



Universidade de Coimbra
Faculdade de Ciências e Tecnologia
Departamento de Engenharia Eletrotécnica e de
Computadores

**Multiple Description Coding
for Path Diversity Video Streaming**

Pedro Daniel Frazão Correia

2012

Multiple Description Coding for Path Diversity Video Streaming

Pedro Daniel Frazão Correia

*Submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy*

Universidade de Coimbra
Faculdade de Ciências e Tecnologia
Departamento de Engenharia Eletrotécnica e de Computadores

under supervision of
Prof. Dr. Vitor Manuel Mendes da Silva
Prof. Dr. Pedro A. Amado Assunção

Copyright© 2012 by PEDRO CORREIA. All rights reserved.

In memory to my mother, Maria Celeste

To Sara, Samuel, Simão and Paula

Acknowledgments

This page remember all of those who have involved, directly or indirectly in this work. First of all, I would like to thank to my scientific supervisors, Prof. Pedro Assunção and Prof. Vitor Silva, by their constant support, pragmatism and also by their fruitful discussions allowing me to pursue the research work with success.

My gratitude to Instituto de Telecomunicações, in special to Leiria branch, for providing me the physical, material and financial conditions necessary to develop the research activities.

I also want to thank to Instituto Politécnico de Tomar, institution that I belong as Assistant, for its support in order to combine the academic activities with the research work, and in this way to make possible this thesis.

My acknowledgment also to Fundação para a Ciência e a Tecnologia (FCT), that has supported this work with grants SFRH/BD/30087/2006 and SFRH/BD/50035/2009, last one under the program "Programa de Apoio à Formação Avançada de Docentes do Ensino Superior Politécnico" (PROTEC).

In a personal plan, I want to thank to my friends Ana Cristina, Ana Vieira and Gabriel Pires by their encouragement and motivation during this period. Also, I want to thank to Lino Ferreira, to his friendship and for his collaboration in part of the presented work. My gratitude to my friends of Missionary Community "Servidores do Evangelho", by their constant presence.

I sincerely want to thank to all my family, in special to my father Daniel, for all support and care that have given in all this time. Finally, my deep gratefulness to my daughter Sara, my sons Samuel and Simão, and my wife Paula, for their constant comprehension, support and love.

I would like to dedicate this work in memory to my mother, Maria Celeste.

Agradecimentos

Esta página lembra todos aqueles que direta e indiretamente colaboraram para que este trabalho pudesse chegar a este ponto. Em primeiro lugar gostaria de agradecer aos meus orientadores científicos, Prof. Pedro Assunção e Prof. Vitor Silva, pelo incentivo constante, pelo seu pragmatismo e pelas frutuosas discussões que permitiram que o trabalho desenvolvido chegasse a bom porto.

O meu agradecimento ao Instituto de Telecomunicações, em especial à Delegação de Leiria, pelas condições físicas, materiais e financeiras proporcionadas para o desenvolvimento do trabalho de investigação.

Agradeço também ao Instituto Politécnico de Tomar, instituição a que pertença como Assistente, pelas condições proporcionadas para conciliar a atividade docente com a atividade de investigação, conducente ao trabalho apresentado nesta dissertação.

Quero agradecer o apoio financeiro da Fundação para a Ciência e a Tecnologia (FCT) através das bolsas SFRH/BD/30087/2006 e SFRH/BD/50035/2009, esta última no âmbito do Programa de Apoio à Formação Avançada de Docentes do Ensino Superior Politécnico (PROTEC).

Num plano mais pessoal quero agradecer aos meus amigos, Ana Cristina, Ana Vieira e Gabriel Pires, pelo encorajamento e motivação ao longo de todo o período que decorreu este trabalho. Ao Lino Ferreira pela amizade e pela colaboração em parte do trabalho realizado. A minha gratidão a todos os amigos da Comunidade Missionária Servidores do Evangelho, pela sua presença constante.

Quero também agradecer a toda a minha família, e em especial ao meu pai Daniel, por todo o apoio e incentivo que dedicaram ao longo deste tempo. E por fim o meu agradecimento profundo, aos meus filhos, Sara, Samuel e Simão e à minha esposa Paula, por toda a sua compreensão, o seu apoio constante, pelo seu amor.

Dedico este trabalho em memória da minha mãe, Maria Celeste.

Abstract

In the current heterogeneous communication environments, the great variety of multimedia systems and applications combined with fast evolution of networking architectures and topologies, give rise to new research problems related to the various elements of the communication chain. This includes, the ever present problem in video communications, which results from the need for coping with transmission errors and losses. In this context, video streaming with path diversity appeared as a novel communication framework, involving different technological fields and posing several research challenges. The research work carried out in this thesis is a contribution to robust video coding and adaptation techniques in the field of Multiple Description Coding (MDC) for multipath video streaming.

The thesis starts with a thorough study of MDC and its theoretical basis followed by a description of the most important practical implementation aspects currently available in literature. Additionally, a review of Multiple Description (MD) video coding is presented, where the relevant methods recently developed regarding this issue are explained, covering different video coding approaches. In MDC, a video signal is typically encoded into several independent descriptions, i.e., compressed streams, where each one can be delivered over different channels making use of path diversity communication environments. A high quality video representation is achieved when all descriptions are available at the decoder, whereas a lower, but still acceptable quality, is obtained when only one description is received. These interesting features of MDC are investigated in this thesis and compared to classic Single Description Coding (SDC).

A research evaluation study of MDC for Advanced Video Coding (AVC) in regard to coding efficiency, distortion and error resilience is presented. Starting with open-loop MD video coding architectures, this study evaluates the effects of distortion propagation that happens when individual decoding of each description is performed. A novel multi-loop architecture for AVC is then proposed, which prevents drift distortion accumulation, by generating a controlled amount of additional information. In particular, the proposed MD video coding architecture is based on Multiple Description Scalar Quantization (MDSQ), including a new method for generating descriptions with different rates, *i.e.* unbalanced descriptions. This research extends the current state-of-the-art methods using balanced MDSQ, developing new MDC capabilities in different application scenarios without losing coding efficiency neither robustness to transmission losses.

This thesis also extends the current concept of multiple description coding (MDC) to the compressed domain, by proposing efficient splitting of standard single description coded (SDC) video into a multi-stream representation. A novel multiple description video splitting (MDVS) scheme was developed to operate at network edges, for increased robustness in path diversity video streaming across heterogeneous communications chains. The proposed scheme is able to effectively control drift distortion in both intra and inter predictive coding, even when only one description reaches the decoder. This is achieved by generating a controlled amount of relevant side information to compensate for drift accumulation, whenever any description is lost in its path, achieving significant quality improvement at reduced redundancy cost.

Additionally, and taking into account the new achievements, a novel high-accurate rate control method is proposed for MDSQ video coding. Known models based on linear relation between the output rate and the percentage of zeros of the quantized transform coefficients are extended to MDSQ video coding. In particular, this research demonstrates that the linear relation is maintained when MDSQ is used, making possible its use in linear rate control methods for MDC. Taking into account the balancing rate between descriptions, i.e. the percentage of the overall rate that is given to each description, this new method has the ability to choose the appropriate coding parameters to accurately achieve a predefined target MD bitrate.

Overall, the new methods investigated in this thesis along with the good performance obtained from the experimental results, demonstrate the relevancy of its contribution to the field of MDC and practical usefulness in new robust multimedia services and applications using path diversity channels.

Resumo

Nos ambientes heterogêneos das redes de comunicação atuais, a grande variedade de tecnologias de rede, assim como a rápida evolução das suas arquiteturas e topologias, tem dado origem ao aparecimento de novos métodos que permitam solucionar os problemas de erros de transmissão e perdas. A difusão de vídeo com diversidade de canais é vista como um esquema interessante de comunicação para fazer face a este cenário, envolvendo vários aspectos tecnológicos e colocando também vários desafios para investigação. Enquadrando-se neste contexto, esta dissertação apresenta novas técnicas de codificação e adaptação de vídeo com robustez a erros usando Codificação de Vídeo com Múltiplas Descrições (MDC)

Esta dissertação faz um estudo das técnicas MDC existentes, abordando o assunto nos seus aspectos teóricos, e descrevendo também os aspectos relacionados com a sua implementação. Adicionalmente, a dissertação apresenta uma revisão da literatura das aplicações MDC no contexto de codificação de vídeo. Esta técnica de codificação consiste na representação de uma fonte de vídeo por vários fluxos independentes *i.e.*, descrições, podendo ser decodificados com uma qualidade aceitável. No entanto, estes fluxos podem ser conjugados entre si, obtendo-se uma qualidade de reconstrução mais elevada. Esta característica interessante das técnicas de MDC é investigada nesta tese, comparando o seu desempenho com a codificação tradicional de descrição única.

Este trabalho, inclui um estudo experimental sobre o desempenho das técnicas MDC em Codificação Avançada de Vídeo(AVC), tendo em conta a eficiência de codificação, a distorção e a resiliência a erros. Partindo de arquiteturas MDC para vídeo em malha aberta e do estudo da influência da propagação de distorção nos sinais decodificados, esta tese propõe uma nova arquitetura multi malha para AVC. Esta permite prevenir a propagação da distorção quando apenas uma descrição é decodificada de forma independente, gerando informação adicional redundante que permite compensar a propagação de erros. Demonstra-se que a qualidade de reconstrução das sequências de vídeo aumenta significativamente quando comparada com arquiteturas de malha aberta. Em particular, a nova arquitetura permite gerar descrições múltiplas baseadas em Quantificação Escalar (MDSQ), cujas descrições possam ser codificadas com débitos distintos. Este novo método apresenta-se como uma evolução do estado da arte relativamente aos métodos atuais de MDSQ em codificação de vídeo, onde as várias descrições são codificadas com o mesmo débito.

O trabalho de investigação que conduziu a esta dissertação introduz também a aplicação de MDC para o domínio comprimido. É proposto um novo método que permite gerar diferentes fluxos de vídeo a partir de um só fluxo de vídeo comprimido normalizado. Este novo esquema permite operar em nós de rede que funcionem como pontos de adaptação, para aumentar a robustez a erros em redes e serviços com diversidade de canais em ambientes heterogéneos. A arquitetura proposta, possui a capacidade de controlar a propagação de erros na descodificação de uma descrição, para os vários modos de predição existentes na codificação. Para isso é gerada informação adicional a partir do fluxo de vídeo existente, que permite o controlo eficiente da propagação de distorção no caso de descodificação independente de uma descrição.

Considerando os possíveis cenários de aplicação dos novos métodos propostos, esta dissertação apresenta também um novo método de controlo de débito para codificação de vídeo usando MDSQ. O método proposto adapta os modelos conhecidos de controlo de débito baseados na relação linear entre o débito de saída e a percentagem de coeficientes quantificados de valor nulo resultantes da utilização de transformadas na codificação. O trabalho desenvolvido mostra que a relação de linearidade se mantém quando a operação MDSQ é colocada na malha de codificação, permitindo a implementação de métodos precisos de controlo de débito no âmbito de codificação robusta de vídeo com MDC. O método proposto tem em consideração o débito que cada descrição deve apresentar, escolhendo os parâmetros de codificação apropriados de modo a atingir de forma precisa o débito global de saída desejado.

Tendo em conta o desempenho dos métodos propostos, é possível afirmar que esta dissertação apresenta contributos relevantes na área de Codificação de Vídeo com Descrições Múltiplas, demonstrando a sua utilidade em novos serviços e aplicações com robustez a erros onde a diversidade de canais de transmissão possa ser utilizada.

Contents

1	Introduction	1
1.1	Motivation	3
1.2	Main Contributions	4
1.3	Outline of the thesis	6
2	Multiple Description Coding Techniques	9
2.1	Theoretical Basis	10
2.1.1	Single Description Rate-Distortion Region	11
2.1.2	Multiple Description Rate-Distortion Region	13
2.1.3	Multiple Description Bounds for a Gaussian Source	16
2.1.4	Representation of MD Rate-Distortion Region	18
2.1.5	Redundancy	20
2.2	Practical Implementation	22
2.2.1	Multiple Description with Correlating Transforms	23
2.2.2	Multiple Description Sub-band Coding	27
2.2.3	Polyphase Decomposition and Selective Quantization	28
2.2.4	Multiple Description Scalar Quantisation	29
2.2.5	MDC Based on Error Correction Codes	34
2.3	Conclusion	35
3	Multiple Description Video Coding and Networking	37
3.1	Basic Video Coding Schemes	38
3.2	MD Video Coding - an overview	42
3.3	MD Architectures for Video Encoding	45
3.3.1	Embedded MDC	46
3.3.2	MDC with pre-processing	51
3.3.3	High dimensional MDC	54
3.3.4	Multiple Description Scalable Coding	55
3.3.5	Stereo/Multiview MDC	57
3.3.6	Unbalanced MDC	59
3.3.7	Multiple Description Distributed Video Coding	61
3.4	Multiple Description Video Streaming	62

3.4.1	MDC Video Streaming over Content Delivery Networks	63
3.4.2	MDC Video Streaming over Peer-to-peer networks	65
3.4.3	MDC video streaming over Wireless Ad-hoc Networks	69
3.4.4	Scheduling in path diversity networks	70
3.5	Conclusion	71
4	MDSQ for Advanced Video Coding	73
4.1	Balanced MDSQ	74
4.2	Unbalanced MDSQ	75
4.3	MD Video Architecture	77
4.3.1	Open-loop MD Encoder	77
4.3.2	Open-loop MD Decoder	78
4.3.3	Drift Analysis	79
4.3.4	Multi-loop MD Encoder	82
4.3.5	Multi-loop MD Decoder	84
4.4	R-D Performance - Two Descriptions	85
4.4.1	Balanced MDSQ	86
4.4.2	Unbalanced MDSQ	92
4.5	R-D Performance - Single Description	95
4.5.1	Balanced MDSQ	96
4.5.2	Unbalanced MDSQ	101
4.5.3	Frame-by-frame Distortion	107
4.6	Conclusion	109
5	Multiple Description Video Splitting of Coded Streams	111
5.1	MDVS Scenario	112
5.2	Classic MDVS	114
5.2.1	Drift Analysis	115
5.3	MDVS with drift compensation	116
5.3.1	MDVS architecture	117
5.3.2	Simplified MDVS	118
5.4	MDVS drift performance	121
5.4.1	Intra predicted frames	122
5.4.2	MC predicted frames	122

5.4.3	Generic regular GOP	124
5.4.4	The overall effect of side information	125
5.5	MDVS streaming with path diversity	127
5.6	MDVS vs MDC: comparative discussion	131
5.7	Conclusion	132
6	MDC Rate Control and Priority Video Streaming	133
6.1	MDC Hypothetical Reference Decoder	134
6.2	The ρ model for MDC	138
6.3	MDC Rate Control	140
6.3.1	GOP Level rate control	142
6.3.2	Frame Level rate control	143
6.3.3	MDC Rate Control Performance Evaluation	145
6.3.4	Performance over lossy channels	149
6.4	Priority MDC video streaming	152
6.4.1	MD priority streaming scenario	152
6.4.2	Priority Streaming	153
6.4.3	Optimal binary frame classification	154
6.4.4	Simulation Results	157
6.5	Conclusion	160
7	Conclusions and future work	163
7.1	Conclusions	163
7.2	Future Work	165
A	Published Papers	169
A.1	Journal Papers	169
A.2	Conference Papers	169
B	Test video sequences	171
	References	175

List of Tables

4.1	Balanced MDSQ, $k=2$	75
4.2	Unbalanced MDSQ, $k=1$ and $Z = 0$	76
4.3	Balancing ratio including side information - Intra Frames (<i>Bus sequence</i>) .	102
4.4	Balancing ratio including side information - MC Frames (<i>Bus sequence</i>) . .	105
5.1	PSNR vs Side Information Redundancy (one description).	126
5.2	Average PSNR in transmission with different packet loss rates for 1Mbps@30Hz	131
6.1	Rate control accuracy	147

List of Figures

2.1	SDC rate-distortion encoder and decoder.	11
2.2	MD source coding model for $N = 2$ descriptions.	14
2.3	MD source coding model for $N = 2$ descriptions.	20
2.4	Lower Bounds on side distortions $D_1 = D_2$ when the base rate is r and fixed distortion $D_0 = 2^{-2r}$ [Goyal 2001b].	22
2.5	MDCT Generic Scheme.	24
2.6	Multiple Description Subband Coding.	27
2.7	MDC with polyphase transform and selective quantisation.	28
2.8	Generic MDSQ scheme.	30
2.9	MDSQ example 1 - Matrix representation.	31
2.10	MDSQ example 1 - Linear representation.	31
2.11	MDSQ example 2 - Matrix representation.	32
2.12	MDSQ example 2 - Linear representation.	32
2.13	MDC FEC - source data partitioning.	34
2.14	MDC FEC - N-description generation.	35
3.1	Hybrid video encoder.	39
3.2	3-D sub-band video encoding.	40
3.3	3-D sub-band decomposition.	41
3.4	Distributed video encoding.	42
3.5	Generic MD encoder	43
3.6	Open-loop multiple description video encoder.	46
3.7	Data Partitioning among channels	47
3.8	Tree-loop multiple description video encoder.	49
3.9	Pre-processing MDC.	51
3.10	An example of high-dimensional MDC for $N=4$ descriptions.	54
3.11	Unbalanced MDC application scenario.	60
3.12	Path diversity in content delivery networks.	64
3.13	MD-P2P based on multiple-tree topology.	68
3.14	MD-P2P based on mesh topology.	69
4.1	Open-loop MDSQ encoder.	78
4.2	MDSQ open-loop video decoder.	79
4.3	Distortion accumulation within an intra frame - <i>Coastguard</i>	81
4.4	Distortion accumulation in MC predicted frames- <i>Coastguard</i>	82

4.5	MDSQ encoder with drift compensation.	83
4.6	MDSQ video decoder.	85
4.7	Central Distortion <i>Foreman</i> sequence, I Frame.	87
4.8	Central Distortion <i>bus</i> sequence, I Frame.	87
4.9	Redundancy of <i>Foreman</i> sequence, I Frame.	88
4.10	Redundancy of <i>bus</i> sequence, I Frame.	88
4.11	Rate distribution for one I Frame SDC.	89
4.12	Central distortion <i>Foreman</i> sequence, IPBB Frame.	90
4.13	Central distortion <i>bus</i> sequence, IPBB Frame.	90
4.14	Redundancy of <i>Foreman</i> sequence, IPBB Frame.	90
4.15	Redundancy of <i>bus</i> sequence, IPBB Frame.	90
4.16	Rate distribution for P Frames-SDC.	91
4.17	Rate distribution for B Frames-SDC.	91
4.18	<i>Bus</i> sequence: $QP=10;k=1$	92
4.19	<i>Bus</i> sequence: $QP=20;k=1$	92
4.20	Central distortion <i>foreman</i> sequence, Intra Frame.	94
4.21	Central distortion <i>bus</i> sequence, Intra Frame.	94
4.22	Redundancy of <i>foreman</i> sequence, Intra Frame	94
4.23	Redundancy of <i>bus</i> sequence, Intra Frame.	94
4.24	Central distortion <i>foreman</i> sequence, IPBB Frame.	95
4.25	Central distortion <i>bus</i> sequence, IPBB Frame.	95
4.26	Redundancy of <i>foreman</i> sequence, IPBB Frame.	96
4.27	Redundancy of <i>bus</i> sequence, IPBB Frame.	96
4.28	Side-distortion <i>Foreman</i> sequence, Intra Frame, MDSQ with 2 diagonals.	97
4.29	Side-distortion <i>bus</i> sequence, Intra Frame, MDSQ with 2 diagonals.	97
4.30	Side-distortion <i>Foreman</i> sequence, Intra Frame CIF, MDSQ with 3 diagonals.	98
4.31	Side-distortion <i>Bus</i> sequence, Intra Frame CIF, MDSQ with 3 diagonals.	98
4.32	Side Distortion <i>Foreman</i> sequence, CIF 15HZ, MDSQ with 2 diagonals.	99
4.33	Side Distortion <i>bus</i> sequence, CIF 15HZ, MDSQ with 2 diagonals.	99
4.34	Side Distortion <i>Foreman</i> sequence, MDSQ with 3 diagonals.	100
4.35	Side Distortion <i>Bus</i> sequence, MDSQ with 3 diagonals.	100
4.36	Side Distortion <i>Foreman</i> sequence, MDSQ with 5 diagonals.	101
4.37	Side Distortion <i>Bus</i> sequence, MDSQ with 5 diagonals.	101
4.38	Side distortion <i>foreman</i> sequence, Intra Frame $QP0=11$	103
4.39	Side distortion <i>Bus</i> sequence, Intra Frame $QP0=11$	103
4.40	Side Distortion <i>foreman</i> sequence, Intra Frame $QP0=15$	104

4.41	Side Distortion <i>Bus</i> sequence, Intra Frame $QP0=15$.	104
4.42	Side Distortion <i>foreman</i> sequence, Intra Frame, $QP0=21$.	104
4.43	Side Distortion <i>Bus</i> sequence, Intra Frame, $QP0=21$.	104
4.44	Side Distortion <i>foreman</i> sequence, IPBB Frame $QP0=11$.	106
4.45	Side Distortion <i>Bus</i> sequence, IPBB Frame $QP0=11$.	106
4.46	Side Distortion <i>foreman</i> sequence, IPBB Frame $QP0=15$.	106
4.47	Side Distortion <i>bus</i> sequence, IPBB Frame $QP0=15$.	106
4.48	Side Distortion <i>foreman</i> sequence, IPBB Frame, $QP0=21$.	107
4.49	Side Distortion <i>Bus</i> sequence, IPBB Frame, $QP0=21$.	107
4.50	Frame-by-frame side PSNR <i>foreman</i> sequence, IPBB Frame, $QP0=15$; $QP1=35$.	108
4.51	Frame-by-frame side PSNR <i>bus</i> sequence, $QP0=15$; $QP1=35$.	108
5.1	MDVS application scenario.	113
5.2	Classic MDVS scheme.	115
5.3	Two-loop MDVS architecture.	117
5.4	Single-loop MDVS architecture.	119
5.5	Equivalent single-loop MDVS architecture.	120
5.6	PSNR of Intra Frame macroblocks.	123
5.7	PNSR for MC predicted frames (IPPP...) for <i>coastguard</i> and <i>foreman</i> .	123
5.8	PNSR for generic regular GOP (IPBBP...) for <i>coastguard</i> .	124
5.9	PNSR for generic regular GOP (IPBBP...) for <i>foreman</i> .	125
5.10	PNSR for generic regular GOP (IPBBP...) for <i>bus</i> .	125
5.11	Average PNSR for <i>bus</i> at 1.25 Mbit/s (Burst length BEL=4).	129
5.12	Average PNSR for <i>bus</i> at 1.25 Mbit/s (Burst length BEL=12).	129
5.13	Average PNSR for <i>bus</i> at 1.8 Mbit/s (Burst length BEL=4).	129
5.14	Average PNSR for <i>bus</i> at 1.8 Mbit/s (Burst length BEL=12).	129
5.15	Average PNSR for <i>bus</i> at 2.16 Mbit/s (Burst length BEL=4).	130
5.16	Average PNSR for <i>bus</i> at 2.16 Mbit/s (Burst length BEL=12).	130
5.17	Frame by frame PSNR for <i>bus</i> sequence.	130
6.1	Generic SDC HRD model.	135
6.2	Generic MDC HRD model.	137
6.3	Relation between percentage of zeros and rate- <i>Bus</i> sequence-I frames.	140
6.4	Relation between percentage of zeros and rate- <i>Bus</i> sequence-P frames.	140
6.5	Relation between percentage of zeros and rate- <i>Foreman</i> sequence-I frames.	140
6.6	Relation between percentage of zeros and rate- <i>Foreman</i> sequence-P frames.	140

6.7	Rate control R-D performance - <i>Foreman</i>	146
6.8	Rate control R-D performance - <i>Bus</i>	146
6.9	Bits/Description (left) and Buffer occupancy (right), <i>foreman</i> sequence- Target bitrate 1Mbit/s.	147
6.10	Bits/Description (left) and Buffer occupancy (right), <i>foreman</i> sequence- Target bitrate 1Mbit/s.	148
6.11	Bits/Description (left) and Buffer occupancy (right), <i>foreman</i> sequence- Target bitrate 1Mbit/s.	148
6.12	Average PSNR for mean burst length L=4, <i>foreman</i> sequence.	150
6.13	Average PSNR for mean burst length L=4, <i>bus</i> sequence.	150
6.14	Average PSNR for mean burst length L=4, <i>foreman</i> sequence.	151
6.15	Average PSNR for mean burst length L=4, <i>bus</i> sequence.	151
6.16	MD priority streaming scenario.	153
6.17	Priority MDC streaming scheme.	154
6.18	Simulation results for <i>Mother & daughter</i> sequence.	159
6.19	Simulation results for <i>Foreman</i> sequence.	160
6.20	Simulation results for <i>Bus</i> sequence.	160
B.1	Sequence <i>Bus</i> (CIF:352 × 288)	172
B.2	Sequence <i>City</i> (CIF:352 × 288)	172
B.3	Sequence <i>Coastguard</i> (CIF:352 × 288)	173
B.4	Sequence <i>Foreman</i> (CIF:352 × 288)	173
B.5	Sequence <i>News</i> (CIF:352 × 288)	174
B.6	Sequence <i>Mother and daughter</i> (CIF:352 × 288)	174

List of Abbreviations

AVC	Advanced Video Coding
B-MDSQ	Balanced Multiple Description Scalar Quantization
CBR	Constant Bit Rate
CDN	Content Delivery Networks
CIF	Common Intermediate Formate
DCT	Discrete Cosine Transform
DPCM	Differential Pulse Coding Modulation
DVC	Distributed Video Coding
DWT	Discrete Wavelet Transform
EZW	Embedded Zerotree Wavelet
FEC	Forward Error Correction
FGS	Fine Granularity Scalability
FMO	Flexible Macroblock Ordering
GOP	Group of Pictures
HP	High Priority
HRD	Hypothetical Reference Decoder
HVS	Human Visual System
IP	Internet Protocol
IPTV	Internet Protocol Television
JPEG	Joint Photographic Experts Group
KLT	Karhunen-Loeve Transform
LAN	Local Area Networks

LC	Layered Coding
LP	Low Priority
MC	Motion Compensation
MCTF	Motion Compensated Temporal Filtering
MD	Multiple Description
MD-FEC	Multiple Description with Forward Error Correction
MDC	Multiple Description Coding
MDCT	Multiple Description with Correlating Transforms
MDDVC	Multiple Description with Distributed Video Coding
MDSQ	Multiple Description Scalar Quantization
MDVS	Multiple Description Video Splitting
ME	Motion Compensation
MP	Matching Pursuits
MPEG	Moving Picture Expert Group
MV	Motion Vector
NALU	Network Abstraction Layer Unit
OBMC	Overlapped Block Motion Compensation
P2P	Peer-to-Peer
PCT	Pairwise Correlating Transform
PLR	Packet Loss Rate
PSNR	Peak to Peak Signal-to-Noise Ratio
QCIF	Quarter Common Intermediate Formate
QoS	Quality of Service

QP	Quantization Parameter
R-D	Rate-Distortion
R-S	Reed-Solomon
SD	Single Description
SDC	Single Description Coding
SVC	Scalable Video Coding
U-MDC	Unbalanced Multiple Description Coding
U-MDSQ	Unbalanced Multiple Description Scalar Quantization
UEP	Unequal Error Protection
VBR	Variable Bit Rate
VoD	Video on Demand
WLAN	Wireless Local Area Networks
WZ	Wyner-Ziv

Introduction

Contents

1.1	Motivation	3
1.2	Main Contributions	4
1.3	Outline of the thesis	6

In the current heterogeneous communication environments, the great variety of multimedia communications combined with fast evolution of networking architectures and topologies, give rise to new research problems related to the various elements of the communication chain. Regarding networking, such as Internet, wireless LAN or last generation wireless access networks, different quality of service constraints exist, namely bandwidth, delay, jitter and packet loss which pose new problems to ensure an appropriate level of error robustness. On the other hand, different terminal capabilities and convergence between traditional TV sets, Internet TV and mobile video streaming, also pose challenges in content adaptation and technology constraints, such as those resulting from limited processing capacity, memory, display size and power autonomy.

Additionally, new emerging video streaming services and applications using both centralized Content Delivery Networks(CDN) and distributed infrastructures, such as Peer-to-Peer (P2P) Networks have been developed in recent years. In this scenario, path diversity is seen as an efficient transmission scheme capable to providing high adaptability to network reconfiguration, and also error robustness and load balancing. In CDN schemes, a high number of access users leads to adoption of streaming schemes using server diversity. On the other hand, P2P networks have to deal with problems related to network management and configuration, because a user is a network receiver, and at the same time, can be a network node to other peer. Also, each node can join/leave the network at any instant,

forcing to a new network reconfigurations. Furthermore, taking into account the network dynamics, the massive dissemination of high quality multimedia content combined with the access ubiquity, pose new challenges to video coding methods in terms of efficiency, error resilience and quality of service.

Multiple Description Coding (MDC) has recently been given particular attention by the research community as a promising approach to improve the quality of multimedia streaming over error-prone networks with path diversity. In MDC, a video signal is typically encoded into several independent descriptions, i.e., compressed streams, where each one can be delivered over a separate channel making use of available path diversity. If joint decoding of all descriptions is done at the receiver, then the quality of the reconstructed signal is higher than that obtained from individual decoding of any single description. These interesting features of MDC are accomplished at the cost of higher coding rate, i.e., redundancy, when compared to classic single description coding (SDC) [Goyal 2001a]. This is essentially driven by networks with multiple available paths from the sender to the receiver (e.g., mesh and overlay networks) and multiple source coding representations (i.e., MDC) that go beyond the classical paradigm of SDC, where one source is encoded into one single representation [Frossard 2008].

The use of MDC in path diversity video networking is a competitive alternative with other network-adaptation schemes, such as scalable video coding (SVC) for robust streaming over time-varying networking conditions. In general, the main difference between SVC and MDC lies in the inter-dependency of SVC layers in contrast with independent MDC descriptions. MDC is also an interesting approach whenever retransmission schemes are not a viable solution, such as when transmission delay is not acceptable. Also, compared with the channel coding approaches usually referred to as forward error correction codes (FEC), MDC techniques are more efficient for high packet loss rates where long code sizes are needed, creating the difficulties associated with overhead delay associated to FEC.

Therefore further research on robust video coding and adaptation techniques is necessary, in order to adapt the quality of coded signals to the constraints and needs posed by network transmission infrastructure. Particularly, in the MDC field, novel coding and adaptation schemes are needed to improve error robustness on live and pre-encoded video signals, aiming to obtain the highest possible video quality at each user terminal.

1.1 Motivation

Traditionally, MDC schemes use descriptions of equal importance due to the principle that each description should be independently decodable and an acceptable quality must be provided by decoding any single description. On the other hand, the heterogeneity of the networks and applications requires adaptability of source coding schemes in terms of the rate-distortion performance. Although several works have been proposed to obtain the best coding parameters which give the minimum distortion based on known channel conditions, these works are mainly focus on balanced descriptions, *i.e* with same rate. Therefore, investigation on unbalanced MDC schemes and rate control methods for MDC solutions have not been fully exploited in the literature, which is an interesting topic to evaluate in order to cope with content delivery chains with different resources.

Whenever a coded video stream is processed, the predictive nature of video coding algorithms must be taken into account, because of distortion accumulation at the end-user decoding terminal (*i.e* drift). In MDC, the effect of drift can be explained as follows. In the absence of errors or data loss, all descriptions are decoded and the reconstructed blocks/frames are then used as reference for others, providing accurate predictions for decoding. However, if any description is lost, then the predictions reconstructed in the decoder do not match those originally used for encoding the original SDC stream. This mismatch is the origin of drift by adding distortion to decoded video, which is further accumulated in the reconstruction loop and propagated throughout all subsequent predicted blocks/frames. New problems need to be solved when the most recent advanced predictive techniques are expanded into MDC methods, which means that further research work needs to be done in order to propose new efficient MDC methods adapted to advanced video coding schemes.

Combination of MDC with path diversity has always been used in communication chains typically comprised of an uncompressed source signal feeding an MDC encoder, followed by multiple transmission paths to the receiver [Mao 2005b], [Akyol 2007] or by streaming multiple complementary descriptions distributed across the edge servers of content delivery networks [Apostolopoulos 2002], [Ahuja 2008]. A shortcoming of such communication model is that it does not take into account the typical scenario of current heterogeneous networking, where single path routes co-exist with multiple paths in the same delivery chain.

The novel concept of multiple description video splitting of coded streams (MDVS) introduced in this thesis fills the existing gap in heterogeneous video communications where an SDC stream is transmitted over a single path network and then needs to be split into several MDC streams. This might be particularly useful at edge nodes to benefit from path diversity over different networks where multiple paths are available. Since MDVS operates on coded streams, any networking node with such processing capability can split an incoming SDC stream into the different outgoing paths that can be used from that particular network node to the end user terminal. A recent work highlighting the advantages of using MDVS for robust video streaming and to deal with handoff over wireless local area networks (WLAN) is presented in [Chen 2007b].

Since MDVS suffers from the same intrinsic problem of drift as MDC, i.e., accumulation of decoding distortion when any description is lost in the network, an open research topic that needs to be addressed is the development of novel processing architectures capable of limiting such type of distortion. New types of low complexity MDVS architectures must be investigated to efficiently generate MDC video streams from the SDC.

Despite the improvement of error resilience in video transmission chains due to path diversity, intermediate nodes must be able to use scheduling algorithms in order to increase the quality of experience under network congestions. Based on scheduling decisions, one network node selects the packet to be either dropped or re-transmitted. Usually these methods use rate-distortion optimization in order to maximize the overall decoding quality under a certain rate constraint. The research on scheduling and prioritization schemes for MDC-based applications is still an open issue and also needs further investigation in order to find efficient methods to adapt the MDC error resilience capabilities to the dynamic nature of path diversity video delivery networks.

1.2 Main Contributions

Drift-free multiple description for advanced video coding A multiple description coding scheme based on a multi-loop structure is proposed, which prevents drift distortion accumulation in both intra and inter predicted slices. The drift is compensated by generating a controlled amount of side information used by the decoder whenever any description is lost. The experimental results show that predicted slices do not suffer from drift and their quality is significantly improved at reduced redundancy cost in comparison

with the open-loop implementation. Papers J1, C6 and C7 (Annex A) results from this investigation.

Unbalanced Multiple Description Scalar Quantisation A new method is proposed to generate unbalanced descriptions using Multiple Description Scalar Quantisation (MDSQ). For this purpose, a novel index assignment method is proposed based on a new set of tables that dynamically generate two descriptions with different unbalancing rates including the traditional balanced case. The performance evaluation results show that the proposed scheme achieves better rate-distortion performance comparing with balanced scheme. Good performance is also obtained in packet loss simulation scenarios. Paper C2 (Annex A) results from this investigation (to be submitted).

Multiple Description Video Splitting An extension of the current MDC concept to the compressed domain is proposed, by developing an efficient method for splitting of standard SDC video into a multi-stream representation. A novel Multiple Description Video Splitting (MDVS) scheme is proposed to operate at network edges, for increased robustness in path diversity video streaming across heterogeneous communications chains. The proposed scheme is able to effectively control drift distortion in both intra and inter predictive coding, even when only one description reaches the decoder. The simulation results show that any individual description can be decoded on its own without producing drift, achieving significant quality improvement at reduced redundancy cost. Papers J1, C4 and C5 (Annex A) results from this investigation.

Rate control for multiple description video coding The problem of rate control in unbalanced MDSQ is investigated and an extension of the ρ model was developed based on the evidence that the linear relationship between the rate of each description and the corresponding percentage of zeros in transform coefficients is maintained when moving from the SDC into the MDC domain. The simulation results that the rate control algorithm exhibits high accuracy for a given target bitrate and unbalancing ratio between two descriptions without having buffer underflow and overflow. The proposed method is useful in MDC video streaming over asymmetric transmission paths, where the rate of each description can be dynamically adapted to different channel conditions. Paper C1 (Annex A) results from this investigation (to be submitted).

Optimal priority MDC video streaming for networks with path diversity A robust video streaming scheme for priority networks with path diversity is proposed, based on a combined approach of MDC with optimal frame classification into two priorities. A binary frame classification algorithm is proposed to define high (HP) and low (LP) priority network abstraction layer units (NALU), which in turn define the packet priorities. An optimization algorithm finds HP pictures based on minimization of the packet loss concealment distortion. The proposed algorithm is able to effectively improve the decoded video quality without increasing the MDC stream redundancy. Paper C3 (Annex A) results from this investigation.¹

1.3 Outline of the thesis

This thesis comprises seven chapters dealing with different aspects of MD video coding research. The thesis consists in two parts: the first part of the thesis is formed by Chapter 2 and Chapter 3 that describe the theory and the state of the art of MD coding respectively. The second part is formed by Chapters 4, 5, 6 and 7 which describes the work and contributions developed in this thesis.

Chapter 2 is focused on MDC techniques. Firstly, the theoretical definitions and optimal rate-distortion bounds are explained. Secondly, the MDC techniques currently used in the literature are described, including, Multiple Description with Correlating Transforms (MDCT), Multiple Description Subband Coding, based on Polyphase Decomposition and Selective Quantisation, Multiple Description Scalar Quantisation (MDSQ), and MDC based on the Error Correction Codes.

Chapter 3 describes the main MDC video architectures for different video coding schemes including scalable coding, multi-view and stereo video coding, and also for the MDC video applications with $N > 2$ descriptions. In the second part of the chapter, the different path diversity topologies and the MDC networking applications are described.

Chapter 4 presents the main contributions of this thesis for MDC in advanced video coding. Firstly the overall MDSQ video architecture is presented. A rate-distortion evaluation of balanced MDSQ is explored, comparing the performance of the proposed architecture with a classic open-loop scheme. Secondly, a new set of index assignment

¹This is a joint work with Lino Ferreira (PhD student) and Luís Cruz (Professor at the University of Coimbra)

tables is proposed to generate unbalanced descriptions. A comparison between balanced and unbalanced schemes is carried out, taking into account, the overall R-D performance and also the side-information coding efficiency.

Chapter 5 addresses MDVS for coded video. This chapter proposes a novel MDVS scheme based on MDSQ, using side information to control drift in both spatial and temporal prediction. The side information is generated from the original stream and its rate is controlled with an independent quantisation parameter which also controls redundancy. Then, a simplified architecture is devised to reduce the overall complexity in regard to the number of processing functions and memory requirements.

Chapter 6 is divided in two different parts. The first part deals with the problem of rate control in unbalanced MD video coding. A new method is proposed to ensure an asymmetric target bit rate in each description. An extension of the ρ model is developed based on the evidence that the linear relationship between the rate of each description and the corresponding percentage of zeros in transform coefficients is maintained when moving from the single description domain (SDC) into the MDC domain. Finally an experimental study on error-resiliency features for distinct packet-loss rates, and available channel bitrates is presented. The second part proposes a robust video streaming scheme for priority networks with path diversity, based on a combined approach MDC with optimal picture classification into two priorities. An optimization algorithm finds high priority frames based on minimization of the packet loss concealment distortion. The proposed algorithm is able to effectively improve the decoded video quality without increasing the MDC stream redundancy.

Finally, Chapter 7 concludes the thesis and discuss several topics for future work in the Multiple Description Video Coding.

Multiple Description Coding Techniques

Contents

2.1	Theoretical Basis	10
2.1.1	Single Description Rate-Distortion Region	11
2.1.2	Multiple Description Rate-Distortion Region	13
2.1.3	Multiple Description Bounds for a Gaussian Source	16
2.1.4	Representation of MD Rate-Distortion Region	18
2.1.5	Redundancy	20
2.2	Practical Implementation	22
2.2.1	Multiple Description with Correlating Transforms	23
2.2.2	Multiple Description Sub-band Coding	27
2.2.3	Polyphase Decomposition and Selective Quantization	28
2.2.4	Multiple Description Scalar Quantisation	29
2.2.5	MDC Based on Error Correction Codes	34
2.3	Conclusion	35

This chapter presents a review of Multiple Description Coding (MDC) principles, where the relevant theoretical definitions and optimal rate-distortion bounds are explained. The main MDC techniques documented in the literature are described, including Multiple Description with Correlating Transforms (MDCT), Multiple Description Subband Coding, based on Polyphase Decomposition and Selective Quantisation, Multiple Description Scalar Quantisation (MDSQ), and MDC based on Error Correction Codes.

2.1 Theoretical Basis

The initial motivation for MDC was driven by the need of high reliability in the telephone networks in order to improve mechanisms to handle outages. Since the use of path diversity is an efficient approach to provide robustness in transmission networks, in the case of telephone networks, this was achieved using redundant standby transmission links. One way to improve robust transmission without using standby links was to split the speech signal among two different links or paths. Hence, in normal operation, the communication would be done with good quality, and when any link or path failed, the other one would guarantee communication with a lower, yet acceptable quality [Goyal 2001a].

The theoretical formulation of this practical constraint was done by Ozarow in [Ozarow 1980], by defining MDC as *the source coding problem with two channels and three receivers*. The starting point is the Shannon rate-distortion theory, which is named as *Single Description (SD) rate-distortion region* in this context. The focus on this theory is on derivation of performance bounds, that is, determining the region of achievable points in the rate-distortion trade-off for certain limited statistical sources.

While traditional rate-distortion theory intends to find the minimum achievable rate for a given distortion, MDC rate-distortion approach increases the dimension of the problem and needs to find the optimal jointly achievable distortions for three different decoded representations that can be obtained from two different descriptions of the source encoded at the respective rates. This is different from the traditional rate-distortion formulation because the distortion achieved by joining two streams does not correspond to the Shannon limits for a single source encoded at the same rate as that of both descriptions added together.

Thus, the MDC theoretical characterization has not been formulated in a closed form, but only some bounds have been found along with some practical approximations. This was done in [Gamal 1982], with the definition of the *Multiple Description (MD) rate-distortion region*, evaluated for Gaussian sources and quadratic distortion. In MDC context, if each description is individually encoded, each one with a different rate, the performance rate-distortion region corresponds to the combination of achievable distortions of each description individually decoded, and at the same time, obtaining a minimum as possible distortion when all descriptions are merged. This formulation is presented in section 2.1.2.

These optimal bounds of MD rate-distortion region cannot be established for situations

of practical relevance since they are taken for certain limited statistical source classes. Moreover, in general these bounds provide useful information to benchmark practical implementations of MDC techniques.

2.1.1 Single Description Rate-Distortion Region

In order to better understand the fundamental MDC theoretical statements, some basic definitions of rate-distortion theory are explained following the approach used in [Cover 2006]. The main concern of rate-distortion theory is how to represent a source with the minimum possible number of bits for a given reproduction quality. This formulation requires a measure of the source fidelity which leads to the definition of a distortion function based on a distance metric between the source and its coarse representation.

Fig.2.1 shows the general rate-distortion model for a SD coding scheme. The source signal is represented by a finite alphabet χ in the form of a vector of n independent identically distributed (*i.i.d.*) random variables X_1, X_2, \dots, X_n with a probability mass function $p(x), x \in \chi$. The encoder describes the source vector X^n by an index $f_n(X^n) \in \{2^0, 2^1, \dots, 2^{nR}\}$, where nR is the *rate* in bits per alphabet symbol. The decoder represents X^n by an estimate $\hat{X}^n \in \hat{\chi}$. Usually, the reproduction alphabet $\hat{\chi}$ is the same as χ .

A distortion measure is a mapping function,

$$d : \chi \times \hat{\chi} \rightarrow \mathfrak{R}^+ \quad (2.1)$$

from the set of source pairs reproduction into the set of non-negative real numbers. The distortion $d(x, \hat{x})$ measures the cost of representing symbol x by symbol \hat{x} .

In general, several different distortion metrics can be used, but usually the squared-error, $d(x, \hat{x}) = (x - \hat{x})^2$ is chosen for rate-distortion analysis. This distortion function can be

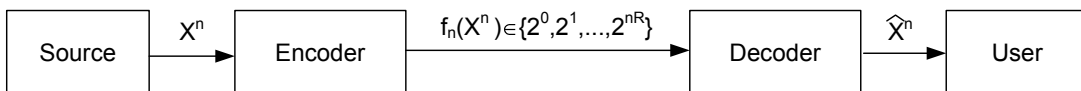


Figure 2.1: SDC rate-distortion encoder and decoder.

extended for vectors of symbols as follows:

$$d(x^n, \hat{x}^n) = \frac{1}{n} \sum_{i=1}^n d(x_i, \hat{x}_i) \quad (2.2)$$

The distortion measure for a sequence is the average of the per symbol distortion taking into account all the elements of the sequence.

A rate-distortion code, defined as $(2^{nR}, n)$, consists in an encoding function,

$$f_n = \mathcal{X}^n \rightarrow \{2^0, 2^1, \dots, 2^{nR}\}, \quad (2.3)$$

and a decoding (reproduction) function,

$$g_n = \{1, 2, \dots, 2^{nR}\} \rightarrow \hat{\mathcal{X}}^n. \quad (2.4)$$

The distortion associated with $(2^{nR}, n)$ code is

$$D = E[d(X^n, g_n(f_n(X^n)))] = \sum_{x^n} p(x^n) d(x^n, g(f(x^n))), \quad (2.5)$$

where the set of n-tuples $g(1), g(2), \dots, g(2^{nR})$, denoted by $\hat{X}(1), \hat{X}(2), \dots, \hat{X}(2^{nR})$ constitutes the *codebook* and $f^{-1}(1), f^{-1}(2), \dots, f^{-1}(2^{nR})$ are the associated *assignment regions*.

A rate-distortion pair (R, D) is said to be *achievable* if there exists a sequence of $(2^{nR}, n)$ rate-distortion codes (f, g) such that

$$\lim_{n \rightarrow \infty} E[d(X^n, g_n(f_n(X^n)))] \leq D. \quad (2.6)$$

After defining the *achievability* concept for a given (R, D) pair, the definition of *SD rate-distortion region* for a source is the closure of the set of the achievable rate-distortion pairs (R, D) .

Finally, let the *rate-distortion function* $R(D)$ be the *infimum* of rates R such that (R, D) lies in the rate-distortion region of the source for a given distortion D .

Another way of looking the *rate-distortion function* is with the definition of the *distortion-*

rate function $D(R)$ which is the *infimum* of all distortions D such that (R, D) is in the *rate-distortion region* of the source for a given R .

Functions $R(D)$ and $D(R)$ both provide equivalent definitions for dealing with the boundary of the *rate-distortion region*.

The *rate-distortion function* for an i.i.d. source X with distribution $p(x)$ and distortion function $d(x^n, \hat{x}^n)$ is equal to

$$R(D) = \inf I(X; \hat{X}) \quad (2.7)$$

where

$$I(X; \hat{X}) = \sum_{x, \hat{x}} p(x, \hat{x}) \log(p(x, \hat{x})/p(x)p(\hat{x})), \quad (2.8)$$

is the *mutual information* between X and \hat{X} . The *infimum* is taken over all joint probability mass functions $p(\hat{x}, x)$ such that

$$\sum_{x, \hat{x}} p(x, \hat{x}) d(x, \hat{x}) \leq D. \quad (2.9)$$

2.1.2 Multiple Description Rate-Distortion Region

The SD rate-distortion definitions can be extended for MDC. The main difference lies in the definition of a quintuple $(R_1, R_2, D_0, D_1, D_2)$ instead of a single pair (R, D) .

Fig. 2.2 shows the generic MD scenario for two descriptions, originally named as the channel-splitting problem [Ozarow 1980]. The sequence of source symbols X^n is encoded into two redundant and mutually refinable descriptions represented by indices $f_{1n}(X^n) \in \{1, 2, \dots, 2^{nR_1}\}$ and $f_{2n}(X^n) \in \{1, 2, \dots, 2^{nR_2}\}$, which are transmitted separately over two different noiseless or error corrected channels: *Channel 1* and *Channel 2*. Three decoders are used at the receiver side. *Decoder 0* (*Central Decoder*) is the one that receives the information from both channels and outputs the estimate of the source X^n , \hat{X}_0^n . *Decoder 1* and *Decoder 2* (*Side Decoders*) receive information only from each individual channel. The outputs are \hat{X}_1^n and \hat{X}_2^n , respectively.

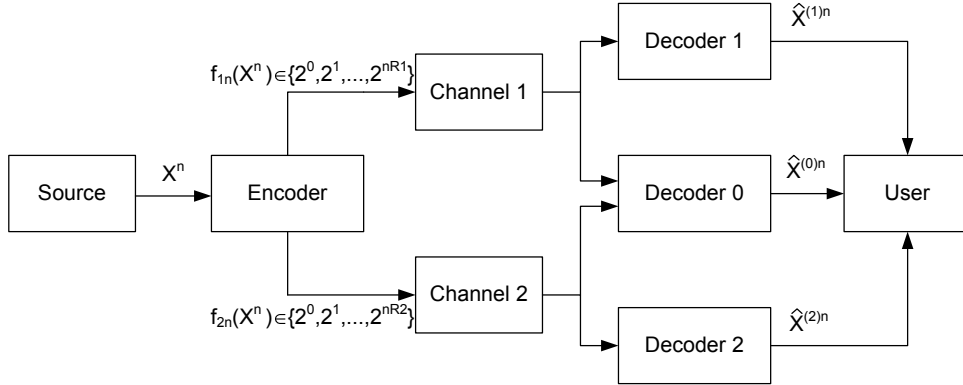


Figure 2.2: MD source coding model for $N = 2$ descriptions.

A $(2^{nR1}, 2^{nR2}, n)$ rate-distortion code consists of an encoding function,

$$f_{1n} = \chi^n \rightarrow \{2^0, 2^1, \dots, 2^{nR1}\}; \quad (2.10)$$

$$f_{2n} = \chi^n \rightarrow \{2^0, 2^1, \dots, 2^{nR2}\}; \quad (2.11)$$

The decoding functions are defined as,

$$g_{0n} = \{\{1, 2, \dots, 2^{nR1}\}, \{2^0, 2^1, \dots, 2^{nR2}\}\} \rightarrow \hat{\chi}_0^n, \quad (2.12)$$

$$g_{1n} = \{2^0, 2^1, \dots, 2^{nR1}\} \rightarrow \hat{\chi}_1^n, \quad (2.13)$$

$$g_{2n} = \{2^0, 2^1, \dots, 2^{nR2}\} \rightarrow \hat{\chi}_2^n. \quad (2.14)$$

This MD model produces three different approximations of the sequence of source symbols X^n , which implicitly requires the definition of three distortion measures, one for each of the three decoded reconstructions \hat{X}_i^n produced by decoder i , $i = 0, 1, 2$, in regard to the same sequence of source symbols X^n .

The distortion measure functions for each reconstruction $i = 0, 1, 2$ is defined as

$$d_i(x^n, \hat{x}_i^n) = \frac{1}{n} \sum_{k=1}^n d(x_k, \hat{x}_{ik}) \quad i = 0, 1, 2. \quad (2.15)$$

Consequently, the distortion associated to $(2^{nR_1}, 2^{nR_2}, n)$ codes are defined as the following set of distortions

$$D_0 = E[d_0(X^n, g_{0n}(f_{1n}(X^n)), f_{2n}(X^n))], \quad (2.16)$$

$$D_1 = E[d_1(X^n, g_{1n}(f_{1n}(X^n)))], \quad (2.17)$$

$$D_2 = E[d_2(X^n, g_{2n}(f_{2n}(X^n)))], \quad (2.18)$$

where D_0 is defined as *central distortion* and $D_i, i = 1, 2$ are the *side distortions*. A rate pair (R_1, R_2) is said to be *achievable* for distortion $\mathbf{D} = (D_0, D_1, D_2)$, if there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ rate-distortion codes

$$\lim_{n \rightarrow \infty} E[d_0(X^n, g_{0n}(f_{1n}(X^n)), f_{2n}(X^n))] \leq D_0. \quad (2.19)$$

$$\lim_{n \rightarrow \infty} E[d_1(X^n, g_{1n}(f_{1n}(X^n)))] \leq D_1, \quad (2.20)$$

$$\lim_{n \rightarrow \infty} E[d_2(X^n, g_{2n}(f_{2n}(X^n)))] \leq D_2. \quad (2.21)$$

The *MD rate-distortion region* for a source is defined as the closure of the set of achievable rate pairs (R_1, R_2) whose reconstructions achieve expected distortions $\mathbf{D} = (D_0, D_1, D_2)$.

Inversely, the MD rate-distortion region is defined as the achievable set of distortions $\mathbf{D} = (D_0, D_1, D_2)$ with rate pairs (R_1, R_2) that satisfies the following cumulative conditions,

$$D_0 \geq D(R_1 + R_2), \quad (2.22)$$

$$D_1 \geq D(R_1), \quad (2.23)$$

$$D_2 \geq D(R_2). \quad (2.24)$$

Achieving equality simultaneously in three equations (2.22), (2.23) and (2.24) would imply that an optimal rate $R_1 + R_2$ can be partitioned into optimal rate R_1 and R_2 . Unfortunately, this is not possible because the individual descriptions with rates R_1 and R_2 are

similar and redundant when combined. Thus, the MD rate-distortion problem differs from the SD classical rate-distortion since it is impossible to satisfy the limits $R_1 = R(D_1)$, $R_2 = R(D_2)$, and at the same time, obtain from the central decoder distortion D_0 such that $R_1 + R_2 = R(D_0)$. The main problem is to find appropriate descriptions sufficiently different, with reduced redundancy when combined together.

2.1.3 Multiple Description Bounds for a Gaussian Source

The definition of the MD rate-distortion region is in general a complex problem, highly dependent of the source statistics and number of descriptions. The characterization of the set of achievable distortions for a Gaussian source described by two descriptions was made by Ozarow in [Ozarow 1980]. For a memoryless Gaussian source of variance σ^2 and squared error-distortion measure, the set of achievable distortions $\mathbf{D}(\sigma^2, R_1, R_2)$ is the one that satisfies the equations [Ozarow 1980].

$$D_1 \geq \sigma^2 2^{-2R_1}, \quad (2.25)$$

$$D_2 \geq \sigma^2 2^{-2R_2}, \quad (2.26)$$

$$D_0 \geq \sigma^2 2^{-2(R_1+R_2)} \gamma_D, \quad (2.27)$$

where,

$$\gamma_D = \begin{cases} \frac{1}{1-(\sqrt{\Pi}-\sqrt{\Delta})^2}, & D_1 + D_2 < \sigma^2(1 + 2^{-2(R_1+R_2)}) \\ 1, & otherwise \end{cases} \quad (2.28)$$

Π and Δ are defined as,

$$\Pi = (1 - D_1/\sigma^2)(1 - D_2/\sigma^2), \quad (2.29)$$

and

$$\Delta = (D_1 D_2)/\sigma^4 - 2^{-2(R_1+R_2)}. \quad (2.30)$$

The bounds (2.25) and (2.26) are referring to the side-distortion and follows from equations (2.23) and (2.24). The inequality (2.27) means that the central distortion must

exceed the rate-distortion bound by the factor γ_D .

The behavior of γ_D was clarified in [Goyal 2001b] for three different cases, considering $\sigma^2 = 1$:

- (a) Both descriptions are individually good;
- (b) Joint description is as good as possible;
- (c) Balanced descriptions, where $R_1 = R_2$ and $D_1 = D_2$.

