in a large number of published papers, including [56] and [57]. Scalable coding produces a number of hierarchical descriptions that provide flexibility in terms of adaptation to user and network requirements. Although a
hierarchical scalable approach is more compression efficient than MDC, the loss of a lower layer prevents all higher layers being decoded, even if they are correctly received, which means that a significant amount of correctly received data may be useless in certain channel conditions.

Scalable MDC aims to combine the flexibility of scalable coding with the robustness of MDC. In Section 5.3.1, scalable side information MDC based on the MPEG4-MAC-MDC is investigated. This method begins with an MDC codec and the side information is made scalable. Even and odd frame scalable MDC based on the SVC is proposed in Section 5.3.2. This proposed algorithm begins with an SVC codec and the base layer is coded into MDC descriptions.

5.3.1 Scalable Side Information

The side information in MPEG4-MAC-MDC is not very useful in error free environments, but can be used up to a certain extent in error prone environments. If the side information is coded in the enhancement layer using MPEG-4-Fine Granularity Scalability (MPEG4-FGS), it can be truncated by an FGS server in error free transmission. Furthermore, using FGS scalability, the video quality of the side reconstruction can be optimised over a given bit rate range.

Initially, MPEG-4-MAC-MDC developed in previous chapters is used to develop scalable MDC that is in category two as explained in Section 5.2.4. The side information is placed in the enhancement layer and coded using MPEG4-FGS style [90]. Figure 5-3 shows the block diagram of the proposed method. As an example, the difference between interpolated frame (of frame 0 and frame 2) and the reconstructed frame 1 is coded using DCT. The DCT coefficient is then coded using bit-plane coding similar to the enhancement layer of the MPEG4-FGS. As an initial implementation, only texture side information is encoded in the enhancement layer and not the depth information.

To evaluate the coding performance of the encoder in error free environment, rate distortion curve is plotted for Orbi. The tests are carried out using CIF format (352x288). The basic encoding parameters are: 300 frames, IPPP... sequence format, 30 frames/s original frame rate, variable length coding (VLC) and without error resilience. The Quantisation parameter (QP) in the configuration file is varied to obtain the bit rate range shown in the rate-distortion curves.

The rate distortion curves show the image quality measured in PSNR against the resulting bit rate when both of the MDC streams are received in error free, also known as central distortion. Figure 5-4 shows the rate-distortion curves for Orbi colour and depth sequences for SDC, MDC-EO, MDC-EOS, MDC-FGS. For MDC-EOS and MDC-FGS, the quantisation parameter for the side information is set to 20.
Chapter 5. Scalable Multiple Description Coding for 3D Video

It can be seen that MDC-FGS has poor performance compared to SDC, MDC-EO and MDC-EOS due to the extra information needed to code the FGS enhancement layer of the side information as in [90]. However, MDC-FGS has the advantage of truncating the enhancement layer to a desired bit rate by the FGS server in error free transmission. With this feature, the video quality of the side reconstruction can be enhanced if more bit rate is available and can be reduced to acceptable quality if bit rate less bit rate is available.

5.3.2 Scalable Multiple Description Coding Using Scalable Video Coding

Using MDC it is possible to mitigate the effects of losing lower layer data, and produce an optimal trade-off for a particular channel error rate, in terms of compression efficiency and error robustness. This is usually achieved by varying the amount of redundancy introduced by the MDC scheme. Thus, in good channels, compression efficiency equivalent to hierarchical scalability can be achieved, while in poor channels, MDC may be used to mitigate the effects of errors.

In the previous section, MPEG-4-MAC-MDC is used to generate scalable side information. However, it is recently found in [103], that using SVC to code texture in base layer and depth in enhancement layer is more efficient than MPEG4-MAC configuration. Hence, SVC will be used in this section to provide scalable MDC. The configuration used to code the texture and depth information using SVC is shown in Figure 5-5.
Figure 5-4: Rate-Distortion curves for ‘Orbi’ sequence (a) Colour image sequence (b) Depth image sequence
The proposed scalable MDC in this section begins with a standard SVC coder and the error resilience of the base layer of SVC is enhanced using temporal MDC. The temporal MDC of the base layer is produced by using the multi-state video coding approach as in [60], which separates the even and odd frames of a sequence into two MDC streams as shown in Figure 5-7.

Figure 5-6 shows the general block diagram of the proposed scalable MDC for stereoscopic 3D video. The depth data, which will be combined with the texture data using DIBR technique [13] to produce left and right views, is placed in the enhancement layer.

SVC can produce scalable layers that can be exploited for MDC. It is mentioned in [91] that SVC can provide a multicast scenario with interlayer prediction switched off. Hence, in the scalable MDC simulation, the even and odd frames are separated before the encoding process for both texture and depth. Then, the even frames for the texture are coded in the base layer (layer 0) and the odd frames for the texture are coded in the enhancement layer (layer 1). With the interlayer prediction switched off, it can be assumed that layer 1 is also the base layer for the scalable MDC. The even frames for the depth are coded in the enhancement layer (layer 2) and the odd frames for the depth are coded in the enhancement layer (layer 3). Table 5-1 shows the description of the layers that can be produced by the scalable MDC for a CIF size image sequence for only one spatial resolution. Example with spatial scalability is given in Section 5.3.2.1.4.

A single standard compliant bit stream is produced from the above configuration. At the decoder, a bit stream extractor is used to extract the even and odd bit streams of the texture and the depth. Each bit stream can then be decoded by the standard SVC decoder. Finally, the decoded even and odd frames are merged together to produce a full resolution decoded sequence.
Table 5-1: Description of layers for the scalable MDC encoder

<table>
<thead>
<tr>
<th>Layer</th>
<th>Resolution</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>352x288</td>
<td>Base layer, even - Texture</td>
</tr>
<tr>
<td>1</td>
<td>352x288</td>
<td>Base layer, odd - Texture</td>
</tr>
<tr>
<td>2</td>
<td>352x288</td>
<td>Enhancement layer, even - Depth</td>
</tr>
<tr>
<td>3</td>
<td>352x288</td>
<td>Enhancement layer, odd - Depth</td>
</tr>
</tbody>
</table>

For both texture and depth, if both the even and odd streams are received, the decoder can reconstruct the coded sequence at full temporal resolution. If only one stream is received, the decoder can still decode the received stream at half the original temporal resolution. Since the...
even frames are predicted from previous even frames (independent from odd frames), there will be no mismatch if one of the streams are lost at the decoder. Additionally, in the case of one stream is received, the decoder can decode at full resolution by interpolating between the received frames or by repeating the frame.

<table>
<thead>
<tr>
<th>Even frames: 0 2 4 6 ...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Errors propagate</td>
</tr>
<tr>
<td>Odd frames: 0 1 3 5 7 ...</td>
</tr>
<tr>
<td>Displayed: 0 1 2 3 5 7 9 11</td>
</tr>
</tbody>
</table>

Reduced frame rate

**Figure 5-8: Example of error in one of MDC stream**

If an error occurs, for example, in one of the frames in the even MDC stream the error will propagate to the next even frame as shown in Figure 5-8. The MDC decoder can then switch to the odd stream and display the frame at a reduced frame rate. Alternatively, the MDC decoder can interpolate between the reduced frame rates or repeat the frame to achieve full temporal resolution. In this section averaging frame interpolation is performed using:

\[ I_{ip}(i,j) = \frac{I_{prev}(i,j) + I_{fut}(i,j)}{2} \]  

(5-1)

where \( I_{ip}(i,j) \) is the frame to be interpolated at location \((i,j)\), \( I_{prev}(i,j) \) is the previous frame and \( I_{fut}(i,j) \) is the future frame.

### 5.3.2.1 Performance Evaluation of the Proposed Scalable MDC

The performance of the proposed scalable MDC is compared with the single description SVC (SDC) in an error free channel and in a wireless error prone WiMAX channel (Appendix C). Single description coding for SVC means that the base and enhancement layer of SVC are used to respectively encode the 2D and depth information of the sequence as shown in Table 5-2. The proposed scalable MDC uses a configuration as shown in Table 5-1 for the simulation.

**Table 5-2: Description of layers for the SDC**

<table>
<thead>
<tr>
<th>Layer</th>
<th>Resolution</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>352x288</td>
<td>Base layer, Texture</td>
</tr>
<tr>
<td>1</td>
<td>352x288</td>
<td>Enhancement layer, Depth</td>
</tr>
</tbody>
</table>
Performance of the proposed scalable MDC in an ideal MDC channel is also presented. The JSVM software [104] has been adapted to produce the even and odd frame MDC scheme of the base layer. 3D stereoscopic test sequences (2D video plus depth), namely Interview and Orbi at CIF resolution of 300 frames is used in the simulation. The base and enhancement layer of SVC are used to respectively encode the 2D and depth information of the sequence.

Throughout the thesis, different wireless and wired simulations have been used, namely UMTS, Wireless LAN, Internet and WiMAX. Initially the research starts with UMTS, but it was found that UMTS could not support high bit rates application with good quality 3D video. Hence, wireless LAN is then used as it can support higher bit rates and is available within the ILAB-CCSR research group. Recently, WiMAX is developed within the SUIT project [84]. As it is the latest wireless simulation and it can also support much higher bit rates than wireless LAN, it is decided to use WiMAX simulation for error prone performance in this chapter. Occasionally, Internet packet loss is also used for the error prone environments.