(a) Both descriptions are individually good Assuming that both descriptions are individually good, distortions D_1 and D_2 are set by minimum distortion as given by equations 2.23) and (2.24, which means that $D_1 = 2^{-2R_1}$ and $D_2 = 2^{-2R_2}$. Replacing D_1 and D_2 in equations 2.29 and 2.30, $\gamma_D = \frac{1}{1-(1-D_1)(1-D_2)}$ and $\Delta = 0$ respectively.

Consequently,

$$D_0 \geq D_1 D_2 \frac{1}{1 - (1 - D_1)(1 - D_2)} = \frac{D_1 D_2}{D_1 + D_2 - D_1 D_2}. \quad (2.31)$$

A further chain of inequalities gives $2D_0 \geq \min\{D_1, D_2\}$, which means that the joint description is only slightly better than the two individual descriptions [Goyal 2001b].

(b) Joint description is as good as possible Considering that joint description is as good as possible, this means that $D_0 = 2^{-2(R_1+R_2)}$. In that case $\gamma_D = 1$, and thus

$$D_1 + D_2 \geq 1 + 2^{-2(R_1+R_2)}. \quad (2.32)$$

An unitary distortion value for a gaussian source means that $R(D) = 0$, which means that no information is sent, and the reconstruction is made using its mean [Cover 2006]. From the right side of inequality of equation (2.32), one can conclude that at least, one of the side decoders has a very poor reconstruction.

Balanced Descriptions The last case is considering balanced descriptions where $R_1 = R_2$ and $D_1 = D_2$. Two different cases are important to be mentioned in order

to qualitatively understand the MD rate-distortion bounds. The first relates the exponential decay of the three distortions, while the second bounds the side distortion based on the gap between D_0 and $2^{-2(R_1+R_2)}$.

Assuming $R_1 = R_2 \gg 1$ and $D_1 = D_2 \ll 1$, then,

$$\gamma_D = \frac{1}{1 - ((1 - D_1) - \sqrt{D_1^2 - 2^{-4R_1}})^2} \approx \frac{1}{1 - ((1 - D_1) - D_1)^2} = \frac{1}{4D_1 - 4D_1^2} \approx \frac{1}{4D_1}. \quad (2.33)$$

With $\gamma_D = (4D_1)^{-1}$ in (2.27) the central distortion is

$$D_0 \geq \frac{2^{-4(R_1)}}{4D_1} \quad (2.34)$$

and the product of central and side distortions D_0D_1 is approximately lower-bounded by $4^{-1}2^{-4R_1}$. With $D_1 = D_2 \approx 2^{-2(1-\alpha)R_1}$ and $0 \leq \alpha \leq 1$, the best central distortion exponential decay is

$$D_0 \approx \frac{2^{-2(1-\alpha)R_1}}{4} \quad (2.35)$$

This result shows that the penalty in the exponential decay of D_1 inserted by parameter α corresponds to increasing of the rate decay of D_0 , which means that increasing the side balanced distortions reduces the corresponding central one.

2.1.4 Representation of MD Rate-Distortion Region

Instead of representing MD rate-distortion region as a function of the minimum distortion for a given rate, as defined in equations (2.25), (2.26) and (2.27), this can be done by describing the dual problem. This consists in representing the rate-distortion region, by finding the minimum rates for each description for a given a distortion value. Thus, given the quadratic distortions $\mathbf{D} = (D_0, D_1, D_2)$, the set of admissible rate pairs R_1, R_2 are defined as [Zamir 1999],

$$R_1 \geq \frac{1}{2} \log\left(\frac{\sigma^2}{D_1}\right), \quad (2.36)$$

$$R_2 \geq \frac{1}{2} \log\left(\frac{\sigma^2}{D_2}\right), \quad (2.37)$$

$$R_1 + R_2 \geq \frac{1}{2} \log\left(\frac{\sigma^4}{D_1 D_2}\right) + \delta, \quad (2.38)$$

where $\delta = \delta(\sigma^2, D_0, D_1, D_2)$ is defined as

$$\delta = \begin{cases} \frac{1}{2} \log\left(\frac{1}{1-\rho^2}\right), & D_0 \leq D_{0max} \\ 0, & D_0 \geq D_{0max} \end{cases} \quad (2.39)$$

and represents the excess rate in each description comparing with a rate-distortion functions for a Gaussian source $X \sim N(0, \sigma^2)$ at distortions D_1 and D_2 .

$$\rho = \begin{cases} -\frac{\sqrt{\pi\varepsilon_0^2 + \gamma} - \sqrt{\pi\varepsilon_0^2}}{(1-\varepsilon_0)\sqrt{\varepsilon_1\varepsilon_2}} & D_0 \leq D_{0max} \\ -\sqrt{\frac{\pi}{\varepsilon_1\varepsilon_2}}, & otherwise \end{cases} \quad (2.40)$$

$$\gamma = (1 - \varepsilon_0)[(\varepsilon_1 - \varepsilon_0)(\varepsilon_2 - \varepsilon_0) + \varepsilon_0\varepsilon_1\varepsilon_2 - \varepsilon_0^2], \quad (2.41)$$

$$\pi = (1 - \varepsilon_1)[(1 - \varepsilon_2)], \quad (2.42)$$

$$\varepsilon_i = \frac{D_i}{\sigma^2}, \quad for \quad i = 0, 1, 2, \quad (2.43)$$

and

$$D_{0max} = \frac{D_1 D_2}{D_1 + D_2 - \frac{D_1 D_2}{\sigma^2}}. \quad (2.44)$$

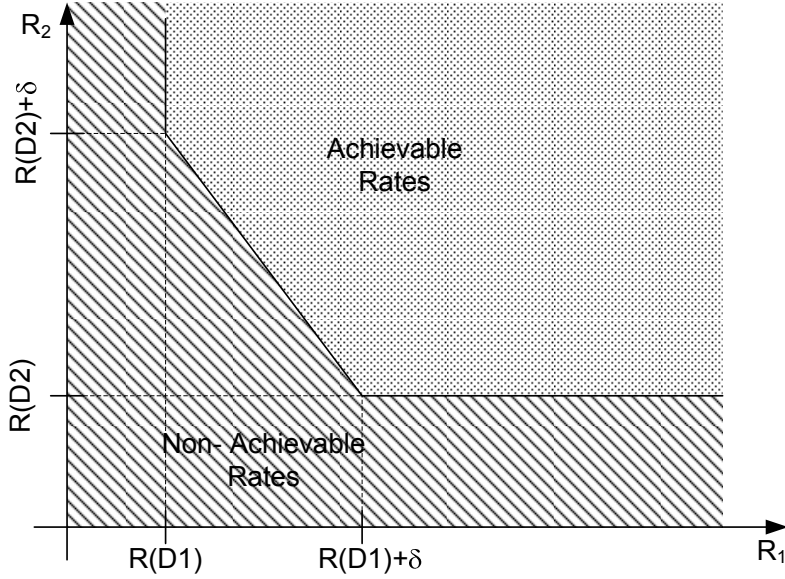


Figure 2.3: MD source coding model for $N = 2$ descriptions.

Fig. 2.3 shows the typical form of the achievable rates $R(\sigma^2, D_0, D_1, D_2)$ for $D_0 \leq D_{0max}$. $\delta, \gamma \geq 0$ and $-1 \leq \rho \leq 0$, for all $D_1, D_2 \leq \sigma^2$ and $D_0 \leq D_{0max}$.

The quantity $\delta = \delta_1 + \delta_2 \geq 0$ is defined as *excess marginal rate* where

$$\delta_i = R_i - \frac{1}{2} \log\left(\frac{\sigma^2}{D_i}\right) \quad i = 1, 2. \quad (2.45)$$

If D_1 and D_2 are held fixed, then the maximum central distortion D_{0max} can be achieved without excess marginal rate, *i.e* with $\delta = 0$, and

$$R_1 + R_2 = R(D_0). \quad (2.46)$$

2.1.5 Redundancy

Following [Goyal 2001b], the tradeoff between central and side distortions is interpreted by rearranging equation (2.27), considering that $D_1 = D_2$ and $\sigma^2 = 1$. Consequently,

$$D_1 \geq \min\left\{\frac{1}{2}[1 + D_0 - (1 - D_0)\sqrt{1 - 2^{-2(R_1+R_2)/D_0}}], 1 - \sqrt{1 - 2^{-2(R_1+R_2)/D_0}}\right\}. \quad (2.47)$$

The overall coding rate $R_1 + R_2$ is comprised of two components. The *base rate* r and *redundancy*, ρ where the base rate corresponds to the quality of the joint description and the redundancy is the additional rate necessary to make the side distortion low.

$D_0=2^{-2r}$ and $\rho=R_1 + R_2 - r$. Substituting in 2.47 gives,

$$D_1 \geq \begin{cases} \frac{1}{2}[1 + 2^{-2r} - (1 - 2^{-2r})\sqrt{1 - 2^{-2\rho}}], & \text{for } \rho \leq r - 1 + \log_2(1 + 2^{-2r}) \\ 1 - \sqrt{1 - 2^{-2\rho}}, & \text{for } \rho \geq r - 1 + \log_2(1 + 2^{-2r}) \end{cases} \quad (2.48)$$

Fig.2.4 shows the graphical representation of equation 2.48. From this figure, one can see that, the curves are partially coincident because the distortion bound is independent of r at high redundancies.

From 2.48, the slope of the low-redundancy D_1 versus the ρ characteristic is defined as

$$\frac{\partial D_1}{\partial \rho} = -\frac{1 - 2^{-2r}}{2} \frac{2^{-2\rho} \ln 2}{\sqrt{1 - 2^{-2\rho}}}. \quad (2.49)$$

At $\rho = 0^+$, the slope is infinite. When designing a system one should look, not only at the central distortion but also at the side distortion. If side distortion is also to be considered, the infinite slope means that a small increase in the overall rate will have higher impact in reducing the side distortion itself than in reduction the central distortion.

As shown in Fig. 2.4, although the central distortion is independent from redundancy, since it depends only of r , the side distortion is reduced with redundancy insertion. Therefore, practical *MDC* redundancy allocation schemes are needed in order to adapt the coding parameters to the channel conditions, and consequently, reduce the overall distortion.

The rate-distortion bounds for non-Gaussian sources are not completely known. Nevertheless, an extension of Shannon rate-distortion bounds was proposed in [Zamir 1999] in order to define inner and outer bounds for MD coding of any continuous-valued memoryless source with squared-error distortion. The authors have shown that codebooks of Gaussian sources provide an efficient mechanism to encode non-gaussian sources.

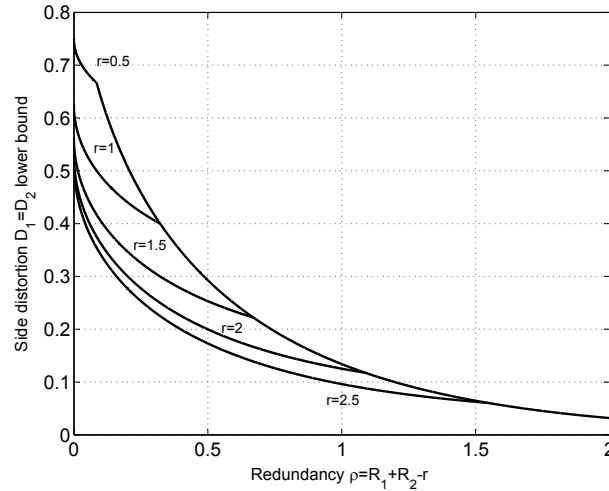


Figure 2.4: Lower Bounds on side distortions $D_1 = D_2$ when the base rate is r and fixed distortion $D_0 = 2^{-2r}$ [Goyal 2001b].

2.2 Practical Implementation

The theoretical foundations described in section 2.1.2 do not answer the question about how to generate two or more redundant and mutually refinable descriptions, but only establish the optimal achievable rates for given joint central and side distortions.

The first practical approach to generate two descriptions was to split the source in odd and even samples, each one independently encoded. The distortion performance of the side decoder at high rates is short due to the interpolation error, but at low rates, its performance is competitive comparing with the central decoder [Goyal 2001a]. Nevertheless, such scheme adds implicit redundancy due to loss of coding efficiency but does not allow to dynamically adapt the redundancy level of MDC.

Other techniques were proposed in the literature, in order to improve the MDC features for source delivery with path diversity in error prone transmission scenarios. Such techniques need to solve several theoretical questions. On the other hand, the description must be independently decoded and needs to be refinable, which means that by joining both descriptions, the overall quality must be improved. Still, the redundancy among descriptions must be dynamically controllable, in order to obtain the best MD rate-distortion achievable points. Thus, the design of MD coders requires some structured approach for trading off source coding efficiency and MDC objectives.

2.2.1 Multiple Description with Correlating Transforms

Most of the source coding techniques are based on transforms, which exploit the redundancy between samples. Transforms concentrate the energy of the source signal in a set of few coefficients, which is one of the most important properties used in efficient source coding. Each coefficient carries its own information which means that each one is not correlated with the others. A good transform should yield an optimal energy concentration which is achieved when the coefficients are totally uncorrelated.

However, the objective of coding efficiency through the use of transforms has a significant conflict with MDC, which does not contribute to increase MDC performance. In fact, in case of transmission errors/losses, uncorrelated coefficients are extremely inefficient in the recovering process of the missing information. This is because statistical dependencies among transform coefficients are necessary for loss concealment, *i.e.*, to estimate the lost coefficients in missing descriptions at the encoder.

MDCT is based on the principle that correlation between coefficients should be controlled for providing the necessary tradeoff between source coding efficiency and MDC goals. For this purpose, this technique uses a linear transform which inserts correlation between coefficients explicitly adding redundancy to each description. The MDCT method was introduced by *Wang et. al.* [Orchard 1997] [Wang 2001] and then the concept was generalized in [Goyal 2001b] for N descriptions.

Generally MDCT follows equation (2.50), where Transform $\hat{\mathbf{T}}$ takes two independent input variables A and B , and outputs two transform variables C and D . $\hat{\mathbf{T}}$ is a 2×2 matrix.

$$\begin{bmatrix} C \\ D \end{bmatrix} = \hat{\mathbf{T}} \begin{bmatrix} A \\ B \end{bmatrix}. \quad (2.50)$$

The transform $\hat{\mathbf{T}}$ controls the correlation between C and D , which in turn gives a measure of the redundancy introduced by this MDCT scheme and at the same time, preserves decorrelation among variables within each description.

The MDCT generic scheme is represented in Fig. 2.5. Two independent Gaussian random independent variables (A,B) with variances σ_A^2 and σ_B^2 are considered. A linear transform

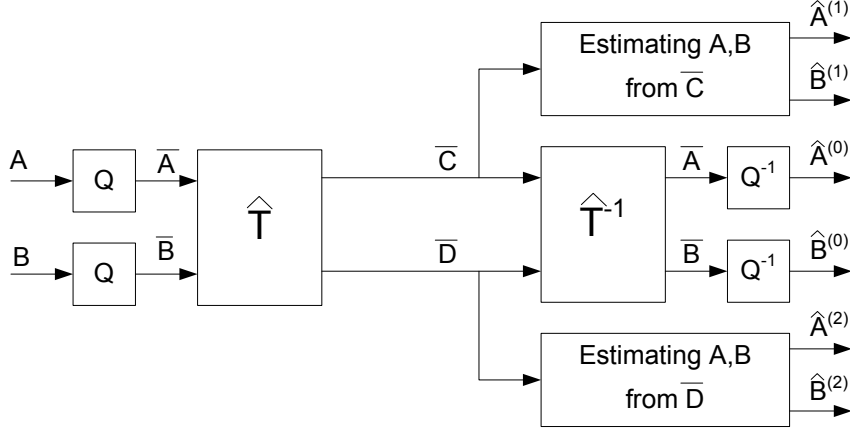


Figure 2.5: MDCT Generic Scheme.

$\hat{\mathbf{T}}$ is applied to the quantized indices (\bar{A}, \bar{B}) in order to obtain the correlated indices (\bar{C}, \bar{D}) . The (\bar{C}, \bar{D}) indices are entropy encoded in independent streams, *i.e.* descriptions.

At the decoder side, if both descriptions are received, then the reconstructed pair (\hat{A}^0, \hat{B}^0) is obtained by using the inverse transform $\hat{\mathbf{T}}^{-1}$. If only one description is received, then the (\hat{A}, \hat{B}) pair is estimated from the received coefficients by exploiting the correlation added at encoder. If only one transform coefficient is available, then it is possible to predict the original missing one. For instance, if \bar{D} is missing, the estimate \hat{D} from the available one \bar{C} is given by

$$\hat{D}(\bar{C}) = \gamma_{D\bar{C}}\bar{C}, \quad (2.51)$$

where $\gamma_{D\bar{C}}$ depends on the variances of C and D and also on the quantization error energy of the two channel encoder. Similarly, if \bar{C} is missing

$$\hat{C}(\bar{D}) = \gamma_{C\bar{D}}\bar{D}, \quad (2.52)$$

and the reconstructed values $\hat{A}^{(i)}, \hat{B}^{(i)}, i = 1, 2$ are obtained from inverse transform

$$\begin{bmatrix} \hat{A}^{(1)} \\ \hat{B}^{(1)} \end{bmatrix} = \mathbf{T}^{-1} \begin{bmatrix} \bar{C} \\ \hat{D} \end{bmatrix}, \quad (2.53)$$

and,

$$\begin{bmatrix} \hat{A}^{(2)} \\ \hat{B}^{(2)} \end{bmatrix} = \mathbf{T}^{-1} \begin{bmatrix} \hat{C} \\ \bar{D} \end{bmatrix}. \quad (2.54)$$

An arbitrary transform $\hat{\mathbf{T}}$ with determinant one is defined as

$$\hat{\mathbf{T}} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad (2.55)$$

where $\hat{\mathbf{T}}$ can be either orthogonal or nonorthogonal. If the transform is nonorthogonal, the quantization in (C, D) transform is not geometrically coincident with quantization in (A, B) domain, increasing the quantization error. In order to overcome this problem the transform is not applied to variables (A, B) but to the integer indices (\bar{A}, \bar{B}) . This is done using a discrete version of the transform in order to yield (\bar{C}, \bar{D}) from (\bar{A}, \bar{B}) . The discretized transform should be designed to allow reversible integer-to-integer mapping, preserving the quantization cells in both domains [Li 1998]. $\hat{\mathbf{T}}$ must have determinant one in order to define a one-to-one correspondence between the index pairs (\bar{A}, \bar{B}) and (\bar{C}, \bar{D}) , thus, maintain the MDC coding efficiency, *i.e.*,

$$\bar{A} = \left\lceil \frac{A}{Q} \right\rceil, \bar{B} = \left\lceil \frac{B}{Q} \right\rceil. \quad (2.56)$$

From the method proposed in [Li 1998] based on LU decomposition, the output pair (\bar{C}, \bar{D}) is determined as *i.e.*,

$$\bar{D} = [dZ] - \bar{A}, \quad (2.57)$$

$$\bar{C} = Z - \left\lceil \frac{1-b}{d} \bar{D} \right\rceil, \quad (2.58)$$

with

$$Z = \bar{B} + \left\lceil \frac{1+c}{d} \bar{A} \right\rceil, \quad (2.59)$$

where $[\cdot]$ is the rounding operation. If (\bar{C}, \bar{D}) values are received, then the corresponding inverse transform is implemented as,

$$\bar{A} = [dZ] - \bar{D}, \quad (2.60)$$

$$\bar{B} = Z - \left\lceil \frac{1+c}{d} \bar{A} \right\rceil, \quad (2.61)$$

with

$$Z = \bar{C} + \left\lceil \frac{1-b}{d} \bar{D} \right\rceil, \quad (2.62)$$

and

$$\hat{A} = \bar{A}Q, \hat{B} = \bar{B}Q. \quad (2.63)$$

The general transform $\hat{\mathbf{T}}$ can be parameterized as,

$$\hat{\mathbf{T}} = \begin{bmatrix} r_2 \cos \theta_2 & -r_2 \sin \theta_2 \\ -r_1 \cos \theta_1 & r_1 \sin \theta_1 \end{bmatrix} = [\mathbf{v}_1 \quad \mathbf{v}_2] \quad (2.64)$$

The transform coefficient is the representation of the original variables using the basis vectors $[\mathbf{v}_1 \mathbf{v}_2]$. The transform parameters are defined such that the determinant of $\hat{\mathbf{T}}$ is one. For that, $r_1 r_2 = 1/\sin(\Delta\theta)$ where $\Delta\theta = \theta_1 - \theta_2$, and $0 < \Delta\theta < \pi$. Considering these parameters, for an output pair (C, D) with variances σ_C and σ_D , the correlation between both of them is given by angle ϕ where,

$$\sigma_C^2 \sigma_D^2 \sin^2 \phi = \sigma_A^2 \sigma_B^2 \quad (2.65)$$

Variables C and D are correlated and more bits are required to represent them in comparison with A and B . Using the rate-distortion functions for Gaussian variables, each pair of rates is given as

$$R = \frac{1}{2} \log_2 \frac{\sigma_C \sigma_D}{D_0} + K, \quad R^* = \frac{1}{2} \log_2 \frac{\sigma_A \sigma_B}{D_0} + K, \quad (2.66)$$

for some constant K . For Gaussian sources and entropy encoding, $K = (1/2) \log_2(\pi e/6)$. Using (2.67) and (2.66) the redundancy $\rho = R - R^*$ is given by

$$\rho = -\frac{1}{2} \log_2 \sin(\phi). \quad (2.67)$$

The optimal transform is the one that minimizes the *side distortion* for a given redundancy level ρ . The definition of the optimal $\hat{\mathbf{T}}$ was presented in [Wang 2001], as a non-orthogonal transform outperforming the one presented in [Orchard 1997]. On the other hand, the same authors have shown that, for orthogonal transforms, the quantization error is small if quantization is applied to the output transform. The orthogonal transforms have the disadvantage of showing smaller range of redundancies in comparison with nonorthogonal ones.

Practical implementations of *MDCT* are found in image and video coding schemes based on block partitioning using transform coding, such as *DCT*. *MDCT* is only used for pairs of variables and optimization algorithms are needed to find the best pair combinations. The optimization criterion is the one that minimizes the MD rate-distortion region, represented as the MDC redundancy and side distortions. In [Wang 2001] an optimization scheme is proposed based on three aspects: 1) optimal redundancy allocation assuming that M pairs are used; 2) optimal pairing strategy and finally; 3) selection of the coefficients to be transformed considering the quantization effect, by pairing only coefficients having a large enough variance.

2.2.2 Multiple Description Sub-band Coding

MDC based on discrete wavelet transform (DWT) or sub-band decomposition was firstly proposed by Yang *et al.* [Yang 2000] and Dragotti *et al.* [Dragotti 2002] to be used on image and video coding. The proposed scheme is shown in Fig. 2.6 for the case of two bands.

The input signal $x[n]$ is passed through low-pass and high-pass filters, with frequency responses $H_0(\omega)$ and $H_1(\omega)$, respectively, then each resulting component are sub-sampled to the new sampling rate thus yielding the sub-band signals $y_1[n]$ and $y_2[n]$. Following this step, each sub-band signal $y_1[n]$ and $y_2[n]$ are separately encoded and transmitted through different channels. To finally reconstruct the signal, each sub-band is up-sampled to the sampling rate of the input. All up-sampled components are passed through the synthesis filters $G_0(\omega)$ and $G_1(\omega)$. Then, each output component is interpolated and subsequently added all together to form the reconstructed signal \hat{x} . With separable filters, multi-dimensional analysis and synthesis can be carried out with stages of uni-directional filters. For example, an image might be first analyzed in its vertical direction followed by the same analysis in the horizontal direction. Missing samples are recovered using linear estimation based on received ones.

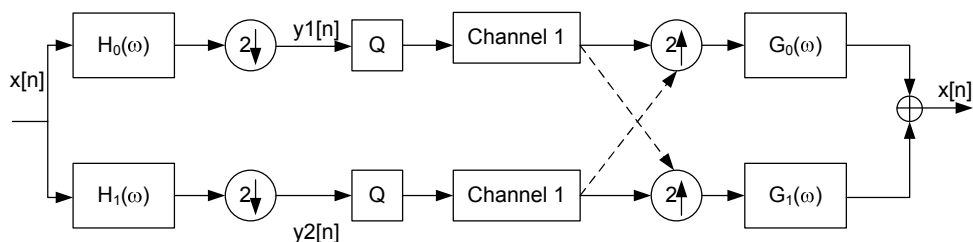


Figure 2.6: Multiple Description Subband Coding.

Sub-band filters are designed to optimize the rate-distortion performance of the MDC scheme. The optimization intends to finding the sub-band filters frequency response for a given redundancy level. The variation of the optimization parameters provides the ability to achieve all points on the theoretical redundancy rate-distortion curve, and associate the optimal filter banks with all redundancy rates. At zero redundancy, the optimization problem becomes the usual coding gain optimization, while for the maximum redundancy solutions, the two subband signals are maximally correlated and have the same stochastic properties. In practice, Multiple Description Sub-band Coding needs one filter bank for each redundancy level, which limits the use of this technique in applications where MDC parameters have to be adapted in function of coding and transmission conditions.

2.2.3 Polyphase Decomposition and Selective Quantization

A particular case of the Multiple Description Sub-band Coding described in section 2.2.2 was proposed by Ortega in [W. Jian 1999] by using the simplest polyphase decomposition filters [Vaidyanathan 1990], which consists in partitioning input samples into even and odd numbered subsets. Such scheme also adds redundancy by coding the original source sequence using different quality resolutions as redundant information. Fig. 2.7 shows the proposed MDC scheme, where the source is decomposed into two sub-sources via polyphase transform.

The source $x[n]$ is split in its polyphase components $y_1[n]$ and $y_2[n]$. Then each component is quantized with different quantization step-sizes, forming two types of information: primary, resultant of a finer quantization, and secondary, from a coarse quantization.

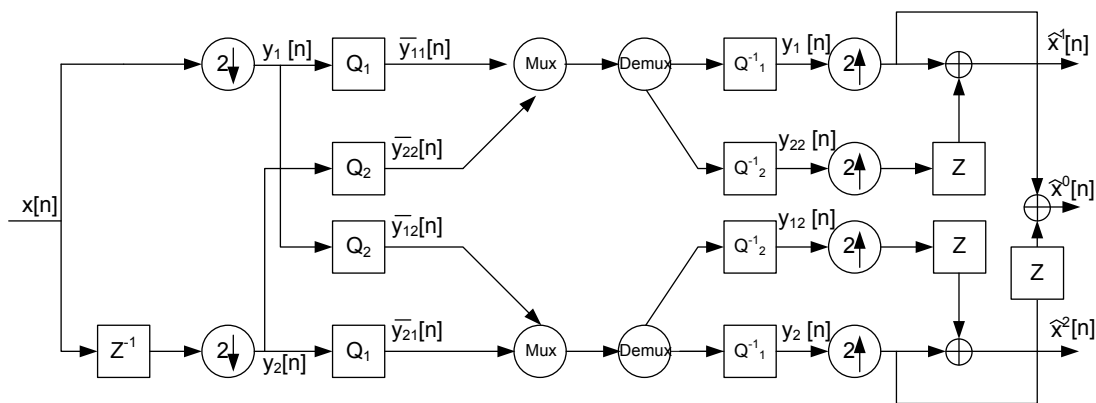


Figure 2.7: MDC with polyphase transform and selective quantisation.

Therefore, the primary information, $\bar{y}_{1,1}[n]$ and $\bar{y}_{2,1}[n]$ of each partition $y_1[n]$ and $y_2[n]$, is independently quantized with a finer quantization parameter Q_1 . The secondary and totally redundant information is formed by a coarser quantized version of each partition with Q_2 , $\bar{y}_{1,2}[n]$ and $\bar{y}_{2,2}[n]$. Primary information of each polyphase component are multiplexed with the complementary secondary partition, generating two descriptions, one with $\bar{y}_{1,1}[n]$ and $\bar{y}_{2,2}[n]$ and the other with $\bar{y}_{2,1}[n]$ and $\bar{y}_{1,2}[n]$.

At the decoder, each description is de-multiplexed, and its components are de-quantized and interpolated. If both descriptions are available, the primary information $\bar{y}_1[n]$ and $\bar{y}_2[n]$ are combined in order to obtain $\hat{x}^0[n]$. The secondary components are fully redundant (they do not contribute with new information to decoding). When only one description is received the, primary and secondary components of the same description are combined, obtaining $\hat{x}^1[n]$ and $\hat{x}^2[n]$.

This MDC scheme was applied in [W. Jian 1999] on image coding using the Said-Pearlman wavelet coder [Said 1996]. The input image is first wavelet transformed and its polyphase components were extracted. Two different types of polyphase transforms were used. The first is obtained by simply group all the even coefficients into one description and all the odd coefficients into the other description. The other polyphase method is to use the zerotree structures separating the even blocks to one description and all the odd blocks to the other description. Such approach have inspired many proposals applied to image and video coding due to his flexibility, as will be described in chapter 3.

2.2.4 Multiple Description Scalar Quantisation

Multiple Description Scalar Quantization (MDSQ) was proposed by Vaishampayan in [Vaishampayan 1993]. MDSQ is usually designed to generate two independent descriptions from the same source signal and it is based on two operations: *scalar quantization* and *index assignment*. A generic MDSQ scheme is shown in Fig. 2.8.

A source x is scalar quantized, and for each index \bar{x} , an index pair (\bar{y}_1, \bar{y}_2) is assigned, each one encoded using an entropy encoder. At the decoder, three types of mapping are used with three different reconstructions of x . When (\bar{y}_1, \bar{y}_2) are available, the output of the mapping function $(\bar{y}_1, \bar{y}_2) \rightarrow \bar{x}$, is inverse quantized and the central reconstruction $\hat{x}^{(0)}$ is obtained. When only \bar{y}_1 or \bar{y}_2 reach the decoder, a mapping $\bar{y}_i \rightarrow \bar{x}^{(i)}, i = 1, 2$ is used, and side reconstruction signals $\hat{x}^{(i)}, i = 1, 2$ are obtained after inverse quantization

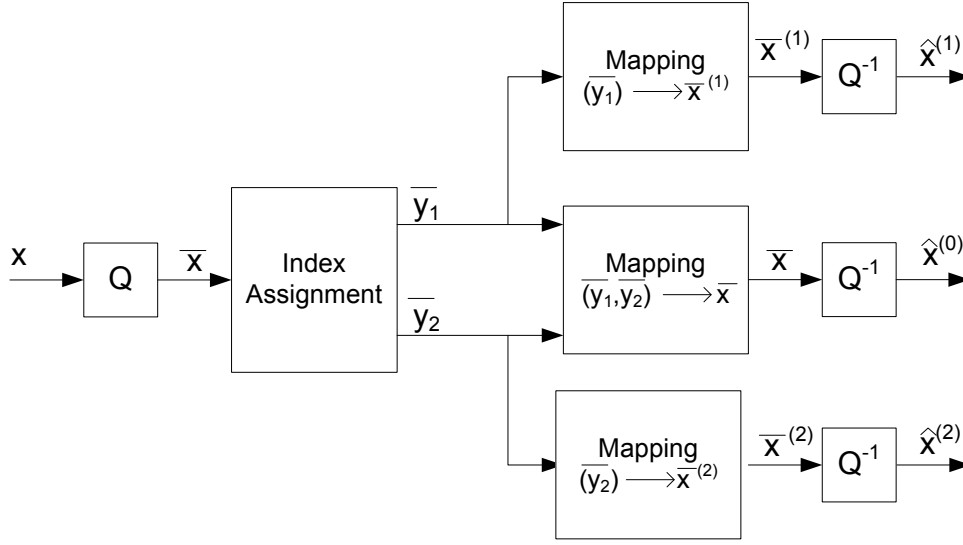


Figure 2.8: Generic MDSQ scheme.

of $\bar{x}^{(i)}$.

Using SD coding as reference, in a M-level scalar quantization, each source sample x is mapped to a reconstruction level \hat{x} , which takes values from the codebook $\hat{\chi} = \{x_1, x_2, \dots, x_M\}$. The M-level scalar quantization is usually regarded as composition of an encoder map $f : \mathbb{R} \rightarrow \{1, 2, \dots, M\}$ whose output is a codeword index and a decoder map $f : \{1, 2, \dots, M\} \rightarrow \mathbb{R}$. The encoder partitions \mathbb{R} in M cells. Therefore, the SDC M-level scalar quantization is completely described by its partition and its codebook. Partition is represented by $A = \{A_1, A_2, \dots, A_M\}$ where $A_i = \{x : f(x) = i\}$, $i = 1, 2, \dots, M$.

The main principle of MDSQ is to construct two uniform quantisers with coarse reconstruction values such that separate inverse quantization is possible. Extending the SDC M-level scalar quantization to MDC, an $(M1, M2)$ -level MDSQ maps the source sample x to reconstruction levels $\hat{x}^{(0)}$, $\hat{x}^{(1)}$ and $\hat{x}^{(2)}$ that take values in the codebook $\chi^0 = \{\hat{x}_{\bar{y}_1, \bar{y}_2}, (\bar{y}_1, \bar{y}_2) \in C\}$, $\chi^1 = \{\hat{x}_{\bar{y}_1}^1, \bar{y}_1 \in I_1\}$ and $\chi^2 = \{\hat{x}_{\bar{y}_2}^2, \bar{y}_2 \in I_2\}$ and C is a subset of $I_1 \times I_2$. An MDSQ scheme can be broken up into two side encoders, $f_1 : \mathbb{R} \rightarrow I_1$ and $f_2 : \mathbb{R} \rightarrow I_2$, which select the indexes \bar{y}_1 and \bar{y}_2 , respectively, and three decoders, $g_0 : C \rightarrow \mathbb{R}$ (central decoder) and $g_1 : I_1 \rightarrow \mathbb{R}$, $g_2 : I_2 \rightarrow \mathbb{R}$ (side decoders), whose outputs are the reconstruction levels with indexes $\bar{y}_1 \bar{y}_2$, \bar{y}_1 , and \bar{y}_2 from codebooks χ^0 , χ^1 and χ^2 respectively. The two encoders impose a partition $A = \{A_{\bar{y}_1, \bar{y}_2}, (\bar{y}_1, \bar{y}_2) \in C\}$ on C , where $A_{\bar{y}_1, \bar{y}_2} = \{x : f_1(x) = \bar{y}_1, f_2(x) = \bar{y}_2\}$. Thus MDSQ is completely described by A , χ^0 , χ^1 and χ^2 .

MDSQ formulation is illustrated in Fig. 2.9 and Fig. 2.10, using two distinct forms: matrix and linear representation respectively. The source x is quantized according to the mapping function $Q_0 : \mathbb{R} \rightarrow 1, \dots, 11$, representing the quantization function of the central decoder, with quantization stepsize Q_0 , which means that are used $M = 11$ quantization cells. For each description, $M = 6$ and the MDSQ encoder output is $Q_i : \mathbb{R} \rightarrow 1, \dots, 6, i = 1, 2$, representing the quantization function of the side decoders with quantization stepsize $Q_i, i = 1, 2$. In matrix representation, the inside values are the quantization indices \bar{x} , and each one have a correspondent index pair, (\bar{y}_1, \bar{y}_2) , the indices of each description. In linear representation, each quantizer has one cell representation. The reconstruction of $Q_i(x) = k_i, i = 0, 1, 2$ is the centroid of the cell $Q_i^{-1}(k_i)$. The central decoder results from the intersection cell $Q_1^{-1}(k_1) \cap Q_2^{-1}(k_2)$ between both side quantisers. The intersection cell is half size of the side quantizer cells, which means that an equivalent quantisation with a finner step size are obtained.

		\bar{y}_2 (Description 2)					
		1	2	3	4	5	6
\bar{y}_1 (Description 1)	1	1	2				
	2		3	4			
	3			5	6		
	4				7	8	
	5					9	10
	6						11

Figure 2.9: MDSQ example 1 - Matrix representation.

\bar{y}_1	1	2	3	4	5	6					
\bar{x}	1	2	3	4	5	6	7	8	9	10	11
\bar{y}_2	1	2	3	4	5	6					

Figure 2.10: MDSQ example 1 - Linear representation.

Differently from single description scalar quantization, the MDSQ side quantizer levels are characterized by having discontinuous partition cells. Figs. 2.11 and 2.12 show an example where the side quantizer cells are not continuous. This can be observed in linear representation (Fig. 2.12) where each side quantizer cells are interleaved between them. Despite, the discontinuity of the quantiser cells of each description, they refine each other in a symmetric way, which means that central decoder have a continuous characteristic with also a finner step size.

		\bar{y}_2 (Description 2)					
		1	2	3	4	5	6
\bar{y}_1 (Description 1)	1	1	2				
	2	3	4	6			
	3		5	7	8		
	4			9	10	12	
	5				11	13	14
	6					15	16

Figure 2.11: MDSQ example 2 - Matrix representation.

\bar{y}_1	1	2	3	2	3	4	5	4	5	6						
\bar{x}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
\bar{y}_2	1	2	1	2	3	4	3	4	5	6	5	6				

Figure 2.12: MDSQ example 2 - Linear representation.

In general, MDC algorithms use balanced descriptions where the respective rates and distortions are approximately the same. Considering two descriptions with rate-distortion pairs (R_1, D_1) and (R_2, D_2) , these two descriptions are generated using side-quantisers from a central quantiser with rate-distortion pair (R_0, D_0) . For balanced MDSQ, $R_1 \approx R_2$ and $D_1 \approx D_2$. Given a certain partition of the equivalent central encoder (the one that would give the reconstruction $\hat{x}^{(0)}$), the main problem is to find side quantizers that produce the optimal combination between central distortion D_0 and side distortions D_1, D_2 .

The first MDSQ implementation, named as level-constrained MDSQ, was proposed in [Vaishampayan 1993] for fixed length binary encoders. In [Vaishampayan 1994] an optimal entropy-constrained MDSQ was proposed by using variable-length coding. The practical implementation is based on the MDSQ matrix representation as shown in Figs. 2.10 and 2.12, which is defined as *index assignment matrix*. The main problem is to design the index assignment function for individual quantizers, where the aim is to find the index assignment scheme that gives optimal combined exponential decay rates for the central and side decoders. Then, the index assignment problem is to find the best scanning sequence for a selected set of index pairs that results in small spread of each cell of side partition, *i.e.*, a small as possible interleaving between quantization cells of side encoders. For a given scanning scheme, the number of central cells is N , $n_1 = n_2$ is the number of index pairs, and $M = n_1 * n_2$. The central and side distortions are adjusted by the index spread k , that represents the number of diagonals above and below to the main

diagonal of index assignment matrix. For instance in Fig. 2.12(a), $N = 16$, $M = 36$ and $k = 1$. Varying k and M , for a given fixed N , means that D_0 does not change, and side distortion D_1 (D_2) increases with higher k . Redundancy is reduced because M is lower. On the other hand, for fixed M and increasing k , the central distortion D_0 is reduced and side distortions $D_1 = D_2$ increase. Therefore, the index spread k is the most important parameter in order to tune both the central and side distortions and, consequently, the redundancy of the MDSQ scheme.

An interesting evaluation of the MDSQ scheme can be done by comparison with the optimal MD rate distortion bounds. The asymptotic analysis of MDSQ was made in [Vaishampayan 1998]. In asymptotic sense, the optimal MD rate distortion bound at large rates for a squared-error distortion measure, and unit-variance memoryless Gaussian source can be approximately given as

$$D_0 D_1 \approx \frac{1}{4} 2^{-2(2R)}. \quad (2.68)$$

The optimum level-constrained MDSQ for a unit-variance Gaussian source is given by

$$D_0 D_1 \approx \frac{2\pi^2}{16} 2^{-2(2R)}. \quad (2.69)$$

Finally, for the entropy-constrained MDSQ the RD bound is given by

$$D_0 D_1 \approx \frac{\pi^2 e^2}{144} 2^{-2(2R)}. \quad (2.70)$$

This results show that for entropy-constrained MDSQ, the RD bound is closer to the optimal level than the level-constrained MDSQ. A simpler MDSQ implementation based on two-stage index assignment with RD performances close to the entropy-constrained MDSQ scheme was proposed in [C. Tian 2004]. A continuum of tradeoff points between the central and side distortions are achieved by controlling quantization parameter of the first stage and redundancy parameter of the second stage. The index assignment is done based on these two steps avoiding extensive training used in entropy-constraint MDSQ on the quantizer thresholds and reconstruction values for a given index assignment matrix.

2.2.5 MDC Based on Error Correction Codes

Most of modern communication networks use Forward Error Correction (FEC) Codes as a channel coding scheme, specially in applications where retransmissions are not available or are not possible due to delay constraints. FEC coding consists in adding redundancy at channel coding level, where the redundant codes provide useful information to recover missing one at decoder. The redundancy level is dependent of the error probability of the channel. An important type of FEC codes are the Reed-Solomon (RS) block coding, which are characterized by the size (N, k) , where N is the length of the block code and k is the length of the source symbols. The $N - k$ symbols are the channel coding redundancy added to the source symbols in order to detect and recover lost information. The main feature of a (N, k) RS code is that, all of the lost symbols can be recovered at the decoder if the number of erasures does not exceed $N - k$.

By extending the FEC properties to the MDC problem, a new MDC scheme is defined as MDC with FEC (MD-FEC), which is suited to be used in MDC systems with $N > 2$ descriptions. The use of such scheme is tailored to benefit multi-resolution or scalable source coding. In fact, the main goal of MD-FEC is to give unequal error protection (UEP) to different coded layers or quality levels, which means that the most important layers or high quality levels are assigned with stronger FEC codes in order to better protect those layers. Fig. 2.13 shows an example of the *MD - FEC* scheme, where the source is divided in N layers or quality levels, $\{1, \dots, N\}$, each one with rates $\{R_1 \dots R_N\}$, respectively.

Fig. 2.14 shows the N -description generation. For each of the N descriptions, description k have $N - k + 1$ source layer blocks. Description 1 contains groups of all layers. In Description 2, the first block is filled with FEC codes, and all the other layers are filled with the corresponding blocks, and so on. As a result of this process, N descriptions are encoded and N RS code blocks are formed. As shown in Fig. 2.14, the k th layer is protected

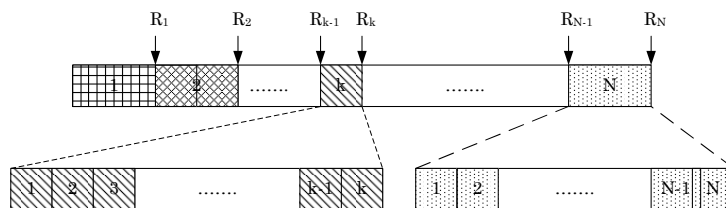


Figure 2.13: MDC FEC - source data partitioning.

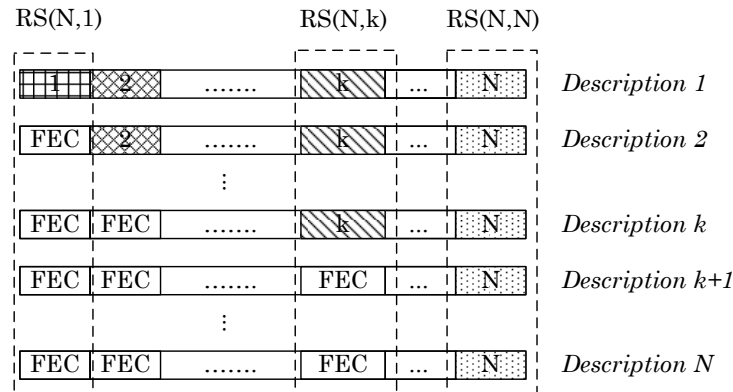


Figure 2.14: MDC FEC - N-description generation.

by a (N, k) RS code, which means that layers $\{1, 2, \dots, N\}$ have $\{(N, 1), (N, 2), (N, N)\}$ RS codes. This structure ensures that the first N source layers are decoded by performing *RS* decoding from any N descriptions.

The MDC redundancy is tuned using the optimal layer partitioning (R_1, R_2, \dots, R_N) . The optimization criteria is the minimization of the overall distortion at the receiver given the total rate constraint, based on loss probability of descriptions K , $K = 1, 2, \dots, N$ [Puri 1999]. Although MD-FEC adds robustness to many layer coding schemes, the delay caused by FEC operations may affect the performance and applicability in systems where such parameter is a critical constraint.

2.3 Conclusion

This chapter presented a review MDC principles, where the relevant theoretical definitions and optimal rate-distortion bounds was explained. The main MDC techniques documented in the literature were described, including MDCT, Multiple Description Subband Coding, based on Polyphase Decomposition and Selective Quantisation, MDSQ, and MDC based on Error Correction Codes. Moreover, these techniques were described for a generic source model. Applications of the MDC techniques to video signals requires definition and implementation of the processing architectures and their corresponding modules. Next chapter addresses MD schemes for video coding and transmissions.

Multiple Description Video Coding and Networking

Contents

3.1	Basic Video Coding Schemes	38
3.2	MD Video Coding - an overview	42
3.3	MD Architectures for Video Encoding	45
3.3.1	Embedded MDC	46
3.3.2	MDC with pre-processing	51
3.3.3	High dimensional MDC	54
3.3.4	Multiple Description Scalable Coding	55
3.3.5	Stereo/Multiview MDC	57
3.3.6	Unbalanced MDC	59
3.3.7	Multiple Description Distributed Video Coding	61
3.4	Multiple Description Video Streaming	62
3.4.1	MDC Video Streaming over Content Delivery Networks	63
3.4.2	MDC Video Streaming over Peer-to-peer networks	65
3.4.3	MDC video streaming over Wireless Ad-hoc Networks	69
3.4.4	Scheduling in path diversity networks	70
3.5	Conclusion	71

This chapter presents a literature review on the most relevant research in the field of Multiple Description (MD) coding and transmission for video signals. The first part is an

overview of Multiple Description Coding (MDC) techniques and architectures proposed in the literature in recent years. MDC video architectures are described for different video coding schemes, based on techniques explained in Chapter 2. These are applied in different video coding approaches such as, advanced video coding, scalable video coding, distributed video coding, multiview and stereo video coding, for several coded descriptions. The second part of this chapter describes MDC video applications in networks with path diversity scenarios. MD video delivery network topologies, architectures and applications are described for content delivery networks and peer-to-peer networks.

3.1 Basic Video Coding Schemes

The main goal of video encoding schemes is to compress the amount of data used to represent visual information, controlling the coding distortion introduced in the original signals. Video coding and compression schemes exploit two main principles: *redundancy* and *irrelevancy* of video signal.

Redundancy is related with the source sample information that is correlated with other neighbor samples. In video sequences, the redundancy exists in both the spatial and temporal dimensions. In the spatial dimension, this is due to the fact that pixels within the same frame are highly correlated. In the temporal dimension, the use of high frame rates, capturing several images per second results in consecutive frames containing very similar information, since temporal evolution of signal contents is generally smooth. In addition, the coded symbols are correlated, which means that entropy redundancy is also exploited in order to compress source information. The characteristic of irrelevancy is directly related to the Human Visual System (HVS), which acts as a filter in regard to some spatio-temporal frequency components. The principle behind lossy coding schemes is to remove the irrelevant components from the video signals, in order to reduce the amount of information to be coded without introducing noticeable distortion.

Therefore, a variety of coding tools to efficiently compress visual information are used, which includes lossless and lossy schemes. These are based on Differential Pulse Coding Modulation (DPCM) schemes where the rate-distortion performance is mainly dependent on the prediction functions, that are responsible to minimize the residue energy of a given sample or block of samples. Associated to the best prediction, most of the video coding architectures are based on transform coding and quantization (in lossy schemes) that

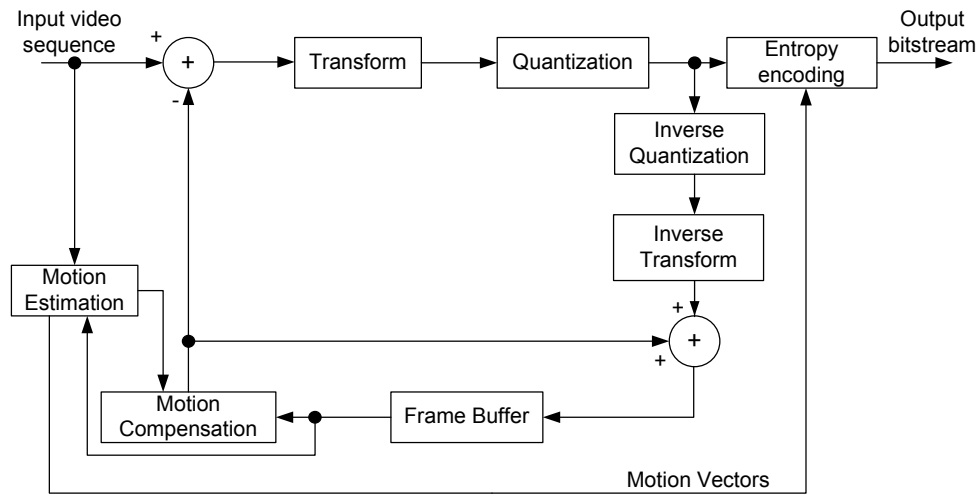


Figure 3.1: Hybrid video encoder.

allows to concentrate the residue energy in few coefficients. Finally, source symbols are entropy encoded, ideally, close to the source entropy. Thus, video coding algorithms are grouped in three main approaches: 1) DCT coding, which is used in modern video coding standards; 2) 3D-subband coding, based on sub-band decomposition in temporal and spatial domains; 3) Distributed Video Coding, which transfer the computational complexity to the decoder, through simpler encoders and using side information to generate decoded video sequences.

DCT based video coding Fig. 3.1 shows the common usual DCT based video coding scheme which consists in a feedback loop of a block-based *motion estimation* (ME) and *motion compensation* (MC) to reduce redundancy followed by a transform (usually DCT) of the prediction error. The transformed residue is then quantized and entropy encoded. One of the main blocks is ME, which finds the motion information between the current frame and the reference ones. This is essential to reduce the error energy of the motion compensated prediction, and consequently, to improve coding efficiency. Irrelevance is exploited by quantization which is characterized by the quantization step-size that maps each transform coefficient into a corresponding index. The quantization step-size is directly related to the distortion of decoded signal. Headers, motion information represented by *motion vectors* and prediction modes are entropy encoded. This encoding scheme is used in standard video codecs, such as, MPEG-2 [ISO/IEC13818-2 2000] and H.264/AVC [ISO/IEC14496-10 2012].

The DCT-based video architecture is also used in scalable video coding tools, which con-

sists in encoding video sequences in several dependent layers. This means that higher layers need lower layers to be decoded. Different layers correspond to different resolutions or quality that can be obtained by partial decoding of the scalable stream. In general, full resolution is obtained when all layers are decoded. Scalability can be classified in tree types: temporal, spatial and quality. In temporal scalability each layer corresponds to different temporal resolutions that are combined to reach the full resolution. In spatial scalability the spatial resolution of the original sequences is down-sampled in order to obtain several resolutions. The quality scalability is based on the refinement of the transform coefficients in order to obtain several quality levels at the decoder.

Sub-band video coding Sub-band video coding extends the spatial sub-band decomposition based on filter banks Discrete Wavelet Transform (DWT) used on image coding to the temporal dimension. Although good performance of DWT for image coding, which has been adopted in JPEG2000 standard [15444-1 2002], sub-band video coding has not been successful for current state of the art video coding standards. Nevertheless, sub-band video coding is inherently scalable which means it is still an interesting video coding paradigm for video coding applications in heterogeneous environments.

Fig. 3.2, shows the block diagram of sub-band video coding. The main functional block is the 3D-analysis, to the sequence in temporal and spatial frequency sub-bands. This is done using a filter banks, composed by temporal filtering in high and low frequencies and, spatial *DWT*, that divides spatial information in frequency sub-bands. Finally, each sub-band coefficient is quantized. The output bitstream results from efficient entropy coding of the temporal and spatial sub-bands. 3-D analysis filtering can be made without motion compensation [Said 1996] or with motion compensation [Ohm 1994, Kim 1997, Choi 1999], which is the Motion Compensated Temporal Filtering (MCTF) video encoder scheme.

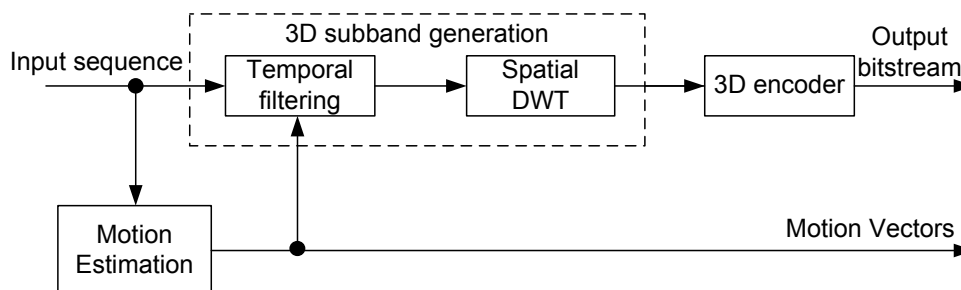


Figure 3.2: 3-D sub-band video encoding.

Fig. 3.3 illustrates an overall 3-D analysis example. The sub-band decomposition is

based on a hierarchical tree-structured scheme. At the first stage, the input video is decomposed of temporal L- and H-sub-bands. Then, the L-sub-band is decomposed again into temporal LL and LH-sub-bands at the second stage. Thus, the resultant four frame unit is composed of temporal decompositions, t-LH, and two t-H sub-bands. Finally, a spatial wavelet transform is applied on each temporal sub-band. Image wavelet sub-band coding uses embedded zero-tree wavelet coding (EZW) schemes based on [Shapiro 1992] which exploits the inter sub-band similarities. In video coding the extension to 3D-EZW is used in order to exploit also the temporal sub-bands [Kim 2000].

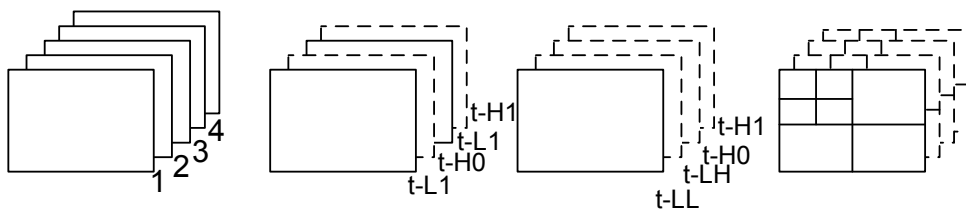


Figure 3.3: 3-D sub-band decomposition.

Distributed Video Coding Hybrid schemes are characterized by complex encoders and simpler decoders. Distributed video coding intends to change this paradigm by transferring the complexity from the encoder to decoder. This is a new approach in video coding paradigms which is based on two theorems from information theory: Slepian-Wolf [Slepian 1973] and Wyner-Ziv theorems [Wyner 1976]. The Slepian-wolf theorem states that two statically dependent signals X and Y can be compressed in a distributed way, i.e. separately encoded and jointly decoded, using approximately the same rate if encoded and decoded together. On the other hand, Wyner-Ziv theorem extends the earlier theorem for lossy encoding and also states that if the correlated signal Y is only available at decoder (the side information at decoder), there is no loss in coding efficiency in X , comparing with the lossy joint encoding of X and Y . These theorems together have provided the basis for the new video coding paradigm that allows to encode two statistically dependent signals independently and decoding them jointly with coding efficiency closer to the predictive encoding schemes.