5.3.2.1.1 Error Free Performance

The rate distortion of MDC versus SDC for the Interview and Orbi sequence at 30 frame/s of the base layer (2D video) is shown in Figure 5-9 and Figure 5-10 respectively. Average PSNR_lumin is used to measure the quality of the 2D video. The SDC is the original JSVM software used to code the Interview sequence. There is an I-frame coded every two seconds to for both MDC and SDC. For SDC the I-frame is inserted every 60 frames and for MDC the I-frame is inserted every 30 frames in each MDC stream. The same fixed QP values are used for the SDC and MDC to achieve the rate distortion curve.

It can be seen from Figure 5-9 and Figure 5-10 that MDC is less efficient than SDC in terms of compression capability. MDC PSNR is about 1-2 dB lower than SDC for the same bit rate. This is due to the prediction from previous two frames that result in larger residual, and the larger motion vectors that need to be coded. MDC also has more I-frames than SDC because there are two I-frames (one for even streams and one for odd streams) inserted every 30 frames in the MDC streams, compared to one every 60 frames in the SDC stream.
5.3.2.1.2 Ideal Multiple Description Coding Channel

In an ideal MDC channel, it is assumed only one of the MDC description is received and the other MDC description is lost or heavily corrupted. In this situation, the proposed scalable MDC decoder interpolates between the received frames to achieve full temporal resolution. Table 5-3 shows the average Y-PSNR obtained from decoding only even frames and both descriptions of the proposed scalable MDC.
It is observed that the proposed scalable MDC has high average PSNR, of more than 35 dB for both sequences even if only one description is decoded. The PSNR difference between both and one description is 2.24 dB and 0.46 dB for Orbi and Interview respectively. No disturbing artifacts such as in [99] are observed when playing the decoded sequences as the interpolation produce a perceptually insignificant averaging effect between the frames.

The subjective result for error prone is shown in Figure 5-11 for 20% packet loss of Internet error pattern [105] applied on the even stream only. The Interview sequence at CIF resolution and 30 frame/s is coded using the original JSVM software (SDC) and the modified JSVM software (MDC) to achieve the same bitrate of about 300 kbit/s by varying the quantisation parameter. Table 5-4 shows the quantisation parameter and the I-frame rate used to achieve the bitrate. In order to achieve the same bitrate as MDC at the same quantisation parameter, SDC used more I-frames than MDC.

<table>
<thead>
<tr>
<th>Coder</th>
<th>QP</th>
<th>I-frame rate</th>
<th>Bitrate (kbit/s)</th>
<th>Average PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDC</td>
<td>30</td>
<td>Every 15 frames</td>
<td>290.89</td>
<td>35.64</td>
</tr>
<tr>
<td>MDC</td>
<td>30</td>
<td>Every 30 frames in each stream</td>
<td>292.85</td>
<td>35.63</td>
</tr>
</tbody>
</table>
Chapter 5. Scalable Multiple Description Coding for 3D Video

Figure 5-11: Subjective results for frame 80 of interview sequence when subjected to 20% packet loss: (a) SDC; (b) MDC-repeat; (c) MDC-interpolate

The 20% packet loss is applied only in the even stream for MDC to simulate ideal MDC channel (only one stream is lost at a time), while in SDC, all frames are subjected to the 20% packet loss. It is assumed that one lost packet will mean that one frame is lost. Figure 5-11 shows the subjective results for frame 80 and a comparison between the interpolated and the repeat frame methods for concealment the reduced frame rate. In SDC, frame copy error concealment is used. In the MDC simulation, if an error occurs in even frame, the decoder switch to odd frame and ignores the rest of the even frame. It then interpolates or repeats the odd frames to achieve full resolution. When playing the sequence, a little bit of jerky motion is observed for MDC-repeat due to the frame repetition. It can be concluded that under ideal MDC channel, MDC is very effective in combating the transmission errors compared to SDC.
5.3.2.1.3 Error Prone Performance

The compressed 3D video is transmitted over a simulated WiMAX (IEEE802.16e) channel developed in SUIT (Scalable, Ultra-fast and interoperable Interactive Television) project (Appendix C) [84].

The network parameter settings used in this paper for the IEEE802.16e simulator are: 16QAM ¾ modulation coding scheme (144 bits per time slot), 16 time slots in the down-link (DL) and IP packet size of 256 bytes. The mobile speed is up to 60km/h and Partially Used Sub Channelisation (PUSC) permutation is used. Convolutional Turbo Coding (CTC) is employed for the channel coding and the ITU Vehicular-A environment is chosen.

Figure 5-12 and Figure 5-13 shows the performance of transmitting the Orbi and Interview sequence respectively using scalable MDC compared to SDC in the WiMAX channel. The result shows the average PSNR for the luminance component. The depth data is assumed to be transmitted error free using the enhancement layer, as the research focus in this section is to provide error resilience of the base layer which is used to code the texture information.

Table 5-5 shows the quantisation parameter, the resulting error free PSNR for luminance (Y-PSNR) and depth (D-PSNR) and the corresponding texture and depth bit rate in kbit/s. The overall bit rate for scalable MDC and SDC is 450 kbit/s and 460 kbit/s respectively for 300 frames of Orbi sequence at CIF resolution. For Interview sequence, the overall bit rate for scalable MDC and SDC is 403 kbit/s and 405 kbit/s respectively. An I-frame is inserted for every 45 frames in SDC and for every 45 frames in each MDC stream, resulting in 6 I-frames in total for each codec.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Codec</th>
<th>QP</th>
<th>Y-PSNR (dB)</th>
<th>D-PSNR (dB)</th>
<th>Bit rate Texture (kbit/s)</th>
<th>Bit rate Depth (kbit/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orbi</td>
<td>SDC</td>
<td>28.4</td>
<td>38.97</td>
<td>41.43</td>
<td>318</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>MDC</td>
<td>30</td>
<td>37.70</td>
<td>40.45</td>
<td>313</td>
<td>137</td>
</tr>
<tr>
<td>Interview</td>
<td>SDC</td>
<td>28</td>
<td>37.06</td>
<td>42.05</td>
<td>270</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>MDC</td>
<td>30</td>
<td>35.63</td>
<td>40.64</td>
<td>278</td>
<td>125</td>
</tr>
</tbody>
</table>
Chapter 5. Scalable Multiple Description Coding for 3D Video

Figure 5-12: Performance of scalable MDC compared to SDC in WiMAX for Orbi

Figure 5-13: Performance of scalable MDC compared to SDC in WiMAX for Interview

It can be seen that scalable MDC has an improved performance of about 1 dB at high error rates (SNR 10 to 16 dB). One of the reasons is that MDC has the advantage of using the other stream when error occurs in one stream. Provided the other stream is not in error, MDC can achieve reasonable results as shown in Figure 5-12. Sometimes, the errors corrupt both streams, hence, an I-frame is needed to stop error propagation. At low error rates (SNR 18-24 dB), SDC performed.
better by about 1 dB, due to its superior performance in error free conditions as shown in Table 5-5.

Similar performance is observed for Interview sequence in Figure 5-13. Subjective results of the 2D video and the 3D stereoscopic video of Orbi sequence at a SNR of 10 dB in Figure 5-14 also show the improvement achieved by the proposed scalable MDC compared to SDC. The DIBR technique in [13] is used to achieve stereoscopic 3D video. In Figure 5-14 (c) and (d), the 3D video is rendered as red and blue images and can be viewed using the red and blue glass.

![Figure 5-14: Subjective results for frame 23 of the Orbi sequence for 2D video: (a) SDC; (b) MDC; and for stereoscopic 3D video: (c) SDC; (d) MDC when subjected to 10 dB SNR WiMAX channel](image)

5.3.2.1.4 Scalable Performance

Table 5-6 shows the spatial scalable result of the proposed scalable MDC for Interview sequence. All the layers are contained in one single bit stream. Layer 0 and Layer 1 are the MDC layers (base layer). The user can select to decode the required bit stream according to their terminal requirement in the virtual collaboration system. For example, two video resolutions, QCIF and CIF can be decoded from the bit stream. If a stereoscopic 3D terminal is available, the user can also decode the depth layer and used a DIBR technique to achieve stereoscopic 3D video.
Chapter 5. Scalable Multiple Description Coding for 3D Video

Table 5-6: Description of layers

<table>
<thead>
<tr>
<th>Layer</th>
<th>Resolution</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>176x144</td>
<td>Base layer, even – Texture</td>
</tr>
<tr>
<td>1</td>
<td>176x144</td>
<td>Base layer, odd – Texture</td>
</tr>
<tr>
<td>2</td>
<td>176x144</td>
<td>Enhancement layer, even – Depth</td>
</tr>
<tr>
<td>3</td>
<td>176x144</td>
<td>Enhancement layer, odd – Depth</td>
</tr>
<tr>
<td>4</td>
<td>352x288</td>
<td>Enhancement layer, even – Texture</td>
</tr>
<tr>
<td>5</td>
<td>352x288</td>
<td>Enhancement layer, odd – Texture</td>
</tr>
<tr>
<td>6</td>
<td>352x288</td>
<td>Enhancement layer, even – Depth</td>
</tr>
<tr>
<td>7</td>
<td>352x288</td>
<td>Enhancement layer, odd – Depth</td>
</tr>
</tbody>
</table>

5.4 Performance Comparison of the Proposed Scalable MDC versus MDC-EOASP

In this section, the performance of the SDC and the MDC developed in this chapter under SVC framework is compared with the best MDC developed using MPEG4-MAC, namely MDC-EOASP. Performance is compared in error free environment and in error prone (WiMAX) channel. The results show the average PSNR for only the luminance component. The depth data is assumed to be transmitted error free. Performance of the algorithms with error prone transmission of the depth is described in Section 5.7.2.

5.4.1 Error Free Environment

Figure 5-15 and Figure 5-16 show the coding efficiency of the SDC, MDC and MDC-EOASP for Orbi and Interview sequence respectively. The average PSNR is obtained for the luminance component. The quantisation parameters used for MDC-EOASP is 5, 10, 15, 20, 25 and 30. The highest QP that can be used in MPEG4 encoder is 30. The quantisation parameters used for SDC and MDC are 20, 30, 35, 40, 45 and 50. It can be seen from both figures that SDC out performs MDC and MDC-EOASP at all bit rates. This is expected as MDC-EOASP is based on MPEG4-MAC, and SDC and MDC is based on SVC. In error free channel, MDC coding efficiency is less than SDC due to the redundant information as shown in Section 5.3.2.1.1.
5.4.2 Error Prone Environment

The performance of the SDC, MDC and MDC-EOASP algorithms are evaluated over an error prone channel (simulated WiMAX channel) as in Section 5.3.2.1.3. The network parameter settings used in this section for the IEEE802.16e simulator are: 16QAM ¾ modulation coding.
scheme (144 bits per time slot) and IP packet size of 256 bytes. The time slots are changed to obtain the physical layer bit rates that match the bit rates obtained from varying the quantisation parameters. The mobile speed is up to 60km/h and Partially Used Sub Channelisation (PUSC) permutation is used. Convolutional Turbo Coding (CTC) is employed for the channel coding and the ITU Vehicular-A environment is chosen. Same quality experiment and same bit rate experiment are performed for the error prone condition.

5.4.2.1 Same Quality Experiment

In this experiment, the three algorithms have almost the same error free average PSNR as shown in Table 5-7 and Table 5-8. QP for MDC-EOASP is chosen so that its average PSNR is the closest to SDC and MDC in error free environment. Examples of the results are shown in Figure 5-17 for Orbi, and Figure 5-18 for Interview sequence.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>QP</th>
<th>Bit rates (kbit/s)</th>
<th>Error free average PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDC</td>
<td>30</td>
<td>329</td>
<td>37.75</td>
</tr>
<tr>
<td>MDC</td>
<td>30</td>
<td>457</td>
<td>37.58</td>
</tr>
<tr>
<td>MDC-EOASP</td>
<td>5</td>
<td>1737</td>
<td>39.04</td>
</tr>
</tbody>
</table>

Figure 5-17: Error prone performance of the algorithms using Table 5-7 configuration for Orbi
Table 5-8: QPs used, the resulting bit rates and the error free average PSNR (dB) for Interview

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>QP</th>
<th>Bit rates (kbit/s)</th>
<th>Error free average PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDC</td>
<td>40</td>
<td>72</td>
<td>28.96</td>
</tr>
<tr>
<td>MDC</td>
<td>40</td>
<td>110</td>
<td>28.98</td>
</tr>
<tr>
<td>MDC-EOASP</td>
<td>20</td>
<td>374</td>
<td>29.42</td>
</tr>
</tbody>
</table>

Figure 5-18: Error prone performance of the algorithms using Table 5-8 configuration for Interview

It can be seen that performance of the SDC and MDC developed using SVC framework is better than MDC-EOASP at high error rate even though the bit rates of MDC-EOASP are higher than SDC and MDC. This is expected as MDC-EOASP is developed using MPEG4-MAC framework which has less coding efficiency than H.264. In the low error rates region (21-24 dB SNR), the MDC-EOASP is slightly better than SDC and MDC because it follows the error free average PSNR as shown in Table 5-7 and Table 5-8.

It can also be seen that MDC outperforms SDC at high error rate. At 38% redundancy, MDC is about 6 dB better than SDC at about 18 dB SNR for Orbi sequence for the quantisation parameters used in Table 5-7. One dB gain is achieved by MDC at about 50% redundancy for Interview sequence at about 15 dB SNR for the quantisation parameters used in Table 5-8.

5.4.2.2 Same Bit Rate Experiment

In this experiment, the three algorithms have almost the same bit rates as shown in Table 5-9 and Table 5-10. QP for MDC-EOASP is chosen so that its bit rate is the closest to SDC and MDC in error free environment. In the simulation, intra frame is assumed not corrupted. For MDC, it is assumed that both of the streams are not corrupted at the same time. This means that some frames
are not corrupted. The same frames are also not corrupted in SDC and MDC-EOASP. Examples of the results are shown in Figure 5-19 for Orbi, and in Figure 5-20 for Interview sequence.

Table 5-9: QPs used, the resulting bit rates and the error free average PSNR (dB) for Orbi

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>QP</th>
<th>Bit rates (kbit/s)</th>
<th>Error free average PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDC</td>
<td>28.4</td>
<td>474</td>
<td>39.04</td>
</tr>
<tr>
<td>MDC</td>
<td>30</td>
<td>457</td>
<td>37.58</td>
</tr>
<tr>
<td>MDC-EOASP</td>
<td>30</td>
<td>456</td>
<td>30.22</td>
</tr>
</tbody>
</table>

![Graph](image)

Figure 5-19: Error prone performance of the algorithms using Table 5-9 configuration for Orbi

Table 5-10: QPs used, the resulting bit rates and the error free average PSNR (dB) for Interview

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>QP</th>
<th>Bit rates (kbit/s)</th>
<th>Error free average PSNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDC</td>
<td>28.4</td>
<td>432</td>
<td>37.00</td>
</tr>
<tr>
<td>MDC</td>
<td>30</td>
<td>423</td>
<td>35.63</td>
</tr>
<tr>
<td>MDC-EOASP</td>
<td>14</td>
<td>411</td>
<td>31.22</td>
</tr>
</tbody>
</table>
Chapter 5. Scalable Multiple Description Coding for 3D Video

Figure 5-20: Error prone performance of the algorithms using Table 5-10 configuration for Interview

It can be seen that performance of the SDC and MDC developed using SVC framework is better than MDC-EOASP at for all error rates. As in the same quality experiment, this is expected as MDC-EOASP is developed using MPEG4-MAC frame work which has less coding efficiency than H.264. It should be noted that all the intra frames in MDC-EOASP, SDC and MDC are assumed not corrupted. However, MDC-EOASP has only one intra frame at the start of the sequence, while SDC and MDC have one intra frame for every 30 frames. Therefore, the performance of MDC-EOASP is poorer than SDC and MDC.

Figure 5-20 also shows that the average PSNR for SDC and MDC becomes flat at low channel SNRs due to the availability of uncorrupted intra frames, and some other frames that are uncorrupted to avoid both MDC streams corrupted at the same time.

It can also be seen that MDC out performs SDC at high error rate. At error rate of less than 17 dB SNR, MDC is about 1 dB better than SDC for Orbi sequence. Similarly, about one dB gain is achieved by MDC for Interview sequence at error rates of 13 dB SNR and below. The even and odd MDC algorithm does not achieved a large gain over SDC at high error rates due to the reduced coding efficiency in error free environment. Hence, in the next section, the reduced resolution depth compression technique developed in Chapter 2 is reintroduced to improve MDC performance.
5.5 Scalable Video Coding with Down-Sampling and Up-Sampling

The DSUS technique in Chapter 2 can be used to reduce the bit rate of coding high definition depth image sequence. In this section, the DSUS algorithm is applied to the enhancement layer of SVC which is used to code the depth information. The block diagram of the proposed method for the encoder is shown in Figure 5-21. At the decoder, the coded depth information from the enhancement layer is up-sampled back to its original resolution.

The depth information is down-sampled from 720x576 resolution to CIF (352x288) resolution. The SVC spatial scalability does not allow the enhancement layer to be used to send image sequence at lower resolution than the base layer. Hence, three layers configuration as shown in Table 5-11 is used in the simulation. For both SVC without DSUS (SVC-Org) and SVC with DSUS (SVC-DSUS), the base layer is used to send the texture information at CIF resolution. Note that for SVC-DSUS, the enhancement layer 1 is used the send the depth at reduced resolution.

The enhancement layer 2 for both SVC-Org and SVC-DSUS is used to send the texture information at full resolution. The inter-layer prediction between layer 2 (720x576, texture) and layer 0 (352x288, texture) is switched on using adaptive inter-layer prediction to exploit the redundancies between the two layers. The rate distortion for the depth information of the Orbi sequence at 720x576 resolution at 30 frames/s is shown in Figure 5-22. The bit rate is varied using the quantisation parameter. Same quantisation parameter is used for texture and depth. The horizontal axis shows the bit rate for the depth and the vertical axis shows the average PNSR for the depth.
Table 5-11: Configuration used for SVC-Org and SVC-DSUS

<table>
<thead>
<tr>
<th>Encoder</th>
<th>Layer</th>
<th>Size</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVC-Org</td>
<td>2 (enhancement)</td>
<td>720x576</td>
<td>Texture</td>
</tr>
<tr>
<td></td>
<td>1 (enhancement)</td>
<td>720x576</td>
<td>Depth</td>
</tr>
<tr>
<td></td>
<td>0 (base)</td>
<td>352x288</td>
<td>Texture</td>
</tr>
<tr>
<td>SVC-DSUS</td>
<td>2 (enhancement)</td>
<td>720x576</td>
<td>Texture</td>
</tr>
<tr>
<td></td>
<td>1 (enhancement)</td>
<td>352x288</td>
<td>Depth</td>
</tr>
<tr>
<td></td>
<td>0 (base)</td>
<td>352x288</td>
<td>Texture</td>
</tr>
</tbody>
</table>

Figure 5-22: Rate distortion for the depth information of Orbi image sequence

It can be seen from Figure 5-22 that the performance of the SVC-DSUS is much better than SVC-Org at the same bit rate in low bit range (less than 120 kbit/s). For example, at about 50 kbit/s, the average PSNR for SVC-Org is about 30 dB and the average PSNR for SVC-DSUS is about 35 dB.