Fig. 3.4 shows the distributed video coding (DVC) scheme. Two different kind of frames are considered, the Wyner-Ziv and Key frames, which are separately encoded. The Wyner-Ziv frames are intraframe encoded based on DCT transform followed by quantization. Each coefficient are bit-plane encoded and each one is encoded using channel

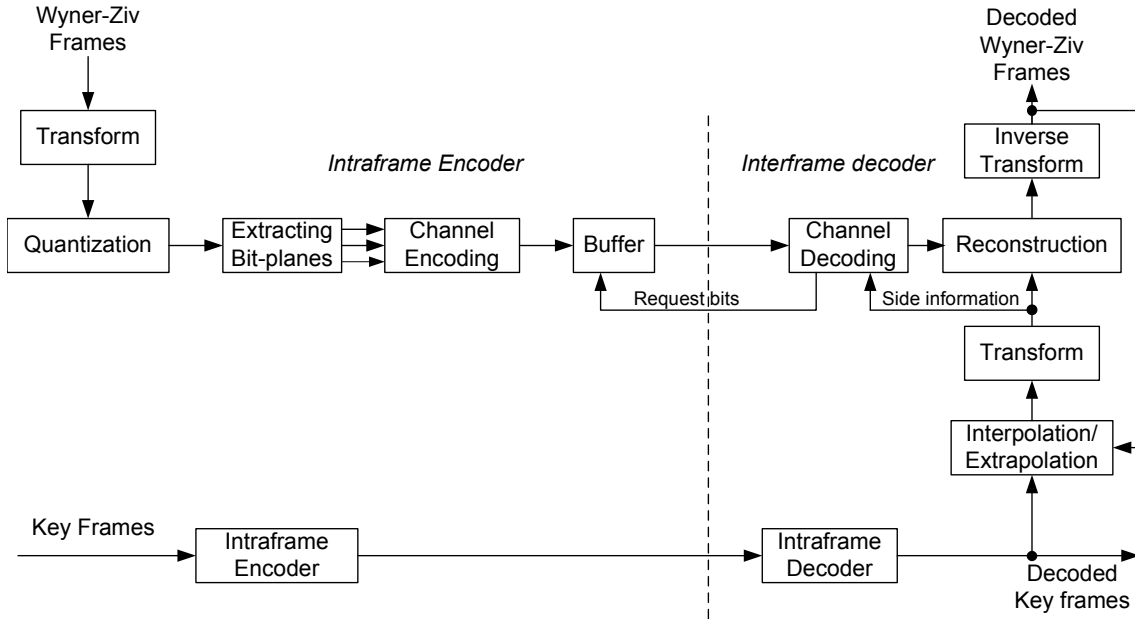


Figure 3.4: Distributed video encoding.

encoding schemes. Key frames are conventionally intraframe encoded and are used as side information to decode the Wyner-Ziv frames. DVC transfer the computational complexity to the decoder, where Wyner-Ziv decoding is done using the key frames as side information. The decoding of Wyner-Ziv frames are based on interpolation/extrapolation of key frames, that combined with encoded bit-planes, reconstruction frames are obtained. The received channel coding information depends of the quality that is needed at decoder, which can interactively request bits to encoder. A review on DVC video coding techniques can be found in [Pereira 2009].

3.2 MD Video Coding - an overview

The video encoding schemes briefly explained in Section 3.1 are the starting point of the development of MD video encoders proposed in the literature by combining Single Description Coding with the MDC techniques referred in Chap. 2. Considering that a MD video encoder outputs M descriptions, these can be generated in different encoding stages: i) pre-processing stage; ii) encoding stage inside the prediction loops; iii) post-processing stages. This defines three types of MDC classes: pre-processing MDC, embedded MDC and post-processing MDC respectively.

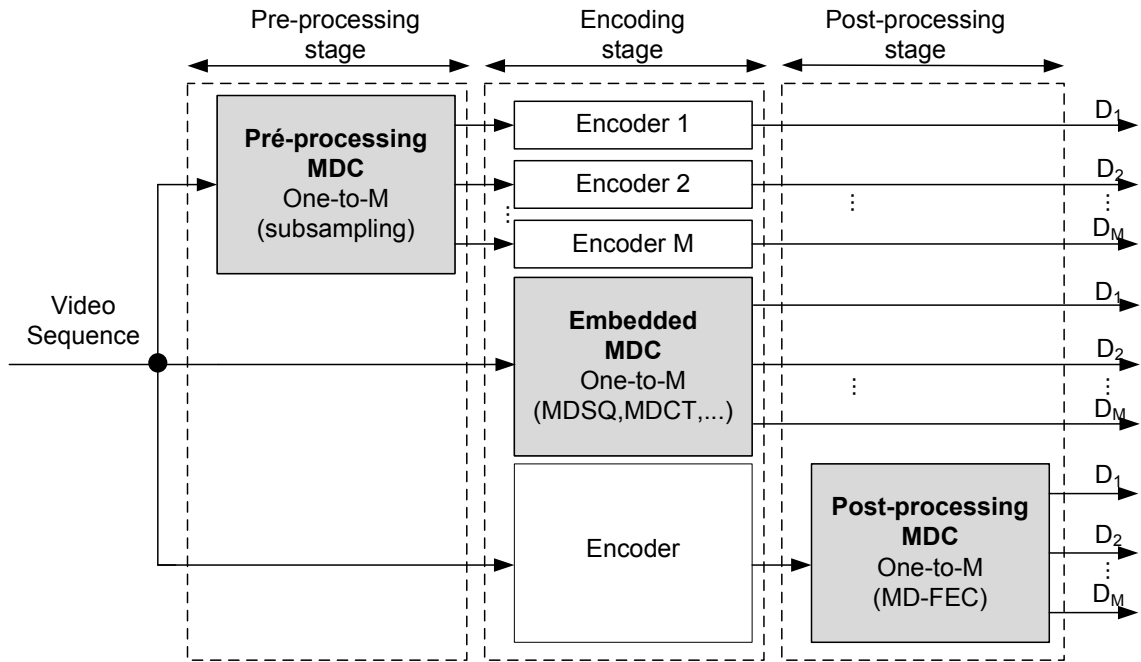


Figure 3.5: Generic MD encoder

Fig. 3.5 shows a generic framework where the different stages can be combined to achieve MD video encoding. In the pre-processing MDC, the video sequences are pre-processed in order to form M descriptions. The input video sequence is first sub-sampled into different representations of the source signal and then, each one is independently encoded. Down-sampling is done in either the spatial or temporal domains. In embedded MDC encoders the M descriptions are generated at the video encoder prediction loop. This category includes MDC based on $MDCT$, $MDSQ$ and down-sampling in frequency domain. Finally, in post-processing MDC, M descriptions are generated to form the output bitstream. MD-FEC techniques are included in this group of MDC encoders.

As an error-resilience technique, the main objective of MD video coding is to improve the robustness of the source video affected by transmission errors, where a controlled amount of redundant information is added in each description. Consequently, the resultant output rates are defined for each description which can be encoded with the same rates, *i.e.* balanced MDC, or to be encoded with different rates, *i.e.*, unbalanced MDC. Usually, MDC techniques use balanced schemes since each description has the same importance, considering that they are transmitted over independent channels. On the other hand, unbalanced descriptions are generated in order to adapt their corresponding rates to constrained channel conditions, such as available rate and packet loss rates. This is done by

adding explicit redundant information, which depends on the robustness of each description based on adequate rate-distortion optimization criteria for error prone environments.

Considering multiple descriptions combined with specific video coding tools, the redundancy components of MD video encoders have many sources, and depend on the specific MD scheme that is applied. Redundancy components include: (i) redundancy introduced by MD prediction error encoder for each description (ρ_{md}); (ii) redundancy introduced by using less efficient prediction compared with SDC (ρ_{ef}); (iii) additional explicit signaling for drift reducing (ρ_{dr}) and finally, (iv) additional excess side information not used in SD encoding, for instance by duplicating motion or header information (ρ_{si}).

$$\rho = \rho_{md} + \rho_{ef} + \rho_{dr} + \rho_{si}. \quad (3.1)$$

The overall rate allocation includes the redundancy, that controls two fundamental characteristics of the MDC scheme: (i) the side distortion (when only one description is decoded) and (ii) the central distortion when all descriptions are decoded in function on packet loss rates and available bit-rates for each channel.

The MD decoder must be able to decode each description individually, producing reconstructions with an acceptable quality, and also to combine both of them forming a higher quality reconstruction. The decoding process includes the use of the redundant information that is used only when some description fails. Additionally, it can include error concealment methods based on interpolation based on the available information. Depending on the MDC scheme, different interpolation techniques are used. In pre-processing MDC, interpolation has to deal with the fail of temporal or spatial signal components. In this case, reconstruction is based on the received information of the same description and also, from the other descriptions. On the other hand, in embedded MDC, usually the spatial and temporal resolutions are not changed, which means that interpolation schemes are applied only when all descriptions have failed, which is an advantage of embedded schemes compared with pre-processing MDC techniques.

In addition, decoders have to deal with prediction mismatch between encoder and decoder and the consequent error propagation which needs to be controlled or mitigated. Due to the predictive nature of most of video coding algorithms, the MD video coding schemes have to deal with the problem of drift introduced when some description does not reach the decoder, which originates mismatch between the decoder and encoder reference frames.

As the main drawback of MD video coding, the drift control is one of the most important issues that MD video coders should solve and also one of the main performance parameter used to evaluate MD video coding schemes. Drift control can be done by introducing different prediction loops at the encoder to prevent all the possibilities of missing descriptions at decoder. Obviously, this approach is only practicable if a reduced number of descriptions is used, because MD video coding architecture complexity increases with the number of descriptions. Alternatively, other less efficient prediction schemes in rate-distortion sense but appropriate in dealing with error propagation can be used, for instance, intra refresh. However, these schemes cannot be sufficient to avoid drift in environments with high packet loss rates. In conclusion, the encoding process has to provide this ability without knowing the available descriptions at each decoding time, which means that decoder must have the sufficient tools in order prevent error drift propagation when transmission errors occur.

3.3 MD Architectures for Video Encoding

This review presented in this section is based on SDC encoding, defining a classification according to the SDC scheme that is in the genesis of the corresponding MDC architecture. The MDC video coding schemes are classified as: (a)MD embedded video encoding; (b)MDC with pre-processing; (c)High Dimensional MDC; (d)Scalable MDC; (e)Multiple Description Distributed Video Coding (MDDV); (f)Stereo/Multiview MDC.

MD embedded video encoding schemes comprise the techniques where the MD functional block is embedded in the encoder prediction loop. These schemes include MDSQ, MDCT and splitting of either transform coefficients or motion vectors. MDC with pre-processing includes all the schemes that use spatio-temporal sub-sampling before encoding of each description, where redundant information can be explicitly inserted improving the interpolation process at the decoder. High dimension MDC includes the schemes with $N > 2$ descriptions. Although these MDC schemes are mainly based on sub-sampling schemes, an independent classification is presented. Scalable MDC includes all MDC schemes based on 3D-subband and DCT based scalable video coding. Novel video coding techniques such as DVC, Stereo and multiview video coding have also been used to MDC.

3.3.1 Embedded MDC

In video encoding with embedded MDC schemes, the MD module is embedded into the prediction loop. These MD schemes include scalar quantisation (MDSQ), correlating transforms (MDTC), splitting of DCT coefficients, sub-sampling of residual and motion information. Embedded MDC schemes are further classified as either open-loop architectures, where descriptions are generated independently from the prediction loop or three-loop architectures that encode the residual difference to reduce or even eliminate the drift at decoder.

Open-loop architectures Fig. 3.6 shows the generic open-loop MD video encoder using two descriptions. The MD encoder is embedded in the prediction loop generating two output independent streams r_1 and r_2 . An MD decoder is used in the loop that merges both descriptions, which means that reference frames used for prediction are resulting from *SDC* decoding. Thus, only with both descriptions the prediction loop at decoder is the same as correspondent encoder.

Several open-loop embedded MD video encoders have been proposed, where the prediction mismatch between encoded and decoder is not controlled. These include splitting of transform coefficients wherein the output of a standard video encoder is split into two correlated streams based on rate-distortion criteria. In such schemes, the redundancy is relatively high because header information, motion vectors and DCT coefficients with magnitudes above a certain threshold are duplicated in the two descriptions, whereas the remaining coefficients are evenly distributed among the two streams. At the decoder,

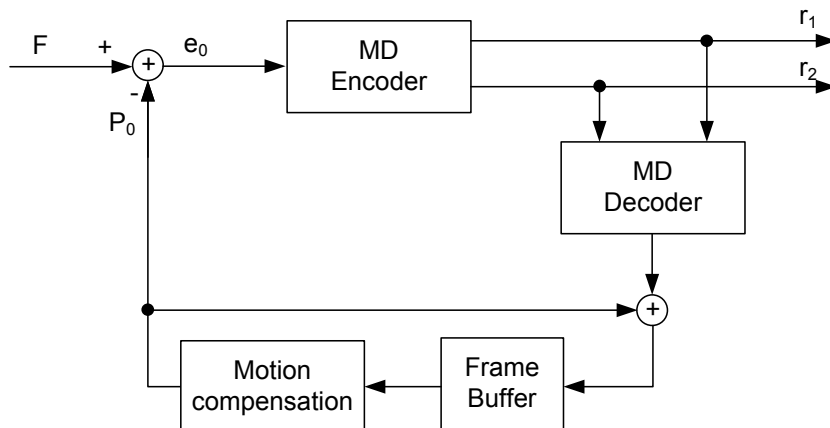


Figure 3.6: Open-loop multiple description video encoder.

if both descriptions are received, then the redundant information is discarded [Reibman 2001] [Matty 2005]. An open-loop architecture based on MDSQ has been proposed in [Vaishampayan 1999] for H.263. In [Campana 2008] an open-loop MDSQ for H.264/AVC has been proposed. In such architecture, drift control is not supported which results in distortion accumulation along the GOP, whenever any description is lost in the transmission network.

Usually, motion information is duplicated in both descriptions, considering that such information has not an expensive cost in the total redundancy, which might not be exactly truth in low rate application environments. Fig.3.7 shows the MD motion coding method proposed in [Kim 2001], where the original motion vector field is sub-sampled using quincunx lattices to form two different descriptions. In both descriptions, the encoded macroblocks and motion vectors are alternated. This approach increases coding efficiency since the motion information has an important contribution to the total rate, mainly at lower rates.

Since motion vectors are split over descriptions, at the decoder, the motion compensation is done with the available motion vectors from received descriptions. In order to avoid visual artifacts in the reconstructed signal, overlapped block motion compensation (OBMC) is used when only one description is present. Since DCT coefficients are sub-sampled to form two descriptions, if one description is not available, the missing *DCT* coefficients are replaced by the corresponding ones in the previous frame. Since this scheme does not use drift compensation, distortion accumulation occurs in decoded sequences, whenever some description fails.

Open-loop MD video encoders have the advantage over other MDC schemes of maintaining the original temporal and spatial source resolutions, which simplifies the decoding

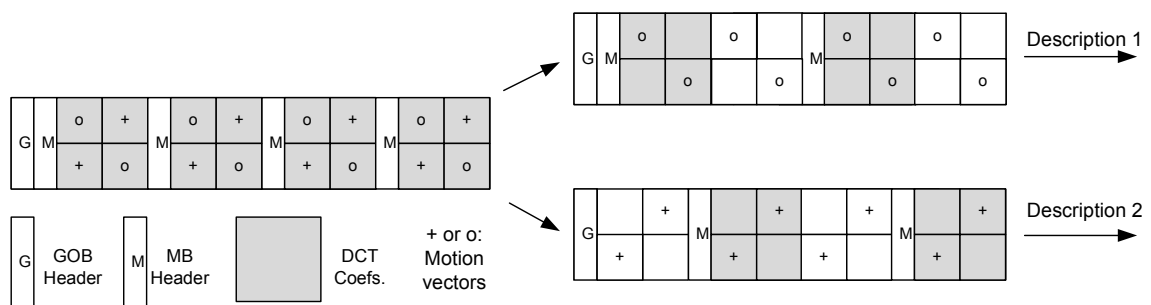


Figure 3.7: Data Partitioning among channels
[Kim 2001]

process, and also achieves smoother quality variations. Redundancy can be relatively low depending on several factors, like quantization step-sizes and redundant information that is duplicated in all descriptions. When coarser step-sizes are used, such overhead information is not negligible in the overall redundancy components. Nevertheless, prediction mismatch that occurs at decoder, is an important disadvantage that affect the overall performance obtained with such schemes.

Three-loop architectures In order to overcome the problem of drift distortion in MDC, multi-loop architectures with additional side information have been proposed in the literature [Wang 2005b] [Reibman 2002]. In comparison with the open loop, in these architectures, an excess rate is added to avoid the degradation of the side distortion, thus improving the overall rate-distortion performance.

The general architecture of a three-loop architecture is presented in Fig. 3.8. The architecture is formed by three encoders: one central encoder and two side encoders which generate a controlled amount of redundancy, directly related with the side distortion. Signal $e_0 = F - P_0$, is the prediction error of the central decoder, where F is the current frame and P_0 is the central motion compensated or spatial prediction. MD function is applied to the signal e_0 and the two output residues r_1 and r_2 are generated and entropy encoded. Two additional prediction loops are used in order to mimic the decoder when only one description is available. The residue r_1 and r_2 are individually decoded in the MDC decoder functional block, resulting in \hat{r}_1 and \hat{r}_2 . The prediction error e_1 is the difference between current frame F , the prediction P_1 and the residue \hat{r}_1 while e_2 is the difference between current frame F , and prediction P_2 . e_1 and e_2 are encoded according to a redundancy allocation scheme generating side information s_1 and s_2 . Description 1 is formed by the multiplexed signals r_1 and s_1 and description 2 is formed by multiplexed signals r_2 and s_2 .

Based on the generic architecture of Fig. 3.8, several MDC schemes have been proposed. The main differences are related to the central encoder and the adopted MDC techniques. For instance, an MDC architecture using *MDCT* was proposed in [Reibman 2002] where the MD video coder is jointly optimized using motion vectors, DCT and side information levels. Only the DCT coefficients are considered to build the different descriptions, while motion vectors and side information are simply duplicated. Redundancy levels of the proposed schemes are relatively high, between 50% and 80%, because it is necessary to add to the redundancy imposed by MDCT method, the side information that is sent to elim-

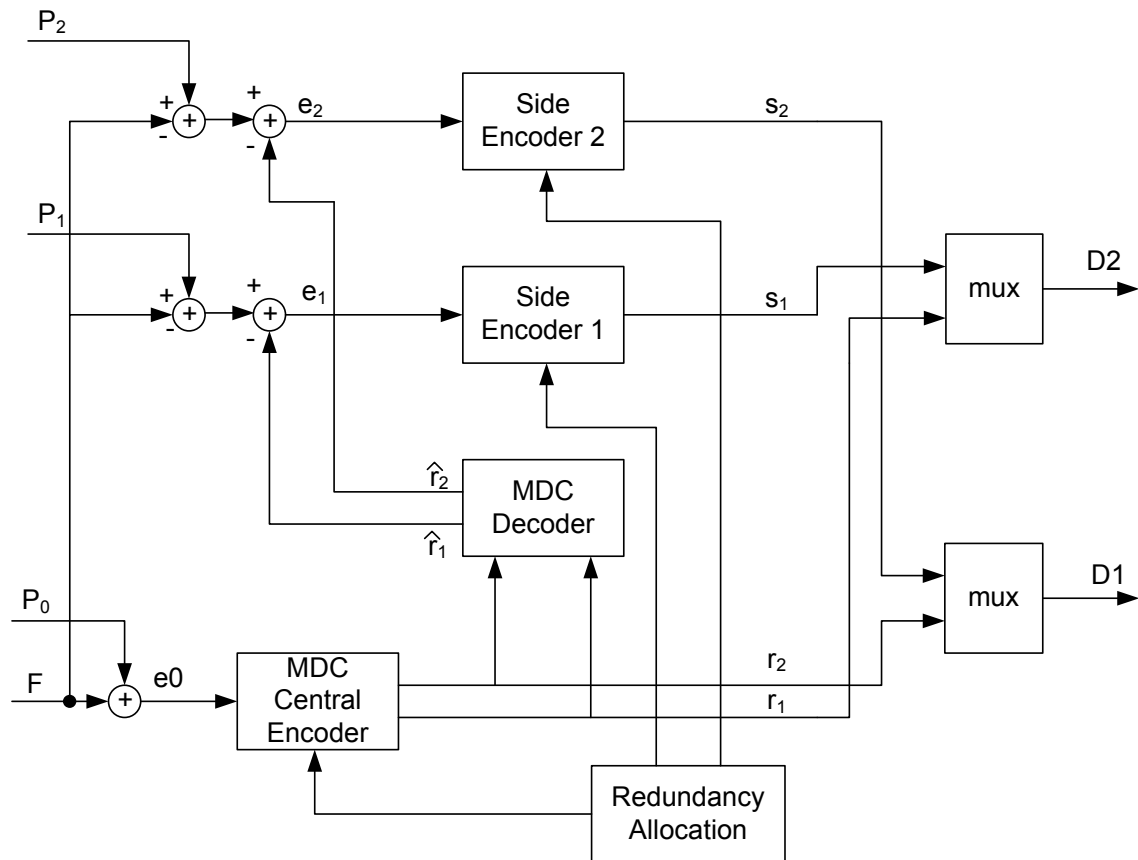


Figure 3.8: Tree-loop multiple description video encoder.

inate the prediction mismatch and this is not negligible in the overall redundancy levels. Even so, better performance than SDC schemes is obtained in packet-loss transmission scenarios. Nevertheless, despite the important role of side information in improving the quality of decoded signals, it is important to achieve an efficient tradeoff between redundancy allocation and distortion due to prediction mismatch. Another similar work has been proposed in [Franchi 2005] using a three prediction loop based on polyphase down-sampling in sub-band video coding, replacing the MDCT module. Instead of using the original frames, the reconstructed ones are used as reference to compute the prediction error of the side information. This scheme achieves a slightly better coding efficiency performance, when compared with earlier scheme.

In [Tang 2002], an MDC scheme is proposed using Matching Pursuits (MP) video coding instead transform coding. Video coding based on Matching Pursuits does not use DCT, and the residue from motion compensation is decomposed onto a basis set larger than the complete set provided by DCT. At each coding stage of the residue, a search for the best

basis function is performed by computing the inner product between the residual and the dictionary of basis function. The residue is subtracted from the best basis function and the iteration is repeated. This ensures that the most important features are coded first. The set of over-complete basis is called MP dictionary which consists in a general family of time-frequency atoms generated by scaling, translating and modulating a single time function. In MP-MDC, the residue is coded into two sets of atoms (one basis function). The first L atoms found during MP iterations are shared by both sets and subsequent atoms are alternatively put into the two sets. As a result, these sets have approximately equal importance, because atoms are found in decreasing order of magnitude in MP iterations. The correlation between these two sets of atoms is controlled by the number of shared atoms L . The authors compare performance of their proposed method with MDTC and show that better performance is obtained. Nevertheless, MD video coding methods based on dictionary are characterized by high computational complexity comparing with transform based video encoders, that limits its use. Another MP-MDC scheme is proposed in [Hsi-Tzeng Chan 2005] where MDSQ is used to encode the MP atom parameters into two descriptors.

Usually, the side information is useless when both descriptions are available, and in this case it leads to significant loss in coding efficiency. In order to improve the coding efficiency, in cases where the error probability is not severe, [Y.-C. Lee 2004] have proposed an alternative three prediction loop. The redundancy is reduced by using the side information to improve the quality of the central description in error-free conditions. Despite the better performance of this method in error free scenario, it has a poor performance for higher packet loss rates, because the quality can only be improved when all predictions are available.

In conclusion, tree-loop schemes have two main redundancy components. The implicit redundancy introduced by the multiple description technique, and redundancy introduced by the side information and duplicated information, such as motion vectors and headers. Redundancy levels can be high, considering the relative importance of each component. Notwithstanding, rate-distortion optimization methods are needed in order to optimally define the MDC parameters (for instance, in MDCT this is the correlation between descriptions) and side information rate usually defined by the quantization step-size [Liu 2005], allocating adequate redundancy levels to a specific application or network transmission scenario to obtain an efficient error resiliency capability.

3.3.2 MDC with pre-processing

In MDC with pre-processing the input video signal is downsampled in order to form two or more descriptions of the source, which are independently encoded and decoded. The full reconstructed sequence is obtained by merging all decoded descriptions. Two types of architectures have been proposed: open-loop and three-loop architectures. In open-loop architectures, each description is separately encoded and decoded. In this case, some proposals send explicit redundancy of complementary descriptions with a coarser quality, independently encoded. In three-loop architectures the complementary information is sent as side information which is generated using different prediction loops one for each description.

Open-loop architectures Fig. 3.9 shows a generic MDC scheme with pre-processing for two descriptions. The whole chain between encoder and reconstructed sequence is shown. The video sequence is pre-processed using either i) spatial down-sampling, where each frame is divided, usually in macroblocks or slices in order to make two or more groups of frames with reduced resolution, or ii) temporal down-sampling, where the video frames are separated to form two or more descriptions with reduced temporal resolution comparing with the original. Usually, each description is individually encoded. However, each encoder is able to use the information of the other description, encoding the complementary group as redundant information.

Each description is individually decoded, and based on the state recovery that depends on the received information of each description, is used to merge both of them. Error recovery at the decoder is done by using the other description already decoded and redundant information that may exist. A simple MDC scheme based on these principles was proposed in [Apostolopoulos 2000] where an odd/even frame splitting is proposed with two independent encoders. The redundancy results from the loss of prediction efficiency

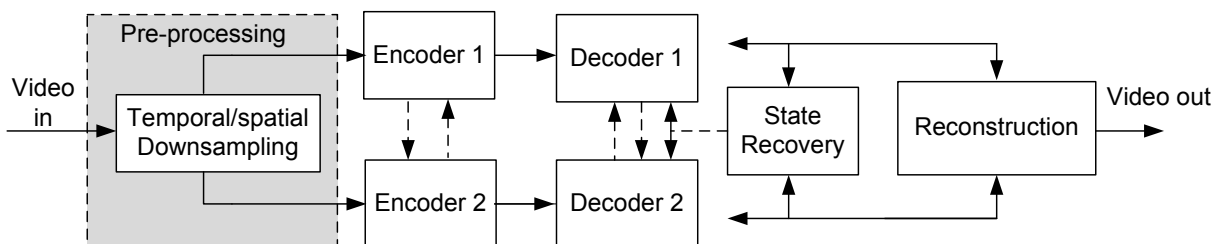


Figure 3.9: Pre-processing MDC.

due to increased temporal distance between frames. The main problem of such schemes is the poor side reconstruction in comparison with the SDC full rate scheme.

Several works exploit standard error resilience tools, such as slice reordering and redundant frames. These MD schemes follow the same principle described in section 2.2.3, where two types of information are encoded after down-sampling: primary information, resultant from a finer quantization, and secondary information (redundant frames), resulting from a coarse quantization. Secondary information is encoded as redundant information to be used whenever the primary information is not available. A redundant frame approach was proposed in [Radulovic 2010] in order to deal with missing frames at the decoder when odd/even separation is used at the encoder. The input video sequence is split into sequences of odd and even source frames. In the encoding process, each primary frame in the even/odd description is predicted only from other frames of the same description, typically the previous frame. In addition, redundant frames are included in the bitstream of each description, thus carrying the information from the alternate description. The time location is the corresponding one in order to replace a lost primary picture. Unlike the primary frames, which use the previous primary frames from the same description as a reference, redundant frames are predicted from the previous frame in the input sequence. The redundancy of this scheme can be high, considering that two sources are included. On the other hand, the frames in each description are temporarily further apart, losing coding efficiency. Thus, redundant picture information depends on how fine they are quantized. Reasonable redundancies are obtained when the Quantization Parameter difference between primary and secondary pictures are relatively high, which means that visual quality can be highly affected due to meaningful differences between decoded primary and redundant frames. Additionally, redundant information only improves the recovery of missing frames and does not deal with prediction mismatch at the decoder, which means that error propagation is not avoided.

A similar approach was proposed in [Tillo 2008], where each frame is split in even/odd slices. For each description, a complementary set of primary and redundant slices are defined. In this case, sub-sampling is done at slice level with full temporal resolution, which means that the extra redundancy introduced is relatively reduced. This additional information is mainly dependent on redundant slice quantization step-size. Additionally, an optimal redundancy allocation at GOP level was also devised based on the packet loss ratio. The scheme was later improved by using the optimization method at slice level in [Peraldo 2010] and more recently at macroblock level [Lin 2011]. In other later works,

redundant slices use the same prediction modes as the primary slices. An independent encoding of the redundant slices are considered in [Schmidt 2011] where a joint optimization mode decision of redundant information is proposed. Such schemes are interesting because of their reduced redundancy, but suffer from the same problems as the other open-loop schemes mentioned before. On the other hand, such schemes are dependent on specific coding tools, such as redundant slices and Flexible Macroblock Ordering (FMO), which cannot be available in the most of applications or profiles.

In addition to the redundant information explicitly sent to MD decoders, concealment and interpolation techniques can be used by exploiting the neighboring received frames when any description is lost [Ma 2008]. An improved concealment method is proposed by the authors to reduce the energy error propagation introduced by concealed frames in the following decoded frames. In order to reduce the error energy, a linear combination between the decoded frames using correctly received residue and interpolated versions are used to obtain decoded frames. Other efficient concealment schemes can be exploited in order to improve the decoded MDC signals since they have sufficient and rich redundant information to be usefully exploited by the decoder.

Three-loop Architectures A three-prediction loop MD video codec using temporal sub-sampling was proposed in [Wang 2002]. The video sequence is partitioned in multiple sets using temporal sub-sampling, and then an independent prediction loop to each set is employed. The original sequence is split in odd and even frames. The overall redundancy is controlled by the predictor coefficients and by the quantisation stepsize used for encoding the mismatch signal. Depending on the network conditions, the amount of redundancy is controlled to achieve a good tradeoff between coding efficiency and error resilience.

Another type of approach is based on the slice reordering feature of H.264/AVC, namely the FMO tool. A slice group scheme is presented in [D. Wang 2005] where three motion compensation loops are used. The video signal is encoded in the central encoder and, then, divided into two descriptions, each one corresponding to one slice group. Each slice group includes redundant information from the other one. Based on this scheme, a rate controlled redundancy-adaptive model that takes into account the effects of error propagation and concealment is proposed in [N. Kamanoonwatana 2010]. A similar approach is proposed in [Su 2008] where the temporal and spatial correlations between macroblocks are exploited to achieve efficient redundant coding. Although they exhibit good performance in terms of redundancy allocation, these schemes are highly dependent

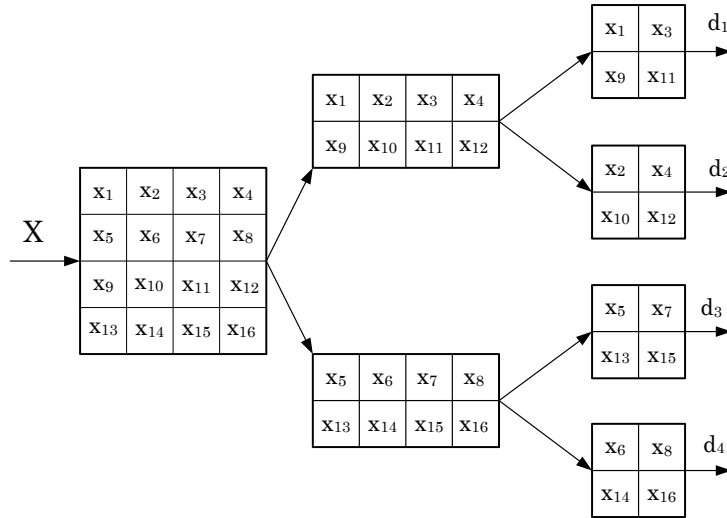


Figure 3.10: An example of high-dimensional MDC for $N=4$ descriptions.

on the specific features of the corresponding SDC coding schemes, and are not always available for all coding profiles.

3.3.3 High dimensional MDC

Usually, MDC video schemes generate two output descriptions. In the high dimensional MDC schemes a higher number of descriptions level are used at source coding. This is achieved applying pre-processing methods to increase the temporal and spatial down-sampling factor. Fig. 3.10 shows a spatial down-sampling example to generate $N = 4$ descriptions. An original image of 4×4 pels are downsampled by rows and columns, generating four descriptions with dimension 2×2 , each one is separately encoded. The main advantage on the increasing of the number of descriptions, is that it is easier to recover the missing information, because each description does not have the same importance than if a lower number are used. Since higher number of descriptions reduce the coding efficiency, the overall quality is not improved comparing with MDC schemes with only two descriptions. Nevertheless, due to its high scalability level, MDC with high number of descriptions can be considered an interesting approach in application scenarios where the source video is highly distributed.

In the recent past, several works have been proposed in order to increase the number of descriptions at source level. In [Franchi 2005] polyphase down-sampling is used to generate four descriptions. Comparing with SDC and 2-description MDC, the error-free perfor-

mance is highly affected because coding efficiency is lost caused by down-sampling. The higher number of descriptions can be advantageous only for very high packet loss rates. Nevertheless, considering the additional redundancy, such approach needs to be carefully evaluated in order to find out whether its is worthwhile for each specific transmission conditions. Another proposal that compares its results with [Franchi 2005] is presented in [Tillo 2010]. A multiple description scheme based on multi-rate coding for H.264/AVC is proposed, where the source data is divided into subsets, followed by encoding at several rates to properly assign different rates to the components of each description. In fact, this method corresponds to encode each source signal with different rates, and then combining each subset to generate several descriptions. The redundancy introduced by this approach is lower than spatial polyphase down-sampling, even so the performance reduction with respect to SDC is about 2.5dB, which is a not negligible value.

Other works have been recently proposed. A hybrid spatial/frequency downsampling scheme is proposed where the source data is spatially down-sampled followed by the splitting of *AC* DCT coefficients [Hsiao 2010]. This proposal is an open-loop embedded MDC scheme, that is highly dependent on drift propagation at the decoder. Also, its coding efficiency is highly affected because all bitstream components, such as prediction modes and motion vectors, are repeated for all descriptions. On the other hand, coefficient splitting makes run-length encoding very inefficient because the regular distribution of null values in each block. Yet, experimental results show that better performance is obtained comparing with polyphase downsampling. An equivalent approach is proposed in [Tsai 2010], where four descriptions are generated by using spatial and temporal downsampling. First, the input sequence is temporally down-sampled, and then, for each temporal component, the DCT coefficients are split, generating four descriptions. Additionally, an error concealment tool is proposed to recover lost temporal components of each description. In this case, better coding efficiency is achieved but the error-resilience capability at higher packet loss rates is affected. This is because the temporal concealment is based on spatially down-sampled frames, affecting the error propagation in the decoded sequence.

3.3.4 Multiple Description Scalable Coding

The scalable nature of the motion compensated video codecs based on sub-band/wavelet transforms were also exploited in order to obtain multiple descriptions. A few proposals have been published in the literature joining open-loop scalable video coding with

MDC. Bajic and Woods [Bajic 2003], apply data-partitioning approaches to the wavelet coefficients and motion vectors produced by the Motion Compensated Temporal Filter (MCTF). Although such scheme is able to give an optimal partitioning of each sub-band coefficients allowing to simplify the concealment at the decoder, the introduced redundancy may not be sufficiently adapted in order to improve the quality of decoded signals, mainly for high packet loss rates. In order to overcome this problem, an Embedded Multiple Description Scalar Quantiser based on MDSQ applied to the spatiotemporal decomposition of the input is proposed in [Verdicchio 2006]. Another approach to insert explicit redundancy in MDC based on 3D-scan based wavelet video encoder is proposed in [M. Pereira 2003] using an optimal redundancy allocation method based on the channel conditions, defining the rate of each redundant wavelet sub-band in order to minimize the overall distortion. Other approaches use temporally filtered frames splitting in order to create multiple representations of the input. For instance, in [van der Schaar 2003], where critical data, such as MVs and the low-pass temporal band are duplicated in all descriptions, and in [Tillier 2007] where a redundant motion-compensated scheme derived from wavelet decomposition is proposed in order to build temporally correlated descriptions in a t+2D video encoder. Wavelet video encoders are very suited to be used in MDC coding, since its scalability allows to refine several resolutions of the same source. This means that lost descriptions can be replaced or estimated from available components without a significant distortion. Nevertheless, this coding approach has not been adopted in most of video coding applications, which limits its use.

Using the scalability tools of the hybrid video encoders, several methods aim to combine the advantages of Layered Coding and MDC, defining the technique as Layered Multiple Description Coding. This can be made using the scalable structure, where multiple descriptions are generated using either the base, enhanced layers or both of them. For instance, in [Wang 2003] MDC is used on base layer encoding two descriptions, and then three versions of the enhanced layer are encoded using FGS scalability. In particular, considering that the base layer can be decoded using either both descriptions or only one of them, three versions of the enhanced layer are encoded using different prediction loop, each one corresponding to each decoding possibility. In [Chen 2007a] FGS scalability scheme is used with an open-loop MDSQ scheme for both layers, whereas other distinct approach was proposed in [Kondi 2005] where only the enhanced layers are multiple description encoded. These approaches are very dependent on the base layer reconstruction, which means that the robustness introduced at enhanced layers can be totally useless if the base layer is corrupted. Nevertheless, the results show that such approaches can

achieve better performance comparing with traditional scalable tools without MDC.

The scalable extension of H.264/AVC (SVC) has also been used to generate multiple descriptions. Since several prediction structures are present in SVC, the variety of combinations that can be used to encoded multiple descriptions also increases. For instance, a multiple description scheme based on distinct quality resolutions is proposed in [Abanoz 2009]. Two balanced descriptions are defined, each one scalable encoded, using several combinations of video segments (GOP or frames) encoded at high and low rates. Also a SVC-based scheme was proposed in [Favalli 2011] where half of the polyphase sub-sampled components are encoded as base layer and other components are predictively encoded as enhanced layers. The use of hierarchical B-frames in MDC is exploited in [Zhu 2009], where two descriptions are generated by duplicating the original sequence and then coded in hierarchical B structures with staggered key frames in the two descriptions. By using different QPs at different levels, their approach enables each frame to have two different quality fidelities in different descriptions. A better performance is achieved in [Tsai 2012] where different levels of redundancy are defined according to the hierarchy level and to the different fidelity requirements. Despite presenting great flexibility in encoding schemes and redundancy allocation, the decoding combination of each component needs to be efficiently exploited to correctly decode all prediction layers, taking into account the control of the drift distortion propagation.

3.3.5 Stereo/Multiview MDC

Considering the multi-view video coding, MDC is an efficient method for transmission over lossy channels. MDC applied to multi-view video coding exploits the additional dimension that is the interview redundancy. If each view is independently encoded, *i.e.* simulcast coding, such scheme could be viewed as an MDC scheme, where each description corresponds to one view. In this case, the introduced redundancy is the loss coding efficiency from interview prediction, that would be associated to the loss of 3-D perception when only one description (view) is available at decoder. Nevertheless, the loss of 3-D view cannot be acceptable in 3D video, which means that such solution cannot be seen as a true MDC scheme, and other solutions need to be explored. Taking the simulcast encoding of each view as reference, two MDC approaches have been introduced in [Norikin 2006]: spatial scaling stereo-MDC and, multi-state stereo MDC schemes. The first approach consists in encoding a redundant left (L_{red}) sub-sampled view predicted

form the main right view (R_{main}) and vice-versa, *i.e.*, encode redundant right (R_{red}) sub-sampled view predicted from the main left view (L_{main}). Each main view is predicted as a monoscopic video losing the interview correlation and then, the sub-sampled versions are totally redundant if both descriptions are available. This scheme can be highly affected by redundancy introduced by the loss of interview prediction between main views. On the other hand, stereo pair visual quality is highly dependent on the quality of the main view, which means that quality of the redundant views can be adjusted according to the overall redundancy level. The second approach consists in temporally split the stereo pair whose frames are then independently encoded. The additional redundancy is due to the temporal distance, increasing with the corresponding loss of coding efficiency. Comparing with earlier solution, this approach does not allow to adapt redundancy between descriptions and can also be affected by temporal discontinuity, which may add important visual constraints in stereo perception. Temporal sub-sampling is also used in [H. Abdul Karim 2009] where an 3-D even and odd MDC scheme with adaptive redundancy added to frames with motion activity higher than a threshold.

An MDC scheme based on scalable coding was proposed in [M.B.Dissanayake 2010], where a redundant version of the enhanced layer is encoded. Only disparity vectors, intra residues and headers are encoded as redundant information and the interview residue is set to zero. A View Interpolation Based Multiple Description Coding scheme is proposed in [Xiu 2009]. Each description is made by the even and odd views, and for each view, redundant information is encoded, which corresponds to the residue between the interpolation of even(odd) views from odd(even) views and even(odd) original ones. A rate-distortion optimization model for stereoscopic video streaming with unequal error protection is presented in [A. Serdar Tan 2009]. The coding structure uses scalability where each layer has different priority and protection. In [Erhan Ekmekcioglu 2010], an MDC scheme is presented using video+depth coding, with viewpoint synthesis. Scalable coding is used and each one of the two viewpoints is encoded with two spatial layers and two temporal layers. In addition to the scalable layers, a redundant bitstream for the base layer is generated where only the more significant foreground objects are redundantly encoded.

The use of the MDC for multiview video delivery is a promising application not only as a error resilience tool, but also as an network adaptation scheme. In fact, the idea of having different views of one scene distributed in different servers that can be used by each user without having the scalability constraints is an usefully application of MDC in a multiview

scenario. The proposed literature in this field is still incipient and usually the available coding tools such as scalable coding are used in order to devise MDC schemes mainly based on stereo video. Actually, the complexity of the multiview video coding tools is not only a restriction of the MDC development, but also a challenge and is definitely an open issue research. For instance, the combination of different multiview coding formats for the same source video that can be combined in function of the user constraints is an open issue that have not been explored in the literature.

3.3.6 Unbalanced MDC

In general, most of the existing MDC architectures use $N = 2$ descriptions with approximately the same rate, *i.e.* balanced. However, due to non-stationarity characteristics of channel conditions in terms of available bandwidth and packet error rates, MDC encoders should dynamically adapt the transmission rate of each description in order to minimize the distortion at decoder. Therefore, an MDC scheme should be able to produce asymmetric bitrate for each description as necessary, *i.e.* unbalanced MDC (U-MDC), where each description may have different resolutions and/or qualities when only one description is decoded.

Fig. 3.11 shows an unbalanced MDC application scenario using two descriptions transmitted over different paths. Streaming server feeds the client with two logical paths. Each path is defined over distinct network branches between client and server, passing through intermediate network nodes. Usually, each branch is characterized by the available bandwidth, delay and packet loss rates, represented by R_i , d_i and p_i , $i = 1..4$, respectively. Moreover, considering that the streaming server needs to adapt the MDC output rates to the available channel bandwidth, the use of unbalanced MDC schemes are very useful to cope with network conditions. Usually, this is done using rate control schemes which can include rate-redundancy distortion models. Furthermore, since bandwidth and packet loss can differ in each branch, the intermediate nodes can perform rate adaptation in each description, supported by the information received from streaming server or by the other network nodes.

The U-MDC schemes use distinct approaches such as adapting the quantisation, temporal or spatial resolutions. An U-MDC scheme is proposed in [Apostolopoulos 2001b] using different frame rates for each description, while maintaining the quality of individual

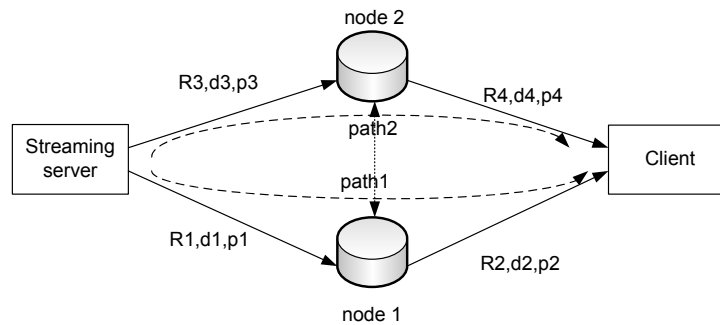


Figure 3.11: Unbalanced MDC application scenario.

frames. The redundancy introduced at lower rate descriptions can be high because the temporal distance between frames increase. In [Comas 2003] and [Bin 2006], an unbalanced MD video coding with rate-distortion optimization method is proposed to generate two descriptions maintaining the original temporal resolution. In particular, one description corresponds to encode the sequence with high resolution (HR) while the other is encoded with low resolution (LR). The main problem of such approach is that the overall distortion corresponds to the HR description, so the LR description is totally redundant, which means that the LR description does not improve the decoded sequence when both of them are received. Nevertheless, such approaches cannot be considered true MDC schemes because the overall quality is not improved when both descriptions are available.

Unbalancing rates can be obtained by adapting the amount of redundancy that is sent in description. For instance, in [Kim 2005], is proposed an unbalanced multiple description coding framework for wavelet-based coders. In particular, the proposed method groups different wavelet coefficient trees for each sub-stream based on the available source coding bit rate for each sender, and then encode each group independently. This work shows that unbalanced MDC combined with rate allocation algorithm achieves higher performance compared to conventional balanced MDC.

In [Kamnoonwatana 2011] an asymmetric rate control scheme is proposed for H.264/AVC, where the MDC redundancy is adaptively allocated based on an end-to-end distortion model. The overall rate is controlled by jointly setting the central and side distortion. Although resultant rates of the encoded descriptions may vary the corresponding redundancy according to channel conditions, this method does not allow changes in each description rate without severe degradation of the central distortion. Overall, further research needs to be done in order to generate unbalanced descriptions without losing error resilience capabilities.

3.3.7 Multiple Description Distributed Video Coding

Distributed Video Coding can be exploited to provide error resilience features for robust improved decoding. The independent data encoded at the transmitter provides robust decoding when combined with predictive correlated information used as side information given a correct decoding [Sehgal 2004]. Therefore, these can be expanded to MDC where resilience obtained from redundancy are combined with error robustness provided by Wyner-Ziv (WZ) coding which derived from its channel coding approach which is named as Multiple Description Distributed Video Coding (MDDVC). Usually, MDDVC approach use the replication, in both descriptions, of the Wyner-Ziv model composed by WZ frames combined with key frames encoding as side information.

In [Milani 2010] an MDDVC scheme based on redundant slices is proposed, where WZ slices are encoded with different lossy syndromes. The performance is increased over MDC based on redundant slices for high loss rates. Although such approach could not be considered as a true MDC scheme, in [Rane 2008] a scheme using redundant slices was proposed where such slices are encoded as WZ frames. At the decoder, to correct the lost primary slice, the ones received without errors, act as side information for WZ redundant slices. In [Fan 2011] an MDDVC scheme based on a pair of staged quantizers and, consequently, a bitplane extraction scheme is proposed in order to make two WZ descriptions. The intra frames are encoding with MDCT. Although such scheme achieves drift reduction, its performance is only better than the classic MDCT scheme for higher packet loss rate due to the high level of redundancy that is introduced. Another proposal using DVC is presented in [Crave 2010] where for both key frame and WZ-frames two descriptions are produced using MDSQ. Finally, a DVC scheme based on three-band motion compensated temporal filtering video encoder (MCTF) is proposed in [Crave 2008]. The original subband layers are split into two descriptions. Each frame is conventionally encoded in one description and WZ-encoded in the other description. Comparing with conventional MDC encoder this proposal has better side-distortion performance due to channel error recovery feature of the WZ frames used only at side decoding.

The DVC approach in MDC field improves the error robustness due to channel encoding of WZ frames. However, the redundancy that must be used does not compensate the loss of central encoder quality, which means that only for high packet loss rates such approach could be competitive. Nevertheless, the MDDVC scheme is an interesting solution in applications where most of the computational complexity is transferred to the decoder.

3.4 Multiple Description Video Streaming

Video streaming is a generic concept used to express the transmission of video signals through packet networks. Real-time video transmission schemes can be grouped in two main categories: bi-directional and unidirectional, which includes generation and transmission in real-time. This can be further divided in live and on-demand. In live streaming, video signals are seen by clients in real-time from one or more servers, usually from a unique multicast session. In on-demand streaming, each user has a different and asynchronous streaming session from one or more servers.

Generally, streaming video is characterized to have strict delay constraint that makes streaming services very sensitive to packet loss and network outages. For example, when receiving a streaming video session, data packets that arrive late are useless and thus considered as lost. Also, streaming video over wireless mobile networks, impose significant challenges to video transmissions since node mobility and lack of infrastructure in the network can lead to frequent link failures and route changes. Furthermore, the link quality is affected by fading and interference in the wireless channels. Since traditional approaches for dealing with packet loss, such as retransmissions, may not be possible in real-time streaming context or in wireless networks due to its error-prone nature, and also the vulnerability of compressed video to packet losses, additional mechanisms are needed to provide streaming media delivery over packet networks.

Due to these issues, the deployment of mechanisms to overcome the effects that errors and failures have on the streaming quality, is a mandatory requirement. In this context, path diversity is a powerful solution to overcome this problem. Therefore, MDC appears as high competitive source/channel coding approach in order to be applied in networks where its infrastructures must be highly dynamic. Considering this, MDC video streaming is included in a wide range of applications where path diversity networks are used to deliver compressed video either from the source (e.g., video codec, streaming server) to the final destination (e.g., user terminals) or from a network node to mobile terminal (e.g., wireless link with path diversity). In particular, MDC is regarded as an efficient error resilience coding scheme to overcome the lack of centralized QoS control that might exist in more classic streaming application scenarios.

The usage of MDC in path diversity video streaming is also a competitive alternative with other network-adaptation schemes, such as scalable video coding for robust streaming

(SVC) over time-varying networking conditions. As mentioned before, the main difference between SVC and MDC lies in the inter-dependency of SVC layers in contrast with independent MDC descriptions. Therefore, MDC clearly has the advantage of higher flexibility and independence from a common base layer, which also contributes to simplify the network management mechanisms [Chakareski 2005]. The comparison between both techniques is greatly dependent on the application scenario, but generally MDC gives a stronger error resilience protection in applications where feedback channels are not available and also in networks with either high packet loss rate or high dynamic topologies like P2P frameworks [Xua 2012].

Different MDC video streaming architectures have been proposed but in general they can be grouped in two main areas: content delivery networks (CDN), which is the traditional client-server streaming approach, and peer-to-peer networks (P2P), where each client acts simultaneously as a peer to deliver video contents to other users. Traditional client-server based technologies offer the possibility to provide good performance and high availability rates. However, those approaches usually need high deployment and maintenance costs. These costs could increase drastically if more videos continue to improve to higher qualities. Therefore, several P2P streaming systems have been proposed to enable live and Video-on-Demand (VoD) streaming to large audiences with better video quality at low deployment and server costs [Liu 2008b] [Liu 2008a]. Furthermore, utilization of the P2P paradigm yields several advantages, such as, inherent bandwidth and resource scalability, network path redundancy, and self organization [Jurca 2007].

3.4.1 MDC Video Streaming over Content Delivery Networks

Content Delivery Networks are point-to-point communications that were developed to overcome network congestions and server overload that happens when the number of users highly increase. In order to enhance the end user performance, the descriptions of the same source content are distributed over networks, at different levels, not only at network core of the topology, but also using caching of the popular contents on edge servers located closer to the users. The main advantage is that it prevents the server overload, since the replicated descriptions can be delivered from the closest edge server and not from the origin server and, secondly, the use of a shortest path reduce the request response time, the packet loss rate and the network resources.

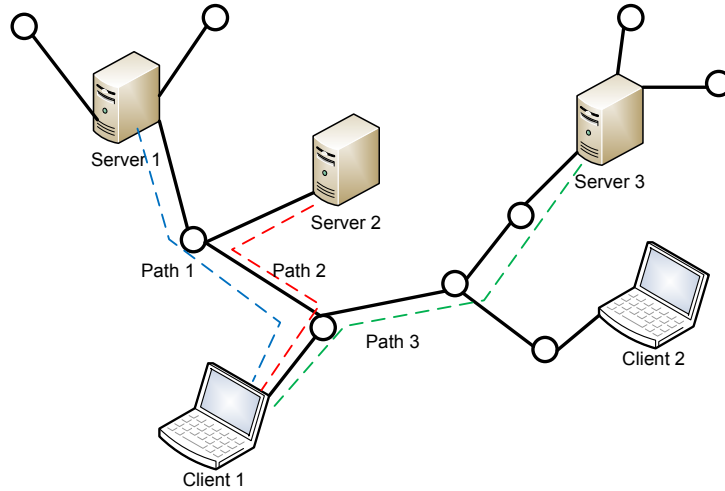


Figure 3.12: Path diversity in content delivery networks.

Video streaming in content delivery networks using path diversity and MDC, extends the traditional streaming CDN approach (single description)(SD-CDN) to the multiple description case (MD-CDN) [Apostolopoulos 2002]. In fact, the traditional scheme deals with the problem that consists in finding the best server or edge proxy to feed one specific user. The optimization criterion is usually the shortest path between client and server.

Fig. 3.12 shows an MD-CDN topology capable of supporting MDC streaming with path diversity. Different servers feed several clients where video source descriptions are distributed using different paths. Hence, in order to achieve path diversity, the same description or the complementary one, needs to be available at different servers, having at least one path between the client and one server. For instance, Client 1 are feed from Server 1 using Path 1. If link goes down, then the streaming server switches to Server 2 through Path 2. Furthermore, in MDC video streaming several descriptions are streamed form different servers using different paths. For instance, Client 1 can stream complementary descriptions from Server 1 and Server 3 using paths 1 and 3 respectively.

Therefore, MD-CDN poses different problems compared with SD-CDN. In case of MD-streaming, different designing issues need to be considered, namely: i) how to distribute the servers in the network, ii) how to distribute different descriptions among the available servers or proxies and iii), how to select for each client multiple servers with complementary descriptions and paths, in order to achieve path diversity. In this regard, MDC streaming offers the advantage of using a distributed architecture with several servers at different locations. In [Apostolopoulos 2002] it is shown that MD streaming performs better than SD streaming using the same CDN infrastructure without any MD optimized

server placement, with fewer number of servers for the same expected distortions. Using the benefits of path diversity and taking into account the definition of disjointness ratio between paths used for server placement, then the reduction of distortion is significantly improved compared with the shortest path criterion as used in SD-CDN.

Furthermore, overlay infrastructures can also be exploited to transmit MD source coding sequences using optimal multi-path routing schemes using disjoint paths, *i.e.*, independent and non-overlapped. In this process, several optimization schemes are used based on selecting multiple paths using the shortest path or maximally link-disjoint criteria. However, the quality of the decoded signals is significantly improved using end-to-end distortion optimization taking into account the channel conditions of each available path, such as packet loss rates [Begen 2005], or channel rates using unbalanced MDC [Kim 2005]. Other methods can be used such as, assessment tools for evaluating the quality of experience (QoE) combined with a multipath routing strategy based on bandwidth availability estimation [Ghareeb 2010].