Another way to measure the performance of the algorithm is by using the rate distortion, measured using left and right PSNR. An original left-and-right image sequence is produced from the original 2D image sequence and its associated original depth information image sequence using DIBR technique described in Chapter 2. Similarly, a compressed left-and-right image sequence is obtained from the compressed 2D and depth image sequence. The left-and-right
average PSNR is then obtained by comparing the original left-and-right image sequence with the compressed left-and-right image sequence.

One problem with the DIBR algorithm is to handle disocclusion problem. This problem happens when the occluded area in the original view might become visible in any of the synthesised left-and-right image. In [19], pre-processing of the depth information is applied to improve the visual quality during the disocclusion. In our software implementation of DIBR, the disocclusion problem is not considered. Hence, our synthesized left-and-right image sequence will have some artifacts such as shown on the right face of the person in Figure 5-23. However, these artifacts will also appear in the compressed left-and-right image sequence. These artifacts can also be reduced when viewed using the Stereoscopic Movie Player [30] by adjusting the distance parameter between left and right image.

![Disocclusion artefacts from DIBR implementation on Interview sequence (Left-image)](image)

The left-and-right average PSNR rate distortions for Orbi and Interview image sequences are plotted in Figure 5-24 and Figure 5-25 respectively.

![Rate distortion for the Orbi image sequence (average left-and-right PSNR)](image)
Figure 5-25: Rate distortion for the Interview image sequence (average left-and-right PSNR)

Figure 5-24 shows that the DSUS algorithm improves the rate distortion of the Orbi sequence for SVC-Org at bit rates less than 1300 kbit/s. This is because SVC-DSUS has less overall bit rate than SVC-Org due to the use of DSUS on the depth information. The up-sampling distortion only reduces the performance of SVC-DSUS at bit rates more than 1300 kbit/s for Orbi sequence. Similar performance is obtained for Interview sequence as in Figure 5-25. The SVC-DSUS outperforms SVC-Org at bit rates less than 1200 kbit/s.

The DSUS technique can also be applied to the texture information and both on the texture and the depth information. Figure 5-26 and Figure 5-27 shows the left-and-right average PSNR rate distortions for Orbi and Interview image sequences when the DSUS technique is applied to depth only (SVC-DSUS-d), texture only (SVC-DSUS-c), both texture and depth (SVC-DSUS-cd), and without DSUS (SVC-Org). Only two layers configuration is used, layer 0 for the texture and layer 1 for the depth. The texture and depth of Orbi and Interview sequence at 720x576 with 30 frames/s is used for original resolution. The sequences are down-sampled to CIF (352x288) resolution when DSUS is applied.

It can be seen that SVC-DSUS-d performance is better than SVC-DSUS-c for all the bit rates. This is due to the more up-sampling distortion in the texture compared to the depth. Application of DSUS to both texture and depth greatly reduces the bit rate, but as shown in the results, SVC-DSUS-d manages to outperform SVC-DSUS-cd except at very low bit rate. SVC-DSUS-cd also has up-sampling distortion in both texture and depth. Based on these results, the use of DSUS on the depth information only is considered in the next section. These results also justify the use of
Figure 5-26: Rate distortion comparison of DSUS application to SVC for Orbi sequence

Figure 5-27: Rate distortion comparison of DSUS application to SVC for Interview sequence

DSUS on the depth information in the early part of this section, Section 2.8 and Section 2.9.4 of the thesis.
5.6 Scalable Multiple Description Coding with Down-Sampling and Up-Sampling

In this section the DSUS technique in Chapter 2 is applied to the scalable MDC developed in this chapter to improve the scalable MDC performance in coding the high definition depth image sequence at low bit rate range. Figure 5-28 shows the application of down-sample to the depth information which will be coded using the enhancement layer of SVC-MDC encoder. At the decoder, the decoded depth information will be up-sampled back to its original resolution and the performance is compared with SVC-MDC without DSUS (SVC-MDC-Org).

A six layer configuration, as shown in Table 5-12, is used in the simulation for SVC-MDC-Org and SVC-MDC-DSUS. Note that for SVC-MDC-DSUS, the enhancement layer 2 and 3 is used to send the even frames and odd frames respectively for the depth information at reduced resolution. The inter-layer prediction between layer 5 (720x576-texture-odd) and layer 1 (352x288-texture-odd), and between layer 4 (720x576-texture-even) and layer 0 (352x288-texture-even) is switched on using adaptive inter-layer prediction to exploit the redundancies between the layers.

![Figure 5-28: Block diagram of SVC-MDC encoder with DSUS (SVC-MDC-DSUS)](image)

The rate distortion for the depth information of Orbi and Interview sequence is plotted in Figure 5-29 and Figure 5-30 respectively. IPPP... sequence format is used and fixed quantisation parameter is employed to obtain the bit rates. The horizontal axis shows the bit rate for the depth and the vertical axis shows the average PNSR for the depth. The sequences are coded according to the layer parameters shown in Table 5-12. The left-and-right average PSNR rate distortions for Orbi and Interview image sequences are plotted in Figure 5-31 and Figure 5-32 respectively.
Chapter 5. Scalable Multiple Description Coding for 3D Video

Table 5-12: Layers configuration for SVC-MDC-Org and SVC-MDC-DSUS

<table>
<thead>
<tr>
<th>Encoder</th>
<th>Layer</th>
<th>Size</th>
<th>Type</th>
<th>Stream</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 (enhancement)</td>
<td>720x576</td>
<td>Texture</td>
<td>Odd</td>
</tr>
<tr>
<td>SVC-MDC-Org</td>
<td>4 (enhancement)</td>
<td>720x576</td>
<td>Texture</td>
<td>Even</td>
</tr>
<tr>
<td></td>
<td>3 (enhancement)</td>
<td>720x576</td>
<td>Depth</td>
<td>Odd</td>
</tr>
<tr>
<td></td>
<td>2 (enhancement)</td>
<td>720x576</td>
<td>Depth</td>
<td>Even</td>
</tr>
<tr>
<td></td>
<td>1 (base)</td>
<td>352x288</td>
<td>Texture</td>
<td>Odd</td>
</tr>
<tr>
<td></td>
<td>0 (base)</td>
<td>352x288</td>
<td>Texture</td>
<td>Even</td>
</tr>
<tr>
<td>SVC-MDC-DSUS</td>
<td>5 (enhancement)</td>
<td>720x576</td>
<td>Texture</td>
<td>Odd</td>
</tr>
<tr>
<td></td>
<td>4 (enhancement)</td>
<td>720x576</td>
<td>Texture</td>
<td>Even</td>
</tr>
<tr>
<td></td>
<td>3 (enhancement)</td>
<td>352x288</td>
<td>Depth</td>
<td>Odd</td>
</tr>
<tr>
<td></td>
<td>2 (enhancement)</td>
<td>352x288</td>
<td>Depth</td>
<td>Even</td>
</tr>
<tr>
<td></td>
<td>1 (base)</td>
<td>352x288</td>
<td>Texture</td>
<td>Odd</td>
</tr>
<tr>
<td></td>
<td>0 (base)</td>
<td>352x288</td>
<td>Texture</td>
<td>Even</td>
</tr>
</tbody>
</table>

Figure 5-29: Rate distortion for the depth information of Orbi image sequence
Chapter 5. Scalable Multiple Description Coding for 3D Video

Figure 5-30: Rate distortion for the depth information of Interview image sequence

It can be seen from Figure 5-29 and Figure 5-30 that performance of SVC-MDC-DSUS is better than SVC-MDC-Org at low bit rates range. For Orbi sequence, below 240 kbit/s SVC-MDC-DSUS outperforms SVC-MDC-Org by 1-5 dB. For Interview sequence, below 70 kbit/s, SVC-MDC-DSUS outperforms SVC-MDC-Org by 1-2 dB.

Figure 5-31: Rate distortion for the Orbi image sequence (average left-and-right PSNR)
5.7 Performance Comparison with Down-Sampling and Up-Sampling

The DSUS technique can be applied to the scalable MDC to improve coding efficiency. In this section the rate distortion performance of SVC-Org, SVC-MDC-Org and SVC-MDC-DSUS is compared in error free environments. It is expected that the rate distortion of the luminance for SVC-MDC-DSUS will be better than SVC-MDC-Org at the expense of the depth information. This section also presents the performance of SVC-Org, SVC-DSUS, SVC-MDC-Org and SVC-MDC-DSUS in an error free condition and error prone internet environment.