3.4.2 MDC Video Streaming over Peer-to-peer networks

MDC video streaming over peer-to-peer networks (MD-P2P) are mostly characterized by having a fully decentralized architecture where all network nodes have equal functionality. The members are connected based on a specific construction policy forming a distributed network topology. On the other hand, streaming peers are usually heterogeneous in terms of their processing and storage capacity, and network bandwidth. Also, each peer node has a dynamic and unpredictable contribution to the streaming topology. Therefore, not only the streaming topology is very dynamic, but also no assumptions should be made about the availability of resources or network paths. Consequently, MD-P2P streaming networks must be able to recover from the unexpected and ungraceful leave of any of its members at any time.

Different from other P2P systems like file sharing, successfully MD-P2P video streaming, needs to have low end-to-end packet loss rate, delay, and delay jitter, similarly to point-to-point live streaming. In order to satisfy these requirements, the transmission bandwidths must be effectively exploited in bandwidth-scarce environments, adapting to the changes in network conditions, such as rate congestion in some of the peer-to-peer transmission paths available for each description. MD-P2P streaming topologies are defined using an

overlay layer, which is responsible for maintaining the network topology at every instant, improving the dynamics of the overall topology. So, resilience in MD-P2P networks is achieved by defining different redundant topologies in order to deal with peer nodes that randomly join and leave the overlay during the streaming session, maintaining the cohesion of the overlay structure, and minimizing the information loss due to node failures.

Path diversity can be achieved by different forms of exploiting the specific network topology, *i.e.* structured (tree-based) or unstructured (mesh-based). Usually in tree-based topology, path redundancy is achieved by setting a multi-tree structure in order to overcome the possible fail or congestion of one tree. In the mesh-based topology, several logical disjoint paths are defined in order to set distinct paths to deliver each description. In the case of MD-P2P applications, the tree-based approach consistently exhibits an inferior performance over the mesh-based approach in their ability to cope with the instability of the peers, who join and leave the system, due to the static mapping of content to a particular overlay tree and diverse placement of peers in different overlay trees [Magharei 2007]. Overall, MD-P2P streaming can be significantly better than in a MD-CDN schemes, despite the high degree of unreliability of the P2P systems [Khan 2004].

Tree-based streaming In tree-based MD-P2P video streaming, the clients are assigned to a multicast session, usually with centralized management. One major drawback of tree-based streaming is their vulnerability to peer churn (peer leaves the network), because if one peer leaves the structure then all the tree must be reconfigured in order to define a new topology. Error resilient streaming is obtained by exploiting multi-path transmission. In tree-based MD-P2P topologies, the error resilience is achieved using multiple-trees transmitting different representations of the source with several multicast trees.

The Splitstream framework described in [Castro 2003] splits the content to be transmitted in several parts, and then multicast each part using a separate tree. The main goal of this system is to overcome the inherent unbalanced forwarding load in conventional tree-based multicast systems using multiple multicast trees. Thus, the forwarding load is distributed over all participating peers respecting the limits of each distinct capacity limits, thereby reducing the bandwidth demands on individual peers. Despite exploring the multi-path error-resilience capabilities, this scheme is not an error resilience scheme since the transmitted source is only split in different parts without taking account video coding constraints.

Nevertheless, this framework can be exploited in the MDC video streaming context. Fig. 3.13 shows an example of redundant multiple-tree topology using two redundant trees for MDC transmission. Each description is streamed from the server through different trees. For each peer, descriptions are delivered through different paths. This fact mitigates the problem of peer churn and reconfiguration because at least one description is available for each node.

In the CoopNet system an MD-P2P multiple-tree framework is proposed based on MD-FEC encoding scheme [Padmanabhan 2003]. The system is based on a centralized management, generating and managing a diverse set of distribution trees. Resilience is exploiting building short and diverse trees, where each peer acts as an interior forwarding node in a few trees and a leaf node in the rest of the trees. Thus, several disjoint paths are set contributing to tree diversity and robustness. CoopNet use MD-FEC in order to generate M descriptions. The proposed scheme presents its effectiveness in comparison with traditional FEC approach. However, such scheme has not the ability to deal with heterogeneous environments allowing MDC source adaptation at servers or peer nodes to available rates in each multicast tree. Moreover, since encoding is done at GOP level, an additional delay is introduced in order to perform the unequal protection block. In [Wei 2007] a multiple-tree transmission scheme for wireless ad-hoc networks was proposed using MDC as source coding based on video coding with matching pursuits. Two disjoint multicast trees are used one for each description, using a parallel multiple nearly-disjoint multicast trees where each receiver is able to always connect to each tree. Thus, a same time routing overhead and construction delay compared with single multicast tree is achieved. Performance evaluation shows that using MDC in multiple trees outperforms the single tree multicast.

Another proposed scheme based on multiple multicast trees is proposed in [Setton 2008] and inner references for P2P live multicast streaming base based on congestion-distortion optimization scheduling. These scheduler is composed in two parts. At the sender, a prioritization scheme determines the transmission order of video packets destined to multiple peers. At the receiver, a feedback mechanism is used to recover missing video packets when a peer is disconnected from one or several multicast trees. However, since this is based on receiver acknowledgments, such scheme is highly affected by network dynamics, dissymmetry between uplink and downlink throughput and unpredictably peer churn that affects congestion prediction and also does not guarantee available feedback channels with acceptable delay.

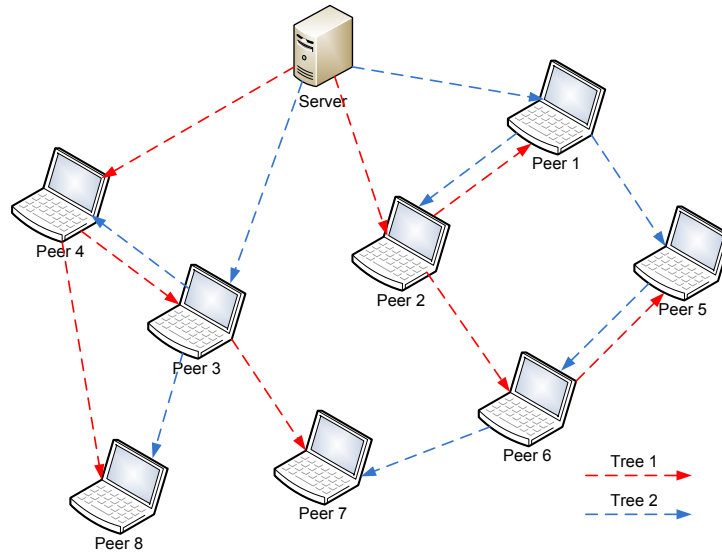


Figure 3.13: MD-P2P based on multiple-tree topology.

Mesh-based streaming In mesh-based MD-P2P streaming system, peers are not confined to a static topology, which improves the robustness to peer churns comparing with tree-based topology. Fig. 3.14, shows a mesh based MD-P2P scheme similar to that proposed in [Zha 2005], the CoolStreaming/Donet system. This is characterized by three key modules: 1) a membership manager, which helps the node to maintain information about other overlay nodes; 2) a partnership manager, responsible for establishing and maintaining the partnership with other nodes; 3) a scheduler which distributes video data among nodes. This framework has been implemented for video distribution over internet. Although its interesting achievements, its performance is highly dependent on the available bandwidth on the edge nodes, and also on the number of nodes in the network. Moreover, the implementation is based on generic internet traffic that does not take into account video adaptation techniques in order to improve decoded video quality.

The NTUStreaming framework [Lu 2007] is a MD-P2P system using mesh based topology for Internet Protocol Television (IPTV) infrastructure adapted to heterogeneous networks. Considering the topology management system, this framework intends to provide a partnership formation to reduce disconnecting time and isolated peers by exploiting the available resources, such as storage and bandwidth among peer users. At source coding level, it uses MDC based on pre-processing down-sampling and adapting each description rate according to the available bandwidth and device capability. Two independent streams resulting from temporal downsampling and two fully redundant streams from spatial downsampling are encoded for each sequence. Despite the effectiveness of proposed

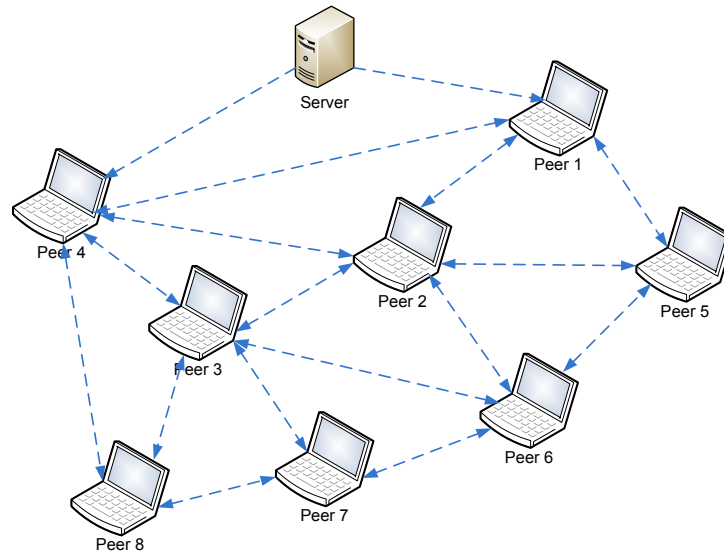


Figure 3.14: MD-P2P based on mesh topology.

MDC scheme due to its flexibility, the interpolation schemes at the decoder severely degrade the video quality when only one temporal and spatial description is available.

3.4.3 MDC video streaming over Wireless Ad-hoc Networks

In Wireless Ad-hoc networks the information exchange among mobile nodes is achieved through multi-hop wireless communications. Like other streaming infrastructures, wireless ad-hoc networks pose a great challenge for video transmission. The infrastructure is not fixed and variability of topology is conditioned by the node mobility. Therefore, wireless links are constantly established and broken and quality is constantly affected by channel fading and interferences from other channels.

As in P2P streaming over internet, MDC is a suited source coding approach in order to cope with variability of the infrastructure and transmission errors, particularly in wireless ad hoc networks, where a single reliable path may not exist and packet loss may be beyond the recovery capability of error control mechanisms. One of the first works that use multistream coding based on MDC is proposed in [Mao 2005a]. The mesh topology is exploited in order to establish multiple paths between source and destination. This work compares three error resilient schemes, namely ARQ, layered coding and MDC. As in other applications, MDC outperforms the other error resilient schemes in low delay applications without feedback channel. In video transmission based on MDC, one of the

main issues in ad-hoc wireless networks is the problem of multipath routing. These intend to find an optimal routing topology that minimizes the application layer video distortion. In [Mao 2006] the authors propose a meta-heuristic approach to solve the cross-layer optimization problem of routing, which minimizes the application layer video distortion. However, computational complexity highly increase with the number of nodes. More recently, is proposed in [Wei 2009] an MDC streaming scheme for multi-path wireless Ad-Hoc networks that minimize the concurrent packet drop probability, improving MDC decoding capabilities. A routing-aware multiple description video coding approach is proposed in [Liao 2011] to support video transmission over mobile ad-hoc networks with multiple path transport. The reference frames are chosen based on estimated packet loss probability from the received routing messages. Despite such method does not require additional feedback channel or extra overhead, it does not be applied to pre-encoded video where reference frame selection is already chosen.

3.4.4 Scheduling in path diversity networks

In path diversity networks, besides the streaming paths, routing and average rate allocation definition, proper scheduling schemes of the media packets are needed in order to guarantee efficient video delivery. Furthermore, not all packets of video stream have the same importance, which means that they do not equally contribute to the video quality at the receiving node. Moreover, a packet is useful at decoder only if its delay is not higher than an acceptable limit, and also if previous packets have been correctly received. The unequal importance of video packets, along with timing constraints, requires the derivation of efficient packet scheduling algorithms that determine which packets should be transmitted at a given time instant on a given streaming path, in order to maximize the overall video quality. One possible approach is to classify the video data based on coding parameters and to add error correction capabilities according with their importance. For instance, in [Qu 2006] the FEC redundancy is set according to the motion activity classification of the source. Also in [Tillo 2011], an UEP scheme is used, where FEC redundancy allocation is a function of the relative importance of each source slice.

Considering packet scheduling solutions, these have been studied in client server architectures with a single channel where rate-distortion optimized packet scheduling strategies are used [Chou 2006]. These approaches have been extended for path diversity environment using MDC. These were proposed in [Chakareski 2005], [Chakareski 2008],

[Chakareski 2006]. Although such approach to give an optimal scheduling in the rate-distortion sense, it needs a feedback channel in order to dynamically adapt the channel model, which can not be available in most of the applications. On the other hand, this approach is computationally intensive since it needs, at the same time, to set the channel conditions in function on acknowledgement information, and also, to find the optimal rate-distortion packet scheduling. Moreover, the method needs an additional delay, limiting its application in real-time applications. A quite different approach is the congestion-distortion optimization scheduling for packet prioritization based on retransmissions in P2P streaming proposed in [Setton 2008] and [Xua 2012]. The importance of each packet is not only adjusted based on rate-distortion issues, but also based on the numbers of descendent nodes in multicast tree that are affected by the loss or late arrival of this packet. However, its performance is highly conditioned if acknowledgement information is not available at each network node.

In [Milani 2008] a packet classification scheme for MDC video streaming is proposed, based on the relation between the number of zeros of transform coefficients used in each description and the corresponding PSNR if description is not received. This information is used to match the QoS classes supported by the IEEE 802.11e MAC layer. This approach works well whenever both packet streams are affected by the same congestion, but is suboptimal in other cases. In order to overcome this problem, a slightly different classification approach based on game theory is proposed in [Milani 2010] to cope with network congestion periods in a video streaming scenario where multiple descriptions are streamed from several servers. This method intends to avoid that one description control all the available bandwidth. However, it is assumed that all descriptions share the same channels which limits its application in path diversity transmission scenarios.

3.5 Conclusion

MDC video streaming is an useful error resilience source coding approach exploiting multipath communications over networks. The source video is encoded in several independent decodable descriptions that can be combined to improve decoded video quality. Several techniques are used being highly dependent on the source coding approach. These can be generated using pre-processing units based on temporal and downsampling schemes, using MDC techniques located at prediction loop and finally in post-processing

units exploiting layered coding schemes. Moreover, independent description decoding has several constraints related to predictive video coding schemes where interpolation and drift control are the main issues to be solved. In order to overcome this inherent drawback, error resilience features are improved by adding additional implicit and explicit controlled redundant information. Therefore, taking into account each source coding paradigm, MDC schemes have to jointly consider coding efficiency, error resilience and decoding schemes. The main challenge is to find the best compromise in order to improve the overall quality.

MDC compete with other error resilience schemes such as scalable coding and retransmission schemes based on feedback channels. The advantages over other methods are dependent on application scenario, but generally MDC outperforms the other methods in networks with low delay constraints, without feedback and with high packet loss rates. In particular, MDC is a suited source coding approach for video streaming applications over ad hoc wireless and peer-to-peer networks exploiting multipath communications. These are characterized by instability of the network topologies, thus suffering with packet loss errors and delay. Related to MDC, cross-layer optimization multipath routing and packet scheduling methods are used in order to improve decoded quality of video sequences taking into account link failure and node congestions. However, further research on MDC methods for heterogeneous environments are needed, based on rate adaptation considering independent decoding of each description, including its application in pre-encoded video sequences.

MDSQ for Advanced Video Coding

Contents

4.1	Balanced MDSQ	74
4.2	Unbalanced MDSQ	75
4.3	MD Video Architecture	77
4.3.1	Open-loop MD Encoder	77
4.3.2	Open-loop MD Decoder	78
4.3.3	Drift Analysis	79
4.3.4	Multi-loop MD Encoder	82
4.3.5	Multi-loop MD Decoder	84
4.4	R-D Performance - Two Descriptions	85
4.4.1	Balanced MDSQ	86
4.4.2	Unbalanced MDSQ	92
4.5	R-D Performance - Single Description	95
4.5.1	Balanced MDSQ	96
4.5.2	Unbalanced MDSQ	101
4.5.3	Frame-by-frame Distortion	107
4.6	Conclusion	109

This chapter proposes a MD coding scheme for Advanced Video Coding based on a MDSQ mixed open-loop/multi-loop structure, which prevents drift distortion accumulation in both intra and motion compensated predicted slices. The drift is compensated by generating a controlled amount of side information used by the decoder whenever any description is lost in the corresponding network path. In order to maintain a reduced redundancy cost, the side information is only applied to anchor slices, which allows to prevent distortion accumulation for single description decoding. New types of index assignment tables are proposed which can be used to dynamically change the rate of each description. The proposed method jointly controls the drift distortion that occurs when only one description is decoded and the redundancy introduced by the side information. The proposed method allows to improve the MDSQ error robustness comparing with open-loop schemes with an expense of introducing a small amount of additional information.

4.1 Balanced MDSQ

Most of MD video encoders based on MDSQ are designed to generate two independent descriptions from the same source signal. The main principle is to construct two uniform quantisers with coarse reconstruction values such that separate inverse quantisation is possible to implement. If joint inverse quantisation of both quantised descriptions is performed, then finer reconstruction values are obtained. Considering only two different descriptions, the goal of MDSQ is to find coarse scalar quantisers for each side encoder such that after being combined together, a finer central quantiser is obtained, producing lower distortion than each individual side decoder. MDSQ is based on two different functions: central quantisation and index assignment. The index assignment is used for mapping each quantisation index of the original transform coefficients i_0 (i.e., central indices) into a pair of side indices (i_1, i_2) which are then entropy encoded. In general, MDC algorithms use balanced descriptions where the respective rates and distortions are approximately the same in all descriptions. Considering two descriptions $S1$ and $S2$, with rate-distortions pairs $(R1, D1)$ and $(R2, D2)$, the two descriptions are generated using side-quantisers from a central quantiser with rate-distortion pair $(R0, D0)$. For balanced MDSQ, $R1 \approx R2$ and $D1 \approx D2$.

The underlying principles used in MDC for Advanced video coding follows the same approach as proposed in [Vaishampayan 1993] as explained in Chap. 2. An index assignment

matrix is defined as shown in Table 4.1, whose elements are the SDC quantiser indices, i.e., central indices, each one corresponding to a pair of side indices defined by the respective column and row. The amount of redundancy introduced by this MDC scheme is controlled by an index spread parameter k where $2k + 1$ is the number of diagonals of the index matrix. In Table 4.1, $k = 2$ which means that 5 diagonals are used.

Table 4.1: Balanced MDSQ, $k=2$

		i2 (Description 2)									
		-4	-3	-2	-1	0	1	2	3	4	5
i1 (Description 1)	-4	-20	-16	-14							
	-3	-17	-15	-12	-8						
	-2	-13	-11	-10	-6	-4					
	-1		-9	-7	-5	-1					
	0			-3	-2	0	1	4			
	1					2	5	6	8		
	2					3	7	10	12	14	
	3						9	11	15	16	18
	4							13	17	20	...
	5								19

At the decoder, if both descriptions are available, then an inverse index assignment process is used to restore a unique central index i_0 to be inverse quantised and inverse transformed. In the case where any description is not available for decoding due to transmission errors/losses, the central index cannot be unambiguously determined because there are several possible values for each individual description index. For instance, if only description $S1$ is decoded and $i_1 = 2$, then the reconstructed index i'_0 can be any of the following values, $i'_0 = 3, 7, 10, 12, 14$. Using the choice of main diagonal index as reconstruction rule, then $i'_0 = 10$, which can be different from the original SDC value i_0 . This leads to a decoding error in received indices and in the case of predictive encoding, as all current standard video coding schemes, such type of error originates a mismatch between the original SDC prediction loop and that of the MDC decoder, producing drift.

4.2 Unbalanced MDSQ

Traditional index assignment tables are defined to produce balanced descriptions as shown in earlier sections. To the best of our knowledge, encoding unbalanced descriptions using MDSQ is a new topic that has not been exploited in the literature. The main purpose is to join two unbalanced descriptions in order to obtain good quality reconstruction, while

a worst reconstruction yet acceptable is achieved when only one description is decoded. In this case a distinct amount of side distortion is generated for each description. Thus, in unbalanced MDSQ two equivalent side encoders are used, with shifted and different coarser quantizers in order to produce different rates in each description and at the same time, to generate an equivalent finer central quantizer.

The specific characteristic of unbalanced MDC is to generate two descriptions $S1$ and $S2$, in which any rate-distortions pair $(R1, D1)$ and $(R2, D2)$ have $R1 \neq R2$ and $D1 \neq D2$. In MDSQ, the central quantiser is also characterized by a different rate-distortion pair $(R0, D0)$. In order to obtain unbalanced descriptions, a new type of index assignment table is proposed to allow dynamic changes in the number of central coefficients that are coded as $i_n = 0, n = 1, 2$.

Table 4.2: Unbalanced MDSQ, $k=1$ and $Z = 0$

		i2 (Description 2)									
		-4	-3	-2	-1	0	1	2	3	4	5
i1 (Description 1)		-11									
	-3	-9	-7								
	-2	-8	-6	-5							
	-1		-4	-3							
	0			-2	-1	0	1	2			
	1							3	5		
	2							4	6	7	
	3								8	9	11
	4									10	12
	5								

The proposed unbalanced index assignment method is defined by k , which defines the number of diagonals $(2k + 1)$ and the central index spread variation parameter Z , respectively. The total spread is expressed by $S = 2 * (2k) + 1 + Z$. The spread S is the number of central indices that are encoded as zero in a given description $i_n = 0, n = 1, 2$. Different values of S can be defined by varying Z and, consequently, several index assignment tables can be used to dynamically unbalance each description. Table 4.2 shows an example of the index assignment mapping with $k = 1$ (3 diagonals) and $Z = 0; S = 5$ ($i_1 = 0$ for central indices $i_0 = -2, -1, 0, 1, 2$).

4.3 MD Video Architecture

Starting with the open-loop MDSQ video encoding, a novel MD architecture is proposed for Advanced Video Coding using MDSQ, based on the three-loop scheme. A new mixed open-loop/multi-loop structure is proposed in order to achieve the best tradeoff between MDC coding efficiency and drift control at decoder. The proposed architecture uses the three-loop scheme for anchor slices, namely, I and P slices, while B slices use the open-loop scheme without adding side information. This approach allows to reduce the side information redundancy, maintaining the drift control since B frames are not used as reference frames in most coding algorithms. Therefore, the proposed architecture is able to deal with drift distortion that is propagated in MDSQ open-loop architectures, and at the same time, to reduce the redundancy introduced by the traditional three-loop architectures.

4.3.1 Open-loop MD Encoder

Fig. 4.1 shows the H.264/AVC open-loop MDSQ video encoding scheme. This is an embedded MDC architecture which is implemented by inserting the MDSQ index assignment module described in sections 4.1 and 4.2 in the encoding loop. Index assignment is operated to the quantized DCT coefficients, where each one is mapped into two indices (i_1, i_2) . In order to generate two different descriptions, each index is independently entropy encoded. At the same time, the other information, such as slice maps, prediction modes and motion vectors are encoded in both descriptions, generating two independent decodable bitstreams. Since encoding loop is independent on indices (i_1, i_2) , the other duplicated information is exactly the same in each description.

The output rate of each description depends on the chosen MDSQ scheme. Traditionally, these are balanced, i.e., with identical rates and distortion levels. However, new MDSQ schemes are proposed generating distinct rates e consequently, different distortion for each description, i.e., unbalanced. The choice of MDSQ scheme, and the correspondent index assignment tables depends on the target balancing ratio, i.e., the percentage of rate that is distributed in each description. Therefore, the individual target rate on each description needs to be included in MD encoding rate control schemes. Taking into account these needs, a new unbalanced rate control method for MDSQ is proposed in this thesis, as described in Chap. 6.

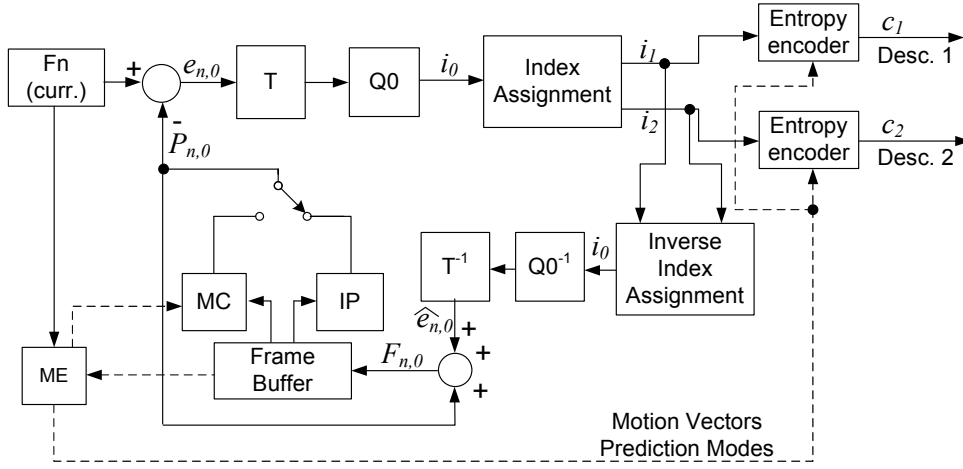


Figure 4.1: Open-loop MDSQ encoder.

4.3.2 Open-loop MD Decoder

The decoder architecture for MDSQ open-loop decoding is represented in Fig. 4.2. Descriptions 1 and 2 are composed by its main streams $c_{i,i=1,2}$. If both descriptions are available, then an Inverse Index Assignment is made to restore the central description index, and then the inverse quantisation and inverse transform are applied. The decoded residue \hat{r}_0 , is added to the corresponding prediction. If any description is lost, the side decoder will decode $\hat{r}_{i,i=1,2}$, using an inverse index assignment with only one description. Thus, the corresponding reconstructed frames $\hat{F}_{n,i}$ are decoded with a poorer quality if compared with the obtained when both descriptions are received. The reconstructed frames $\hat{F}_{n,i}$ is defined as

$$\hat{F}_{n,i} = P_{n,i} + \hat{r}_i, i = 1, 2. \quad (4.1)$$

The side decoders use only one description to decode video sequence. Therefore, adding to the poor reconstruction quality given by coarser quantization, the influence of drift distortion in side distortion needs to be evaluated, because prediction error does not match with the one that is obtained at encoder. This problem is exposed and evaluated in the next section.

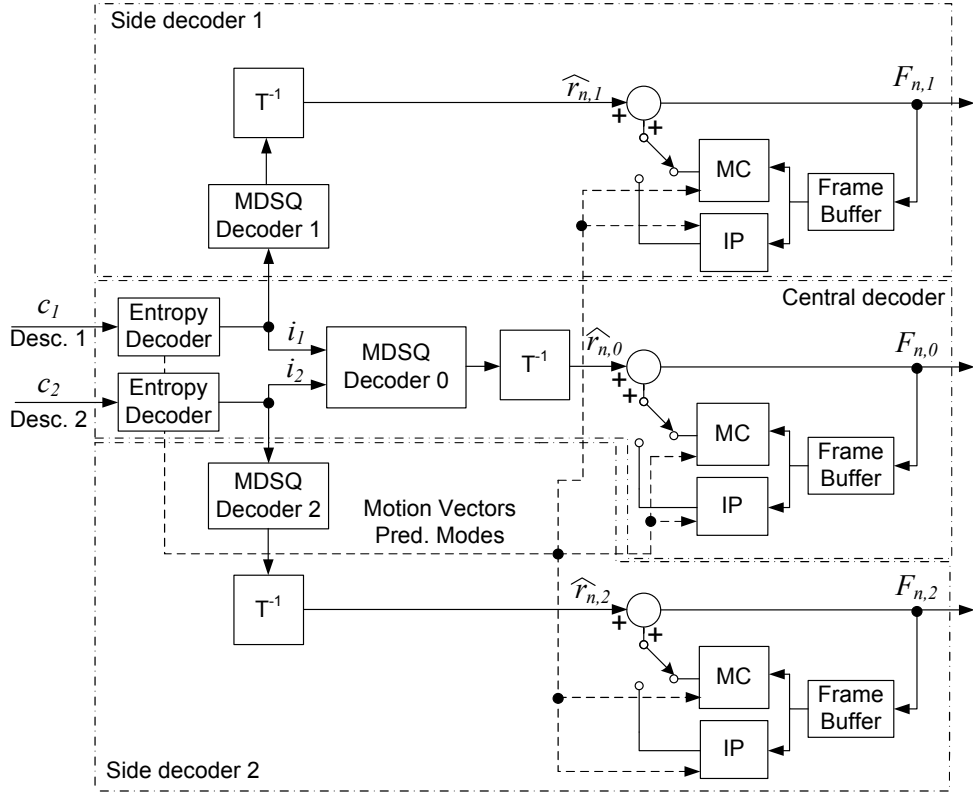


Figure 4.2: MDSQ open-loop video decoder.

4.3.3 Drift Analysis

If the MDSQ open-loop encoder scheme of Fig. 4.1 is used to encode two descriptions of the same source video, then drift is introduced at the decoder whenever any description is lost. The drift distortion component can be determined from the relevant signals involved in encoding and decoding. When two descriptions are received, both side indices (i_1, i_2) are decoded and merged into the corresponding central index i_0 . In this case, for each block n , the reconstructed central pixel values $F_{n,0}$ are given by,

$$F_{n,0} = r_{n,0} + P_{n,0} \quad (4.2)$$

where $r_{n,0}$ is the decoded residue and $P_{n,0}$ its associated prediction either from intra prediction or motion compensation, formed from decoding both descriptions. If only one description j (either $j = 1$ or $j = 2$) is decoded, then the reconstructed pixel values $F_{n,j}$ are given by,

$$F_{n,j} = r_{n,j} + P_{n,j}, \quad (4.3)$$

where $r_{n,j}$ is the decoded residue and $P_{n,j}$ its prediction formed from decoding description j only. Since $r_{n,j}$ results from inverse index assignment using with only one description as input, the difference between the central encoder and that decoded from only one description produces a reconstruction error ε , i.e.,

$$r_{n,j} = r_{n,0} + \varepsilon. \quad (4.4)$$

Substituting (4.4) in (4.3),

$$F_{n,j} = r_{n,0} + \varepsilon + P_{n,j}, \quad (4.5)$$

and, then using (4.2) in (4.5), $F_{n,j}$ becomes,

$$\begin{aligned} F_{n,j} &= F_{n,0} + P_{n,j} - P_{n,0} + \varepsilon \\ &= F_{n,0} + d_P + \varepsilon, \end{aligned} \quad (4.6)$$

where

$$d_P = P_{n,j} - P_{n,0}, \quad (4.7)$$

is the drift component due to mismatch between the predictions used in the encoder and those reconstructed at the final decoder from only one description. Note that the above analysis is valid for both spatial and temporal drift components, though these can be identified as separate contributors to the overall drift distortion.

The actual impact of the drift component given by (4.7) and the objective video quality was experimentally evaluated for the open-loop scheme of Fig. 4.1. The results for intra predicted and motion compensated (MC) frames, are shown in Figs. 4.3 and 4.4, respectively.

Intra Prediction - I Frames Fig. 4.3 shows the drift effect over one intra frame from the *coastguard* sequence, encoded using H.264/AVC with all intra prediction modes enabled. The Peak Signal Noise Ratio (PSNR) is shown for each macroblock decoded from only one description and also from both of them. The SDC trace provides a reference for comparison at the same rate as the single description (1.59 bpp). Fig. 4.3 shows that drift distortion introduced by open-loop MDSQ encoder of an I frame yields unacceptable quality when only one description is decoded. The more macroblocks are decoded, the higher is the accumulated distortion, leading to continuous drop of PSNR along each row of macroblocks. The peaks in PSNR correspond to reset the accumulated drift distortion

to zero at the beginning of each row of macroblocks because the first macroblock is not predicted from the previous ones.

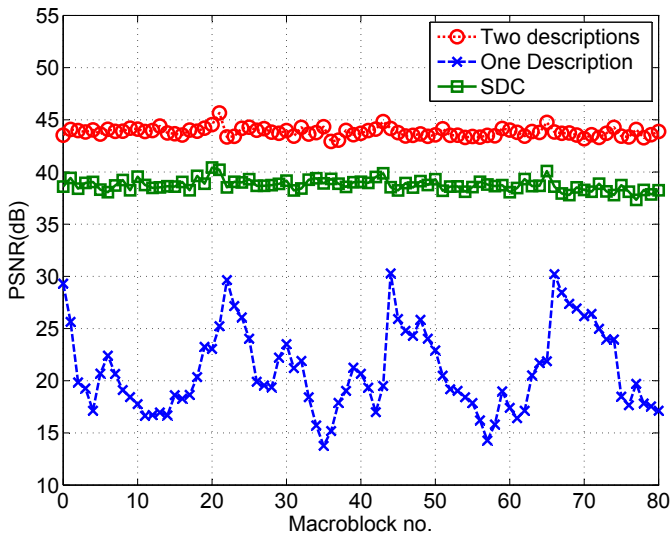


Figure 4.3: Distortion accumulation within an intra frame - *Coastguard*

MC prediction - P Frames Fig. 4.4 shows the effect of drift accumulated over one GOP, comprised of one initial I frame followed by 20 P frames. Both descriptions of the initial I frame are fully decoded in order to avoid drift in subsequent P frames. The PSNR is shown for each frame decoded from only one description and also from both of them. For comparison, the PSNR of the SDC stream using the same bit rate as that of a single description (3.9Mbit/s) is also shown. Fig. 4.4 shows that drift distortion introduced by open-loop MDSQ encoder over one GOP also yields unacceptable quality when only one description is decoded. The effect of drift is quite evident from the rapid decrease of PSNR due to distortion accumulation along the GOP. Since the SDC encoded sequence has the same rate as the single description, the continuous decreasing of PSNR observed in the latter does not depend on the actual coding rate. Rather it is due to prediction mismatch between encoder and decoder.

The above analysis and results clearly show that open-loop MDC encoder cannot be used without drift compensation. An efficient solution is described in the next sections.

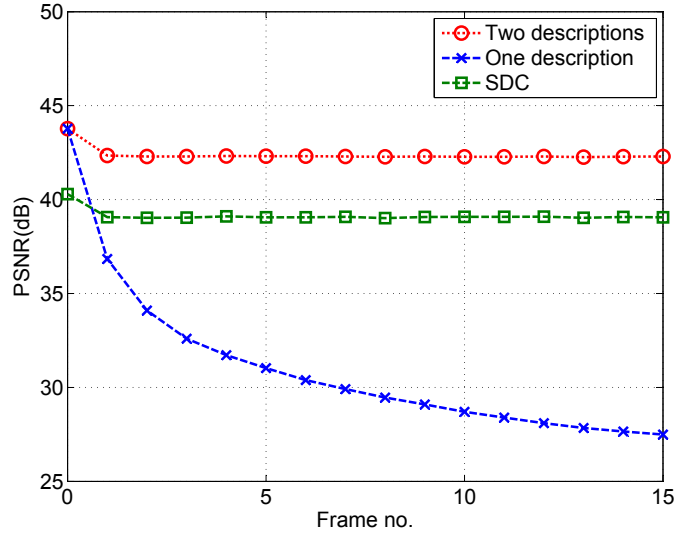


Figure 4.4: Distortion accumulation in MC predicted frames-*Coastguard*

4.3.4 Multi-loop MD Encoder

Figure 4.5 shows the proposed H.264/AVC drift-free MD video encoding scheme is based on a three prediction loop architecture. This MD architecture uses MDSQ in order to generate two distinct independent decodable H.264/AVC bitstreams that can be combined in order to achieve different output quality levels. Each description is composed by the corresponding MDSQ stream and also by the side information that is generated for drift compensation at the decoder. For I and P frames, the drift-free MD for advanced video encoder is composed by one central encoder and two side encoders. The central encoder includes the MDSQ module that produces two descriptions and the corresponding streams. These two distinct bitstreams are run-length and entropy encoded. All the syntax elements, namely headers and prediction modes are duplicated in both descriptions.

The two side encoders control the redundant information which is generated as side information in order to guarantee a drift-free acceptable decoded frame.

The side information s_i can be defined as,

$$s_i = Q_i\{T\{F_n - P_{n,i} - \hat{r}_i\}\}, i = 1, 2, \quad (4.8)$$

where Q_i is the quantisation operation with side quantiser QP_i that determines the amount

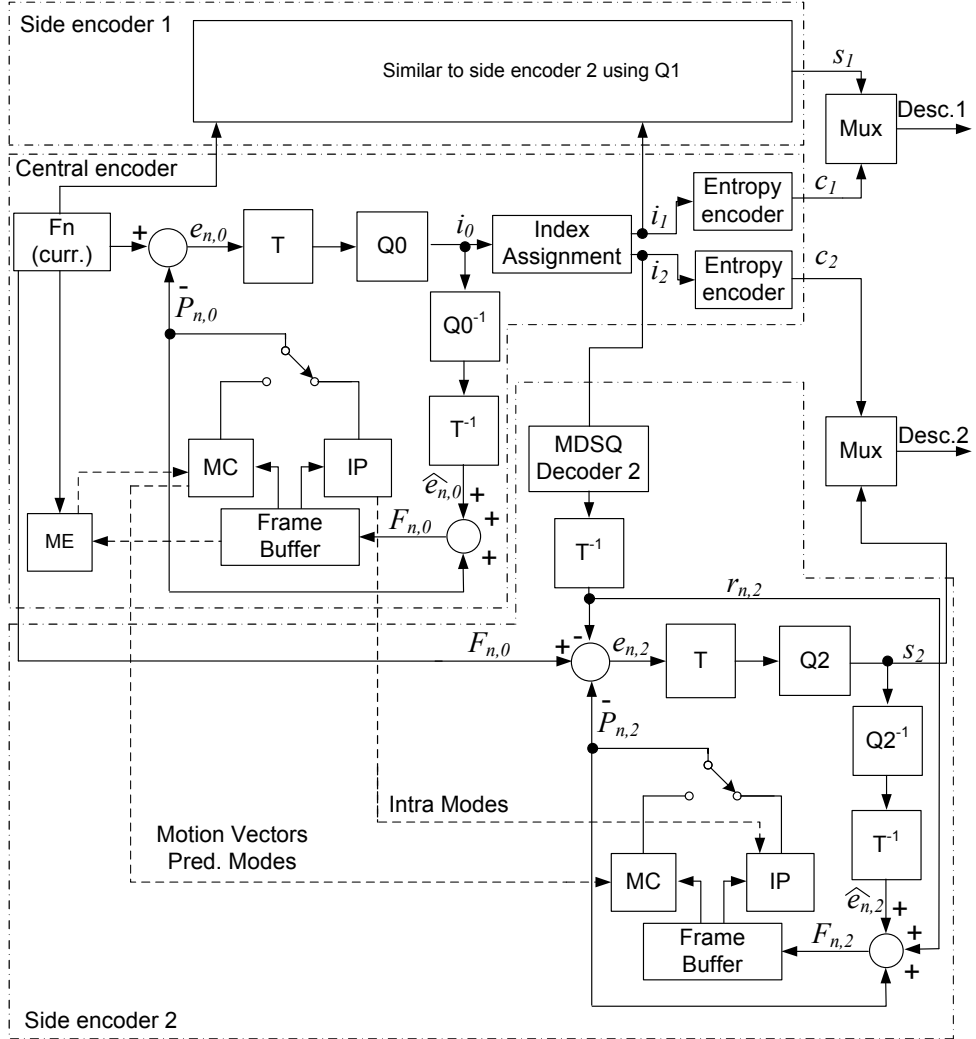


Figure 4.5: MDSQ encoder with drift compensation.

of redundant information, and T is the transform operation. The current frame F_n and $P_{n,i}$ are the prediction value from each respective side encoders. \hat{r}_i is defined as

$$\hat{r}_i = T^{-1}\{Q_0^{-1}\{A_i^{-1}(r_i)\}\}, i = 1, 2, \quad (4.9)$$

which represents the residue available at decoder if only one description is correctly received. The mapping A_i^{-1} represents the inverse index assignment operation if only one description exists. In this case, it is necessary to define a reconstruction rule for each side index because each one corresponds to several possible central indices. In this work, the main diagonal central value is used as reconstruction rule. Operation Q_0^{-1} is the inverse quantisation with the central encoder quantisation parameter QP_0 .

This side information is considered as redundant because it is used for decoding of only one description and represents the residual that compensates the prediction mismatch accumulation between encoder and decoder when only one description is decoded. Nevertheless, only the residual signal is encoded, without further information such as prediction modes and motion information. This is because side information is encoded with the same modes as central information and these elements have already been duplicated in both descriptions. The amount of side information depends on the quantization parameter of each side encoder. Since \hat{r}_i is the residual signal to compensate mismatch accumulation in each description, the prediction modes are the same as used in central encoder. Therefore, some coding efficiency is lost because the prediction modes cannot be adapted to quantization parameter level that needs to be chosen to achieve an acceptable redundancy level. However, this loss of coding efficiency is not very relevant since redundancy level is relatively small.

4.3.5 Multi-loop MD Decoder

The decoder architecture is represented in Fig. 4.6. Descriptions 1 and 2 are composed by its main streams and correspondent side information $s_i, i = 1, 2$. If both descriptions are available, then an Inverse Index Assignment is made to restore the central description index. Then inverse quantisation and inverse transform are applied. If some description is lost, for I and P frames, the side decoder will decode a drift-free frame, though with poorer quality than if both descriptions were received. The reconstructed frame $\hat{F}_{n,i}$ can be defined as

$$\hat{F}_{n,i} = P_{n,i} + \hat{r}_i + T^{-1}\{Q_i^{-1}\{s_i\}\}, i = 1, 2. \quad (4.10)$$

The drift-free decoded frames are used to decode the subsequent central frames whenever both descriptions are correctly decoded. For B frames side decoding is used in open-loop, which means that for each decoded residue, $s_i = 0, i = 1, 2$ and mismatch between encoder and decoder is introduced. Drift propagation is mitigated by I and P frames, since usually B frames are not used as reference.

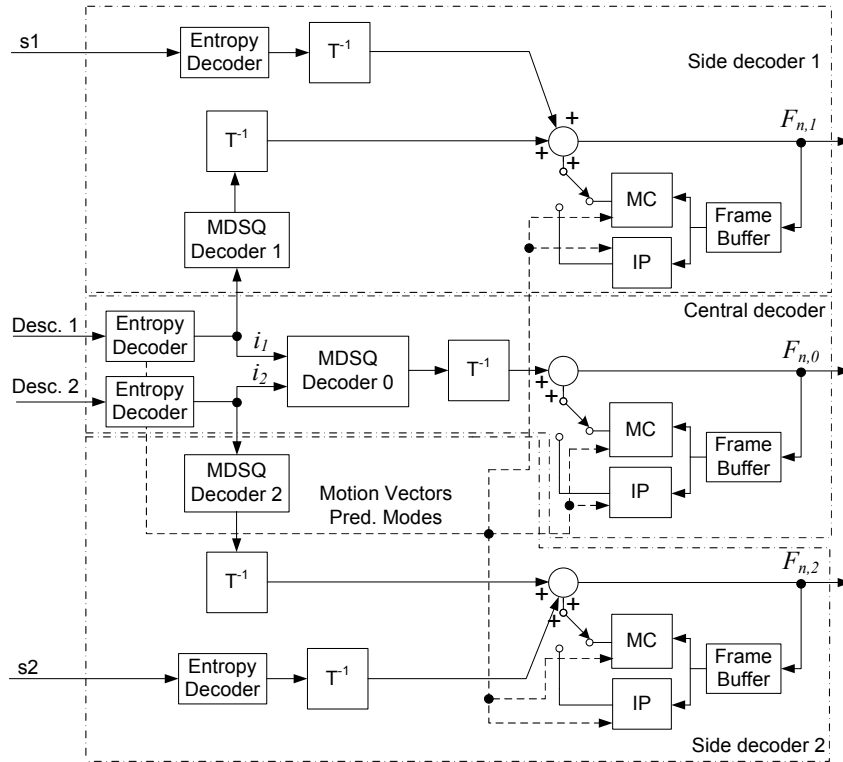


Figure 4.6: MDSQ video decoder.

4.4 R-D Performance - Two Descriptions

This section evaluates the R-D performance of the proposed MDSQ for advanced video coding for joint decoding of two descriptions. The R-D performance of the proposed MDSQ scheme is compared with SDC. Thus, the quality penalty and redundancy levels that are introduced by MDSQ schemes comparing with SDC is evaluated. In regard to this aspect, an effective study of the redundancy components introduced by each MDSQ scheme is important to evaluate the impact of each component, finding solutions to improve the MDSQ coding efficiency. Only redundancy introduced by MDSQ is considered on this evaluation because if two descriptions are jointly decoded, then no side information is used. In other words, this study is done using the open-loop MDSQ embedded architecture for balanced and unbalanced MDSQ. Therefore, an important issue that needs proper evaluation is the redundancy obtained for different rate regimes and its components, such as, the influence of duplicated information on each description. It is also important to know the behavior of texture encoding in MDSQ for different quantization parameter values.

4.4.1 Balanced MDSQ

The R-D performance evaluation compares different balanced MDSQ index assignment tables with 2, 3, 5, and 7 diagonals for open-loop architecture described in section 4.3. MDC performance is compared with single description coding (SDC) and also with duplicated SDC. Headers, prediction modes and motion vectors are duplicated in both descriptions which are included in the overall rate. Evaluation is done using sequences *foreman* and *bus*, CIF format, 15Hz. The resulting rates are obtained from coding with fixed quantization parameters QP between 10 and 40, without using rate-control schemes. Also, the redundancy obtained with the proposed scheme is evaluated by comparing its rate performance with SDC. The use of MDSQ in intra frames and motion prediction frames has different relevant aspects to be considered since behind the residual redundancy components, the rate resulting from motion information in motion predicted frames is duplicated. However, the weight of motion information in the global rate needs to be evaluated since this is mainly dependent on the sequence contents and quantization parameter levels.

R-D for Intra Frames Figs. 4.7 and 4.8 shows the R-D performance for intra frames. The total rate is composed by duplicated headers, prediction modes and MDSQ indices for each description. From the graphical analysis, evaluation of each MDSQ scheme is compared with the SDC curve, which is obtained considering that each frame is repeated in both descriptions, corresponding to duplicate the SDC stream. This is the MDC redundancy limit, since one source can be repeated for transmission over different channels without losing the SDC quality. Comparing the MDSQ performance for index assignment with different diagonals, it is seen that at low rates, no differences in reconstructed quality are noted. However, at higher rates, in particular higher than 1 bit per pixel (bpp), using 2 diagonals, some losing of coding efficiency is seen in the figures, where the PSNR value is worse about 1dB at 1.5 bpp comparing with MDSQ with 3 diagonals. Evaluating the comparison with 3, 5 and 7 diagonals, the decoding quality with two descriptions are approximately the same, which means that the overall MDSQ redundancy has the same order of magnitude. Clearly, the amount of side distortion that is obtained in each solution must be measured, which is done in the next section. On the other hand, it is noted that the PSNR difference between SDC and MDSQ, is around 1.5dB for higher rates and 2dB for lower rates, which means that redundancy increases at lower rates. Nevertheless, by comparing with the difference that is obtained comparing with duplicated SDC, this is around 5-6 dB.

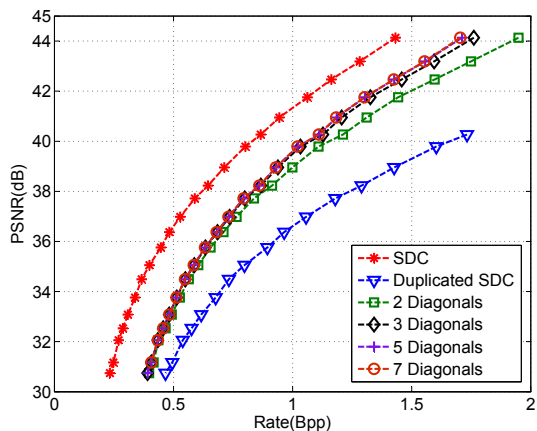


Figure 4.7: Central Distortion *Foreman* sequence, I Frame.

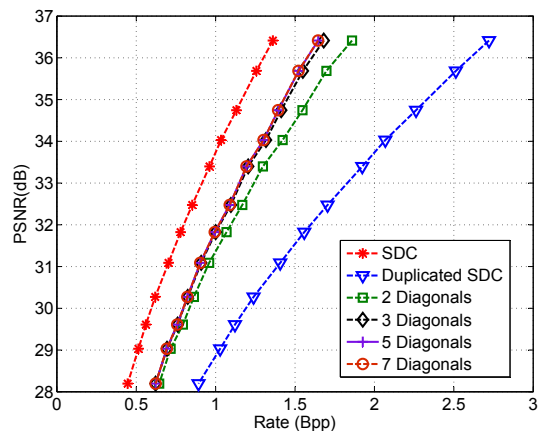


Figure 4.8: Central Distortion *bus* sequence, I Frame.

The PSNR results obtained from decoding of two descriptions is directly related to the MDSQ redundancy obtained for each scheme. Since, redundancy is the excess rate introduced by MDSQ scheme, it can be measured as a percentage, given by,

$$\psi = \left(\frac{R_1 + R_2}{R_{SDC}} - 1 \right) \times 100, \quad (4.11)$$

where $R_i, i = 1, 2$ is the rate of each description and R_{SDC} is the rate of the SDC stream.

Other perspective of earlier results are represented in Figs. 4.9 and 4.10 where redundancy percentage levels are shown for different quantization parameters. Redundancy levels are mainly dependent on the excess rate introduced by MDSQ in residual information. From theoretical point of view, MDSQ redundancy controlled by the number of diagonals in index assignment matrix. Experimental results for advanced video coding show that this is effectively true, mainly for lower values of quantization parameter. Also, MDSQ with 2 diagonals has a distinct behavior when compared with MDSQ using higher number of diagonals. With 2 diagonals the redundancy values are around 45-55% for higher rates. In this case, despite the relatively high redundancy, with the increase of QP, the redundancy level decreases. Then, for lower rates, the redundancy tends to increase. The other group of MDSQ schemes were developed using 3, 5 and 7 diagonals. Firstly, the case of QP below 25 is analyzed, where the redundancy is maintained below 30-35% for 3 diagonals and 20-30% for 5 and 7 diagonals.

Mainly for 2 and 3 diagonals, the redundancy decreases when QP increase. In finer

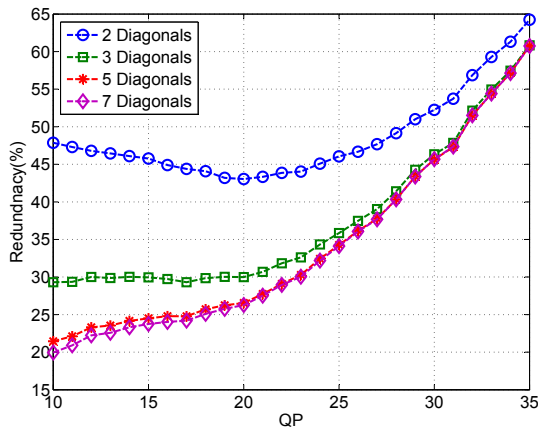


Figure 4.9: Redundancy of *Foreman* sequence, I Frame.

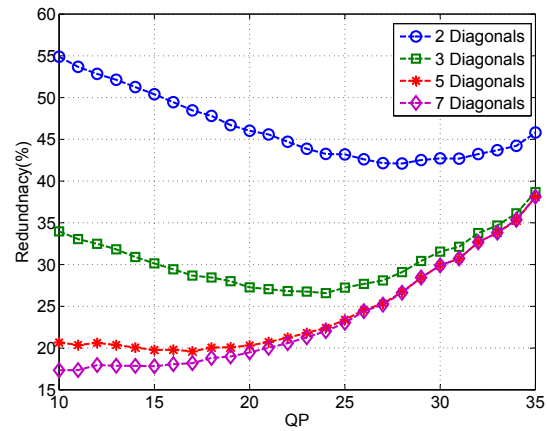


Figure 4.10: Redundancy of *bus* sequence, I Frame.

quantization, the number of non-null coefficients is very high which in turn reduces the coding efficiency resultant from index assignment, mainly due to source statistics related to efficiency of entropy encoding. Secondly, for coarser quantization, coding efficiency is increased because the number of null indices increase with the same number of diagonals in index assignment, turning run-length encoding more efficient. This phenomena is not so evident for higher number of diagonals, because this corresponds to a coarser quantization where fewer coefficients are indexed as zero.

The redundancy rapidly increases with higher QP values, namely for $QP > 35$, independently from the number of diagonals. There are two main reasons for this fact. First is that the redundancy of residual information highly increases, regardless of the number of diagonals. The reason is that for higher QP's, the central quantization step-size is sufficiently large making side quantization less efficient. This is because MDSQ is applied to a small number of coefficients, whose values are similarly indexed independently of the number of diagonals. This results in a decreasing number of non-null transform coefficients in SDC mode and, consequently, a relative rate increase. Other reason for this is that the relative importance of repeated information increase with coarser quantization. This fact is illustrated in Fig. 4.11, that shows the rate percentage of each component of SDC encoding. The residual information is nearly 100% for $QP < 20$ and correspondent percentage decrease rapidly for $QP > 25$. The remaining information, is mainly due to the encoding of *cbp* that increases with the number of null blocks, which is approximately 10% for higher QP values. It can be seen in the figure that the overall remaining information is approximately 20% of the total rate, which means this is an important component

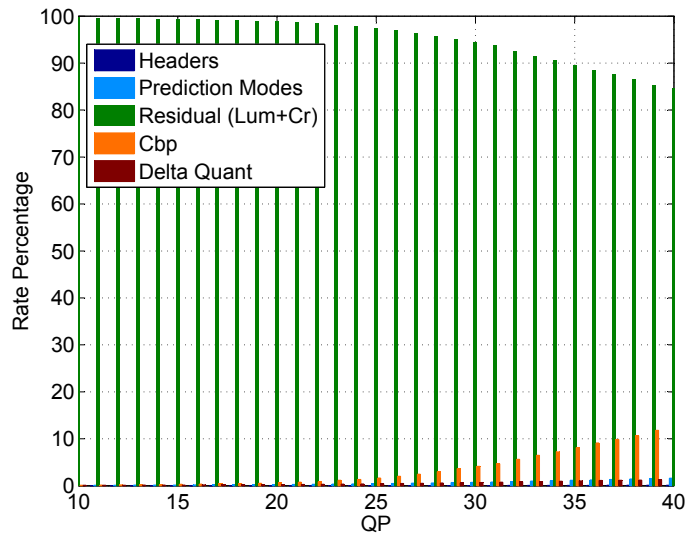


Figure 4.11: Rate distribution for one I Frame SDC.

of redundancy in coarse quantization.

R-D performance for motion predicted frames The rate-distortion performance of the proposed scheme is found for motion prediction frames using a GOP structure IPPB..., with GOP=16 frames. The experimental evaluation is done using sequences *foreman* and *bus*, CIF format, 15Hz. The resulting rates are obtained from coding with fixed quantization parameters QP between 10 and 40, without using rate-control schemes. MDSQ is applied to transform coefficients, and other information is repeated in both descriptions.