5.7.1 Error Free Environment

The Orbi sequence at 720x576 is used in the error free comparison with IPPP… sequence format and a fixed quantisation parameter to obtain the bit rates. An I-frame is inserted every 45 frames. The average PSNR for luminance and depth is respectively plotted over the overall bit rate in
Chapter 5. Scalable Multiple Description Coding for 3D Video

Figure 5-33: Rate distortion for the luminance information of the Orbi image sequence (average PSNR versus overall bit rates)

Figure 5-34: Rate distortion for the depth information of the Orbi image sequence (average PSNR versus overall bit rates)
Figure 5-35: Rate distortion for the depth information of the Orbi image sequence (average PSNR versus depth bit rates)

Figure 5-33 and Figure 5-34. The average PSNR for the depth information against the depth bit rate is shown in Figure 5-35. It can be seen from Figure 5-33 that the coding efficiency of SVC-MDC-DSUS has improved over SVC-MDC-Org and become quite close to coding efficiency of SVC-Org. This is because SVC-MDC-DSUS has less overall bit rate than SVC-MDC-Org due to the used of down-sampling for depth information in SVC-MDC-DSUS.

The improved coding efficiency of the luminance comes at the expense of reduced coding efficiency of the depth for SVC-MDC-DSUS at high overall bit rates (above 400 kbit/s) as shown in Figure 5-34. At overall bit rates below 400 kbit/s, performance of SVC-MDC-DSUS for Orbi sequence is comparable to SVC-MDC-Org because, as the overall bit rates of SVC-MDC-DSUS are reduced, the depth average PSNR for SVC-MDC-DSUS is also reduced due to the up-sampling distortion.

To show the improved rate distortion for the depth information at low bit rates, the average PSNR versus depth bit rate is plotted in Figure 5-35. At depth bit rates below 150 kbit/s, SVC-MDC-DSUS performs better than SVC-Org and SVC-MDC-Org by about 1 dB to 3 dB.

The left-and-right PSNR for SVC-Org, SVC-DSUS, SVC-MDC-Org and SVC-MDC-DSUS is plotted over the overall bit rates in Figure 5-36 and Figure 5-37 for Orbi and Interview respectively. Configuration in Table 5-11 is used for SVC-Org and SVC-DSUS, and configuration in Table 5-12 is used for SVC-MDC-Org and SVC-MDC-DSUS.
Figure 5-36: Rate distortion for the Orbi image sequence (average left-and-right PSNR)

It can be seen from Figure 5-36 that the coding efficiency of SVC-DSUS has improved over SVC-Org for Orbi image sequence for bit rates up to 1300 kbit/s. SVC-MDC-DSUS performs better than SVC-MDC-Org until 1800 kbit/s, and become quite close to coding efficiency of SVC-Org at bit rates less than 400 kbit/s. This is because SVC-MDC-DSUS has less overall bit rate than SVC-MDC-Org due to the used of down-sampling for the depth information.

Figure 5-37: Rate distortion for the Interview image sequence (average left-and-right PSNR)
Similar performance is obtained for Interview sequence in Figure 5-37. The coding efficiency of SVC-DSUS has improved over SVC-Org for Orbi image sequence for bit rates up to 1200 kbit/s. SVC-MDC-DSUS performs better than SVC-MDC-Org until 1700 kbit/s, and become quite close to coding efficiency of SVC-Org at bit rates less than 800 kbit/s due to the used of down-sampling the depth information, which reduces the overall bit rates.

### 5.7.2 Error Prone Environment

In this section, performance on SVC-Org, SVC-DSUS, SVC-MDC-Org and SVC-MDC-DSUS are evaluated under error prone environment. The compressed 3D video is transmitted over a simulated internet channel [105]. This time, internet error patterns are used instead of the WiMAX error pattern. It is found in Section 5.3.2.1.3 that the gain achieved by the MDC in WiMAX error pattern is quite small at high error rates. Furthermore, the error rates for the WiMAX are very high at SNR below 18.5 dB and can be up to 99% packet loss.

The internet channel in [105] has four internet packet loss error patterns, namely 3%, 5%, 10% and 20%. In the simulation, the loss of one packet is assumed to be loss of one video frame. Frame copy error concealment is used for SVC-Org and SVC-DSUS. SVC-MDC-Org and SVC-MDC-DSUS use frame interpolation in case of error. Orbi and Interview image sequences are used in the simulation. An I-frame is inserted for every 45 frames in SDC and for every 45 frames in each MDC stream.

Figure 5-38, Figure 5-39, Figure 5-40 and Figure 5-41 show the rate distortions of Orbi sequence for packet loss 3%, 5%, 10% and 20% respectively. Figure 5-42, Figure 5-43, Figure 5-44 and Figure 5-45 show the rate distortions of Interview sequence for packet loss 3%, 5%, 10% and 20% respectively. In these figures, the decoded left-and-right PSNR for SVC-Org, SVC-DSUS, SVC-MDC-Org and SVC-MDC-DSUS is plotted over the overall bit rates.

From Figure 5-38 to Figure 5-45, it can be seen that most of the time, SVC with MDC has better performance compared to standard SVC, particularly for packet losses of 5% and above. As an example, for the Orbi sequence, in high bit rate range, SVC-MDC-DSUS has better performance than the other schemes at all packet loss rates. This is due to the properties of the MDC scheme, which is able to effectively mitigate the effects of errors. SVC-MDC-DSUS also has better error free coding efficiency compared to SVC-MDC-Org in the low bit rate range. The worst performance is given by SVC-Org at high and low bit rates for the Orbi sequence. Sometimes, SVC-DSUS performs better than others with 3% packet loss in the low bit rate range, but performs comparatively to SVC-MDC-DSUS.
Chapter 5. Scalable Multiple Description Coding for 3D Video

For the Interview sequence, SVC-MDC-DSUS performs better than others. SVC-DSUS performs worst at high bit rate range probably due to the up-sampling distortion in the depth image and its inability to recover quickly from errors. In the low bit rate range, SVC-MDC-DSUS performs better and comparable to SVC-DSUS most of the time. Similar to the Orbi sequence results, the worst performance is given by SVC-Org most of the time in the low bit rate ranges.
Figure 5-40: Rate distortion for Orbi sequence at 10% packet loss

Figure 5-41: Rate distortion for Orbi sequence at 20% packet loss

Figure 5-46 shows example of the stereoscopic 3D video of Interview sequence at qp=35 for 5% packet loss. It shows the improvement achieved by the SVC-MDC-Org and SVC-MDC-DSUS over SVC-Org and SVC-DSUS. The DIBR technique in [13] is used to achieve stereoscopic 3D video. In Figure 5-46, the 3D video is rendered as red and blue images and can be viewed using the red and blue glass. Table 5-13 shows the bit rates and error free and error prone average left-right PSNR in dB for the sequence in this figure.
It can be seen that, with about 7% redundancy, SVC-MDC-DSUS performs better than SVC-Org by 3 dB in the 5% packet loss. With 43% redundancy SVC-MDC-Org is 2 dB better than SVC-Org. The lowest bit rate is achieved by SVC-DSUS, but its error prone performance is not better than SVC-MDC-Org and SVC-MDC-DSUS.
The packet loss performance for Orbi and Interview sequence at \( q_p = 35 \) is shown in Figure 5-47 and Figure 5-48. For Orbi sequence, it can be seen that the MDC algorithms perform better especially at high packet loss rate, with SVC-MDC-DSUS performs the best. For example at 5% packet loss, SVC-MDC-DSUS is about 3 dB better than SVC-Org.
Table 5-13: Algorithms performance (average left-right PSNR) at qp=35 for 5% packet loss

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Error free</th>
<th>Error prone</th>
<th>Bit rates (kbit/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVC-Org</td>
<td>33.75</td>
<td>28.25</td>
<td>610</td>
</tr>
<tr>
<td>SVC-DSUS</td>
<td>33.53</td>
<td>29.10</td>
<td>437</td>
</tr>
<tr>
<td>SVC-MDC-Org</td>
<td>33.78</td>
<td>30.31</td>
<td>872</td>
</tr>
<tr>
<td>SVC-MDC-DSUS</td>
<td>33.59</td>
<td>31.26</td>
<td>653</td>
</tr>
</tbody>
</table>

Similar performance is obtained for Interview sequence. As an example, at 5% packet loss, SVC-MDC-DSUS is nearly 4 dB better than SVC-Org. SVC-DSUS performs worst than SVC-Org at 20% packet loss because from the rate-distortion curve of the 20% packet loss (see Figure 5-45), SVC-Org performs better than SVC-DSUS at this quantisation parameter.

### 5.8 Conclusion

In this chapter, scalable MDC methods are proposed for scalability and error resilience purposes. For MPEG-4-MAC-MDC, the side information is placed in the enhancement layer and coded using MPEG4-FGS. The algorithm is named MDC-FGS. The simulation results show that MDC-FGS has poor error free performance compared to SDC, MDC-EO and MDC-EOS due to the extra information needed to code the FGS enhancement layer of the side information. However, MDC-FGS can truncate the enhancement layer to a desired bit rate by the FGS server in error free transmission. With this feature, the video quality of the side reconstruction can be enhanced if more bit rate is available and can be reduced to acceptable quality if bit rate less bit rate is available.

A scalable MDC scheme base on even and odd frames is also proposed in this chapter for the SVC targeting 3D video applications. The proposed algorithm generates two descriptions for the base layer of SVC based on even and odd frame separation, thus reducing its coding efficiency by about 1-2 dB compared to SDC. However, in an ideal MDC channel, the proposed scheme achieved good quality even if only one of the descriptions is received. Objective and 2D/3D subjective evaluation in an error prone WiMAX channel shows an improved performance of about 1 dB by the proposed algorithm at high error rates when compared to SDC. The proposed scheme can provide error resilience through MDC and scalability through SVC for 3D video applications.