Figs. 4.12 and 4.13 show the R-D performance for IPBBP GOP structure where the overall rates include duplicated headers, prediction modes and motion information. Comparing the different MDSQ schemes, they show similar R-D performance for index assignment matrices with 2, 3, 5 and 7 diagonals. These figures also, show that the quality obtained for different MDSQ schemes is very similar, which leads to conclude that the amount of redundancy is on the same order of magnitude. The PSNR difference is around 2dB, which indicates a relatively high redundancy. Using 2 diagonals, the PSNR decrease for higher rates if compared with the others MDSQ schemes. Furthermore, at lower rates, the contribution of motion information for the overall rate increase and the resulting rates become approximately by the twice of SDC, which is the maximum for MDC R-D overall performance. Therefore, MDSQ with 2 and 3 diagonals can be a good solution to be used in advanced video coding, because better side-distortion performance is expected in

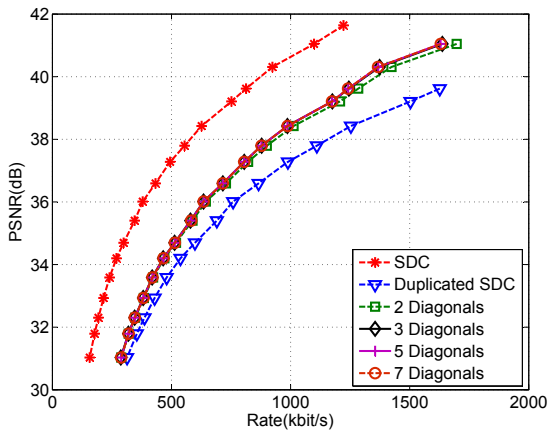


Figure 4.12: Central distortion *Foreman* sequence, IPBB Frame.

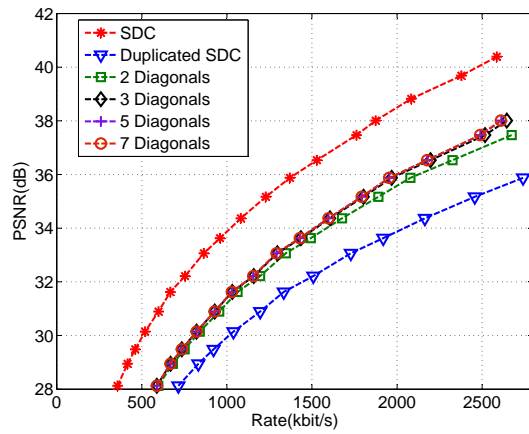


Figure 4.13: Central distortion *bus* sequence, IPBB Frame.

comparison with schemes with more diagonals. This is shown in the next section.

Looking at the redundancy levels for these sequences, Figs. 4.14 and 4.15 show that the redundancy is relatively small in the regions of high rate, and for $QP > 30$, the redundancies are very similar in MDSQ with 3, 5 and 7 diagonals. Similarly in Intra frames, the redundancy for MDSQ with 2 diagonals is higher than in the other schemes.

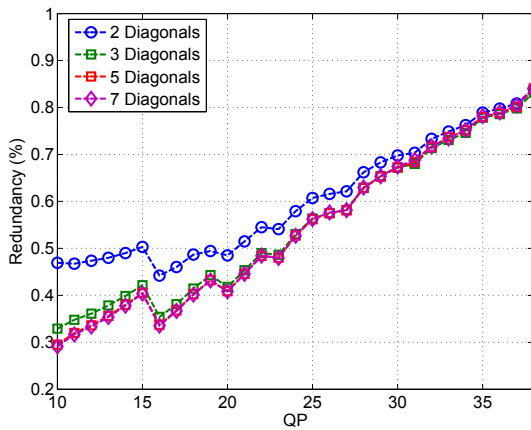


Figure 4.14: Redundancy of *Foreman* sequence, IPBB Frame.

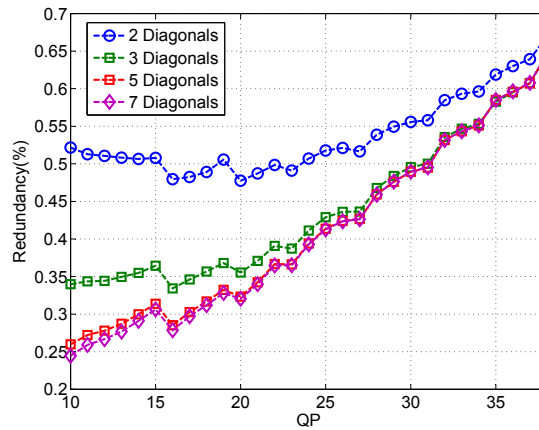


Figure 4.15: Redundancy of *bus* sequence, IPBB Frame.

To better understand the redundancy level and its components, Figs. 4.16 and 4.17 show the components of the overall rate including the motion information in P and B frames respectively. These graphics clearly show that in motion compensated prediction, the motion information and prediction modes have a relevant weight in the overall rate, which increases in B frames. As the figures show, for higher QP values, MDSQ is applied to only half of the SDC rate, which means that repeated information has an important role in the overall redundancy. For instance, in Figs. 4.16 and 4.17, is shown that using $QP = 35$, the rate percentage for texture information (luminance and chrominance) is 60% (P frames) and 50% (B frames) of the total rate. At higher rates, for instance using $QP = 20$, the texture information rate is nearly 75% and 78% the total rate, for P and B frames respectively, which means that the duplication effect of motion information, prediction modes and headers is slightly mitigated.

Furthermore, the characteristics of MDSQ highlighted for Intra frames, are quite similar for motion predicted frames, which means that the overall redundancy is relatively high, mainly due to the repetition of the motion information and prediction modes. Overall, MDSQ is more suitable for high rates, where finer step-sizes are used and more differentiation between methods and redundancy are achieved. Nevertheless, MDSQ provides efficient tools to be used as a valid MDC technique for video coding for a wide range of rates and redundancy.

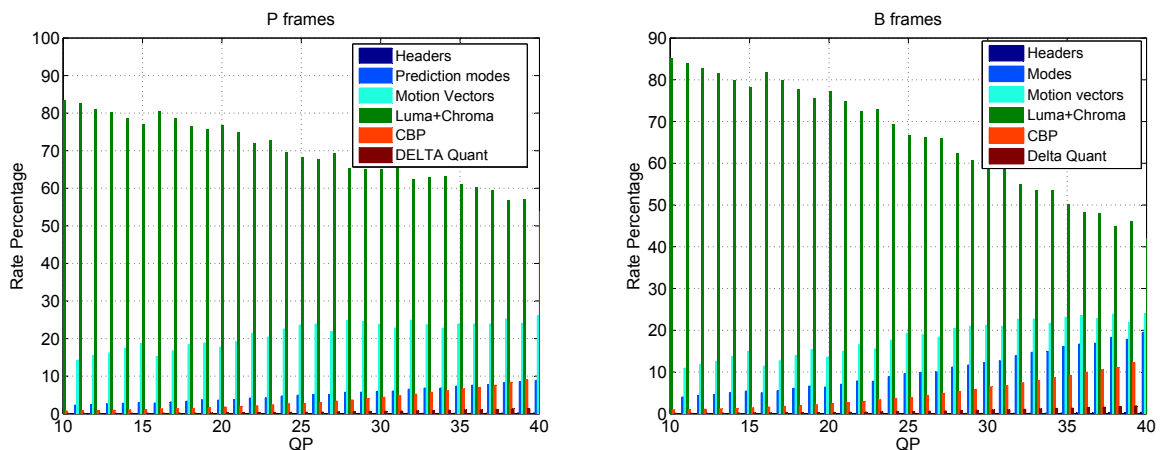


Figure 4.16: Rate distribution for P Frames-SDC. Figure 4.17: Rate distribution for B Frames-SDC.

4.4.2 Unbalanced MDSQ

Unbalanced MDSQ is based on two main parameters, *i.e.*, the number of diagonals and the value of Z which determines the balancing ratio between descriptions. Taking the results of balanced MDSQ as reference, this section presents an evaluation of the unbalanced MDSQ performance for a given number of diagonals used in index assignment matrices. Unbalanced MDSQ using 3 diagonals for several balancing ratios is used in this study. The choice of 3 diagonals is because it presents a good tradeoff between redundancy and distortion when compared with the other number of diagonals used in balanced MDSQ index assignment matrix.

Balancing ratio Figs. 4.18 and 4.19 show the balancing ratio percentages considering the total rate $R = R_1 + R_2$ and $R_1 \neq R_2$ as functions of the Z parameter. For I, P and B frame types, headers, motion vectors and macroblock modes are duplicated. It is noted that by varying the Z parameter from -1 to 3, the rates of each description are set to different rate balancing ratios. Analyzing the results, one can conclude that the balancing ratio among descriptions can reach variations in about 10 – 15% for several values of Z . Defining (π_1, π_2) as the balancing percentages of the total rate for description 1 and description 2, respectively, this means that the proposed set of index assignment tables are capable of producing unbalanced descriptions between (50, 50) (balanced case) and (25, 75) for different types of frames.

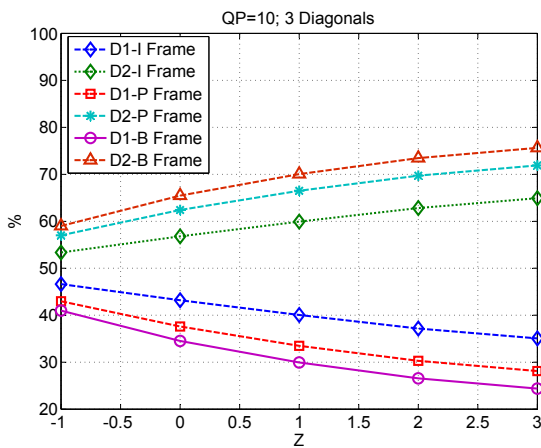


Figure 4.18: *Bus* sequence: $QP=10; k=1$.

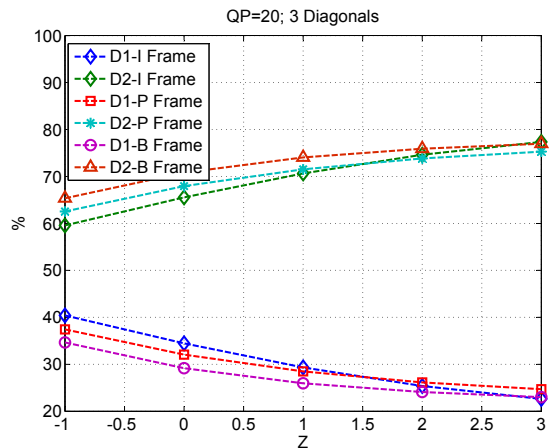


Figure 4.19: *Bus* sequence: $QP=20; k=1$.

Intra Frames Figs. 4.20 and 4.21 show the R-D performance for Intra frames. Total rate is composed by duplicated headers and prediction modes and MDSQ indices for each description. Evaluation is made by comparing SDC, balanced and unbalanced MDSQ for different balancing ratios, obtained by controlling the Z parameter. Considering different values of Z , PSNR values are closer to SDC when balancing ratio increases, which means that redundancy is reduced for higher values of Z . In regard with balanced MDSQ, the PSNR obtained for decoding of both descriptions is in the same range of values. Nevertheless, better quality is achieved for unbalanced MDSQ comparing with balanced MDSQ. Clearly, such differences are more evident for higher rates. At lower rates, the overall performance is very similar.

Figs. 4.22 and 4.23 show the redundancy results for each sequence. Redundancy for lower QP values is greater than balanced MDSQ, regardless the Z value, which is noted for $QP < 25$. When the QP values increase, the unbalanced MDSQ redundancy decreases consistently. This is because, the Z parameter directly affects the indexed coefficients in each description, increasing the coding efficiency, close to SDC in high frequency coefficients. On the other hand, using higher QP values, the redundancy is mainly affected by duplicated information rather than transform coefficients. Another important result is that with higher QP values, the redundancy of unbalanced MDSQ is lower than balanced MDSQ. This fact leads to a reduction of robustness of MDC encoding. Nevertheless, this loss can be compensated using a reduced amount of additional information.

IPPB GOP The R-D performance evaluation is carried out by comparing different balancing ratios for MDSQ with 3 diagonals. The MDC performance of unbalanced MDC is compared with balanced MDC, SDC and duplicated SDC. Headers, prediction modes and motion vectors are duplicated in both descriptions which are included in the overall rate. Performance evaluation is carried out using sequences *foreman* and *bus*, CIF format, 15Hz. Fixed quantization parameters are used with QP between 10 and 40, without rate-control schemes.

Figs. 4.24 and 4.25 show the R-D performance using an IPBBP GOP structure. For this case, the overall rates include the motion information in both descriptions. The new unbalanced MDSQ R-D performance is evaluated for several central index spread variation parameters Z in index assignment matrices with 3 diagonals, with $Z = 0, 1, 2, 3$. The figures show that the PSNR results of unbalanced MDSQ are better than balanced MDC, mainly at higher rates. On the other hand, the decoded quality increases with

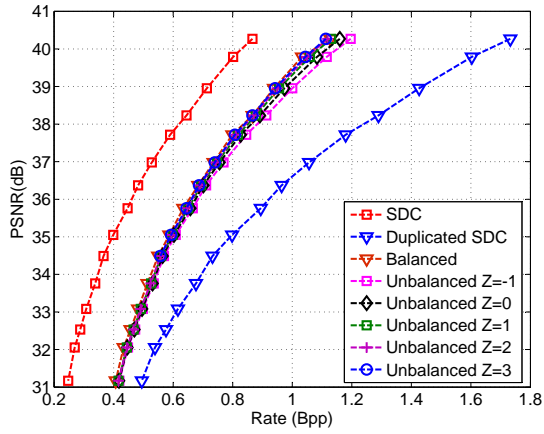


Figure 4.20: Central distortion *foreman* sequence, Intra Frame.

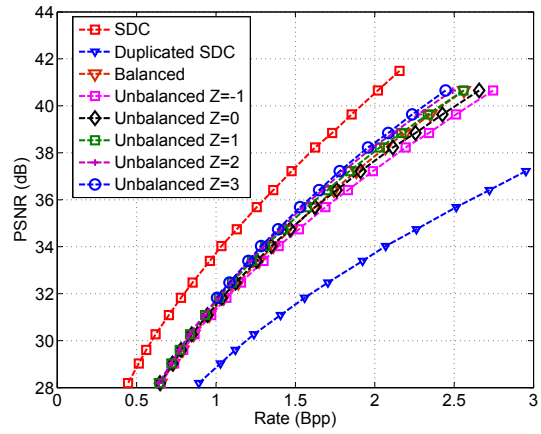


Figure 4.21: Central distortion *bus* sequence, Intra Frame.

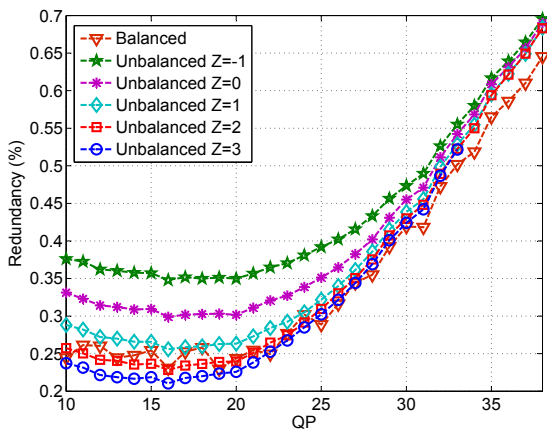


Figure 4.22: Redundancy of *foreman* sequence, Intra Frame

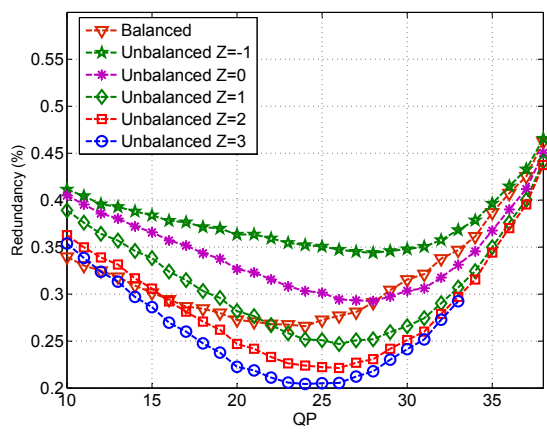


Figure 4.23: Redundancy of *bus* sequence, Intra Frame.

the balancing ratio. This is better noted in the *bus* sequence, but also found in *foreman* sequence. This is because the MDSQ redundancy decreases with the unbalancing ratio, due to better entropy coding efficiency.

Figs. 4.26 and 4.27 show the comparison of redundancy levels for balanced and unbalanced MDSQ. Generally, unbalanced MDSQ has lower redundancy compared with balanced one. Redundancy increases with QP and is higher compared with obtained in Intra coding. This is related to the fact that motion information, prediction modes and headers files, have a non-negligible effect in the global redundancy. For higher QP redundancy is around 50% and 40% for QP=25, for sequences *foreman* and *bus*, respectively. For $QP > 25$ redundancy increase, to 80%-70% with QP=35. On the other hand, the redundancies for different schemes of unbalanced MDSQ are reduced with increasing of Z . Nevertheless, redundancies are remaining at an acceptable value for a wide range of QP values which means that these schemes can be used in many applications.

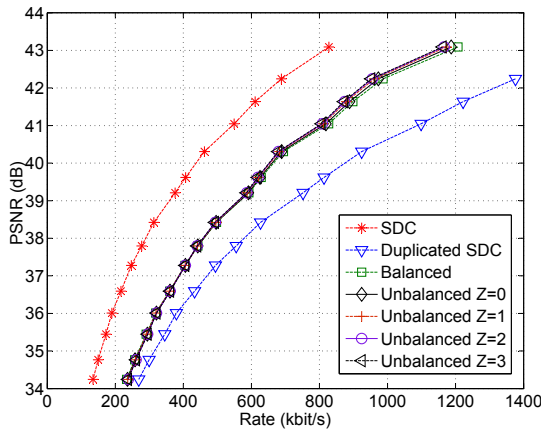


Figure 4.24: Central distortion *foreman* sequence, IPBB Frame.

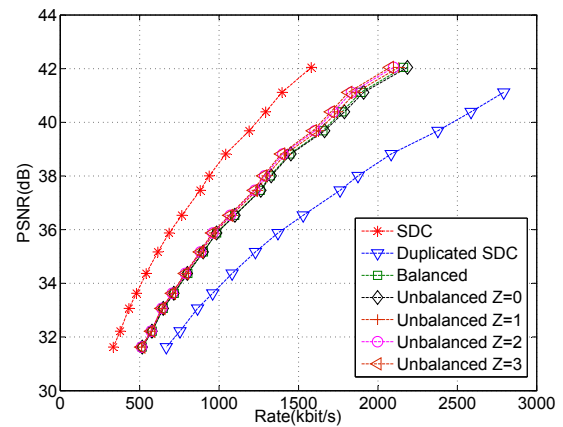


Figure 4.25: Central distortion *bus* sequence, IPBB Frame.

4.5 R-D Performance - Single Description

Single description performance is one of the most important features in MDC schemes since it shows the single description quality for a given rate or redundancy level. R-D performance of the proposed close loop balanced MDSQ (Fig. 4.5) is compared with the state of the art open-loop balanced MDSQ (Fig. 4.2). Furthermore, the novel unbalanced MDSQ R-D performance evaluation is compared with balanced MDSQ.

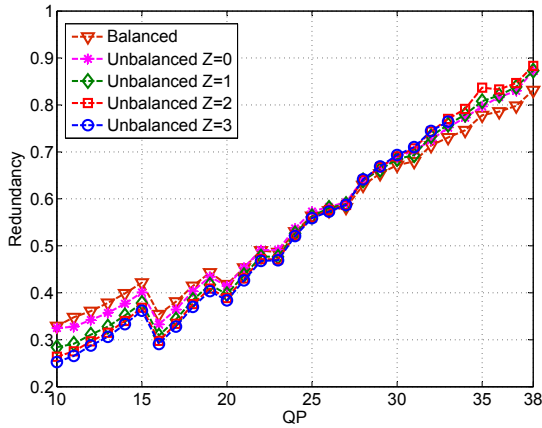


Figure 4.26: Redundancy of *foreman* sequence, IPBB Frame.

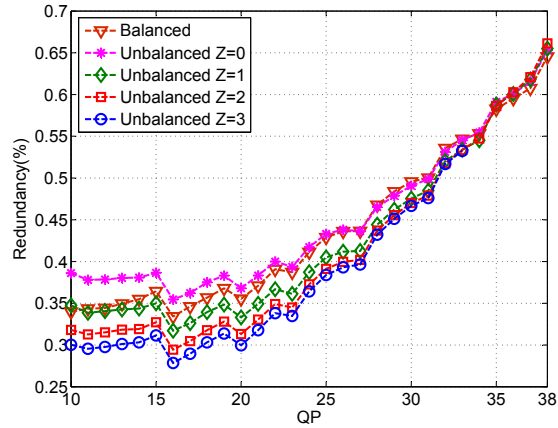


Figure 4.27: Redundancy of *bus* sequence, IPBB Frame.

4.5.1 Balanced MDSQ

The side distortion evaluation defines the MDC performance measured as the PSNR obtained when only one description is decoded. The proposed scheme is compared with the open-loop MDSQ. The excess rate introduced by side information is only used for side decoding, so it is important to study the gains introduced by the additional redundancy in the overall decoding quality. This study is done in Intra frames and also in IPPB GOP structures.

4.5.1.1 Intra Slices

Intra prediction in Advanced Video Coding poses new challenges to MDC schemes, mainly because drift distortion is very significant within each coded frame as it was shown in earlier section. The proposed MDSQ architecture shown in Fig. 4.5 for intra frames is an efficient solution to overcome this problem, by adding side information that allows to have a drift-free reconstruction with an acceptable quality at side decoder when only one description is available. An experimental study of intra drift-free MDSQ architecture is done for different MDSQ schemes, with index assignment matrices with 2, 3, 5 diagonals, where the effectiveness of the solution is shown in comparison with the open-loop architectures. Sequences *foreman* and *bus* are used in the simulations. The side PSNR is defined as the mean value of the PSNR resulting from two descriptions individually decoded, and redundancy results from MDSQ plus side information considering the both descriptions.

The results of the drift-free proposed solution are obtained using different levels of side information for a given central distortion. These results show the side PSNR using central $QP_0=10,15$ and 21. For each QP_0 , several levels of side information resulting from variation of QP_1 and QP_2 are shown.

MDSQ with 2 Diagonals Figs. 4.28 and 4.29 show the side PSNR versus redundancy from proposed Intra MDSQ scheme compared with MDSQ Intra open loop scheme using 2 diagonals. These results show that gains in decoded signals due to side information can be from 5-10 dB depending of the redundancy level. For instance, in Fig. 4.28 shows that using $QP_0=15$, the open-loop side PSNR is 26dB using redundancy approximately 45%. The close-loop scheme using the same $QP_0 = 15$, has side PSNR varying from 30-40 dB, depending on the side encoder QP values, which results in a redundancy around 70%. Generally, the side information redundancies are 10-40% ahead of the MDSQ redundancy increasing with higher QP values because the drift compensation signal error energy also increases, thus more bits are needed for encoding.

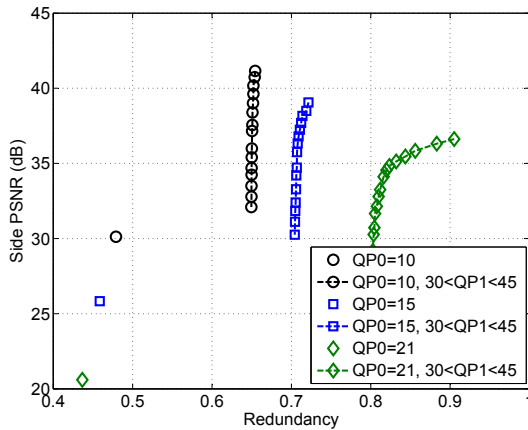


Figure 4.28: Side-distortion *Foreman* sequence, Intra Frame, MDSQ with 2 diagonals.

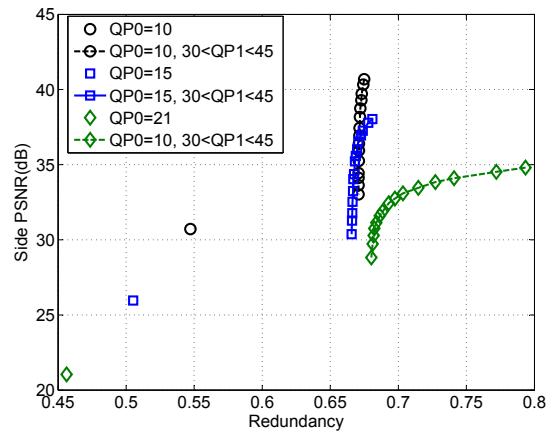


Figure 4.29: Side-distortion *bus* sequence, Intra Frame, MDSQ with 2 diagonals.

MDSQ with 3 Diagonals Figures 4.30 and 4.31 show the proposed Intra MDSQ scheme compared with MDSQ Intra open-loop scheme for 3 diagonals. The results are similar when compared with MDSQ with 2 diagonals. Nevertheless, a lower MDSQ redundancy is achieved and also lower side PSNR values comparing with MDSQ with 2 diagonals. This is because, with the number of diagonals, the prediction error encoded

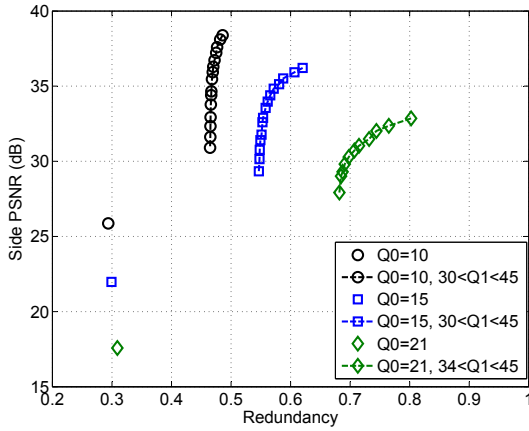


Figure 4.30: Side-distortion *Foreman* sequence, Intra Frame CIF, MDSQ with 3 diagonals.

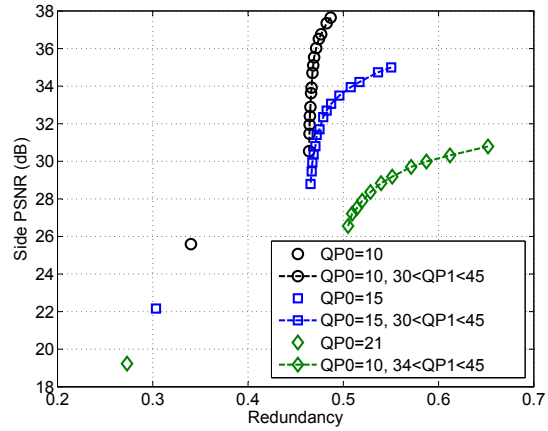


Figure 4.31: Side-distortion *Bus* sequence, Intra Frame CIF, MDSQ with 3 diagonals.

as side information increases, losing coding efficiency. Also, the prediction modes of side-information are the same as the original ones, which means that some coding efficiency is lost due to this fact. Nevertheless, the gains are sufficiently expressive, to conclude that the proposed scheme is an effective solution in order to have side reconstructions with an acceptable quality, when MDSQ with 3 diagonals is used.

4.5.1.2 Motion Compensation Predicted Frames

Performance evaluation of the proposed MDC architecture shown in Fig. 4.5, is done for Motion Compensation (MC) predicted slices. The main novelty of this proposal is the use of drift compensation at anchor slices I and P that are used on prediction for other slices. This allows to reduce the amount of side-information and, at the same time, to control drift-distortion in the whole GOP. In this section a performance evaluation of the proposed scheme is done by comparing MDSQ with 2, 3, and 5 diagonals. Sequences *foreman* and *bus*, CIF format, with GOP type IPBBP and GOP size of 16 frames are used in the simulations. Central encoding was done with fixed quantization parameters QP_0 between 10,15 and 21, and $QP_i, i = 1, 2$ between 25 and 40 without using rate-control schemes. Proposed MDSQ for Advanced Video Coding scheme is compared with MDSQ open-loop.

MDSQ with 2 Diagonals Figs. 4.32 and 4.33 show the R-D performance for *foreman* and *bus* sequences, *CIF* format, for MDSQ with 2 diagonals. In order to reduce the excess rate, only I and P frames use side information. For MDSQ without side information, the side PSNR is highly dependent on the quantisation parameter value. For higher QP, the side PSNR has poor performance. The proposed MDC with side information has performance gains comparing with the open loop approach. By comparing the proposed method with open-loop MDC, the side PSNR gains are about 3-15 dB, depending on introduced redundancy which is 3-15% higher than MDSQ central redundancy. For instance, in Fig. 4.33 shows that using $QP_0=15$, the open-loop side PSNR is 30dB using redundancy approximately 51%. The close-loop scheme using the same $QP_0 = 15$, has side PSNR varying between 33-37 dB, depending on the side encoder QP values, which results in a redundancy between 55-57%. This value is lower than compared with Intra frames because, B frames have not side information which allows to reduce significantly the amount of redundancy. Considering the performance gains that are achieved, the conclusion is that the introduced redundancy has an residual increase in the overall rate.

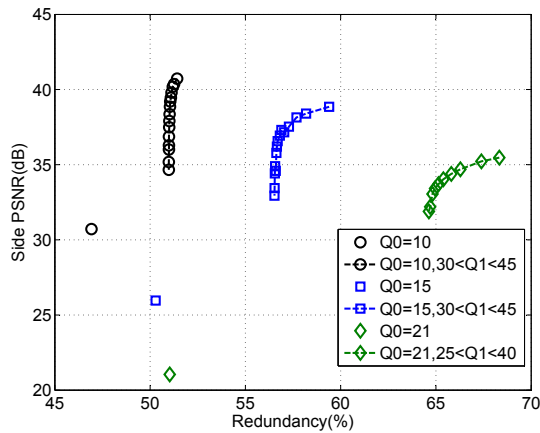


Figure 4.32: Side Distortion *Foreman* sequence, CIF 15HZ, MDSQ with 2 diagonals.

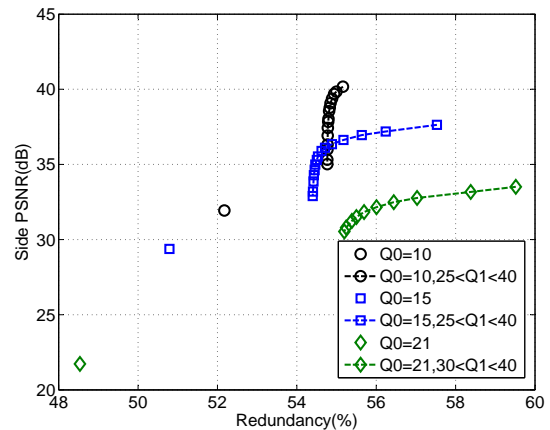


Figure 4.33: Side Distortion *bus* sequence, CIF 15HZ, MDSQ with 2 diagonals.

MDSQ with 3 Diagonals Figs. 4.34 and 4.35 show the R-D performance for *foreman* and *bus* sequences with 3 diagonals. Comparing the proposed method with open-loop MDC, the side PSNR gains are in the range of 5-10 dB, with additional redundancy of 5-20%. For instance, in Fig. 4.35 shows that using $QP_0=10$, the open-loop side PSNR is 28 dB using redundancy approximately 35%. The close-loop scheme using the same

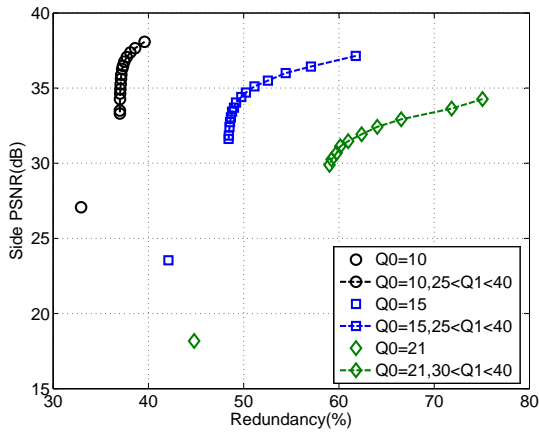


Figure 4.34: Side Distortion *Foreman* sequence, MDSQ with 3 diagonals.

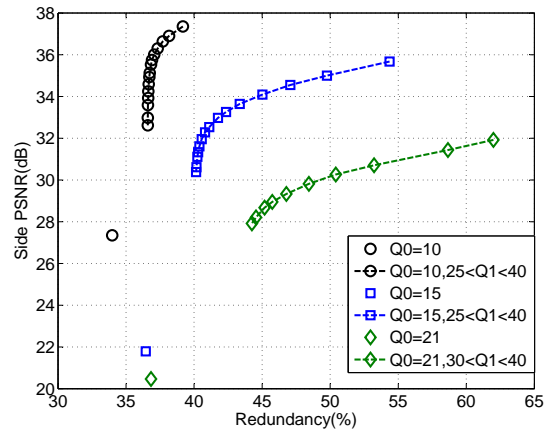


Figure 4.35: Side Distortion *Bus* sequence, MDSQ with 3 diagonals.

$QP_0 = 10$, has side PSNR varying between 32-38 dB, depending on the side encoder QP values, which results in a redundancy between 36-39%, i.e., more 3-4% if compared with open-loop scheme. The total redundancy and side PSNR are lower if compared with MDSQ with 2 diagonals. The redundancy levels are mainly dependent on the central MDSQ component which is lower in this case. The open loop results show that increasing the number of diagonals, the side distortion increases, resulting in a poor reconstruction quality. On the other hand, the coding efficiency of the side information decreases because drift compensation signal increases, resulting in a poor side-information R-D coding performance.

MDSQ with 5 Diagonals Figs. 4.36 and 4.37 show the R-D performance for *foreman* and *bus* sequences for MDSQ with 5 diagonals. The side distortion of open-loop MDSQ is around 15-20 dB which is not an acceptable quality. Comparing the proposed method with MDC open-loop the gains are about 15 dB depending on the redundancy level increase. For instance, in Fig. 4.36 shows that using $QP_0=15$, the open-loop side PSNR is 16 dB using redundancy about 31%. The close-loop scheme using the same $QP_0 = 15$, has side PSNR varying between 26-30 dB, depending on the side encoder QP values, which results in a redundancy between 37-52%, i.e., more 6-19% if compared with open-loop scheme. Comparing with MDSQ with 2 and 3 diagonals, the side information coding efficiency is lower because the error energy increases. This leads to loss of coding efficiency because the lower redundancy that is obtained with MDSQ using 5 diagonals is partially lost by the encoding inefficiency of the side information.

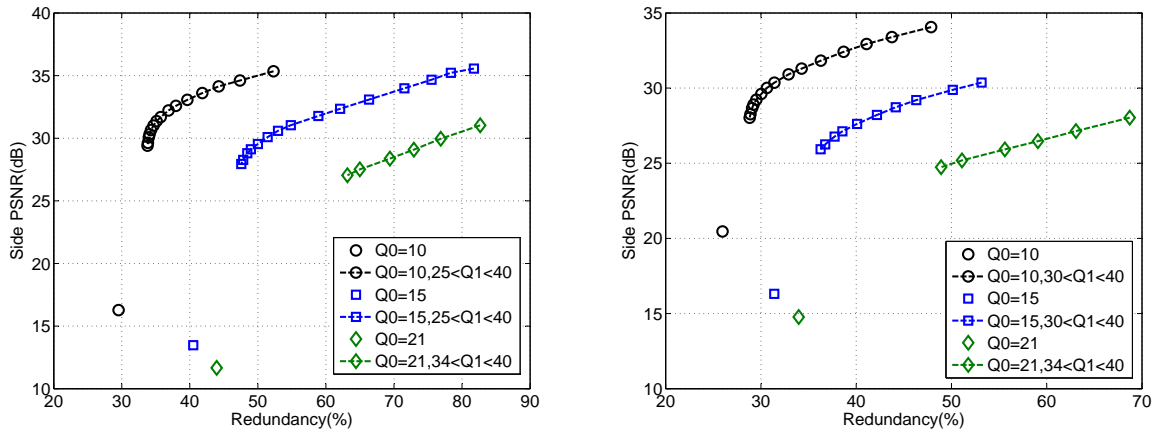


Figure 4.36: Side Distortion *Foreman* sequence, MDSQ with 5 diagonals. Figure 4.37: Side Distortion *Bus* sequence, MDSQ with 5 diagonals.

4.5.2 Unbalanced MDSQ

Unbalanced MDSQ (U-MDSQ) is intended to produce descriptions controlling the balancing ratio between respective rates. As it was shown in the last section, different balancing ratios are achieved by changing parameters of the index assignment matrix, namely, the Z parameter that controls the central SDC indices that are close to zero. It is important to note, as pointed out in Chapter 2, that in case of unbalanced descriptions, at least one of the side distortions is very poor, which in practice can make such scheme unfeasible. Thus, applying this method in highly predictive coding schemes such as Advanced Video Coding, its usage needs to be evaluated in order to obtain an acceptable side-distortion performance.

Hence, the proposed unbalanced MDC use additional information assuring an acceptable side-distortion performance, maintaining the balancing ratio between descriptions. The problem is distinct from the balanced case, because unbalanced MDSQ has to combine central rates with side information in order to have a desired balancing ratio considering all contributions to the overall rate. In particular, unbalanced rates can be obtained by three different methods. The first one is to use balanced MDSQ with different rates in redundant side-information. The second method consists in using unbalanced MDSQ with similar side-information rates in each description and finally, the third one is to use unbalanced MDSQ with different side information rates allocated to each description. In this case, each description is assumed to be inherently unbalanced, where unbalanced MDSQ is used to produce descriptions with the same side information rates in each description.

Therefore, the performance evaluation is focused on the effect introduced by the side-information in decoded quality, for a given ratio between descriptions. On the other hand, it is important to understand how the side-information affects the global balancing ratio for a given redundancy level. The results evaluate the side PSNR obtained by using fixed $Q_i, i = 1, 2$ in side encoders between 25 and 40 applied only to anchor frames. Balanced and unbalanced MDSQ are compared regarding to the overall redundancy and the side PSNR obtained in each description.

Intra Frames Table 4.3 shows the balancing ratio obtained in U-MDSQ including the side information applied to Intra frames. Considering that the same quantization parameter is used in each description at side encoders, the redundancy levels in each description have different relative weight, hence the overall redundancy is considered rather than its individual parts. Thus, the balancing ratio is mainly dependent on the U-MDSQ central encoding parameters, because the side information is relatively small in comparison with central information. Using different values of Z , balancing rates can vary between approximately 0.46 – 0.54 and 0.25 – 0.75, which follows the results presented in Figs 4.18 and 4.19 concerning to the balancing ratios obtained without side information.

Table 4.3: Balancing ratio including side information - Intra Frames (*Bus sequence*)

Z	QP0=11		QP0=15		QP0=18		QP0=21	
	Desc. 1	Desc. 2	Desc. 1	Desc. 2	Desc. 1	Desc. 2	Desc. 1	Desc. 2
-1	0.46	0.54	0.45	0.55	0.43	0.57	0.44	0.56
0	0.44	0.56	0.40	0.60	0.38	0.62	0.40	0.60
1	0.40	0.60	0.36	0.64	0.33	0.67	0.31	0.68
2	0.36	0.64	0.33	0.67	0.30	0.69	0.28	0.72
3	0.35	0.65	0.30	0.70	0.28	0.72	0.25	0.75

Figs. 4.38, 4.39, 4.40, 4.41, 4.42 and 4.43 show the obtained side PSNR versus redundancy. Redundancy includes the component introduced by the MDSQ scheme and the overall side-information rate from both descriptions. The side PSNR represents the decoded quality of each description individually decoded. The results show that, for high balancing ratios, the higher differences occur between side PSNR in description 1 and 2. The PSNR differences are around 4-6dB. This is because, the unbalancing of each description, introduce asymmetry in side quantization from MDSQ scheme, thus producing asymmetric reconstruction qualities. Therefore, adding the side information to compensate the drift distortion, does not improve the quality difference between descriptions, because in the coarser quantized description most of the original residual information is

eliminated. It is important to note that the PSNR values correspond to an acceptable quality, which means that unbalancing descriptions does not compromise any decodable description quality.

The analysis of results between balanced MDSQ (B-MDSQ) and U-MDSQ is done by comparing the side PSNR and redundancy. Regarding to side PSNR, the conclusion that can be drawn is that better results are obtained in U-MDSQ case, specially when the description 2 is the finer quantized one. The main reason for this is that B-MDSQ produces an equivalent side quantizer that is coarser than the central quantizer which is common to both descriptions. This leads to an equivalent side reconstruction quality in both descriptions that is worse when compared with the higher rate description in U-MDSQ. Considering the total redundancy levels, the redundancy decrease with the increasing of the balancing ratios, which are mostly dependent on the U-MDSQ central redundancy. Overall, U-MDSQ with side information leads to an effective unbalanced MDC scheme for Intra frames where an efficient coding of the side-information yields an effective side reconstruction quality even with better results than B-MDSQ.

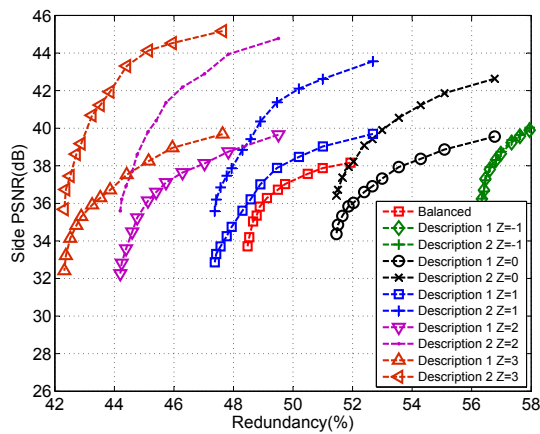


Figure 4.38: Side distortion *foreman* sequence, Intra Frame $QP0=11$.

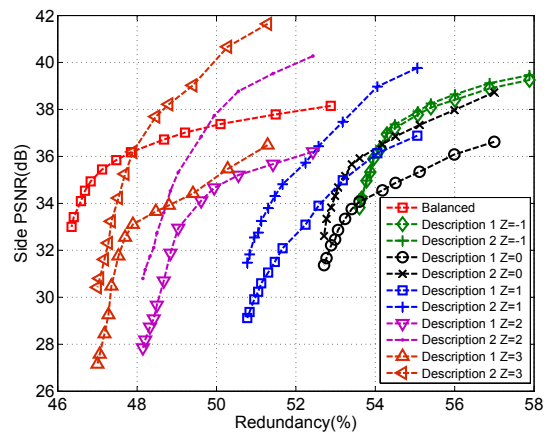


Figure 4.39: Side distortion *Bus* sequence, Intra Frame $QP0=11$.

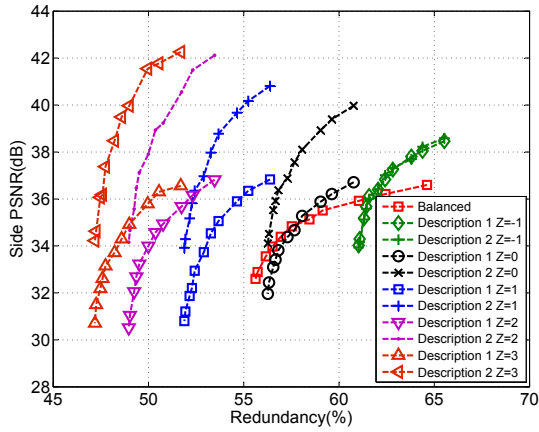


Figure 4.40: Side Distortion *foreman* sequence, Intra Frame $QP0=15$.

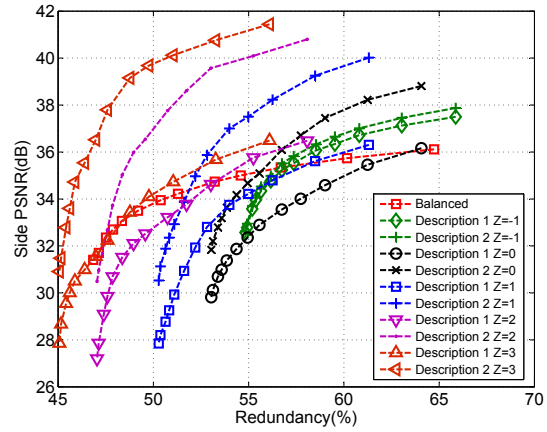


Figure 4.41: Side Distortion *Bus* sequence, Intra Frame $QP0=15$.

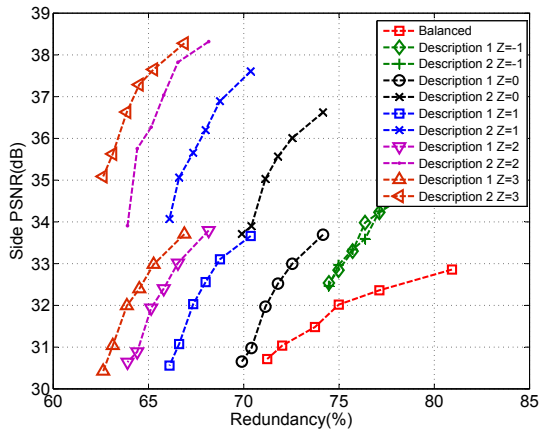


Figure 4.42: Side Distortion *foreman* sequence, Intra Frame, $QP0=21$.

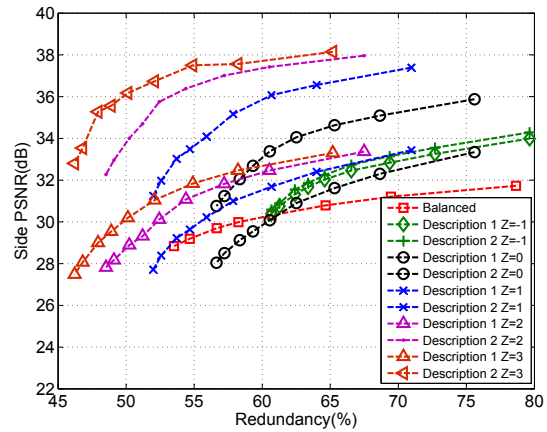


Figure 4.43: Side Distortion *Bus* sequence, Intra Frame, $QP0=21$.

Motion Predicted Frames Table 4.4 shows the balancing ratio obtained for U-MDSQ including the side information on motion compensated frames, considering that quantization parameters used in side-information are the same on both descriptions. As in Intra frames, balancing rates are mainly dependent on the U-MDSQ central encoding parameters, because the side information is relatively small comparing with central information. Using different values of Z , balancing rates can vary between approximately 0.4 – 0.6 and 0.3 – 0.7, respectively. In comparison with the results obtained in Intra frames (see Table 4.3), these balancing ratios are relatively higher. This is because the residue energy in motion prediction frames is lower than Intra prediction yielding higher efficiency in splitting of high frequency transform coefficients between descriptions.

Table 4.4: Balancing ratio including side information - MC Frames (*Bus sequence*)

Z	QP0=11		QP0=15		QP0=18		QP0=21	
	Desc. 1	Desc. 2	Desc. 1	Desc. 2	Desc. 1	Desc. 2	Desc. 1	Desc. 2
-1	0.43	0.57	0.41	0.59	0.40	0.60	0.39	0.61
0	0.37	0.63	0.36	0.64	0.35	0.65	0.33	0.67
1	0.33	0.67	0.32	0.68	0.31	0.69	0.30	0.70
2	0.30	0.70	0.30	0.70	0.29	0.71	0.29	0.71
3	0.27	0.73	0.27	0.73	0.27	0.73	0.27	0.73

Figs. 4.44, 4.45, 4.46, 4.47, 4.48 and 4.49 show the evaluation results representing the side PSNR versus total redundancy introduced by the MDSQ scheme including the side information of each description. The side PSNR represents the corresponding value of each decoded description.

The experimental results show that overall redundancy is reduced when comparing with balanced case. The results clearly show that redundancy can have at least 10% reduction in comparison with balanced MDSQ. There are two main reasons for this fact. First, the MDSQ redundancy is lower in U-MDSQ compared with B-MDSQ. Secondly, the side information in the U-MDSQ scheme is more efficiently encoded than in the case of the balanced case. This is because high frequency coefficients have smaller values, and when are encoded with U-MDSQ the energy of the compensation error signal is smaller, leading to increase the side information coding efficiency when comparing with balanced MDSQ. Regarding to side PSNR results, U-MDSQ clearly outperforms balanced MDSQ in almost central QP values and redundancy levels. The PSNR differences between descriptions in U-MDSQ are around 4-6 dB. This is an interesting result because, despite this difference, none of the descriptions have average PSNR values with a poor reconstruction. On the

contrary, even the coarser quantized description has at least the same average PSNR than balanced MDSQ. This derived from the fact that only anchor frames use side information, which leads to a worse quality reconstruction in B frames in balanced scheme. This is a very interesting result because with this method, is possible to generate unbalanced descriptions with relatively small redundancy levels, and at the same time, to achieve side distortions that are smaller comparing with balanced case.

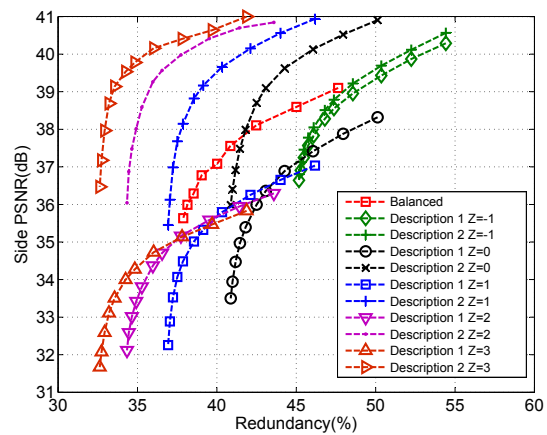
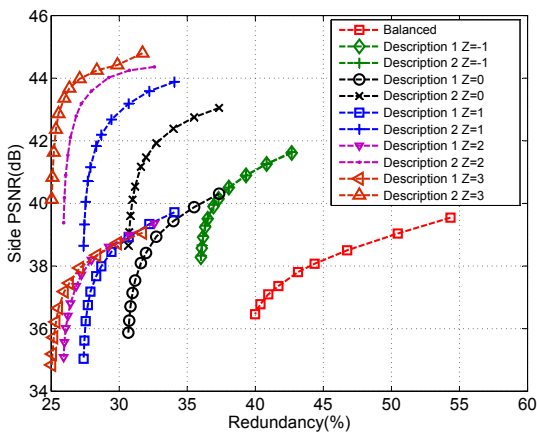


Figure 4.44: Side Distortion *foreman* sequence, IPBB Frame $QP0=11$. Figure 4.45: Side Distortion *Bus* sequence, IPBB Frame $QP0=11$.

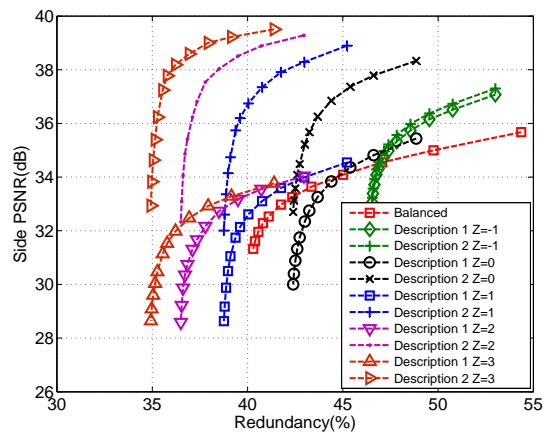
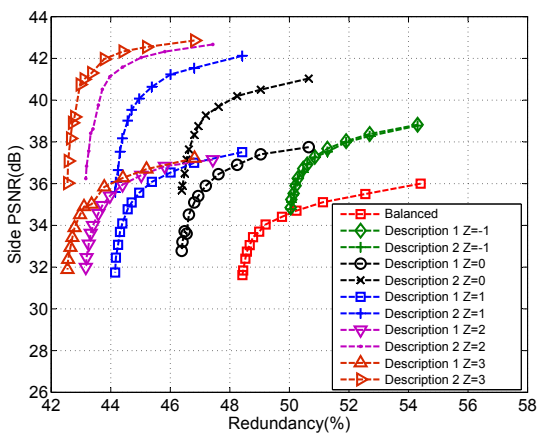


Figure 4.46: Side Distortion *foreman* sequence, IPBB Frame $QP0=15$. Figure 4.47: Side Distortion *bus* sequence, IPBB Frame $QP0=15$.

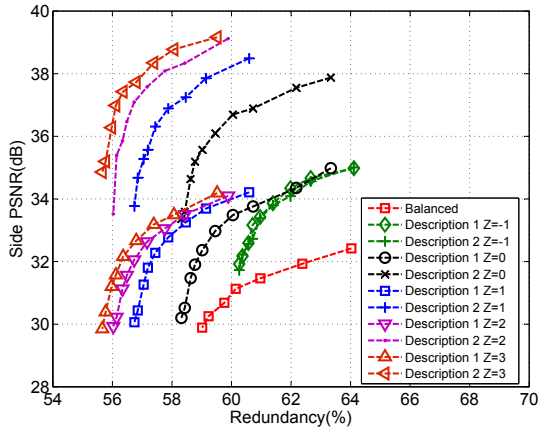


Figure 4.48: Side Distortion *foreman* sequence, IPBB Frame, $QP0=21$.

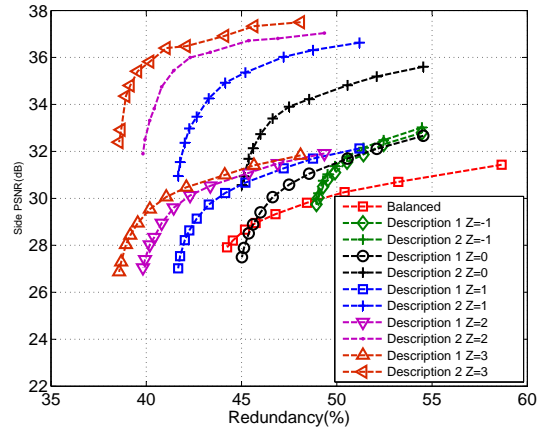


Figure 4.49: Side Distortion *Bus* sequence, IPBB Frame, $QP0=21$.

4.5.3 Frame-by-frame Distortion

Figs. 4.50 and 4.51 show the frame-by-frame distortion for *foreman* and *bus* for the first 60 frames. Fig. 4.50 shows that differences between unbalanced descriptions is around 3-4 dB, as also has been shown at section 4.5.2. An interesting observed feature is that each description is drift free as shown in frame-by-frame. Also, the PSNR that is maintained nearly constant taking the anchor frames as reference. Furthermore, this is mainly dependent on the sequence content. As the figure shows, the side PSNR is maintained roughly constant, despite B frames being decoded without side information. This is due to the fact that the motion prediction residue energy in B-frames is relatively small which results in most of the higher frequency coefficients being encoded into only one description. This leads to a reconstruction degradation in the coarser description, as noted in some parts of the sequences, where the residual energy is higher, *i.e.*, the scene content change.

Fig. 4.51 shows equivalent results in *bus* sequence. The results show the PSNR difference about 6 dB on unbalanced descriptions depending on the balancing ratio. Furthermore, it clearly shows that the anchor frames maintain a drift-free decoded sequences from both unbalanced and balanced MDSQ. The effect of open loop decoding at B frames is also shown. In particular, a worse decoded quality in coarser descriptions is obtained when the balancing ratio increases, because the remaining residual energy is relatively small, producing a poor reconstruction. On the other hand, in finer descriptions, the decoded video quality is not significantly affected in B frames because almost all of the residual

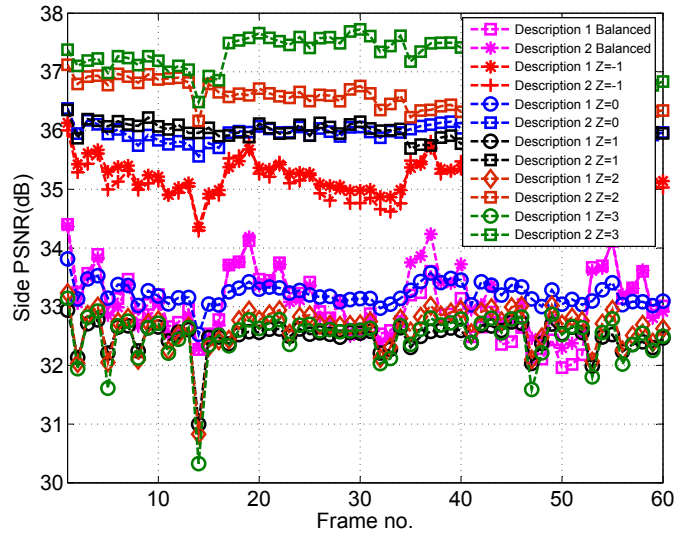


Figure 4.50: Frame-by-frame side PSNR *foreman* sequence, IPBB Frame, $QP_0=15$; $QP_1=35$.

information is encoded. In regard to the B-MDSQ case, Fig. 4.51 shows that resulting side PSNR is between unbalanced descriptions values, which is an expected result and where anchor frames maintain a drift-free decoded sequence.

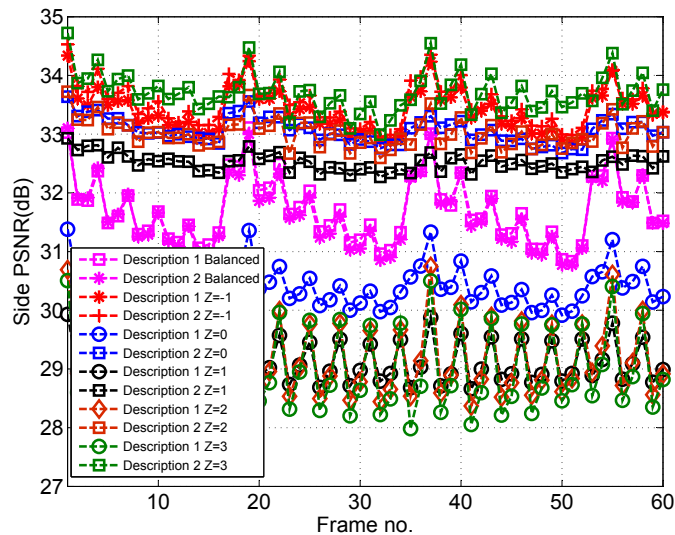


Figure 4.51: Frame-by-frame side PSNR *bus* sequence, $QP_0=15$; $QP_1=35$.

4.6 Conclusion

This chapter proposes a new MDSQ architecture for H.264/AVC with drift-control for both balanced and unbalanced descriptions. A study on the main issues related to MDSQ developing is presented, where drift distortion, redundancy, and overall rate-distortion performance is evaluated. The proposed architecture allows to improve the open-loop MDSQ architecture in terms of side-distortion performance with a reduced amount of additional information. The novel unbalanced MDSQ, improves the overall performance comparing with the balanced close-loop MDSQ scheme, and, consequently to the open-loop architecture. It is shown that the proposed unbalanced MDSQ scheme increases the coding efficiency by correctly allocating the side-information without changing the balancing ratio, and improving the overall decoded quality.

Multiple Description Video Splitting of Coded Streams

Contents

5.1	MDVS Scenario	112
5.2	Classic MDVS	114
5.2.1	Drift Analysis	115
5.3	MDVS with drift compensation	116
5.3.1	MDVS architecture	117
5.3.2	Simplified MDVS	118
5.4	MDVS drift performance	121
5.4.1	Intra predicted frames	122
5.4.2	MC predicted frames	122
5.4.3	Generic regular GOP	124
5.4.4	The overall effect of side information	125
5.5	MDVS streaming with path diversity	127
5.6	MDVS vs MDC: comparative discussion	131
5.7	Conclusion	132

This chapter extends the current concept of MDC to the compressed domain, by proposing efficient splitting of standard SDC video into a multi-stream representation. A novel multiple description video splitting (MDVS) scheme is proposed to operate at network edges, for increased robustness in path diversity video streaming across heterogeneous communications chains. It is shown that poor performance of existing methods is mainly due to distortion accumulation, i.e., drift, when decoding is carried out with missing descriptions. The proposed scheme is able to effectively control drift distortion in both intra and inter predictive coding, even when only one description reaches the decoder. This is achieved by generating a controlled amount of relevant side information to compensate for drift accumulation, whenever any description is lost in its path. The simulation results show that any individual description can be decoded on its own without producing drift, achieving significant quality improvement at reduced redundancy cost. The overall performance evaluation, carried out by simulating video streaming over lossy networks with path diversity, also demonstrates that MDVS enables higher quality video in such heterogeneous networking environments, for a wide range of packet loss rates.

The MDVS scheme is based on MDSQ, using side information to control drift in both spatial and temporal prediction. The side information is generated from the original stream and its rate is controlled with an independent quantisation parameter which also controls redundancy. Then, a simplified architecture is devised to reduce the overall complexity in regard to the number of processing functions and memory requirements. No additional information is needed from the original SDC encoder in order to generate such side information at any MDVS-enabled network node. The main novel aspects of the proposed MDVS scheme comprise: (i) a two-loop MDVS architecture with drift control in both intra and inter predictive coded slices; (ii) an equivalent single-loop architecture; (iii) a method to generate side information from SDC video; (iv) the capability of controlling the amount of side information according to the expected decoder drift; (v) an overall performance similar to MDC using uncompressed video.

5.1 MDVS Scenario

Fig. 5.1 shows a possible video streaming scenario where MDVS might be useful. A single video stream (i.e., SDC stream) is distributed from the streaming server to several user terminals over heterogeneous networks, some of them having disjoint separate paths

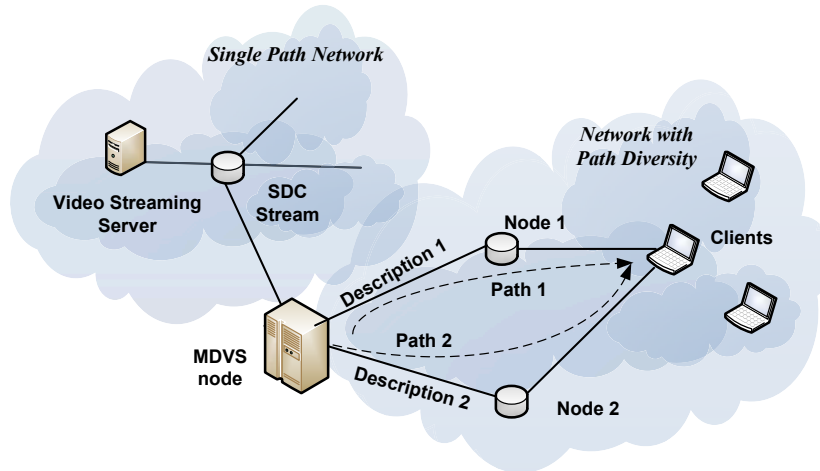


Figure 5.1: MDVS application scenario.

from an intermediate node to the user terminal. Since the server storage capacity and streaming bandwidth required for SDC video is less than that of equivalent MDC (i.e., same quality), due to the inherent redundancy of MDC, SDC is a more efficient coded format for storage and distribution over single path networks. The advantage of MDVS is to introduce a further level of flexibility in SDC video streaming, in order to benefit from transmission over multiple paths where these are available along the delivery chain.

Therefore, MDVS can be seen as a novel adaptation functionality of edge nodes in the heterogeneous video streaming environments of the future media internet. In comparison with the existing MD schemes previously cited, this chapter addresses a different MD problem, which consists in splitting compressed video streams rather than producing MD from uncompressed video signals. Moreover, since MDVS suffers from the same intrinsic problem of drift as MDC, i.e., accumulation of decoding distortion when any description is lost in the network, a novel aspect of the proposed scheme in this chapter is its capability for limiting such type of distortion, by generating a controlled amount of side information, specifically for this purpose.

Other methods capable of mitigated the drift. In [S. Lin 2001], the reference frames for motion compensated prediction are selected according to feedback information received from the transmission paths. A similar principle is used in [Liao 2011], where routing messages are used in order to estimate packet loss error rate, by dynamically selecting the best reference frame in order to alleviate error propagation. If the drift compensation process is considered as a bit stream switching problem (i.e., switching from decoding

with two descriptions to decoding with a single description), then periodic switching frames (e.g., H.264/AVC SI/SP slices) might be used to enhance MDC error resilience [Wang 2005a]. However, this mechanism cannot be used directly with compressed streams because SI/SP frames would need to be dynamically computed by full MDC encoders at MDVS network nodes.

Since the multi-loop drift compensation methods mentioned before, lead to high complexity implementations, a different approach is followed in the proposed MDVS scheme, which is based on a single-loop MDSQ. This comprises a novel MDVS architecture with low processing complexity, achieved by reusing most coding parameters of the incoming SDC video streams.

Two rather different approaches, based either on channel coding or source coding, have been followed in video streaming applications using MDVS. In [Puri 2001] the channel coding approach is proposed for an end-to-end video communication system where MDVS, based on forward error correcting codes (FEC), is integrated in a congestion control framework for video streaming over the internet. Other examples of previous work in MDVS based on channel coding are reported in [Gan 2006] and [Essaili 2007]. Also, MDVS based on the source coding approach was also addressed in the past, but mainly focussed on its application in some networking scenarios without taking in account the rate-distortion efficiency of actual MDVS processing architectures. In particular, drift free MDVS architectures cannot be found in the available literature. Related work can be found in [Kim 2003], where an MDVS scheme is proposed based on redundancy rate-distortion optimisation for splitting DCT coefficients of the incoming bitstream. In [Kim 2006] another MDVS scheme is proposed, based on replication and interleaving of DCT coefficients among all descriptions. However, these are open-loop schemes with no drift compensation, also resulting in higher levels of rate redundancy.

5.2 Classic MDVS

MDVS can be regarded as a data partitioning scheme, capable of generating two descriptions from an SDC video stream. Since current coded video formats convey a great deal of the source information in transform coefficients, MDSQ is a good candidate to design low complexity MDVS systems. Fig. 5.2 shows a classic MD video splitting scheme where each transform coefficient is represented by two different values, which result in divid-

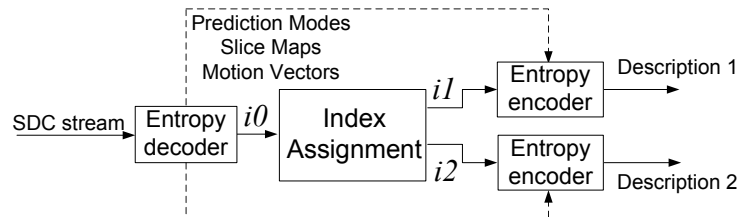


Figure 5.2: Classic MDVS scheme.

ing an SDC stream into two descriptions. For instance, this MDVS method was used in [Chen 2007b] where the coding information embedded in the original SDC stream, such as slice maps, prediction modes and motion vectors are duplicated in the two resultant descriptions. In the scheme shown in Fig. 5.2, an index assignment module is used for mapping each quantisation index of the original transform coefficients i_0 (i.e., central indices) into a pair of side indices (i_1, i_2) which are then entropy encoded.

5.2.1 Drift Analysis

If the classic MDVS scheme of Fig. 5.2 is used for adapting coded video streams to networks with path diversity, then drift is introduced at the decoder whenever any description is lost. The drift distortion component can be determined from the relevant signals involved in MD splitting and decoding. When two descriptions are received, both side indices (i_1, i_2) are decoded and merged into the corresponding central index i_0 . In this case, for each block n , the reconstructed central pixel values $F_{n,0}$ are given by,

$$F_{n,0} = r_{n,0} + P_{n,0}, \quad (5.1)$$

where $r_{n,0}$ is the decoded residue and $P_{n,0}$ its associated prediction either from intra prediction or motion compensation, formed from decoding both descriptions. If only one description j (either $j = 1$ or $j = 2$) is decoded, then, the reconstructed pixel values $F_{n,j}$ are given by,

$$F_{n,j} = r_{n,j} + P_{n,j}, \quad (5.2)$$

where $r_{n,j}$ is the decoded residue and $P_{n,j}$ its prediction formed from decoding description j only. Since $r_{n,j}$ results from inverse index assignment using only one description as input, the difference between the original SDC residue and that decoded from only one

description produces a reconstruction error ε , i.e.,

$$r_{n,j} = r_{n,0} + \varepsilon. \quad (5.3)$$

Substituting (5.3) in (5.2),

$$F_{n,j} = r_{n,0} + \varepsilon + P_{n,j} \quad (5.4)$$

and, then, using (5.1) in (5.4), $F_{n,j}$ becomes,

$$\begin{aligned} F_{n,j} &= F_{n,0} + P_{n,j} - P_{n,0} + \varepsilon \\ &= F_{n,0} + d_P + \varepsilon, \end{aligned} \quad (5.5)$$

where

$$d_P = P_{n,j} - P_{n,0}, \quad (5.6)$$

is the drift component due to mismatch between the SDC stream predictions used in the original encoder and those reconstructed at the final decoder from only one description. Note that the above analysis is valid for both spatial and temporal drift components, though these can be identified as separate contributors to the overall drift distortion. Since classic MDVS is applied to the DCT coefficients of the original SDC stream, this problem is equivalent to the exposed in Chap. 4 for open-loop MDC encoders. Hence, decoding drift effects due to prediction mismatch in intra and motion compensated predicted slices are equivalent in both schemes.

5.3 MDVS with drift compensation

A novel MDVS architecture and an equivalent simplified version are proposed to overcome the problem of spatial and temporal drift. In comparison with the classic scheme of Fig. 5.2, the proposed splitting architecture generates an additional amount of side information for the specific purpose of preventing drift when only one description is decoded. As explained in the next subsections, the side information is generated for each description by further encoding the difference between the original SDC signal and the one decoded from the description itself.

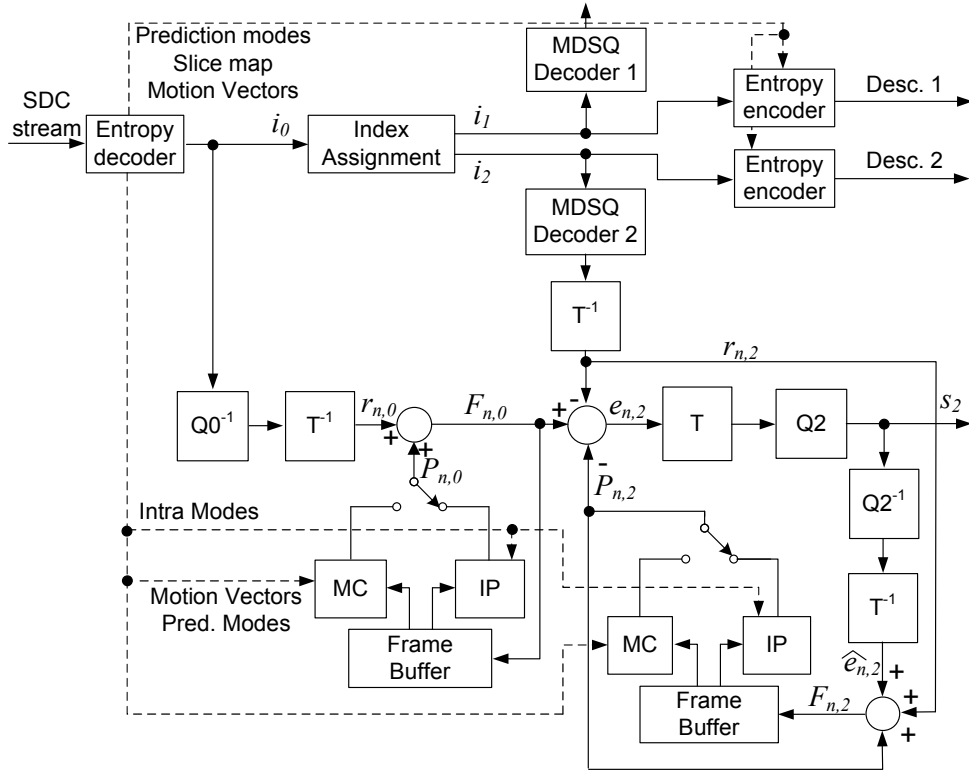


Figure 5.3: Two-loop MDVS architecture.

5.3.1 MDVS architecture

The proposed MDVS architecture using side information for drift compensation is shown in Fig. 5.3. Such an architecture is used for drift free MDVS of both intra predicted and motion compensated predicted frames. However, the B frames that are not used as references may be processed with the classic MDVS scheme of Fig. 5.2 because they do not contribute for drift accumulation. Similarly to the classic MDVS, the headers, prediction modes and motion vectors are duplicated in both descriptions. Also, the side information is encoded by using the same prediction modes and motion vectors.

In the MDVS architecture shown in Fig. 5.3 the side information s_j for each description j , with $j = 1, 2$, is defined as,

$$s_j = Q_j\{T(e_{n,j})\}, \quad (5.7)$$

where

$$e_{n,j} = F_{n,0} - r_{n,j} - P_{n,j}, \quad (5.8)$$

and Q_j is the quantisation with side quantiser Q_{P_j} that determines the amount of side information sent to the decoder, T is the transform kernel, $F_{n,0}$ is the current frame and $P_{n,j}$ is the motion compensated prediction from side encoder j . The residue $r_{n,j}$, available at the decoder when only one description is correctly received, is given by,

$$r_{n,j} = T^{-1}(Q_0^{-1}\{A_j^{-1}(i_j)\}), \quad (5.9)$$

where $A_j^{-1}(\cdot)$ denotes the inverse index assignment operation when only description j is available and Q_0^{-1} is the inverse quantisation using the SDC quantisation parameter Q_{P_0} . The central index i'_0 is assigned to the value located in the main diagonal of the index assignment matrix which corresponds to either the row or column of side index i_j .

At the decoder, if both descriptions are available, then, the inverse index assignment operation is responsible to restore the exact value of the original SDC index, which is then inverse quantised and inverse transformed. However, if only one description is decoded, then frame $F_{n,j}$ is reconstructed without drift as follows,

$$F_{n,j} = P_{n,j} + r_{n,j} + \hat{e}_{n,j}, \quad (5.10)$$

with

$$\hat{e}_{n,j} = T^{-1}(Q_j^{-1}\{s_j\}). \quad (5.11)$$

By comparing (5.2) with (5.10), one concludes that $e_{n,j}$ is the drift compensation signal that is encoded as side information and sent to the decoder. Note that in the classic MDVS scheme shown in Fig. 5.2, only signal $r_{n,j}$ is sent for decoding which is not enough to prevent drift.

5.3.2 Simplified MDVS

The MDVS architecture shown in Fig. 5.3 can be simplified by assuming that prediction is a linear operation (this is valid except for rounding and truncation arithmetic). Using equations (5.1) and (5.8), the following relation can be derived,

$$e_{n,j} = P_{n,0} - P_{n,j} + r_{n,0} - r_{n,j}. \quad (5.12)$$

Considering (5.12) and the previous assumption of linearity, the simplified architecture

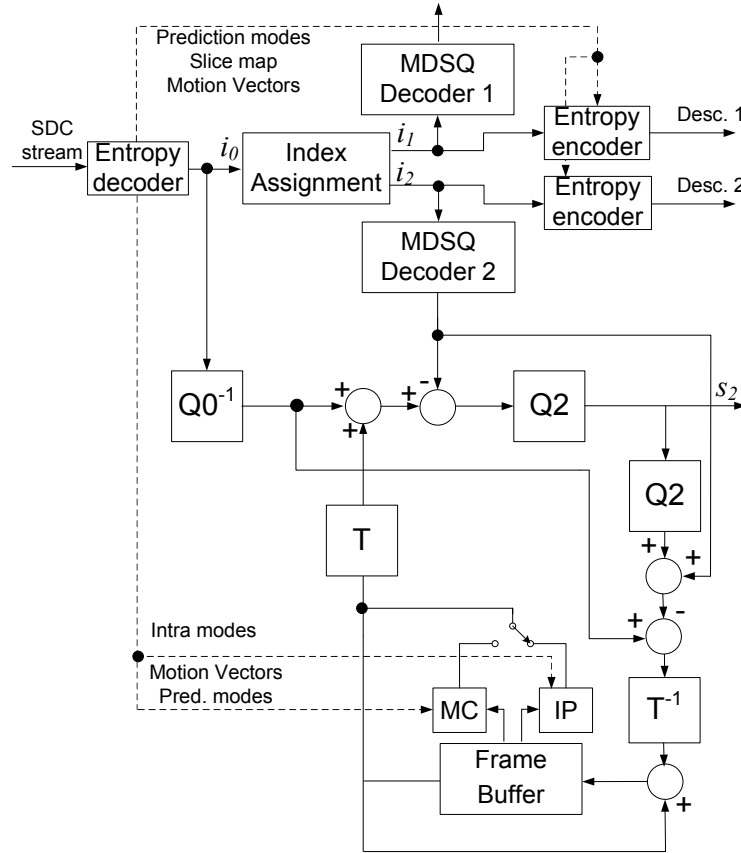


Figure 5.5: Equivalent single-loop MDVS architecture.

$$\begin{aligned}
 F_{nT,j} &= F_{n,0} - P_{n,0} + \\
 &\quad - (r_{n,j} + F_{n,j} - P_{n,j} - r_{n,j}) + P_{nT,j} \\
 &= F_{n,0} - F_{n,j} - P_{n,0} + P_{n,j} + P_{nT,j}.
 \end{aligned} \tag{5.15}$$

Considering the linearity of prediction $P_{nT,j} = P_{n,0} - P_{n,j}$ then,

$$F_{nT,j} = F_{n,0} - F_{n,j}. \tag{5.16}$$

Equation (5.16) represents the difference between frames reconstructed from the two prediction loops shown in Fig. 5.3. In the simplified architecture shown in Fig. 5.4 such difference is accumulated in only one loop and the result is used in the same manner as shown in Fig. 5.3, which demonstrates that both architectures are equivalent.

Moreover, since transform and quantisation can be performed as independent operations, the architecture shown in Fig. 5.4 can be further simplified to that shown in Fig. 5.5. For each description, this scheme only uses one frame buffer and two transforms while that of Fig. 5.3 needs two frame buffers and four transforms. Note that in H.264/AVC, the scheme shown in Fig. 5.5 needs to use scaling coefficients in the quantisation and inverse quantisation in order to make them independent from the transform. This can be easily done as described in [Lefol 2006].

Although some operations may involve non-linear arithmetic, such as clipping operation in transform/quantisation, rounding in sub-pel MC interpolation and deblocking filtering, the actual effect on the drift performance of the simplified MDVS is mostly negligible. However, it might be slightly more significant in high motion sequences.

The proposed MDVS architectures were developed using on the reference software of H.264/AVC available online. Each coded description produced by MDVS is standard-compliant and the corresponding coded data is encapsulated into Video Coding Layer (VCL) Network Adaptation Layer (NAL) units. To include the side information in the standard syntax, a new type of VCL NAL unit must be defined for such coded data. This can be done by extending the existing NAL types using different approaches. For instance, in H.264/SVC [Ye-Kui 2007], new NAL unit types were defined to accommodate several layers and associated information and, in [Lamy-Bergot 2010], a new type of NAL unit is proposed for embedding redundant information inside standard video streams.

5.4 MDVS drift performance

In this section, performance results were obtained to prove the advantages of using the proposed MDVS method by evaluating the impact of drift accumulation when one description is totally lost, either in intra or inter predicted coded frames. The drift performance of MDVS was evaluated in two different aspects: (i) distortion accumulation in decoded video when only one of the two descriptions reaches the decoding terminal and (ii) the extra redundancy of the side information produced by the proposed MDVS to compensate for drift, also in one of the two descriptions. The reference used for comparison is one description obtained from the classic MDVS scheme of Fig. 5.2. The spatial and temporal drift performance were evaluated by using coded streams with intra predicted and MC predicted frames (P and B), respectively. In order to obtain a comparable evaluation, all

streams were encoded at the same rate.

The original headers, prediction modes, slice maps and motion vectors are duplicated into both descriptions. The side information is encoded using the same coding modes as the corresponding descriptions. Note that coding modes and motion vectors are not included in the side information because they are available from the respective coded description.

5.4.1 Intra predicted frames

The benefit of drift compensation in intra predictive MDVS is shown in Fig. 5.6 where the PSNR of each macroblock of one frame, in one of the two descriptions (*bus* sequence), is shown for classic MDVS and for two-loop MDVS (i.e., Fig. 5.3) at the same bit rate. The same rate is ensured by an average central quantiser $Q_{P0} = 15$ for Classic MDVS and $Q_{P0} = 18$, $Q_{P1,2} = 33$ for proposed MDVS. Fig. 5.6 clearly shows that the proposed MDVS produces much higher and smoother PSNR along the I frame than in the case where drift compensation is not done. In the case of no drift compensation, i.e., classic MDVS, the lowest PSNR is below 20dB, which is definitely not acceptable.

5.4.2 MC predicted frames

A different experiment was carried out to evaluate the performance of temporal drift compensation. A GOP structure with high number of predicted frames was used to provide a worst case scenario in regard to temporal drift, i.e., a sequence of P frames using only one reference, i.e., IPPP... The GOP size was set to 20 frames. The entire loss of one description is simulated in the path diversity network, for the initial I frame and also for all subsequent P frames (i.e., only one description is decoded). In the error-free descriptions of all streams the initial I frame is sent with side information in order to not affect the quality of subsequent P frames. Both the two-loop and the single-loop MDVS architectures were used in the experiment in order to evaluate the effect of the non-linearity of motion compensation in the drift accumulation over a significant number of temporally predicted frames.

Fig. 5.7 shows the PSNR for *coastguard* and *foreman* sequences at the same bit rate, using an average central quantiser $Q_{P0} = 16$ for classic MDVS and $Q_{P0} = 18$, $Q_{P1,2} = 30$ for proposed MDVS. It is quite evident that the proposed MDVS architectures are drift-free,

while temporal drift accumulation is responsible for severe quality degradation in classic MDVS. At the end of the GOP, the PSNR obtained by using the proposed MDVS is about $6dB$ higher than classic MDVS. For classic MDVS, two-loop and single-loop, the average PSNR of *coastguard* is, respectively, $28.96dB$, $32.25dB(+3.46)$ and $32.42dB(+3.29)$ and that of *foreman* is $31.41dB$, $34.34dB(+2.93)$ and $34.43dB(+3.02)$. These results also show that non-linearity of motion compensation is negligible, since the PSNR obtained from the single-loop MDVS architecture are quite similar to those obtained from two-loop MDVS. Therefore, full decoding of the incoming stream is not necessary to achieve drift free MDVS and these results validate the proposed single-loop architecture.

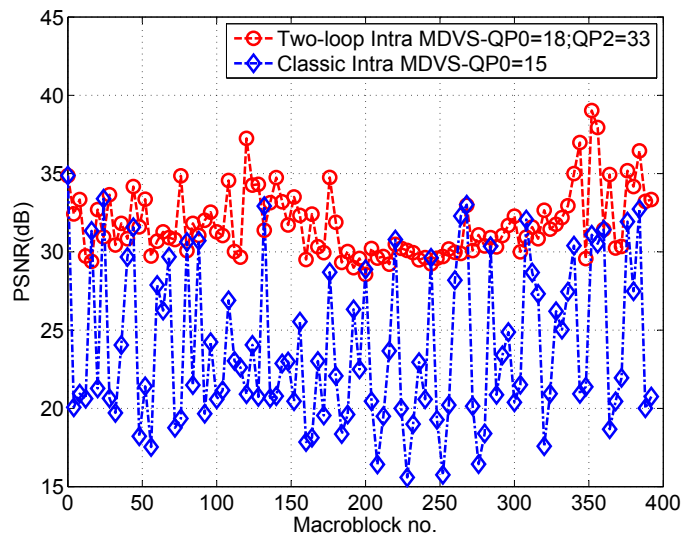


Figure 5.6: PSNR of Intra Frame macroblocks.

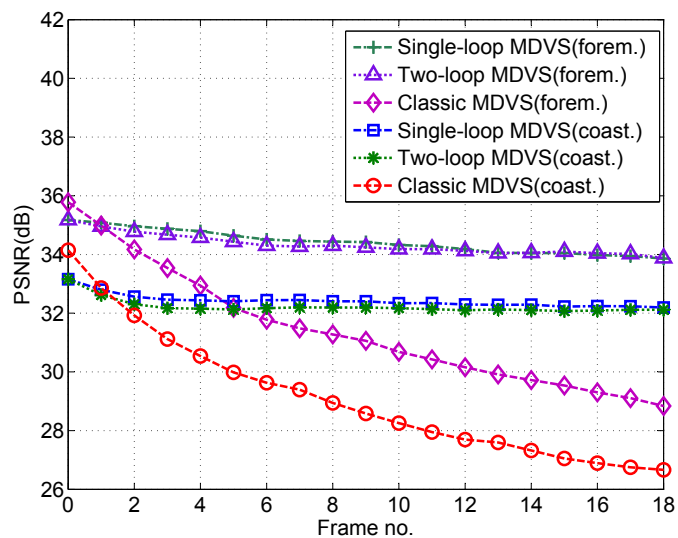


Figure 5.7: PSNR for MC predicted frames (IPPP...) for *coastguard* and *foreman*.

5.4.3 Generic regular GOP

The overall performance of the proposed MDVS using generic IBBP GOP structures was also evaluated (GOP size of 20 frames) using the same three sequences, i.e., *coastguard*, *foreman* and *bus*. In the proposed MDVS, drift compensation was used for both I and P frames but not for B frames because these are not used as references for prediction. Performance evaluation was carried out by splitting SDC streams of each sequence into two descriptions and then simulating that only one description reaches the decoder for all frames in the GOP. In all streams, the initial I frame is always sent with side information in order to not influence the quality of subsequent predicted frames.

Figs. 5.8, 5.9 and 5.10 shows the PSNR obtained from the proposed MDVS architectures and classic MDVS. The central and side quantisers used in these experiments were $Q_{P0} = 18$ for classic MDVS and the proposed MDVS, which also used $Q_{P1,2} = 30$ for *coastguard* and *foreman* and using an average central quantiser $Q_{P0} = 15$ and $Q_{P1,2} = 28$ for classic MDVS and the proposed MDVS for *bus* sequence. The results in Figs. 5.8, 5.9 and 5.10 clearly confirm the effectiveness of the proposed architecture to eliminate the drift, and, consequently to achieve significant quality improvement in MD adaptation of coded video streams. The proposed architecture compensates for the drift in P frames which results in a significant overall quality improvement. Comparing PSNR of both the two-loop and single-loop MDVS architectures, these are very similar, which further validates the effectiveness of the simplified single-loop.

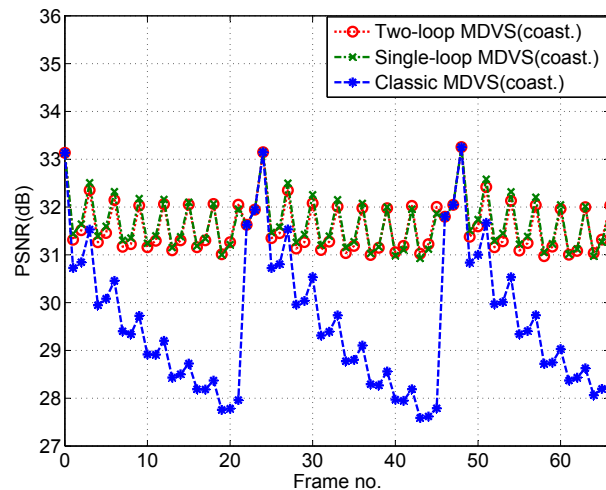
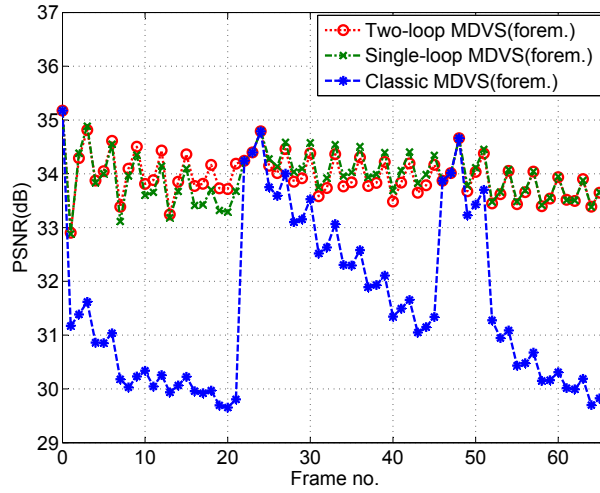
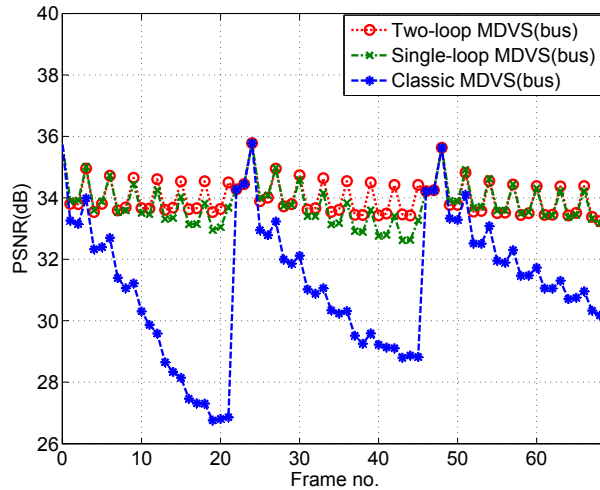


Figure 5.8: PSNR for generic regular GOP (IPBBP...) for *coastguard*.

Figure 5.9: PSNR for generic regular GOP (IPBBP...) for *foreman*.Figure 5.10: PSNR for generic regular GOP (IPBBP...) for *bus*.

5.4.4 The overall effect of side information

The overall effect of the side information in the rate and video quality for different combinations of average Q_{P_0} and $Q_{P_{1,2}}$ is presented in Table 5.1, where the average PSNR and extra redundancy are shown. This extra redundancy is due to the side information and it is measured as the percentage of total bit rate increase in each description, using the SDC rate obtained at the same Q_{P_0} as reference. Therefore, this is the actual cost of the side information for achieving drift compensation in MDVS. Without such extra redundancy, the overall redundancy is equal to that of classic MDVS and it is in line with various MDC schemes, as discussed in 5.6. Note that, as previously pointed out, the side information does not include coded motion vectors neither prediction modes.

Table 5.1: PSNR vs Side Information Redundancy (one description).

Seq.	QP_0	$QP_{1,2}$	Classic MDVS PSNR(dB)	Two-loop MDVS PSNR(dB)	Single-loop MDVS PSNR(dB)	Red. (%)
foreman	10	20	36.9	40.1(+3.2)	40.0(+3.1)	8.1
		25	36.2	38.0(+1.8)	38.0(+1.8)	2.9
		28	35.8	36.9(+1.1)	36.9(+1.1)	2.0
		30	35.5	36.3(+0.8)	36.3(+0.8)	1.8
	15	28	33.4	35.5(+2.1)	35.4(+2.0)	4.6
		30	33.1	34.6(+1.5)	34.6(+1.5)	3.4
		33	32.6	33.6(+1.0)	33.7(+1.1)	2.8
		35	32.2	32.9(+1.0)	32.9(+1.0)	2.6
	18	30	31.5	34.0(+2.5)	33.9(+2.4)	6.5
		33	31.1	32.8(+1.7)	32.9(+1.8)	4.9
		35	30.7	32.0(+1.3)	32.0(+1.3)	4.4
		38	30.0	30.9(+0.9)	30.8(+0.8)	4.1
21	30	30.8	33.7(+2.9)	33.5(+2.7)	14.0	
	33	30.3	32.5(+2.2)	32.3(+2.0)	8.4	
	35	29.9	31.5(+1.6)	31.4(+1.5)	6.8	
	38	29.2	30.2(+1.0)	30.2(+1.0)	6.0	
coastguard	10	20	35.9	39.1(+3.2)	39.1(+3.2)	6.4
		25	35.4	36.9(+1.5)	37.0(+1.6)	2.1
		28	35.0	35.8(+0.8)	35.9(+0.9)	1.2
		30	34.6	35.2(+0.6)	35.2(+0.6)	1.0
	15	28	31.5	33.6(+2.1)	33.7(+2.2)	3.6
		30	31.3	32.7(+1.4)	32.8(+1.5)	2.3
		33	30.9	31.7(+0.8)	31.7(+0.8)	1.6
		35	30.4	31.0(+0.6)	31.0(+0.6)	1.3
	18	30	29.1	31.7(+2.6)	31.7(+2.6)	4.8
		33	28.8	30.4(+1.6)	30.5(+1.7)	2.8
		35	28.4	29.6(+1.2)	29.7(+1.3)	2.2
		38	27.8	28.6(+0.8)	28.6(+0.8)	1.8
21	30	27.3	31.4(+4.1)	30.8(+3.5)	11.0	
	33	26.7	29.7(+3.0)	29.4(+2.7)	5.9	
	35	26.5	28.8(+2.3)	28.5(+2.0)	4.0	
	38	26.0	27.5(+1.5)	27.4(+1.4)	2.9	
bus	10	20	36.0	39.3(+3.3)	39.0(+3.0)	6.9
		25	35.4	37.2(+1.8)	37.1(+1.7)	2.3
		28	35.0	36.1(+1.1)	36.1(+1.1)	1.4
		30	34.7	35.5(+0.8)	35.5(+0.8)	1.2
	15	28	31.5	34.0(+2.5)	33.9(+2.4)	3.7
		30	31.2	33.1(+1.9)	33.0(+1.8)	2.5
		33	30.8	32.1(+1.3)	32.1(+1.3)	1.8
		35	30.4	31.6(+1.2)	31.5(+1.1)	1.6
	18	30	29.3	32.1(+2.8)	32.0(+2.7)	5.1
		33	28.9	30.9(+2.0)	30.8(+1.9)	3.1
		35	28.6	30.1(+1.5)	30.1(+1.5)	2.5
		38	28.1	29.1(+1.0)	29.0(+1.0)	2.2
21	30	28.0	31.9(+3.9)	31.4(+3.4)	11.5	
	33	27.4	30.2(+2.8)	29.9(+2.5)	5.7	
	35	27.0	29.3(+2.3)	29.1(+2.1)	4.1	
	38	26.5	28.0(+1.5)	27.9(+1.4)	3.1	

When only a single description is received, these results demonstrate that the proposed MDVS can significantly improve the video quality at a small cost in additional redundancy. The table 5.1 shows that extra redundancy due to side information ranges from 1% to 14%, while PSNR benefits from increases between 0.6dB and 4.1dB, in comparison with classic MDVS. As previously pointed out, such PSNR improvement is due to drift compensation. Although this is dependent on the type of sequence, it is worthwhile to note that PSNR obtained from the proposed MDVS is consistently better for acceptable levels of extra redundancy. For instance, for the *foreman* sequence with $QP_0 = 10$ and an excess rate of 2.9%, the mean PSNR improves 1.8dB and for *bus* sequence with $QP_0 = 15$ and with an excess rate of 3.7%, the mean PSNR improves 2.5dB. Better improvements can be achieved with higher redundancy values. This is the case, for example, of the *bus* sequence where the quality improves by 3.4dB if 11.5% of excess rate is used by side information.

The results in Table 5.1 also show that redundancy of side information increases with Q_{P0} . This is because the amount of side information to be encoded increases for higher values of Q_{P0} , which in turn is due to the larger differences between SDC and each description (i.e., difference between $F_{n,0}$ and $r_{n,j}$, $j = 1, 2$ in Fig. 5.3). Moreover, for each Q_{P0} , the extra redundancy decreases with $Q_{P1,2}$. This is due to the fact that $Q_{P1,2}$ is the quantiser used to encode the side information itself. Thus, the higher the value of $Q_{P1,2}$, the smaller the respective coded rate is. These results provide useful insight for future design of efficient MDVS rate control algorithms.

5.5 MDVS streaming with path diversity

The MDVS performance using the proposed architectures was evaluated in a simulated path diversity scenario where an SDC video stream is split into two descriptions at the network edge for streaming over different paths (e.g., Fig. 5.1). After splitting the SDC stream at the MDVS edge node, each description is streamed over independent paths subject to the same packet loss rates (PLR) and average burst error length (BEL). The side information is then multiplexed and packetised along with the corresponding description. Therefore, when a packet is lost, the coded description and side information are both lost. In the simulations, the packet size was set to 1000 bytes. The reference used for performance comparison of the proposed MDVS is SDC streaming under the same networking conditions, i.e., using the same amount bandwidth and suffering from equal PLR and BEL.

Burst packet loss was simulated using a Gilbert-Elliott 2-state Markov model in order to generate different average packet loss rates and mean burst duration [Li 2009]. In order to obtain statistically meaningful results, the transmission of each sequence was simulated 100 times under the same network conditions, i.e., average PLR values of 3%, 5%, 7% and 10% and average BEL of 4 and 12 packets.

Five test sequences were used, *Bus*, *Foreman*, *Mother-daughter*, *News* and *City*, CIF@30Hz. The GOP structure was IBBPBBP... with GOP=15 frames. In all cases, an index assignment matrix of 3 diagonals ($k = 1$) was used to generate the two descriptions from the compressed SDC stream. In this case, a single index encoded in a description represents 3 coefficients in the original SDC stream. Frame-copy error concealment was used whenever one packet is lost. Note that, for this type of performance evaluation, such

low performance concealment method is preferable over more efficient ones, because the quality results do not include masking effects due to concealment.

Figures 5.11-5.16, show the average PSNR obtained for different PLR (3%, 5%, 7% and 10%), BEL and rates (1.25, 1.8 and 2.16 Mbit/s) using *bus* CIF@30Hz sequence. For PLR higher than 3%, the simulation results show that proposed method achieve better average PSNR than classic MDVS and SDC. For longer burst length and higher PLR, gains are significantly increased, particularly for longer error burst lengths. Considering BEL=12 and PLR=7% and 10%, the gains comparing with SDC are 2-3dB and comparing with classic MDVS are 1-2dB, considering all rates. For BEL=4, the proposed MDVS architecture improves the decoded video quality, where it is most significantly for PLR=5% and PLR=7%. Note that for higher PLR, the probability of losing both descriptions simultaneously is also higher, which tends to increase the influence of the error concealment at the decoder and to reduce the advantages of MDC streaming. Figs 5.11-5.16 also show a critical point around PLR=3%, which indicates that for lower values of PLR it is better to keep SDC instead of MDVS. These results suggest that such switching point should be used for no serious loss network conditions. The optimal computation of such singularity under different networking conditions is an open issue that deserves further investigation. Note that similar switching points are referred in literature (e.g., [Franchi 2005]). Other recent MDC schemes based on DVC (e.g., [Milani 2010, Fan 2011]) also exhibit similar behaviour at relatively higher PLR (e.g., 5-15%).

As expected, in the lossless case (i.e., PLR=0%), both the SDC and classic MDVS achieve better PSNR in comparison with the proposed MDVS. This is due to the overhead required to encode the side information, since the PSNR of the three streams is compared at exactly the same overall bit rate. The difference of about 2-3 dB is in line with other MDC schemes available in the literature, as discussed in section 5.6.

Nevertheless, in the presence of packet loss, the perceptual quality is better because the variation of PSNR is much lower in the MDVS case. This is shown in Fig. 5.17 for *bus* sequence, by comparing classic MDVS with proposed MDVS, affected by the same lost packets. During the periods affected by packet loss, the proposed MDVS achieves PSNR gains of about 3-4dB and much lower quality variation, i.e., about 3dB variation in comparison with 7dB of classic MDVS.

Finally, Table 5.2, shows simulation results for different sequences for PLR=3%, 5%, 7% and 10% and average BEL of 4 and 12 packets, at 1Mbit/s. Comparing average PSNR

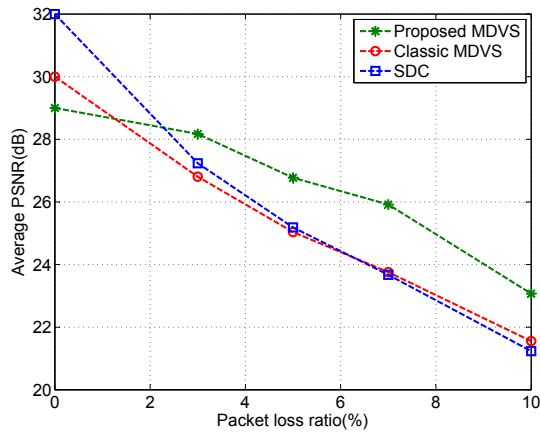


Figure 5.11: Average PSNR for *bus* at 1.25 Mbit/s (Burst length BEL=4).

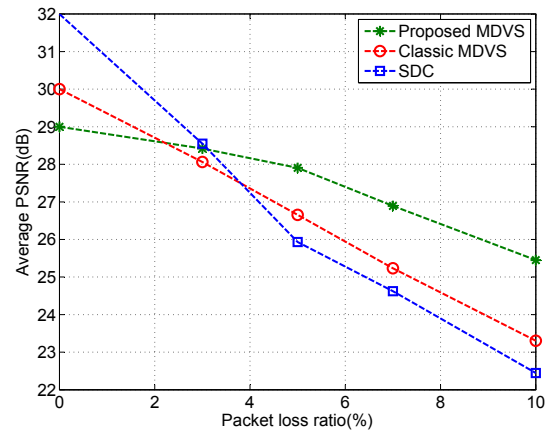


Figure 5.12: Average PSNR for *bus* at 1.25 Mbit/s (Burst length BEL=12).

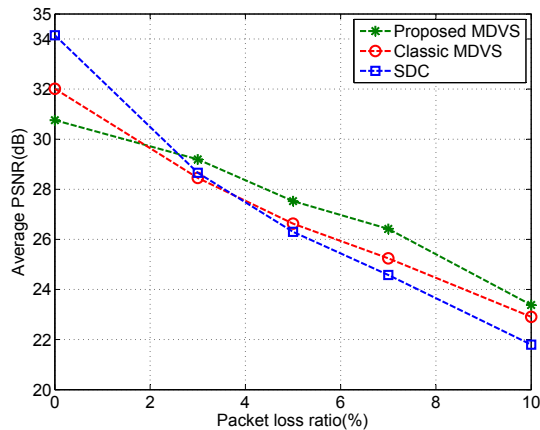


Figure 5.13: Average PSNR for *bus* at 1.8 Mbit/s (Burst length BEL=4).

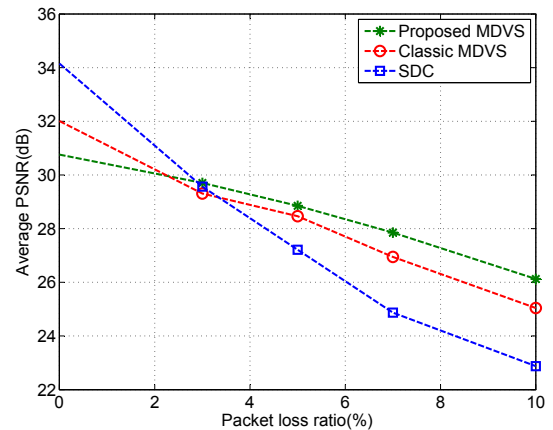


Figure 5.14: Average PSNR for *bus* at 1.8 Mbit/s (Burst length BEL=12).

gains for BEL=4 packets of the proposed MDVS over classic MDVS, these are 1-2dB for PLR=3%, 1.8-2.6dB for PLR=5%, 2-2.5dB for PLR=7% and 1.1-2.6dB for PLR=10%. These results show that the proposed scheme achieves better quality while the small excess rate is beneficial by avoiding drift in case of packet loss. In comparison with SDC the average PSNR gains are -0.3-0.6 dB for PLR=3%, 0.4-1.5 dB for PLR=5%, 0.1-2.6 dB for PLR=7% and 0.1-2dB for PLR=10%. For average BEL=12 packets and comparing the proposed MDVS over classic MDVS, the gains are from 0.1-0.3dB for PLR=3%, 0.9-1dB for PLR=5%, 1.1-1.5dB for PLR=7% and 1.6-2.4dB for PLR=10%. In comparison with SDC the average PSNR gains are 0-1 dB for PLR=5%, 0.2-1.3 dB for PLR=7% and 0-3dB for PLR=10%. Higher gains are obtained in sequences with high motion and texture

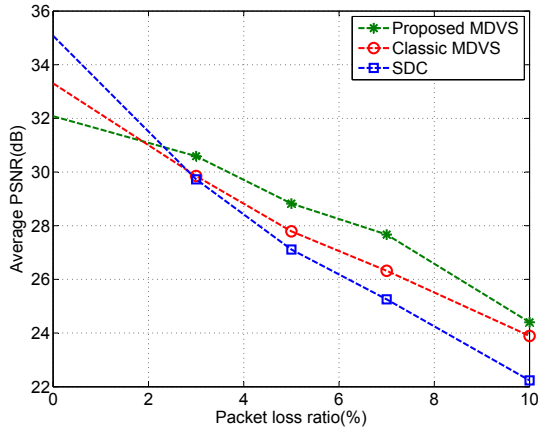


Figure 5.15: Average PNSR for *bus* at 2.16 Mbit/s (Burst length BEL=4).

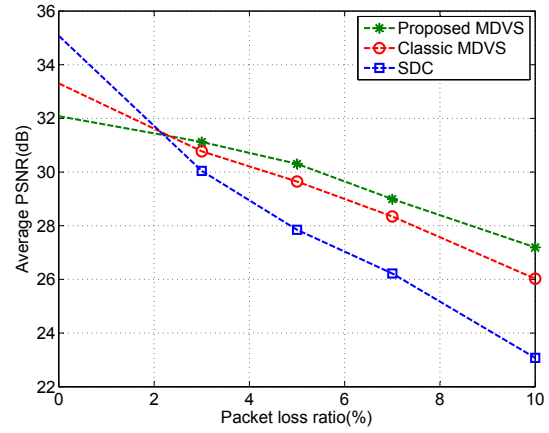


Figure 5.16: Average PNSR for *bus* at 2.16 Mbit/s (Burst length BEL=12).

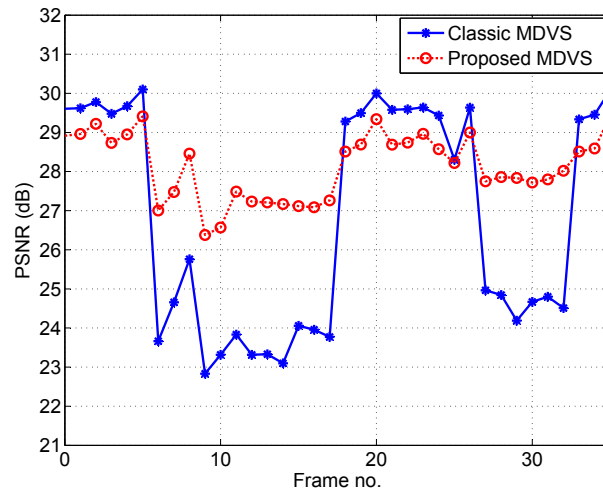


Figure 5.17: Frame by frame PSNR for *bus* sequence.

complexity, like *bus* and *foreman* sequences. This means that the proposed architecture is more efficient to reduce the overall drift distortion in sequences with high motion and texture complexity.

Overall, these results show that network-adaptive MDVS for path diversity consistently improves the robustness of video streaming across networks with multiple paths. Furthermore, the proposed MDVS exhibits better performance than existing classic schemes because of its improved drift characteristics. During packet loss periods, significantly higher average PSNR is obtained at the expense of acceptable redundancy which is necessary for drift compensation.

Table 5.2: Average PSNR in transmission with different packet loss rates for 1Mbps@30Hz

Sequence	BEL=4											
	Packet Loss Rate											
	3%			5%			7%			10%		
	SDC	Classic MDVS	Proposed MDVS	SDC	Classic MDVS	Proposed MDVS	SDC	Classic MDVS	Proposed MDVS	SDC	Classic MDVS	Proposed MDVS
	PSNR(dB)			PSNR(dB)			PSNR(dB)			PSNR(dB)		
Bus	26.9	25.7	26.9	24.9	24.1	25.9	22.2	22.5	24.7	21.0	20.8	23.0
Foreman	33.1	32.5	33.7	30.7	30.4	32.3	27.8	28.4	30.4	27.1	26.4	28.2
Mother	39.9	38.2	39.8	37.2	35.7	38.3	34.9	33.4	35.9	33.8	31.4	34.0
News	37.9	36.7	37.6	34.8	33.6	35.2	31.8	31.5	33.0	29.4	29.4	30.5
City	31.9	30.1	31.9	30.0	28.3	30.9	28.7	26.3	28.8	26.7	24.5	26.8
Sequence	BEL=12											
	Packet Loss Rate											
	3%			5%			7%			10%		
	SDC	Classic MDVS	Proposed MDVS	SDC	Classic MDVS	Proposed MDVS	SDC	Classic MDVS	Proposed MDVS	SDC	Classic MDVS	Proposed MDVS

5.6 MDVS vs MDC: comparative discussion

Despite the inherent differences between MDVS and MDC, due to the different nature of their input signals (i.e. compressed vs uncompressed video in MDVS and MDC, respectively), the rate-redundancy performance of the proposed MDVS can still be compared and discussed in the light of that obtained in previous MDC schemes at PLR=0% (same overall rate in both MDC and SDC), namely, those based on multi-loop approaches that cope better with drift than classic MDVS. The overall performance of MDVS in comparison with SDC was found to have a PSNR drop of about 1.6dB and 3.1dB for rate-redundancies between 50% and 90% (including extra redundancy of side information), using the set of sequences referred to above. Under the same conditions, the multi-loop MDC architectures proposed in [Reibman 2002], [Franchi 2005] and [Tang 2002], exhibit rate-redundancies from 40% to 100% for PSNR drops about 1.69dB to 4dB. Also in [Matty 2005], an open loop MDC scheme is proposed where the results show rate-redundancies between 45% and 100% for PSNR drops between 1.81dB and 3.35dB.

The MDC scheme proposed in [Su 2008], uses a multi-loop approach based on a spatial slice partitioning method previously used in [D. Wang 2005]. This scheme has a temporal partitioning counterpart, also using multiple-loops, proposed in [Wang 2002]. In these papers, the overall rate-redundancy distortion performance (at PLR=0%) was found to be better than that obtained in MDSQ based architectures. However, the coding approaches used by such MDC schemes cannot be applied in MDVS without fully decoding the input SDC video followed by independent MDC encoding. Therefore, the use of these multi-loop MDC methods in the same networking scenarios as MDVS is highly complex, which is a

significant disadvantage in comparison with MDVS. A wavelet-based MDC scheme was recently proposed in [M. Biswas 2009], but exhibits lower performance than MDVS. The results in [M. Biswas 2009] must be combined with those achieved in [Tillier 2007] in order to find out that 4dB quality drop is obtained at a relatively low rate-redundancy (i.e., 30%). Similar conclusions can be derived by comparing MDVS with another recent work in MDC [Fan 2011]. Overall, even though the rate-redundancy distortion performance of the proposed MDVS is affected by the coding distortion present in the input signal, the global rate-redundancy distortion of MDVS is inline with that of different MDC schemes.