The performance of the scalable MDC is also compared with MDC-EOASP which is developed using MPEG4 in error free and error prone condition. Simulation results show that performance of the SDC and MDC developed using SVC framework is much better than MDC-EOASP in error free environment. This is also true at high error rate situation, even though the bit rates of MDC-
EOASP are higher than SDC and MDC. These results are expected as MDC-EOASP is developed using MPEG4-MAC framework which has less coding efficiency than H.264.
The DSUS technique in Chapter 2 can be used to reduce the bit rate of coding high definition depth image sequence. In this chapter, the DSUS algorithm is applied to the enhancement layer of SVC which is used to code the depth information. At the decoder, the coded depth information from the enhancement layer is up-sampled back to its original resolution. Performance of the SVC-DSUS is much better than SVC-Org at the same bit rate in low bit range (less than 120 kbit/s).
In this chapter also the DSUS technique is applied to the scalable MDC to improve the scalable MDC performance in coding the high definition depth image sequence at low bit rate range. Performance is compared with SVC-MDC without DSUS (SVC-MDC-Org). Performance of SVC-MDC-DSUS is better than SVC-MDC-Org at low bit rates range. For Orbi sequence, below than 240 kbit/s SVC-MDC-DSUS out performs SVC-MDC-Org by 1-5 dB. For Interview sequence, below 70 kbit/s, SVC-MDC-DSUS out performs SVC-MDC-Org by 1-2 dB.

This chapter also compares the rate distortion performance of SVC-Org, SVC-MDC-Org and SVC-MDC-DSUS in error free environment. It is found that the rate distortion of the luminance for SVC-MDC-DSUS is better than SVC-MDC-Org at the expense of the depth information and become quite close to coding efficiency of SVC-Org. This is because SVC-MDC-DSUS has less overall bit rate than SVC-MDC-Org due to the used of down-sampling for the depth information in SVC-MDC-DSUS. The improved coding efficiency of the luminance comes at the expense of reduced coding efficiency of the depth for SVC-MDC-DSUS at high overall bit rates (above 400 kbit/s).

The overall performance of SVC-Org, SVC-DSUS, SVC-MDC-Org and SVC-MDC-DSUS is also compared in this chapter. In general, the coding efficiency of SVC-DSUS is the best for low bit rates up to around 1000 kbit/s depending on the sequence. SVC-MDC-DSUS performs better than SVC-MDC-Org for low bit rates up to around 1500 kbit/s, and become quite close to coding efficiency of SVC-Org at bit rates less than around 900 kbit/s due to the used of down-sampling the depth information, which reduces the overall bit rates.

Finally, this chapter presents the performance of SVC-Org, SVC-DSUS, SVC-MDC-Org and SVC-MDC-DSUS for stereoscopic 3D video in error prone Internet environment. Simulation results show that most of the time, SVC-MDC algorithms (original or DSUS) perform better in error prone than SVC-DSUS and SVC-Org especially at high error rates. At the same error free quality, it is shown that the redundancies introduced by the MDC algorithms manage to do better than the original SVC and down-sampled SVC in error prone network.
Chapter 6

Conclusion and Future Work

6.1 Overview

This thesis involves 3D visual data compression for transmission over error prone networks. Issues such as resilience and scalability have been taken into account to mitigate the effects of transmission errors and information loss on the compressed 3D video stream.

The works in this thesis can be concluded in three main areas as follows:

- 3D video compression
- 3D video error resilience
- 3D video scalability

Each will be explained in the following three sections.

6.2 Summary of the Work on 3D Video Compression

One of the objectives of the research is to compress the stereoscopic 3D video data for transmission purposes. A format of stereoscopic video, namely 2D video plus depth has been chosen to represent the stereoscopic video. Using the depth based image rendering technique, a left and right image sequence from the 2D image and its associated depth information is produced which can be rendered to obtain a stereoscopic 3D video. Compression of this type of data can be accomplished by using an available video coding standard such as MPEG-4. With regard to the depth compression, down-sampling prior to encoding and up-sampling of the depth data after decoding is proposed. The use of down-sampling prior to encoding and up-sampling after decoding introduces up-sampling distortions beside the quantisation errors. However, the simulation results on high resolution image sequence demonstrated that if the resolution of the depth image sequence is reduced prior to encoding and up-sampled back to its original resolution after decoding, far better objective and subjective quality could be achieved compared to compressing using the original resolution of the depth image at low bit rates.
Chapter 6. Conclusion and Future Work

The 2D video and depth data can be compressed using MPEG4 version 2. Objective and subjective performance of MPEG4 to code the 3D video (2D video plus depth) have been explored. Two architectures of MPEG4, namely MPEG4-MAC and MPEG4-Separate are used for performance comparison. MPEG4-MAC uses one of the multiple auxiliary components to code the depth information. In MPEG4-Separate configuration, the 2D video is encoded separately from the depth data. Using an MPEG4 video codec, the depth data is supplied as the luminance channel. Another MPEG4 video codec is used to encode the ordinary 2D video. Simulation results with 2D video plus depth image sequences show that both MPEG4-MAC and MPEG4-Separate perform at par in coding the 2D video at the same target bit rate. MPEG4-Separate is slightly better in coding the depth information. More importantly, the resulting 3D perception of both architectures does not seem to show much difference in terms of the subjective quality achieved. Hence, MPEG4-MAC is an attractive solution to compress the 3D video. The one bit stream from MPEG4-MAC can be fed into a Multiple Description Coding (MDC) encoder to increase its error resilience.

Finally, the application of DSUS to the depth information in MPEG4-MAC is also investigated. The results show that if the resolution of the depth image sequence is reduced prior to MAC encoding and up-sampled back to its original resolution after decoding, far better luminance performance could be achieved compared to compressing using the original resolution of the depth image. This is due to the overall reduced bit rate obtained by down-sampling the depth information. However, performance of the depth information degrades as the bit rate increases beyond 1 Mbit/s because of the up-sampling distortion.

6.3 Summary of the Work on 3D Video Error Resilience

Transmitting stereoscopic 3D compressed video over error prone channels is a challenging task. Hence, error resilience tools such as developed within MPEG-4 standard are needed to mitigate the effect of the error. The error resilience tools are extended to the depth data in stereoscopic 3D video. It was demonstrated that the performance of the 2D video, additionally the depth data improved under error prone condition with data partitioning and RVLC.

Furthermore, multiple description coding (MDC) is introduced as a form of error resilience tools for 3D video. An even and odd (EO) frame based MDC (MDC-EO) method was developed and applied to 3D video. Simulation results in error free environment show a decrease of 1-2 dB average PSNR over SDC. Although a decrease in coding efficiency is a drawback, MDC-EO has performance advantages in error prone environments. In frame loss situation, it was found that SDC is more vulnerable to errors compared to MDC-EO. SDC’s PSNR quality drops is larger
Chapter 6. Conclusion and Future Work

than MDC-EO. As long as both of the MDC-EO streams are not corrupted at the same time, the uncorrupted stream can be used to speed up recovery from error.

Simulation results in an error prone wireless channel, specifically Wireless LAN, show that MDC-EO is objectively better than SDC for both texture and depth data at very high error rates, especially for the depth data. This is due to the depth data that has a lot of low frequency component which allow more effective error concealment using frame interpolation that frame copy in error prone condition. Hence, the average PSNR of MDC-EO is largely better than SDC for depth information, but slightly better than SDC for texture information in high error rates. In terms of the 3D stereoscopic video, it was found that the subjective quality achieved by MDC-EO is shown to be better than SDC at high error rates.

Novel MDC-EOS and MDC-EOSP are proposed as improvements to MDC-EO. It is found that their coding efficiency is reduced compared to MDC-EO and SDC in error free and error prone environments due to the large amount of side information. Nevertheless, in frame loss situations and in ideal MDC channels, MDC-EOS and MDC-EOSP feature lower PSNR variation compared to MDC-EO and SDC. MDC-EOS also performs better than SDC in a controlled error prone wireless LAN environment and in a UMTS channel.

MDC-EOSB is also developed with the objective of reducing the amount of side information by using B-frame interpolation method. Comparable rate distortion performance with MDC-EOS is achieved at high bit rate, but at low bit rates, MDC-EOSB performance deteriorates because of inefficient coding of the B-frame motion vector in the side information. On the other hand, MDC-EOSB performance is better than SDC at high bit rates in an error prone wireless environment.

In a comparison simulation in UMTS environment, MDC-EOSP is also found to be better than SDC, MDC-EO, MDC-EOSB and SDC. MDC-EOSP is at least comparable with MDC-EOS. The simulation result shows that MDC with or without side information is very useful in the error prone environment.

The side information in MDC-EOS and MDC-EOSP contributes to the high redundancy of these algorithms. Hence, it is also proposed to send the side information adaptively according to the motion in the sequence. Large motion will make the algorithm send the side information and no side information is sent if the motion is low. Novel MDC-EOAS and MDC-EOASP are developed to send the adaptive side information. The coding efficiency of these two algorithms is better than MDC-EOS and MDC-EOSP and very close to MDC-EO. Their frame PSNR variations during error are less than MDC-EO. In frame lost situation, the performance is better than SDC, and better than MDC-EO if the side information is received by the decoder. The error prone performance of MDC-EOAS and MDC-EOASP is better than SDC and MDC-EO at high packet
loss objectively and subjectively. Similar to MDC-EO, the gain achieved by MDC-EOAS and MDC-EOASP for depth is larger than the gain achieved for luminance.

In conclusion, MDC with side information is promising approach to combat channel errors for stereoscopic 3D video transmission, but the side information should be carefully sent as it can cause huge redundancies. It can be sent adaptively according to motion, and ideally, according to the network condition. The novelty in this section is about sending and using the side information for stereoscopic 3D video. Trade off between central distortion (rate distortion performance in error free) and side distortion (rate distortion performance in ideal MDC channel) also need to be taken into account in developing an effective MDC algorithm.

### 6.4 Summary of the Work on 3D Video Scalability

In this section, scalable MDC methods are proposed for scalable and error resilience purposes. For MPEG-4-MAC-MDC, the side information is placed in the enhancement layer and coded using MPEG4-FGS. The algorithm is named MDC-FGS. The simulation results show that MDC-FGS has poor error free performance compared to SDC, MDC-EO and MDC-EOS due to the extra information needed to code the FGS enhancement layer of the side information. However, MDC-FGS has the advantage of truncating the enhancement layer to a desired bit rate by the FGS server in error free transmission. With this feature, the video quality of the side reconstruction can be enhanced if more bit rate is available and can be reduced to acceptable quality if bit rate less bit rate is available.

A scalable MDC scheme base on even and odd frames is also proposed for SVC targeting 3D video applications. The proposed algorithm generates two descriptions for the base layer of SVC based on even and odd frame separation, thus reducing its coding efficiency by about 1-2 dB compared to SDC. However, in an ideal MDC channel, the proposed scheme achieved good quality even if only one of the descriptions is received. Objective and 2D/3D subjective evaluation in an error prone WiMAX channel shows an improved performance of about 1 dB by the proposed algorithm at high error rates when compared to SDC. The proposed scheme can provide error resilience through MDC and scalability through SVC for 3D video applications.

The MDC-EOASP, which is developed using MPEG4, is also compared with the scalable MDC developed using SVC standard. Simulation is performed in error free and error prone condition. Simulation results show that performance of the SDC and MDC developed using SVC framework is much better than MDC-EOASP in error free environment. This is also true at high error rate situation, even though the bit rates of MDC-EOASP are higher than SDC and MDC. These results are expected as MDC-EOASP is developed using MPEG4-MAC framework which has less coding efficiency than H.264.
The DSUS technique can be used to reduce the bit rate of coding high resolution depth image sequences. In this thesis, the DSUS algorithm is applied to the enhancement layer of SVC which is used to code the depth information. At the decoder, the coded depth information from the enhancement layer is then up-sampled back to its original resolution. Performance of the SVC-DSUS is much better than SVC-Org at the same bit rate in low bit rate range.

In this thesis also, the DSUS technique is applied to the scalable MDC to improve the scalable MDC performance in coding the high resolution depth image sequence at low bit rates. Performance is compared with SVC-MDC without DSUS (SVC-MDC-Org). Performance of SVC-MDC-DSUS is better than SVC-MDC-Org at low bit rates. For Orbi sequence, below than 240 kbit/s, SVC-MDC-DSUS out performs SVC-MDC-Org by 1 dB to 5 dB. For Interview sequence, below 70 kbit/s, SVC-MDC-DSUS out performs SVC-MDC-Org by 1 dB to 2 dB.

This thesis also compares the rate distortion performance of SVC-Org, SVC-DSUS, SVC-MDC-Org and SVC-MDC-DSUS in an error free environment. Simulation results show that SVC-DSUS is the best at low bit rates because of the reduced depth resolution. It is also found that the rate distortion of the luminance for SVC-MDC-DSUS is better than SVC-MDC-Org at the expense of the depth information and becomes quite close to coding efficiency of SVC-Org. This is because SVC-MDC-DSUS has less overall bit rate than SVC-MDC-Org due to the use of down-sampling for the depth information in SVC-MDC-DSUS. The improved coding efficiency of the luminance comes at the expense of reduced coding efficiency of the depth for SVC-MDC-DSUS at high bit rates.

Finally, this thesis presents the performance of SVC-Org, SVC-DSUS, SVC-MDC-Org and SVC-MDC-DSUS in an error prone Internet environment. Simulation results show that, overall, SVC-MDC (Org or DSUS) performance in error prone is better than SVC-Org and SVC-DSUS especially at high error rates. In conclusion, it has been demonstrated that the redundancies introduced by MDC algorithms are very useful for error prone transmission of stereoscopic 3D video.

6.5 Original Achievements

The work presented in this thesis involved extensive studies on the video coding standard, multiple description video coding and scalable multiple description video coding for stereoscopic 3D video application. The main novelty claims are summarised as follows categorised into three main areas:
Chapter 6. Conclusion and Future Work

- 3D video compression
  - Compression of depth information using DSUS method – A new compressing tool for depth information based on pre-processing and post-processing is proposed. The proposed scheme can be applied without any modification to the existing video coding standard.
  - Compression of 2D and depth using MPEG4-MAC – It is proposed to exploit the existing tools in the video coding standard to perform compression on stereoscopic 3D video based on 2D video and depth information. The MAC components in MPEG4-version 2, which are used to transmit disparity in the literature, are exploited to transmit the depth information. The performance is comparable to separately encoding the 2D video and depth information.
  - Application of DSUS to the depth information in MPEG4-MAC – The DSUS tool, proposed before, is applied to the depth information in the MAC to improve stereoscopic 3D video coding efficiency. The performance is better than the original MPEG4-MAC at limited quality of the depth information at high bit rates.

- 3D video error resilience
  - Extension of RVLC and data partitioning to depth information – An extension of RVLC and data partitioning to the depth information is proposed. This is necessary as the MAC in the original MPEG4 version 2, which is used to transmit the depth information, is not protected with error resilience methods. The performance of the depth information in error prone environment is found to be better with the applied RVLC and data partitioning.
  - Application of even and odd frame based MDC to 3D video – It is proposed to apply MDC to the MPEG4 version 2 reference software. The applied MDC is also extended to the depth information for stereoscopic 3D video. So far, the MDC method has been applied to 2D video in the literature, but not on 2D video and depth stereoscopic 3D video. The performance is better than SDC at high error rates in error prone environment.
  - Development of even and odd frame based MDC with side information – Novel MDC schemes with side information are developed and are applied to stereoscopic 3D video. Different methods of obtaining the side information are proposed to improve coding performance in error free and error prone
environment. Compared to SDC, the proposed MDC methods perform better especially at high error rates in error prone channel.

- Development of even and odd frame based MDC with adaptive side information
  - The adaptive side information method further improved the novel MDC schemes developed before for stereoscopic 3D. The side information is adaptively sent according to the video sequence motion. This method improves the coding performance of the previous MDC in error free and error prone environment.

- 3D video scalability
  - Development of even and odd frame based MDC with scalable side information using FGS for stereoscopic 3D – In this proposed method, the side information is placed in the enhancement layer and coded using MPEG4-FGS. The novelty in this method is that the video quality of the side reconstruction can be optimised over a given bit range. Hence, the side reconstruction video quality will be better when more bit rates are available in the channel.

  - Development of scalable MDC based on even and odd frames MDC for SVC – In this method, it is proposed to extend the even and odd MDC method to SVC for error resilience and scalability of the stereoscopic 3D video. Compared to the original SVC, the proposed scalable MDC has better stereoscopic 3D video error prone performance especially at high loss rates. At the same time, the scalability features in the original SVC is still maintained.

  - Application of DSUS to the depth information in SVC – In this proposed method, the application of DSUS results in an improvement in the stereoscopic 3D video coding performance of the original SVC, particularly in the low bit rate range. The performance of the proposed method in error prone environment is better than the original SVC, particularly in the low bit rate range.

  - Application of DSUS to the depth information in the scalable MDC – This novel method helps to improve the stereoscopic 3D video coding performance when coded using scalable MDC. In error prone channel, the proposed method performs better than original SVC, particularly in the presence of high channel error rates and for low bit-rate video.
6.6 Areas of Future Research

This section describes some of the issues, which remain to be tackled in the provision of stereoscopic 3D video wireless communication.

6.6.1 3D Video Coding

In Chapter 2, a basic depth image base rendering technique is implemented to convert the 2D plus depth image sequence to stereoscopic video. Further research can be carried out to improve the implemented depth image base rendering technique taking into consideration the occlusion problems.

In this thesis, only a single view 2D plus depth is considered. The Moving Picture Expert Groups (MPEG) is developing standard for coding of multi-view, multiple synchronised video scenes that show the same scene from different viewpoints. They are currently enabling the 3D depth perception of the observed scenes. Research in this direction can focus on exploiting the nature of the depth image for compression and also the redundancies between the 2D image and the depth image. In this thesis, this redundancy is explored through the re-use luminance motion vector for the depth in MPEG4-MAC and reduced spatial depth resolution before coding. Other possibilities might be the down-sampled of the depth in DCT domain and the use of model based video coding for the depth information. Further studies are also required on the effects of the down-sampling of the depth information on the synthesized stereoscopic 3D video.

Another possible future work is to investigate the use of distributed video coding (DVC) to 2D plus depth video for stereoscopic 3D application. The distributed video coding for the 2D plus depth can shift the encoder computational complexity burden to the decoder. Distributed source coding has already been applied to the multi-view video coding, as it enables to exploit the correlation among distributed sources at the decoder, by performing disjoint encoding but joint decoding. Further extension of DVC to 2D plus depth multi-view might be a fair research direction.