5.7 Conclusion

This chapter demonstrates that splitting of compressed video streams into multiple descriptions using a classic MDC architecture leads to unacceptable drift accumulation, which severely affects the quality of decoded video when only one description reaches the decoder. Novel MDVS architectures were proposed to overcome the drift problem. The proposed schemes are effective to prevent drift by using a controlled amount of side information. The experimental results provide evidence that the decoded video quality can be significantly improved at the expense of an acceptable redundancy increase in comparison with classic MDVS, for channels with distinct packet loss rates. Overall the proposed MDVS architecture finds application in multimedia networking heterogeneous environments, where lossy networks with single and multiple available paths co-exist along the same delivery chain.

MDC Rate Control and Priority Video Streaming

Contents

6.1	MDC Hypothetical Reference Decoder	134
6.2	The ρ model for MDC	138
6.3	MDC Rate Control	140
6.3.1	GOP Level rate control	142
6.3.2	Frame Level rate control	143
6.3.3	MDC Rate Control Performance Evaluation	145
6.3.4	Performance over lossy channels	149
6.4	Priority MDC video streaming	152
6.4.1	MD priority streaming scenario	152
6.4.2	Priority Streaming	153
6.4.3	Optimal binary frame classification	154
6.4.4	Simulation Results	157
6.5	Conclusion	160

This chapter addresses the problem of rate control in unbalanced MDC video coding based on Multiple Description Scalar Quantisation (MDSQ). A new rate control method is proposed to dynamically generate two unbalanced descriptions for a given target bit-rate and predefined balancing ratios between descriptions. The proposed rate control method combines the new unbalanced MDSQ method described on Chapter 4 with an extension

of the ρ model which was developed based upon the linear relationship between the rate of each description and the corresponding percentage of zeros of transform coefficients. In particular, it is shown that this linear dependency is maintained when moving from the single description domain (SDC) into the MDC domain. The simulation results show that the proposed rate control algorithm exhibits high accuracy for different target bitrates and balancing ratios between two descriptions without having buffer underflow and overflow. The proposed method is useful in MDC video streaming over asymmetric transmission paths, where the rate of each description can be dynamically adapted to different channel conditions.

This chapter also proposes a robust video streaming scheme for priority networks with path diversity, based on a combined approach of multiple description coding (MDC) with optimal frame classification into two priorities. A binary classification algorithm is proposed to define high (HP) and low (LP) priority network abstraction layer units (NALU), which in turn define the packet priorities. An optimization algorithm is used to find HP frames, based on dynamic programming and relying on minimisation of the packet loss concealment distortion. Hence, the proposed algorithm is able to effectively improve the decoded video without increasing the MDC stream redundancy. The overall performance evaluation, carried out by simulating MDC video streaming over lossy networks with path diversity, demonstrates that the proposed algorithm yields higher video quality for a wide range of packet loss rates (PLR).

6.1 MDC Hypothetical Reference Decoder

The general model in video streaming comprises a video encoder, a streaming server, a transmission channel and a streaming client. In this communication chain, several rate and delay constraints must be considered. Firstly, the video source is compressed with variable bit-rate because the amount of coded data generated by the encoder is dependent on the video content itself. Secondly, the available channel rates depend on the network conditions, which might impose severe constraints during some periods. In general transmission is made over variable bit-rate (VBR), but in particular applications, constant bit-rate (CBR) is also used.

Therefore, since the encoding output rates are not equal to transmission rates, it is necessary to use buffering in order to deal with different variable rates taking into account

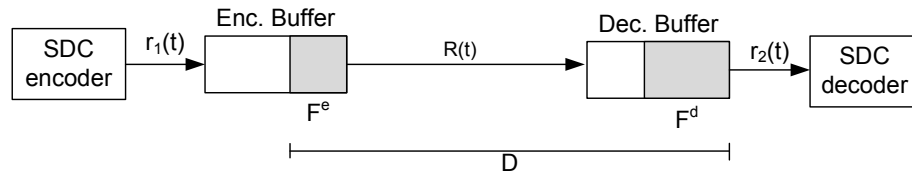


Figure 6.1: Generic SDC HRD model.

the temporal decoding and displaying timing. In CBR applications buffering is used to convert the VBR produced by encoders into the CBR of the transmission channel. In VBR applications, buffering may not be strictly necessary, since these systems only need to feed the decoder at each decoding time. However, buffers are used in VBR to smooth the streaming server output rates. In streaming applications, this buffering capability is normally expressed as a start-up or initial delay. This start-up delay D is defined as the time elapsed since the first bit arrives at the decoder buffer until the first frame is removed from that buffer to be decoded and displayed [Reibman 1992]. This delay D is maximum in video download applications corresponding to the downloading time, because decoding and displaying starts only after receiving the complete sequence. Furthermore, such delay is minimum in low-delay real-time applications, such as video conference applications. In the middle, streaming applications have an initial playback delay, which depends on the output rate, i.e., the decoded video quality and channel rate. Finally, delay is defined as $D = F/R$ where F is the initial buffer fullness (in bits) and R is transmission rate.

Fig. 6.1 shows the buffering model of an SDC transmission chain. The encoder output rate is $r_1(t)$, and the initial buffer fullness of the encoder is F^e . In the particular case of $R(t) = R$, the encoder output feeds the decoder buffer with rate R , where the initial decoder buffer fullness is defined as F^d . Bits are removed from decoder buffer at rate $r_2(t)$ which is the same as $r_1(t)$. Considering a simplified model with null transmission delay, the $r_2(t) = r_1(t - D)$ where $D = F^e/R + F^d/R$, i.e. the overall system has constant delay and encoder and decoder buffers smoothes the variable bitrate variations [Assunção 2000].

In order to maintain the temporal schedule, the decoder buffer must have store sufficient bits to display data at each decoding and display time, i.e., decoder buffer cannot underflow. On the other hand, the encoder buffer has to smooth encoded data at any time, saving all encoded rate, i.e., cannot overflow. Therefore, in SDC streaming the encoding parameters, buffer delay and decoding scheduling have to be set in order to guarantee that buffers never underflow at the decoder or overflow at the encoder. Additionally, the decoder buffer must accommodate network jitter delay that needs to be maintained in

controlled bounds [Stockhammer 2004].

Since the amount of data fetched from the decoder buffer at each decoding/display time depends on the output rates generated by the encoder, this means that the encoder itself needs to know or estimate what will happen at the decoder side. The overall buffering delay is constant between encoder and decoder, while the transferring variable rate from the encoder to its buffer is the as that from the decoder buffer to the decoder itself. This means that the decoder buffer dynamics is complementary to the encoder buffer. Taking this behaviour into account, the encoder uses an hypothetical decoder comprised of a virtual decoder buffer, which is conceptually connected to the encoder output. This virtual decoder is known as the hypothetical reference decoder (HRD) and it is used by the encoder to ensure that the generated bit stream is supported by any possible decoder without producing buffer overflow neither underflow. In this decoder model an SDC stream is transmitted over a channel with either fixed or variable rate and then is decoded after a predefined delay, which is dependent on the buffer size.

Usually, the HRD model is characterized by three parameters (R,B,F) , where R is the peak transmission rate (in bits per second) from the encoder buffer to the decoder buffer, B is the encoder and decoder buffer sizes and F is the initial decoder buffer fullness (in bits) before the decoder can start removing bits from its buffer, *i.e.*, F^e . Thus, the initial buffer delay D is defined by F/R . Usually, to maintain a minimum delay the buffer size is defined as $B = F^e + F^d$. In H.264/AVC an extension of the HRD model with (R,B,F) parameters is proposed in order to deal with variable output rates. For each model a minimum buffer occupancy level is defined in function of several peak output rates in order to avoid buffer underflow and overflow. [Ribas-Corbera 2003].

Based on the SDC model of Fig. 6.1, the MDC communication model of Fig. 6.2 is defined for two descriptions transmitted over different channels. This generic model, considers two encoders, two different channels and three decoders. A pair of encoder and decoder buffers is used for each channel. The MDC encoder buffer 1 receives description 1 with rate $r_1(t)$ and Decoder 1 receives from channel 1 with channel rate $R_{D1}(t)$. In order to maintain constant delay, $B_1 = F^{e1} + F^{d1}$. Additionally, the MDC encoder buffer 2 receives description 2 with output rate $r_2(t)$, and decoder 2 receives data with rate $R_{D2}(t)$, where $B_2 = F^{e2} + F^{d2}$. At the MD decoder each description is independently decoded using Decoder 1 and Decoder 2. The central Decoder 0 merges the two received descriptions in the high quality representation. Taking into account the transmission model and buffering delays, the decoding and display timing at each decoder must be the same. Therefore,

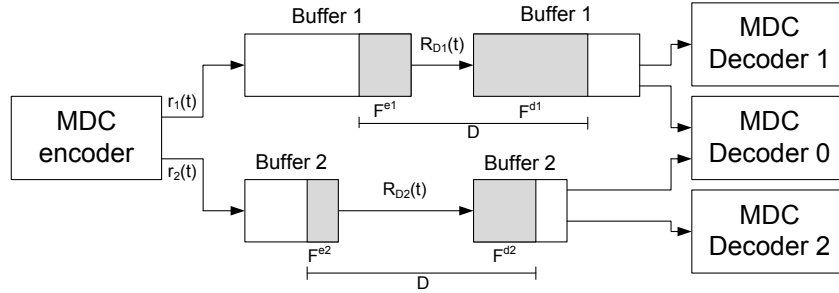


Figure 6.2: Generic MDC HRD model.

considering that each description is transmitted over different channels with different rates, MDC transmission must have the same buffering delay for all descriptions regardless of the individual output rates, which means that the two buffers must have different sizes. On the other hand, since transmission packet delays in each channel are not correlated, buffering needs also to consider an additional bounded delay, whose maximum must be common for each description and limited by decoding and playout timing.

Since the MDC model uses two encoders and three decoders, some modifications in SDC HRD model (R, B, F) must be done. In addition, tree parameter set, *i.e.*, (R_1, B_1, F_1) , (R_2, B_2, F_2) and (R_0, B_0, F_0) are defined, one for each decoder. (R_0, B_0, F_0) depends on the individual decoders where $R_0 = R_1 + R_2$, $B_0 = B_1 + B_2$ and $F_0 = F_1 + F_2$. Therefore, $D = F_0/R_0 = F_1/R_1 = F_2/R_2$ *i.e.*, the same buffer delay is obtained for each decoder. Since decoder buffers must have the same timing constraints, each description can be independently decoded if only one description is available and at the same time, jointly decoded if both descriptions are available. Therefore, considering that each description is transmitted over different channels with one buffer for each decoder, if none of the individual buffer underflows, then the joint buffer does not underflow, maintaining the common timing for decoding and displaying.

For each MDC HRD model, an equivalent encoder buffer size must be defined, *i.e.*, B_0, B_1, B_2 . However, in the MDC model only two encoders are used, each one connected to a different channel. Since the central decoder buffer size is $B_0 = B_1 + B_2$, then the same buffer size is used at encoder. In practice, the encoder uses an equivalent buffer B_0 with input rate $r_1(t) + r_2(t)$ and output rate $R_1 + R_2$. Since the encoder output rate is $r_1(t) + r_2(t)$, then the encoder uses the B_0 dynamics to adapt MDC coding parameters in order to comply with output rate constraints. Therefore, by controlling the B_0 buffer overflow at the encoder, none of the decoder buffers underflows.

6.2 The ρ model for MDC

From rate-distortion theory, distortion of coded signals is directly related with the rate used by a given source. Particularly, in transform coding of images and video this is the most important factor that should be optimized in order to tune the encoding parameters for a set of rate and distortion constraints, *i.e.*, to find the best $R - D$ pairs of the given image or video sequence. The major problem is to determine a quantization parameter QP to achieve the target coding bitrate or target picture quality. To this end, it is necessary to analyze and estimate the $R - D$ behaviour of the video sequences. This behavior is obtained characterizing the rate-quantization ($R - Q$) and distortion quantization (R-D) functions. Consequently, the R-D functions are necessary to implement efficient methods for computing appropriate QP values, which correspond to a target output rate or quality.

This can be done by using two different approaches: the operational and model based. In the operational based approach, several R-D points of the source are obtained, and then interpolated in order to define the R(D) function and then to find the best coding parameters for a given rate-distortion tradeoff. However, this approach is computationally intensive because it is necessary to find a very large number of R-D points for each coding unit. Other distinct approach is to use rate-quantization (R-Q) or distortion quantization (D-Q) models relating the rate or distortion for a given source to the corresponding quantization parameter. Considering video coding, these models are applied at different coding unit levels, *i.e.*, frame, slice or macroblock.

Regarding model based approaches for video coding, two main different models are used. The most used approach is the quadratic model, that relates the texture rate for a given Mean Absolute Difference (MAD) with the quantization parameter (QP) [Chiang 1997]. Usually, this model is continuously updated in order to improve the target rate estimation. On the other hand, MAD values are based on previous coding units, depending on the coding level. Other proposed R-Q model is based on the so-called ρ model [He 2001]. The ρ model is based on the linear dependence between the coding rate and the percentage of null transform coefficients ρ , after quantization. This model is very simple and can be expressed by the following equation, using a single model parameter ϕ ,

$$R(\rho) = \phi(1 - \rho), \quad (6.1)$$

where $R(\rho)$ is the rate obtained in function of ρ value, *i.e.* the percentage of null co-

efficients. The model parameter ϕ is obtained from perviously coded units. The R-Q function is indirectly found by relating the source statistics with quantization stepsize. Therefore, the source statistics need to be computed which can be achieved based on parametric models [Milani 2008]. Other approach is to use previously coded units to find an operational source statistics allowing to find an adequate QP for a given ρ value.

The ρ model was extensively studied in SDC context, and applied to different sources, such as image and video using several transform coding schemes. However it has not been applied to the MDC context, in particular to MDSQ. Using MDSQ it is necessary to find the coding parameters that achieve the target output rates for each description. Therefore, an open issue to be investigated is the relation between the coding rate in each description, the quantization parameter and the MDSQ parameters. If a linear relation between coding rate and percentage of null coefficients does not change with MDSQ insertion, then ρ model can be used to rate control schemes for MDSQ video coding. Further investigations have been carried out to study this linear relation in order to find out whether this is maintained in each description after applying a cascade of quantisation and MDSQ to H.264/AVC transform coefficients. After an exhaustive experimental study using several different sequences, the conclusion is that there is also a linear relation between the percentage of zeros before MDSQ and the rate of each balanced (*i.e.* the same rate) and unbalanced description (*i.e.* different rates).

Relevant results regarding these issues are shown in Figs. 6.3, 6.4, 6.5 and 6.6, obtained from encoding I and P frames. In this figures, the represented bits/description/frame are the total bits used to encode each description. It would be possible to use only the texture bits for this analysis, but considering the overall rate, a direct function between rate and luminance features allows to simplify rate control algorithms since is no necessary to consider each rate component separately. These figures show that in both cases, the linear model can be used for unbalanced descriptions with the interesting characteristic of keeping the proportionality fairly constant among different unbalancing ratios. This feature is quite evident in balanced MDSQ and also for high rate description in unbalanced MDSQ. Since the rate is measured as total bits/description/frame, the proportionality in lower rate descriptions is not so evident because the (overhead (non-texture) bits have a higher relative weight in the overall rate. These results also show that MDSQ does not change the linear relation between rate and percentage of zeros, which is of great practical interest to be used in rate control algorithms. Therefore, using the linear model based on the higher rate description, only target rate for each description needs to be taken into

account to set the coding parameters to a given target rate.

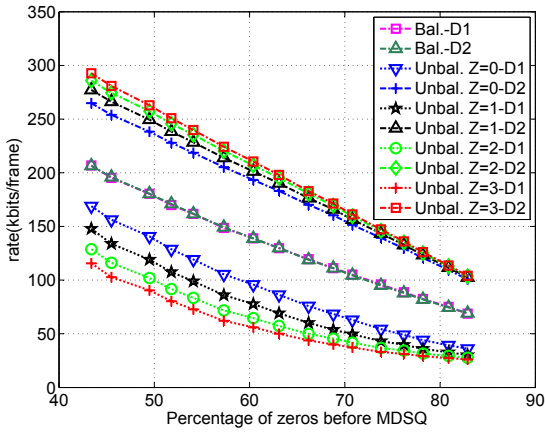


Figure 6.3: Relation between percentage of zeros and rate-*Bus* sequence-I frames.

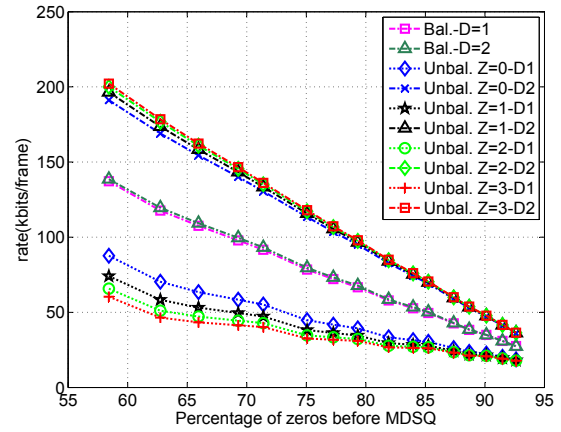


Figure 6.4: Relation between percentage of zeros and rate-*Bus* sequence-P frames.

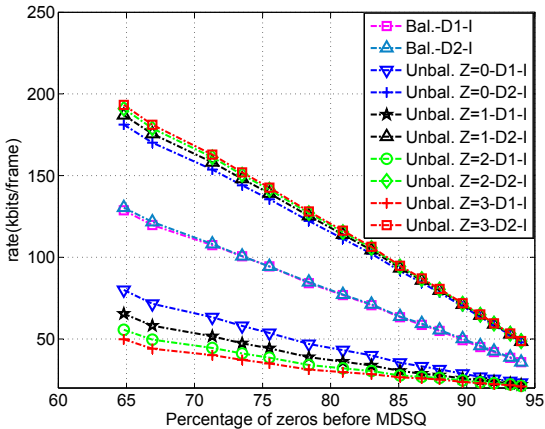


Figure 6.5: Relation between percentage of zeros and rate-*Foreman* sequence-I frames.

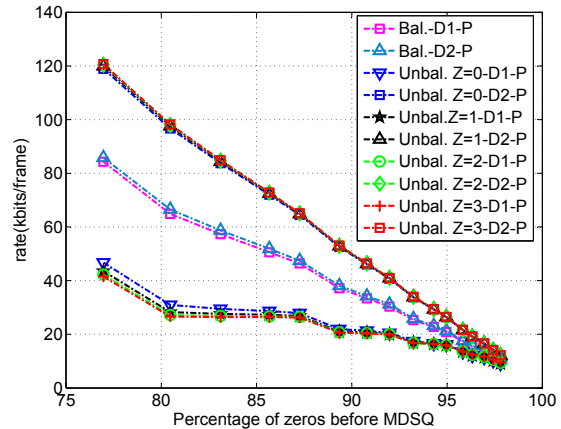


Figure 6.6: Relation between percentage of zeros and rate-*Foreman* sequence-P frames.

6.3 MDC Rate Control

The previous analysis on MDC HRD allows to define the necessary conditions to joint rate control schemes considering the overall rates generated by each description. As shown before, since the decoding time of both descriptions must be synchronized, an equivalent joint encoder buffer for all descriptions can be considered instead of individual encoder buffers for each description. Therefore, this assumption can be exploited in rate

control schemes considering an overall target bit rate for MDC considered all of the rate components.

Usually, MDC rate control schemes define a certain redundancy level for a given packet rate loss. Nevertheless, each channel has rate constraints that need to be considered, which means that traditional rate control schemes should take into account the specific MDC scheme. In particular, MDSQ uses an equivalent quantisation stepsize for each description that is indirectly related with the quantization parameter that defines central distortion. Consequently, rate control schemes based on the mean absolute difference used in Advanced Video Coding schemes are not suited to define the QP value for a given target bit rate, opening room for further research in rate control schemes capable of effectively adapting the MDC output rates to match a given target.

A new approach for MDSQ-based MDC is proposed using the concept of index assignment for asymmetric descriptions as reference. This approach relies on the new unbalanced MDC scheme based on MDSQ for video coding, where the rate of each description is dynamically allocated by changing the index assignment tables of MDSQ. The overall rate is controlled by an MDC rate control method capable of producing two descriptions with different rates based on the evidence of the linear dependency between the percentage of zeros of transform coefficients and the rate achieved each description after MDSQ. Based on these two findings, a new rate control scheme is proposed in order to find the QP values to be used on MDSQ scheme for advanced video encoding for a given target rate. This rate control scheme was defined to be independent of the redundancy level allocated for any given packet loss rate. Thus, it only depends on the overall target rate for each description, which can include rate-redundancy allocation.

The rate control method comprises two levels: GOP Level and Frame Level. In the GOP Level, the rate control algorithm is able to find the rate budget in order to keep an appropriate buffer occupancy. This method follows the reference model based on [Li 2006] used on H.264/AVC. The main difference is that an additional target for the balancing ratio is included to define the rate of each description, given the overall target rate. At the frame level, the rate control method is comprised of two stages. The first stage defines an overall target rate for each frame and considering a predefined balancing ratio between descriptions the target rate for each one is also defined. The second stage finds the coding parameters to be used, such as the quantization parameter QP for a given target rate per frame. An MDSQ-adapted ρ model based on the linear relation between the obtained rate of each description and the percentage of null transform coefficients existing before

MDSQ is used. These findings provide the ability to easily find an accurate QP value capable of matching the overall target rate.

6.3.1 GOP Level rate control

The GOP level rate control defines the total number of bits allocated to each GOP and the initial quantization parameter of each GOP considering the individual target rates defined for each description. The total number of bits allocated for one GOP is determined as defined in the JVT-G012 rate control method [Li 2006]. The unbalance between descriptions is achieved by using the index assignment tables described in section 4.2.

The initial target rate T_i defined for the i th GOP is defined as

$$T_i = \frac{(R_{D1_i} + R_{D2_i})}{f} N_i + (T_{i-1}(N_{i-1}) - \frac{B}{8}), \quad (6.2)$$

where T_i is the rate allocated for all GOP, $R_{D1} + R_{D2}$ is the output rate considering both descriptions, f is the frame rate, N_i is the number of frames in GOP i and $T_{i-1}(N_{i-1})$ the remaining bit budget of the $(i-1)$ th GOP. A minimum buffer occupation should be kept at $\frac{B}{8}$. Considering variable bit rate (VBR) transmission, the output rates are not constant and

$$T_i(j) = T_i(j-1) - b_i(j-1) + \frac{(R_{D1_i}(j) + R_{D2_i}(j)) - (R_{D1_i}(j-1) + R_{D2_i}(j-1))}{f} (N_i - j + 1). \quad (6.3)$$

If constant bit rate (CBR) is considered, the output rates $(R_{D1_i}(j) + R_{D2_i}(j)) = (R_{D1_i}(j-1) + R_{D2_i}(j-1))$ and

$$T_i(j) = T_i(j-1) - b_i(j-1), \quad (6.4)$$

where $b_i(j-1)$ is the number of bits used to encode both descriptions in the $j-1$ frame. Using this approach, the MDC rate control scheme is able to generate a constant output

rate which is independent from the target bitrate of each description.

At GOP level the initial quantization parameter of the i th GOP, $QP_i(1)$ is defined as:

$$QP_i(1) = \max\left\{\min\left\{\frac{\sum_{j=1}^{N_p(i-1)} \overline{QP}_{i-1}(jL+1)}{N_p(i-1)} - \min\left\{2, \frac{N_{i-1}}{15}\right\}, QP_{i-1}(1) + 2\right\}, QP_{i-1}(1) - 2\right\}, \quad (6.5)$$

where $N_p(i-1)$ is the total number of I and P frames in the $i-1$ GOP, L is the number of B frames between P frames, $\overline{QP}_{i-1}(jL+1)$ is the mean quantization parameter values of the j th GOP I and P frames in the $(i-1)$ th GOP. In conclusion, $QP_i(1)$ is adapted in function of the GOP length and the available channel bandwidth, jointly considering the available rate for all descriptions.

6.3.2 Frame Level rate control

The rate control at frame level determines the overall target rate for each frame, taking into account the individual rates of each description. The QP values for B frames are obtained from interpolation of neighboring anchor frames (I and P) and depends on its distance. The QP difference between adjacent frames cannot be higher than 2. Furthermore, the QP values for P frames are obtained based on two different steps. This first step is to determine the target bit rate for each frame, which will be distributed among descriptions afterwards. The second step is to find the corresponding QP for the given target.

Target rate for P frames The target bit rate for each P frame is defined according to the rate for each description, i.e., is the sum of the individual description rates. The target bit rate takes into account the buffer occupancy after coding each frame, and also the complexity weights of P and B frames and QP values. Since complexity weights are dependent on the generated bits of corresponding frames, in MDSQ this corresponds to the sum of bits used in each description, which means this is not a different problem than in SDC coding.

Given the target bit rate for each anchor frame $T_i(j)$, it is necessary to find the corre-

sponding rate for each description. Since unbalancing ratio is defined at GOP level, and defining the target balancing percentages as (π_1, π_2) for description k , ($k = 1, 2$), the target rate for each description is defined as,

$$\begin{cases} T_i(j)_1 = T_i(j) * \pi_1 \\ T_i(j)_2 = T_i(j) * \pi_2 \end{cases} \quad (6.6)$$

QP based on ρ model for MDC Based on the findings described in section 6.2, a ρ model was devised to determine the quantisation parameter (QP) for a given target rate for each description. Since the target rate for each description depends on the balancing ratio and is computed based on overall target rate for the current frame, the ρ model is only applied to one description. Taking into account the results previously obtained, the model has higher linearity for the description with higher rate. Therefore, in order to find the QP value, the target rate and corresponding model parameter is based on the description with higher rate, named as main description in this context. For each frame, the model parameter is computed based on the main description by using the source statistics of the last frame with the same type. Thus, the model parameter ϕ_i for frame i is computed by,

$$\phi_i = R_{i-1} / (1 - \rho_{i-1}). \quad (6.7)$$

After computing ϕ_i for a given frame and target bitrate for the main description, the corresponding ρ value (i.e., percentage of zeros) is determined using equation 6.1. Then the corresponding QP value is found by using the histogram of the transform coefficients. For each quantisation step-size δ , the corresponding percentage of zeros is determined from the histogram and a lookup table is built, establishing the relationship between ρ and δ . Then, using the quantisation step-size previously found, the quantisation parameter QP is determined from the following expression, already used in H.264/AVC,

$$\delta = 0.67 \cdot 2^{\left(\frac{QP}{6}\right)}, 0 \leq QP \leq 51. \quad (6.8)$$

6.3.3 MDC Rate Control Performance Evaluation

The performance evaluation of the proposed MDC rate control method is focused on three aspects: i) the R-D performance; ii) the rate control accuracy; iii) buffer occupancy. The sequences *foreman* and *bus*, CIF resolution, frame rate= 15Hz, GOP=IPBBP... and 16 frames/GOP was used in the simulation. The MDC video architecture described in Chap. 4 was used to implement the proposed rate control method for unbalanced MDC. A buffer size of $bitrate * 1.5$ (in bits) with an initial delay of $buffer\ size * 0.8$ equivalent to 1.2 sec delay was used in these simulations which is an acceptable delay to streaming applications. The coding efficiency of the proposed unbalanced MDC is compared with the balanced MDC and SDC. The overall target bitrate and the target balancing ratio defined as the balancing percentages (π_1, π_2) are defined as setting parameters.

Figs. 6.7 and 6.8 show the PSNR obtained from both unbalanced descriptions for different balancing ratios. The figures show that unbalanced MDC has equivalent overall RD performance in comparison with balanced MDC, since the rate control scheme is mainly dependent on the buffer occupancy at each coding instant, independently of the target rate for each description. Nevertheless, it is noted that for higher rates and higher balancing ratios, the rate-distortion performance is improved. The unbalanced MDC improves the coding efficiency comparing with balanced MDC, because the number of zero coefficients in one description is increased while the other one carries the full value of a higher number of non-zero ones, i.e., more coefficients are not split between the two descriptions. The same behavior for unbalanced MDC was found in [Kim 2005], though in a different context. The loss in coding efficiency in comparison with SDC is about 1.5-2dB which is inline with the obtained using constant QP values.

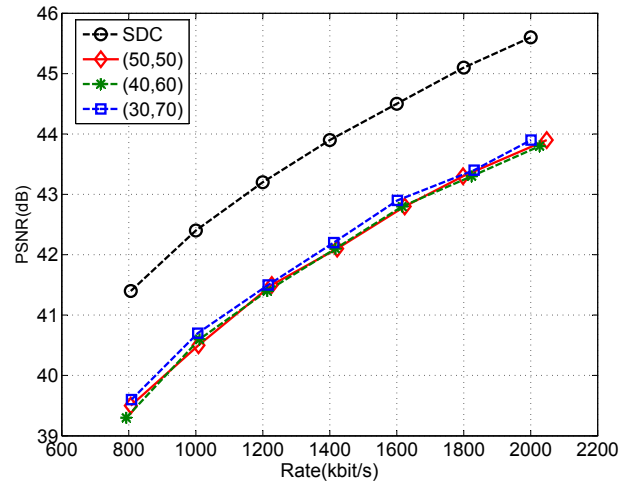
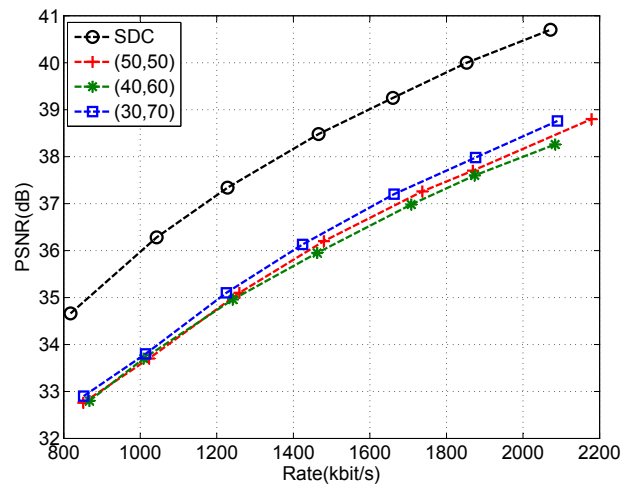
Figure 6.7: Rate control R-D performance - *Foreman*.Figure 6.8: Rate control R-D performance - *Bus*.

Table 6.1: Rate control accuracy

Seq.	Target Rate (kbit/s)	Target Bal. $(\pi_1, \pi_2)=(50,50)$			Target Bal. $(\pi_1, \pi_2)=(40,60)$			Target Bal. $(\pi_1, \pi_2)=(30,70)$		
		R_{D1} (kbit/s)	R_{D2} (kbit/s)	(π_1, π_2)	R_{D1} (kbit/s)	R_{D2} (kbit/s)	(π_1, π_2)	R_{D1} (kbit/s)	R_{D2} (kbit/s)	(π_1, π_2)
Foreman	800	399	407	(49,51)	303	489	(38,62)	275	533	(34,66)
	1000	501	507	(50,50)	381	630	(38,62)	328	678	(33,67)
	1200	608	619	(50,50)	447	766	(37,63)	379	837	(31,69)
	1400	704	718	(49,51)	514	903	(36,64)	435	977	(31,69)
	1600	804	820	(49,51)	579	1037	(36,64)	482	1120	(30,70)
	1800	890	908	(49,51)	645	1179	(35,65)	535	1296	(29,71)
	2000	1014	1034	(49,51)	714	1313	(35,65)	565	1436	(29,71)
Bus	800	427	424	(50,50)	326	541	(38,62)	274	578	(32,68)
	1000	525	529	(50,50)	391	657	(37,63)	320	706	(31,69)
	1200	637	642	(50,50)	470	800	(37,63)	371	874	(30,70)
	1400	754	762	(50,50)	558	953	(37,63)	424	1043	(29,71)
	1600	833	842	(50,50)	609	1049	(37,63)	456	1162	(28,72)
	1800	962	972	(50,50)	679	1171	(37,63)	517	1365	(27,73)
	2000	1045	1056	(50,50)	782	1334	(37,63)	565	1476	(28,72)

The accuracy of the proposed rate control method was evaluated in order to find the magnitude of deviations from the specified target. Table 6.1 shows the rate obtained for description 1 (R_{D1}), description 2 (R_{D2}), and the resulting balancing (π_1, π_2) between descriptions. The results show that achieved bitrate closely follows the specified target, as well as the unbalancing ratio defined for both descriptions. From table results, target deviation is around (1-6)%, which is a good performance, considering that ρ model is only applied to I and P frames. Using B frames the distance between frames is increased, thus reducing the model accuracy. Better results are achieved if coding profiles using GOPs with only I and P frames.

Figs. 6.9, 6.10 and 6.11 show the buffer occupancy and bits/description/frame for a target

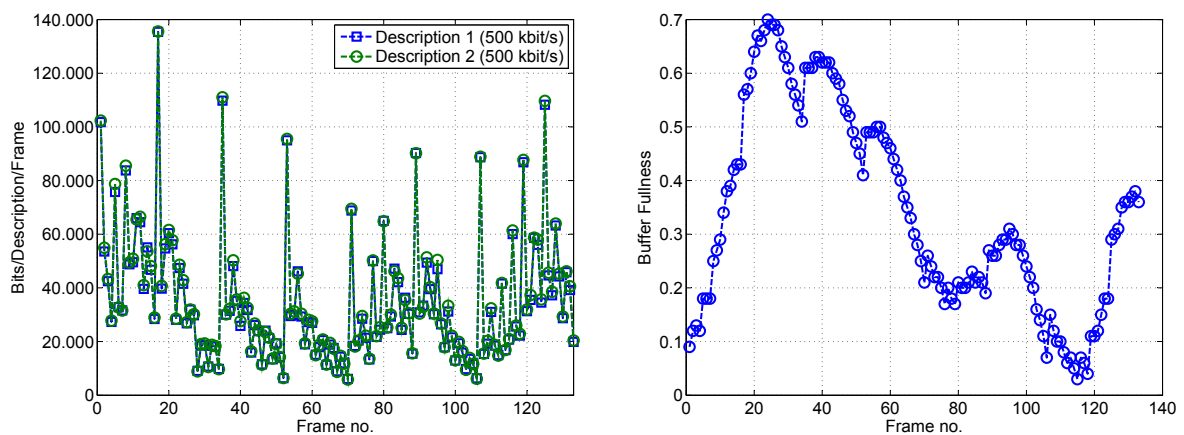


Figure 6.9: Bits/Description (left) and Buffer occupancy (right), *foreman* sequence-Target bitrate 1Mbit/s.

rate of 1Mbit/s and several unbalancing ratios. In Regard to buffer occupancy, this is maintained within appropriate levels for all balancing ratios, which means that there was not any occurrence of buffer underflow/overflow. Note that B_{min} , *i.e.*, the minimum buffer occupancy is set to $B_{max}/8$, which is the minimum obtained, for instance in Fig. 6.10. The evaluation results show also the target distribution between descriptions and buffer dynamics. Description 2 is defined as main description in the ρ model where output rate resulting from description 1 is dependent on the MDSQ scheme. This fact determines distinct buffer dynamics for each MDSQ scheme, without significantly changes in the overall target rate, maintaining the balancing ratio between descriptions.

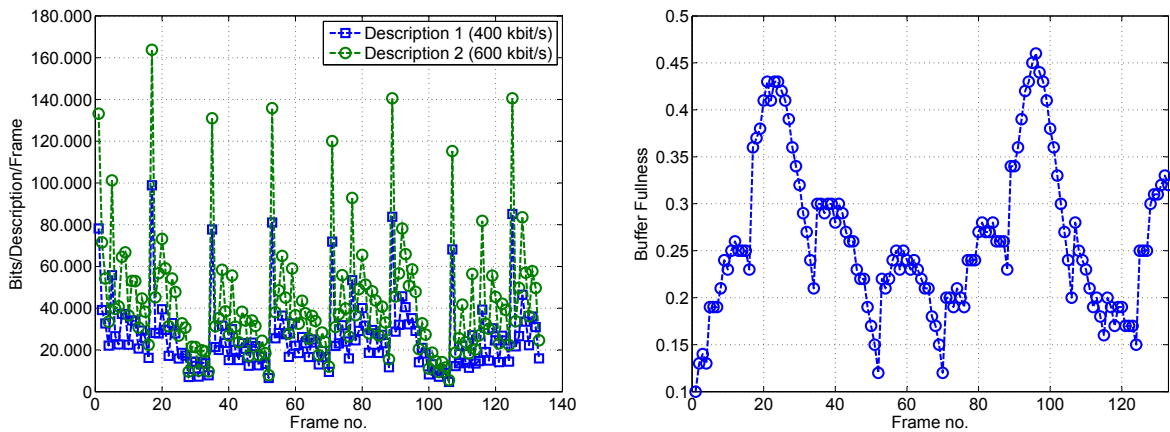


Figure 6.10: Bits/Description (left) and Buffer occupancy (right), *foreman* sequence-Target bitrate 1Mbit/s.

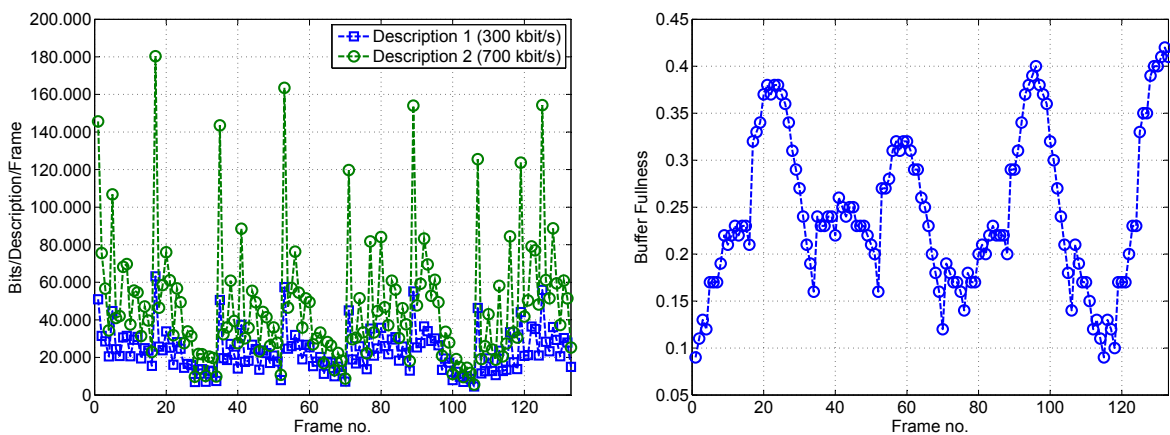


Figure 6.11: Bits/Description (left) and Buffer occupancy (right), *foreman* sequence-Target bitrate 1Mbit/s.

6.3.4 Performance over lossy channels

The proposed rate control method to the central encoder rate of unbalanced MDSQ scheme was evaluated in packet loss environments. In this simulation study, the open loop MDSQ scheme is evaluated for different channel rates and compared with SDC transmission. Moreover, since the proposed MDSQ architecture uses additional side information in order to compensate for drift distortion at the decoder, an evaluation of closed-loop MDSQ is carried out for channels with different available bandwidth and packet loss rates. In this case the side information is separately coded using a fixed rate per description. The evaluation intends to show that unbalancing rates can be effectively used when available rates are not the same at each channel under different packet loss rates.

Channels with the same PLR In this section different channel rates are simulated under the same packet loss rate. Experimental results were obtained using the *foreman* and *bus* sequences. The total rate was defined as 1.2 Mbit/s, 15Hz, using a fixed number of 10 packets per frame. Burst packet loss was simulated using a Gilbert-Elliott 2-state Markov model in order to generate different average packet loss rates and mean burst duration [Li 2009]. In order to obtain statistically meaningful results, the transmission of each sequence was simulated 50 times under the same network conditions, i.e., average PLR values of 3%, 5%, 7% and 10% and average burst error length of 4 packets.

Figs. 6.12 and 6.13 compare the performance over lossy channels of open-loop MDSQ (balanced and unbalanced) with SDC, when each channel has the same PLR. The results show that open-loop MDSQ achieves better results in comparison with SDC. These gains are dependent on the used balancing ratios, where the average PSNR decrease for higher balancing ratios. This is an expected result since the lower rate description has a worse quality, and consequently, worsening the average PSNR. Nevertheless, the results show that using open-loop unbalanced MDSQ schemes, results in average results better than SDC, around 2dB for PLR=5% and 3dB for (50,50) and (40,60) balancing ratio, respectively and close to 4dB for higher PLR values. Overall, the performance of unbalancing open-loop MDSQ in channels with different rates is highly dependent on the amount of information available in each description because mainly of the side reconstruction in lower rate description.

Figs. 6.14 and 6.15 show the performance results using close-loop MDSQ with fixed rate in side information. The output rate was set to 1Mbit/s with 20% side informa-

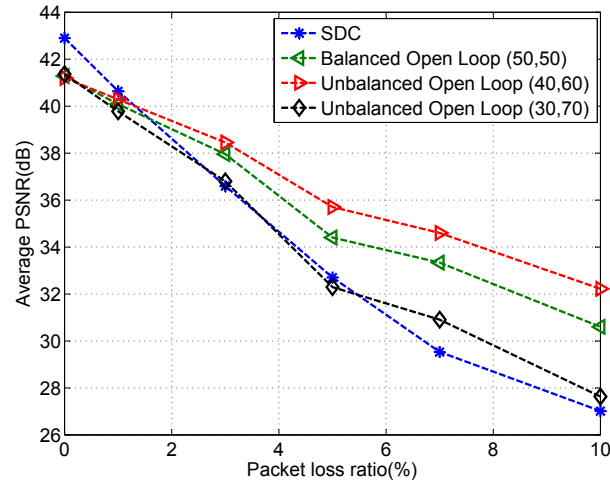


Figure 6.12: Average PSNR for mean burst length $L=4$, *foreman* sequence.

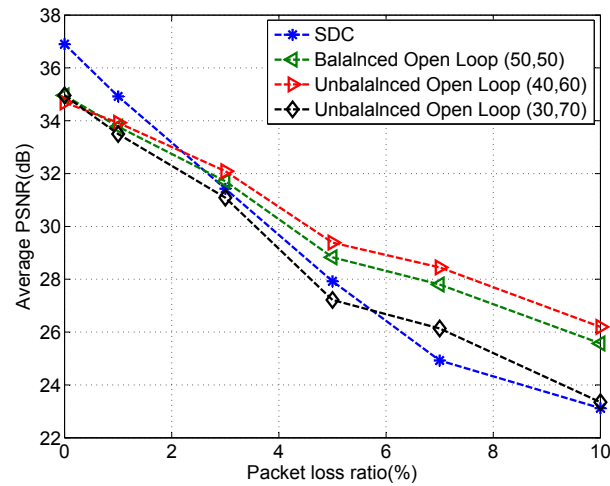


Figure 6.13: Average PSNR for mean burst length $L=4$, *bus* sequence.

tion redundancy equally distributed over each description, making the overall rate of 1.2 Mbit/s. The side information is distributed by the two descriptions in order to maintain the balancing ratio from rate control scheme. The simulation results show that a small amount of side information in each description, and maintaining the balancing ratio between descriptions, achieves significantly better results than SDC and open-loop MDSQ. For instance, for PLR=5% the PSNR gains obtained in comparison with SDC are near 5 dB and 6dB for (50,50) and (30,70) using MDSQ schemes in *foreman* sequence respectively. For the same PLR, the PSNR gains are between 3-4 dB for *bus* sequences under the same conditions. Additionally, these quality can reach gains can achieve almost 6-8 dB for PLR=10%.

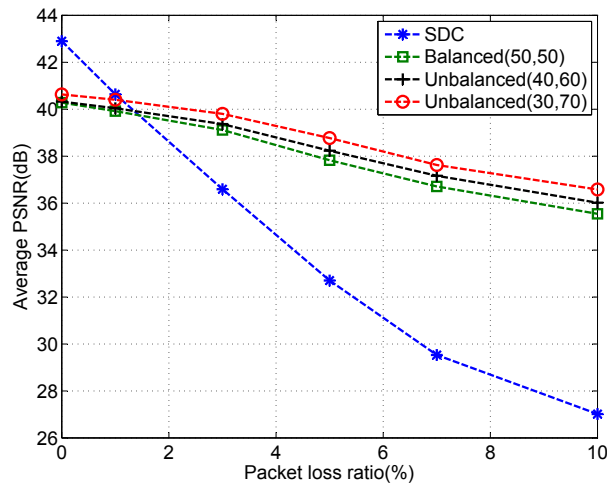


Figure 6.14: Average PSNR for mean burst length $L=4$, *foreman* sequence.

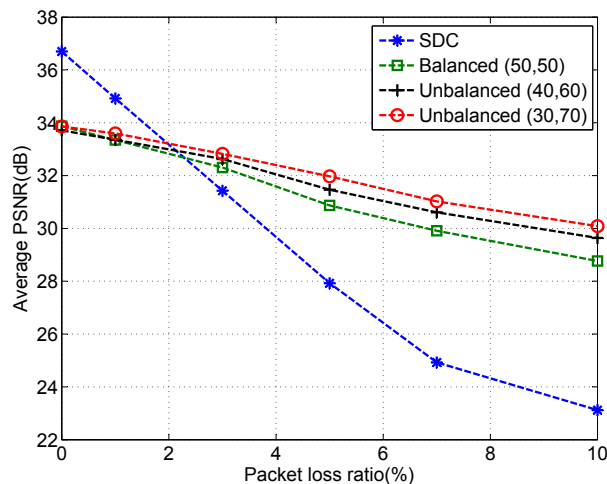


Figure 6.15: Average PSNR for mean burst length $L=4$, *bus* sequence.

Overall, comparison between balancing and unbalancing MDSQ show that although having different rates in each description, the average PSNR for unbalanced MDSQ is slightly better than balanced MDSQ at the same rate. This fact is explained by two main reasons. Firstly, the side information is sufficiently efficient to guarantee a side description decoding with a good quality. Secondly, as shown in Chap. 4, the side reconstruction in unbalanced MDSQ is more efficient which means that better quality in side reconstruction is obtained considering the same output rate. This results show that proposed unbalanced MDSQ scheme is an efficient MDC scheme to be used in path diversity applications based on channels with different available capacity.

6.4 Priority MDC video streaming

This section proposes a method to improve the robustness of MDC video streaming over multiple path channels without increasing the MDC source redundancy, using a binary frame classification approach for prioritised transmission. In the past, several methods have been proposed in the literature to classify coded video units as was explained in Chapter 3. This new approach is different from existing ones, by using pre-processing to classify frames before encoding, based on minimization of a distortion measure that takes into account the loss of non-priority frames. The proposed scheme uses an optimal packet priority classification method that identifies a limited set of high priority frames within a temporal sliding window. These high priority packets carry those frames which minimize the distortion of the decoded video affected by packet loss.

The optimization criterion is based on the minimization of the frame loss concealment distortion assuming the conservative concealment model of frame-copy method. Prioritised network abstraction layer units (NALU) are then packetised according to the NALU priority of each description and the corresponding packets are classified as either high-priority (HP) or low-priority (LP). This classification is used to set different levels of Quality of Experience (QoE) for HP and LP packets, which means that the most important NALU are always received at the decoder in order to maximise the decoded video quality. A novel aspect of this method is the independence from other source coding RD optimized error resilient tools, which allows its use regardless of the type of video encoder. The overall performance evaluation shows the effectiveness of such approach for different ratios of HP frames and a wide range of packet loss ratios.

6.4.1 MD priority streaming scenario

MDC video streaming has inherent error/loss robustness, nevertheless, better QoE can be achieved when the multiple paths available for transmission comprise channels with different transmission priorities. In this case, the open problem for research is how to find the best packet priority assignment at the MDC encoding/streaming server, based on the relevance of video content.

Figure 6.16 shows a video streaming scenario where the proposed priority scheme might be useful. An MDC streaming scenario with two disjoint paths is assumed, including an optimal priority frame classification scheme used to packetise each NALU according

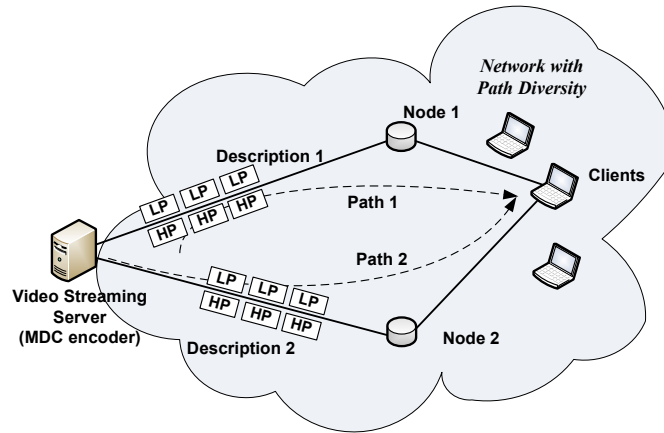


Figure 6.16: MD priority streaming scenario.

to its importance for the receiver. Then those packets classified as HP are delivered with no losses, while the LP ones are prone to transmission losses. Thus, packet priority classification is used to assign different levels of QoS to the source packets. Although only two channel paths are considered, the proposed method can be used in networks with high diversity, i.e., more available paths. Nevertheless, the use of two independently streams in MDC is regarded as a good compromise between capability for error recovery and compression efficiency [Apostolopoulos 2001a].

6.4.2 Priority Streaming

Figure 6.17 shows the proposed optimal MDC priority streaming scheme. The optimal binary frame classification described in section 6.4.3 is applied to a temporal window comprising several frames. Prioritised NALU are then packetised according to the NALU priority of each description and the corresponding packets are classified as either high-priority (HP) or low-priority (LP). This is achieved by setting the first byte in the NAL unit packet, the *nal_ref_idc* (NRI) bits, that identifies the relative transport priority of each NAL unit [Wenger 2005].

After reordering the received packets, the decoder knows which packets are lost in any description and then decides how to decode the received stream. If both corresponding packets are received, then the MDSQ decoder merges the respective decoded indices in order to obtain the original transform coefficients. If only the packet of one description is received, then the decoder uses the side information in order to decode the corresponding slice with controlled mismatch. In the case where both packets are lost, then error

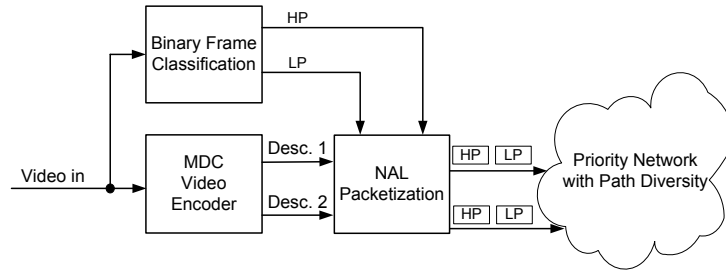


Figure 6.17: Priority MDC streaming scheme.

concealment must be used (in this work we use the conservative frame-copy method).

6.4.3 Optimal binary frame classification

The optimal binary frame classification algorithm identifies a set of m HP frames and its complementary set of $n-m$ LP ones, from a temporal video segment with n frames, defined by a sliding window over the original sequence. The classification of HP frames is based on minimization of the frame loss concealment distortion at the decoder. Given the set of HP frames, the whole temporal segment can be reconstructed using any interpolation algorithm for concealment. The frame-copy method was chosen for this work, i.e., zero-order interpolation. The algorithm finds the set of $m < n$ HP frames which is the best representation of its corresponding temporal segment. This means that the temporal segment is fully reconstructed from the HP set of frames with minimum concealment distortion, i.e., assuming that all LP frames are missing and concealed by interpolation of the HP ones.

6.4.3.1 Definitions and Formulations

Let a temporal segment of n frames be denoted by $F = \{f_0, f_1, \dots, f_{n-1}\}$, where the subscripts represent the temporal order of frames. The corresponding HP frame set denoted by $P = \{f_{l_0}, f_{l_1}, \dots, f_{l_{m-1}}\}$, where l_k is the frame index (referred to the source set F) of the k^{th} element of the set P . P is determined by the index frame selection l_0, l_1, \dots, l_{m-1} , such that $l_0 < l_1 < \dots < l_{m-1}$ are defined in the temporal segment F . For example, given a temporal segment $F = \{f_0, f_1, f_2, f_3, f_4\}$ a possible HP frame set would be $P = \{f_0, f_3\}$, with $l_0 = 0$ and $l_{m-1} = 3$.

The reconstructed set $F' = \{f'_0, f'_1, \dots, f'_{n-1}\}$ from the HP frame set P is obtained using

frame-copy by substituting missing frame with the most recent frame that belongs to P , i.e., the same frame is repeated along the time where HP frames do not exist, that is

$$f'_k = f_i, \quad i = \max(l) \quad : \quad s.t. \quad l \in \{l_0, l_1, \dots, l_{m-1}\}, i \leq k. \quad (6.9)$$

The frame concealment distortion $d(f_k, f'_k)$ quantifies the difference between frame f_k of F and its corresponding frame f'_k from F' . The mean squared error (MSE) between the two frames is used as the measure for computing the concealment distortion, but other metric can be used, since the optimal solution is independent from the distortion criteria. Note that if f_k is belongs to P , then $d(f_k, f'_k) = 0$. The MSE measure can be calculated as:

$$d(f_k, f'_k)_{MSE} = \frac{1}{h \times w} \sum_{y=0}^{h-1} \sum_{x=0}^{w-1} (f_k(x, y) - f'_k(x, y))^2. \quad (6.10)$$

The concealment distortion is computed as the average frame distortion between the temporal segment F and its approximation set F' , i.e., the one concealed from the HP frame set P . This concealment distortion is given by:

$$D(P) = 1/n \sum_{k=0}^{n-1} d(f_k, f'_k). \quad (6.11)$$

The HP ratio $R(P)$ is the ratio between the number of frames m belonging to the HP frame set P and the total number of frames n of temporal segment F . Therefore,

$$R(P) = m/n. \quad (6.12)$$

6.4.3.2 Problem Formulation

Using the definition presented in the previous subsection, we formulate the following binary priority frame classification algorithm as a HP ratio-distortion concealment optimization problem where the objective function is to find a subset of frames in each temporal segment that provides its best representation within a given maximum HP ratio R_{max} , i.e., without using more than $m = R_{max} * n$ HP frames. Given the HP ratio constraint R_{max} , the optimal HP frame set P^* is the one that minimizes the concealment

distortion given by,

$$P^* = \arg \min_P D(P), \quad (6.13)$$

$$s.t. \quad R(P) \leq R_{max},$$

where $R(P)$ and $D(P)$ are defined by (6.12) and (6.11) respectively. For example, given a temporal segment F of $n = 10$ frames and a priority ratio $R(P) = 0.2$, the proposed algorithm classifies at most 2 frames as high priority, i.e. $m = 2$. Thus, the constraint R_{max} is a user-defined parameter to define the number of HP frames in each temporal segment.

6.4.3.3 Dynamic Programming Solution

In general, there are $\binom{n-1}{m-1} = \frac{(n-1)!}{(m-1)!(n-m)!}$ feasible solutions to above problem, assuming the first frame of temporal segment always belongs to set P since it is an Intra frame and consequently the most important one. When n and m are large, an exhaustive search solution would be prohibitive. Alternatively to this solution, the dynamic programming method was used to solve the priority ratio-distortion concealment optimization problem by breaking it down into simpler subproblems in a recursive manner [Bertsekas 1987] [Ferreira 2011]. The concealment distortion stage D_t^k is defined as the minimum total concealment distortion incurred by HP frame set that has t frames and ends with the frame $f_k(l_{t-1} = k)$. Therefore,

$$D_t^k = \min_{l_1, l_2, \dots, l_{t-2}} \sum_{j=0}^{n-1} d(f_j, f_i), \quad \text{with} \quad (6.14)$$

$$i = \max(l) : s.t. \quad l \in \{0, l_1, l_2, \dots, l_{t-2}\}, \quad i \leq j.$$

Note, $l_0 = 0$ and $l_{t-1} = k$ are removed from the optimization process. Since $0 < l_1 < l_2 < \dots < l_{t-2} < k$, and $i \leq j$. After some math manipulation, the current concealment distortion stage D_t^k can be broken into two parts. (6.15). The first part is the previous concealment distortion stage $D_{t-1}^{l_{t-2}}$ already computed, and it represents the minimum total concealment distortion produced by HP frame set with $t-1$ frames and ending with frame corresponding to index l_{t-2} . In addition, the second part, $e^{l_{t-2}, k}$ represents the previous concealment distortion stage reduction, if frame k is selected into the HP frame set of

$t - 1$ frames ending in frame with the l_{t-2} index. Therefore, we have

$$D_t^k = \min_{l_{t-2}} \{D_{t-1}^{l_{t-2}} - e^{l_{t-2},k}\}, \quad (6.15)$$

where the distortion reduction was defined by,

$$e^{l_{t-2},k} = \sum_{j=k}^{n-1} [d(f_j, f_{l_{t-2}}) - d(f_j, f_k)]. \quad (6.16)$$

Since it was assumed that the first frame is always in set P , the initial concealment distortion stage D_1^0 is given as

$$D_1^0 = \frac{1}{n} \sum_{j=1}^{n-1} d(f_0, f_j). \quad (6.17)$$

At this point, we can compute the current concealment distortion stage D_t^k for any HP frame set of t frames ending with the frame k by the recursion in (6.15) with the initial concealment distortion stage given by (6.17).

Finally, the HP frame set P is given by the calculation of frame index l_0, l_1, \dots, l_{m-1} found by backtracking the computation already performed of D_t^k (6.15). Therefore,

$$\begin{aligned} l_0 &= 0 \\ l_{m-1} &= \arg \min_k \{D_m^k\} \end{aligned} \quad (6.18)$$

where l_0 and l_{m-1} represent the first and last frame index selected into the high priority frame set $P = \{f_{l_0}, f_{l_1}, \dots, f_{l_{m-1}}\}$, respectively. In this work, each temporal segment F was defined as being coincident with the group of frames (GOP) used by the MDC encoder and each frame corresponds to one NALU.

6.4.4 Simulation Results

The performance of the MDC priority streaming scheme described in the previous sections, was evaluated assuming path diversity communication with several PLR. Two main

scenarios were compared: MDC with and without binary frame classification and MDC without binary frame classification.

A random packet loss generator with uniform distribution is used to drop packets according to the required packet loss rate. Each description is streamed over two independent paths with the same PLR. The packets classified as HP are not lost. In order to obtain statistically sound results, the transmission of each sequence (*mother & daughter*, *foreman* and *bus*, CIF resolution, frame rate= 30Hz) was simulated 50 times under the same network conditions, i.e., average PLR of 3%, 5%, 7%, 10% and 15%. The GOP structure is IPPPP... with GOP=15 frames, i.e., constant temporal segment $F = 15$ frames. A fixed QP set was used to encode each sequence, i.e., *mother & daughter* was encoded with $QP_I = 23$ and $QP_P = 24$, *foreman* was encoded with $QP_I = 30$ and $QP_P = 31$ and *bus* was encoded with $QP_I = 37$ and $QP_P = 38$. The resulting MDC rate for both descriptions added together is about 1 Mbps, including the inherent redundancy.

The frame-copy error concealment was used whenever packets carrying the same NALU in both descriptions are simultaneously lost. In order to evaluate the impact of priority frame classification, the HP constraining ratio was set into two fixed levels, $R_{max} = 0.13$ and $R_{max} = 0.26$, which corresponds to 2 and 4 HP frames per GOP respectively. In all experiments, the I frames were always set as HP frames. The proposed method is compared with i) MDC with Intra Frame Binary Classification: only I frames are classified as HP and ii) MDC Without Binary Frame Classification methods: i.e., corresponding to use MDC without any frame classification method.

Figures 6.18, 6.19 and 6.20, show the video quality, i.e., average peak signal-to-noise ratio (PSNR) and 95% confidence intervals, obtained after simulation for the sequences *mother&daughter*, *foreman* and *bus* respectively, under random packet loss. The simulation results show that the average PSNR improves for all PLR when the proposed optimal binary frame classification is used. For instance, considering the *mother&daughter* sequence and comparing the proposed method, using $R_{max} = 0.26$, with MDC without HP frames, the average PSNR gains are 1.03dB for PLR=3%, 1.16dB for PLR=5%, 1.30dB for PLR=7%, 1.49dB for PLR=10% and 2.3dB for PLR=15%. Comparing MDC where only I frames are HP with the no priority scheme, the PSNR gains are 0.85dB for PLR=3%, 0.82dB for PLR=5%, 0.75dB PLR=7%, 0.89dB for PLR=10% and 1.55dB for PLR=15%. On the other hand, the confidence intervals show that average PSNR have smoother variations in the case where frame classification and priority is used.

These results show that classifying intra frames as HP significantly improves the overall video quality, as expected. Also, comparing the simulation results of the proposed method with $R_{max} = 0.26$ and MDC with only I frames as HP, the PSNR gains are 0.18dB for PLR=3%, 0.35dB for PLR=5%, 0.55dB for PLR=7%, 0.60dB for PLR=10% and 0.74dB for PLR=15%, which means that optimal choice of few priority frames over a GOP, improves the average video quality received at the user end.

Note that in any case, packet loss always leads to distortion propagation along the affected GOP in the video decoder because HP frames are not intra-coded, i.e., they are P frames. This means that, even with a very small amount of priority packets the overall quality of decoded video is significantly improved. Identical results are obtained for *foreman* and *bus* sequences in Figures 6.19 and 6.20. Moreover MDC schemes usually have higher error resilience capabilities in comparison with other error resilience methods, such as adaptive intra refresh [Radulovic 2010]. Therefore, since the proposed scheme improves the performance of classic MDC, one might infer that better performance is also obtained in comparison with those methods. Overall, these results show that the optimal priority MDC streaming based on binary frame classification can be used to improve the quality of decoded sequences in MDC video streaming over multipath networks. In particular, the proposed scheme exhibits consistently better performance over a non-priority MDC scheme.

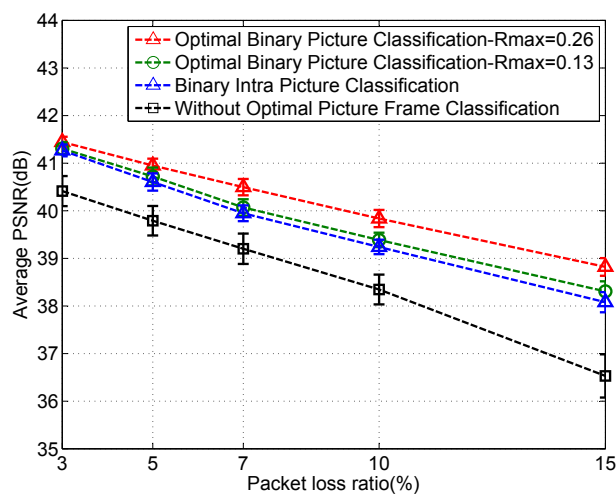
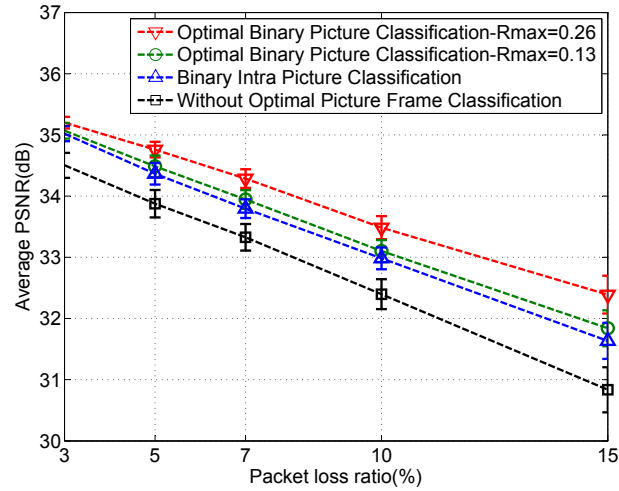
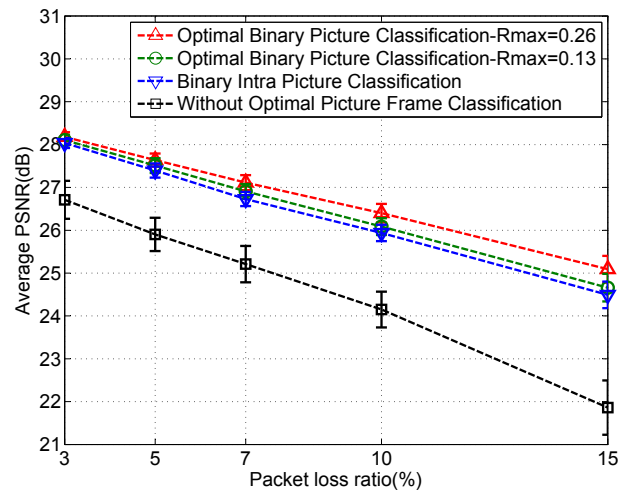


Figure 6.18: Simulation results for *Mother & daughter* sequence.

Figure 6.19: Simulation results for *Foreman* sequence.Figure 6.20: Simulation results for *Bus* sequence.

6.5 Conclusion

In this chapter an extension of the ρ model rate control for unbalance MDSQ video coding was proposed and evaluated. The method was shown to be effective in unbalanced rate control of MDSQ descriptions by using a set of new index assignment tables. The experimental results show that overall RD performance is improved using unbalanced MDC in comparison with balanced MDC. Therefore, the proposed scheme finds application in video delivery over path diversity networks with asymmetric rate constraints.

Also, this chapter demonstrates that improved quality can be achieved in video streaming over path diversity networks by using MDC combined with optimal binary frame clas-

sification for setting packet priority. The proposed scheme was shown to be effective in increasing the QoE by classifying NALUs into two levels of priority, according to their relevance for the concealment distortion. The experimental results show that the decoded video quality can be significantly improved under different channel loss conditions. Therefore, the proposed scheme finds application in video delivery over path diversity networks where different levels of priorities can be implemented.

Conclusions and future work

Contents

7.1	Conclusions	163
7.2	Future Work	165

7.1 Conclusions

Throughout this thesis new MDC schemes for advanced video coding and MDC video adaptation have been proposed.

In Chapter 2, the theoretical definitions and optimal rate-distortion bounds were discussed and also the most relevant MDC techniques proposed in the literature were studied, such as Multiple Description with Correlating Transforms (MDCT), Multiple Description Subband Coding, based on Polyphase Decomposition and Selective Quantisation, Multiple Description Scalar Quantisation, and MDC based on the Error Correction Codes.

In Chapter 3, the main MDC video architectures were described for the different video coding schemes. The state of the art of MDC for video was analysed and discussed including applications using scalable coding, multi-view, stereo video coding, Wyner-Ziv coding and also for high-dimensional MDC. Different path diversity topologies and the MDC networking applications were also discussed.

Chapter 4 describes the research work on a multiple description coding scheme for Advanced Video Coding based on a MDSQ mixed open-loop/multi-loop structure. An experimental study about drift distortion accumulation in both intra and motion compensated predicted slices was carried out. A method was proposed to use side information only

in anchor slices maintaining a reduced redundancy cost. It was demonstrated that such method is capable of preventing drift propagation for single description decoding. A new MDSQ scheme for unbalanced MDC was proposed based on new types of index assignment tables to dynamically change the rate of each description and also controlling drift distortion at the decoder. The simulation results have shown that proposed method has the advantage of reducing the redundancy of MDSQ. The method is also able to add side information to unbalanced descriptions. Experimental results show that using a small amount of side information, results in a good performance comparing the side-distortion with a closed-loop balanced scheme. It was found that this is due to the increasing of coding efficiency in both central and side information in unbalanced MDSQ compared with balanced MDSQ.

In Chapter 5 a new MDVS scheme was developed with the purpose of adapting an SDC compressed stream into a MDC one with two descriptions to act in any network edge node. The main novel aspects of the proposed MDVS scheme comprise a two-loop MDVS architecture with drift control in both intra and inter predictive coded slices. An equivalent single-loop architecture with reduced complexity was also developed. The proposed architecture generates side-information from the compressed streams without full decoding, thus reusing most of the original coding parameters and keeping the coding structure, such as slice maps. A evaluation study is presented with discussion of the proposed MDVS and comparison with other results obtained from the literature in MDC context. From this research, it was concluded that the overall performance is similar to MDC using uncompressed video. Therefore, this is a valid approach to be used in network edge nodes where MDVS is possible to implement.

In Chapter 6 a new rate control method for MDSQ video coding was developed, by generating an asymmetric target bitrate in each description extending the traditional ρ model rate control method for MDC applications. The linear relationship between the rate of each description and the corresponding percentage of zeros of transform coefficients for MDSQ indices was shown to be maintained when moving from the single description domain (SDC) into the MDC domain. Based on this evidence, a new rate control method was proposed. This new method allows to dynamically balance each description rate by choosing adequate MDC coding parameters. The experimental results show that the proposed rate control method has a good accuracy for a given target bitrate and balancing ratio among descriptions. Furthermore, the proposed method efficiently controls the buffer encoder taking into account the overall output rate to avoid buffer overflow

at each encoding time, and consequently buffer underflow at decoder buffers. Overall, the proposed method presents an effective evolution in rate control methods for MDC schemes, in particular using MDSQ.

Also in Chapter 6 has been developed a robust video streaming scheme for priority networks with path diversity, based on a combined approach MDC with optimal frame classification into two priorities. An optimization algorithm finds high priority frames based on minimization of the packet loss concealment distortion. It was shown that the quality improvement at the decoder can be achieved in video streaming over path diversity networks by using MDC combined with optimal binary picture classification for setting packet priority. The experimental results show that the decoded video quality can be significantly improved under different channel loss conditions. Therefore, the proposed scheme finds application in video delivery over path diversity networks where different levels of priorities can be implemented.

7.2 Future Work

The research work carried out in the scope of this thesis leaves further research in several topics addressed in the pervious chapters. Some of such open topics for future work are the following.

MDC adaptation using distributed computing The proposed MDVS method use the incoming SDC stream to generate additional redundant information in order to avoid drift distortion at the decoder. However, further investigation is needed in new re-quantization schemes of the original SDC bitstream associated to MDC in order to adapt the overall output rate. Distributed computing, namely cloud computing, is being largely deployed within an increasing number of access network technologies. Adapting video contents at some network edge or mobile device can be done using distributed computing, where some tasks can be done in another device with different processing capabilities. The new concept of MDC video adaptation with drift control introduced in this thesis can be evaluated in this new context, possibly exploiting more complex models in terms of rate-distortion optimization and coding dependencies, which implies full-decoding operations. Also several prediction representations provided by re-encoding operations improve the flexibility of MDC in terms of drift control and decoding descriptions. Furthermore,

new joint low delay source-channel coding approaches can be added in order to improve the error resilience of the video sources.

MDC redundancy rate-distortion models Further investigation is needed to optionally find the overall rate and redundancy allocation based on optimized redundancy rate-distortion models combining each rate component and at the same time, adapting the coding parameters to the channel conditions, namely the available bandwidth and the packet-loss rate. Also, this thesis have shows that the main idea of prioritizing each packet of each description to be used in schedule systems is an effective solution to be used in MDC video context. Considering particular application scenarios, MDC rate-distortion models can be improved using prioritization schemes where packet dependencies are taken into account.

MDC for High Efficiency Video Coding The test model of next generation coding standard called high efficiency video coding (HEVC) has adopted new coding structures, such as the usage of larger coding unit and transform blocks dimensions. The coding unit in the range of 16×16 to 64×64 provides great bit rate saving while the coding complexity increases dramatically due to nested data structures. This new coding units will give new spatial and temporal prediction dependencies which means that new MDC schemes must be developed in order to improve the coding efficiency and error resilience capabilities. For instance, new schemes based on redundant information with different qualities adapted for HEVC structures can be investigated.

MDC for 3D video Multi-view video coding is a new research topic where the main video coding techniques are improved taking into account the additional dimension given by the interview prediction. Currently, the main research effort has been done in coding efficiency and also in 3D rendering not only in video+deph representation, but also in view synthesis. Multi-view video transmission needs a more research effort in order to improve error resiliency of this new video format characterized by its high complexity prediction dependencies. MDC can be an important approach in order provide error resilience in 3D video transmission. This can be done by simplifying the prediction dependencies, introducing some loss of coding efficiency, but also to exploit disjoint different interview predictions, introducing some redundancy but adding several decoding combinations. In particular, MDC applied in motion and disparity information is an open issue to further

investigation. Also, several quality representations can be combined in a rate-distortion sense in order to improve quality by exploiting the specific subjective factors of 3D video that poses new problems in perceived quality.

Optimal Solutions in High Dimensional Spaces Multiple descriptions transmitted over multipath lossy networks gives rise to a multidimensional problem for which no complete solution is yet known. The problem that can be posed is to find the optimal number of descriptions, as well the bit rates that should be allocated to each one under known network conditions (e.g., packet loss statistics, bandwidth), such that the total average distortion per user is minimized. New optimization methods are needed to solve this multidimensional problem that can be find application in new schemes of streaming where video contents can be divided in small chunks that can be distributed in several network nodes (e.g. peer-to-peer networks).

Published Papers

A.1 Journal Papers

- J1.** Pedro D. F. Correia, Pedro A. Amado Assunção, Vitor Silva, "Multiple Description of Coded Video for Path Diversity Streaming Adaptation", *IEEE Transactions On Multimedia*, vol. 14, no. 2-3, pp. 923-935, June 2012.

A.2 Conference Papers

- C1.** Pedro D. F. Correia, Pedro A. Amado Assunção, Vitor Silva, "Rate Control for Unbalanced Multiple Description Video Coding", *IEEE International Conference on Image Processing-ICIP*, Melbourne, Australia, September 2013 (to be submitted).
- C2.** Pedro D. F. Correia, Pedro A. Amado Assunção, Vitor Silva, "Unbalanced Multiple Description Video Coding", *Conference on Telecommunications - ConfTele*, Castelo Branco, Portugal, May 2013 (to be submitted).
- C3.** Pedro D. F. Correia, L.F. Ferreira, Pedro A. Amado Assunção, Luís Cruz, Vitor Silva, "Optimal priority MDC video streaming for networks with path diversity", *Proc. International Conference on Telecommunications and Multimedia (TEMU)*, Heraklion, Greece, pp.54-59, July 2012.
- C4.** Pedro D. F. Correia, Pedro A. Amado Assunção, Vitor Silva, "Enhanced H.264/AVC Video Streaming using Network-adaptive Multiple Description Coding", *Proc. IEEE Region 8 EUROCON 2011 International Conference on Computer as a Tool*, Lisbon, Portugal, April 2011.
- C5.** Pedro D. F. Correia, Pedro A. Amado Assunção, Vitor Silva, "Multiple Description Video Transcoding with Temporal Drift Control", *Proc. Picture Coding Symposium-PCS*, Nagoya, Japan, Vol. 1, pp. 558-561, December 2010.

- C6.** Pedro D. F. Correia, Pedro A. Amado Assunção, Vitor Silva, "Drift-free Multiple Description Intra Video Coding", *Proc. IEEE International Conference on Image Processing-ICIP*, Cairo, Egypt, Vol. 1, pp. 625-628, November 2009.
- C7.** Pedro D. F. Correia, Pedro A. Amado Assunção, Vitor Silva, "Multiple Description Coding Scheme for H.264/AVC Intra Slices", *Proc. Conference on Telecommunications - ConfTele*, Santa Maria da Feira, Portugal, May 2009.

APPENDIX B

Test video sequences



Figure B.1: Sequence *Bus* (CIF:352 × 288)



Figure B.2: Sequence *City* (CIF:352 × 288)

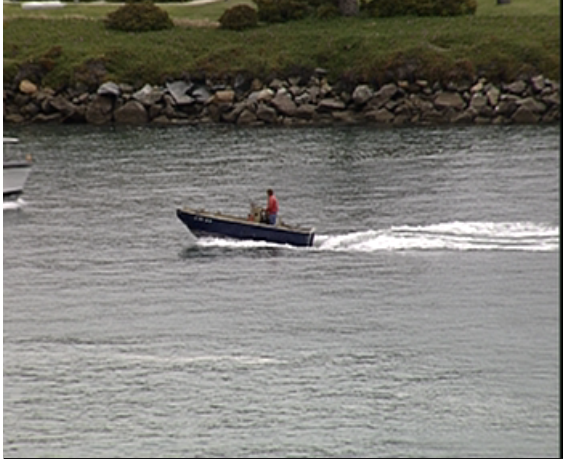


Figure B.3: Sequence *Coastguard* (CIF:352 × 288)



Figure B.4: Sequence *Foreman* (CIF:352 × 288)



Figure B.5: Sequence *News* (CIF:352 × 288)



Figure B.6: Sequence *Mother and daughter* (CIF:352 × 288)

References

- [15444-1 2002] Rec. ITU-T T.802—ISO/IEC 15444-1. *Recommendation ITU-T T.802: JPEG 2000 image coding system*, 2002.
- [A. Serdar Tan 2009] Godze Akar A. Serdar Tan Anil Aksay and Erdal Arıkan. *Rate Distortion Optimization for Stereoscopic Video Streaming with Unequal Error Protection*. Eurasip Journal On Advances in Signal Processing, 2009.
- [Abanoz 2009] T. Berkin Abanoz and A. Murat Tekalp. *SVC-Based scalable multiple description video coding and optimization of encoding configuration*. Signal Processing: Image Communication, no. 24, pages 691–701, 2009.
- [Ahuja 2008] S. Ahuja and M. Krunz. *Algorithms for Server Placement in Multiple-Description-Based Media Streaming*. IEEE Transactions on Multimedia, vol. 10, no. 7, pages 1382–1392, Nov. 2008.
- [Akyol 2007] E. Akyol, A. M. Tekalp and M. R. Civanlar. *A Flexible Multiple Description Coding Framework for Adaptive Peer-to-Peer Video Streaming*. IEEE Journal of Selected Topics in Signal Processing, vol. 1, no. 2, pages 231–245, 2007.
- [Apostolopoulos 2000] J.G. Apostolopoulos. *Error-resilient video compression through the use of multiple states*. In Image Processing, 2000. Proceedings. 2000 International Conference on, volume 3, pages 352–355 vol.3, 2000.
- [Apostolopoulos 2001a] J. G. Apostolopoulos. *Reliable Video Communication over Lossy Packet Networks using Multiple State Encoding and Path Diversity*. Proceedings of Visual Communications and Image Processing, Jan. 2001.
- [Apostolopoulos 2001b] J.G. Apostolopoulos and S.J. Wee. *Unbalanced multiple description video communication using path diversity*. In Proceedings. 2001 International Conference on Image Processing, volume 1, pages 966–969 vol.1, 2001.
- [Apostolopoulos 2002] J. Apostolopoulos, T. Wong, W. tian Tan and S. Wee. *On Multiple Description Streaming with Content Delivery Networks*. In IEEE INFOCOM, 2002.
- [Assunção 2000] Pedro A. Amado Assunção and Mohammed Ghanbari. *Buffer analysis and control in CBR video transcoding*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, no. 1, pages 83–92, 2000.

- [Bajic 2003] I. V. Bajic and J. W. Woods. *Domain based multiple description coding of images and video*. IEEE Transactions on Image Processing, vol. 12, no. 10, pages 1211–1225, Oct. 2003.
- [Begen 2005] Ali C. Begen, Yucel Altunbasak, Ozlem Ergun and Mostafa H. Ammar. *Multi-path selection for multiple description video streaming over overlay networks*. Signal Processing: Image Communication, vol. 20, no. 1, pages 39 – 60, 2005.
- [Bertsekas 1987] Dimitri P. Bertsekas. *Dynamic programming: deterministic and stochastic models*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1987.
- [Bin 2006] Li Bin, Huang Feng, Sun Lifeng and Yang Shiqiang. *Optimized Rate Allocation for Unbalanced Multiple Description Video Coding Over Unreliable Packet Network*. In 2006 IEEE International Conference on Multimedia and Expo(ICME), pages 105 –108, july 2006.
- [C. Tian 2004] S. S. Hemami C. Tian. *Universal Multiple Description Scalar Quantization: Analysis and Design*. IEEE Transactions On Information Theory, vol. 50, no. 9, pages 2089–2102, Sept. 2004.
- [Campana 2008] O. Campana, R. Contiero and G. A. Mian. *An H.264/AVC Video Coder Based on a Multiple Description Scalar Quantizer*. IEEE Transactions On Circuits and Systems For Video Technology, vol. 18, no. 2, pages 268–272, Feb. 2008.
- [Castro 2003] Miguel Castro, Peter Druschel, Anne-Marie Kermarrec, Animesh Nandi, Antony Rowstron and Atul Singh. *SplitStream: high-bandwidth multicast in cooperative environments*. In Proceedings of the nineteenth ACM symposium on Operating systems principles, SOSP '03, pages 298–313, New York, NY, USA, 2003. ACM.
- [Chakareski 2005] J. Chakareski, S. Han and B. Girod. *Layered coding vs. multiple descriptions for video streaming over multiple paths*. Multimedia Systems, vol. 10, no. 4, pages 275–285, 2005.
- [Chakareski 2006] J. Chakareski and P. Frossard. *Rate-distortion optimized distributed packet scheduling of multiple video streams over shared communication resources*. Multimedia, IEEE Transactions on, vol. 8, no. 2, pages 207 – 218, april 2006.

- [Chakareski 2008] J. Chakareski and P. Frossard. *Distributed Collaboration for Enhanced Sender-Driven Video Streaming*. Multimedia, IEEE Transactions on, vol. 10, no. 5, pages 858–870, aug. 2008.
- [Chen 2007a] Chih-Ming Chen, Chien-Min Chen, Chia-Wen Lin and Yung-Chang Chen. *Error-Resilient video streaming over wireless networks using combined scalable coding and multiple-description coding*. Signal Processing: Image Communication, vol. 22, pages 403–420, 2007.
- [Chen 2007b] Chih-Ming Chen, Chia-Wen Lin, Hsiao-Cheng Wei and Yung-Chang Chen. *Robust video streaming over wireless LANs using multiple description transcoding and prioritized retransmission*. Journal of Visual Communication & Image Representation, vol. 18, pages 191–206, 2007.
- [Chiang 1997] Tihao Chiang and Ya-Qin Zhang. *A new rate control scheme using quadratic rate distortion model*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 7, no. 1, pages 246–250, feb 1997.
- [Choi 1999] Seung-Jong Choi and J.W. Woods. *Motion-compensated 3-D subband coding of video*. IEEE Transactions on Image Processing, vol. 8, no. 2, pages 155–167, feb 1999.
- [Chou 2006] P.A. Chou and Zhouong Miao. *Rate-distortion optimized streaming of packetized media*. Multimedia, IEEE Transactions on, vol. 8, no. 2, pages 390–404, april 2006.
- [Comas 2003] David Comas, Raghavendra Singh, Antonio Ortega and Ferran Marques. *Unbalanced Multiple Description Video Coding with Rate-Distortion Optimization*. In EURASIP Journal on Applied Signal Processing, pages 81–90, 2003.
- [Cover 2006] Thomas M. Cover and Joy A. Thomas. Elements of information theory 2nd edition (wiley series in telecommunications and signal processing). Wiley-Interscience, July 2006.
- [Crave 2008] Olivier Crave, Christine Guillemot, Béatrice Pesquet-Popescu and Christophe Tillier. *Distributed temporal multiple description coding for robust video transmission*. EURASIP J. Wirel. Commun. Netw., vol. 2008, January 2008.

- [Crave 2010] O. Crave, B. Pesquet-Popescu and C. Guillemot. *Robust Video Coding Based on Multiple Description Scalar Quantization With Side Information*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, no. 6, pages 769–779, June 2010.
- [D. Wang 2005] N. Canagarajah D. Wang and D. Bull. *Slice Group based multiple description video coding with three motion compensation loops*. In IEEE International Symposium on Circuits and Systems, pages 960–963, May 2005.
- [Dragotti 2002] P.L. Dragotti, S.D. Servetto and M. Vetterli. *Optimal filter banks for multiple description coding: analysis and synthesis*. Information Theory, IEEE Transactions on, vol. 48, no. 7, pages 2036–2052, jul 2002.
- [Erhan Ekmekcioglu 2010] Maheshi Dissanayake Erhan Ekmekcioglu Banu Günel, Stewart T. Worrall and Ahmet M. Kondoz. *A Scalable Multi-view Audiovisual Entertainment Framework With Content-Aware Distribution*. In IEEE International Conference On Image Processing, ICIP, 2010, pages 2401–2404, Sept. 2010.
- [Essaili 2007] Ali El Essaili, Shaoib Khan, Wolfgang Kellerer and Eckehard Steinbach. *Multiple Description Video Transcoding*. In IEEE International Conference on Image Processing (ICIP), volume 6, pages 77–78, 2007.
- [Fan 2011] Yuhua Fan, Jia Wang and Jun Sun. *Distributed Multiple Description Video Coding on Packet Loss Channels*. IEEE Transactions on Image Processing, vol. 20, no. 6, pages 1768–1773, June 2011.
- [Favalli 2011] Lorenzo Favalli and Marco Folli. *ILPS: a scalable multiple description coding scheme for H.264*. Multimedia Tools and Applications, vol. 54, pages 609–634, 2011. 10.1007/s11042-010-0577-0.
- [Ferreira 2011] Lino Ferreira, Luís Cruz and Pedro Assuncao. *Efficient scalable coding of video summaries using dynamic GOP structures*. In IEEE International Conference on Computer as a Tool, EUROCON 2011, April 2011.
- [Franchi 2005] N. Franchi, M. Fumagalli, R. Lancini and S. Tubaro. *Multiple Description Video Coding for Scalable and Robust Transmission Over IP*. IEEE Transactions On Circuits and Systems For Video Technology, vol. 14, no. 3, pages 321–334, March 2005.

- [Frossard 2008] P. Frossard, J.C. de Martin and M. Reha Civanlar. *Media Streaming With Network Diversity*. Proceedings of the IEEE, vol. 96, no. 1, pages 39–53, Jan. 2008.
- [Gamal 1982] A.E. Gamal and T. Cover. *Achievable rates for multiple descriptions*. Information Theory, IEEE Transactions on, vol. 28, no. 6, pages 851–857, nov 1982.
- [Gan 2006] Tong Gan, Lu Gan and Kai-Kuang Ma. *Reducing video-quality fluctuations for streaming scalable video using unequal error protection, retransmission, and interleaving*. IEEE Transactions on Image Processing, vol. 15, no. 4, pages 819–832, April 2006.
- [Ghareeb 2010] M. Ghareeb and C. Viho. *A Multiple Description Coding Approach for Overlay Multipath Video Streaming Based on QoE Evaluations*. In Multimedia Information Networking and Security (MINES), 2010 International Conference on, pages 39–43, Nov. 2010.
- [Goyal 2001a] V. Goyal. *Multiple Description Coding: Compression Meets the Network*. IEEE Signal Processing Magazine, vol. 18, no. 5, pages 74–93, Sept. 2001.
- [Goyal 2001b] V. Goyal and J. Kovacevic. *Generalized Multiple Description Coding With Correlated Transforms*. IEEE Transactions on Information Theory, vol. 47, no. 6, pages 2199–2224, Sept. 2001.
- [H. Abdul Karim 2009] S. Worrall H. Abdul Karim A. Sali, Abdulk H. Sadka and A.M. Kondo. *Multiple Description Video Coding for Stereoscopic 3D*. Consumer Electronics, IEEE Transactions on, vol. 55, no. 4, pages 2048–2056, November 2009.
- [He 2001] Zhihai He and S.K. Mitra. *A unified rate-distortion analysis framework for transform coding*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 11, no. 12, pages 1221–1236, dec 2001.
- [Hsi-Tzeng Chan 2005] Chih-Ming Fu Hsi-Tzeng Chan and Chung lin Huang. *A New Error Resilient Video Coding Using Matching Pursuit and Multiple Description Coding*. IEEE Transactions On Circuits and Systems For Video Technology, vol. 15, no. 8, pages 1047–1052, Aug. 2005.

- [Hsiao 2010] C.-W. Hsiao and W.-J. Tsai. *Hybrid Multiple Description Coding Based on H.264*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, no. 1, pages 76–87, Jan. 2010.
- [ISO/IEC13818-2 2000] ISO/IEC13818-2. *Generic coding of moving pictures and associated audio information: Video*, 2000.
- [ISO/IEC14496-10 2012] ISO/IEC14496-10. *Coding of audio-visual objects – Part 10: Advanced Video Coding*, 2012.
- [Jurca 2007] Dan Jurca, Jakob Chakareski, Jean-Paul Wagner and Pascal Frossard. *Enabling Adaptive Video Streaming in P2P Systems*. IEEE Communications Magazine, vol. 45, no. 6, pages 108–114, 2007.
- [Kamnoonwatana 2011] Nawat Kamnoonwatana, Dimitris Agraotis and Nishan Canagarajah. *Rate controlled redundancy-adaptive multiple description video coding*. Signal Processing Image Communication, vol. 26, no. 5, pages 205–219, 2011.
- [Khan 2004] S. Khan, R. Schollmeier and E. Steinbach. *A performance comparison of multiple description video streaming in peer-to-peer and content delivery networks*. In Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on, volume 1, pages 503–506 Vol.1, june 2004.
- [Kim 1997] Beong-Jo Kim and W.A. Pearlman. *An embedded wavelet video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)*. In Data Compression Conference, 1997. DCC '97. Proceedings, pages 251–260, mar 1997.
- [Kim 2000] Beong-Jo Kim, Zixiang Xiong and W.A. Pearlman. *Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT)*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 10, no. 8, pages 1374–1387, dec 2000.
- [Kim 2001] C.-Su Kim and S.-Uk Lee. *Multiple Description Coding of Motion Fields for Robust Video Transmission*. IEEE Transactions On Circuits and Systems For Video Technology, vol. 11, no. 9, pages 999–1010, Sept. 2001.
- [Kim 2003] Il Koo Kim, Nam Ik Cho and Jeho Nam. *Error Resilient Video Transcoding Dased on the Optimal Multiple Description of DCT Coefficients*. In International Workshop on Advanced Image Technology, 2003.

- [Kim 2005] Joohee Kim, R.M. Mersereau and Y. Altunbasak. *Distributed video streaming using multiple description coding and unequal error protection*. IEEE Transactions on Image Processing, vol. 14, no. 7, pages 849–861, July 2005.
- [Kim 2006] Il Koo Kim and Nam Ik Cho. *Video Transcoding for Packet Loss Resilience Based on the Multiple Descriptions*. In International Conference on Multimedia and Expo(ICME), pages 109–112, 2006.
- [Kondi 2005] L. P. Kondi. *A Rate-Distortion Optimal Hybrid Scalable/Multiple-Description Video Codec*. IEEE Transactions On Circuits and Systems For Video Technology, vol. 15, no. 7, pages 921–927, July 2005.
- [Lamy-Bergot 2010] Catherine Lamy-Bergot and Benjamin Gadat. *Embedding Protection Inside H.264/AVC and SVC Streams*. EURASIP Journal on Wireless Communications and Networking, vol. 2010, 2010.
- [Lefol 2006] D. Lefol, D. Bull and N. Canagarajah. *Performance Evaluation of Transcoding Algorithms for H.264*. IEEE Transactions on Consumer Electronics, vol. 52, no. 1, pages 215–221, 2006.
- [Li 1998] Xin Li, Bo Tao and M.T. Orchard. *On implementing transforms from integers to integers*. In Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on, pages 881–885 vol.3, Oct. 1998.
- [Li 2006] Z.G. Li, W. Gao, F. Pan, S.W. Ma, K.P. Lim, G.N. Feng, X. Lin, S. Rahardja, H.Q. Lu and Y. Lu. *Adaptive rate control for H.264*. Journal of Visual Communication and Image Representation, vol. 17, no. 2, pages 376–406, 2006.
- [Li 2009] Zhicheng Li, J. Chakareski, Xiaodun Niu, Yongjun Zhang and Wanyi Gu. *Modeling of distortion caused by Markov-model burst packet losses in video transmission*. In IEEE International Workshop on Multimedia Signal Processing (MMSP), pages 1–6, Oct. 2009.
- [Liao 2011] Y. Liao and J. D. Gibson. *Routing-aware multiple description video coding over mobile ad-hoc networks*. IEEE Transactions On Multimedia, vol. 13, no. 1, pages 132–142, 2011.
- [Lin 2011] Chunyu Lin, T. Tillo, Yao Zhao and Byeungwoo Jeon. *Multiple Description Coding for H.264/AVC With Redundancy Allocation at Macro Block Level*. IEEE

- Transactions on Circuits and Systems for Video Technology, vol. 21, no. 5, pages 589–600, may 2011.
- [Liu 2005] Yilong Liu and S. Oraintara. *Drift-free multiple description video coding with redundancy rate-distortion optimization*. In Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on, pages 4034–4037 Vol. 4, may 2005.
- [Liu 2008a] Jiangchuan Liu, S.G. Rao, Bo Li and Hui Zhang. *Opportunities and Challenges of Peer-to-Peer Internet Video Broadcast*. Proceedings of the IEEE, vol. 96, no. 1, pages 11–24, jan. 2008.
- [Liu 2008b] Yong Liu, Yang Guo and Chao Liang. *A survey on peer-to-peer video streaming systems*. Peer-to-Peer Networking and Applications, vol. 1, no. 1, pages 18–28, March 2008.
- [Lu 2007] Meng-Ting Lu, Jui-Chieh Wu, Kuan-Jen Peng, P. Huang, J.J. Yao and H.H. Chen. *Design and Evaluation of a P2P IPTV System for Heterogeneous Networks*. Multimedia, IEEE Transactions on, vol. 9, no. 8, pages 1568–1579, dec. 2007.
- [M. Biswas 2009] J. Arnold M. Biswas M. Frater and M. Pickering. *Improved Resilience for Video Over Packet Loss Networks with MDC and Optimized Packetization*. IEEE Transactions On Circuits and Systems For Video Technology, vol. 19, no. 10, pages 1556–1560, Oct. 2009.
- [M. Pereira 2003] M. Antonioni M. Pereira and M. Barlaud. *Multiple Description Image and Video Coding for Wireless Channels*. Elsevier Signal Processing: Image Communications, vol. 18, pages 925–945, 2003.
- [Ma 2008] Mengyao Ma, O.C. Au, Liwei Guo, S.-H.G. Chan and P.H.W. Wong. *Error Concealment for Frame Losses in MDC*. IEEE Transactions on Multimedia, vol. 10, no. 8, pages 1638–1647, dec. 2008.
- [Magharei 2007] N. Magharei, R. Rejaie and Yang Guo. *Mesh or Multiple-Tree: A Comparative Study of Live P2P Streaming Approaches*. In INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE, pages 1424–1432, May 2007.
- [Mao 2005a] S. Mao, S. Lin, Y. Wang, S.S. Panwar and Y. Li. *Multipath video transport over wireless ad hoc networks*. IEEE Wireless Communications Magazine, vol. 12, no. 4, pages 42–49, Aug. 2005.

- [Mao 2005b] Shiwen Mao, Y.T. Hou, Xiaolin Cheng, H.D. Sherali and S.F. Midkiff. *Multiple description video in wireless ad hoc networks*. In INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE, volume 1, pages 740–750, March 2005.
- [Mao 2006] Shiwen Mao, Y.T. Hou, Xiaolin Cheng, H.D. Sherali, S.F. Midkiff and Ya-Qin Zhang. *On Routing for Multiple Description Video Over Wireless Ad Hoc Networks*. IEEE Transactions on Multimedia, vol. 8, no. 5, pages 1063–1074, Oct. 2006.
- [Matty 2005] K. R. Matty and L. P. Kondi. *Balanced Multiple Description Video Coding Using Optimal Partitioning of the DCT Coefficients*. IEEE Transactions On Circuits and Systems For Video Technology, vol. 15, no. 7, pages 928–934, July 2005.
- [M.B.Dissanayake 2010] S.T.Worrall M.B.Dissanayake D.V.S.X. De Silva and W.A.C. Fernando. *Error Resilience Technique for Multi-view Coding Using Redundant Disparity Vectors*. In IEEE International Conference On Multimedia and Expo, ICME, 2010, pages 1712–1717, 2010.
- [Milani 2008] S. Milani, L. Celetto and G. A. Mian. *An Accurate Low-Complexity Rate Control Algorithm Based on ρ -Domain*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, no. 2, pages 257–262, February 2008.
- [Milani 2010] S. Milani and G. Calvagno. *Multiple Description Distributed Video Coding Using Redundant Slices and Lossy Syndromes*. IEEE Signal Processing Letters, vol. 17, no. 1, pages 51–54, Jan. 2010.
- [N. Kamoonwatana 2010] D. Agrafiotis N. Kamoonwatana and C.N. Canagarajah. *Rate-Controlled Redundancy Multiple Description Coding For Video Transmission Over MIMO Systems*. In IEEE 17th International Conference on Image Processing (ICIP), pages 1269–272, Sept. 2010.
- [Norkin 2006] Andrey Norkin, Anil Aksay, Cagdas Bilen, Gozde Akar, Atanas Gotchev and Jaakko Astola. *Schemes for Multiple Description Coding of Stereoscopic Video*. In Multimedia Content Representation, Classification and Security, volume 4105 of *Lecture Notes in Computer Science*, pages 730–737. Springer Berlin / Heidelberg, 2006.

- [Ohm 1994] J.-R. Ohm. *Three-dimensional subband coding with motion compensation*. Image Processing, IEEE Transactions on, vol. 3, no. 5, pages 559–571, sep 1994.
- [Orchard 1997] M.T. Orchard, Y. Wang, V. Vaishampayan and A.R. Reibman. *Redundancy rate-distortion analysis of multiple description coding using pairwise correlating transforms*. In Image Processing, 1997. Proceedings., International Conference on, volume 1, pages 608–611 vol.1, oct 1997.
- [Ozarow 1980] L. Ozarow. *On a Source Coding Problem with Two Channels and Three Receivers*. The Bell System Technical Journal, vol. 59, no. 10, pages 1909–1921, Dec. 1980.
- [Padmanabhan 2003] V.N. Padmanabhan, H.J. Wang and P.A. Chou. *Resilient peer-to-peer streaming*. In Network Protocols, 2003. Proceedings. 11th IEEE International Conference on, pages 16–27, nov. 2003.
- [Peraldo 2010] L. Peraldo, E. Baccaglini, E. Magli, G. Olmo, R. Ansari and Y. Yao. *Slice-level rate-distortion optimized multiple description coding for H.264/AVC*. In IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pages 2330–2333, March 2010.
- [Pereira 2009] Fernando Pereira. *Distributed video coding: basics, main solutions and trends*. In Proceedings of the 2009 IEEE international conference on Multimedia and Expo, ICME'09, pages 1592–1595, Piscataway, NJ, USA, 2009. IEEE Press.
- [Puri 1999] R. Puri and K. Ramchandran. *Multiple description source coding using forward error correction codes*. In Signals, Systems, and Computers, 1999. Conference Record of the Thirty-Third Asilomar Conference on, volume 1, pages 342–346 vol.1, 1999.
- [Puri 2001] R. Puri, K. won Lee, K. Ramchandran and V. Bharghavan. *An Integrated Source Transcoding and Congestion Control Paradigm for Video Streaming in the Internet*. IEEE Transactions on Multimedia, vol. 3, pages 18–32, 2001.
- [Qu 2006] Qi Qu, Yong Pei and J.W. Modestino. *An Adaptive Motion-Based Unequal Error Protection Approach for Real-Time Video Transport Over Wireless IP Networks*. Multimedia, IEEE Transactions on, vol. 8, no. 5, pages 1033–1044, oct. 2006.

- [Radulovic 2010] I. Radulovic, P. Frossard, Y.-K. Wang, M. M. Hannuksela and A. Halapuro. *Multiple Description Video Coding with H.264/AVC Redundant Pictures*. IEEE Transactions On Circuits and Systems For Video Technology, vol. 20, no. 1, pages 144–148, 2010.
- [Rane 2008] S. Rane, P. Baccichet and B. Girod. *Systematic Lossy Error Protection of Video Signals*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 18, no. 10, pages 1347–1360, oct. 2008.
- [Reibman 1992] A.R. Reibman and B.G. Haskell. *Constraints on variable bit-rate video for ATM networks*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 2, no. 4, pages 361–372, Dec. 1992.
- [Reibman 2001] A. Reibman, H. Jafarkhani, Yao Wang and M. Orchard. *Multiple description video using rate-distortion splitting*. In Image Processing, 2001. Proceedings. 2001 International Conference on, volume 1, pages 978–981 vol.1, 2001.
- [Reibman 2002] A. R. Reibman, H. Jafarkhami, Y. Wang and M. T. Orchard. *Multiple-Description Video Coding Using Motion-Compensated Temporal Prediction*. IEEE Transactions On Circuits and Systems For Video Technology, vol. 12, no. 3, pages 193–204, March 2002.
- [Ribas-Corbera 2003] J. Ribas-Corbera, P.A. Chou and S.L. Regunathan. *A generalized hypothetical reference decoder for H.264/AVC*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pages 674–687, July 2003.
- [S. Lin 2001] Y. Wang S. Lin S. Mao and S. Panwar. *A Reference Picture Selection Scheme for video Transmission over Ad-HOC Networks using multiple paths*. In IEEE International Conference on Multimedia and Expo (ICME), pages 96–99, Aug. 2001.
- [Said 1996] A. Said and W.A. Pearlman. *A new, fast, and efficient image codec based on set partitioning in hierarchical trees*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 6, no. 3, pages 243–250, Jun. 1996.
- [Schmidt 2011] J.C. Schmidt and K. Rose. *Jointly Optimized Mode Decisions in Redundant Video Streaming*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 4, pages 513–518, april 2011.

- [Sehgal 2004] A. Sehgal, A. Jagmohan and N. Ahuja. *Wyner-Ziv coding of video: an error-resilient compression framework*. IEEE Transactions on Multimedia, vol. 6, no. 2, pages 249 – 258, April 2004.
- [Setton 2008] E. Setton, P. Baccichet and B. Girod. *Peer-to-Peer Live Multicast: A Video Perspective*. Proceedings of the IEEE, vol. 96, no. 1, pages 25 –38, jan. 2008.
- [Shapiro 1992] J.M. Shapiro. *An embedded wavelet hierarchical image coder*. In Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on, volume 4, pages 657 –660 vol.4, mar 1992.
- [Slepian 1973] D. Slepian and J. Wolf. *Noiseless coding of correlated information sources*. Information Theory, IEEE Transactions on, vol. 19, no. 4, pages 471 – 480, jul 1973.
- [Stockhammer 2004] T. Stockhammer, H. Jenkac and G. Kuhn. *Streaming video over variable bit-rate wireless channels*. IEEE Transactions on Multimedia, vol. 6, no. 2, pages 268 – 277, April 2004.
- [Su 2008] C.-C. Su, Homer H. Chen, Jason J. Yao and P. Huang. *H.264/AVC-based multiple description video coding using dynamic slice groups*. Image Commun., vol. 23, no. 9, pages 677–691, 2008.
- [Tang 2002] X. Tang and A. Zakhor. *Matching Pursuits Multiple Description Coding for Wireless Video*. IEEE Transactions On Circuits and Systems For Video Technology, vol. 12, no. 6, pages 566–575, June 2002.
- [Tillier 2007] C. Tillier, T. Petrisor and B. P.-Popescu. *A Motion-Compensated Over-complete Temporal Decomposition for Multiple Description Scalable Video Coding*. EURASIP Journal on Image and Video Processing, 2007.
- [Tillo 2008] T. Tillo, M. Grangetto and G. Olmo. *Redundancy Slice optimal Allocation for H.264 Multiple Description Coding*. IEEE Transactions On Circuits and Systems For Video Technology, vol. 18, no. 1, pages 59–70, 2008.
- [Tillo 2010] T. Tillo, E. Baccaglini and G. Olmo. *Multiple Descriptions Based on Multi-rate Coding for JPEG 2000 and H.264/AVC*. IEEE Transactions on Image Processing, vol. 19, no. 7, pages 1756 –1767, July 2010.

- [Tillo 2011] T. Tillo, E. Baccaglioni and G. Olmo. *Unequal Protection of Video Data According to Slice Relevance*. Image Processing, IEEE Transactions on, vol. 20, no. 6, pages 1572–1582, June 2011.
- [Tsai 2010] W.J. Tsai and J.Y. Chen. *Joint Temporal and Spatial Error Concealment for Multiple Description Video Coding*. IEEE Transactions on Circuits and Systems for Video Technology, vol. PP, no. 99, page 1, 2010.
- [Tsai 2012] Wen-Jiin Tsai and Hao-Yu You. *Multiple Description Video Coding Based on Hierarchical B Pictures Using Unequal Redundancy*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 22, no. 2, pages 309–320, feb. 2012.
- [Vaidyanathan 1990] P.P. Vaidyanathan. *Multirate digital filters, filter banks, polyphase networks, and applications: a tutorial*. Proceedings of the IEEE, vol. 78, no. 1, pages 56–93, Jan. 1990.
- [Vaishampayan 1993] V. A. Vaishampayan. *Design of Multiple Description Scalar Quantizers*. IEEE Transactions On Information Theory, vol. 39, no. 3, pages 821–834, May 1993.
- [Vaishampayan 1994] V.A. Vaishampayan and J. Domaszewicz. *Design of entropy-constrained multiple-description scalar quantizers*. Information Theory, IEEE Transactions on, vol. 40, no. 1, pages 245–250, jan 1994.
- [Vaishampayan 1998] V.A. Vaishampayan and J.-C. Batllo. *Asymptotic analysis of multiple description quantizers*. Information Theory, IEEE Transactions on, vol. 44, no. 1, pages 278–284, jan 1998.
- [Vaishampayan 1999] V.A. Vaishampayan and S. John. *Balanced interframe multiple description video compression*. In Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on, volume 3, pages 812–816 vol.3, 1999.
- [van der Schaar 2003] M. van der Schaar and D.S. Turaga. *Multiple description scalable coding using wavelet-based motion compensated temporal filtering*. In Proceedings. 2003 International Conference on Image Processing, 2003. ICIP 2003., volume 3, pages III – 489–92 vol.2, sept. 2003.
- [Verdicchio 2006] F. Verdicchio, A. Munteanu, A. I. Gavrilescu, J. Cornelis and P. Schelkens. *Embedded Multiple Description Coding of Video*. IEEE Transactions On Image Processing, vol. 15, no. 10, pages 3114–3130, Oct. 2006.

- [W. Jian 1999] A. Ortega W. Jian. *Multiple Description Coding via Polyphase Transform and Selective Quantization*. In SPIE Visual Communications and Image Processing Conference (VCIP 99), 1999.
- [Wang 2001] Y. Wang, M. T. Orchard, V. Vaishampayan and A. R. Reibman. *Multiple Description Coding Using Pairwise Correlating Transforms*. IEEE Transactions On Image Processing, vol. 10, no. 3, pages 351–366, March 2001.
- [Wang 2002] Y. Wang and S. Lin. *Error-Resilient Video Coding Using Multiple Description Motion Compensation*. IEEE Transactions On Circuits and Systems For Video Technology, vol. 12, no. 6, pages 438–452, June 2002.
- [Wang 2003] H. Wang and A. Ortega. *Robust Video Communication by Combining Scalability and Multiple Description Coding Techniques*. Proc. of SPIE Image and Video Communications and Processing, vol. 1, Jan. 2003.
- [Wang 2005a] D. Wang, N. Canagarajah and D. Bull. *S frame design for multiple description video coding*. In IEEE International Symposium on Circuits and Systems, ISCAS'2005, pages 2719 – 2722 Vol. 3, May 2005.
- [Wang 2005b] Y. Wang, A. R. Reibman and S. Lin. *Multiple Description Coding for Video Delivery*. Proceedings of the IEEE, vol. 93, no. 1, pages 57–70, Jan. 2005.
- [Wei 2007] W. Wei and A. Zakhor. *Multiple Tree Video Multicast Over Wireless Ad Hoc Networks*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 17, no. 1, pages 2 –15, jan. 2007.
- [Wei 2009] Wei Wei and A. Zakhor. *Interference Aware Multipath Selection for Video Streaming in Wireless Ad Hoc Networks*. Circuits and Systems for Video Technology, IEEE Transactions on, vol. 19, no. 2, pages 165 –178, Feb. 2009.
- [Wenger 2005] S. Wenger, M.M. Hannuksela, T. Stockhammer, M. Westerlund and D. Singer. *RTP Payload Format for H.264 Video*. RFC 3984, Feb. 2005.
- [Wyner 1976] A. Wyner and J. Ziv. *The rate-distortion function for source coding with side information at the decoder*. Information Theory, IEEE Transactions on, vol. 22, no. 1, pages 1 – 10, jan 1976.
- [Xiu 2009] Xiaoyu Xiu and Jie Liang. *View Interpolation Based Multiple Description Coding of Multiview Images*. In Asilomar,2009, pages 876–880, 2009.

-
- [Xua 2012] Yuanyuan Xua, Ce Zhua, Wenjun Zengb and Xue Jun Lia. *Multiple description coded video streaming in peer-to-peer networks*. Signal Processing: Image Communication, vol. 27, no. 5, pages 412 – 429, 2012.
- [Y.-C. Lee 2004] Y. Altunbasak Y.-C. Lee and R. M. Mersereau. *An Enhanced Multiple Description Video Coder With Drift Reduction*. IEEE Transactions On Circuits and Systems For Video Technology, vol. 14, no. 1, pages 122–127, Jan. 2004.
- [Yang 2000] Xuguang Yang and K. Ramchandran. *Optimal subband filter banks for multiple description coding*. Information Theory, IEEE Transactions on, vol. 46, no. 7, pages 2477 –2490, nov 2000.
- [Ye-Kui 2007] Wang Ye-Kui, M.M. Hannuksela, S. Pateux, A. Eleftheriadis and S. Wenger. *System and Transport Interface of SVC*. IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 9, pages 1149 –1163, Sept. 2007.
- [Zamir 1999] R. Zamir. *Gaussian codes and Shannon bounds for multiple descriptions*. Information Theory, IEEE Transactions on, vol. 45, no. 7, pages 2629 –2636, nov 1999.
- [Zha 2005] CoolStreaming/DONet: a data-driven overlay network for peer-to-peer live media streaming, volume 3, March 2005.
- [Zhu 2009] C. Zhu and M. Liu. *Multiple Description Video Coding Based on Hierarchical B Pictures*. IEEE Transactions On Circuits and Systems For Video Technology, vol. 19, no. 4, pages 511–521, April 2009.