6.6.2 3D Video Error Resilience

In this thesis, error resilience and concealment for stereoscopic 3D video are explored using multiple description video coding. For future work, the depth information might be useful for the error concealment of the texture information. The depth information should provide an indication of object boundaries, and the motion of the objects. This can be used during the concealment process in conjunction with the depth map segmentation algorithms.
Chapter 6. Conclusion and Future Work

The MDC with the side information algorithm developed in Chapter 4 for stereoscopic 3D video can be further optimised by using a better estimation of the motion and ideally according to the network conditions. Rate distortion optimisation of the algorithm is necessary for optimum performance. A distortion model that accurately models the overall distortion, including quantisation distortion and channel distortion can be used to optimise the MDC algorithm performance. A joint-source channel coding scheme for MDC for stereoscopic 3D video is also an interesting future works to be carried out.

In Chapter 4, the side information in MDC is exploited to improve central prediction and to be used for error concealment. Further research is needed so that the side information can be effectively used in error free and error prone conditions.

Furthermore, application of MDC for stereoscopic 3D video over a challenging ad hoc network is also an interesting future research direction. Ad hoc networks pose a great challenge to video transport as there is no fixed infrastructure and the topology is frequently changing due to node mobility. Therefore, links are continuously established and broken. MDC may be useful here if there is no feedback channel available.

6.6.3 3D Video Scalability

In Chapter 5, scalable MDC based on MPEG-FGS is developed to provide scalable side information. The FGS scalable side information is only applied to luminance information. Future works can extend the scalability to the depth information as the depth is required for stereoscopic 3D video application. An interesting future work would also be to compare the performance of MDC-FGS and MDC-SNR (Signal to noise ratio), where in MDC-SNR, the side information is coded using SNR scalability. In [90] it was found that MPEG4-FGS performed better than MPEG4-SNR scalability because the bit-plane coding of the DCT coefficient in MPEG4-FGS has better coding efficiency than run-level coding in MPEG4-SNR.

The MDC algorithm developed for the scalable H.264 in Chapter 5 only uses even and odd frames technique without the side information to produce two descriptions. As a future work, some side information should also be included in each description to compensate for losses of a description. Side information can be put in user defined SEI NALU’s, which will be ignored by standard decoders, but can be exploited by “enhanced” decoders to improve the quality of the received video. The addition of the side information introduces redundancy that can be varied according the channel conditions to optimise the received quality. In theory, it should be possible to produce an algorithm that can jointly optimise the overall amount of redundancy and allocate the redundancy amongst the different schemes to minimise the distortion.
However, when dealing with scalability, using a simple computational distortion measure is not necessarily the best option, as the choice of whether to provide better SNR, frame rate, or resolution, can be subjective, and can depend on both user preferences and content type. Therefore, to determine the way forward it will be necessary to perform some subjective tests. These subjective tests can be used to suggest appropriate methods of providing the combined MDC optimisation (e.g. incorporate some weighting factors derived from a set of user preferences and content characteristics).
Bibliography


[104] JSVM 8.7 reference software from CVS server, garcon.ient.rwth-aachen.de/cvs/jvt


Appendix A – UMTS Simulation Model

The UMTS error patterns used in this thesis are obtained from the simulated UMTS channel described in [53]. The UMTS channel model simulates the UMTS air interface. Figure A-6-1 shows the simplified block diagram of the simulated physical layer at the transmitter side. Table A-6-1 briefly described the parameters used in the UMTS simulation channel to obtain the error patterns.

![Figure A-6-1: UMTS Physical layer model](image)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRC attachment (before channel coding)</td>
<td>24, 16, 12, 8, 0</td>
</tr>
<tr>
<td>Channel coding</td>
<td>No coding, 1/2 rate convolutional code, 1/3 rate convolutional code, 1/3 rate turbo code</td>
</tr>
<tr>
<td>Interleaving</td>
<td>1st interleaving: Block interleaver with inter-frame permutation. 2nd interleaving: Block interleaver with inter-columns permutation.</td>
</tr>
<tr>
<td>Rate matching</td>
<td>Bit puncturing or bit repetition.</td>
</tr>
<tr>
<td>Transport channel (TrCH) multiplexing</td>
<td>Experiments were conducted for two TrCH</td>
</tr>
<tr>
<td>Transport format detection</td>
<td>Transport Format Combination Indicator (TFCI) based detection</td>
</tr>
<tr>
<td>Spreading factor</td>
<td>512, 256, 128, 64, 32, 16, 8, 4</td>
</tr>
<tr>
<td>Transmission Time Interval (TTI)</td>
<td>10 ms, 20 ms, 40 ms, 80 ms</td>
</tr>
<tr>
<td>Interference/Noise characteristics</td>
<td>Variance of AWGN source</td>
</tr>
<tr>
<td>Fading characteristics</td>
<td>Rayleigh fading mobile channel</td>
</tr>
<tr>
<td>Multipath characteristics</td>
<td>Vehicular, Pedestrian</td>
</tr>
<tr>
<td>Mobile terminal velocity</td>
<td>User definable. Constant for the simulation run</td>
</tr>
<tr>
<td>Chip rate</td>
<td>3.84 Mcps</td>
</tr>
<tr>
<td>Carrier frequency</td>
<td>2000 Mhz</td>
</tr>
<tr>
<td>Antenna characteristics</td>
<td>0 dB gain for both transmitter and receiver</td>
</tr>
<tr>
<td>Receiver characteristics</td>
<td>Rake receiver</td>
</tr>
<tr>
<td>Transmission diversity characteristic</td>
<td>Closed loop fast power control</td>
</tr>
</tbody>
</table>
Appendix B – WLAN Simulation Model

The WLAN error patterns used in this thesis are obtained from the simulated WLAN channel described in [84]. Figure B6-1 shows the implemented WLAN channel at the transmitter side.

![Baseband system model of IEEE 802.11g](image)

The system parameters for the WLAN IEEE802.11g are presented in Table B6-2. The combination of channel coding and modulation schemes (MCS) produces several transmission modes with different data rate as shown in Table B6-3. Several channel models are adopted with different environments and delay spreads. Rayleigh fading mobile channel is used and the environment characteristics include small office, medium office, large office and outdoor with line of sight (LOS) or without LOS.

<table>
<thead>
<tr>
<th>System Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carrier Modulation</td>
<td>OFDM</td>
</tr>
<tr>
<td>FFT Size</td>
<td>64</td>
</tr>
<tr>
<td>Carrier Frequency</td>
<td>2.4 GHz</td>
</tr>
<tr>
<td>Sampling Rate</td>
<td>20 MHz</td>
</tr>
<tr>
<td>Channel Coding</td>
<td>Punctured Convolutional Coding</td>
</tr>
<tr>
<td>Modulation</td>
<td>BPSK, QPSK, 16QAM, 64QAM</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MCS</th>
<th>Modulation</th>
<th>Bit Rate (Mbit/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>BPSK</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>BPSK</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>QPSK</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>QPSK</td>
<td>18</td>
</tr>
<tr>
<td>4</td>
<td>16-QAM</td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>16-QAM</td>
<td>36</td>
</tr>
<tr>
<td>6</td>
<td>64-QAM</td>
<td>48</td>
</tr>
<tr>
<td>7</td>
<td>64-QAM</td>
<td>54</td>
</tr>
</tbody>
</table>
Appendix C – WiMAX Simulation Model

The WiMAX error patterns used in this thesis are obtained from the simulated WiMAX channel [84]. Figure C6-3 shows the implemented WiMAX channel at the transmitter and receiver side. Table C6-4 briefly described some of the systems parameters used in the WiMAX simulator.

Channel coding procedures include randomization, FEC encoding and bit interleaving. The data shall follow the coding chain up to the QAM mapping. The data outputted from the QAM mapper shall be loaded onto the block of pre-allocated sub-channels for transmission. The sub-channel allocation follows one of the sub-carrier permutation schemes, e.g. Full Usage of Sub-Channels (FUSC) or Partial Usage of Sub-Channels (PUSC). After that, multiple antenna signal processing (MIMO) is applied if available in the system. Finally the data is passed to the OFDM transceiver (IFFT block) for transmission. The available FFT points are 2048, 1024, 512 and 256.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duplexing</td>
<td>Time Division Duplex (TDD) frame length of 5ms</td>
</tr>
<tr>
<td>Subcarrier permutation</td>
<td>PUSC</td>
</tr>
<tr>
<td>Channel coding</td>
<td>Convolutional Turbo Coding (CTC)</td>
</tr>
<tr>
<td>Carrier Frequency</td>
<td>2.3 GHz</td>
</tr>
<tr>
<td>Channel Bandwidth</td>
<td>8.75 MHz</td>
</tr>
<tr>
<td>Subcarrier Spacing</td>
<td>9.765625 KHz</td>
</tr>
<tr>
<td>OFDM Duration</td>
<td>115.2 µs</td>
</tr>
<tr>
<td>No. Symbols in a Frame</td>
<td>42</td>
</tr>
<tr>
<td>Length of Trace (s)</td>
<td>15</td>
</tr>
<tr>
<td>Terminal Speed (km/h)</td>
<td>60, 120</td>
</tr>
<tr>
<td>Test Environment</td>
<td>ITU Vehicular A</td>
</tr>
<tr>
<td>MCS Mode</td>
<td>QPSK-1/2, QPSK-3/4, 16QAM-1/2, 16QAM-3/4, 64QAM-1/2</td>
</tr>
<tr>
<td>SNR Range</td>
<td>0 to 30 dB which will have 5 to 7 data points for each MCS mode</td>
</tr>
</tbody>
</table>

Figure C-6-3: Physical layer of IEEE802.16e standard
Red and blue glass to view the 3D stereoscopic images in this thesis